

Buổi 1

Tổng quan về CSDL

I. Các khái niệm về CSDL

1. Dữ liệu và cơ sở dữ liệu

Dữ liệu (*data*) là thông tin của 1 đối tượng (người, vật, sự việc, khái niệm...) được lưu trữ trên máy tính.

Dữ liệu có thể được mô tả dưới nhiều dạng khác nhau (kí tự, hình ảnh, âm thanh...). Mỗi cách mô tả dữ liệu sẽ gắn với một ngữ nghĩa nào đó. Tuy nhiên, dữ liệu về cùng 1 đối tượng có thể khác nhau tùy vào mục đích quản lí.

Ví dụ: Với đối tượng Sinh viên:

- Để quản lí điểm, cần có dữ liệu Mã số SV, Tên, Điểm môn 1, Điểm môn 2...
- Để quản lí lí lịch, cần có dữ liệu Tên, Ngày sinh, Quê quán...

Khi lượng dữ liệu lớn, cần tổ chức quản lí dữ liệu để dễ dàng lưu trữ, truy xuất hoặc cập nhật. Với cách lưu dữ liệu vào file, có một số nhược điểm sau:

- Dễ dẫn đến dư thừa và mâu thuẫn dữ liệu.
- Truy xuất khó khăn.
- Dữ liệu được lưu trữ rời rạc, không có mối liên hệ với nhau.
- Độ an toàn và bảo mật kém

➔ Do đó, chúng ta sẽ quản lí dữ liệu dưới dạng cơ sở dữ liệu.

Cơ sở dữ liệu¹ (*database*) là tập hợp dữ liệu được tổ chức có cấu trúc liên quan với nhau và được lưu trữ trong máy tính.

CSDL được thiết kế, xây dựng cho phép người dùng dễ dàng lưu trữ dữ liệu, truy xuất thông tin hoặc cập nhật dữ liệu.

2. Mô hình dữ liệu quan hệ

Có nhiều mô hình dữ liệu (*data model*), ví dụ như mô hình dữ liệu phẳng, mô hình dữ liệu phân cấp, mô hình dữ liệu mạng, mô hình dữ liệu quan hệ, mô hình dữ liệu hướng đối tượng...

Mô hình dữ liệu quan hệ (*Relational Data Model*) là mô hình dữ liệu phổ biến nhất. Trong mô hình này, dữ liệu được tổ chức dưới dạng bảng với các dòng và các cột².

3. Ngôn ngữ SQL và hệ quản trị CSDL

Để lưu trữ, truy xuất và cập nhật dữ liệu, chúng ta sử dụng ngôn ngữ SQL³.

Để quản lí cũng như tương tác với CSDL bằng SQL, cần có 1 hệ quản trị CSDL⁴ (*Database Management System*). Có nhiều hệ quản trị CSDL tương ứng với các mô hình dữ liệu khác nhau:

- Microsoft SQL Server, Microsoft Access, MySQL, Oracle là các HQT CSDL cho mô hình dữ liệu quan hệ.

¹ Viết tắt là CSDL (tiếng Việt) hoặc DB (tiếng Anh)

² Xem phần I.4.

³ Structured Query Language – Ngôn ngữ truy vấn có cấu trúc

⁴ Viết tắt là HQT CSDL (tiếng Việt) hoặc DBMS (tiếng Anh)

- IMS của hãng IBM là HQT CSDL cho mô hình dữ liệu phân cấp.
- IDMS là HQT CSDL cho mô hình dữ liệu mạng.

4. Bảng và thuộc tính

Một CSDL có thể có nhiều bảng (*table*). Mỗi bảng lưu thông tin về các đối tượng cùng loại và được đặt tên riêng, không trùng nhau (thường là đặt tên theo loại đối tượng đó),

Trong bảng có thể gồm nhiều dòng (*row*) và nhiều cột (*column*). Mỗi dòng mô tả 1 bộ dữ liệu⁵, gọi là các bản ghi (*record*), mỗi cột mô tả 1 thuộc tính⁶, gọi là các trường dữ liệu (*field*).

Tên bảng và tên thuộc tính là danh từ, không có dấu tiếng Việt, không có khoảng trắng. **Khuyến cáo** đặt tên bảng bằng chữ cái và dấu gạch dưới.

Ví dụ: Bảng KHACHHANG lưu thông tin các khách hàng của 1 cửa hàng.

MAKH	TENKH	DIACHI	DT	EMAIL
KH01	Nguyễn Thị Bé	Tân Bình	8457895	bnt@yahoo.com
KH02	Lê Hoàng Nam	Bình Chánh	9878987	bnt@yahoo.com
KH03	Trần Thị Chiêu	Tân Bình	8457895	
KH04	Mai Thị Quế Anh	Bình Chánh		
KH05	Lê Văn Sang	Quận 10		sanglv@hcm.vnn.vn
KH06	Trần Hoàng Khải	Tân Bình	8457897	

Trong đó:

- Bảng KHACHHANG lưu thông tin các khách hàng, tức là các đối tượng cùng loại.
- Mỗi dòng là 1 bộ dữ liệu đầy đủ thuộc tính của 1 khách hàng.
- Mỗi cột là tất cả các giá trị có thể có của 1 thuộc tính.

5. Khóa (key)

a) Khóa chính

Khóa chính (*primary key*) dùng để định danh 1 dòng dữ liệu, và để phân biệt tất cả các dòng dữ liệu trong 1 bảng.

Mỗi bảng có đúng 1 khóa chính, và khóa chính có thể bao gồm 1 hoặc nhiều thuộc tính.

Khóa chính phải đảm bảo các điều kiện sau:

- Không được phép mang giá trị NULL.
- Giá trị của khóa chính không được phép trùng nhau ở tất cả các dòng dữ liệu.
- Khóa chính nên mang tính chất định danh, và tránh sự thay đổi giá trị sau này.

Ví dụ: Với CSDL dân cư VN, bảng NGUOIDAN sẽ có khóa chính là số CMND vì nó mang tính chất định danh, không thể trùng nhau và khó có thể thay đổi sau này. Số ĐT hoặc email cũng không thể trùng nhau nhưng có thể thay đổi sau này và không mang tính chất định danh, không nên chọn 2 thuộc tính này làm khóa chính. Họ tên có thể trùng nhau nên không thể là khóa chính.

b) Khóa ngoại

Khóa ngoại⁷ (*foreign key*) dùng để tạo mối quan hệ giữa các bảng. Khóa ngoại trỏ tới khóa chính của 1 bảng khác.

Mỗi bảng có thể có nhiều khóa ngoại. Khóa ngoại có thể bao gồm 1 hoặc nhiều thuộc tính.

⁵ Tuple

⁶ Attribute

⁷ Khóa ngoại sẽ được giải thích kĩ hơn trong buổi 2

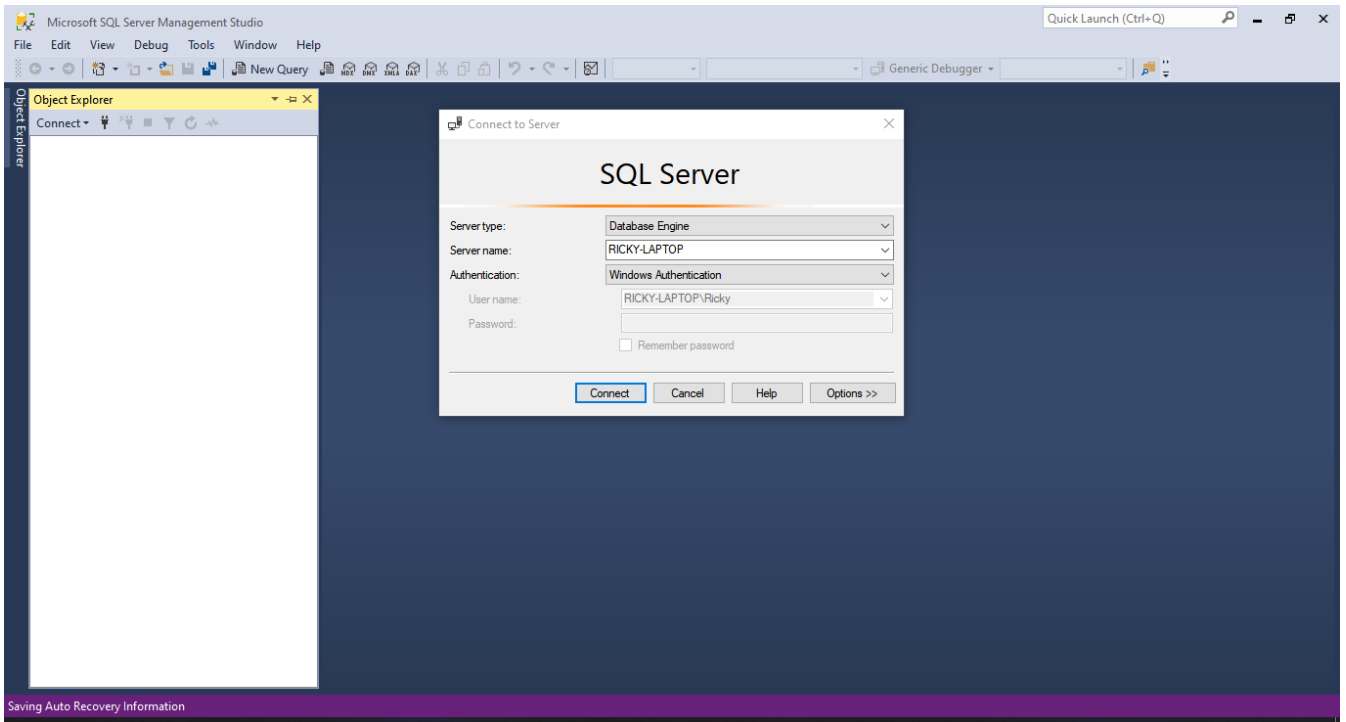
II. Các công cụ, phần mềm

Trong khuôn khổ môn học này, chúng ta sử dụng HQT CSDL Microsoft SQL Server⁸.

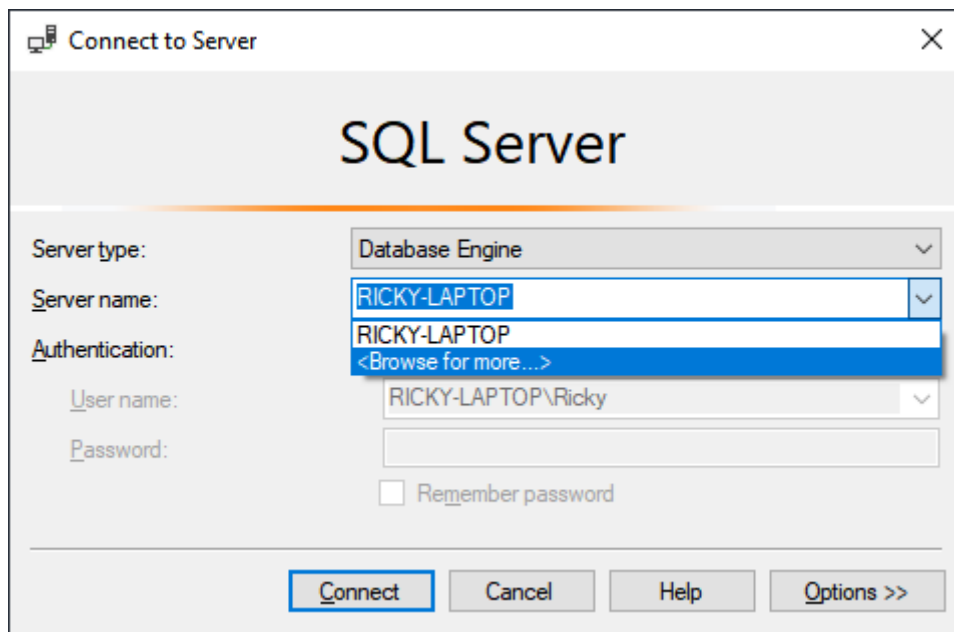
Địa chỉ tải SQL Server chính thức từ Microsoft (miễn phí với phiên bản Developer):
<https://www.microsoft.com/en-us/sql-server/sql-server-downloads>.

Với các phiên bản SQL Server gần đây, sau khi cài đặt chúng ta cần phải tải thêm 1 phần mềm có tên là SQL Server Management Studio (SSMS) nếu như máy tính chưa có.

Giao diện khi mở SSMS⁹ như sau:



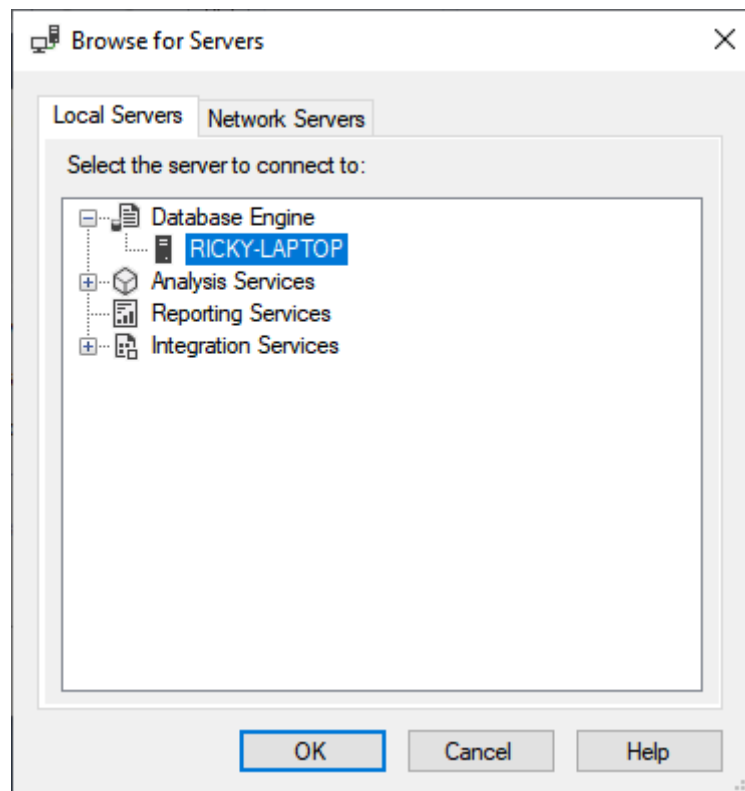
Để sử dụng được SQL Server, bước đầu tiên cần phải đăng nhập vào máy chủ ảo do SQL Server tạo ra. Mục *Server Name* mặc định là tên máy tính hiện tại. Nếu không hiển thị Server Name thì nhấn vào menu xổ xuống chọn *<Browse for more...>*:



Trong cửa sổ hiện ra, mở mục *Database Engine*, chọn tên máy tính và nhấn OK:

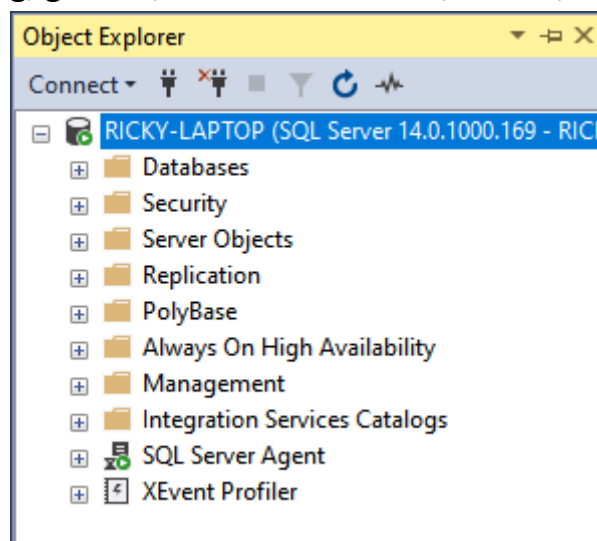
⁸ Gọi tắt là SQL Server hoặc MSSQL

⁹ Để cho dễ hiểu, kể từ nay, khi nói "mở SQL Server" tức là "mở SSMS"

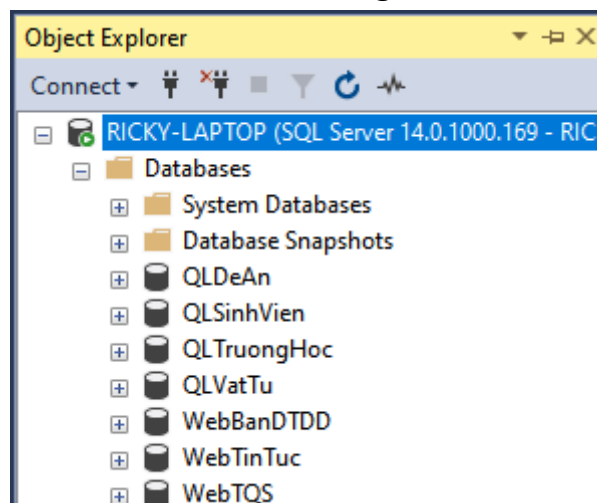


Cuối cùng, nhấn nút *Connect* để kết nối vào máy chủ ảo.

Sau khi kết nối thành công, giao diện SQL Server sẽ hiện ra mục *Object Explorer*¹⁰ như sau:



Mục *Databases* chứa tất cả các CSDL hiện có trong SQL Server:

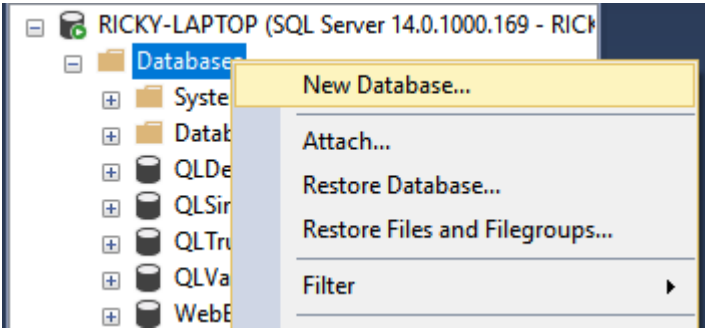


Lưu ý: Không xóa bất kì mục nào có sẵn của SQL Server.

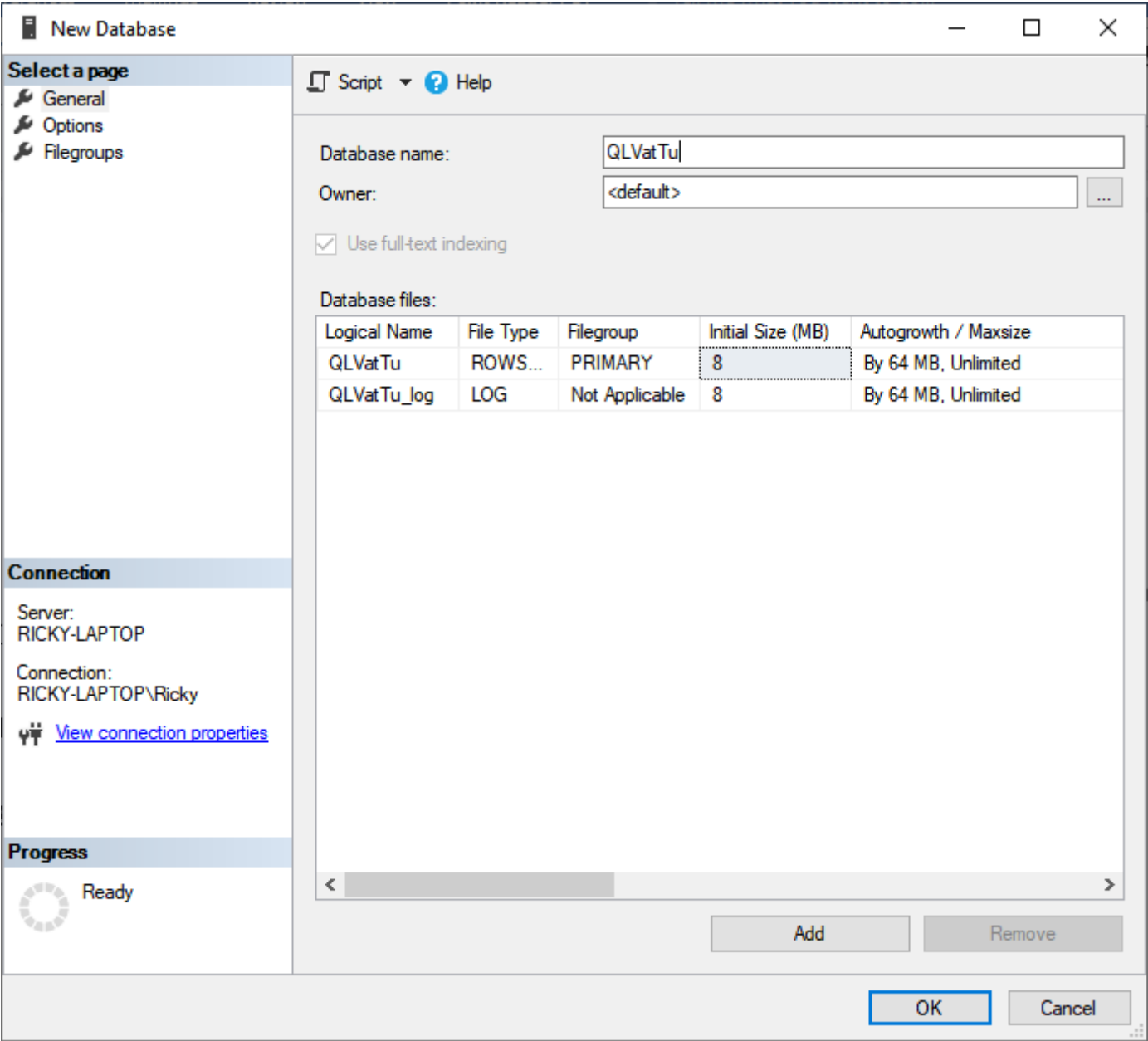
¹⁰ Nếu không hiển thị mục này, vào menu View chọn Object Explorer (hoặc nhấn phím tắt F8)

III. Tạo CSDL với Microsoft SQL Server

Để tạo 1 CSDL, click chuột phải vào mục *Databases* và chọn *New Database...*



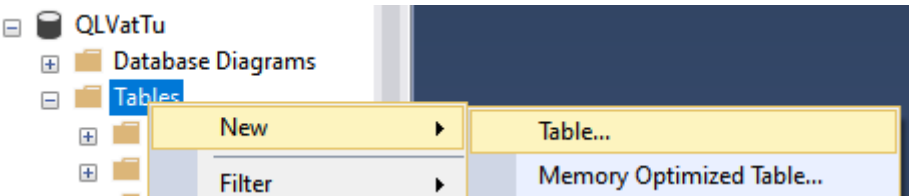
Trong cửa sổ hiện ra, điền tên CSDL vào mục *Database name* và nhấn OK. **Khuyến cáo** đặt tên CSDL, tên bảng và tên thuộc tính không có dấu tiếng Việt, không có khoảng trắng:



1. Tạo cấu trúc bảng

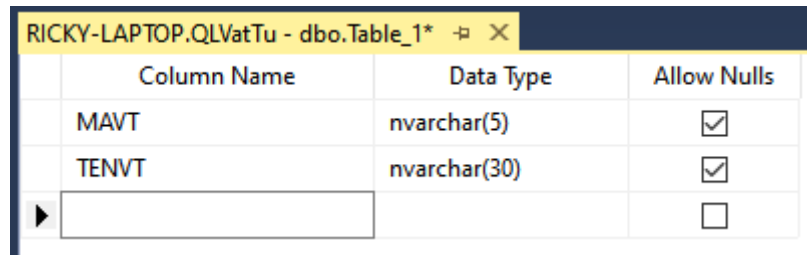
Nhấn để mở danh sách các mục trong CSDL. Mục *Tables* chứa danh sách các bảng.

Để tạo bảng mới, click chuột phải vào mục *Tables* và chọn *New* → *Table...*



Giao diện tạo bảng hiện ra, gồm có 3 mục:

- *Column Name* là tên thuộc tính/tên cột. Khuyến cáo đặt tên thuộc tính
- *Data Type* là kiểu dữ liệu.
- *Allow Nulls* quy định thuộc tính đó có được nhận giá trị NULL hay không.

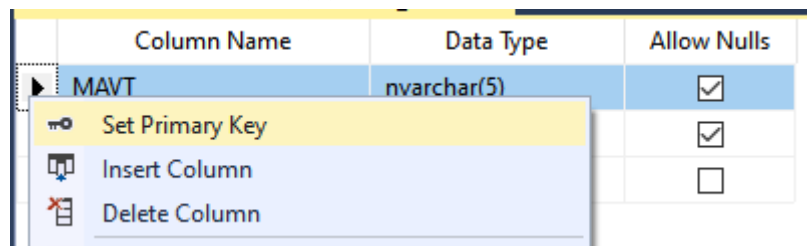


	Column Name	Data Type	Allow Nulls
	MAVT	nvarchar(5)	<input checked="" type="checkbox"/>
	TENVT	nvarchar(30)	<input checked="" type="checkbox"/>
▶			<input type="checkbox"/>

Các kiểu dữ liệu chính trong SQL Server:

- `varchar(x)`: Kiểu chuỗi không dấu. x cho biết độ dài chuỗi (x là 1 con số hoặc x là MAX).
- `nvarchar(x)`: Kiểu chuỗi có dấu tiếng Việt.
- `int`: Kiểu số nguyên.
- `float`: Kiểu số thực.
- `bit`: Kiểu logic (1 tương đương với True, 0 tương đương với False).
- `date`: Ngày theo định dạng yyyy-mm-dd.
- `datetime`: Ngày giờ theo định dạng yyyy-mm-dd hh:mm:ss.

Sau khi đã tạo xong cấu trúc bảng, trước khi lưu bảng, cần chọn khóa chính bằng cách click chuột phải vào đầu dòng chứa thuộc tính muốn đặt làm khóa chính và chọn *Set Primary Key* (nếu khóa chính gồm nhiều thuộc tính thì dùng phím *Ctrl* hoặc *Shift* để chọn nhiều thuộc tính):



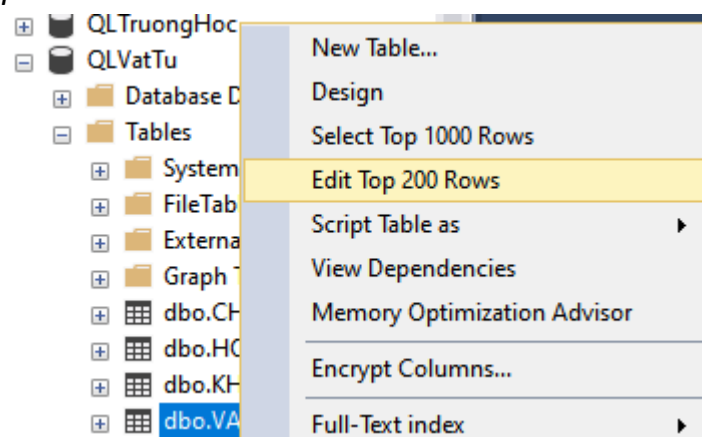
	Column Name	Data Type	Allow Nulls
▶	MAVT	nvarchar(5)	<input checked="" type="checkbox"/>
			<input checked="" type="checkbox"/>
			<input type="checkbox"/>

- Set Primary Key
- Insert Column
- Delete Column

Để lưu bảng, nhấn phím tắt *Ctrl + S*, sau đó đặt tên bảng ở cửa sổ hiện ra.

2. Thêm dữ liệu vào bảng

Sau khi đã tạo xong cấu trúc bảng, cần thêm dữ liệu cho các bảng bằng cách click chuột phải vào bảng và chọn *Edit Top 200 Rows*:

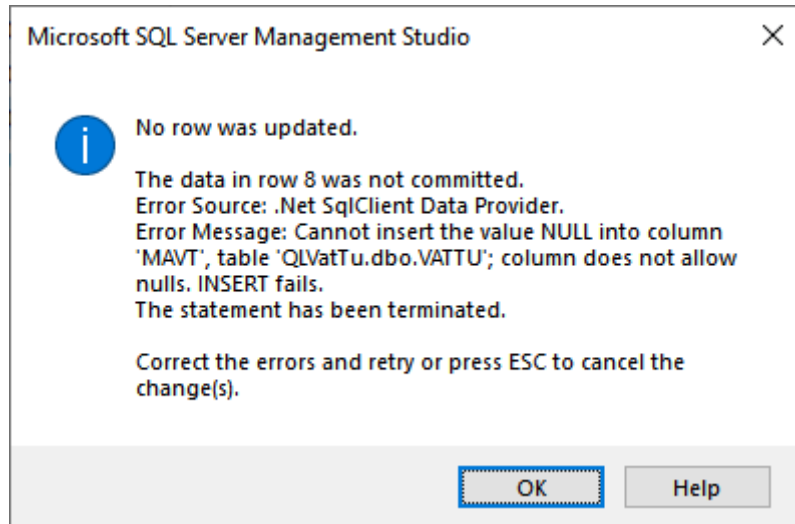


Sau khi nhập dữ liệu xong cho 1 dòng, nhấn *Enter* để lưu lại dòng đó, hoặc nhấn *Esc* để hủy bỏ dòng đó.

3. Một số lỗi khi thêm dữ liệu vào bảng

Các lỗi thường gặp:

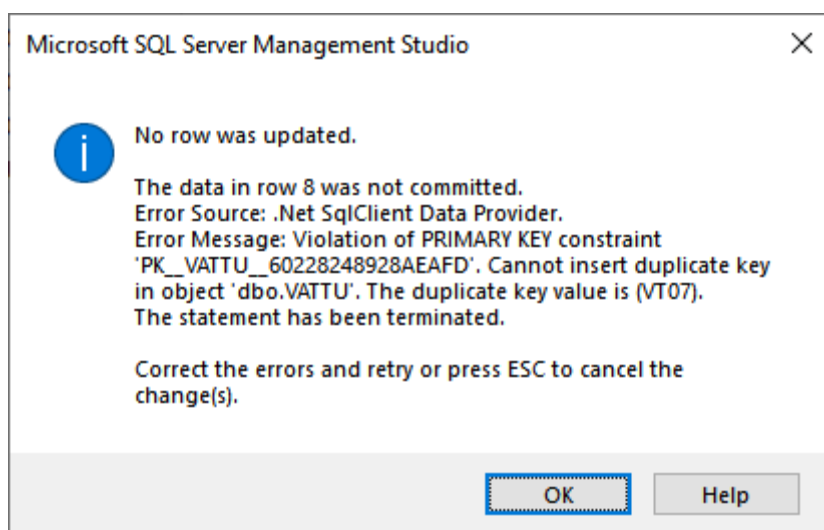
a) *Cannot insert the value NULL into column 'X', table 'Y'; column does not allow nulls:*



Lỗi này xảy ra khi thực hiện thêm giá trị NULL vào 1 cột không cho phép NULL (Mục Allow Nulls không được chọn). **Lưu ý:** Khóa chính của bảng không được phép mang giá trị NULL.

Cách khắc phục: Bổ sung dữ liệu cho cột đó.

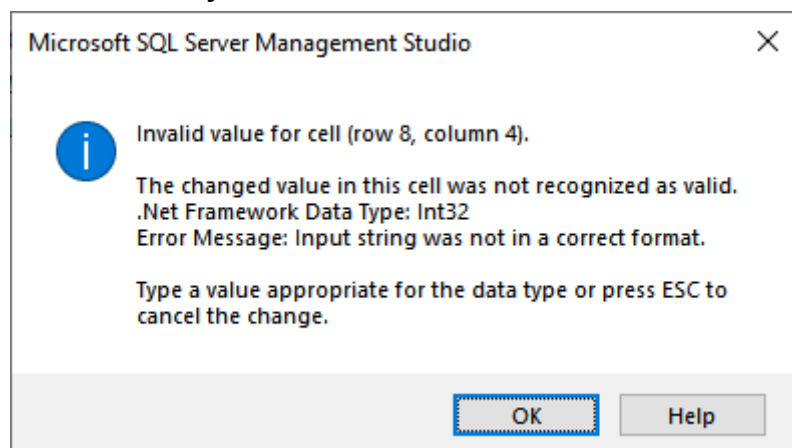
b) *Violation of PRIMARY KEY constraint 'X'. Cannot insert duplicate key in object 'Y'. The duplicate value is (Z):*



Lỗi này xảy ra khi khóa chính của dòng muốn thêm trùng với khóa chính của dòng khác. **Lưu ý:** Giá trị của khóa chính không được phép trùng nhau ở tất cả các dòng dữ liệu.

Cách khắc phục: Thay đổi giá trị của khóa chính.

c) *Input string was not in a correct format:*



Lỗi này xảy ra khi dữ liệu của 1 cột không đúng kiểu dữ liệu hoặc không đúng định dạng của cột đó. Ví dụ: Cột có kiểu dữ liệu là int nhưng khi nhập có chữ cái, hoặc nhập sai định dạng ngày...

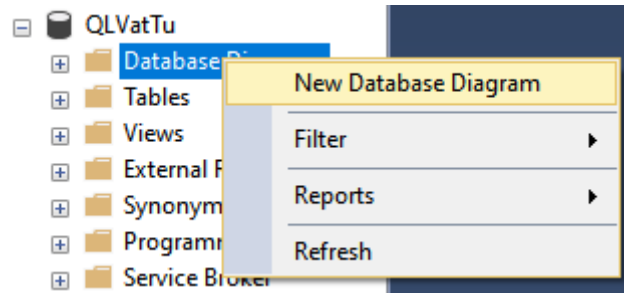
Cách khắc phục: Nhập lại dữ liệu đúng kiểu và đúng định dạng.

IV. Biểu diễn CSDL quan hệ

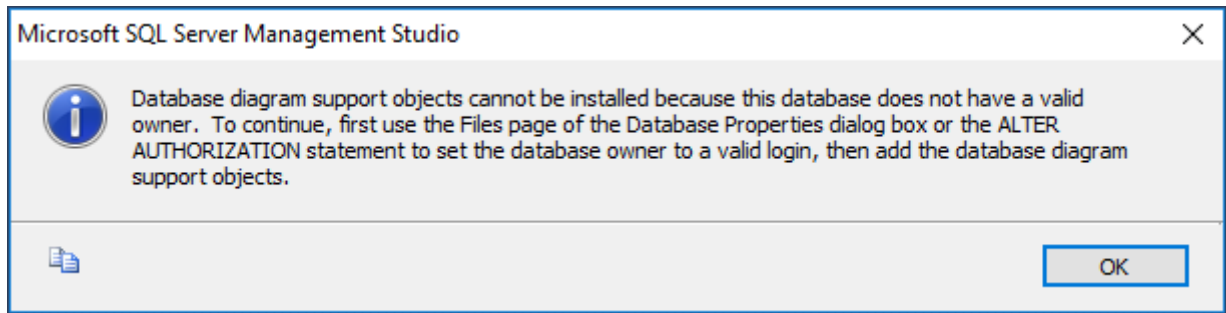
1. Lược đồ CSDL

Lược đồ CSDL (*Relationship Diagram*) cho biết cấu trúc của từng bảng (danh sách thuộc tính và kiểu dữ liệu) cũng như mối quan hệ giữa các bảng (khóa ngoại¹¹).

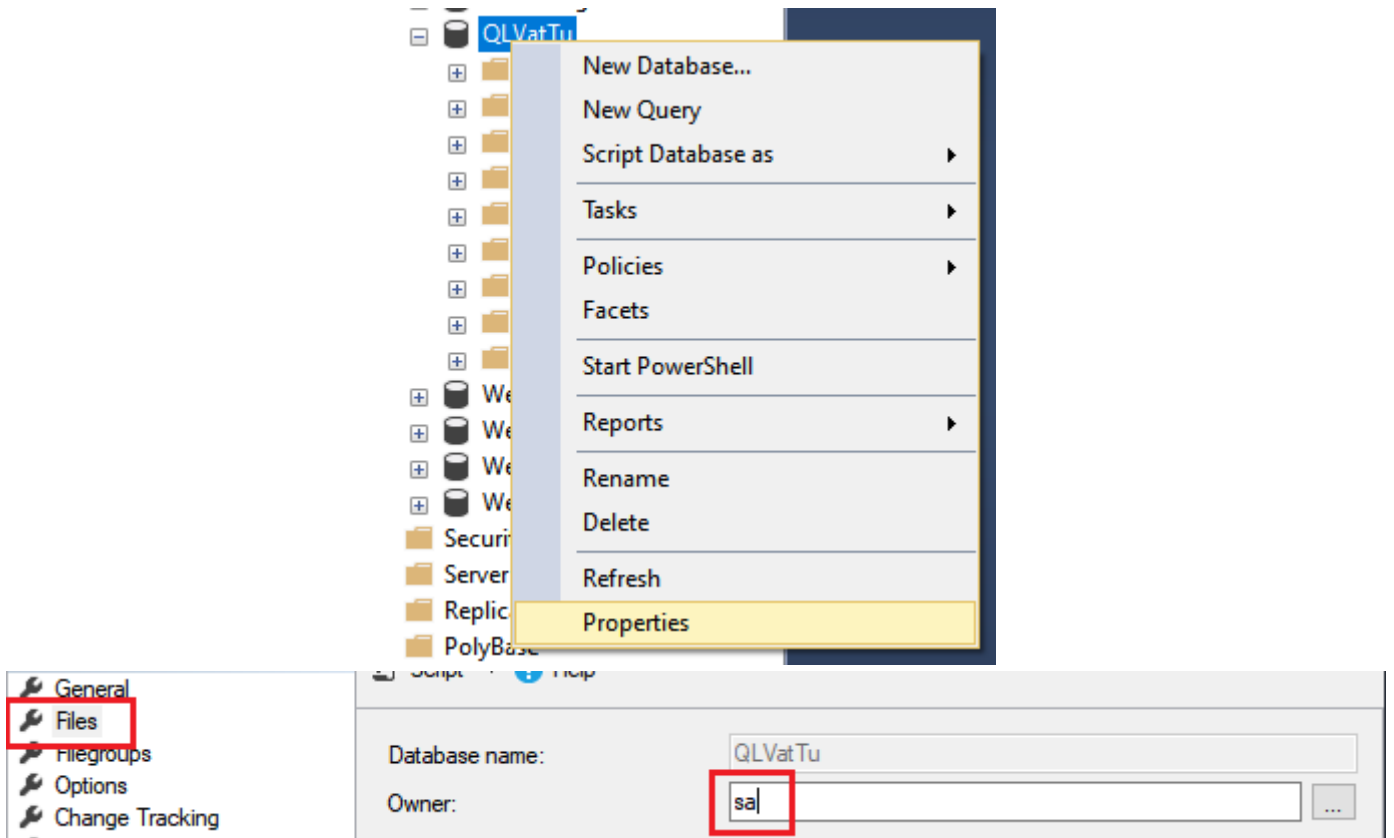
Để tạo lược đồ CSDL, click chuột phải vào *Database Diagrams*, chọn *New Database Diagram*:



Nếu gặp thông báo lỗi như sau nghĩa là CSDL đang thuộc về 1 tài khoản không hợp lệ:



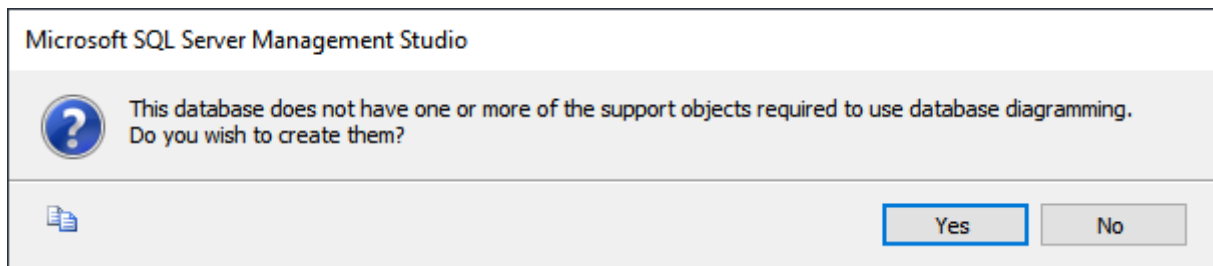
Cách khắc phục: click chuột phải vào tên CSDL, chọn *Properties*. Ở menu bên trái chọn *Files* và điền vào mục *Owner* giá trị *sa*¹², sau đó nhấn OK:



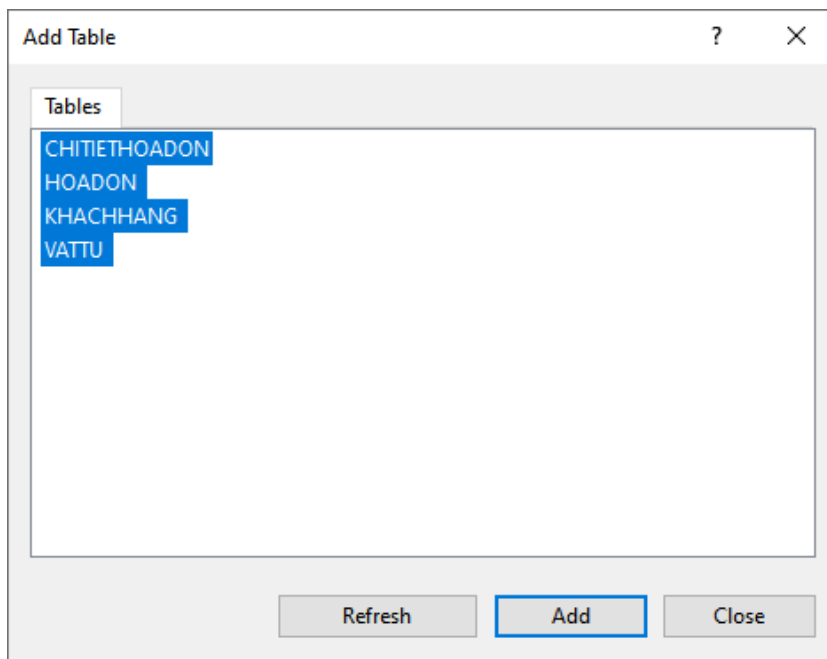
¹¹ Khóa ngoại sẽ được trình bày kỹ hơn trong các buổi sau

¹² sa (viết tắt của system admin) là tài khoản quản trị có quyền cao nhất của SQL Server

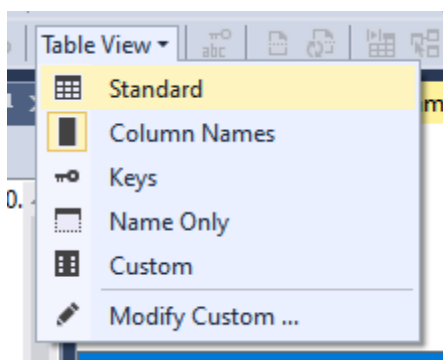
Nếu SQL Server yêu cầu cài đặt công cụ vẽ lược đồ thì nhấn *Yes*:



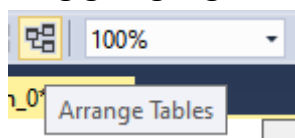
Trong cửa sổ *Add Table*, chọn các bảng muốn hiển thị trong lược đồ (dùng thêm phím *Ctrl* hoặc *Shift* để chọn nhiều bảng), sau đó nhấn *Add*:



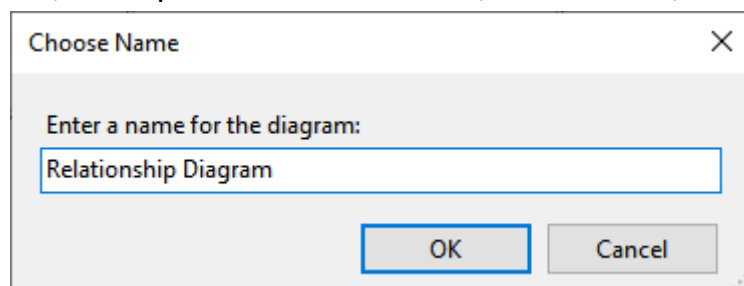
Để hiển thị kiểu dữ liệu của thuộc tính, chọn bảng (dùng thêm phím *Ctrl* hoặc *Shift* để chọn nhiều bảng, hoặc nhấn phím tắt *Ctrl + A* để chọn tất cả các bảng). Sau đó, ở mục *Table View* chọn *Standard*:



Để yêu cầu SQL Server sắp xếp các bảng gọn gàng, nhấn nút *Arrange Tables*:



Để lưu lại lược đồ CSDL, nhấn phím tắt *Ctrl + S* và đặt tên cho lược đồ, sau đó nhấn *OK*:



2. Bảng dữ liệu

Bảng dữ liệu dùng để thể hiện các dữ liệu được chứa trong các bảng của CSDL. Các dòng thể hiện các bản ghi, các cột thể hiện các thuộc tính:

MAKH	TENKH	DIACHI	DT	EMAIL
KH01	Nguyễn Thị Bé	Tân Bình	8457895	bnt@yahoo.com
KH02	Lê Hoàng Nam	Bình Chánh	9878987	bnt@yahoo.com

Trên thực tế, chúng ta sử dụng song song cả lược đồ CSDL và bảng dữ liệu để thể hiện cấu trúc và dữ liệu của CSDL.

V. Sao lưu và phục hồi CSDL

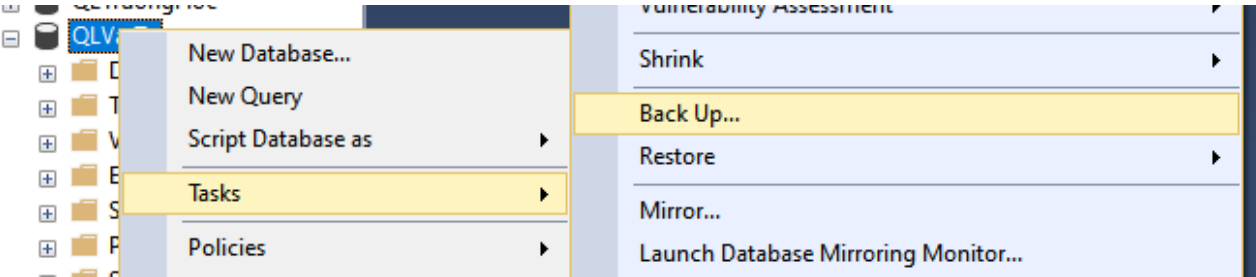
Trong quá trình vận hành các ứng dụng có sử dụng CSDL, việc sao lưu lại CSDL để tránh sự cố là rất quan trọng. Tần suất sao lưu càng nhiều thì sẽ càng giảm thiểu rủi ro mất dữ liệu khi xảy ra sự cố, nhưng bù lại sẽ tốn dung lượng để lưu trữ các bản sao lưu này.

1. Backup & Restore

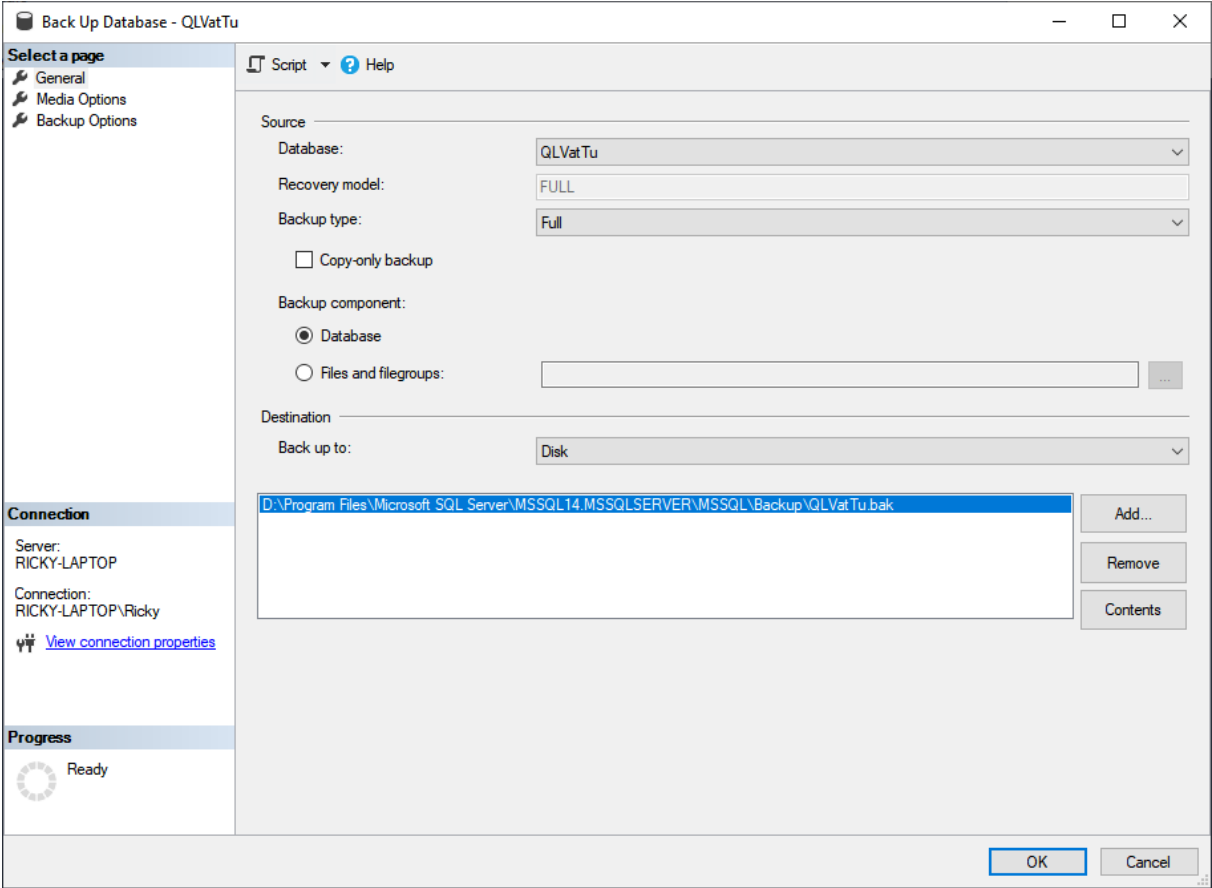
Đây là cách chính thức để sao lưu CSDL.

Thao tác backup như sau:

- Bước 1: Click chuột phải vào CSDL, chọn *Tasks* → *Backup*:



- Bước 2: Trong cửa sổ hiện ra, chọn CSDL muốn sao lưu tại mục *Source – Database*, sau đó nhấn *OK*. Mặc định file backup (.bak) sẽ được đặt trong thư mục *Backup* nằm trong thư mục cài SQL Server. Nếu muốn thay đổi nơi lưu file, dùng nút *Add* và *Remove*:

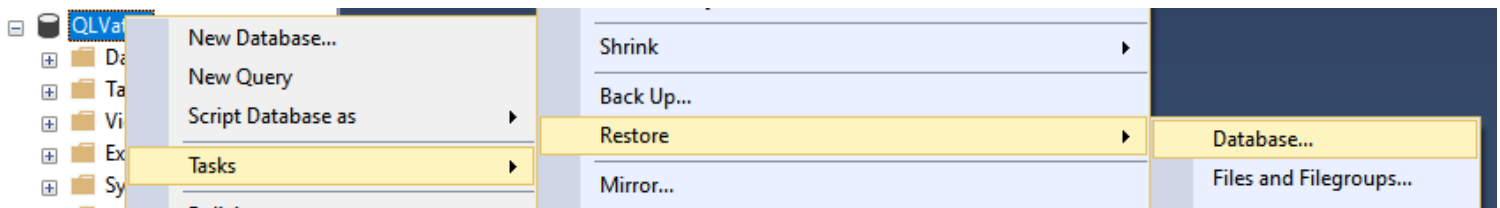
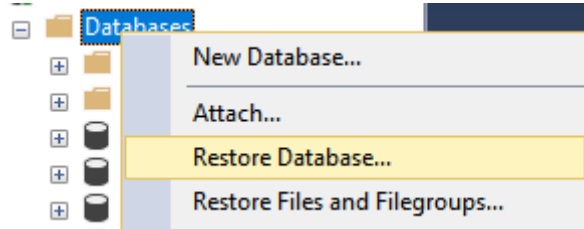


Lưu ý: Mục *Backup type* có thể chọn *Full* (sao lưu toàn bộ) hoặc *Differential* (chỉ sao lưu những phần khác biệt).

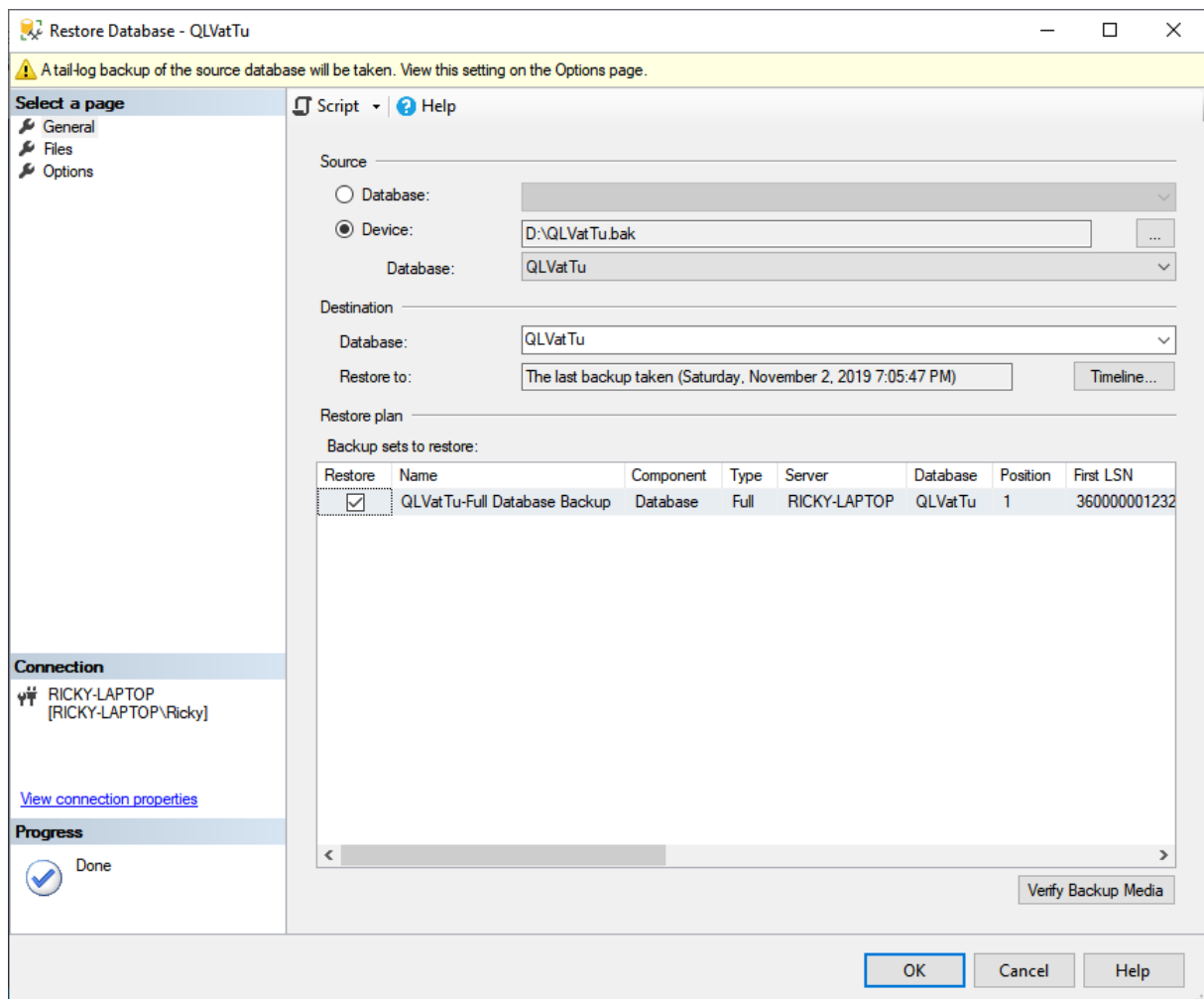
Sau khi đã có file .bak, có thể khôi phục (*restore*) lại.

Thao tác restore như sau:

- Bước 1: Click chuột phải vào mục *Databases*, chọn *Restore Database...* Tuy nhiên nếu muốn restore trực tiếp lên 1 CSDL cụ thể, click chuột phải vào CSDL đó, chọn *Tasks* → *Restore* → *Database...*:



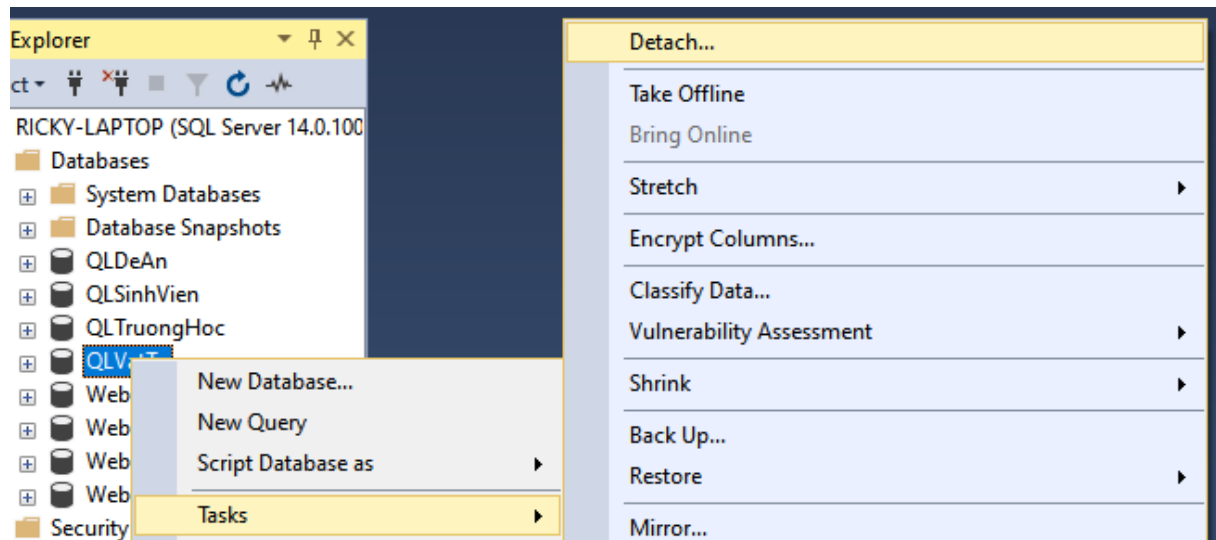
- Bước 2: Trong cửa sổ hiện ra, ở mục *Source* chọn *Device* và tìm đến file .bak cần phục hồi. Thông tin của CSDL sẽ được hiển thị. Sau khi kiểm tra thông tin, nhấn OK:



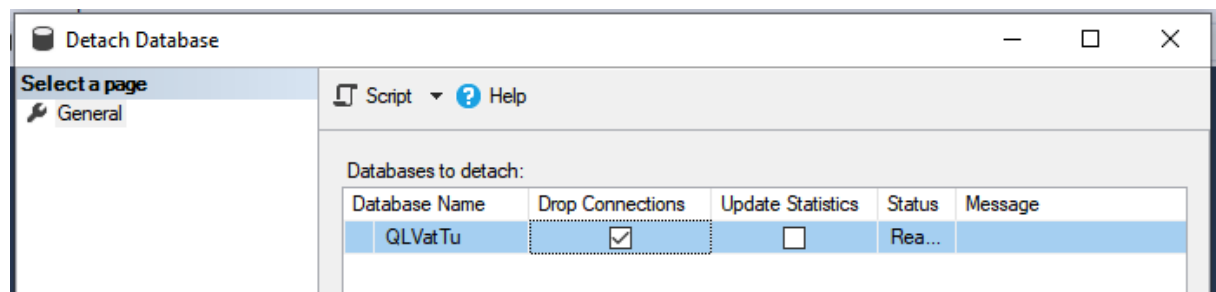
2. Detach & Attach

Về bản chất, đây là cách để copy CSDL từ 1 máy sang 1 máy khác, không phải là cách chính thức để sao lưu CSDL.

Đầu tiên, cần gỡ bỏ CSDL ra khỏi SQL Server bằng cách click chuột phải vào CSDL, chọn *Tasks* → *Detach...*:

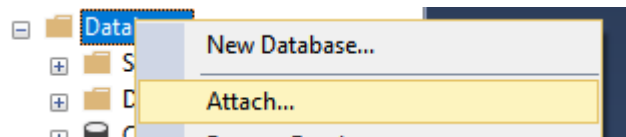


Trong cửa sổ *Detach Database*, check vào mục *Drop Connections* để ngắt các kết nối hiện có tới CSDL:

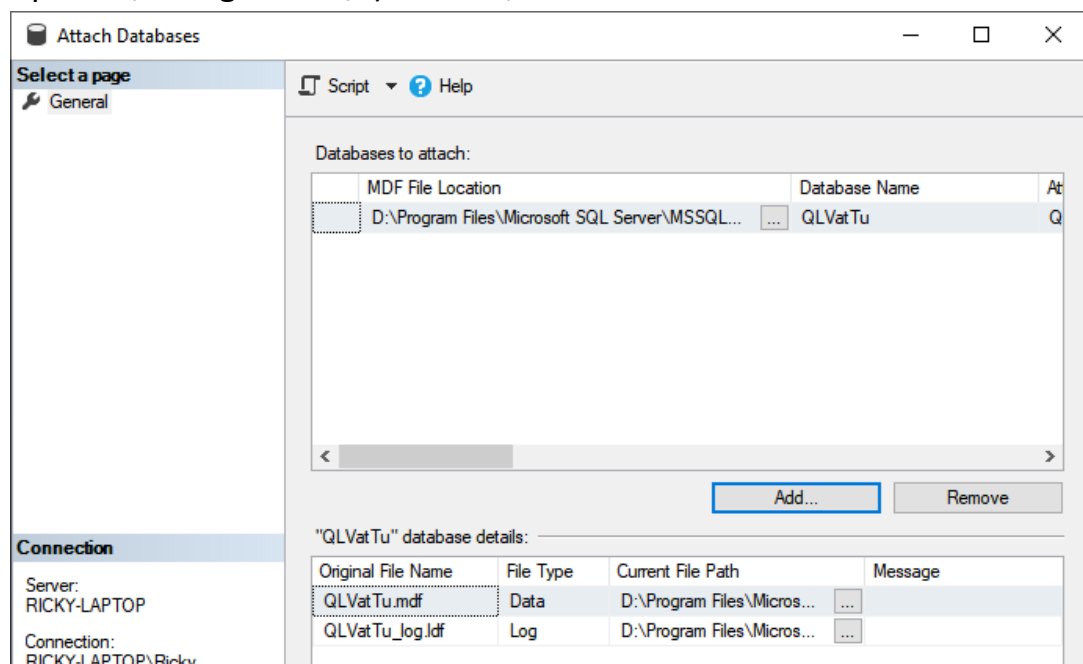


1 CSDL trong SQL Server được lưu bằng 2 file có định dạng *.mdf* và *.ldf* đặt trong thư mục *DATA* nằm trong thư mục cài SQL Server. Do đó nếu muốn chép sang máy khác, cần chép cả 2 file này.

Để gắn kết lại CSDL vào SQL Server, click chuột phải vào mục *Databases*, chọn *Attach...*:



Trong cửa sổ *Attach Databases*, nhấn *Add...* và tìm đến file *.mdf*. File *.ldf* sẽ được tự động tìm thấy (cả 2 file phải đặt cùng thư mục). Sau đó, nhấn *OK*:



3. So sánh Backup/Restore và Detach/Attach

Backup & Restore	Detach & Attach
Bản chất là sao lưu CSDL.	Bản chất là copy CSDL sang máy khác.
Có thể thực hiện mà không cần ngắt kết nối tới CSDL (chế độ online).	Phải ngắt kết nối tới CSDL trước khi thực hiện (chế độ offline).
Thời gian sao lưu lâu (nhất là khi backup Full). Thời gian phục hồi lâu gấp khoảng 3 lần thời gian sao lưu.	Thời gian detach và attach rất nhanh.
Có thể chỉ sao lưu 1 phần.	Phải sao chép toàn bộ CSDL.
Có thể phục hồi trực tiếp lên 1 CSDL có sẵn.	Không thể ghi đè lên CSDL có sẵn.
Có nhiều tùy chọn khác hỗ trợ việc sao lưu.	Không có các tùy chọn nâng cao.

Lưu ý: Với cả 2 cách, file được tạo từ SQL Server phiên bản mới hơn sẽ không thể Restore hoặc Attach trên phiên bản cũ hơn.

VI. Bài tập

Trong các bài tập sau đây, bỏ qua các thao tác có liên quan tới khóa ngoại:

1. Tạo cấu trúc và thêm dữ liệu cho CSDL QLVatTu – Quản lý cửa hàng vật tư. Lược đồ và dữ liệu của CSDL có trong file *[QLVatTu] Mo ta CSDL.pdf*.
2. Tạo cấu trúc và thêm dữ liệu cho CSDL QLDeAn – Quản lý đề án công ty. Lược đồ và dữ liệu của CSDL có trong file *[QLDeAn] Mo ta CSDL.pdf*.