

Bài báo nghiên cứu

**ỨNG DỤNG KỸ THUẬT LẬP CHỈ MỤC KHÔNG GIAN
TRONG XÂY DỰNG CƠ SỞ DỮ LIỆU KHOÁNG SẢN**Nguyễn Sách Thành^{1*}, Nguyễn Như Hùng¹, Đậu Thanh Bình²¹Trường Đại học Lê Quý Đôn, Việt Nam²Sở Tài nguyên và Môi trường Nghệ An, Việt Nam*Tác giả liên hệ: Nguyễn Sách Thành – Email: thanhsns.geo@lqdtu.edu.vn

Ngày nhận bài: 16-9-2020; ngày nhận bài sửa: 30-9-2020; ngày duyệt đăng: 25-12-2020

TÓM TẮT

Quản lý tài nguyên khoáng sản là một bài toán quan trọng trong chiến lược phát triển bền vững của mỗi quốc gia; trong đó, cơ sở dữ liệu không gian khoáng sản là một thành phần chính của hệ thống quản lý. Ngày nay, với sự phát triển của các công nghệ thu thập và xử lý thông tin, thì dữ liệu không gian về tài nguyên khoáng sản ngày càng lớn. Điều này, đã đặt ra cho bài toán truy vấn nhanh dữ liệu không gian với hàng triệu bản ghi trở nên phức tạp. Để giải quyết bài toán dữ liệu không gian lớn, kỹ thuật lập chỉ mục không gian thường được sử dụng. Trong bài báo này, nhóm tác giả sẽ ứng dụng kỹ thuật lập chỉ mục không gian trong xây dựng cơ sở dữ liệu khoáng sản trên PostGIS; đồng thời, đánh giá khả năng truy vấn không gian trong trường hợp sử dụng và không sử dụng chỉ mục. Kết quả cho thấy, thời gian truy vấn khi sử dụng chỉ mục không gian nhanh hơn rất nhiều.

Từ khóa: cơ sở dữ liệu tài nguyên khoáng sản; PostGIS; chỉ mục không gian

1. Giới thiệu

Việt Nam có nguồn tài nguyên khoáng sản phong phú, đa dạng với hơn 5000 điểm khai thác mỏ. Điều này đóng vai trò quan trọng trong việc phát triển kinh tế – xã hội đất nước (Vo, 2012). Trong đó, một số loại có trữ lượng lớn như bô xít (672,1 triệu tấn), apatit (0,778 triệu tấn), titan (15,71 triệu tấn), than (3,52 triệu tấn), đất hiếm (1,1 triệu tấn) và đá granit (15 tỉ m³)... Với nguồn tài nguyên khoáng sản phong phú như vậy, thì việc quản lý chúng trở thành một bài toán quan trọng trong chiến lược phát triển bền vững của quốc gia. Điều này càng trở nên cấp bách hơn khi mà cơ sở dữ liệu về tài nguyên khoáng sản ngày càng lớn; đòi hỏi cần phải có những giải pháp kỹ thuật hợp lý trong xây dựng cơ sở dữ liệu tài nguyên khoáng sản.

Cite this article as: Nguyen Sach Thanh, Nguyen Nhu Hung, & Dau Thanh Binh (2020). Application of spatial indexing techniques for mineral resource database. *Ho Chi Minh City University of Education Journal of Science*, 17(12), 2120-2129.

Năm 2010, 12 quốc gia thuộc châu Âu, bao gồm: Đan Mạch, Hà Lan, Bỉ... đã thành lập dự án EuroGeoSource. Đây là dự án xây dựng cơ sở dữ liệu không gian về tài nguyên khoáng sản như dầu, khí, than đá... Nguồn dữ liệu này được lưu trữ trên cơ sở dữ liệu không gian PostGIS. Với việc xây dựng chỉ mục không gian, người dùng có thể thực hiện truy vấn nhanh thông tin dữ liệu khoáng sản của EuroGeoSource thông qua hệ thống phân tán. Năm 2014, trong khuôn khổ chương trình hỗ trợ kỹ thuật của Nhật Bản cho các nước thành viên ASEAN thuộc lĩnh vực địa chất khoáng sản, Cục Địa chất Khoáng sản Nhật Bản đã xây dựng cơ sở dữ liệu không gian về tài nguyên khoáng sản của các quốc gia Đông Nam Á. Nguồn dữ liệu này bao gồm các thông tin về vị trí các mỏ khoáng sản, loại khoáng sản, trữ lượng dự kiến... Hệ thống dữ liệu đã tận dụng cách tạo chỉ mục không gian của PostGIS, nhằm đảm bảo việc truy vấn thông tin nhanh trong môi trường WebGIS. Ở Việt Nam, cũng có một số nghiên cứu về xây dựng cơ sở dữ liệu không gian tài nguyên khoáng sản. (Nguyen, Le, Pham, & Nguyen, 2009) đã xây dựng cơ sở dữ liệu địa chất khoáng sản, địa chất môi trường và tai biến địa chất các vùng biển Việt Nam. Cơ sở dữ liệu này được thiết kế, xây dựng trong hệ thống GIS phân tán nhằm cho phép người sử dụng tra cứu các dữ liệu (dạng bảng, dạng bản đồ, dạng ảnh...). Tuy nhiên, các nghiên cứu này chưa trình bày cụ thể về ứng dụng kỹ thuật chỉ mục không gian trong xây dựng cơ sở dữ liệu tài nguyên khoáng sản gian nhằm đảm bảo việc truy xuất nhanh đối với dữ liệu lớn.

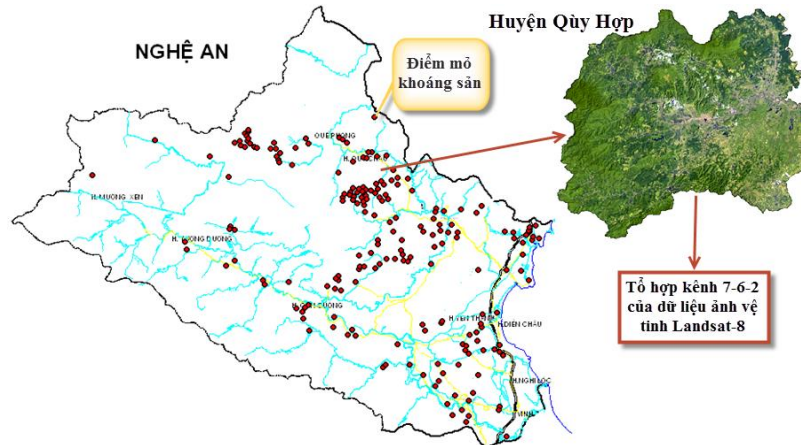
Với sự phát triển mạnh mẽ của các công nghệ tìm kiếm, thăm dò và các kỹ thuật phân tích, xử lý thông tin đã tạo ra nguồn dữ liệu không gian tài nguyên khoáng sản trở nên vô cùng lớn. Điều này đòi hỏi giải pháp thiết kế, xây dựng cơ sở dữ liệu này phải hợp lý (Du, 2009). Sự phát triển của các thuật toán về các cấu trúc dữ liệu không gian trên máy tính đã tạo ra các kỹ thuật xử lý nhanh việc truy xuất dữ liệu lớn. Một trong những kỹ thuật hay được sử dụng là tạo chỉ mục không gian. Chỉ mục giúp việc thao tác một cơ sở dữ liệu không gian lớn trở lên dễ dàng. Ngược lại, nếu không sử dụng chỉ mục, quá trình tìm kiếm sẽ thực hiện duyệt tuần tự tất cả các bản ghi. Điều này dẫn tới thời gian tìm kiếm lâu hơn và bộ nhớ máy tính cần phải lớn hơn (Ernest et al. 2015; Neelabh et al. 2015). Trong nghiên cứu này nhóm tác giả sẽ trình bày về kỹ thuật tạo chỉ mục không gian GiST cho cơ sở dữ liệu tài nguyên khoáng sản được lưu trữ trên PostGIS và đánh giá thời gian thực hiện truy vấn cơ sở dữ liệu lớn trong hai trường hợp: chỉ mục và không sử dụng chỉ mục.

2. Đối tượng và phương pháp nghiên cứu

2.1. Cơ sở dữ liệu tài nguyên khoáng sản

Cơ sở dữ liệu tài nguyên khoáng sản là một nguồn dữ liệu cốt lõi của hệ thống quản lý tài nguyên khoáng sản. Ngoài cơ sở dữ liệu tài nguyên khoáng sản còn có các nguồn thông tin dữ liệu khác như bản đồ địa chất, bản đồ tiềm năng khoáng sản; bản đồ quy hoạch thăm dò, khai thác, chế biến và sử dụng khoáng sản; bản đồ khu vực cấm và khu vực tạm cấm hoạt động khoáng sản; bản đồ vị trí các khu vực mỏ khoáng sản và các dạng dữ liệu khác như dữ liệu raster, dữ liệu độ cao, dữ liệu thông tin thuộc tính, dữ liệu phục vụ

công tác quản lí, khai thác... Trên Hình 1, ảnh bên trái là bản đồ thể hiện sự phân bố khoáng sản tỉnh Nghệ An được biên tập theo bản đồ địa chất và khoáng sản tỉ lệ 1:200.000. Những dấu chấm đỏ thể hiện các điểm mỏ khoáng sản lớn của tỉnh Nghệ An. Bản đồ này được lấy từ địa chỉ website của Tổng cục Địa chất Khoáng sản Việt Nam (<http://dgm.v.gov.vn/>). Ảnh bên phải là tổ hợp kênh địa chất sử dụng SWIR-2 (kênh 7), SWIR-1 (kênh 6) và màu xanh lam (kênh 2). Tổ hợp màu này đặc biệt hữu ích để xác định cấu tạo địa chất, khoáng sản.

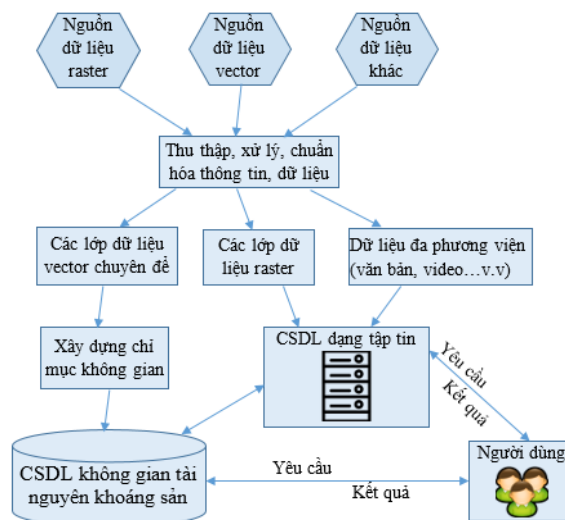


Hình 1. Bản đồ khoáng sản tỉnh Nghệ An

Hình 2 thể hiện mô hình tổng quan về xây dựng cơ sở dữ liệu khoáng sản với dữ liệu đầu vào bao gồm:

- Các nguồn dữ liệu raster: bao gồm các loại dữ liệu viễn thám, bản đồ khoáng sản dạng raster, các bản đồ ảnh thể hiện phân bố khoáng sản từ quá trình xử lí dữ liệu ảnh viễn thám...
- Các nguồn dữ liệu vector: bao gồm các lớp dữ liệu về mỏ khoáng sản (dạng point, line, polygon...), lớp giao thông, lớp ranh giới, lớp thủy hệ...

Các nguồn dữ liệu khác: bao gồm các loại dữ liệu như hình ảnh, văn bản, video...



Hình 2. Mô hình tổng quan xây dựng cơ sở dữ liệu khoáng sản

Đầu ra của mô hình là cơ sở dữ liệu không gian tài nguyên khoáng sản và cơ sở dữ liệu dạng tập tin. Trong đó, cơ sở dữ liệu không gian tài nguyên khoáng sản đóng vai trò cốt lõi, đảm bảo việc phân tích, truy vấn không gian... Cơ sở dữ liệu không gian này được xây dựng từ các lớp dữ liệu dữ liệu vector chuyên đề. Chẳng hạn như Hình 3 thể hiện một số lớp dữ liệu chuyên đề đối với khu vực huyện Quỳnh Hợp – Nghệ An. Các lớp dữ liệu vector chuyên đề khi được nhập vào trong cơ sở dữ liệu không gian tài nguyên khoáng sản sẽ được tạo các chỉ mục không gian nhằm tối ưu (tăng tốc) quá trình truy vấn.

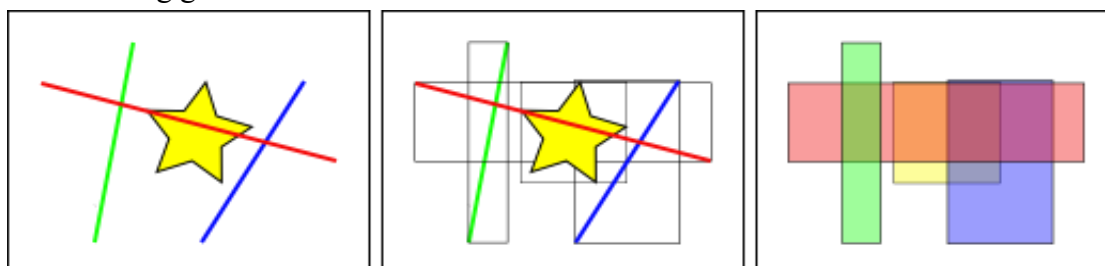
Lớp đường (dạng đường)		Lớp mỏ khoáng sản (dạng điểm)		Lớp dân cư (dạng điểm)	
Lớp sông suối (dạng đường)		Lớp thực phủ chưa thành rừng (dạng vùng)		Lớp khoáng sản dạng vùng	

Hình 3. Một số lớp dữ liệu vector chuyên đề (khu vực Quỳnh Hợp – Nghệ An)

2.2. Kỹ thuật chỉ mục trong cơ sở dữ liệu không gian

Cơ sở dữ liệu không gian địa lí (geospatial databases) là yếu tố trung tâm trong cơ sở hạ tầng dữ liệu không gian (spatial data infrastructures). Ưu điểm chính so với lưu trữ dữ liệu dựa trên tập tin là cơ sở dữ liệu không gian được cấu trúc để bao gồm các khả năng của các hệ thống quản lí cơ sở dữ liệu quan hệ, bao gồm hỗ trợ ngôn ngữ truy vấn có cấu trúc (Structured Query Language – SQL) và khả năng tạo các truy vấn không gian địa lí phức tạp (Angel, 2015; Chengkun et al., 2020).

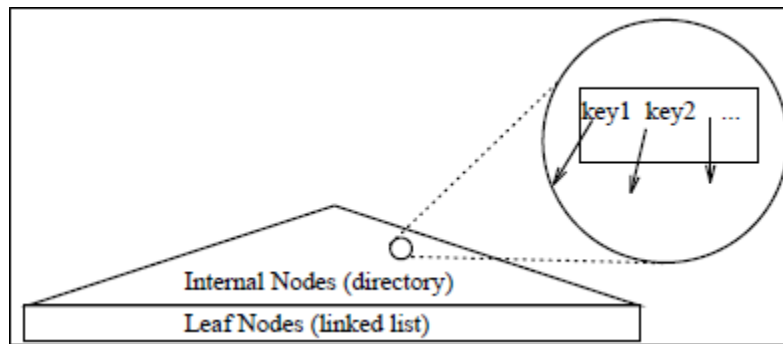
PostGIS là cơ sở dữ liệu không gian trong cho hệ quản trị cơ sở dữ liệu quan hệ đối tượng PostgreSQL. Nó bao gồm các chức năng phân tích không gian và hỗ trợ lập chỉ mục. Chỉ mục rất quan trọng đối với các bảng lưu trữ nhiều đối tượng, bởi vì nó giúp cho việc tìm kiếm thông gian nhanh hơn.



Hình 4. Cách chỉ mục không gian làm việc

Hình 4 mô tả cách thức chỉ mục không gian làm việc. Trên hình 4, có thể thấy rằng số lượng các đường giao cắt với ngôi sao là 1, đó là đường màu đỏ, nhưng các hộp đặc trưng giao với hộp màu vàng là 2, gồm hộp màu đỏ và màu xanh dương. Cách mà cơ sở dữ liệu trả lời hiệu quả câu hỏi “đường nào giao cắt với ngôi sao màu vàng”, trước tiên sẽ trả lời câu hỏi “những hộp nào giao với hộp màu vàng” bằng cách sử dụng chỉ số và sau đó thực hiện tính toán một cách chính xác đường nào sẽ giao với ngôi sao màu vàng. Chỉ mục giúp việc thao tác cơ sở dữ liệu không gian lớn trở lên dễ dàng hơn. Nếu không có việc đánh chỉ mục, thì bất kì việc tìm kiếm nào cũng phải yêu cầu việc “duyet tuần tự” tất cả các bản ghi có trong cơ sở dữ liệu (Vikram et al., 2019).

Chỉ mục không gian GiST (Cây tìm kiếm tổng quát – Generalized Search Tree) được mô tả ban đầu bởi (Hellerstein et al., 1995). Nó là một cấu trúc chỉ mục cung cấp phương pháp truy cập cấu trúc dạng cây cân bằng. Ngoài việc cung cấp khả năng mở rộng cho các loại dữ liệu không gian, GiST đã hợp nhất các cấu trúc khác nhau trước đây như B-tree và R-tree (Nguyen, 2009).



Hình 5. Phác thảo cây tìm kiếm cơ sở dữ liệu không gian sử dụng chỉ mục GiST

Ngày nay, các hệ quản trị cơ sở dữ liệu nói chung, cũng như cơ sở dữ liệu không gian nói riêng, đều hỗ trợ các thuật toán xây dựng chỉ mục không gian. Đối với cơ sở dữ liệu không gian PostGIS, để tạo chỉ mục không gian GiST, có thể sử dụng câu lệnh SQL như sau:

Create index <Tên chỉ số> on <Tên bảng> using gist (<Tên trường hình học>);

Việc tạo chỉ mục không gian trong cơ sở dữ liệu sẽ giải quyết các bài toán truy vấn không gian hay xác định mối quan hệ giữa các đối tượng. Một số mối quan hệ không gian thường được sử dụng như: mối quan hệ giao nhau, mối quan hệ bên trong – khoảng cách, mối quan hệ vượt qua.

Mối quan hệ giao nhau (ST_Intersects): mối quan hệ giao nhau giữa hai đối tượng A và B, trong PostGIS sử dụng câu lệnh:

boolean ST_Intersects(geometry geomA , geometry geomB);

Hàm *ST_Intersects* trả về true nếu A và B giao nhau, ngược lại trả về false. Độ sai số trong mối quan hệ này là 0,00001 m. Trong thực tế, mối quan hệ giao nhau có thể được sử

dụng để xác định các vùng trọng điểm, mật độ tập trung của các mỏ hay khu vực mỏ khoáng sản trong một phạm vi địa lí.

Mối quan hệ bên trong – khoảng cách (ST_Dwithin):

boolean ST_DWithin(geometry g1, geometry g2, double precision distance_of_srid);

Hàm *ST_Dwithin* trả về true nếu khoảng cách giữa *g1* và *g2* nhỏ hơn *d*, ngược lại trả về false. Đơn vị khoảng cách ở đây được xác định theo hệ tham chiếu không gian của *g1* và *g2*. Trong thực tế, mối quan hệ này được sử dụng để tìm kiếm số lượng các điểm mỏ trong một lân cận bán kính nào đó, hay xác định mật độ giao thông (dạng đường) để từ đó phục vụ bài toán quản lí, quy hoạch việc phát triển mở rộng khu vực mỏ.

Mối quan hệ vượt qua (ST_Crosses): Để xét mối quan hệ vượt qua giữa hai đối tượng A và B, trong PostGIS sử dụng câu lệnh:

boolean ST_Crosses(geometry A, geometry B)

Hàm *ST_Crosses* trả về true nếu phần giao nhau của A và B tạo ra dạng hình học có chiều nhỏ hơn chiều lớn nhất của A và B; đồng thời phần giao nhau này nằm bên trong của A và B. Chiều của dạng hình học được quy chuẩn như sau: dạng điểm có chiều là 0, dạng đường có chiều là 1, dạng vùng có chiều là 2. Trong thực tế, mối quan hệ này được sử dụng để xác định khả năng giao cắt giữa các lớp đối tượng dạng tuyến, dạng vùng...

3. Kết quả và thảo luận

3.1. Kết quả nghiên cứu

Để thực hiện việc đánh giá khả năng truy vấn nhanh của chỉ mục không gian đối với cơ sở dữ liệu lớn, nhóm tác giả đã xây dựng nhiều bộ dữ liệu mẫu (thể hiện trên các cột 2, cột 3 của các Bảng 1-4). Trong đó, các bộ dữ liệu này được xây dựng với các dạng hình học khác nhau (dạng điểm, đường, vùng) và các bảng dữ liệu có từ vài nghìn cho tới vài triệu bản ghi. Các thực nghiệm được tiến hành trên máy tính có cấu hình như sau: hệ điều hành Windows 10, Intel core™ i7- 4510U 2.0GHz, 8GB RAM. Kết quả nghiên cứu được thể hiện trên các Bảng 1-4. Trong đó:

- N_{TK}-Bản ghi (đường): Thể hiện số bản ghi của lớp dữ liệu dạng đường. Đây là lớp dữ liệu được sử dụng để truy vấn. TK là viết tắt của từ “Tìm Kiếm”

- N_{ĐK}-Bản ghi (điểm): Thể hiện số bản ghi của lớp dữ liệu dạng điểm. Đây là lớp dữ liệu được sử dụng để làm điều kiện của phép truy vấn. ĐK là viết tắt của từ “Điều Kiện”

- T_{GiST} (ms): Thể hiện thời gian của phép truy vấn được thực hiện trong trường hợp thiết lập chỉ mục không gian GiST. Đơn vị thời gian là mili giây.

- T_{No_GiST}(ms): Thể hiện thời gian của phép truy vấn được thực hiện trong trường hợp không thiết lập chỉ mục không gian GiST. Đơn vị thời gian là mili giây.

- Kịch bản 1: Thực hiện phép phân tích không gian sử dụng mối quan hệ bên trong – khoảng cách (*ST_Dwithin*) giữa hai tập dữ liệu dạng đường và dạng điểm. Kết quả được thể hiện như trên Bảng 1.

- Kịch bản 2: Thực hiện phép phân tích không gian sử dụng mối quan hệ bên trong – khoảng cách (*ST_Dwithin*) giữa hai tập dữ liệu dạng vùng và dạng điểm. Kết quả được thể hiện như trên Bảng 2.

- Kịch bản 3: Thực hiện phép phân tích không gian sử dụng mối quan hệ giao nhau (*ST_Intersects*) giữa hai tập dữ liệu dạng đường và dạng điểm. Kết quả được thể hiện như trên Bảng 3.

- Kịch bản 4: Thực hiện phép phân tích không gian sử dụng mối quan hệ vượt qua (*ST_Crosses*) giữa hai tập dữ liệu dạng đường và dạng đường. Kết quả được thể hiện như trên Bảng 4.

Bảng 1. Bảng kết quả với *ST_Dwithin* giữa hai tập dữ liệu dạng đường và dạng điểm

STT	N _{TK} -Bản ghi (đường)	N _{ĐK} -Bản ghi (điểm)	T _{GIST} (ms)	T _{No_GIST} (ms)
1	43,775	477	15	405
2	79,529	2,323	27	758
3	118,774	15,166	47	1,068
4	849,199	2,401	63	6,396
5	1,047,502	2,628	67	9,472
6	1,646,394	8,510	85	10,319
7	2,693,896	8,510	94	17,226

Bảng 2. Bảng kết quả với *ST_Dwithin* giữa hai tập dữ liệu dạng vùng và dạng điểm

STT	N _{TK} -Bản ghi (vùng)	N _{ĐK} -Bản ghi (điểm)	T _{GIST} (m)	T _{No_GIST} (ms)
1	1,270	267	6	39
2	20,783	1,071	10	611
3	50,971	2,783	17	1682
4	120,082	4,158	21	3658
5	273,975	1,294	25	8023
6	503,902	600	28	9039
7	663,713	2,401	33	19935

Bảng 3. Bảng kết quả với *ST_Intersects* giữa hai tập dữ liệu dạng đường và dạng điểm

STT	N _{TK} -Bản ghi (đường)	N _{ĐK} -Bản ghi (điểm)	T _{GIST} (m)	T _{No_GIST} (ms)
1	49,028	391	8	42
2	75,264	2,174	10	67
3	101,858	17,912	12	91
4	881,107	1,870	23	342
5	1,237,829	2,951	27	437
6	1,546,311	6,749	29	624
7	2,848,743	8,163	41	1,092

Bảng 4. Bảng kết quả với ST_Crosses giữa hai tập dữ liệu dạng đường và dạng đường

STT	N _{TK} -Bản ghi (đường)	N _{ĐK} -Bản ghi (đường)	T _{GiST} (m)	T _{No_GiST} (ms)
1	41,697	384	12	42
2	82,074	2,709	19	71
3	120,375	2,357	23	79
4	901,174	2,401	35	363
5	1,207,891	2,628	42	442
6	1,779,408	4,176	47	973
7	2,801,526	5,526	55	1,056

3.2. Thảo luận

Hình 6 thể hiện biểu đồ thời gian xử lý trong trường hợp sử dụng và không sử dụng chỉ mục không gian GiST: hình 6a là kết quả của kịch bản 1 (Bảng 1); Hình 6b là kết quả của kịch bản 2 (Bảng 2); hình 6c là kết quả của kịch bản 3 (Bảng 3); Hình 6d là kết quả của kịch bản 4 (Bảng 4). Trên mỗi Hình 6 (a, b, c, d), có thể nhận thấy:

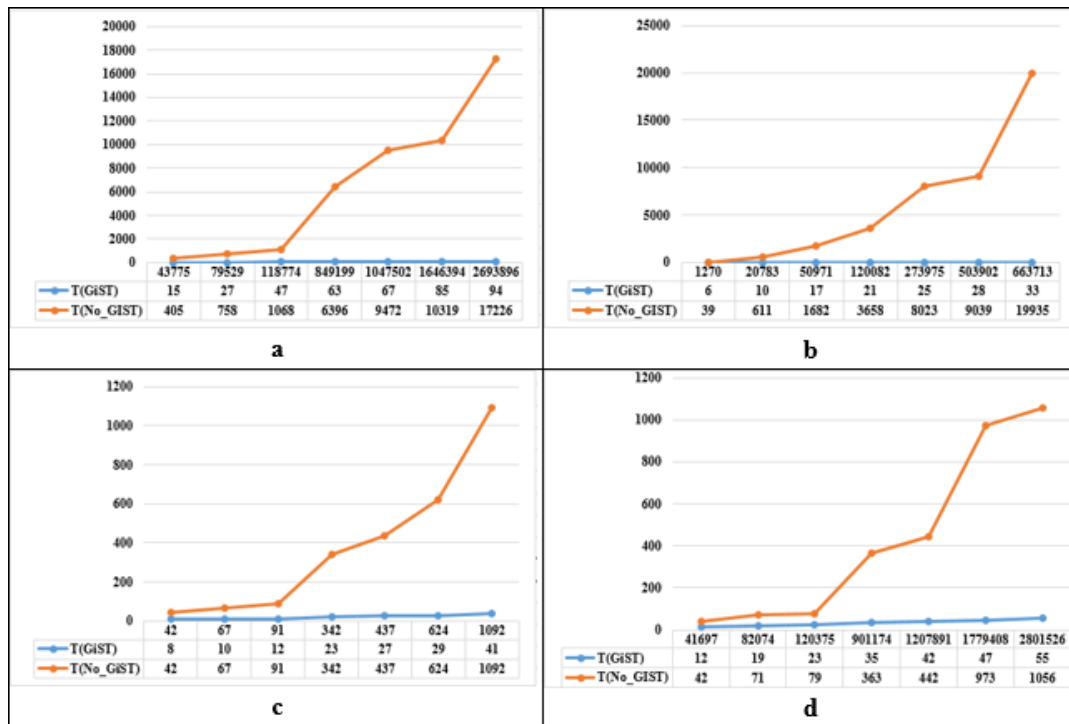
- Đường màu cam – kí hiệu là T(GiST): Thể hiện thời gian thực hiện của các phép tìm kiếm không gian trong trường hợp thiết lập chỉ mục không gian GiST.

- Đường màu xanh lục – kí hiệu là T(No_GiST): Thể hiện thời gian thực hiện của các phép tìm kiếm không gian trong trường hợp không thiết lập chỉ mục không gian GiST.

- Đường màu cam có vị trí gần sát với đường trục hoành hay có khi gần như một đường kẻ ngang. Điều này thể hiện rằng: đối với trường hợp thiết lập chỉ số không gian GiST, thời gian thực hiện truy vấn không gian là ổn định và rất nhanh. Thời gian thực hiện phép truy vấn không gian nhỏ hơn 1 giây cho dù số lượng bản ghi lên tới 2,7 triệu.

- Đường màu xanh lục ban đầu cũng có vị trí thấp (đối với trường hợp số lượng bản ghi ít); tuy nhiên, khi số lượng bản ghi lớn (lên đến vài trăm nghìn, triệu bản ghi) thì đường màu xanh lục sẽ cao dần trở nên gần như dốc đứng. Như vậy, đối với trường hợp không thiết lập chỉ mục không gian, thời gian thực hiện phép truy vấn tăng nhanh đột biến khi dữ liệu lên tới lớn.

Như vậy, kết quả nghiên cứu có thể khẳng định rằng đối với các cơ sở dữ liệu không gian lớn (số lượng bản ghi bắt đầu từ vài trăm nghìn bản ghi) thì việc thiết lập chỉ mục không gian như GiST có ý nghĩa vô cùng quan trọng. Nó thực hiện các phép tìm kiếm, tính toán không gian nhanh hơn rất nhiều so với trường hợp không thiết lập chỉ số. Điều này rất quan trọng đối với việc nghiên cứu và thiết kế cơ sở dữ liệu nói chung, cũng như cơ sở dữ liệu không gian tài nguyên khoáng sản nói riêng.



Hình 6. Biểu đồ thể hiện kết quả thực hiện của các phép truy vấn không gian trong trường hợp sử dụng/không sử dụng chỉ mục GiST.

4. Kết luận

Kỹ thuật chỉ mục không gian nói chung, cũng như chỉ mục GiST nói riêng là một trong những đặc trưng và quan trọng của cơ sở dữ liệu. Nó giúp cho việc thực hiện các phép truy vấn không gian hay phân tích dữ liệu lớn trở nên nhanh hơn. Trong quá trình xây dựng hệ thống, kỹ thuật này cần được tích hợp vào cơ sở dữ liệu không gian. Kết quả nghiên cứu sẽ là một trong những cơ sở quan trọng, giúp cho quá trình xây dựng hệ thống quản lý dữ liệu không gian về tài nguyên khoáng sản được tốt hơn.

❖ **Tuyên bố về quyền lợi:** Các tác giả xác nhận hoàn toàn không có xung đột về quyền lợi.

TÀI LIỆU THAM KHẢO

Angel Marquez (2015). *PostGIS Essentials*. Birmingham: Packt Publishing.

Chengkun, L., Rui, M. & Zhang, L. (2020). Analysis of Spatial Indexing Mechanism and Its Application in Data Management: A Case Study on Spatialite Database. Retrieved from https://www.researchgate.net/publication/338657937_Analysis_of_Spatial_Indexing_Mechanism_and_Its_Application_in_Data_Management_A_Case_Study_on_Spatialite_Database

Du, V. T. (2009). Nghiên cứu chính sách quản lý tài nguyên khoáng sản biển tại Việt Nam [Research policy on management of marine mineral resources in Vietnam]. Retrieved from https://www.researchgate.net/publication/280942930_Dinh_huong_chinh_sach_QLTN_va_EITI_khoang_san_bien_VN

- Ernest, O., Ogbonna, A., Alli-Shehu, B. & Maduakolam, C. (2015). Mineral resources management information system. *European Journal of Computer Science and Information System*, 3(2), 13-23.
- Hellerstein, J. M., Jeffrey, F. N. & Avi, P. (1995). Generalized Search Trees for Database Systems. *Proceedings of the 21th International Conference on Very Large Data Bases*, 562-573.
- Nguyen, H. P., Le, A. T., Pham, Q. H., Nguyen, K. V. (2009). Ung dung cong nghe WebGIS xay dung co so du lieu dia chat khoang san, dia chat moi truong va tai bien dia chat cac vung bien Viet Nam [Application of WebGIS technology to build a database of geological minerals, environmental geology and geological assets in Vietnam's territorial waters]. Retrieved from <http://www.idm.gov.vn/Data/TapChi/2009/A315/A71.htm>
- Nguyen, T. T. (2009). Indexing PostGIS databases and spatial Query performance evaluations. *International Journal of Geoinformatics*, 5(3), 1-9.
- Neelabh, P., Mohammadhani, F., Ramez, E. & Kulsawasd, J. (2015). Performance comparison of spatial indexing structures for different query types. *Proceedings of 57th IRF International Conference*, 43-50.
- Vikram, N., Jialin, D., Mohammad, A. & Tim, K. (2019). Learning Multi-dimensional Indexes. *SIGMOD '20: Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data*, 985-1000.
- Vo, Q. M. (2012). Danh gia thuc trang va tien nang khai thac mot so san pham tu nguon tai nguyen set o DBSCL [Assessing the current status and potential of exploiting some products from clay resources in the Mekong Delta]. Retrieved from https://www.researchgate.net/publication/266494553_Danh_gia_thuc_trang_va_tiem_nang_khai_thac_mot_so_san_pham_tu_nguon_tai_nguyen_set_o_DBSCL

APPLICATION OF SPATIAL INDEXING TECHNIQUES FOR MINERAL RESOURCE DATABASE

Nguyễn Sách Thành^{1*}, Nguyễn Nhu Hưng¹, Đào Thanh Bình²

¹Le Quy Don Technical University, Vietnam

²Department of Natural Resources and Environment of Nghe An Province, Vietnam

*Corresponding author: Nguyễn Sách Thành – Email: thanhns.geo@lqdtu.edu.vn

Received: September 16, 2020; Revised: September 30, 2020; Accepted: December 25, 2020

ABSTRACT

Mineral resource management is an important issue in each country's sustainable development strategies, of which the mineral spatial database is a major component of the management system. Nowadays, with the development of information collection and processing technologies, spatial data on mineral resources are bigger and bigger. This has led to a complicated demand of how to quickly query spatial data with millions of records. To solve the problem of big spatial data, spatial indexing technique is often used. In this paper, the authors will apply spatial indexing techniques in building mineral databases on PostGIS and evaluate the ability to query of using and not using spatial indexes. The results show that query time when using spatial index is much faster.

Keywords: mineral resource database; PostGIS; spatial indexing