

XÂY DỰNG BẢN ĐỒ ĐIỆN TỬ SỬ DỤNG SQL SERVER SPATIAL

Nguyễn Thị Tuyết

Trường Đại học Quy Nhơn

Email: nguyenthituyet@qnu.edu.vn

Ngày nhận bài: 20/02/2023; Ngày nhận đăng: 01/06/2023

Tóm tắt

Cơ sở dữ liệu không gian là một lĩnh vực đã và đang được các nhà khoa học tập trung nghiên cứu và đưa ra nhiều ứng dụng trong thực tiễn. Nó giải quyết bài toán quản lý và phân tích dữ liệu không gian như Hệ thống thông tin địa lý (GIS). Bên cạnh đó, hệ quản trị cơ sở dữ liệu không gian (SDBMS – Spatial Database Management System) là một hệ quản trị cơ sở dữ liệu, có hỗ trợ mô hình dữ liệu không gian, kiểu dữ liệu không gian (SDT – Spatial Data Types) và ngôn ngữ truy vấn SQL không gian. Trong bài báo này, chúng tôi kết hợp cơ sở dữ liệu không gian và sử dụng SQL Server Spatial để truy vấn dữ liệu không gian, truy xuất bản đồ điện tử của Việt Nam và các tỉnh thành trong nước.

Từ khóa: Cơ sở dữ liệu không gian, hệ quản trị cơ sở dữ liệu không gian, truy vấn không gian, kiểu dữ liệu không gian.

Designing electronic maps based on SQL Server Spatial

Nguyen Thi Tuyet

Quy Nhơn University

Received: February 20, 2023; Accepted: June 01, 2023

Abstract

Spatial database is a field that scientists have been focusing on researching and offering many practical applications. It solves the problems of managing and analyzing spatial data such as Geographic Information Systems (GIS). Besides, the Spatial Database Management System (SDBMS – Spatial Database Management System) is the one which supports the spatial data models, spatial data types (SDT - Spatial Data Types) and spatial SQL query languages. In this paper, spatial database and SQL Server Spatial are combined to query spatial data and to retrieve electronic maps of Vietnam and of domestic provinces.

Keywords: Spatial Database, Spatial Database Management System, Spatial Query, Spatial Data Types.

1. Mở đầu

Mặc dù cơ sở dữ liệu truyền thống đã phát triển mạnh trong nhiều năm qua, nhưng việc lưu trữ và quản lý các loại dữ liệu phức tạp như dữ liệu không gian vẫn

còn nhiều thách thức. Dữ liệu không gian biểu diễn dữ liệu đa chiều với các điểm, đường thẳng và đa giác; một số cơ sở dữ liệu không gian xử lý các cấu trúc phức tạp hơn như đối tượng 3D, đường trung bình

tôpô, mạng tuyến tính,...Cùng với sự phát triển của công nghệ thông tin, các hệ thống quản lý cơ sở dữ liệu không gian đã làm cho việc quản lý dữ liệu trở nên dễ dàng và tự nhiên hơn đối với người dùng qua các ứng dụng như quy hoạch đô thị, giao thông và viễn thám.(Philippe Rigaux & cs)

Với cơ sở dữ liệu không gian khi áp dụng với bất kỳ ngôn ngữ hay hệ quản trị nào thì việc đánh chỉ mục (index) cho loại dữ liệu này là khá quan trọng, nó giúp ích cho việc tối ưu hóa truy vấn dữ liệu, giảm thời gian truy vấn và giảm bộ nhớ lưu trữ... Trong bài báo này, chúng tôi tìm hiểu về cơ sở dữ liệu không gian kết hợp với hệ quản trị cơ sở dữ liệu Microsoft SQL Server để xây dựng bản đồ điện tử và các truy vấn trên bộ dữ liệu không gian mà chúng tôi đã tìm hiểu.

2. Kiến thức liên quan

2.1. Tổng quan về cơ sở dữ liệu không gian

Xuất phát từ nhu cầu công việc và xử lý dữ liệu địa lý như điểm, đường, đa giác và các đối tượng hình học đại diện cho các đối tượng trong không gian của thế giới thực đã hình thành loại cơ sở dữ liệu không gian.

Cơ sở dữ liệu không gian là một cơ sở dữ liệu trong đó có chứa đối tượng không gian. Đối tượng không gian được biểu diễn thông qua các đối tượng hình học cơ bản (điểm, đường và vùng) hoặc được biểu diễn dạng ảnh (raster). Dữ liệu không gian biểu diễn thông tin vị trí địa lý và hình dạng của các đối tượng trên trái đất.

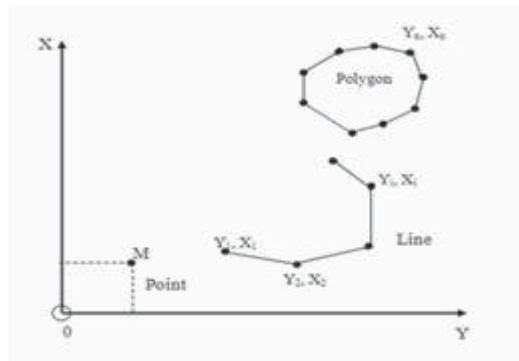


Hình 1. Minh họa CSDL không gian

Dữ liệu không gian có hai mô hình lưu trữ: **mô hình dữ liệu raster** và **mô**

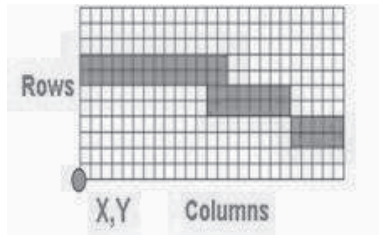
hình dữ liệu vector.

Mô hình dữ liệu Vector: thông tin về điểm, đường, vùng (đa giác) được mã hóa và lưu dưới dạng tập hợp các tọa độ x,y . Một điểm đại diện cho một đối tượng lưu trữ dưới dạng tọa độ (x,y) hoặc (x,y,z) trong không gian. Đối tượng dạng đường như đường giao thông, sông suối... được lưu trữ dưới dạng tập hợp các tọa độ điểm $x_1y_1, x_2y_2, \dots, x_ny_n$ hoặc là một hàm toán học, tính được chiều dài. Đối tượng dạng vùng như khu vực buôn bán, nhà cửa,... được lưu như một vòng khép kín của các điểm tọa độ, tính được chu vi và diện tích vùng (James Gray, 2016).

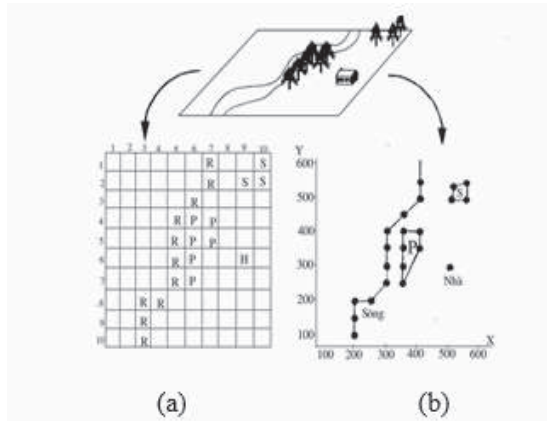


Hình 2. Biểu diễn thông tin điểm, đường, vùng theo cấu trúc vector

Mô hình dữ liệu Raster: Trong cấu trúc dữ liệu Raster, đối tượng được biểu diễn thông qua các ô (cell) hay điểm ảnh (pixel) của một lưới các ô. Trong máy tính, các ô lưới này được lưu trữ dưới dạng ma trận, trong đó mỗi ô lưới là giao điểm của một hàng và một cột trong ma trận. Điểm được xác định bởi một pixel (giá trị nhỏ nhất trong cấu trúc Raster), đường được xác định bởi một chuỗi các ô có cùng thuộc tính kề nhau có hướng nào đó, còn vùng được xác định bởi một số các pixel cùng thuộc tính phủ lên trên một diện tích nào đó (James Gray, 2016).



Hình 3. Biểu diễn thông tin điểm, đường, vùng theo cấu trúc Raster



Hình 4. Thể giới thực trên hai mô hình raster (a) và vector (b)

2.2. Ứng dụng của CSDL không gian

- Quy hoạch đô thị:

Quy hoạch đô thị đòi hỏi phải phân tích dữ liệu để mô hình hóa. Việc sử dụng cơ sở dữ liệu không gian để phân tích giao thông, phân tích sử dụng đất, phân tích ô nhiễm,... giúp ta có thể lập kế hoạch để sử dụng tốt hơn các tài nguyên và dịch vụ trong quy hoạch đô thị (N. R. Adam and A. Gangopadhyay, 1997).

- Các hoạt động quân sự:

Dữ liệu không gian có tầm quan trọng rất lớn đối với hoạt động quân sự vì nó đưa ra quyết định trong việc lập kế hoạch, quản lý các căn cứ quân sự nhằm nâng cao hiệu quả nhiệm vụ chính trị, phân tích các chiến lược thay thế và lưu trữ kiến thức thể chế (N. R. Adam and A. Gangopadhyay, 1997).

- Nuôi trồng:

Cơ sở dữ liệu không gian rất hữu ích trong việc lập bản đồ và mô tả những thay

đổi hiện tại và tương lai về lượng mưa, nhiệt độ, sản xuất cây trồng... Bằng cách lập bản đồ các đặc điểm địa lý và địa chất, các nhà khoa học và nông dân có thể làm việc cùng nhau trong việc tạo ra các kỹ thuật canh tác hiệu quả và năng suất hơn (N. R. Adam and A. Gangopadhyay, 1997).

- Dự báo thời tiết:

Cơ sở dữ liệu không gian là thành phần quan trọng trong các hệ thống xử lý thời tiết, theo đó các hình ảnh vệ tinh và địa hình có thể được mô hình hóa - hỗ trợ rất nhiều cho kỹ năng của người dự báo thời tiết (N. R. Adam and A. Gangopadhyay, 1997).

2.3. Đặc trưng của CSDL không gian

Cơ sở dữ liệu không gian sử dụng đánh chỉ mục không gian để tăng tốc hoạt động của cơ sở dữ liệu.

Ngoài các truy vấn SQL điển hình như câu lệnh SELECT, CSDL không gian có thể thực thi đa dạng các thao tác không gian như:

- Các phép đo không gian (Spatial measurements): đo khoảng cách, diện tích và thể tích...

- Hàm không gian (Spatial function): thực hiện phân tích không gian; kết hợp các tệp không gian với dữ liệu ở các định dạng khác,... Ví dụ hàm `DISTANCE(<Spatial Point1>,<Spatial Point2>,"<units>")` Trả về số đo khoảng cách giữa hai điểm Point1, Point2 trong units. Trong đó units là mét (m), kilômét (km), dặm (mi), bộ (ft).

- Xác thực không gian (Spatial validation): nhằm đánh giá độ chính xác của bản đồ.

- Hàm tạo (Create features): tạo ra các hình mới, như tìm ra các điểm nút có thể tạo nên đường, hay nếu điểm đầu và điểm cuối trùng nhau, chúng có thể tạo nên một đa giác.

- Hàm theo dõi: các câu truy vấn trả

về thông tin cụ thể như: vị trí tâm của một đường tròn hay điểm đầu, điểm cuối của một đường.

Hiện nay có rất nhiều hệ quản trị CSDL không gian như: DB2 (phần mở rộng IBM DB2 Spatial Extender), Oracle (Oracle Spatial), PostgreSQL (phần mở rộng PostGIS), Sqlite (phần mở rộng SpatialLite) và Microsoft SQL Server phiên bản 2008 trở lên.

2.4. SQL Server Spatial

Hệ quản trị cơ sở dữ liệu không gian (SDBMS – Spatial Database Management System) là một hệ quản trị cơ sở dữ liệu có hỗ trợ mô hình dữ liệu không gian, kiểu dữ liệu không gian (SDT – Spatial Data Types) và ngôn ngữ truy vấn SQL không gian. Cơ sở dữ liệu không gian là một cơ sở dữ liệu trong đó có chứa đối tượng không gian. Đối tượng không gian có được biểu diễn thông qua các đối tượng hình học cơ bản (điểm, đường và vùng) hoặc được biểu diễn dạng ảnh (raster). Dữ liệu không gian biểu diễn thông tin vị trí địa lý và hình dạng của các đối tượng trên trái đất (N. R. Adam and A. Gangopadhyay, 1997).

SDBMS cung cấp các tính năng sau:

- Tập hợp các kiểu dữ liệu không gian để biểu thị các kiểu dữ liệu không gian nguyên thủy (điểm, đường, diện tích), các kiểu dữ liệu không gian phức tạp (như đối tượng 3D, đường trung bình tậpô, mạng tuyến tính,...) và các thao tác trên các kiểu dữ liệu này như phép đo không gian (tính độ dài đường thẳng, diện tích đa giác, khoảng cách giữa các hình học,...), chức năng không gian (ví dụ như sửa đổi các đối tượng địa lý hiện có để tạo các đối tượng địa lý mới), dự đoán không gian (cho phép truy vấn đúng/sai về mối quan hệ không gian giữa các hình), v.v.

- Hệ thống đã cung cấp các cải tiến hiệu suất như các chỉ mục để xử lý các truy

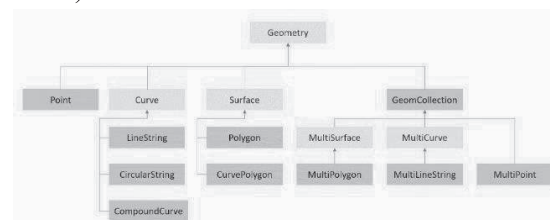
vấn không gian, xử lý song song, v.v..., có sẵn cho dữ liệu phi không gian.

Hiện nay có rất nhiều SDBMS như: DB2 (phần mở rộng IBM DB2 Spatial Extender), Oracle (Oracle Spatial), PostgreSQL (phần mở rộng PostGIS), Sqlite (phần mở rộng SpatialLite), và Microsoft SQL Server phiên bản 2008 trở lên. Trong SQL Server 2008 giới thiệu các 2 kiểu dữ liệu không gian là:

- Kiểu dữ liệu hình học (Geometry):

Là kiểu dữ liệu không gian cơ bản được sử dụng để biểu diễn các đối tượng trong hệ tọa độ Euclide (tọa độ phẳng), đối tượng không có thông tin về SRID (Spatial Reference Identifiers, đây là một hệ quy chiếu đã được định nghĩa trước ứng với một số nguyên).

- Kiểu dữ liệu địa lý (Geography): Là kiểu dữ liệu không gian cơ bản được sử dụng để biểu diễn các đối tượng trong hệ tọa độ trái đất (tọa độ Ellipsoid). Đối tượng có chứa thông tin về SRID (Brian Gough, 2010).



Hình 5. Cấu trúc phân cấp hình học dựa trên các kiểu dữ liệu hình học và địa lý.

Với sự hỗ trợ làm việc với các đối tượng không gian, người sử dụng Microsoft SQL Server 2008 có thể thực hiện việc lưu trữ dữ liệu không gian cũng như thực hiện truy vấn không gian mà các hệ quản trị cơ sở dữ liệu truyền thống (không chứa thành phần không gian – non spatial database) không thể thực hiện được. Tính năng này sẽ cung cấp cho các nhà quản lý trong các lĩnh vực liên quan đến yếu tố không gian một hệ quản trị cơ sở dữ liệu phù hợp, đặc biệt là

những nhà quản lý về tài nguyên, môi trường, nhà quản lý giao thông, các hệ thống điện - cấp thoát nước, nhà nghiên cứu về địa lý, cũng như các nhà quy hoạch phát triển kinh tế.

3. Xây dựng bản đồ điện tử trên SQL Server spatial

Bước 1: Cài đặt SQL Server Spatial Tools

Bước 2: Upload Shapefile vào SQL Server để tạo cơ sở dữ liệu không gian

Link cơ sở dữ liệu để vẽ bản đồ Việt Nam (<https://data.opendevlopmentmekong.net>).

Bước 3: Chuyển đổi kiểu dữ liệu hình học (Geometry) sang kiểu dữ liệu địa lý (Geography)

Trong bước này, chúng tôi sẽ sử dụng hàm biến đổi Affine từ các công cụ không gian của máy chủ SQL để mở rộng và định vị tọa độ, bởi vì hàm biến đổi affine chỉ áp dụng cho kiểu dữ liệu hình học, chúng ta cần chuyển đổi địa lý thành hình học, thực hiện chuyển đổi affine và chuyển đổi hình học trở lại địa lý.

Bước 4: Kết quả thu được

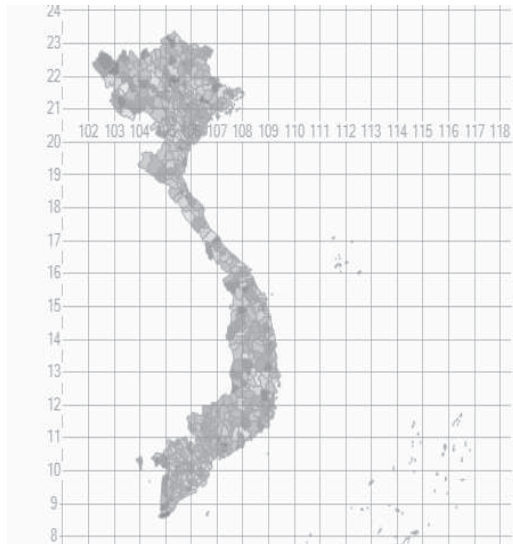
Bảng 1. Dữ liệu vẽ bản đồ

ID	OBJECTID	Province	Dist	Pop_2009	Code_p	Lat_Lon	Lat_DMS	Lon_DMS	Mac_Lon	Latitude	Longitude	geom
1	1	AD02	Ca Mau	66229	96370	01479.70'N, 1050227.71'E	01479.70'N	1050227.71'E	01050440.97048	8.78462991	105.04102963	0a00000000104
2	2	AD02	Ca Mau	180208	96370	01521.64'N, 1051641.69'E	01521.64'N	1051641.69'E	01052440.97025	8.97267746	105.24409242	0a00000000104
3	3	AD02	Ca Mau	123796	96369	01611.60'N, 105270.00'E	01611.60'N	105270.00'E	01053000.97071	9.3671029	105.30000474	0a00000000104
4	4	AD02	Bac Lieu	142990	95960	07222.98'N, 1052147.42'E	07222.98'N	1052147.42'E	01053600.97028	9.12394462	105.36017249	0a00000000104
5	5	AD02	Ca Mau	298196	96364	08481.94'N, 1051712.50'E	08481.94'N	1051712.50'E	01053600.97019	9.15850202	105.36000000	0a00000000104
6	6	AD02	Bac Lieu	142955	95964	07125.97'N, 1051294.07'E	07125.97'N	1051294.07'E	01053600.97028	9.15850202	105.36000000	0a00000000104
7	7	AD02	Bac Lieu	157248	95959	071548.24'N, 1052191.65'E	071548.24'N	1052191.65'E	01053600.97028	9.26239966	105.36017028	0a00000000104
8	8	AD02	Bac Lieu	106782	95961	071838.97'N, 1052843.89'E	071838.97'N	1052843.89'E	01054122.97770	9.27797070	105.41218379	0a00000000104
9	9	AD02	Bac Lieu	38159	95958	072191.99'N, 1054210.12'E	072191.99'N	1054210.12'E	01054200.97020	9.37096970	105.42000000	0a00000000104
10	10	AD02	Soc Trang	163000	94950	07247.98'N, 1055943.09'E	07247.98'N	1055943.09'E	01055796.97000	9.37687188	105.57969261	0a00000000104
11	11	AD02	Bac Lieu	147055	95954	07125.97'N, 1052948.07'E	07125.97'N	1052948.07'E	01054000.97020	9.35721270	105.40000000	0a00000000104
12	12	AD02	Qu. Tầm	89526	94924	04267.42'N, 1053711.07'E	04267.42'N	1053711.07'E	01054400.97020	4.44702999	105.37000000	0a00000000104

Có các kiểu thuộc tính sau:

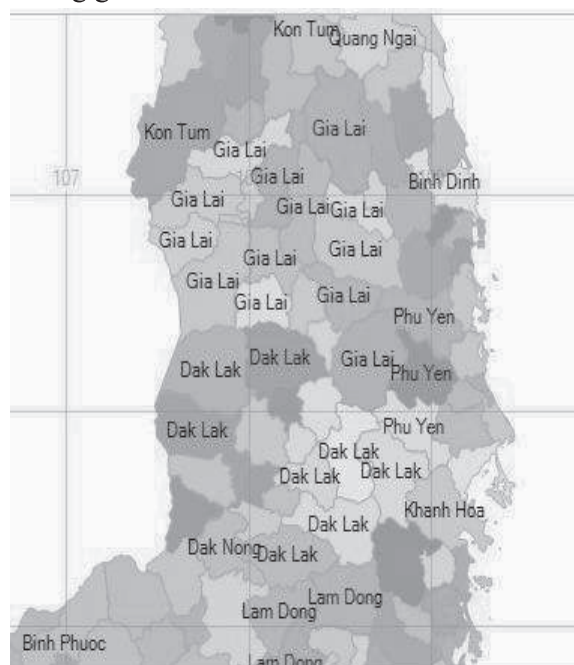
- ID, OBJECTID: Quy định số thứ tự của từng đối tượng và cũng được chọn làm đối tượng chỉ mục của bảng
- F_code: Mã phân biệt
- Province: Các tỉnh

- District: Các huyện
- Pop_2009: Dân số một vùng
- Code_re: Mã vùng
- Latitude, Longitude: Vĩ độ, Kinh độ
- Geom: Tọa độ được trả về dạng ngôn ngữ đánh dấu kiểu nhị phân của đối tượng hình học (WKB).

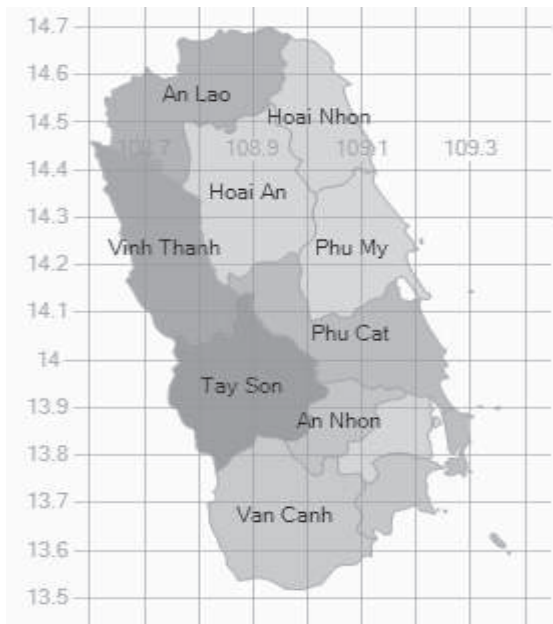


Hình 6. Truy xuất bản đồ Việt Nam

Bước 5: Các kết quả truy vấn trên dữ liệu không gian



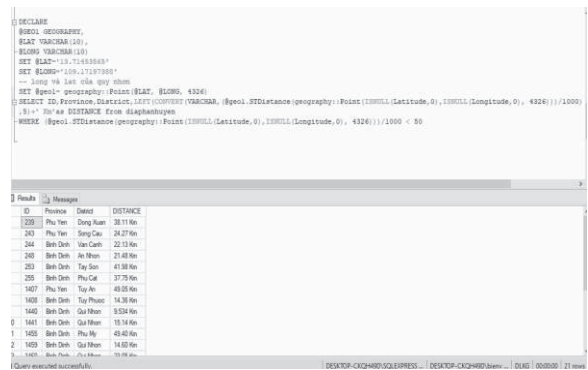
Hình 7. Bản đồ hiển thị địa phận tỉnh



Hình 8. Bản đồ hiển thị tỉnh Bình Định



Hình 9. Bản đồ hiển thị thành phố Quy Nhon



Hình 10. Khoảng cách từ thành phố Quy Nhon đi tới nơi khác giới hạn không vượt quá 50km.

3. Kết luận và hướng nghiên cứu trong tương lai

Trong bài báo này, chúng tôi đã tìm hiểu về cơ sở dữ liệu không gian, trình bày các bước để xây dựng bản đồ điện tử trên hệ quản trị cơ sở dữ liệu Microsoft SQL Server với bộ dữ liệu quốc gia Việt Nam. Thực nghiệm của chúng tôi có thể dùng để truy xuất ra một vùng hoặc nhiều vùng khác nhau trong bản đồ đã xây dựng.

Vì cơ sở dữ liệu không gian được ứng dụng rất nhiều trong thực tế nên trong thời gian tới, chúng tôi tiếp tục nghiên cứu cơ sở dữ liệu không gian, truy vấn dữ liệu để định vị các địa điểm trên bản đồ □

TÀI LIỆU THAM KHẢO

Philippe Rigaux, Michel Scholl, and Agnes Voisard, “Spatial Databases With Application to GIS”, Morgan Kaufmann Publishers.

N. R. Adam and A. Gangopadhyay, “Database Issues in Geographic Information Systems”. Vol. 6 of Advances in Database Systems. Kluwer Academic Publishers, New York, 1997.

Brian Gough, “The PostgreSQL 9.0 Reference Manual” - Volume 1A - SQL Language Reference: The PostgreSQL Global Development Group, 2010.

<https://data.opendevlopmentmekong.net/dataset/a-phn-tnh/resource/5969feda-a2a7-436fb140-54092f66328c>