

Tóm tắt: Ngày nay, do nhu cầu về công việc và cuộc sống, dân số tại các tỉnh/thành phố lớn như TP. Hồ Chí Minh, Hà Nội, Đà Nẵng... tăng lên rất nhanh chóng. Quy hoạch và phát triển khu đô thị, khu dân cư cũng tăng nhanh. Song song, hệ thống giao thông đô thị phát triển, mở rộng và cũng đã đặt ra những yêu cầu, thách thức mới trong công tác quản lý hệ thống giao thông đô thị, đặc biệt là quản lý hệ thống các số đo liên u. Vấn đề đặt ra là hệ thống giao thông phải làm việc với nhiều dữ liệu khác nhau. Bài báo này trình bày một giải pháp để tích hợp toàn bộ dữ liệu giao thông thành hệ thống dữ liệu hybrid. Hệ thống này sẽ tích hợp 3 hệ thống cơ sở dữ liệu khác nhau (dữ liệu địa lý, dữ liệu quan hệ và dữ liệu đa phương tiện) để thích nghi với việc biến đổi hệ thống giao thông trong quá trình phát triển. Hệ thống tích hợp các số đo liên u hybrid này để mô phỏng để thích nghi (học để áp dụng) cho các bài toán quản lý khác, chẳng hạn như: quản lý vận tải, quản lý hành chính thành phố.

1 Giới thiệu

1.1 Ngữ cảnh

Tại Thành phố Đà Nẵng, đơn vị trực tiếp quản lý hệ thống giao thông là Công ty quản lý cầu đường, trực thuộc Sở Giao thông vận tải. Hiện tại, Công ty chủ yếu thực hiện lưu trữ dữ liệu hệ thống giao thông trên giấy (hồ sơ hoàn công, hồ sơ quản lý công trình), một số dữ liệu rời rạc (nhờ bên ngoài hoàn công công trình đường, cầu...) trên máy tính và một số dữ liệu được số hóa sử dụng công nghệ GIS (quản lý Hồ Chí Minh), các file hình ảnh và clip... phục vụ công tác quản lý. Tuy nhiên, dữ liệu số hóa này lại phân tán, chưa đầy đủ và chính xác, chưa được cập nhật thường xuyên. Vì vậy, việc khai thác phần dữ liệu số hóa này còn những hạn chế nhất định. Việc cập nhật, tìm kiếm, thống kê, khai thác... gặp nhiều khó khăn, chi phí, thời gian và nhân lực rất lớn.

Ngoài ra, dữ liệu hệ thống giao thông của một Thành phố rất lớn và phức tạp, gồm hệ thống đường bộ, đường thủy và các công trình khác gắn liền trên đó. Hơn nữa, các dữ liệu này liên tục biến đổi do việc xây dựng mới, cải tạo, sửa chữa... Sự phát triển nhanh chóng của hệ thống công trình hệ thống kết cấu đã đặt ra yêu cầu về hệ thống các phương pháp quản lý thông tin một cách khoa học, chính xác cho công tác quản lý, điều hành các công trình hệ thống kết cấu giao thông. Vì vậy, việc số hóa dữ liệu hiện trạng công trình hệ thống kết cấu giao thông là rất cần thiết và là công cụ hữu ích trong việc tích hợp các số đo liên u gắn kết toàn bộ hệ thống công

trình hệ thống quản lý trên địa bàn thành phố (kết hợp các công trình ngầm); hỗ trợ cho việc quản lý tập trung và khai thác hiệu quả cơ sở dữ liệu, nâng cao hiệu quả công tác quản lý, Giám sát giám tra cứu, tìm kiếm, so sánh số liệu và hiển thị; Tối ưu hóa thời gian, công sức thẩm định, thanh kiểm tra vì phạm công trình hệ thống quản lý do được cung cấp thông tin nhanh chóng, đầy đủ và khách quan.

□□ Xét trên khía cạnh khoa học, dữ liệu lý tưởng hệ thống dữ liệu, chúng ta phải bắt nguồn từ bản chất của dữ liệu hiển thị. Tuy nhiên, chúng ta có những hướng nhìn khác nhau về bản chất dữ liệu và do đó những giới pháp riêng biệt cho từng hướng nhìn này làm cho hệ thống trở nên quá phức tạp dữ liệu. Một cách tổng quát, chúng ta có thể xem xét dữ liệu theo ba hướng tiếp cận như đây.

□□ - Tiếp cận hệ thống thông tin địa lý (GIS): Hướng tiếp cận này rất hữu hiệu cho việc thực hiện các dự án địa lý trong mạng lưới giao thông, gồm các tuyến đường, công nghệ, các công trình phố và toàn bộ thông tin về tình trạng của môi trường địa lý này. Tuy nhiên, việc tích hợp các thông tin về hệ số công trình, hoàn công và các thông tin chi tiết khác thì hệ thống GIS sẽ gặp khó khăn xử lý.

□□ - Tiếp cận cơ sở dữ liệu quan hệ (Relational Database): Hướng tiếp cận này rất tốt cho việc thực hiện dữ liệu doanh nghiệp và hoàn toàn phù hợp với việc quản lý thông tin về hệ số và việc chi tiết toàn bộ quá trình xử lý của các dự án giao thông. Tuy nhiên, hướng tiếp cận này sẽ gặp vấn đề khi hệ thống trở nên phức tạp, đặc biệt là việc xây dựng thêm các tuyến đường mới, các cây cầu mới kèm với các chi tiết hệ số, xây dựng và kiểm tra kèm theo.

□□ - Tiếp cận cơ sở dữ liệu đồ thị (Graph Database): Đây là hướng tiếp cận đột phá (phá vỡ) thực hiện dữ liệu giao thông. Về bản chất khoa học, hệ thống giao thông cũng là một đồ thị. Đồ thị là một đối tượng toán học tổng quát để giải quyết các bài toán giao thông. Đồ thị thể hiện một hệ thống giao thông sẽ hoàn toàn thích nghi được với mọi thay đổi thực tế của hệ thống.

□□ Bài toán đặt ra của đề tài là một giới pháp thực hiện toàn bộ dữ liệu giao thông thành hệ thống dữ liệu hybrid. Hệ thống này sẽ tích hợp 3 hướng tiếp cận khác nhau (từng hướng về bản

chức năng của dữ liệu hiện tại) để thích nghi với việc thay đổi giao thông trong quá trình phát triển. Hệ thống tiếp cận dữ liệu hybrid này đem lại nhiều ưu điểm thích nghi (hỗ trợ các áp dụng) cho các bài toán quản lý khác, chẳng hạn như: quản lý vận tải, quản lý hành chính công sở ...

1.2 Bài toán cần giải quyết

Mục tiêu của đề tài này là việc kết hợp các hệ thống tiếp cận khoa học để với dữ liệu để tạo chức năng toàn bộ dữ liệu hiện tại thành một thể thống nhất. Hệ thống áp dụng của đề tài này là số hóa dữ liệu hệ thống kết thu thập các công trình giao thông tại thành phố Đà Nẵng; cụ thể:

- Nghiên cứu tính đặc thù của mỗi loại dữ liệu trong toàn bộ hệ thống dữ liệu và giao thông hiện tại. Mỗi loại dữ liệu sẽ có hệ thống tiếp cận cụ thể để xử lý.

- Nghiên cứu khả năng tích hợp các loại dữ liệu với nhau. Thực tế là mỗi loại dữ liệu với đặc tính chức năng thành một thể thống nhất, tuy nhiên có mối liên hệ với nhau. Do đó, chúng ta cần tạo chức năng các loại dữ liệu này thành một thể thống nhất để đem lại nhiều ưu điểm quản lý và thích nghi khi có sự thay đổi.

- Thiết kế mô hình các sơ đồ lưu hybrid cho hệ thống kết thu thập các công trình giao thông, gồm các thông số kết thu thập các công trình đường bộ (các nút, các tuyến giao thông), các cấu trúc và xây dựng bản đồ mạng lưới giao thông.

2 Các sơ lý thuyết và các công trình liên quan

Đề tài thu thập lĩnh vực ứng dụng công nghệ thông tin vào công tác quản lý bằng cách xây

Dòng mô tả phần mềm quản lý. Hiện tại, các phần mềm quản lý thông tin được chia thành hai dòng: dòng thống nhất xoay quanh nghiệp vụ ít thay đổi theo thời gian (như quản lý bán hàng, quản lý tác vụ, ..), dòng thống nhất liên quan đến các hoạt động mà nhu cầu xử lý thay đổi theo thời gian và tầm chí liên quan đến các loại dữ liệu khác nhau (ví dụ: phần mềm quản lý việc khám chữa bệnh và chăm sóc sức khỏe)

Hiện nay, hệ thống giao thông ngày càng phát triển đi theo kịp với sự phát triển của xã hội. Vì thế, phần mềm quản lý hệ thống giao thông luôn phải được nâng cấp theo thời gian. Điều này dẫn đến việc phát sinh các chức năng mới của hệ thống. Đây là thách thức đầu tiên đối với việc xây dựng phần mềm quản lý dòng này.

Bên cạnh hệ thống giao thông còn làm việc với nhiều loại dữ liệu khác nhau. Ví dụ: đối với việc lưu trữ và xử lý bản đồ, chúng ta cần hệ thống dữ liệu GIS; đối với dữ liệu mô tả thông tin kết nối các công trình, chúng ta cần hệ thống các số dữ liệu quan hệ; đối với số liệu quan trọng về cảm biến hoặc camera, chúng ta cần xử lý dữ liệu thời gian thực... Do đó, việc tích hợp hệ thống các số dữ liệu mềm dẻo để có thể tích hợp tất cả các loại dữ liệu trên là một thách thức hiện nay.

Đề tài này hướng tới việc “Xây dựng mô hình dữ liệu hybrid với sự kết hợp giữa tất cả dữ liệu có thể có với các số dữ liệu đa thể”. Các số dữ liệu này giúp cho phần mềm có khả năng dễ nâng cấp và bổ sung thêm chức năng để đáp ứng với nhu cầu mới.

Hiện tại, nhóm nghiên cứu đang tham gia phát triển hệ thống mô hình hoá kiến trúc thông tin dựa lý (GIS Agent-based Modeling Architecture, gọi tắt là GAMA [Grignard et al]). Hệ thống GAMA đã đưa vào ứng dụng và đang được sử dụng rộng rãi để mô hình hoá. Cũng thế, GAMA là một hướng tiếp cận cho phép mô phỏng các tác tử trong đó GAMA cho phép tương tác và chuyển số các đối tượng thông tin dựa lý.

3 Đề xuất phương pháp

3.1 Hình thức cấu trúc đồ thị

Đồ thị vô hướng hoặc đồ thị G là một cặp không có thứ tự (unordered pair) $G=(V, E)$, trong đó:

- V , tập các đỉnh hoặc nút,

- E , tập các cặp không thứ tự của các đỉnh phân biệt, đôi một gọi là cạnh. Hai đỉnh thuộc một cạnh đôi một gọi là các đỉnh đầu cuối của cạnh đó.

Các thao tác trên đồ thị :

Các phép toán đồ thị

- Đồ thị đường (Line graph) (tập đồ thị mới bằng cách chuyển cạnh thành đỉnh và tập các cạnh thành đường)

- Đồ thị đối ngẫu (Dual graph) (tập đồ thị mới từ một đồ thị phẳng bằng cách tập một đỉnh cho mỗi miền mặt phẳng và các cạnh đôi một nối giữa hai đỉnh thuộc hai miền kề nhau)

- Đồ thị bù (Complement graph)

• Các phép toán hai ngôi

• Tích Descartes các đồ thị (Cartesian product of graphs)

• Tích Tensor các đồ thị (Tensor product of graphs)

• Các suy rộng

• Trong siêu đồ thị (hypergraph), một tập con có thể chứa nhiều hơn hai đỉnh.

• Một đồ thị vô hướng có thể được coi là một phức đơn hình (simplicial complex) bao gồm các đơn hình 1 chiều (các cạnh) và các đơn hình 0 chiều (các đỉnh). Như vậy, đa hình là suy rộng của đồ thị do chúng cho phép các đơn hình nhiều chiều hơn.

• Mọi đồ thị đều cho một matroid, nhưng nói chung, không thể tạo ra đồ thị từ matroid của nó, do đó, matroid không phải là suy rộng của đồ thị.

• Trong lý thuyết mô hình (model theory), một đồ thị chỉ là một cấu trúc. Nhưng khi đó, không có giới hạn về số cạnh: nó có thể là một số đếm bất kỳ.

3.2 Giải pháp cơ sở dữ liệu để thực

Định nghĩa:

a. Giải pháp cơ sở dữ liệu để thực là một đơn vị dữ liệu

b. Giải pháp nối hai đơn vị cơ sở dữ liệu để thực là liên kết giữa hai đơn vị dữ liệu đó.

c. Trường hợp cơ sở dữ liệu để thực chính là mối quan hệ giữa các liên kết dữ liệu giữa hai đơn vị này.

Các toán tử đối với dữ liệu để thực

a. Toán tử thêm: Thêm dữ liệu vào cơ sở dữ liệu: thêm một đơn vị vào để thực mà đơn vị này sẽ nối với một đơn vị bên trong để thực.

b. Toán tử sửa: Sửa dữ liệu trong cơ sở dữ liệu: thay đổi giá trị của một đơn vị cơ sở dữ liệu

c. Toán tử xóa: Xóa một đơn vị trong cơ sở dữ liệu: xóa một đơn vị trong để thực và các liên kết nối với đơn vị đó. Việc này sẽ ảnh hưởng đến toàn bộ để thực và có thể sẽ mất dữ liệu.

d. Toán tử tìm kiếm: Tìm kiếm một đối tượng trong cơ sở dữ liệu: duyệt qua các đối tượng của đối tượng tìm kiếm đối tượng theo mẫu yêu cầu.

e. Toán tử truy vấn: Truy vấn dữ liệu trong cơ sở dữ liệu để xây dựng một báo cáo: tìm kiếm đi ngược lại để tìm kiếm các đối tượng theo mẫu để kiểm tra các thuộc tính.

3.3 Chương minh tính toán trên cơ sở dữ liệu:

Ở đây, chúng ta cần chương minh rằng mọi dữ liệu trong hệ thống phải truy xuất được. Nghĩa là chúng ta có thể tìm thấy tất cả các dữ liệu có trong cơ sở dữ liệu.

Chương minh bằng phần chứng:

a. Giả sử tồn tại một đối tượng trong dữ liệu mà không thể truy vấn được

b. Suy ra tồn tại một đối tượng của đối tượng không có liên kết với các đối tượng khác.

c. Nghĩa là vi phạm phép toán thêm dữ liệu vào đối tượng.

d. Vì theo vi phạm tính liên thông của đối tượng

Vậy, mọi dữ liệu bên trong các hệ thống lưu trữ đều có thể truy vấn được.

3.4 Giải pháp tích hợp các hệ thống lưu trữ hybrid

Hiện tại, đang có nhiều các hệ thống lưu trữ đang được sử dụng tại đơn vị nghiên cứu. Nếu chúng ta có thể chuyển tất cả dữ liệu này vào các hệ thống lưu trữ thì mọi thứ xem như được tích hợp quy mô toàn diện. Tuy nhiên, quá trình chuyển đổi dữ liệu (data migration) là một quá trình phức tạp và tốn thời gian vì các lý do:

- Tất cả các hệ thống lưu trữ đều phức tạp

- Các các hệ thống lưu trữ khác nhau thì không đồng bộ nhau (có hệ thống a, có hệ thống liên kết)

- Trong quá trình chuyển đổi, sẽ có nhiều dữ liệu được thêm vào

Vì thế, bất kỳ một giải pháp nào cũng có một giai đoạn quá độ mà tất cả các các hệ thống lưu trữ cùng tồn tại song song. Tuy nhiên vẫn phải có thêm một các hệ thống lưu trữ (ở đây các hệ thống lưu trữ được tích hợp) liên kết tất cả các hệ thống lưu trữ này lại. Hệ thống gồm tất cả các các hệ thống lưu trữ đang có kết hợp với các hệ thống lưu trữ mới là các hệ thống lưu trữ hybrid.

Các hệ thống lưu trữ hybrid sẽ hoạt động theo các tiêu chí sau:

• Kết nối tất cả cơ sở dữ liệu vào cơ sở dữ liệu đích.

• Mọi truy vấn được thông qua cơ sở dữ liệu đích trước tiên. Nếu dữ liệu cần truy vấn có trên đích thì sẽ dùng. Nếu không, câu truy vấn sẽ được chuyển đến các cơ sở dữ liệu liên quan.

• Liên tục chuyển thông tin từ các cơ sở dữ liệu liên quan vào bên trong cơ sở dữ liệu đích.

• Mọi thông tin thêm vào đích thêm vào cơ sở đích và các cơ sở dữ liệu liên quan.

• Một khi toàn bộ dữ liệu của một cơ sở dữ liệu nằm trong cơ sở dữ liệu đích, cơ sở dữ liệu hybrid sẽ không truy vấn với cơ sở dữ liệu đó nữa.

4 Đánh giá

4.1 Đánh giá định tính

Đánh giá định tính trong tình huống này trình bày kết quả khi thực hiện các câu truy vấn. Một câu truy vấn sẽ được thực hiện và trả về danh sách các nút và các cạnh liên quan. Hình 1 trình bày kết quả của ví dụ truy vấn các cạnh quan và các cá nhân cùng với các hình sơ đồ pháp lý liên quan đến công trình đường Lý Thường Kiệt. Bên cạnh các nút chứa thông tin liên quan đến công tác quản lý thì các tuyến đường được nối với nhau bằng công trình được thực hiện ngay trên nó. Với mỗi công trình thì được liên kết đến các nút liên quan đến các Công Quan và Cá nhân có liên quan. Đường thì các công trình còn được liên kết với nhau bằng Hệ Sơ Đồ Pháp Lý được chia thành 12 mô hình trong phần Khái quát về dữ liệu đã đề cập.

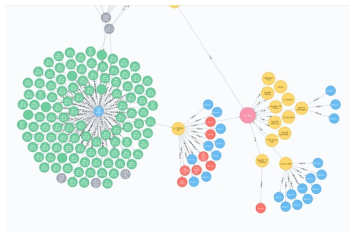


Fig. 1. Kết quả truy vấn hệ số pháp lý của một tuyến đường

Chúng tôi thực hiện rất nhiều câu truy vấn khác nhau liên quan đến việc quản lý nghiệp vụ, tất cả đều được kết quả theo yêu cầu. Ví dụ về các câu truy vấn bao gồm: liệt kê tất cả các nút đường; liệt kê tất cả thông tin liên quan đến một tuyến đường; liệt kê hệ số pháp lý của một tuyến đường; liệt kê nhật ký báo cáo đường của một tuyến đường; liệt kê nhật ký tuồn tra của một tuyến đường;

4.2 Đánh giá hiệu năng

Thực tế, chúng tôi thực hiện việc đánh giá trên một server chuyên dành riêng cho cơ sở dữ liệu. Server này có bộ CPU core i7 2,4Ghz, bộ nhớ 16GB RAM, ổ cứng SSD 156GB. Chúng tôi thực hiện việc truy vấn tất cả các nút trong cơ sở dữ liệu.

Thực nghiệm này chúng tôi bắt đầu đo tốc độ đọc và dữ liệu bản đồ quản lý Hồ Chí Minh. Dữ liệu này gồm 2550 nút đường đi để cho dữ liệu địa lý thể hiện các điểm giao nhau của các tuyến đường. Với mỗi nút, chúng tôi giới thiệu thêm 1000 nút đường đi để cho dữ liệu quan hệ giữa dữ liệu địa lý.

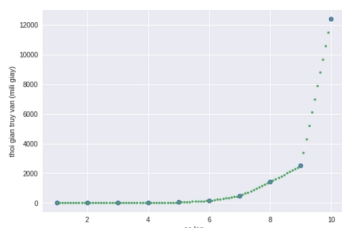


Fig. 2. Thời gian thực hiện truy vấn toàn bộ dữ liệu

Mỗi trường hợp thực nghiệm tiếp theo, chúng tôi nhân số lượng dữ liệu của các nút dữ liệu này lên 2 lần, 3 lần, ... 10 lần. Với mỗi trường hợp, chúng tôi thực hiện truy vấn 100 lần và lấy kết quả trung bình. Hình trên mô tả thời gian truy vấn đối với dữ liệu và thời gian trường hợp thực nghiệm.

Kết quả cho thấy rằng, chúng ta nhân bộ dữ liệu lên 8 lần thì thời gian truy vấn vẫn dưới 2 giây. Nghĩa là nếu chúng ta mở rộng thành phần ra gấp 8 lần thì thời gian truy vấn vẫn dưới 2 giây. Đây là điều chúng ta cần. Tuy nhiên, nếu chúng ta nhân dữ liệu lên gấp 10 lần thì thời gian truy vấn sẽ nhích lên 12 giây. Điều này dẫn đến việc hình ảnh là phần nghiên cứu song song hoá dữ liệu để có thể truy vấn kịp thời.

5 Kết luận

5.1 Kết quả đạt được

Trước hết, đề tài này đã đưa xuất một giải pháp xây dựng cơ sở dữ liệu hybrid để thống nhất các dạng dữ liệu trong hệ thống quản lý giao thông. Cơ sở dữ liệu này sẽ được dữ liệu để thực hiện kết nối với tất cả các loại dữ liệu khác như dữ liệu GIS, dữ liệu quan hệ, và các loại dữ liệu khác.

Đề tài này còn đưa xuất một giải pháp kết hợp tầm thời điểm với hệ thống hiện tại đó là hệ thống hybrid. Hệ thống hybrid là sự kết hợp tầm thời gian của cơ sở dữ liệu hybrid và tất cả các hệ thống phần mềm khác. Hệ thống này sẽ tìm cách sao lưu dữ liệu từ các nguồn vào cơ sở dữ liệu để thực hiện có thể thực hiện truy vấn, thống kê và báo cáo.

5.2 Hướng phát triển:

Trước hết, việc chuyển đổi là phải nhập dữ liệu từ các nguồn dữ liệu vào trong cơ sở dữ liệu hybrid. Đây là một việc đòi hỏi nhân lực và thời gian. Ngoài ra, việc này còn yêu cầu một giao diện nhập liệu thân thiện giúp cho người sử dụng nhập liệu. Điều này mở ra một hướng xây dựng phần mềm hoàn thiện đúng quy trình.

Tiếp theo là hướng phát triển đối với dữ liệu lớn. Chúng ta đưa biết rằng dữ liệu giao thông rất phức tạp và đòi hỏi các phần liên tục. Do đó, dữ liệu sẽ ngày càng nhiều và số người tìm kiếm cũng có một hướng song song để việc nhập và báo cáo được kịp thời.

Tiếp đến là hướng phát triển chia sẻ. Dữ liệu giao thông là dữ liệu xuyên ngành hỗ trợ cho tất cả các ngành khác. Do đó, dữ liệu này cần phát triển thành dạng dịch vụ đám mây để có thể chia sẻ cho các đơn vị khác và các ngành khác cùng khai thác số liệu.

Cuối cùng là hướng phát triển an toàn dữ liệu. Ở đây có hai yếu tố an toàn: bảo mật dữ liệu và bảo vệ dữ liệu. Bảo mật nghĩa là không để đơn vị nào có quyền khai thác một số thông tin của họ nào đó. Điều này mở ra hướng nghiên cứu về phân quyền để việc sử dụng phần mềm mang tính thành phố. Yếu tố tiếp theo là bảo vệ dữ liệu. Nghĩa là dữ liệu cần được đảm bảo là không thể thay thế và giả mạo. Điều này mở ra hướng nghiên cứu về công nghệ block-chain cho việc lưu trữ dữ liệu quy trình, khi mà dữ liệu chỉ nên ghi thêm chứ không nên được xóa.

Tóm tắt: khoa học dữ liệu; cơ sở dữ liệu; dữ liệu; truy vấn; truy vấn; truy vấn.

Chức nhiệm đề tài: Th.S Ngô Văn Quý