

BỘ GIÁO DỤC VÀ ĐÀO TẠO
ĐẠI HỌC ĐÀ NẴNG

NGUYỄN THỊ HUYỀN

**XÂY DỰNG CƠ SỞ DỮ LIỆU PHÂN TÁN CHO
HỆ THỐNG THI TRẮC NGHIỆM TẠI TRƯỜNG
CAO ĐẲNG KINH TẾ-KỸ THUẬT KIÊN GIANG**

Chuyên ngành : Khoa học máy tính
Mã số: 60.48.01

TÓM TẮT LUẬN VĂN THẠC SĨ KỸ THUẬT

Đà Nẵng - Năm 2013

Công trình được hoàn thành tại
ĐẠI HỌC ĐÀ NẴNG

Người hướng dẫn khoa học: PGS.TS. LÊ VĂN SƠN

Phản biện 1: PGS. TS. VÕ TRUNG HÙNG

Phản biện 2: PGS.TS. TRẦN CAO ĐỆ

Luận văn được bảo vệ tại Hội đồng chấm luận văn tốt nghiệp Thạc sĩ kỹ thuật họp tại Đại học Đà Nẵng vào ngày 8 tháng 6 năm 2013.

** Có thể tìm hiểu luận văn tại:*

- Trung tâm Thông tin - Học liệu, Đại học Đà Nẵng

MỞ ĐẦU

1. Tính cấp thiết của đề tài

Trong thời đại của xã hội công nghệ thông tin và nền kinh tế tri thức, mọi hoạt động của các tổ chức muốn đạt hiệu quả cao, giành được thắng lợi trong thế cạnh tranh gay gắt thì nhất thiết phải có những phương pháp để có được những thông tin, tri thức cần thiết một cách nhanh và chính xác.

Đối với các hệ thống lớn thì vấn đề xử lý thông tin nhanh cho kết quả chính xác là vấn đề quan tâm. Quá trình xử lý thông tin trong các hệ thống lớn hoạt động trên hệ thống mạng, để tăng tốc độ xử lý thì không những xây dựng hạ tầng mạng có đường truyền lớn, mà chúng ta cần phải nghĩ đến việc phân tán dữ liệu như thế nào trong hệ thống bởi vì nó quyết định rất lớn đến kết quả xử lý thông tin. Do vậy với cùng một hạ tầng mạng có tốc độ đường truyền như nhau nếu hệ thống đặt CSDL phân tán thành nhiều nơi thì quá trình xử lý thông tin cho kết quả nhanh hơn quá trình xử lý thông tin mà hệ thống đặt CSDL tại một nơi. Chính vì những lý do trên, chúng ta cần phải nghiên cứu về CSDL làm thế nào để phân tán dữ liệu ra nhiều nơi dựa vào tần suất sử dụng và khoảng cách địa lý của các đối tượng khai thác dữ liệu để hệ thống xử lý thông tin cho kết quả nhanh đáp ứng được sự chờ đợi của người sử dụng. Đó là lý do mà tôi chọn đề tài này

2. Mục tiêu nghiên cứu

Mục tiêu mà đề tài hướng đến là trình bày được về tổng quan về CSDL phân tán, thiết kế CSDL phân tán cho hệ thống thi trắc nghiệm và triển khai ứng dụng trên hệ thống mạng LAN. Để đáp ứng được mục tiêu đặt ra đề tài cần giải quyết những vấn đề chính sau: Tìm hiểu lý thuyết về CSDL phân tán, thiết kế CSDL phân tán; Thu

thập thông tin và nghiên cứu quy trình mạng nội bộ và quy trình thi trắc nghiệm của trường để phân tích thiết kế hệ thống; Thiết kế CSDL phân tán cho hệ thống, nghiên cứu công cụ để quản trị CSDL phân tán, nghiên cứu công cụ để thiết kế các giao diện và triển khai được trên môi trường mạng phục vụ cho nhiều đối tượng sử dụng hệ thống một cách hiệu quả nhất.

3. Đối tượng và phạm vi nghiên cứu

Đối tượng nghiên cứu của đề tài gồm: Nghiên cứu về CSDL phân tán, thiết kế CSDL phân tán như: các vấn đề về thiết kế phân tán, các phương pháp thiết kế phân tán, phương pháp phân mảnh, cấp phát cho các mảnh, ... tìm hiểu hệ thống thi trắc nghiệm của trường. Ứng dụng lý thuyết thiết kế CSDL phân tán vào hệ thống. Ứng dụng khả năng quản trị CSDL phân tán của SQL Server và Visual Studio vào hệ thống.

4. Phương pháp nghiên cứu

Để thực hiện được mục tiêu và nhiệm vụ đặt ra trong đề tài, tôi áp dụng hai phương pháp nghiên cứu đó là: phương pháp nghiên cứu lý thuyết và phương pháp nghiên cứu thực nghiệm. Đối với phương pháp nghiên cứu lý thuyết: tôi tiến hành nghiên cứu và thu thập các tài liệu liên quan đến CSDL phân tán, thiết kế CSDL phân tán, các công cụ có thể triển khai quản trị CSDL phân tán, các công cụ thiết kế giao diện chạy được trên hệ thống mạng. Tiếp đến tôi thu thập các tài liệu liên quan đến các quy trình tổ chức thi trắc nghiệm tại trường. Đối với phương pháp thực nghiệm: tôi phân tích yêu cầu thực tế của hệ thống và để xác định được các chức năng, quy trình hoạt động của hệ thống. Tiếp theo vận dụng cơ sở lý thuyết liên quan như CSDL phân tán, thiết kế CSDL phân tán vào thiết kế hệ thống, sử dụng

công cụ SQL Server và Visual Studio để quản trị CSDL phân tán và công cụ thiết kế giao diện và cuối cùng đánh giá kết quả đạt được.

Đề tài đã đạt được một số kết quả như sau: đề tài đã trình bày một cách hệ thống, chi tiết theo hướng thực nghiệm là một tài liệu cần thiết, bổ ích cho những người muốn tìm hiểu, thiết kế CSDL phân tán và triển khai ứng dụng phân tán. Qua kết quả thực nghiệm, bước đầu cài đặt ứng dụng cho kết quả khả quan. Trong tương lai dựa vào lý thuyết thiết kế CSDL phân tán này có thể ứng dụng vào các hệ thống lớn và có đối tượng sử dụng rộng.

5. Bố cục đề tài

Báo cáo luận văn được tổ chức thành 3 chương: Chương 1, tôi trình bày các vấn đề liên quan đến CSDL phân tán, đặc biệt làm rõ ưu và nhược điểm của CSDL phân tán, các mô hình có thể triển khai hệ thống phân tán, các loại truy xuất, các mức trong suốt của CSDL phân tán. Chương 2, trình bày các chiến lược phân tán, các phương pháp thiết kế, các vấn đề thiết kế, các phương pháp phân mảnh, cấp phát cho các mảnh. Chương 3, trình bày quy trình thiết kế CSDL phân tán cho hệ thống thi trắc nghiệm, trình bày công cụ quản trị CSDL phân tán và các bước triển khai trên hệ thống mạng.

6. Tổng quan tài liệu nghiên cứu

Tài liệu nghiên cứu về cơ sở dữ liệu phân tán, thiết kế CSDL bằng các công cụ quản trị CSDL; Nghiên cứu tính năng Replication trong SQL Server 2005; Nghiên cứu hiện trạng mạng nội bộ và quy trình thi trắc nghiệm; Cài đặt cấu hình máy tính và tạo các CSDL lưu ở các Server đúng quy định.

Ngoài ra tham khảo các tài liệu giáo viên hướng dẫn về cơ sở dữ liệu, hệ quản trị CSDL, phân tích thiết kế hệ thống và các trang web.

CHƯƠNG 1

TỔNG QUAN VỀ CƠ SỞ DỮ LIỆU PHÂN TÁN

Trong chương này tôi sẽ trình bày những vấn đề liên quan đến CSDL phân tán. Nội dung của chương bao gồm giới thiệu CSDL phân tán, hệ quản trị CSDL phân tán, các đặc trưng của CSDL phân tán, so sánh CSDL phân tán với CSDL tập trung, các hình thức tổ chức hệ thống phân tán, đặc trưng của loại hình hệ thống phân tán, ưu nhược điểm các loại truy xuất, các mức trong suốt của CSDL phân tán, thuật toán nhiều bản sao và đồng bộ dữ liệu.

1.1. CƠ SỞ DỮ LIỆU PHÂN TÁN

1.1.1. Định nghĩa

CSDL phân tán là một tập hợp nhiều CSDL có liên quan logic và được phân bố trên một mạng máy tính.

1.1.2. Các đặc trưng của CSDL phân tán

Vô hình kết mạng: Trong môi trường phân tán, người dùng được tách khỏi mọi chi tiết hoạt động của mạng.

Vô hình nhân bản: Vì lý do về hiệu năng, độ tin cậy và tính sẵn sàng, người ta mong muốn có thể nhân dữ liệu thành nhiều bản trên các máy trạm.

Vô hình phân mảnh: Phân hoạch dữ liệu cho các vị trí khác nhau là yêu cầu tất yếu của hệ phân tán.

1.2. HỆ QUẢN TRỊ CSDL PHÂN TÁN

1.2.1. Định nghĩa

Hệ quản trị CSDL phân tán là một hệ thống phần mềm cho phép quản lý các hệ CSDL phân tán và làm cho việc phân tán trở nên vô hình đối với người sử dụng.

1.2.2. Các thành phần của hệ quản trị CSDL phân tán

Gồm các thành phần sau: Phần quản trị dữ liệu; Phần truyền thông dữ liệu; Phần từ điển dữ liệu và Phần CSDL phân tán.

1.3. SO SÁNH CSDL PHÂN TÁN VÀ CSDL TẬP TRUNG

Dựa trên 4 yếu tố để so sánh: Điều khiển tập trung, độc lập dữ liệu, giảm dư thừa dữ liệu, biệt lập và bảo mật dữ liệu.[3]

1.3.1. Điều khiển tập trung

Trong CSDL tập trung: Điều khiển tập trung là động cơ mạnh nhất cho việc ra đời CSDL.

Trong CSDL phân tán: Điều khiển tập trung ít được nhấn mạnh hơn, điều này phụ thuộc vào kiến trúc của CSDL phân tán.

1.3.2. Độc lập dữ liệu

Trong CSDL phân tán, độc lập dữ liệu cũng quan trọng giống như trong CSDL truyền thống. Tuy nhiên, một khía cạnh mới được thêm vào trong ý niệm của độc lập dữ liệu là trong suốt phân tán.

1.3.3. Giảm dư thừa dữ liệu

Trong CSDL truyền thống, dữ liệu dư thừa được giảm đến mức tối thiểu bởi hai lý do: (i) Sự không tương thích giữa nhiều bản sao của cùng một tập dữ liệu; (ii) Tiết kiệm không gian lưu trữ bằng cách loại bỏ các dư thừa.

Trong CSDL phân tán, việc giảm dư thừa phức tạp hơn vì ngoài hai lý do trên còn nhiều lý do để giảm dư thừa như: (i) Hoạt động của các trình ứng dụng có thể bị tăng lên khi dữ liệu được sao lại tất cả các vị trí nơi trình ứng dụng cần có; (ii) Tính thường trực của hệ thống sẽ tăng lên bởi vì khi có lỗi xảy ra ở một trạm nào nó sẽ không dừng việc thực hiện các ứng dụng của trạm khác nếu dữ liệu đã được sao chép lại.

1.3.4. Biệt lập và bảo mật

Trong CSDL truyền thống, hệ quản trị CSDL tập trung có thể bảo đảm chỉ truy cập đến dữ liệu đã được ủy quyền.

Trong CSDL phân tán, hệ quản trị dữ liệu địa phương thực chất phải đối mặt với các vấn đề giống như hệ quản trị CSDL trong CSDL truyền thống.

1.4. CÁC HÌNH THỨC TỔ CHỨC HỆ THỐNG PHÂN TÁN

1.4.1. Mô hình Peer-to-Peer

Các máy tính cá nhân và máy trạm có thể được sử dụng như một hệ thống độc lập. Mỗi thành viên trong mạng có vai trò ngang nhau.

1.4.2. Mô hình File Server

Có một số máy dịch vụ file được gắn trực tiếp vào mạng LAN. Mỗi máy cá nhân được phân chia một dung lượng cố định trên ổ cứng của File Server, chương trình ở các máy tính cá nhân có thể tham chiếu đến các file trên phần đĩa tương ứng của nó bằng một đặc tả đường dẫn.

1.4.3. Mô hình Client/Server

Server có chức năng điều khiển, lưu trữ CSDL, xử lý truy vấn và quản lý việc khai thác tài nguyên trên mạng của các máy tính khác. Client được sử dụng để chỉ người khai thác tài nguyên mạng.

1.5. CÁC ĐẶC TRƯNG CỦA CÁC LOẠI HÌNH HỆ THỐNG PHÂN TÁN

1.5.1. Đặc trưng của hệ thống File Server và kiến trúc Client/Server

1.5.2. Các chức năng của kiến trúc Client/Server

a) Trình diễn thông tin phân tán

b) Trình diễn từ xa

c) Quản lý dữ liệu từ xa

d) Phân tán chức năng

1.6. ƯU VÀ NHƯỢC ĐIỂM CỦA PHÂN TÁN

1.6.1. Ưu điểm:

Đáp ứng nhanh hầu hết các ứng dụng sử dụng dữ liệu tại các trạm.

Tăng cường các đơn thể ứng dụng và CSDL mà không làm cản trở người sử dụng hiện tại.

Kiểm soát dữ liệu địa phương theo hướng hoàn thiện sự tích hợp và quản trị dữ liệu từ xa.

Tăng cường khả năng của hệ thống liên quan đến sự dư thừa dữ liệu.

1.6.2. Nhược điểm:

Phần mềm đắt và phức tạp

Phải xử lý các thay đổi thông báo trong mọi địa điểm

Khó kiểm soát tính toàn vẹn dữ liệu với nhiều bản sao dữ liệu được phân bố khắp mọi nơi.

Đáp ứng chậm nhu cầu của các trạm trong trường hợp các phần mềm ứng dụng không được phân bố phù hợp với việc sử dụng chung.

1.7. CÁC LOẠI TRUY XUẤT CSDL PHÂN TÁN

1.7.1. Truy xuất từ xa thông qua các tác vụ cơ bản

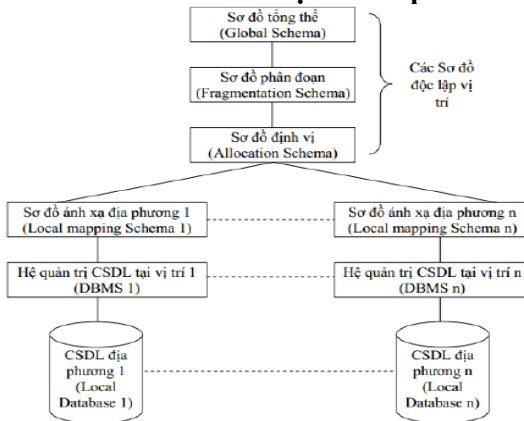
Ứng dụng phát ra một yêu cầu truy xuất CSDL ở một vị trí nào đó. Yêu cầu này sẽ được hệ quản trị CSDL phân tán gửi đến vị trí chứa dữ liệu đó. Thực hiện xong sẽ gửi kết quả về.[3]

1.7.2. Truy xuất từ xa thông qua chương trình phụ trợ

Một ứng dụng yêu cầu thực hiện một chương trình phụ trợ đặt tại vị trí từ xa. Chương trình phụ trợ này sẽ truy xuất CSDL từ xa và trả lại kết quả cho ứng dụng đang yêu cầu.[3]

1.8. CÁC MỨC TRONG SUỐT CỦA CSDL PHÂN TÁN

1.8.1. Kiến trúc cơ bản của một CSDL phân tán



Hình 1.8. Kiến trúc cơ bản của CSDL phân tán

a) *Sơ đồ tổng thể*

b) *Sơ đồ phân đoạn*

c) *Sơ đồ định vị*

d) *Sơ đồ ánh xạ địa phương*

1.8.2. Các đặc điểm chính của hệ phân tán

a) *Chia sẻ tài nguyên*

Chia sẻ tài nguyên được thực hiện thông qua mạng. Mỗi tài nguyên cần phải được quản lý bởi một chương trình có giao diện. Các tài nguyên có thể được truy nhập, cập nhật một cách tin cậy và nhất quán.

b) *Tính mở*

Tính mở của hệ thống phân tán là tính dễ dàng mở rộng phần cứng của nó.

c) Khả năng song song

Khả năng làm việc song song trong hệ thống phân tán được thể hiện qua hai tình huống sau:

- Nhiều người sử dụng đồng thời đưa ra các lệnh hay các tương tác với chương trình ứng dụng.
- Nhiều tiến trình Server chạy đồng thời, mỗi tiến trình phải đáp ứng yêu cầu từ các Clients.

d) Khả năng mở rộng

Khả năng mở rộng của hệ thống phân tán được đặc trưng bởi tính không thay đổi phần mềm hệ thống và phần mềm ứng dụng khi hệ được mở rộng.

e) Khả năng thứ lỗi

Việc thiết kế dựa trên hai giải pháp sau:

- Dùng khả năng thay thế để đảm bảo sự hoạt động liên tục và hiệu quả.
- Dùng các chương trình hồi phục dữ liệu khi xảy ra sự cố.

f) Đảm bảo tính tin cậy và nhất quán

Hệ phân tán phải đảm bảo về khả năng bảo mật của dữ liệu, các chức năng khôi phục hư hỏng phải đảm bảo. Ngoài ra, các yêu cầu của hệ thống về tính nhất quán cũng thể hiện như: không có mâu thuẫn trong nội dung cơ sở dữ liệu.

1.8.3. Tính trong suốt phân tán

Tính trong suốt của một hệ thống phân tán được hiểu như là việc che khuất đi các thành phần riêng biệt của hệ đối với người sử dụng và những người lập trình ứng dụng.

a) Trong suốt phân đoạn

b) Tính trong suốt về vị trí

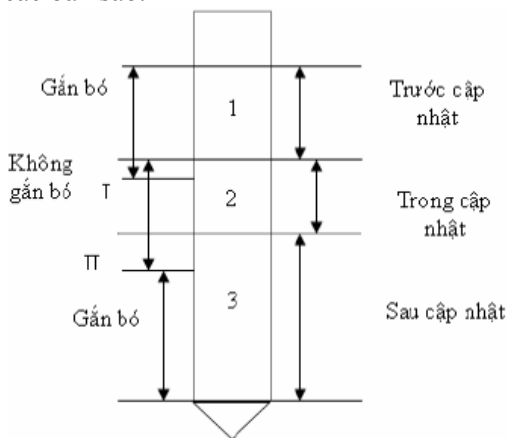
c) Trong suốt ánh xạ địa phương

1.9. THUẬT TOÁN NHIỀU BẢN SAO VÀ ĐỒNG BỘ HÓA DỮ LIỆU

1.9.1. Nhu cầu thực tế và vấn đề quan tâm

1.9.2. Mô hình giải thuật

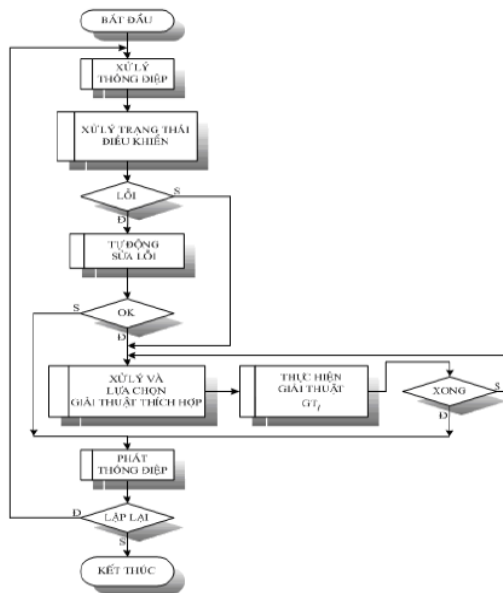
Mô hình và giải thuật nghiên cứu phải đảm bảo các đặc tính toàn vẹn và tuân thủ theo các giai đoạn thể hiện trong hình vẽ sau. Mỗi giai đoạn cập nhật thể hiện những đặc điểm riêng biệt về sự gắn bó dữ liệu giữa các bản sao:



Hình 1.13. Trạng thái dữ liệu trong các giai đoạn cập nhật

1.9.3. Tổng quát

Việc phát hiện và xử lý lỗi trong quá trình xử lý được tiến hành ngay sau khi tác tử nhận được yêu cầu cập nhật. Nếu mọi cố gắng sửa lỗi không thể thực hiện có kết quả, thì thông điệp sẽ được phát đi để yêu cầu tác tử gửi phát lại thông tin. Trong trường hợp công việc cập nhật kết thúc tốt đẹp, một thông điệp khẳng định cũng được phát đi bởi tác tử nhận. Sau khi phát đi thông điệp, tác tử gửi chuyển sang trạng thái chờ thông điệp mới, còn tác tử nhận thông điệp chỉ chuyển sang trạng thái chờ khi đã nhận đủ các thông điệp khẳng định.[6]



Hình 1.15. Sơ đồ tổng quát xử lý nhiều bản sao

1.9.4. Đồng bộ hóa dữ liệu

Đồng bộ hóa dữ liệu Replication cho phép chia sẻ dữ liệu giữa các CSDL tại nhiều điểm khác nhau. Với một dữ liệu gốc, ta có thể tạo nhiều bản sao và sử dụng tại nhiều nơi. Dữ liệu gốc thay đổi, các dữ liệu bản sao cũng sẽ thay đổi và ngược lại.

Tóm lại, trong chương này tôi đã trình bày được các vấn đề liên quan đến CSDL phân tán. Từ những lý thuyết trên giúp người sử dụng biết được CSDL, hệ quản trị CSDL là gì? Và cũng cho chúng ta biết được các đặc trưng của CSDL phân tán, điểm khác với CSDL tập trung, các ưu điểm của CSDL phân tán, các hình thức tổ chức hệ thống phân tán và các loại truy xuất CSDL phân tán. Đặc biệt đối với các mức trong suốt của CSDL phân tán giúp người sử dụng biết được kiến trúc cơ bản của CSDL phân tán và đây cũng là điều kiện tốt để tôi nghiên cứu lý thuyết thiết kế CSDL phân tán ở chương sau.

CHƯƠNG 2

THIẾT KẾ CƠ SỞ DỮ LIỆU PHÂN TÁN VÀ CÁC VẤN ĐỀ VỀ THI TRẮC NGHIỆM

Thiết kế một hệ thống máy tính phân tán cần phải chọn những vị trí đặt dữ liệu và các chương trình trên một mạng máy tính và rất có thể phải kể luôn cả việc thiết kế hệ thống mạng.

Đối với hệ quản trị CSDL phân tán, việc phân tán các ứng dụng đòi hỏi hai điều: Phân tán hệ quản trị CSDL và phân tán các chương trình ứng dụng chạy trên các hệ quản trị đó.

Trong chương này, tôi tập trung nghiên cứu về thiết kế CSDL phân tán. Nội dung của chương trình bày các vấn đề gồm nội dung thiết kế một hệ thống phân tán, các chiến lược phân tán và các vấn đề về thi trắc nghiệm của trường như: Nghiên cứu hiện trạng mạng nội bộ, nghiên cứu chương trình và CSDL của hệ thống thi trắc nghiệm, nghiên cứu tính Replication trong SQL Server 2005. Các nội dung nghiên cứu này làm cơ sở lý thuyết vững chắc để thiết kế CSDL phân tán cho hệ thống phân tán.

2.1. NỘI DUNG THIẾT KẾ MỘT HỆ THỐNG PHÂN TÁN

2.1.1. Các công việc cần phải làm để thiết kế hệ thống phân tán

Xác định kiến trúc mô hình phân tán tổng thể

Định vị các địa phương cần phân tán, loại hình phân tán sử dụng cho mỗi địa phương.

Tiến hành cân đối các yếu tố được phân tán bao gồm các phần tử dữ liệu và các hoạt động xử lý trên mỗi trạm.

Thiết kế CSDL phân tán.

Thiết kế các chương trình ứng dụng.

2.1.2. Các sản phẩm yêu cầu sau khi phân tích thiết kế

2.2. CÁC CHIẾN LƯỢC PHÂN TÁN DỮ LIỆU

Có 4 chiến lược phân tán dữ liệu cơ bản: Tập trung dữ liệu, chia nhỏ dữ liệu, sao lập dữ liệu, phương thức lai.

2.2.1. Tập trung dữ liệu

Tất cả các dữ liệu được tập trung tại một chỗ, cách này đơn giản nhưng có 3 nhược điểm: (i) Dữ liệu không sẵn sàng cho người sử dụng truy nhập từ xa; (ii) Chi phí truyền thông lớn, thường làm cực đại việc truy nhập dữ liệu tới nơi tập trung; (iii) Toàn bộ hệ thống ngưng khi CSDL bị sự cố.

2.2.2. Chia nhỏ dữ liệu

CSDL được chia thành các phần nhỏ liên kết nhau (không trùng lặp). Mỗi phần dữ liệu được đưa đến các trạm một cách thích hợp để sử dụng.

2.2.3. Sao lập dữ liệu

CSDL được nhân thành nhiều bản từng phần hoặc đầy đủ và được đặt ở nhiều trạm trên mạng.

2.2.4. Phương thức lai

CSDL được phân thành 2 phần:

Phần ít quan trọng được lưu giữ một nơi

Phần quan trọng được lưu trữ ở nhiều nơi khác.

2.3. CÁC VẤN ĐỀ THIẾT KẾ PHÂN TÁN

2.3.1. Các lý do phân mảnh

Khung nhìn hoặc đơn vị truy xuất của các ứng dụng không phải là toàn bộ quan hệ mà thường là một mảnh.

Việc phân rã một quan hệ thành nhiều mảnh, mỗi mảnh được xử lý như một đơn vị, sẽ cho phép thực hiện nhiều giao dịch đồng thời.

2.3.2. Các kiểu phân mảnh

Có ba kiểu phân mảnh khác nhau:

Phân mảnh ngang (Horizontal fragmentation)

Phân mảnh dọc (Vertical fragmentation)

Phân mảnh hỗn hợp (Hibrid fragmentation)

2.3.3. Mức độ phân mảnh

Mức độ phân mảnh có thể là từ thái cực không phân mảnh nào đến thái cực phân mảnh thành từng bộ (trường hợp phân mảnh ngang) hoặc thành từng thuộc tính (trường hợp phân mảnh dọc).

2.3.4. Các quy tắc phân mảnh

Có ba qui tắc trong khi phân mảnh đảm bảo CSDL sẽ không thay đổi ngữ nghĩa khi phân mảnh: tính đầy đủ, tính tái thiết và tính tách biệt.

2.3.5. Các kiểu cấp phát

Một CSDL không nhân bản gọi là *cơ sở dữ liệu phân hoạch*, có chứa các mảnh được cấp phát cho các vị trí, trong đó chỉ tồn tại một bản duy nhất cho mỗi mảnh trên mạng.

Kiểu *cơ sở dữ liệu nhân bản* có hai dạng:

Cơ sở dữ liệu nhân bản hoàn toàn, trong đó toàn bộ CSDL đều có bản sao ở mỗi vị trí.

Cơ sở dữ liệu nhân bản một phần, trong đó các mảnh được phân tán đến các vị trí, mỗi mảnh có thể có nhiều bản sao nằm ở các vị trí khác nhau.

2.3.6. Các yêu cầu thông tin

Các thông tin cần cho thiết kế phân tán có thể phân chia thành bốn loại: Thông tin CSDL, thông tin ứng dụng, thông tin về mạng và thông tin về hệ thống máy tính. Hai loại sau có bản chất hoàn toàn định lượng và được sử dụng trong các mô hình cấp phát chứ không phải trong các thuật toán phân mảnh.

2.4. PHƯƠNG PHÁP PHÂN MẢNH

2.4.1. Phân mảnh ngang

Phân mảnh ngang chia một quan hệ theo các bộ, vì vậy mỗi mảnh là một tập con của quan hệ. Có hai loại phân mảnh ngang: phân mảnh ngang nguyên thủy và phân mảnh ngang dẫn xuất.

2.4.2. Phân mảnh dọc

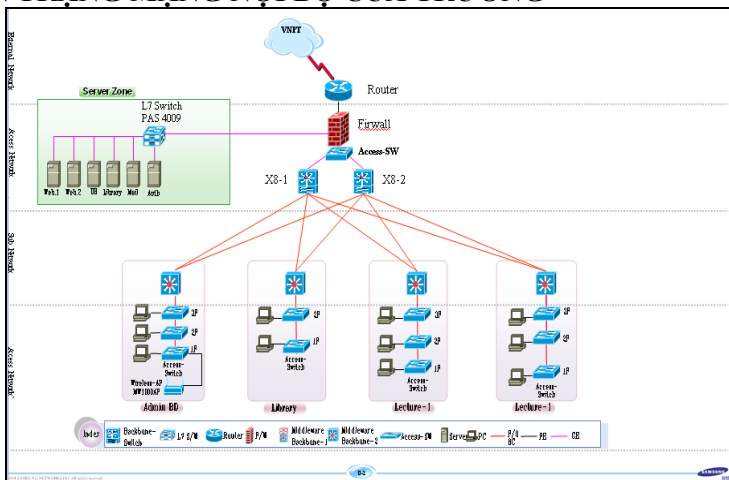
Mục đích của phân mảnh dọc là phân hoạch một quan hệ thành một tập các quan hệ nhỏ hơn để nhiều ứng dụng có thể chỉ chạy trên một quan hệ.

2.4.3. Phân mảnh hỗn hợp

Cách đơn giản nhất để phân mảnh hỗn hợp gồm có:

Áp dụng phân mảnh ngang cho các mảnh phân chia theo chiều dọc; phân mảnh dọc cho các mảnh phân chia theo chiều ngang.

2.5. SỬ DỤNG MẠNG NỘI BỘ HIỆN CÓ VÀ NGHIÊN CỨU HIỆN TRẠNG MẠNG NỘI BỘ CỦA TRƯỜNG



Hình 2.1: Mô hình hệ thống Mạng của trường

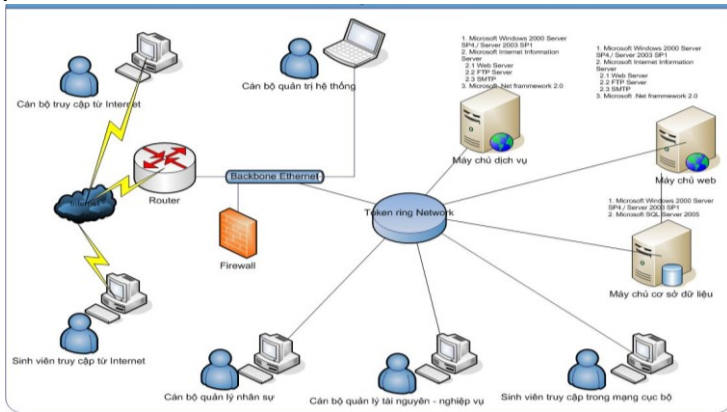
2.4. NGHIÊN CỨU CHƯƠNG TRÌNH VÀ CSDL CỦA HỆ THỐNG THI TRẮC NGHIỆM CỦA TRƯỜNG

2.4.1. Giới thiệu về hình thức thi trắc nghiệm

2.4.2. Tổng quan về hệ thống

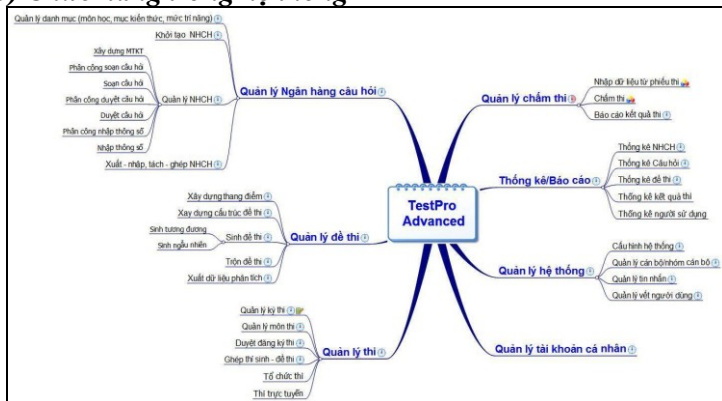
a) Mô hình hệ thống Testpro

Hệ thống TestPro được triển khai trên mạng Internet và mạng cục bộ.



Hình 2.2. Mô hình hệ thống

b) Chức năng trong hệ thống



Hình 2.3. Mô hình các chức năng của hệ thống

2.5. NGHIÊN CỨU TÍNH NĂNG REPLICATION TRONG SQL SERVER 2005

2.5.1. Giới thiệu tính năng Replication trong SQL Server 2005

Replication là một tập hợp các công nghệ cho việc sao chép và phân phối các dữ liệu và các đối tượng CSDL từ một CSDL khác và sau đó đồng bộ giữa các CSDL để duy trì tính nhất quán. Sử dụng Replication, bạn có thể phân phối dữ liệu đến các địa điểm khác nhau và cho người dùng từ xa hoặc qua mạng điện thoại di động khu vực địa phương và rộng, kết nối dial-up, kết nối không dây và mạng Internet.

2.5.2. Các kiểu Replication

2.5.3. Thực hiện Replication trên SQL Server với CSDL thử nghiệm

a) Cấu hình cho Local Publications

b) Cấu hình cho Local Subscriptions

Tóm lại, trong chương 2 tôi đã trình bày được lý thuyết về thiết kế CSDL phân tán và các vấn đề về thi trắc nghiệm. Nội dung chính của chương giúp người sử dụng hiểu được các chiến lược phân tán, nghiên cứu về mạng cục bộ, quy trình thi trắc nghiệm và thực hiện cấu hình của tính năng Replication trên SQL Server với CSDL thử nghiệm.

Từ nội dung của chương giúp tôi có cơ sở lý thuyết vững chắc để xây dựng CSDL phân tán cho hệ thống thi trắc nghiệm và triển khai ứng dụng ở chương sau.

CHƯƠNG 3

GIẢI PHÁP XÂY DỰNG PHÂN TÁN CƠ SỞ DỮ LIỆU

THI TRẮC NGHIỆM CỦA TRƯỜNG CAO ĐẲNG

KINH TẾ - KỸ THUẬT KIÊN GIANG

Trong chương này tôi đưa ra các giải pháp xây dựng phân tán CSDL thi trắc nghiệm của trường. Nội dung gồm nguyên lý thiết kế, đề xuất giải pháp, lựa chọn vị trí đặt CSDL và phân nhóm người sử dụng, từ đó đề nghị hướng phát triển tiếp theo của luận văn.

3.1. NGUYÊN LÝ THIẾT KẾ

Để phục vụ cho việc xây dựng một hệ CSDL phân tán cho hệ thống thi trắc nghiệm tại trường Cao đẳng Kinh tế - Kỹ thuật Kiên Giang, luận văn này dùng lý thuyết đã trình bày ở các chương trước làm tiêu chuẩn. Do đó, việc thiết kế một hệ CSDL phân tán cần phải làm các công việc sau:

- Thiết kế CSDL phân tán
- Dịch các câu hỏi đáp tổng thể về CSDL thành câu hỏi đáp về các mảnh để từng phần của hệ thống chỉ làm việc trên các mảnh.
- Tối ưu hóa chiến lược truy cập
- Quản trị các giao tác phân tán
- Điều khiển tương tranh.
- Quản trị CSDL phân tán.

Hệ thống CSDL này thiết kế dựa trên CSDL SQL Server, công cụ để kết nối CSDL, hệ quản trị CSDL đã có sẵn. Vì vậy công việc thiết kế còn lại chỉ là thiết kế CSDL phân tán và quản trị CSDL phân tán.

3.2. ĐỀ XUẤT GIẢI PHÁP

3.2.1. Mô tả thiết kế CSDL phân tán

Sau quá trình nghiên cứu thực tế mạng nội bộ và chương trình thi trắc nghiệm TestPro của trường, kết hợp với cơ sở lý thuyết đã nghiên cứu tôi xin đưa ra một số đề xuất như sau:

- Khi sinh viên thực hiện thi trắc nghiệm trên máy thì chỉ tương tác với các table trong CSDL TestProPlus như sau: ThiSinh, ThiSinhDapAn, ThiSinhIPAddress, ThiSinhKyThi, ThiSinhMonThi, DeThiCauHoi, DeThi, DapAn, KyThi. Nếu để sinh viên truy cập trực tiếp vào CSDL chính của trường để thực hiện thì dẫn đến hiện tượng nghẽn mạng, tốc độ thực hiện chậm, bảo mật không cao,... vì vậy nên phân tán những table mà sinh viên tương tác khi tiến hành thi trắc nghiệm ra thành một CSDL phân tán, như vậy thì không ảnh hưởng đến những table khác. Tùy theo yêu cầu phát triển hệ thống có thể mở rộng nhiều CSDL có cấu trúc và nhiệm vụ giống nhau CSDL phân tán nói trên.

- Tạo một CSDL bản sao của CSDL chính TestProPlus. Hai CSDL này cùng lưu ở Server của trường, tại Trung tâm CNTT.

3.2.2. Lựa chọn vị trí đặt CSDL và phân nhóm người sử dụng

Đánh giá vị trí đặt CSDL theo một số tiêu chuẩn sao cho vị trí đặt CSDL tiện lợi nhất:

- Tần suất sử dụng CSDL
- Số lần liên kết CSDL
- Các tham chiếu đến CSDL để cập nhật, đọc hay thay đổi.

Dựa trên các tiêu chuẩn trên và tính chất của hệ thống quản lý thi trắc nghiệm có thể lựa chọn hệ thống theo ba nhóm chính tương đương với ba CSDL:

a) Cơ sở dữ liệu Bản sao

Đây là CSDL dành cho việc nhập và thay đổi dữ liệu hệ thống. CSDL này đặt tại Server của trường tại Trung tâm CNTT. Do yêu cầu của công việc giữa các nhóm có thể lấy một số phần thông tin của nhau theo quyền.

b) Cơ sở dữ liệu Phân tán

Đây là CSDL dành cho việc thi trắc nghiệm của sinh viên, theo yêu cầu phát triển hệ thống có thể mở rộng nhiều CSDL có cấu trúc và nhiệm vụ giống như CSDL Phân tán này.

Ở mỗi một dãy lầu tại khu thực hành B chọn một máy để cài SQL Server và thực hiện Replication tại máy này mọi thông tin của sinh viên về bài làm sau khi nộp bài sẽ được lưu tạm tại đây, sau thời gian chỉ định trước, những thông tin này sẽ được cập nhật về CSDL TestProPlus, do phòng Đào tạo quản lý, đối với CSDL Bản sao giành cho Giảng viên thì sẽ không nhận được những thay đổi này vì lý do bảo mật, những thông tin của sinh viên, những thông tin về kết quả của bài thi chỉ được cập nhật cho CSDL chính thức của Phòng Đào tạo. Như thế này đã giải quyết được vấn đề về bảo mật.

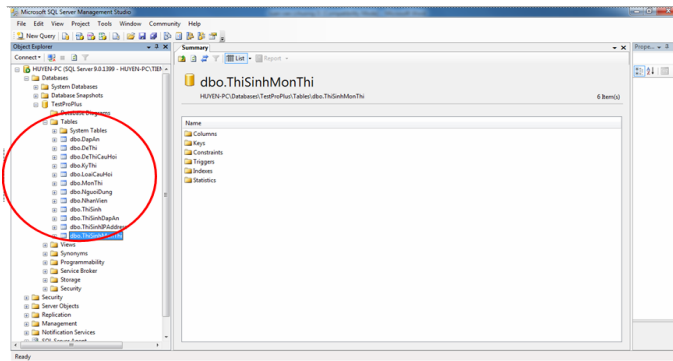
Nếu cùng một lúc cả 3 dãy lầu cùng hi trắc nghiệm thì thời gian ở 3 máy cài Replication khi gửi dữ liệu về cho Server chính sẽ khác nhau, như vậy lưu lượng đường truyền trên mạng tại một lúc chỉ còn 1/3 so với ban đầu, như vậy đã giải quyết được vấn đề về tốc độ đường truyền.

c) CSDL trung tâm TestProPlus

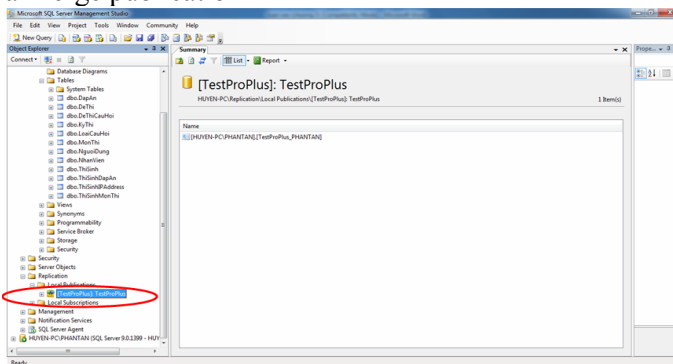
Là CSDL chính phục vụ cho việc truy cập và sử dụng của Phòng Đào tạo. Phòng Đào tạo có thể đọc tất cả thông tin trong CSDL này và có thể thêm quyền sửa đổi với phần báo cáo.

3.3. YÊU CẦU HỆ THỐNG

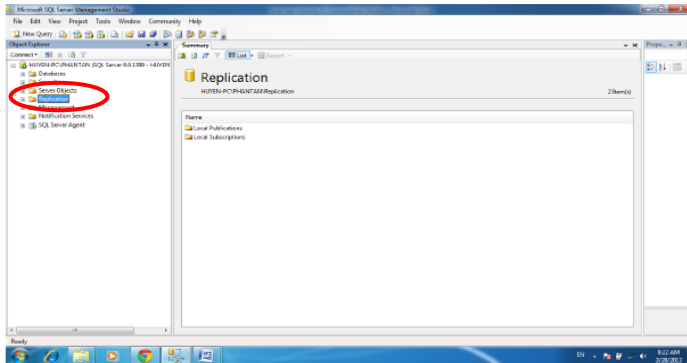
3.4. XÂY DỰNG CÁC SERVER CSDL



Trên con Server này ta sẽ tạo một Publications (theo hướng dẫn trong chương 2) có tên gọi là TestProPlus. Ở đây ta chọn kiểu sao chép là Merge publication



Sau đó trên Server Phân tán có tên là HUYEN-PC\PHANTAN



Ta sẽ tạo một Subscriptions của Publications (TestProPlus) trên Server chính HUYEN-PC (như hướng dẫn chương 1)

Sau khi đã tạo xong Replication cho 2 Server là HUYEN-PC và HUYEN-PC\BANSAO. Tôi đã thực hiện Test dữ liệu xem dữ liệu có đồng bộ trên 2 Server hay không. Kết quả là đồng bộ.

3.5. ĐÁNH GIÁ KẾT QUẢ

Hệ thống đã được cài đặt và vận hành thử một số máy tính tại Khoa CNTT của trường nhìn chung đáp ứng được yêu cầu đề ra. Ưu điểm của hệ thống:

- Thông tin cập nhật nhanh nhất.
- Thực hiện các giao dịch trong khoảng thời gian ngắn nhất.
- Chi phí để xây dựng cho hệ thống nhỏ nhất có thể.
- Giảm được khối lượng xử lý và vận hành trong mạng lớn tại máy trung tâm và hệ thống sẽ phân tán ra các nơi để giải quyết nhu cầu khai thác của người sử dụng.

Tóm lại, trong chương này tôi đã nghiên cứu và thực hiện được giải pháp xây dựng phân tán CSDL thi trắc nghiệm tại trường Cao đẳng Kinh tế - Kỹ thuật Kiên Giang.

KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

Từ việc nghiên cứu cơ sở lý thuyết hệ phân tán, tìm hiểu hệ thống thi trắc nghiệm và hệ thống mạng của trường cao đẳng Kinh tế - Kỹ thuật Kiên Giang, tôi đã cải tiến và xây dựng cơ sở dữ liệu phân tán cho hệ thống thi trắc nghiệm của trường.

1. Những việc đã làm được:

Đã hoàn thành được mục tiêu của luận văn đề ra là nghiên cứu được cơ sở lý thuyết và áp dụng phân tán trên CSDL thi trắc nghiệm của trường. Hệ thống CSDL tôi đưa ra đã phần nào đáp ứng được các yêu cầu sau:

Về lý thuyết: Đã trình bày được tổng quan về CSDL phân tán, thiết kế CSDL phân tán và phần mềm thi trắc nghiệm Testpro 2008. Sau khi đã nghiên cứu hệ quản trị CSDL SQL Server 2005 cho thấy đây là một hệ CSDL hiện đại, hỗ trợ rất mạnh cho CSDL phân tán. Luận văn là cơ sở lý thuyết vững chắc để người sử dụng muốn tìm hiểu và nghiên cứu về CSDL phân tán, giúp người sử dụng có thể áp dụng để phát triển các hệ thống lớn trong môi trường phân tán và phạm vi sử dụng rộng.

Về ứng dụng: Xây dựng thành công CSDL phân tán cho hệ thống thi trắc nghiệm của trường CĐ Kinh tế - Kỹ thuật Kiên Giang. Hệ thống được thiết kế theo hướng phân tán dữ liệu, dữ liệu được đồng bộ tại các Server nên làm tăng tốc độ xử lý và giảm nghẽn mạng khi có nhiều đối tượng khai thác hệ thống.

2. Những việc chưa làm được:

Hệ thống chưa được triển khai thực tế trên mạng LAN của trường.

Chưa khai thác triệt để khả năng quản trị CSDL SQL Server 2005, nghiên cứu về phần mềm thi trắc nghiệm TestPro 2008 còn nhiều hạn chế.

3. Hướng phát triển:

Để hoàn thành đề tài luận văn cần phát triển theo hướng sau:

- Về mặt lý thuyết cần nghiên cứu thêm ủy thác phân tán, điều khiển đồng thời phân tán theo nhãn thời gian, tối ưu hóa vấn tin, điều khiển tương tranh trong CSDL phân tán. Nghiên cứu sâu hơn phần mềm Testpro 2008 và SQL Server 2005.

- Về mặt ứng dụng thì hệ thống triển khai được trên mạng của trường và có thể phạm vi rộng hơn trong khu vực. Kết hợp được các thuật toán điều khiển tương tranh, tối ưu hóa vấn tin vào hệ thống để hệ thống hoạt động tốt đáp ứng nhu cầu của người sử dụng và cho kết quả như mong muốn.