

Tổng quan hệ quản trị CSDL SQL Server

TS. LẠI HIỀN PHƯƠNG

EMAIL: LHPHUONG@TLU.EDU.VN

Nội dung

- Giới thiệu tổng quan về hệ quản trị CSDL
- Giới thiệu về hệ quản trị CSDL SQL Server
- Kiến trúc của SQL Server
- Các thành phần của SQL Server
- Mô hình dữ liệu quan hệ

Tổng quan HQT CSDL

Giới thiệu chung

- Ngày nay, sự thành công của một tổ chức phụ thuộc vào:
 - khả năng thu thập dữ liệu chính xác và kịp thời về hoạt động của họ
 - khả năng quản lý, sử dụng dữ liệu này để phân tích và đưa ra quyết định hoạt động phù hợp
- Thời đại bùng nổ thông tin
 - Cần có các hệ quản trị dữ liệu mạnh mẽ và linh hoạt giúp đơn giản hóa việc quản lý dữ liệu và trích rút các thông tin hữu ích một cách kịp thời.
- Một hệ CSDL (Database system) bao gồm một CSDL (Database) và một hệ quản trị CSDL (Database Management System)

Cơ sở dữ liệu là gì?

- Cơ sở dữ liệu (CSDL - Database) là một tập hợp dữ liệu được tổ chức và lưu trữ theo một cấu trúc chặt chẽ, thường mô tả hoạt động của một hay nhiều tổ chức liên quan.
- Ví dụ: một CSDL của một trường đại học bao gồm:
 - Các thực thể (entities): sinh viên, giảng viên, các khóa học, các lớp học
 - Mối quan hệ (relationship) giữa các thực thể: việc ghi danh của sinh viên vào các môn học, giảng viên giảng dạy các môn học, các phòng học được sử dụng cho các môn học.

Các mô hình CSDL

- **Mô hình dữ liệu file:** dữ liệu được lưu trữ 1 cách rời rạc trên các file. Thường dùng cho các mô hình đơn giản chỉ có một vài file.
- **Mô hình phân cấp:** dữ liệu được tổ chức liên kết vật lý theo mô hình cây (quan hệ nút cha nút con). 1 bản ghi cha có thể có nhiều bản ghi con nhưng 1 bản ghi con chỉ có 1 bản ghi cha.
- **Mô hình mạng:** dữ liệu được tổ chức liên kết vật lý theo dạng tập hợp. 1 bản ghi cha có quyền có nhiều bản ghi con và một bản ghi con có quyền có nhiều bản ghi cha.
- **Mô hình quan hệ:** dữ liệu được tổ chức thành các bảng dữ liệu. Mỗi bảng dữ liệu gồm có các cột (các trường) và các dòng (các bản ghi). Giữa các cột trong bảng và giữa các bảng dữ liệu khác nhau có các ràng buộc.
- **CSDL quan hệ là CSDL sử dụng mô hình quan hệ**

Hệ quản trị cơ sở dữ liệu là gì?

- **Hệ quản trị CSDL** (DataBase Management System – DBMS) là một công cụ phần mềm được thiết kế nhằm hỗ trợ việc lưu trữ, truy xuất, quản trị và tương tác với các CSDL.
- Có nhiều loại HQTCSDL khác nhau, tuy nhiên trong môn học này chúng ta chỉ quan tâm đến các HQTCSDL quan hệ.
 - Hệ quản trị CSDL dùng để quản lý các CSDL quan hệ
- Một số HQTCSDL phổ biến: Oracle, SQL Server, DB2, Access, MySQL.

Lịch sử phát triển

- Đầu những năm 1960, HQT CSDL đầu tiên tạo ra bởi Charles Bach được đặt tên là IDS (Integrated Data Store). Đây là mô hình dữ liệu mạng (network data model) cơ bản.
- Cuối những năm 1960, IBM phát triển HQT CSDL tên là IMS (Information Management System). Đây là mô hình quản lý dữ liệu kiểu phân cấp (hierarchical data model).
- Vào năm 1970, Edgar Codd (IBM) đã đưa ra mô hình biểu diễn dữ liệu kiểu quan hệ (relational data model) đầu tiên, kéo theo sự phát triển nhanh chóng của các HQT CSDL kiểu quan hệ và chúng trở thành kiểu HQT CSDL thống trị cho đến ngày nay.
- Vào những năm 1980, ngôn ngữ truy vấn có cấu trúc (Structured Query Language – SQL) cho các CSDL quan hệ được phát triển bởi IBM với dự án SYSTEM-R. SQL được chuẩn hóa và được công nhận bởi lần đầu tiên vào năm 1986 (chuẩn SQL-86) bởi Viện tiêu chuẩn quốc gia Mỹ (ANSI) và tổ chức tiêu chuẩn thế giới (ISO)
- Tới nay, SQL đã qua 3 lần chuẩn hóa lại để mở rộng các phép toán và tăng cường khả năng bảo mật và toàn vẹn dữ liệu. Hiện nay, đa phần các ngôn ngữ truy vấn CSDL dựa trên chuẩn SQL-92.

Lịch sử phát triển (tiếp)

- Vào cuối những năm 1980 và những năm 1990, các hệ thống dữ liệu tiếp tục phát triển vượt bậc với các ngôn ngữ truy vấn mạnh mẽ hơn và các mô hình dữ liệu phức tạp hơn. Một số HQT CSDL (DB2 của IBM, Oracle 8, Informix UDS) cho phép lưu trữ các kiểu dữ liệu ảnh hay văn bản, cũng như cho phép các truy vấn phức tạp hơn. Một số hệ thống chuyên biệt còn cho phép tạo các kho dữ liệu, hợp nhất dữ liệu từ các nguồn khác nhau hay cho phép các phân tích chuyên biệt.
- Một số HQT CSDL (của Baan, Oracle, PeopleSoft, Siebel) còn tích hợp thêm các gói hỗ trợ quản lý doanh nghiệp thực hiện các công việc như quản lý kho hàng, lập kế hoạch nhân sự, phân tích tài chính, ...
- Ngày nay, các HQT CSDL đều có tính năng cho phép dữ liệu lưu trữ trong CSDL được truy cập thông qua các trình duyệt web trong đó các truy vấn được sinh bởi các hình thức truy cập web và dữ liệu trả về được định dạng dưới dạng các ngôn ngữ kiểu HTML.
- Ngày nay, HQT CSDL ngày càng được phát triển theo hướng cho phép quản lý các CSDL đa phương tiện, video tương tác, thư viện số, ...

Tại sao cần HQT CSDL?

- **Xét ví dụ:** một công ty có một lượng lớn (khoảng 500GB) dữ liệu về nhân viên, các phòng ban, các sản phẩm, các giao dịch bán hàng, ... Một số yêu cầu quản lý dữ liệu của công ty này gồm có:
 - Dữ liệu được truy vấn đồng thời bởi nhiều nhân viên
 - Các câu hỏi về dữ liệu cần được giải đáp nhanh chóng
 - Các thay đổi trên dữ liệu được thực hiện bởi nhiều người dùng phải được nhất quán
 - Một số dữ liệu (ví dụ: lương, thưởng, ...) chỉ được hạn chế truy cập bởi một số người dùng nhất định.

Tại sao cần HQT CSDL? (tiếp)

- Việc lưu trữ dữ liệu của công ty này dưới dạng tập hợp các file hệ thống sẽ gặp phải một số vấn đề như sau:
 - Dữ liệu thường không thể lưu trữ hết trong bộ nhớ chính mà phải được lưu trữ trong các thiết bị nhớ thứ cấp. Do đó, các ứng dụng phải có cơ chế đọc dữ liệu từ bộ nhớ thứ cấp vào bộ nhớ chính để xử lý khi cần thiết.
 - Phải viết các chương trình riêng cho từng câu hỏi về dữ liệu của người dùng
 - Việc bảo vệ dữ liệu khỏi những thay đổi không nhất quán bởi các người dùng khác nhau khó được thực hiện, làm tăng sự phức tạp của ứng dụng
 - Phải đảm bảo dữ liệu được khôi phục khi hệ thống bị treo khi đang thực hiện các thay đổi
 - Cơ chế bảo mật của các hệ điều hành thường không đủ mềm dẻo để cho phép thực thi các chính sách bảo mật trong đó các người dùng khác nhau được phép truy cập các phần dữ liệu khác nhau
- Các vấn đề trên được thực hiện một cách dễ dàng nếu sử dụng các HQT CSDL

Ưu điểm của HQT CSDL

- **Đảm bảo tính độc lập dữ liệu** hay sự bất biến của các chương trình ứng dụng đối với các thay đổi về cấu trúc lưu trữ và chiến lược truy xuất trong mô hình dữ liệu.
- **Khả năng truy cập dữ liệu một cách hiệu quả.** Tính năng này đặc biệt quan trọng khi dữ liệu được lưu trữ trên các thiết bị lưu trữ bên ngoài.
- **Đảm bảo tính toàn vẹn và bảo mật dữ liệu.** HQT CSDL thực thi kiểm soát các ràng buộc toàn vẹn trên dữ liệu cũng như kiểm soát các truy cập vào dữ liệu đối với các lớp người dùng khác nhau.

Ưu điểm của HQT CSDL (tiếp)

- **Quản lý dữ liệu:** HQT CSDL hỗ trợ quản lý dữ liệu hiệu quả khi có nhiều người cùng sử dụng chung một bộ dữ liệu. Quản trị viên sẽ chịu trách nhiệm tổ chức biểu diễn dữ liệu sao cho giảm thiểu dư thừa dữ liệu cũng như cho phép truy cập dữ liệu hiệu quả.
- **Cho phép truy cập đồng thời:** HQT CSDL cung cấp khả năng truy cập đồng thời vào CSDL từ nhiều người dùng tại cùng một thời điểm.
- **Phục hồi dữ liệu:** HQT CSDL có khả năng phục hồi dữ liệu, bảo vệ người dùng khỏi các mất mát dữ liệu đối với các lỗi hệ thống.
- **Giảm thời gian phát triển ứng dụng** do HQT CSDL hỗ trợ sẵn các chức năng phổ biến thường dùng trong các ứng dụng truy cập dữ liệu trong CSDL.

Nhược điểm của HQT CSDL

- Phức tạp
- Thao tác dữ liệu hỗ trợ bởi ngôn ngữ truy vấn không phải lúc nào cũng đáp ứng được các yêu cầu của một số ứng dụng
- Giá thành cao

Đặc điểm của HQT CSDL

- **Sự trừu tượng hóa dữ liệu:** HQT CSDL hỗ trợ che dấu những phức tạp trong cấu trúc CSDL thông qua một số mức trừu tượng để đơn giản hóa các tương tác của người sử dụng đối với hệ thống.
 - *Mức vật lý:* mô tả các cấu trúc dữ liệu mức thấp nhất.
 - *Mức logic:* người quản trị CSDL quyết định các dữ liệu nào được lưu trữ và các mối quan hệ giữa các dữ liệu này
 - *Mức khung nhìn:* người sử dụng làm việc với các khung nhìn khác nhau của cùng một CSDL, mỗi khung nhìn mô tả một phần của CSDL.

Đặc điểm của HQT CSDL (tiếp)

- HQT CSDL cung cấp hai loại ngôn ngữ CSDL:
 - *Ngôn ngữ định nghĩa dữ liệu (Data Definition Language – DDL)*: định nghĩa sơ đồ CSDL (các bảng dữ liệu, các mối quan hệ).
 - *Ngôn ngữ thao tác dữ liệu (Data Manipulation Language – DML)*: giúp người dùng truy nhập hay thao tác dữ liệu (tìm kiếm, thêm, sửa, xóa thông tin trong CSDL)

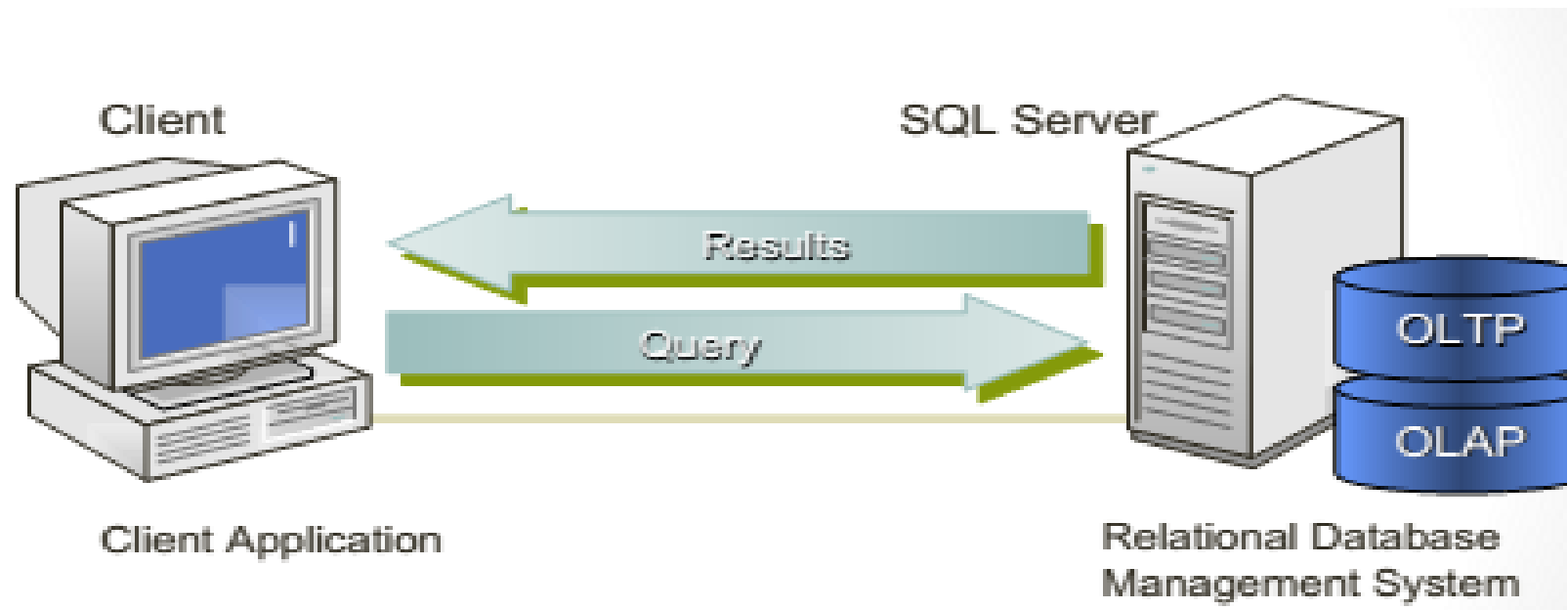
Đặc điểm của HQT CSDL (tiếp)

- HQT CSDL cho phép quản trị giao dịch, phân quyền dữ liệu đồng thời đảm bảo tính nhất quán của CSDL khi thực hiện các giao dịch.
- Quản lý lưu trữ: HQT CSDL phải tổ chức dữ liệu vật lý sao cho tốt, tối thiểu hóa số yêu cầu chuyển dữ liệu giữa đĩa từ vào bộ nhớ chính.

Giới thiệu về HQTCSDL SQL Server

SQL Server

- SQL Server là hệ thống quản trị CSDL quan hệ (Relational Database Management System – RDBMS) sử dụng các lệnh giao chuyển Transaction-SQL để trao đổi dữ liệu giữa Client computer và Server computer.



Đặc tính của SQL Server

- Cho phép quản trị một hệ CSDL lớn (lên đến vài tera byte), có tốc độ xử lý dữ liệu nhanh đáp ứng yêu cầu về thời gian.
- Cho phép nhiều người cùng khai thác trong một thời điểm (lên đến vài chục ngàn user).
- Có hệ thống phân quyền bảo mật.
- Hỗ trợ việc triển khai CSDL phân tán và phát triển ứng dụng trên Internet
- Cho phép lập trình kết nối với nhiều ngôn ngữ lập trình khác dùng xây dựng các ứng dụng đặc thù (Visual Basic, C, C++, ASP, ASP.NET, XML,...).
- Sử dụng câu lệnh truy vấn dữ liệu Transaction-SQL (Access là SQL, Oracle là PL/SQL).

TRANSACTION-SQL

- Là ngôn ngữ SQL mở rộng dựa trên chuẩn ANSI (American National Standard Institute)
- T-SQL bao gồm 3 nhóm:
 - Ngôn ngữ định nghĩa dữ liệu (DDL): CREATE, ALTER, DROP
 - Ngôn ngữ thao tác dữ liệu (DML): SELECT, INSERT, UPDATE
 - Ngôn ngữ điều khiển dữ liệu (DCL): DENY, GRANT, REVOKE

Các phiên bản SQL Server

- Enterprise Manager Edition: là phiên bản đầy đủ của SQL Server, có các dịch vụ phân tích dữ liệu Analysis Service. Phù hợp với các doanh nghiệp, tổ chức có yêu cầu xử lý các giao dịch trực tuyến trên diện rộng, khả năng phân tích dữ liệu phức tạp.
- Standard Edition: giống như Enterprise nhưng bị hạn chế một số tính năng cao cấp. Phù hợp với các doanh nghiệp, tổ chức vừa và nhỏ.
- Developer Edition: tương tự như Enterprise nhưng có hạn chế số lượng người dùng kết nối đến. Dùng để phát triển và kiểm tra ứng dụng.
- Workgroup Edition: dùng cho các doanh nghiệp, tổ chức nhỏ
- Express Edition: phiên bản miễn phí, dễ sử dụng, quản trị CSDL đơn giản.

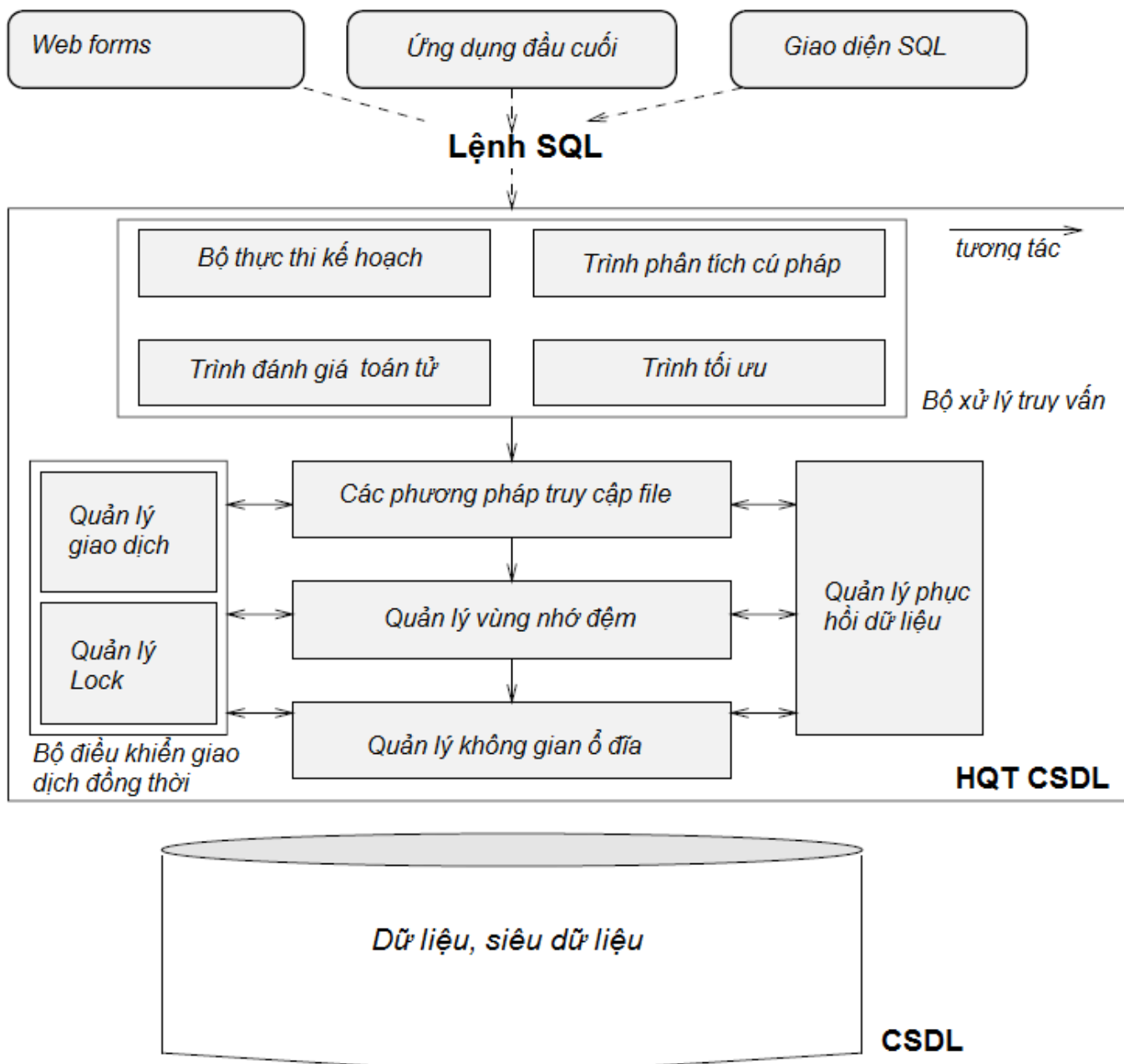
Các phiên bản SQL Server (tiếp)

- Personal: chủ yếu chạy trên PC.
- Desktop Engine: là phiên bản chỉ chạy trên máy tính để bàn và không có giao diện người dùng (GUI), kích thước CSDL tối đa là 2GB.
- Win CE: sử dụng cho các ứng dụng chạy trên Window CE
- Trial: phiên bản dùng thử, bị giới hạn bởi thời gian sử dụng
- SQL Client: phiên bản dùng cho máy khách, khi thực hiện khai thác sẽ kết nối đến phiên bản SQL Server. Phiên bản này cung cấp giao diện GUI khai thác cho người sử dụng.
- SQL Connectivity Only: phiên bản chỉ sử dụng cho các ứng dụng kết nối đến SQL Server. Phiên bản này không cung cấp giao diện GUI cho người dùng khai thác SQL Server.

Kiến trúc của SQL Server

Người dùng thông thường (khách hàng, công ty, ...)

Người dùng chuyên gia, nhà lập trình ứng dụng, người quản trị CSDL



Kiến trúc HQT CSDL quan hệ (tiếp)

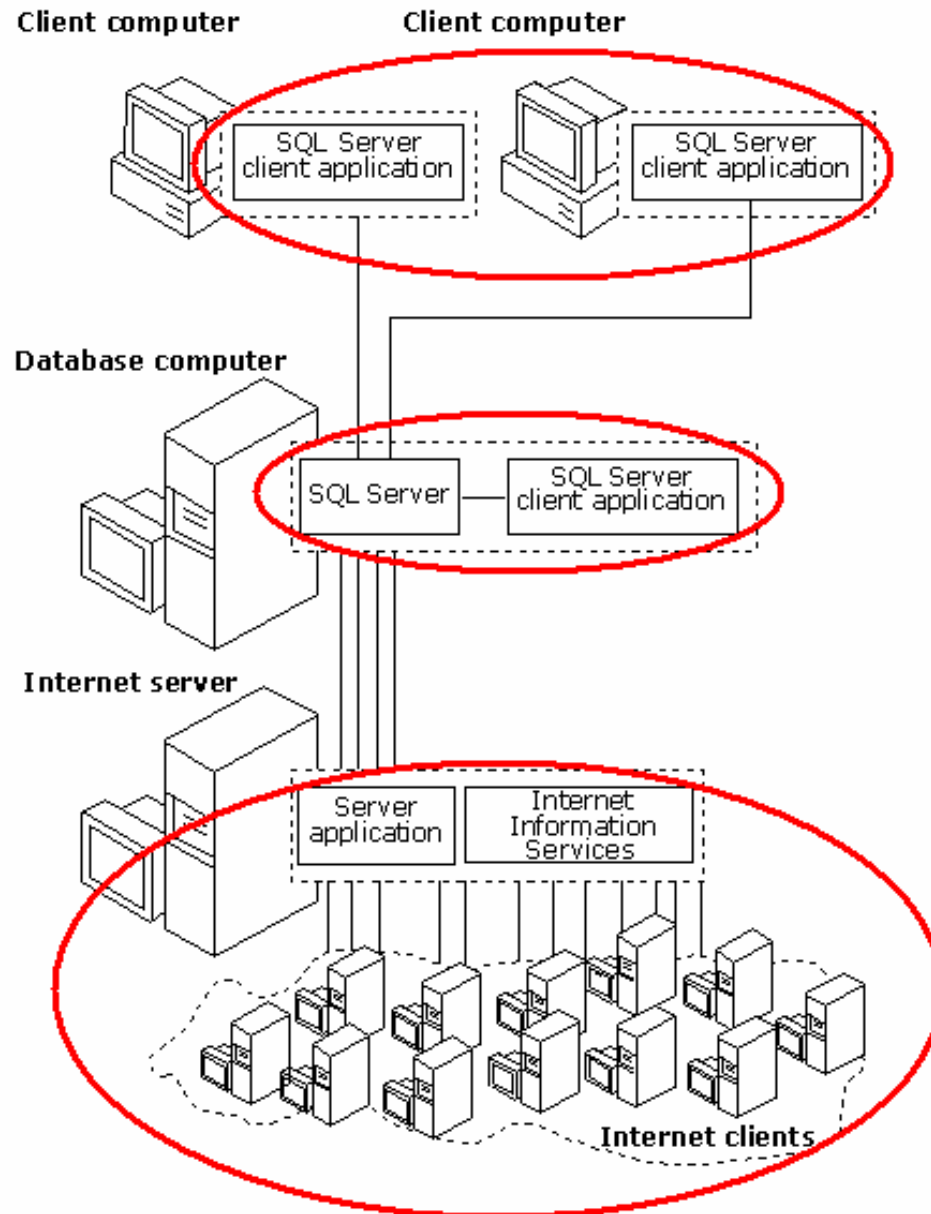
- Dữ liệu cũng như siêu dữ liệu về cấu trúc CSDL được lưu trong các thiết bị nhớ ngoài.
- HQT CSDL nhận các lệnh SQL giao diện người dùng, lập kế hoạch đánh giá các truy vấn, thực hiện các truy vấn trên CSDL và trả về kết quả.

Kiến trúc HQT CSDL quan hệ (tiếp)

- Bộ xử lý truy vấn: nhận truy vấn, phân tích cú pháp và gửi cho bộ tối ưu. Dựa trên cách dữ liệu được lưu trữ, bộ tối ưu sẽ lập kế hoạch thực hiện dưới dạng một cây các toán tử quan hệ.
- Lớp các phương pháp truy cập file: gồm các phần mềm hỗ trợ thiết kế, tổ chức, truy cập các file (tập các trang/bản ghi)
- Bộ quản lý vùng nhớ đệm: đưa các trang dữ liệu từ ổ đĩa vào bộ nhớ chính để trả lời các yêu cầu đọc dữ liệu.
- Bộ quản lý không gian ổ đĩa: thực hiện cấp phát, giải phóng bộ nhớ, đọc, ghi dữ liệu.
- Bộ quản lý phục hồi dữ liệu: duy trì nhật ký các thay đổi đối với CSDL và phục hồi hệ thống về một trạng thái nào đó khi xảy ra sự cố.
- Bộ điều khiển giao dịch đồng thời lên kế hoạch thực hiện các giao tác để tránh đụng độ khi có các giao dịch đồng thời, khi cần thiết có thể phải làm trễ một số thao tác nhất định

Mô hình hoạt động của SQL Server trên mạng máy tính

Mô hình chung SQL Server trên mạng



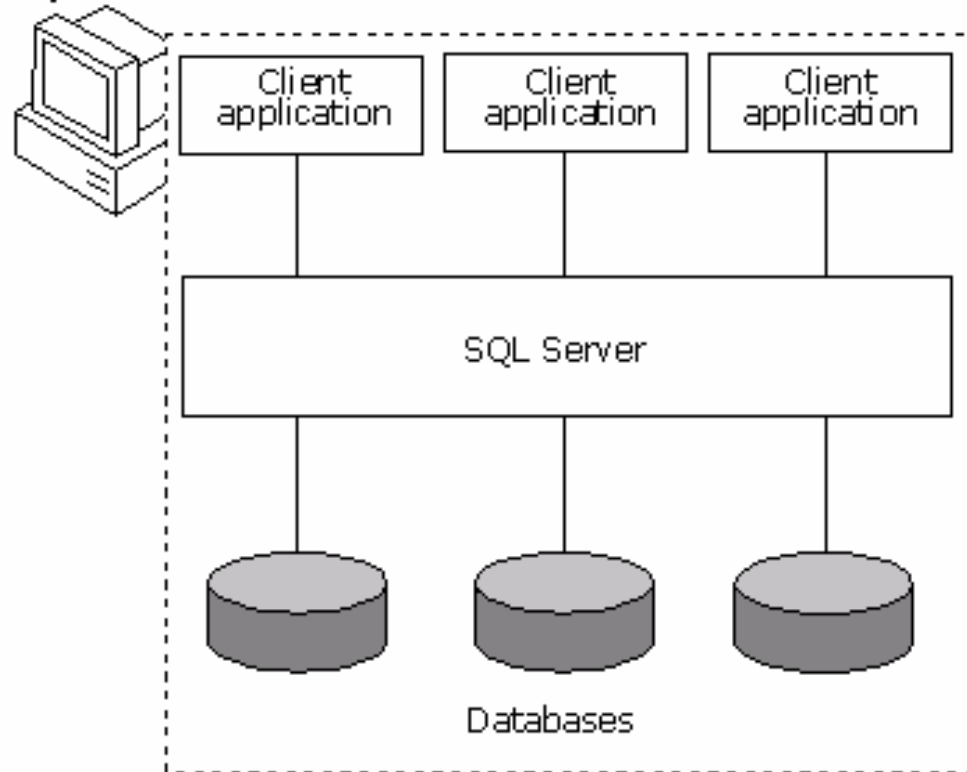
Mô hình chung SQL Server trên mạng (tiếp)

- 3 kiểu kết nối ứng dụng đến SQL Server:
 - Kết nối trên Desktop: các ứng dụng khách nằm trên cùng một máy tính với SQL Server.
 - Kết nối qua mạng nội bộ: các ứng dụng khách nằm trên các máy tính khác và kết nối với máy tính chạy SQL Server qua mạng nội bộ.
 - Kết nối qua mạng Internet: các ứng dụng khách kết nối thông qua máy chủ Internet, dịch vụ IIS thực hiện các ứng dụng trên Internet

Mô hình Desktop

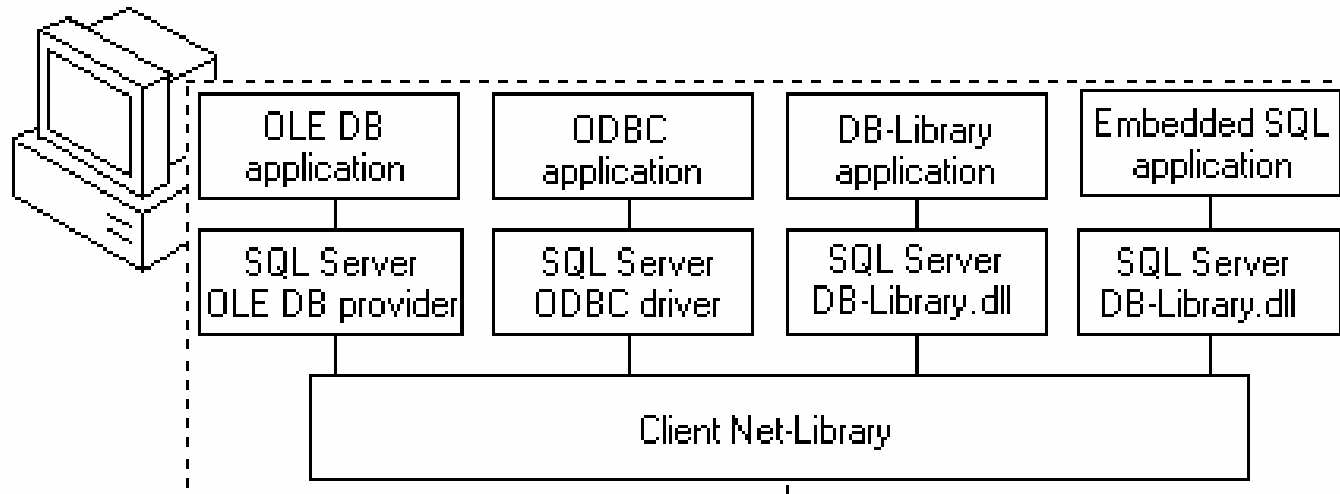
- Trên một desktop có thể có nhiều ứng dụng, thông qua SQL Server để thaco tác với một hay nhiều CSDL

Desktop computer



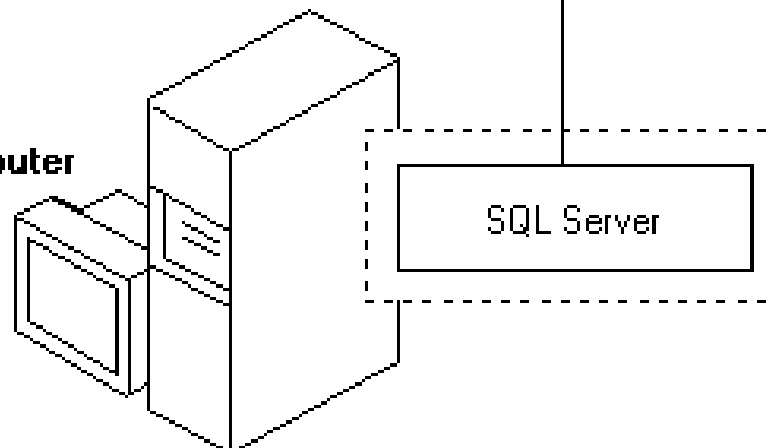
Mô hình Client/Server

Client computer

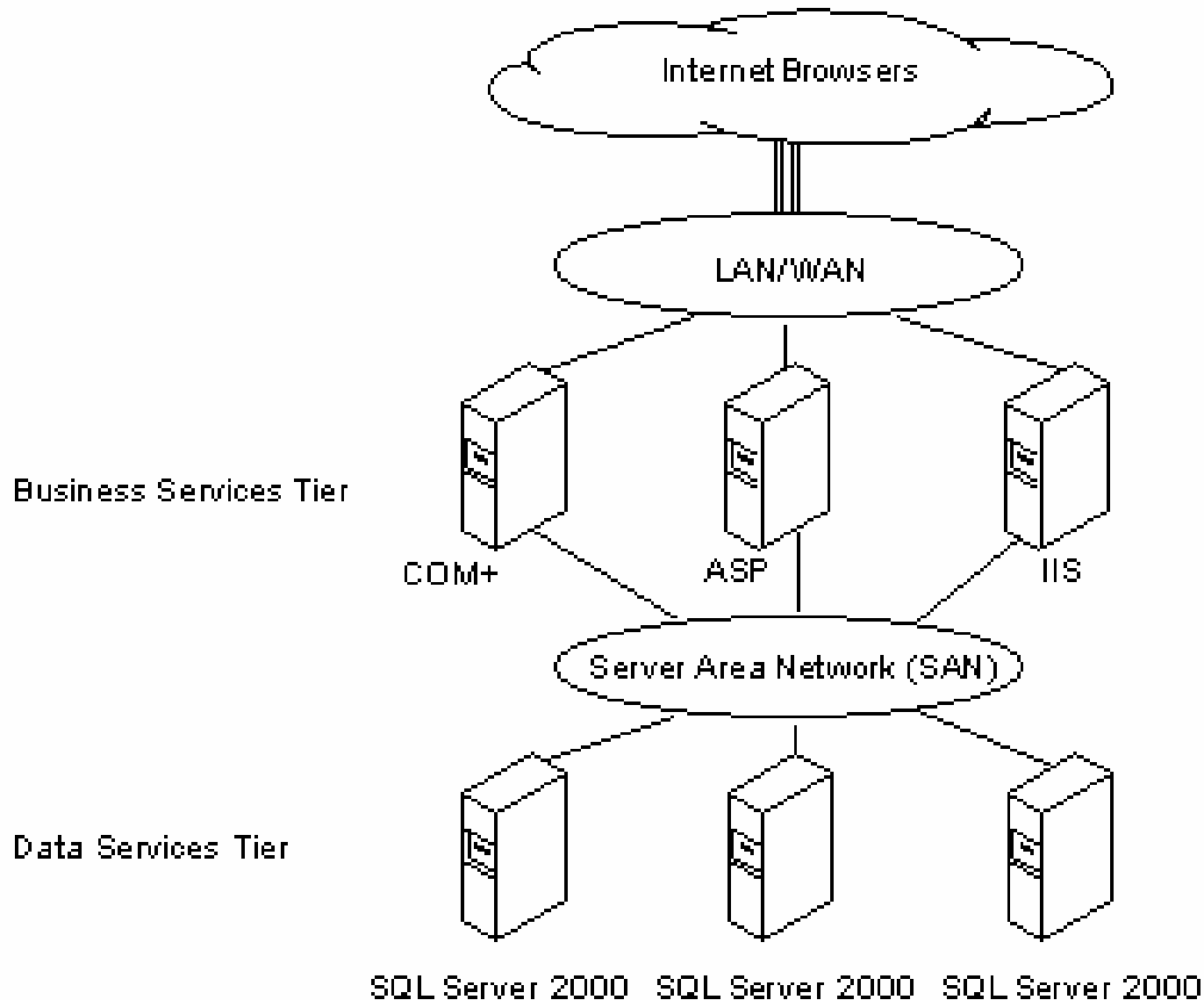


Network

Server computer



Mô hình kết nối trên mạng Internet



Các thành phần của SQL Server

CÁC THÀNH PHẦN CỦA SQL SERVER

- Relational Database Engine: chứa dữ liệu dưới nhiều quy mô khác nhau theo dạng bảng, hỗ trợ nhiều phương thức kết nối (ADO, OLE DB, ODBC)
- Data Transformation Service: cho phép chuyển dữ liệu giữa các server quản trị CSDL khác nhau (SQL server sang Oracle, Access, DB, ...) cũng như chuyển dữ liệu từ dạng này sang dạng khác
- Analyse Service: công cụ khai phá và phân tích dữ liệu, giúp đưa ra những “tri thức” (nhận định, phân tích, đánh giá, dự đoán) từ tập dữ liệu có sẵn

CÁC THÀNH PHẦN CỦA SQL SERVER (tiếp)

- Replication: công cụ nhân bản dữ liệu, giúp tạo ra 1 server khác với bộ dữ liệu giống với server chính, đồng thời tự đồng bộ dữ liệu trên 2 server. Mục đích giúp giảm tải cho server chính, nâng cao số lượng người dùng cũng như phân giao dịch.
- English Query: Truy vấn dữ liệu bằng tiếng Anh
- Meta Data Service: dịch vụ này giúp cho việc lưu trữ và thao tác với meta data.

CÁC THÀNH PHẦN CỦA SQL SERVER (tiếp)

- SQL Server Configuration Management
 - Cho phép cấu hình SQL Server
 - Quản trị các dịch vụ của SQL Server
 - Thiết lập cấu hình kết nối mạng cho SQL Server

CÁC THÀNH PHẦN CỦA SQL SERVER (tiếp)

- SQL Server Management Studio:
 - Mới có từ SQL Server 2005
 - Tích hợp các tính năng của Enterprise Manager, Query Analyser và Analysis Manager
 - Cung cấp truy xuất, cấu hình và quản trị các thành phần của SQL Server
 - Thiết lập và quản lý CSDL
 - Đăng ký server, cấu hình server cục bộ và từ xa
 - Quản lý người dùng: thêm, sửa, xóa, phần quyền cho người dùng

Mô hình dữ liệu quan hệ

Mô hình dữ liệu thực thể - quan hệ

- Mô hình dữ liệu thực thể - quan hệ (Entity-Relationship (ER) data model) mô tả dữ liệu trong thế giới thực bằng các thực thể và các mối quan hệ giữa chúng
- Mô hình này được sử dụng rộng rãi để thiết kế các CSDL quan hệ

Thực thể và tập thực thể

- Thực thể (Entity): là một đối tượng cụ thể hay trừu tượng trong thế giới thực mà tồn tại và có thể phân biệt được với các đối tượng khác
 - VD: “Sinh viên” là một thực thể trừu tượng; “Sinh viên Nguyễn Văn A” là một thực thể cụ thể
- Tập thực thể (Entity set): là một tập các thực thể tương tự nhau.
 - VD: toàn bộ sinh viên, toàn bộ giáo viên

Thực thể và tập thực thể (tiếp)

- Thuộc tính (Attributes): Mỗi thực thể được đặc trưng bởi một tập các đặc tính, gọi là thuộc tính.
 - Tất cả các thực thể trong cùng một tập thực thể phải có cùng một tập các thuộc tính
 - Tập các giá trị có thể của một thuộc tính được gọi là miền giá trị (domain) của thuộc tính
 - Mỗi tập thực thể có một khóa là một tập **tối thiểu** các thuộc tính mà giá trị của nó xác định duy nhất mỗi thực thể trong tập các thực thể.

Thực thể và tập thực thể (tiếp)

- Ví dụ: các thực thể
 - LOP(MaLop, TenLop, Khoa)
 - SINHVIEN(MaSV, HoTen, NgSinh, MaLop)
 - MONHOC(MaMon, TenMon, MoTa)
 - KETQUA(MaSV, MaMon, Diem)

Mối quan hệ

- Mối quan hệ (Relationship): thể hiện liên kết giữa 2 hay nhiều thực thể.
 - VD: mối quan hệ giữa sinh viên và lớp thể hiện sinh viên A sẽ thuộc một lớp B nào đó

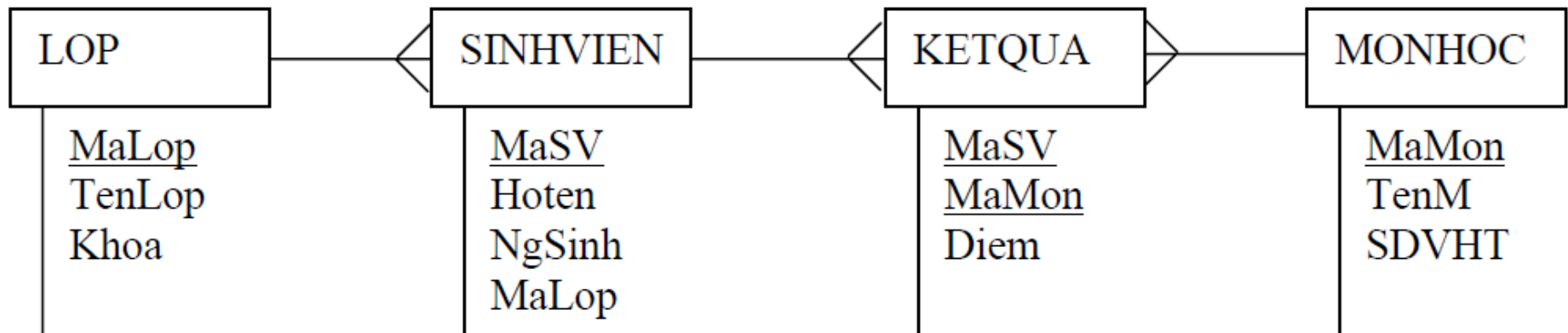
Mối quan hệ (tiếp)

- Các kiểu liên kết:

- Liên kết một – một: mỗi thực thể trong tập thực thể này có nhiều nhất một thực thể trong tập thực thể kia liên kết với nó và ngược lại.
- Liên kết một - nhiều: tập thực thể A liên kết 1-nhiều với tập thực thể B nếu:
 - Mỗi thực thể trong A được liên kết với 0 hay nhiều thực thể trong B
 - Mỗi thực thể trong B được liên kết với nhiều nhất 1 thực thể trong A
- Liên kết nhiều – nhiều: mỗi thực thể trong tập thực thể này có thể liên kết với 0 hay nhiều thực thể trong tập thực thể kia và ngược lại

Mối quan hệ (tiếp)

- VD: Mối quan hệ giữa tập các thực thể



Mô hình dữ liệu quan hệ

- Mô hình dữ liệu quan hệ làm việc trên bảng hay trên quan hệ trong đó:
 - Mỗi bảng tương ứng với một tập các thực thể
 - Mỗi cột trong bảng là một thuộc tính
 - Mỗi dòng trong bảng là một bộ (một bản ghi) tương ứng với một thực thể trong tập

Mô hình dữ liệu quan hệ (tiếp)

- Ưu điểm của mô hình dữ liệu quan hệ:
 - Cấu trúc dữ liệu dễ dùng, không cần hiểu biết sâu về kỹ thuật cài đặt
 - Cải thiện tính độc lập dữ liệu và chương trình
 - Cung cấp ngôn ngữ thao tác phi thủ tục
 - Tối ưu hóa cách truy xuất dữ liệu
 - Tăng tính bảo mật và toàn vẹn dữ liệu

Mô hình dữ liệu quan hệ (tiếp)

- Khóa của quan hệ:

- Khóa chính của quan hệ (primary key): là tập các thuộc tính dùng để phân biệt hai bộ bất kỳ trong quan hệ
- Khóa ngoại của quan hệ (Foreign Key): Một thuộc tính được gọi là khóa ngoại của quan hệ nếu nó trỏ đến khóa chính của quan hệ khác

Các đối tượng CSDL trong SQL Server

- Tables: bảng chứa dữ liệu
- Columns: các cột trong bảng
- Rows: các dòng trong bảng
- Data Types: các kiểu dữ liệu
- Constraint: các ràng buộc dữ liệu
- Defaults: giá trị mặc định của cột nào đó
- Rules: các luật được thiết lập trên dữ liệu
- Index: chỉ số
- View: khung nhìn
- Store procedures, Functions, Triggers