

developerWorks_®

Hướng dẫn **SQL & XQuery cho IBM DB2**, Phần **1**: Nhập **môn Aroma và SQL**

Pat Moffatt (pmoffatt@ca.ibm.com)

20 05 2005

Giám đốc Chương trình quản lý thông tin, Sáng kiến học đường IBM

Bruce Creighton (bcreight@ca.ibm.com)

Chuyên viên lập kế hoạch các phần kỹ năng IBM

Jessica Cao

Chuyên viên phát triển các công cụ đào tạo IBM

Loạt bài hướng dẫn này giới thiệu cho bạn đọc về cơ sở dữ liệu Aroma để áp dụng kiến thức về ngôn ngữ truy vấn có cấu trúc (SQL) và ngôn ngữ vấn tin XML (XQuery) nhằm giải quyết các câu hỏi nghiệp vụ tiêu biểu. Kèm theo có cả các chỉ dẫn chi tiết để cài đặt DB2® và cơ sở dữ liệu Aroma. Hướng dẫn này là Phần 1 của loạt bài hướng dẫn SQL & XQuery cho IBM DB2.

Xem thêm bài trong loat bài này

Trước khi bạn bắt đầu

Về loạt **bài này**

Loạt bài hướng dẫn này dạy các chủ đề SQL từ cơ bản đến nâng cao và các chủ đề XQuery cơ bản. Nó cũng chỉ ra cách diễn đạt các câu hỏi nghiệp vụ thường gặp bằng các truy vấn cơ sở dữ liệu sử dụng các câu truy vấn SQL hay các câu truy vấn XQuery. Những người phát triển ứng dụng và những người quản trị cơ sở dữ liệu có thể sử dụng hướng dẫn này để nâng cao các kỹ năng truy vấn cơ sở dữ liệu của mình. Các thành viên tham gia sáng kiến học đường IBM có thể sử dụng loạt hướng dẫn này như một phần trong chương trình giảng dạy cơ sở dữ liệu của họ.

Tất cả các ví dụ trong bài này là dựa trên **Aroma**, một cơ sở dữ liệu mẫu chứa các dữ liệu doanh thu về các sản phẩm cà phê và chè đã bán trong các cửa hàng trên khắp nước Mỹ. Mỗi ví dụ gồm có ba phần:

- Một câu hỏi kinh doanh dưới dạng ngôn ngữ thường ngày
- Môt hoặc nhiều truy vấn ví du, được biểu diễn bằng SQL hay Xquery
- Một bảng các kết quả trả về từ cơ sở dữ liệu

Hướng dẫn này được thiết kế để cho phép các học viên học tập ngôn ngữ SQL và XQuery. Cũng như học bất cứ cái gì khác, việc bổ sung thêm các bài tập thực hành là rất quan trọng. Dữ liệu và các định nghĩa bảng sẽ tạo điều kiện thuận lợi cho các bài tập thực hành này.

Đối với các sinh viên sử dụng tài liệu này như một phần của khoá học, hãy nhận từ thầy hướng dẫn của mình các chỉ dẫn để kết nối tới cơ sở dữ liệu Aroma và tìm hiểu mọi sự khác biệt giữa hướng dẫn và cài đặt tại máy cục bộ của mình.

Hướng dẫn này được viết cho DB2 Express-C 9 trên nền UNIX®, Linux® và Windows® (trước đây gọi là Viper).

Về hướng dẫn **này**

Hướng dẫn này giới thiệu cho các học viên làm quen với cơ sở dữ liệu Aroma, đặc biệt là nhóm các bảng bán lẻ, các đinh nghĩa bảng, dữ liêu mẫu và quan hệ giữa các bảng.

Các bảng này được tạo ra để biểu diễn một môi trường mà tại đó:

- Có ít bảng
- Các tên cột sử dụng từ vựng của người phân tích
- Các cột phản ánh hoạt động kinh doanh với kích thước tự nhiên của nó

Đây là một kho dữ liệu nhỏ, tiêu biểu, sử dụng lược đồ hình sao. Một kiểu cơ sở dữ liệu khác là để xử lý giao dịch trực tuyến (OLTP). Các cơ sở dữ liệu xử lý giao dịch trực tuyến chứa đựng các giao dịch ở mức độ chi tiết nhất.

Hướng dẫn này cũng giới thiệu cơ bản về SQL và gợi ý một số các câu hỏi kinh doanh tiêu biểu, có thể sử dụng IBM DB2 9 để giải đáp.

Các yêu cầu hệ thống

Để sử dụng hướng dẫn này đầy đủ nhất, bạn cần phải cài đặt IBM DB2 9. Tải xuống DB2 Express-C, một phiên bản DB2 Express miễn phí dành cho cộng đồng nhưng vẫn có đủ các đặc tính dữ liệu cốt lõi giống như phiên bản DB2 Express và cung cấp một cơ sở vững chắc để xây dựng và triển khai các ứng dụng. Bạn cũng sẽ cần cài đặt cơ sở dữ liệu Aroma, có thể được tải xuống ở đây.

Cài đặt DB2 Express- C 9

Để cài đặt DB2 9, thực hiện theo các bước sau (hãy dùng các tuỳ chọn mặc định, trừ phi được chỉ dẫn khác đi):

1. Tìm tới thư mục chứa các tệp được tải xuống - db2exc_viper_WIN_x86.zip (chúng ta sẽ giả sử tệp được tải xuống chứa trong thư mục C:\ DownloadDirector).

- 2. Giải nén tệp.
- 3. Tìm tệp setup.exe trong thư mục đã được giải nén và nhấp đúp chuột trên tệp. Bảng khởi chạy cài đặt DB2 (DB2 Setup Launchpad) sẽ xuất hiện.
- 4. Bên trong Bảng khởi chạy cài đặt DB2, nhấp chuột vào nút Install a Product (cài đặt sản phẩm) trên bảng ở bên trái.
- 5. Tiếp tục nhấp chuột vào nút Install New trong DB2 Express.
- 6. Trong trình cài đặt, nhấp chuột vào nút **Next** để xem thỏa thuận cấp phép. Bạn sẽ phải **Accept** (chấp nhân) thỏa thuân để tiếp tuc và nhấn tiếp nút **Next**.
- 7. Chọn cài đặt **Typical** và nhấp nút **Next**.
- 8. Đánh dấu chọn vào "Install DB2 Universal Database Express Edition on this computer and save my settings in a response file" để xác nhận rằng bạn muốn "Cài đặt DB2 Universal Database Express Edition trên máy tính này và ghi lưu các giá trị cài đặt vào tệp", sau đó nhấp nút **Next**.
- 9. Xác định thư mục cài đặt và nhấp nút **Next**.
- 10. Trên màn hình "Configure DB2 instances" (Cấu hình thể hiện DB2), nhấp nút Next.
- 11. Trong màn hình "User Information" (thông tin người sử dụng), hãy sử dụng trình đơn kéo xuống để đặt Domain là **None -- use local user account** (Không -- sử dụng tài khoản người dùng cục bộ); đặt User name là **db2admin** và đặt một mật khẩu.

 (Bạn có một lựa chọn khác là sử dụng tài khoản **LocalSystem** (hệ thống cục bộ), nhưng nếu dùng tùy chọn này thì sẽ có một số hạn chế về sau. Nhấp nút **Help** để tìm hiểu thêm.)
- 12. Kiểm tra để chắc chắn rằng hộp chọn **Use the same user name and password for the remaining DB2 services** (Áp dụng cùng tên người dùng và mật khẩu cho các dịch vụ DB2 còn lại) đã được đánh dấu chọn và nhấp nút **Next**.
- 13. Nhấp nút **Next** để bỏ qua màn hình "Prepare the DB2 tools catalog" (Chuẩn bị danh sách các công cụ DB2).
- 14. Để đơn giản hóa việc cài đặt, không đánh dấu chọn vào mục **Set up your DB2 server to send notifications** (Cài đặt để máy chủ DB2 gửi các khai báo) và nhấp nút **Next**.
- 15. Kiểm tra để chắc chắn rằng hộp chọn **Enable operating system security checkbox** (Bật an ninh của hệ điều hành) đã được đánh dấu chọn. Để nguyên các giá trị mặc định cho nhóm quản tri DB2 là DB2ADMNS và nhóm người dùng DB2 là DB2USERS. Nhấp nút **Next**.
- 16. Nhấp nút **Finish** để bắt đầu cài đặt sản phẩm DB2 khi bạn nhìn thấy màn hình "Start copying files and create response file" (Bắt đầu sao chép các tệp và tạo tệp response).
- 17. Khi cài đặt xong, một cửa sổ với thông báo "Setup is complete" (Cài đặt đã xong) sẽ xuất hiện. Nhấp nút **Finish** để hoàn tất quá trình cài đặt.
- 18. Bây giờ, một cửa sổ có tiêu đề "DB2 First Steps" (Các bước đầu tiên với DB2) sẽ xuất hiện. Nhấp nút **Create profile**. Bộ duyệt Web của bạn sẽ mở một trang gọi là DB2 First Steps các bước đầu tiên với DB2. Hướng dẫn này có nhiều mối liên kết đến các thông tin hữu ích về DB2, nhưng ta sẽ không dùng đến trong hướng dẫn này. Đóng bộ duyệt Web và tiếp tục làm theo hướng dẫn.

Cài đặt cơ sở dữ liệu Aroma

- 1. Kiểm tra để chắc rằng DB2 Express-C 9 đã được cài đặt.
- 2. Khởi động cửa sổ lệnh từ Start menu ==> all Programs ==> IBM DB2 ==> Command Line Tools ==> Command Windows.

- 3. Cho chay create_aroma.bat trong cửa sổ lệnh.
 - Chú ý: Trước đó đừng quên sử dụng lênh cơ để di chuyển tới thư mục có chứa têp này.
- 4. Hai tệp, create.txt and report.txt, sẽ được tạo ra. Mở tệp report.txt và so sánh với các kết quả được mong đợi sẽ như dưới đây:

TABLE	RECORDS
aroma.class	9
aroma.deal	9
aroma.line_items	182
aroma.market	17
aroma.orders	27
aroma.period	821
aroma.product	59
aroma.promotion	194
aroma.sales	69941
aroma.store	18
aroma.supplier	9

Hỗ trợ XML

XML và cơ bản về XQuery sẽ được giới thiệu trong phần 7 của loạt bài này. Để hỗ trợ XML, một cơ sở dữ liệu aromadb đã được tạo ra bằng lệnh sau:

```
db2 create db aromadb using codeset utf-8 territory us
```

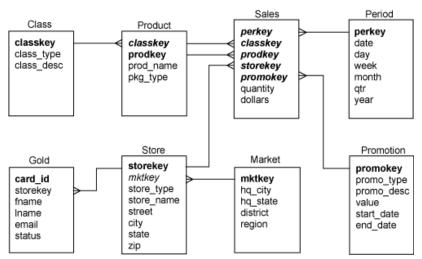
Hãy chú ý đến chỉ thị chọn UTF-8 là bộ mã ký tự cho cơ sở dữ liệu. Việc sử dụng các đặc tính XML bị hạn chế chỉ cho cơ sở dữ liệu được định nghĩa với bộ mã UTF-8 và chỉ có một phân hoạch cơ sở dữ liệu. Trong tương lai, khi tạo lập các cơ sở dữ liệu của riêng mình, bạn đừng quên chỉ thị chọn tham số này, nếu không cơ sở dữ liệu của bạn sẽ không hỗ trợ các đặc tính XML!

Nhóm các bảng của cơ sở dữ liệu Aroma

Hầu hết các ví dụ trong hướng dẫn này dựa trên dữ liệu từ cơ sở dữ liệu Aroma, chứa các dữ liệu gộp về doanh số bán lẻ hàng ngày tại các cửa hàng của công ty Cà phê và Chè Aroma. Cơ sở dữ liệu Aroma chứa các bảng ta sẽ dùng để nắm bắt tình hình bán hàng của công ty Cà phê và Chè Aroma. Có sáu bảng được sử dụng để tạo ra các nhóm khác nhau để hiểu rõ doanh thu. Các bảng đó là: Sales (Doanh thu), chứa các giao dịch hay còn gọi là các sự kiện; Period (Thời kỳ), Product (Sản phẩm), Store (Cửa hàng), Promotion Class (Loại khuyến mại), và Market (Thị trường). Các bảng này được gọi là nhóm các bảng bán lẻ.

Một tập hợp các bảng khác chứa dữ liệu để nắm được tình hình mua sắm của công ty Cà phê và Chè Aroma. Tập hợp các bảng này được gọi là nhóm các bảng mua sắm và có thể được sử dụng để hiểu rõ việc mua sắm từ các nhà cung cấp. Nhóm các bảng này không được sử dụng trong bất kỳ ví dụ nào trong hướng dẫn này. Nó có mặt trong Phụ lục chỉ để cung cấp một bức tranh đầy đủ hơn về công ty Cà phê và Chè Aroma.

Hình 1. Nhóm các bảng bán lẻ



Ký hiệu chân chim trong biểu đồ này biểu thị mối quan hệ một - nhiều giữa hai bảng. Ví dụ, mỗi giá trị phân biệt trong cột **Perkey** của bảng **Period** chỉ có thể xuất hiện một lần trong bảng đó nhưng lại xuất hiện nhiều lần trong bảng Sales. Các tên cột in chữ **đậm** là các cột khoá chính. Các tên cột in chữ *nghiêng* đậm là các cột khóa chính *v*à khoá ngoài.

Dữ liệu mẫu từ mỗi bảng được trình bày trong các phần sau để cho bạn có thể thấy được các mối quan hệ khóa chính - khóa ngoài làm việc như thế nào.

Khóa chính là khóa duy nhất và là một phần của định nghĩa của bảng. Một bảng không thể có nhiều hơn một khóa chính, và các cột thuộc khóa chính không thể nhận giá trị null. (Sau này, ta sẽ nói chi tiết hơn về giá trị null. Nếu giá trị là null, điều đó có nghĩa là có một chỗ để lưu giữ dữ liệu, nhưng không có gì ở đó - điều này khác với một khoảng trống). Khóa chính là tùy chọn và có thể được định nghĩa trong câu lệnh CREATE TABLE hoặc ALTER TABLE. Khoá chính rất quan trọng do một giá trị chỉ có thể xuất hiện một lần; sẽ luôn luôn lựa chọn đúng cùng một hàng. Có thể vài thành phố cùng tên; tuy nhiên, việc lựa chọn bản ghi dựa vào khóa chính sẽ bảo đảm luôn luôn chọn đúng cùng một hàng.

Một *khóa ngoài* sẽ được chỉ rõ trong định nghĩa ràng buộc tham chiếu. Một bảng có thể không có hay có nhiều khoá ngoài. Khóa ngoài phức hợp sẽ có giá trị null nếu một thành phần bất kỳ có giá trị null. Các khóa ngoài là tuỳ chọn và có thể được định nghĩa trong câu lệnh CREATE TABLE hoặc câu lệnh ALTER TABLE.

Cơ sở về **các** bảng **Aroma**

Các bảng và các cột được đặt tên bằng các thuật ngữ kinh doanh quen thuộc, làm cho chúng trở nên dễ hiểu và sử dụng. Một tập hợp các bảng được thiết kế tốt sẽ mang lại các lợi ích sau cho những người phát triển ứng dụng và những người dùng cuối:

- Các câu hỏi nghiệp vụ được biểu diễn dễ dàng bằng các truy vấn SQL
- Thực hiện các truy vấn nhanh và kết quả trả về nhất quán.

Bạn sẽ nhận thấy rằng cơ sở dữ liệu Aroma không ghi từng giao dịch đã xảy ra. Dữ liệu đã được gộp lại phản ánh tổng doanh số tại mỗi cửa hàng trong một ngày đã định, đối với một sản phẩm đã cho, hưởng một ưu đãi khuyến mại cụ thể.

Dữ liệu gộp là dạng tiêu biểu cho các kho dữ liệu. Ý tưởng là tổng hợp dữ liệu sao cho không có quá nhiều hàng. Bình thường, việc này sẽ do người quản trị cơ sở dữ liệu (DBA) hay chương trình máy tính thực hiện. Quá trình gộp dữ liêu thường được gọi là ETL (Trích ra, Biến đổi, Nap vào).

Các bảng chiều Period, Product, và Class

Bảng **Period**

Bảng sau đây hiển thị một vài hàng đầu tiên của bảng **Period** . Cột khóa chính là cột **Perkey**

Perkey	Date	Day	Week	Month	Qtr	Year
1	2004-01-01	TH	1	JAN	Q1_04	2004
2	2004-01-02	FR	1	JAN	Q1_04	2004
3	2004-01-03	SA	1	JAN	Q1_04	2004
4	2004-01-04	SU	2	JAN	Q1_04	2004
5	2004-01-05	МО	2	JAN	Q1_04	2004
6	2004-01-06	TU	2	JAN	Q1_04	2004

Các bảng Product và Class

Bảng sau đây hiển thị một vài hàng đầu tiên của bảng **Product**. Khóa chính là kết hợp hai giá trị của cột **Classkey** và của cột **Prodkey**:

Classkey	Prodkey	Prod_Name	Pkg_Type
1	0	Veracruzano	No pkg
1	1	Xalapa Lapa	No pkg
1	10	Colombiano	No pkg
1	11	Espresso XO	No pkg
1	12	La Antigua	No pkg
1	20	Lotta Latte	No pkg

Chú ý: Nhớ lại rằng khoá chính là một khóa duy nhất và là một phần của định nghĩa bảng. Bình thường, khóa chính là một cột trong bảng. Trong trường hợp này, khóa chính là phức hợp gồm hai cột. Điều này có nghĩa là chỉ có một kết hợp như thế trong một bảng.

Bảng sau đây hiển thi một vài hàng đầu tiên của bảng **Class** table.

		4
Classkey	Class_Type	Class_Desc

1	Bulk_beans	Bulk coffee products
2	Bulk_tea	Bulk tea products
3	Bulk_spice	Bulk spices
4	Pkg_coffee	Individually packaged coffee products
5	Pkg_tea	Individually packaged tea products
6	Pkg_spice	Individually packaged spice products

Các bảng Store, Market, và Promotion

Các bảng Store, Market, và Promotion chứa các thông tin mô tả mà những người phân tích dữ liệu sẽ sử dụng khi truy vấn cơ sở dữ liệu. Ví dụ, bảng Store chứa tên và địa chỉ các cửa hàng; bảng Product chứa thông tin về sản phẩm và cách đóng gói; bảng Period chứa các giá trị theo tháng, quý, và năm. Mọi bảng có một khóa chính gồm có một hoặc nhiều cột; mỗi hàng trong một bảng được nhận biết duy nhất bằng giá trị (hay các giá trị) khóa chính của nó.

Các bảng Store và Market

Bảng sau đây hiển thị một vài hàng đầu tiên của bảng **Store** (một số cột bị cắt bớt để vừa chiều rộng trang). Cột khóa chính là **Storekey**; **Mktkey** là khoá ngoài tham chiếu đến bảng **Market**.

Storekey	Mktkey	Store_Type	Store_Name	Street	City	State	Zip
1	14	Small	Roasters, Los Gatos	1234 University Ave	Los Gatos	CA	95032
2	14	Large	San Jose Roasting	5678 Bascom Ave	San Jose	CA	95156
3	14	Medium	Cupertino Coffee	987 DeAnza Blvd	Cupertino	CA	97865
4	3	Medium	Moulin Rouge	898 Main Street	New Orleans	LA	70125
5	10	Small	Moon Pennies	98675 University	Detroit	МІ	48209
6	9	Small	The Coffee Club	9865 Lakeshore Bl	Chicago	IL	06060
	•	•	•	•	•	•	•

Bảng sau đây hiển thị một vài hàng đầu tiên của bảng Market .

Mktkey	Hq_city	Hq_state	District	Region
1	Atlanta	GA	Atlanta	South
2	Miami	FL	Atlanta	South
3	New Orleans	LA	New Orleans	South
4	Houston	TX	New Orleans	South
5	New York	NY	New York	North

Bảng **Promotion**

Bảng sau đây hiển thị một vài hàng đầu tiên của bảng **Promotion**. Cột khóa chính là **Promokey**.

Promokey	Promo_Type	Promo_Desc	Value	Start_Date	End_Date
0	1	No Promotion	0.00	9999-01-01	9999-01-31
1	100	Aroma catalog coupon	1.00	2004-01-01	2004-01-31
2	100	Aroma catalog coupon	1.00	2004-02-01	2004-02-29
3	100	Aroma catalog coupon	1.00	2004-03-01	2004-03-31
4	100	Aroma catalog coupon	1.00	2004-04-01	2004-04-30
5	100	Aroma catalog coupon	1.00	2004-05-01	2004-05-31

Cột *Value* xác định giá trị tiền mặt chi cho khuyến mại đã tiến hành. Người sử dụng có thể đánh giá chi phí khuyến mại bằng tổng các giá trị này.

Bảng **Sales**

Bảng sau đây hiển thị 20 hàng đầu tiên của bảng Sales.

Perkey	Classkey	Prodkey	Storekey	Promokey	Quantity	Dollars	Comments
2	2	0	1	116	8	34.00	
2	4	12	1	116	9	60.75	
2	1	11	1	116	40	270.00	
2	2	30	1	116	16	34.00	
2	5	22	1	116	11	30.25	
2	1	30	1	116	30	187.50	
2	1	10	1	116	25	143.75	
2	4	10	2	0	12	87.00	
2	4	11	2	0	14	115.50	
2	4	22	2	0	18	58.50	
2	4	0	2	0	17	136.00	
2	5	0	2	0	13	74.75	
2	4	30	2	0	14	101.50	
2	2	10	2	0	18	63.00	
2	1	22	3	0	11	99.00	
2	6	46	3	0	6	36.00	
2	5	12	3	0	10	40.00	
2	1	11	3	0	36	279.00	
2	5	1	3	0	11	132.00	
2	5	10	3	0	12	48.00	
	·		•	•	•	•	*

Khóa chính được định nghĩa là phức hợp gồm năm giá trị từ các cột:

perkey, classkey, prodkey, storekey, promokey

Về bảng

Bảng Sales chứa dữ liệu có thể dễ dàng truy nhập thông qua các thuộc tính nghiệp vụ được định nghĩa trong bảng mà nó tham chiếu và nó lưu giữ một số lượng lớn các thông tin thống kê về các thuộc tính đó. Bảng Sales là bảng lớn nhất trong cơ sở dữ liệu này.

Cột cuối cùng, **Comments**, lưu giữ thông tin về ý kiến phản hồi của khách hàng, sử dụng định dạng dữ liêu XML. XML sẽ được giới thiêu trong phần 7 của loạt bài này.

Khóa chính gồm nhiều phần

Bảng **Sales** chứa một khóa chính nhiều phần. Mỗi cột trong số năm cột của khóa này là một khóa ngoài tham chiếu đến khóa chính của bảng khác:

perkey, classkey, prodkey, storekey, promokey

Khóa chính này liên kết các dữ liệu Sales với các bảng Period, Product, Store, và Promotion.

Bảng **Gold**

Bảng Gold chứa thông tin về các khách hàng "Thẻ Vàng" để cho phép họ được hưởng nhiều loại giảm giá khác nhau, các quà tặng miễn phí và nhắc nhở qua thư điện tử về các hoạt động khuyến mại. Bảng Gold sẽ do các học viên tạo ra khi đến phần 6 của loại bài này. Nó sẽ có các cột sau: số ID duy nhất nhận dạng thẻ, khóa chính của bảng Store, họ, tên, địa chỉ thư điện tử và trạng thái.

SQL là gì?

Ngôn ngữ truy vấn có cấu trúc

Ngôn ngữ truy vấn có cấu trúc (SQL) là ngôn ngữ chuẩn hóa để định nghĩa và xử lý dữ liệu trong một cơ sở dữ liệu quan hệ. Tất cả các hệ quản trị cơ sở dữ liệu quan hệ (RDBMS) đều hiểu được SQL.

Dưới đây là một số thông tin về SQL:

- SQL được đặc biệt tạo ra như một phần của lý thuyết quan hệ. Bạn cần một ngôn ngữ để trao đổi với cơ sở dữ liệu, để lấy ra dữ liệu, hoặc để sửa đổi dữ liệu.
- SQL là một ngôn ngữ không dẫn đường. Nói cách khác, với SQL, bạn không phải ra lệnh cho DB2 hãy tìm một bản ghi, hãy đọc con trỏ của nó và lần theo con trỏ đó tới hàng có liên quan trong bảng. Chỉ cần đơn giản nói DB2 phải làm gì và DB2 sẽ biết cách xử lý lệnh. Đơn giản thế thôi!
- Bạn có thể sử dụng SQL để sửa đổi diện mạo của một bảng mà không phải đưa nó ra ngoại tuyến (offline), có thể thêm các cột mới vào một bảng hiện có hay thêm các quy tắc nghiệp vụ vào định nghĩa của bảng.

Các câu hỏi hay gặp

Dưới đây là một mẫu các câu hỏi nghiệp vụ mà bạn có thể sử dụng SQL để giải đáp. Nhờ các câu trả lời, các quyết đinh kinh doanh trong yếu sẽ được thực hiện đúng lúc.

Các câu hỏi dễ

- Doanh số bán hàng tuần của mặt hàng cà phê hiệu Lotta Latte ở San Jose trong năm vừa qua là bao nhiêu?
- Doanh số bán trung bình hàng tháng của tất cả các sản phẩm cà phê ở bờ Tây trong mỗi tháng của năm vừa qua là bao nhiêu?

Các câu hỏi khó ở mức vừa phải

- Doanh số của Lotta Latte ở San Jose so với doanh số của nó ở Los Angeles và New York là như thế nào?
- Thị phần hàng tháng của Lotta Latte thay đổi thế nào trong hai năm qua trên tất cả các thị trường?
- Những nhà cung cấp nào đòi giá cao nhất cho các sản phẩm chè rời?
- Đợt khuyến mại thành công nhất trong tháng mười hai vừa qua ở California là gì?

Các câu hỏi nghiệp vụ rất khó nhưng có thể dùng DB2 9 để giải đáp.

- Tổng tích lũy doanh thu Lotta Latte theo suốt mỗi tháng của năm qua là bao nhiêu?
- Tỷ lệ doanh thu hàng tháng trên tổng doanh thu (tính theo phần trăm) của sản phẩm Lotta Latte trong cùng kỳ là bao nhiêu?
- Mười thành phố nào có doanh thu cà phê thấp nhất trong năm 2004 tính theo đô la và tính theo lượng bán ra?
- Những cửa hàng Aroma nào thuộc tóp 25 phần trăm xếp theo lợi tức bán hàng của quý I năm 2005? Những cửa hàng nào thuộc nhóm 50 phần trăm ở giữa và những cửa hàng nào rơi vào nhóm đáy 25 phần trăm còn lai?

Các kiểu dữ liệu

Kiểu dữ liệu quy định các giá trị có thể chấp nhận được đối với các hằng số, các cột, các biến chính, các hàm số, các biểu thức, và các thanh ghi đặc biệt. Dưới dạng sơ lược nhất, cũng phải định nghĩa dữ liệu hoặc là có kiểu số hoặc là có kiểu văn bản. Điều này là quan trọng bởi hai lý do chính. Đầu tiên, với các kiểu dữ liệu số, có thể áp dụng các hàm số toán học ví như cộng hay trừ các trường với nhau. Thứ hai, kích thước mục dữ liệu có thể định trước để bảo đảm cấp phát lượng không gian nhớ thích hợp. Đã bao lần bạn phải điền vào một mẫu đơn gì đó và không có đủ chỗ để viết mọi thứ được yêu cầu? Tuy nhiên, việc tăng kích thước tất cả các trường tới cực đại dẫn tới có quá nhiều khoảng trống trong cơ sở dữ liệu. Phần này sẽ mô tả các kiểu dữ liệu được đề cập trong các ví dụ. Nếu muốn biết danh sách toàn bộ và mô tả đầy đủ các kiểu dữ liệu khác, hãy tra cứu Hướng dẫn tham khảo SQL.

Các kiểu số

Tất cả các số có một dấu và một độ chính xác. Độ chính xác là số bit hay số chữ số, trừ dấu.

SMALLINT	A SMALLINT (small integer)(số nguyên nhỏ) là một số nguyên hai byte với độ chính xác 5 chữ số.
INTEGER	An INTEGER (large integer) (số nguyên lớn) là một số nguyên bốn byte với độ chính xác 10 chữ số.
BIGINT	A BIGINT (big integer) (số nguyên lớn hơn) là một số nguyên tám byte với độ chính xác 19 chữ số.

REAL	A REAL (single-precision floating-point number) (số dấu phảy động độ chính xác đơn) là biểu diễn gần đúng dài 32 bit của một số thực.
DOUBLE	A DOUBLE (double-precision floating-point number) (số dấu phảy động độ chính xác kép) là biểu diễn gần đúng dài 64 bit của một số thực. DOUBLE cũng được gọi là FLOAT.
DECIMAL(p,s)	A DECIMAL DECIMAL là một số phập phân. Vị trí của dấu chấm thập phân được xác định bởi độ chính xác (p) và thang độ (s) của số. Độ chính xác là tổng số chữ số và phải nhỏ hơn 32. Thang độ là số các chữ số trong phần phân số và luôn luôn nhỏ hơn hay bằng giá trị của độ chính xác. Kiểu thập phân mặc định có độ chính xác 5 và thang độ 0 nếu những tham số này không được chỉ rõ.

Chuỗi **ký** tự

Một chuỗi $k\acute{y}$ tự là một chuỗi các byte. Độ dài của chuỗi là số byte trong chuỗi. Nếu độ dài bằng không, giá tri này được gọi là chuỗi rỗng.

Chuỗi ký tự có độ dài cố định	CHAR(x) là một chuỗi ký tự có độ dài cố định. Thuộc tính độ dài x phải nằm trong khoảng từ 1 đến hết 254
Chuỗi ký tự có độ dài thay đổi	Các chuỗi ký tự có độ dài thay đổi gồm ba kiểu: VARCHAR, LONG VARCHAR, và CLOB. Kiểu VARCHAR(x) là các chuỗi có độ dài thay đổi, vì vậy một chuỗi có độ dài là 9 có thể được chèn vào trong VARCHAR(15) nhưng vẫn giữ nguyên độ dài chuỗi là 9. Xem mục các đối tượng lớn (LOBs) để biết chi tiết về CLOB.

Chuỗi đồ họa

Một chuỗi đồ họa là một chuỗi dữ liệu ký tự hai byte.

GRAPHIC(x) là một chuỗi có độ dài cố định. Thuộc tính độ dài x phải nằm trong khoảng từ 1 đến hết 254.
Các chuỗi đồ họa có độ dài thay đổi gồm ba kiểu : VARGRAPHIC, LONG VARGRAPHIC, và DBCLOB. Xem mục các đối tượng lớn (LOBs) để biết chi tiết về DBCLOB.

Chuỗi nhị **phân**

Một Chuỗi nhị *phân* là một chuỗi các byte. Nó thường được dùng để lưu giữ các dữ liệu không truyền thống, ví dụ như hình ảnh. Đối tượng lớn nhị phân (BLOB) là một chuỗi nhị phân. Xem mục các đối tương lớn (LOBs) để biết thêm thông tin.

Các giá tri ngày giờ

Các giá trị kiểu ngày giờ biểu diễn ngày tháng, thời giờ và thời ấn (là một chuỗi ký tự 14 ký số biểu diễn ngày tháng và thời giờ hợp lệ dưới dạng yyyyxxddhhmmss). Các giá trị kiểu ngày giờ có thể được sử dụng trong một số phép toán số học và phép toán chuỗi nhất định và tương thích với một số chuỗi ký tự nhất định; tuy nhiên, chúng không phải là chuỗi ký tự mà cũng không là các số.

Date	Một date là một giá trị có ba phần (năm, tháng và ngày).
Time	Một <i>time</i> là một giá trị có ba phần (giờ, phút và giây) chỉ rõ một thời điểm trong ngày với 24 giờ.
Timestamp	Một <i>timestamp</i> là một giá trị có bảy phần (năm, tháng, ngày, giờ, phút, giây và micrô-giây) chỉ rõ ngày tháng và thời giờ.

Giá trị null

Giá trị null là một giá trị đặc biệt, khác với tất cả các giá trị khác null. Nó có nghĩa là không có mặt bất kỳ một giá trị nào khác tại cột đó trên hàng đang xét. Giá trị null tồn tại với tất cả các kiểu dữ liêu.

XML

DB2 9 là máy chủ dữ liệu lai đầu tiên ở mức công nghiệp, phục vụ quản lý dữ liệu theo cả hai mô hình cấu trúc, mô hình quan hệ và mô hình thuần XML™. DB2 9 đã đưa ra một máy lưu trữ dữ liệu tối ưu hóa cho dữ liệu XML song song với máy lưu trữ dữ liệu quan hệ hiện có.

XML là ngôn ngữ tự mô tả theo nghĩa rằng ý nghĩa của dữ liệu được bao gồm trong bản ghi. Cột Comments sử dụng kiểu dữ liệu XML để lưu trữ các ý kiến của khách hàng. Xem phần 7 của loạt bài này để biết thêm chi tiết.

Các cách thực hiện SQL

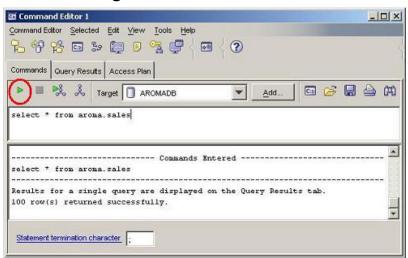
Muốn làm việc với DB2, bạn cần dùng một công cụ, chuyên biệt cho từng nền hệ thống, để nhập vấn tin SQL. Các công cụ này đã được gắn kèm khi bạn cài đặt DB2. Để đạt mục tiêu của loạt bài hướng dẫn này, hãy sử dụng Trung tâm điều khiển hoặc Bộ xử lý dòng lệnh (CLP).

Trung tâm điều khiển

Các lệnh SQL, các lệnh DB2 và các lệnh hệ điều hành có thể được chạy từ Trung tâm điều khiển. Mở Trung tâm điều khiển từ **Start menu ==> all Programs ==> IBM DB2 ==> General Administration Tools ==> Control Center**. Từ trình đơn ở đỉnh của giao diện, chọn **Tools ==> Command Editor**. Một cửa sổ soạn thảo lệnh sẽ bật lên và cho phép các bạn nhập vào các lệnh SQL.

Bạn có thế chạy các lệnh theo phương thức tương tác hoặc ghi lưu các lệnh thành một kịch bản lệnh. Để thực hiện các lệnh này, hãy nhấp chuột vào biểu tượng **Play** ở góc trên - trái của cửa sổ (được khoanh tròn trong hình 2 ở dưới).

Hình 2. Trung tâm điều khiển



Bộ xử **lý dòng** lệnh

Các lệnh cũng có thể được chạy từ Bộ xử lý dòng lệnh (CLP). Mở CLP từ **Start menu ==> all Programs ==> IBM DB2 ==> Command Line Tools ==> Command Line Processor**.

Hình 3. Bộ xử lý dòng lệnh

```
For general help, type: ?.
For command help, type: ?.
For command help, type: ? command, where command can be the first few keywords of a database manager command. For example: ? CATALOG DATABASE for help on the CATALOG DATABASE command ? CATALOG DATABASE for help on all of the CATALOG COMMAND.

To exit db2 interactive mode, type QUIT at the command prompt. Outside interactive mode, all commands must be prefixed with 'db2'. To list the current command option settings, type LIST COMMAND OPTIONS.

For more detailed help, refer to the Online Reference Manual.

db2 => connect to aromadb

Database connection Information

Database server = DB2/NT 9.0.0
SQL authorization ID = JCAO
Local database alias = AROMADB

db2 => select * from aroma.sales_
```

Xuất phát

Kết nối tới một cơ sở dữ liệu

Bạn cần kết nối tới một cơ sở dữ liệu trước khi có thể sử dụng các câu lệnh SQL để truy vấn hay xử lý dữ liêu. Lênh CONNECT kết hợp một kết nối cơ sở dữ liêu với một tên người sử dụng.

Để kết nối tới cơ sở dữ liệu aromadb, gỗ lệnh sau đây vào trong bộ xử lý dòng lệnh DB2:

```
CONNECT TO aromadb USER userid USING password
```

Nếu bạn đang sử dụng hướng dẫn này như một phần của khóa học, hãy hỏi thầy hướng dẫn để lấy userid và password dùng cho kết nối tới cơ sở dữ liêu.

Nếu máy không hỏi userid và password, hoặc nếu bạn sử dụng hướng dẫn này theo chế độ tự học và dùng máy tính riêng của mình thì chỉ cần gõ vào lệnh sau đây:

```
CONNECT TO aromadb
```

Thông điệp sau báo cho bạn biết rằng đã tạo kết nối thành công:

```
Database Connection Information
Database server
= DB2/NT 9.0.0
SQL authorization ID = USERID
Local database alias = AROMADB
```

Khi đã kết nối được, bạn có thể bắt đầu sử dụng cơ sở dữ liệu.

Trung tâm thông tin DB2

Ngoài bộ tài liệu tiêu chuẩn có sẵn kèm với DB2 9 còn có một sưu tập các thông tin DB2 có thể tìm kiếm trực tuyến, nhờ đó mà các tác vụ với DB2 9 được thực hiện dễ dàng hơn. Trung tâm thông tin DB2 cho phép người sử dụng có được sự trợ giúp về mọi vấn đề phát sinh.

Có thể tiếp cận Trung tâm thông tin DB2 qua bộ duyệt Web thông thường hay qua các tiện ích gắn sẵn trong một số giao diện người dùng đồ họa đi kèm với DB2 9. Hiện tại, các trình duyệt Microsoft Explorer (Phiên bản 5 và mới hơn) và Netscape Navigator (Phiên bản 6.1 và mới hơn) đều hỗ trợ.

Các nội dung của Trung tâm thông tin DB2 được cập nhật theo từng phiên bản phát hành của sản phẩm hay gói sửa lỗi. Bạn cần nắm chắc đang chạy phiên bản DB2 nào thì mới nhận được thông tin đúng. Trung tâm thông tin DB2 càng có ích hơn khi bạn sử dụng nó càng nhiều. Khi đã quen với các tùy chọn có sẵn, bạn có khả năng thu hẹp vùng tìm kiếm một cách hiệu quả hơn và sẽ nhận được giải đáp cho các câu hỏi của mình hiệu quả hơn nữa.

- Bạn có thể dễ dàng truy nhập Trung tâm thông tin DB2 từ **Start menu ==> all Programs ==> IBM DB2 ==> Information ==> Information Center**.
- Một cách khác, bạn có thể sử dụng url này: http://publib.boulder.ibm.com/infocenter/db2luw/ v9/index.jsp

Ví dụ

Trong phần 3 của loạt bài này bạn sẽ gặp một lỗi khi thực hiện một truy vấn có chứa kiểu dữ liệu không phù hợp. Ban sẽ nhân được thông báo lỗi dưới đây:

```
[IBM][CLI Driver][DB2/LINUX] SQL0401N Ki#u d# li#u c#a các toán h#ng trong
phép tính ">" không t##ng thích. SQLSTATE=42818
```

Để hiểu rõ hơn ý nghĩa của thông báo này, cần nghiên cứu các chi tiết của lỗi và tìm hiểu các thao tác sửa lỗi sẽ phải thực hiện.

- 1. Di chuyển tới Trung tâm thông tin DB2.
- 2. Nhập mã lỗi có trong thông báo lỗi (SQL0401N) vào trong hộp Search ở góc trên bên trái.
- 3. Nhấp chuột vào nút **GO**.
- 4. Có hai kết quả trả về được hiển thị. Kết quả đầu tiên là "100% SQL0401", nói rằng nó thích hợp 100%, trong khi kết quả thứ hai chỉ là thích hợp 8%. Chọn kết quả đầu tiên.
- 5. Đọc các kết quả tìm kiếm.

Metadata

Metadata (Siêu dữ liệu) có thể định nghĩa là dữ liệu về cấu trúc của bản thân cơ sở dữ liệu. Dữ liệu này bao gồm các thông tin như danh sách của các bảng, các chỉ mục, các cột và nhiều thứ khác có trong cơ sở dữ liệu.

Bạn sẽ thường thấy siêu dữ liệu được gọi bằng các tên khác, ví dụ như dữ liệu danh mục hay các bảng hệ thống. Siêu dữ liệu được DB2 9 thu thập tự động bất cứ khi nào người sử dụng thực hiện một hành động tạo, xóa, hay sửa đổi một số mục nào đó của cơ sở dữ liêu.

Các chuẩn lưu trữ dữ liệu được thiết lập cho SQL truyền thống không áp dụng với lưu trữ siêu dữ liệu. Mỗi nhà cung cấp phần mềm cơ sở dữ liệu có các phương pháp duy nhất của mình để lưu trữ và xử lý siêu dữ liệu. Biết cách lấy ra ra các thông tin này sẽ có tác dụng hỗ trợ rất lớn khi làm việc với cơ sở dữ liêu.

Tuy nhiên, Metadata được nhiều người quản trị cơ sở dữ liệu coi là có gì đó bí mật về bản chất. Do đó, nhiều người sử dụng sẽ không có quyền truy nhập tới sưu tập các siêu dữ liệu. Bạn sẽ phải tự tìm hiểu ở vị trí của chính mình xem có các quyền truy nhập siêu dữ liệu như thế nào.

Các bảng hệ thống

DB2 9 lưu trữ siêu dữ liệu trong gần 100 bảng hệ thống. Tất cả được chứa trong lược đồ SYSIBM. Lược đồ này được sinh ra tự động mỗi khi một cơ sở dữ liệu mới được tạo ra. Các bảng hệ thống cũng tuân thủ một quy ước đặt tên tiêu chuẩn. Tên mỗi bảng bắt đầu bằng SYS, ví dụ như SYSTABLES, SYSINDEXES, và SYSUSERAUTH. (Vì lý do này, tên các bảng do người dùng định nghĩa không được bắt đầu bằng SYS.)

Người sử dụng lấy ra siêu dữ liệu từ các bảng hệ thống theo cùng một cách giống như lấy dữ liệu từ các bảng thông thường. Các câu lệnh SQL hoạt động giống nhau trong cả hai trường hợp. Chỉ cần biết về cấu trúc của các bảng là đủ. Nếu được cấp ủy quyền thích hợp, một người sử dụng có thể nhận được danh sách của tất cả các bảng hệ thống chứa siêu dữ liệu bằng cách truy vấn cơ sở dữ liệu.

Ví dụ truy vấn

SELECT name FROM sysibm.systables WHERE name LIKE 'SYS%';

Kết quả

NAME
SYSATTRIBUTES
SYSBUFFERPOOLNODES
SYSBUFFERPOOLS

SYSTABLES	
···	
SYSXSROBJECTS	

Bạn có thể thực hiện *truy* vấn *trên hay không tùy* thuộc *vào* sự cấp *phép* của người quản *lý* hệ thống *và/*hoặc người quản trị cơ sở dữ liệu *cho* bạn

Nếu thực hiện truy vấn trên mà không có mệnh đề WHERE, bạn sẽ nhận được một danh sách tất cả các bảng trong cơ sở dữ liệu. Tập hợp kết quả sẽ được sắp thứ tự theo Creator- bộ tạo (lược đồ) và trong cùng một Creator sẽ sắp xếp tên bảng theo thứ tự chữ cái. Tên cột "Creator" đã giả thiết rằng lược đồ của mỗi bảng có cùng tên như Creator của nó.

Ví dụ truy vấn

SELECT name, creator FROM sysibm.systables;

Kết quả

NAME	CREATOR
CLASS	AROMA
	AROMA
SUPPLIER	AROMA
SYSATTRIBUTES	SYSIBM
	SYSIBM
SYSXMLSTATS	SYSIBM

Bằng điều kiện ràng buộc đối với tên lược đồ, bạn có thể giới hạn tập kết quả chỉ gồm các bảng là một phần của lược đồ đã cho.

Ví dụ truy vấn

SELECT name, creator FROM sysibm.systables WHERE creator='AROMA';

Kết quả

NAME	CREATOR
CLASS	AROMA
	AROMA
SUPPLIER	AROMA

Tương tự, bạn có thể liệt kê thông tin về các cột trong một bảng cụ thể bằng cách truy nhập tới bảng hệ thống sysibm.syscolumns.

Ví dụ truy vấn

SELECT * FROM sysibm.syscolumns WHERE tbname='CLASS' AND tbcreator='AROMA';

Kết quả

NAME	TBNAME	TBCREATOR	REMARKS	COLTYPE	
CLASS_DESC	CLASS	AROMA		CHAR	
CLASS_TYPE	CLASS	AROMA		CHAR	
CLASSKEY	CLASS	AROMA		INTEGER	

Hãy xem xét nội dung của bảng này bằng cách thực hiện lệnh. Tập hợp kết quả trả về sẽ có nhiều cột hơn nhiều so với trình bày trong bảng trên.

Tóm tắt

Hướng dẫn này đã giới thiệu nhóm các bảng bán lẻ và đề xuất một số câu hỏi nghiệp vụ tiêu biểu mà có thể dùng IBM DB2 9 để giải đáp. Các bảng này được tạo ra để biểu diễn một môi trường mà ở đó:

- Có ít bảng
- Các tên cột sử dụng từ vựng của người phân tích
- Các côt phản ánh hoạt đông kinh doanh với kích thước tư nhiên của nó.

Phần chủ yếu của loạt bài này là các ví dụ chi tiết nhằm chỉ ra cách viết các truy vấn SQL để giải đáp các câu hỏi nghiệp vụ như thế nào. Hầu hết các ví dụ này dựa trên nhóm bảng bán lẻ Aroma. Trừ vài trường hợp ngoại lệ, khi cần đưa thêm vào các bảng rất nhỏ nhằm minh họa những khái niêm đặc biệt.

Các kiểu dữ liệu thường được sử dụng nhiều nhất đã được mô tả trong hướng dẫn này:

Chuỗi ký tự	SMALLINT, INTEGER, BIGINT, REAL, DOUBLE, DECIMAL	
Chuỗi ký tự	Chuỗi ký tự có độ dài cố định (CHAR), Chuỗi ký tự có độ dài thay đổi (VARCHAR)	
Chuỗi đồ họa	Chuỗi đồ họa có độ dài cố định (GRAPHIC), Chuỗi đồ họa có độ dài thay đổi (VARGRAPHIC, LONG VARGRAPHIC, và DBCLOB)	
Chuỗi nhị phân	Đối tượng nhị phân lớn (BLOB)	
Các giá trị thời gian ngày tháng	Date, Time, Timestamp	
Giá trị Null	Null	
XML	Các tài liệu XML	

Các tải về

Mô tả	Tên	Kích thước
Cơ sở dữ liệu Aroma	Aroma_Data.zip	1MB

Tài nguyên

Học tập

- Đọc "Hướng dẫn đánh giá XML DB2" (developerWorks, 06.2006), một tài liệu hướng dẫn từng bước giới thiệu cho người đọc về máy chủ dữ liệu Viper DB2 trên nền Windows, sử dụng khả năng lưu trữ và tìm kiếm XML (SQL/XML, XQuery) sẵn có để hỗ trợ các ứng dụng thế hệ kế tiếp.
- Xem bài này và bài "Hứng khởi bắt đầu nhanh với DB2 Viper" (developerWorks, 03.2006).
- Tìm hiểu cách làm thế nào để "Truy vấn dữ liệu XML trong DB2 với XQuery" (developerWorks, 04.2006).
- Tìm hiểu cách làm thế nào để "Truy vấn dữ liệu XML trong DB2 với SQL" (developerWorks, 03.2006).
- Đọc Tập san các hệ thống của IBM và kỷ niệm 10 năm của XML.
- Xem Tài liệu tham khảo SQL, tập 1 để biết các thông tin bổ sung.
- Xem Tài liêu tham khảo SQL, tập 2 để biết các thông tin bổ sung.
- Xem Trung tâm thông tin DB2 để khắc phục sư cố.

Lấy sản phẩm và công nghệ

 Hiện nay bạn có thể sử dụng DB2 miễn phí. Hãy tải xuống DB2 Express-C, một phiên bản DB2 Express miễn phí dành cho cộng đồng nhưng vẫn có đủ các đặc tính dữ liệu cốt lõi giống như phiên bản DB2 Express và cung cấp một cơ sở vững chắc để xây dựng và triển khai các ứng dụng.

Thảo luân

Xem Diễn đàn hỗ trơ trực tuyến DB2 9.

Đôi nét về các tác giả

Pat Moffatt

Pat Moffatt là giám đốc Chương trình quản lý thông tin thuộc Sáng kiến học đường IBM. Thông qua chương trình sáng kiến học đường này, bà muốn làm ra sẵn những nguồn tài nguyên thích hợp phục vụ giảng dạy về quản lý thông tin để giúp ca´c giảng viên đại học tích hợp chủ đề phần mềm quản lý thông tin vào trong chương trình đào tạo của họ

Bruce Creighton

Bruce Creighton là chuyên viên lập kế hoạch các phần kỹ năng trong khoa Phát triển và Lập kế hoạch đào tạo Quản lý thông tin. Với vai trò này, ông lập kế hoạch đầu tư về nội dung đào tạo và cân bằng đầu tư giữa các lĩnh vực mà IBM có thu lợi nhuận với các lĩnh vực mà tại đó yêu cầu về phát triển các kỹ năng là quan trọng, cần cung cấp đào tạo miễn phí

Jessica Cao

Jessica Cao là một sinh viên ngành Nghệ thuật - Khoa học và ngành Khoa học máy tính tại Trường đại học McMaster. Chị sẽ tốt nghiệp văn bằng kép hạng danh dự vào tháng Tư 2009. Jessica đang làm việc trong chương trình hỗ trợ và lập kế hoạch kênh kỹ năng quản lý thông tin DB2 của phòng thí nghiệm IBM Toronto, nhằm chuyên chú sâu hơn vào lập trình, biên tập và kỹ năng viết

© Copyright IBM Corporation 2005

(www.ibm.com/legal/copytrade.shtml)

Nhẫn hiệu đăng ký

(www.ibm.com/developerworks/vn/ibm/trademarks/)