



Xây dựng dịch vụ web định vị và truy vấn thông tin ảnh vệ tinh trong kho dữ liệu lớn

Đào Khánh Hoài¹, Trương Văn Tuấn¹, Trần Thị Tuyết Vinh²

¹Học viện Kỹ thuật Quân sự, Hà Nội, Việt Nam

²Nhóm nghiên cứu phát triển công nghệ quản lý và phân tích dữ liệu không gian địa lý (GMA),
Trường đại học Mỏ - Địa chất

Email tác giả liên hệ: geogroup2008@gmail.com

<https://doi.org/10.5281/zenodo.xxxxxxx>

Tóm tắt:

Trong những năm gần đây, dữ liệu ảnh vệ tinh ngày càng được sử dụng rộng rãi trong nhiều lĩnh vực của đời sống kinh tế và xã hội. Theo thời gian, các kho dữ liệu ảnh tích lũy đã tăng nhanh cả về số lượng và dung lượng. Điều này mang lại cho các tổ chức nghiên cứu, ứng dụng và sản xuất dữ liệu không gian địa lý nhiều cơ hội hơn để khai thác thông tin, nhưng cũng đặt ra một số thách thức trong việc quản lý chúng. Một nhu cầu cơ bản của người dùng trực tuyến là tìm kiếm, định vị và tra cứu thông tin về ảnh vệ tinh trong các kho dữ liệu mà họ quan tâm. Đóng góp chính trong nghiên cứu này bao gồm hai khía cạnh. Thứ nhất, một cách tiếp cận khoa học thực tiễn hiệu quả để thiết kế kiến trúc dữ liệu để lưu trữ dữ liệu ảnh vệ tinh và quản lý siêu dữ liệu ảnh vệ tinh dựa trên ba công nghệ: Hadoop, Hbase và PostGis đã được đề xuất và thử nghiệm. Thứ hai, một dịch vụ Web cho phép tìm kiếm và tra cứu nhanh thông tin siêu dữ liệu ảnh vệ tinh trong các kho dữ liệu lớn đã được triển khai và thử nghiệm. Các kết quả thử nghiệm trên kho ảnh Sentinel2 đã cho thấy kết quả khả quan. Ảnh vệ tinh được truy vấn và định vị trong cơ sở dữ liệu ở chế độ gần thời gian thực. Giao diện máy khách thông minh, dễ sử dụng tạo ra trải nghiệm mượt mà, API Web nhanh và liền mạch (www.geoview.vn).

Keywords: Hệ thống phân tán; Dữ liệu không gian địa lý; Kho dữ liệu lớn; Truy vấn; Định vị không gian.

Ngày nhận bài: 24/03/2026

Ngày sửa lại: xx/04/2026

Ngày chấp nhận đăng: xx/04/2026

Ngày xuất bản: 30/04/2026

Bổ sung tiếng Anh

Submission received: 24/03/2026

Revised: xx/04/2026

Accepted: xx/04/2026

Published: 30/04/2026

1. Giới thiệu

Theo dữ liệu từ Union of Concerned Scientists (UCS), hiện có khoảng 1.200 vệ tinh quan sát Trái Đất đang hoạt động trên quỹ đạo. Các vệ tinh này cung cấp một nguồn ảnh vệ tinh đa dạng và phong phú. Các kho lưu trữ ảnh vệ tinh không ngừng gia tăng cả về số lượng lẫn dung lượng. Sự mở rộng này mang lại nhiều cơ hội cho các tổ chức hoạt động trong lĩnh vực nghiên cứu viễn thám và sản xuất dữ liệu không gian địa lý khai thác tiềm năng của dữ liệu, nhưng đồng thời cũng đặt ra nhiều thách thức trong việc quản lý các tập dữ liệu lớn này [1–3].

Hiện nay, các nhà cung cấp ảnh vệ tinh trên toàn thế giới đang cung cấp cho người dùng trực tuyến khả năng truy cập và khai thác thông tin ảnh vệ tinh từ các kho dữ liệu thương mại thông qua các dịch vụ web. Một ví dụ tiêu biểu là dịch vụ web NASA Worldview của NASA, cho phép người dùng trực quan khám phá cả quá khứ và hiện tại của hành tinh năng động của chúng ta từ góc nhìn vệ tinh [4]. United States Geological Survey (USGS) cung cấp một bộ API giúp người dùng chủ động xây dựng các truy vấn ảnh vệ tinh phức tạp bằng cách kết hợp nhiều tiêu chí trong một giao diện thân thiện [5]. Airbus, một trong những nhà cung cấp ảnh vệ tinh hàng đầu thế giới, cũng cung cấp các dịch vụ cho phép người dùng xem

trước ảnh mẫu, đánh giá dữ liệu ảnh vệ tinh nâng cao và khám phá các sản phẩm giá trị gia tăng trong kho dữ liệu của họ nhằm hỗ trợ tốt nhất cho các dự án của người dùng [6]. Sentinel Hub Browser cung cấp các công cụ cơ bản để khám phá dữ liệu quan sát Trái Đất, phát triển các kịch bản tùy chỉnh cho phân tích đơn giản và tích hợp các kịch bản từ bên thứ ba. Tương tự, nhà cung cấp ảnh vệ tinh UP42 cho phép người dùng dễ dàng khám phá và thu thập dữ liệu, ước tính chi phí một cách minh bạch và xem trước hình ảnh, qua đó hỗ trợ việc so sánh và tổ chức dữ liệu không gian địa lý.

Tại Việt Nam, lĩnh vực thông tin địa không gian đã được hình thành và phát triển trong nhiều năm, gắn liền với nhiều lĩnh vực của đời sống kinh tế - xã hội. Cộng đồng nghiên cứu về trắc địa, bản đồ, viễn thám và hệ thống thông tin địa lý đã hình thành một mạng lưới nghiên cứu bao gồm các viện nghiên cứu và các trường đại học. Do đó, việc nghiên cứu và xây dựng các công nghệ thông tin địa không gian nhằm chia sẻ các nguồn dữ liệu tham chiếu không gian, giúp các nhà nghiên cứu có cơ hội nhanh chóng tiếp cận các nguồn dữ liệu miễn phí là cần thiết và hữu ích.

Các phần tiếp theo của bài báo sẽ tập trung làm rõ hai nội dung chính sau:

(i) Thiết kế và triển khai một kiến trúc dữ liệu thực tiễn và hiệu quả để lưu trữ dữ liệu ảnh vệ tinh và quản lý siêu dữ liệu ảnh vệ tinh, kết hợp ba công nghệ: Apache Hadoop, Apache HBase và PostGIS. Trong đó, hệ thống phân tán Hadoop lưu trữ dữ liệu gốc, cơ sở dữ liệu phân tán dạng cột tốc độ cao HBase lưu trữ siêu dữ liệu ảnh, và cơ sở dữ liệu không gian PostGIS lưu trữ dữ liệu hình học (footprint) của ảnh vệ tinh. Cách tiếp cận này dựa trên thực tế công nghệ rằng Spatial Hadoop đã ngừng phát triển, HBase không hỗ trợ chi mục không gian, do đó PostGIS được lựa chọn để xử lý các truy vấn không gian phức tạp nhằm nhanh chóng xác định các định danh của ảnh vệ tinh cần truy vấn thông tin.

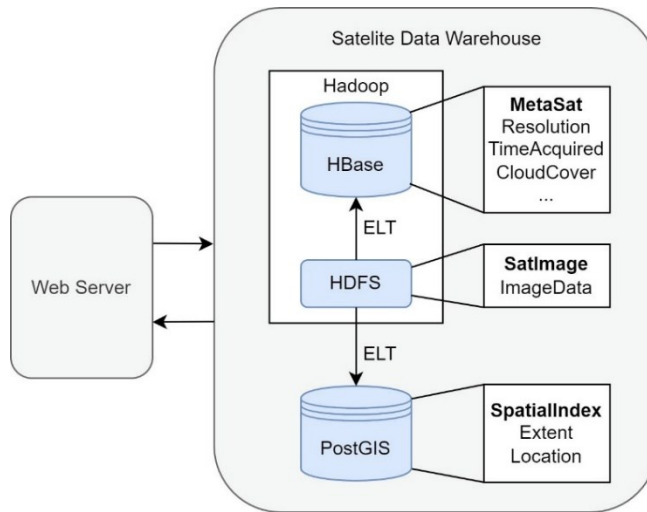
(ii) Thiết kế và xây dựng một dịch vụ web tra cứu thông tin ảnh vệ tinh hoàn chỉnh (www.geoview.vn) với giao diện thân thiện và tốc độ truy vấn nhanh, được triển khai và vận hành thực tế phục vụ nhiều mục đích từ nghiên cứu, học tập công nghệ địa không gian đến các hoạt động tìm kiếm cứu nạn hoặc tình báo không gian. Dịch vụ web này được thiết kế và phát triển thông qua các API mô phỏng các hành vi người dùng phổ biến và điển hình được quan sát trong các hệ thống thông tin địa lý hiện đại đang được sử dụng rộng rãi trên thị trường.

2. Kiến trúc dịch vụ web

2.1. Kiến trúc dữ liệu

Các thành phần dữ liệu trong một hồ dữ liệu vệ tinh quy mô lớn bao gồm ảnh gốc, siêu dữ liệu và một số loại dữ liệu đặc thù khác, tất cả đều được lưu trữ trong cùng một thư mục dưới một định danh ảnh duy nhất do mỗi nhà cung cấp ảnh gán. Các dữ liệu này được tổ chức thành các tệp ở định dạng TIF hoặc XML.

Siêu dữ liệu ảnh bao gồm các thông tin mô tả về ảnh, bao gồm: loại cảm biến, thời gian thu nhận của các cảnh ảnh, hướng của cảm biến, độ phân giải ảnh, hệ tọa độ và phép chiếu, tỷ lệ mây che phủ, các hệ số độ nhạy và độ trễ của cảm biến, tọa độ các góc ảnh, tâm ảnh và nhiều thông tin khác.



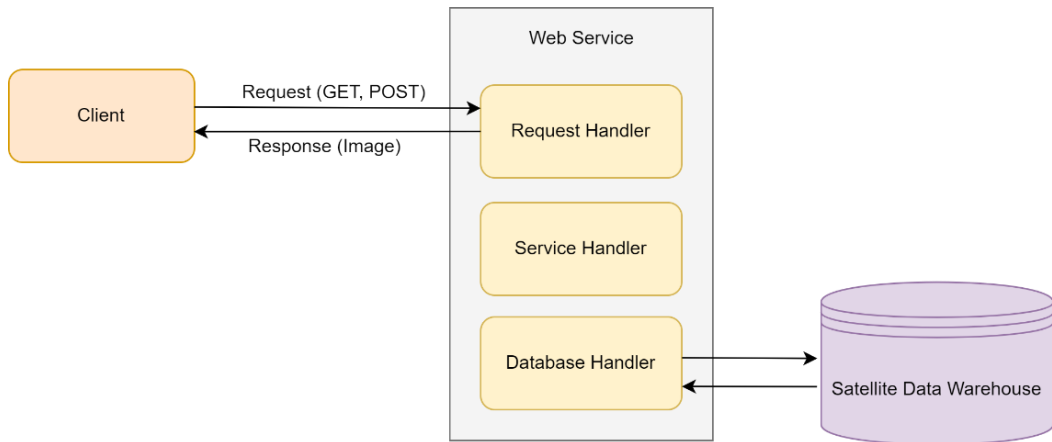
Hình 1. Kiến trúc hồ dữ liệu.

Trong kho dữ liệu, dữ liệu ảnh vệ tinh được mô hình hóa thành các thực thể chính, bao gồm: Ảnh vệ tinh (SatImage), chứa định danh ảnh và địa chỉ lưu trữ ảnh trong hệ thống tệp; Siêu dữ liệu ảnh (MetaSat), chứa các thông tin mô tả về ảnh; và Chỉ mục không gian (SpatialIndex), dùng để lưu trữ hình học footprint của ảnh nhằm hỗ trợ việc định vị ảnh nhanh trong hệ thống cơ sở dữ liệu. Dữ liệu ảnh gốc và siêu dữ liệu được lưu trữ trên các hệ thống phân tán như HBase hoặc hệ thống tệp phân tán Hadoop, trong khi chỉ mục không gian được quản lý bởi cơ sở dữ liệu hỗ trợ lập chỉ mục không gian như PostGIS. Toàn bộ dữ liệu của các thực thể trong kho dữ liệu được trích xuất từ ảnh vệ tinh gốc và được hoàn thiện thông qua quy trình ELT.

Khi người dùng thực hiện truy vấn theo một khu vực địa lý, hệ thống sẽ tìm kiếm trong SpatialIndex để xác định danh sách các ảnh liên quan, sau đó nhanh chóng truy xuất dữ liệu ảnh cần thiết từ hồ dữ liệu lớn. Cấu trúc này đảm bảo hệ thống có thể xử lý đồng thời nhiều truy vấn phức tạp trong thời gian ngắn, đồng thời dễ dàng mở rộng khi khối lượng ảnh vệ tinh tăng lên theo thời gian. Mô hình các thành phần chính của hồ dữ liệu được thể hiện trong Hình 1.

2.2. Kiến trúc dịch vụ web

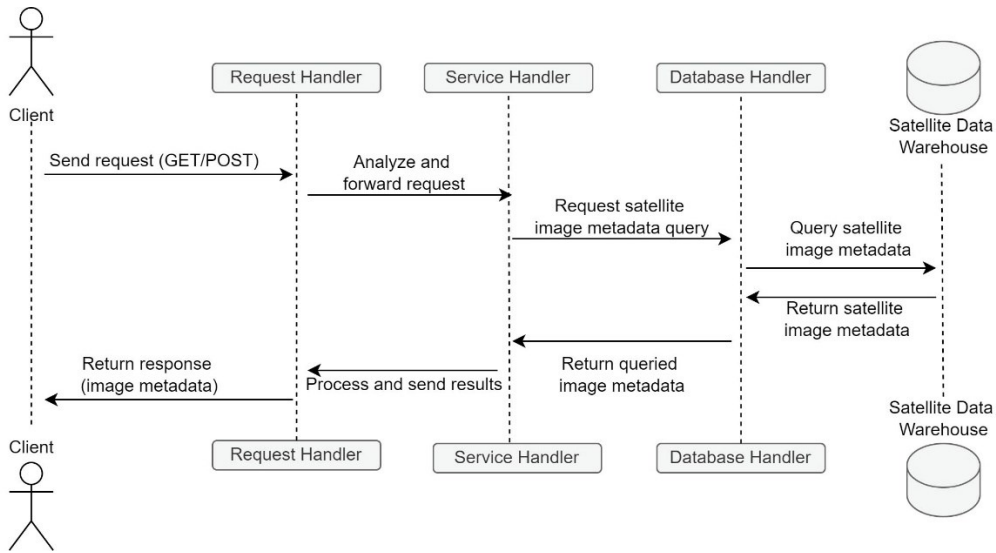
Trong phần này, chúng tôi mô tả một kiến trúc dịch vụ web nhằm định vị và truy vấn thông tin ảnh vệ tinh từ hồ dữ liệu thông qua mạng. Dịch vụ này gồm hai thành phần chính: Web Client và Web Server. Web Client là giao diện người dùng hoạt động trên trình duyệt phía máy khách. Người dùng sử dụng các ứng dụng web trên trình duyệt để gửi yêu cầu và tìm kiếm thông tin dữ liệu dựa trên nhiều tiêu chí khác nhau trên bản đồ (truy vấn không gian, truy vấn thuộc tính). Web Client chịu trách nhiệm định vị dữ liệu trên bản đồ và hiển thị siêu dữ liệu, đồng thời cung cấp các công cụ tương tác để làm việc với dữ liệu bản đồ như phóng to, đo đạc và điều hướng. Web Client được phát triển bằng HTML, CSS, JavaScript và các thư viện như ReactJS và OpenLayers nhằm hỗ trợ các chức năng WebGIS.



Hình 2. Mô hình tổng thể của dịch vụ web.

Web Server là hệ thống phân tán phía máy chủ, chịu trách nhiệm xử lý các yêu cầu truy vấn dữ liệu viễn thám từ Web Client. Web Server thực hiện các nhiệm vụ chính như định vị dữ liệu, truy vấn siêu dữ liệu ảnh vệ tinh từ cơ sở dữ liệu, tối ưu hóa hiệu năng truy vấn và trả kết quả về cho Web Client. Web Server được xây dựng để kết nối với hệ thống cơ sở dữ liệu phân tán như Apache HBase. Mô hình tổng thể của dịch vụ web phục vụ truy vấn và trực quan hóa siêu dữ liệu ảnh vệ tinh được thể hiện trong Hình 2.

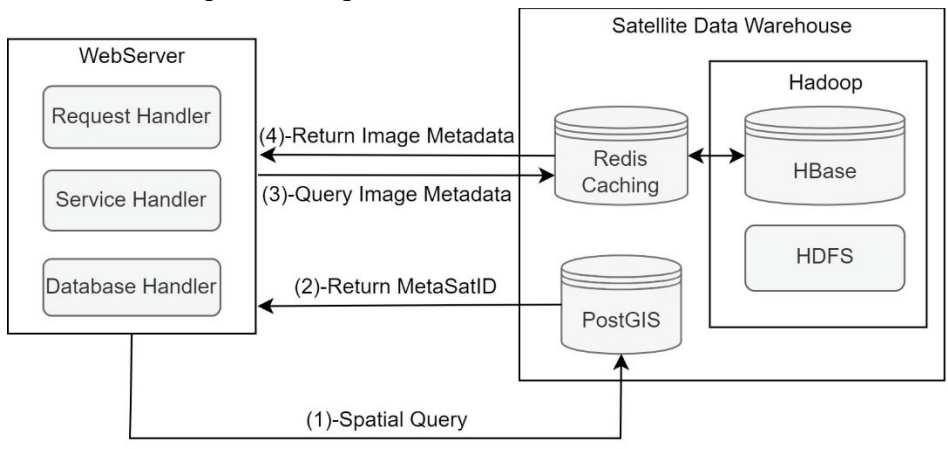
Khi người dùng tương tác với trình duyệt, chẳng hạn như tìm kiếm ảnh vệ tinh hoặc xem chi tiết ảnh, một yêu cầu sẽ được gửi đến Web Service thông qua giao thức HTTP, thường sử dụng các phương thức như GET hoặc POST. Bộ xử lý yêu cầu (Request Handler) tiếp nhận các yêu cầu từ người dùng và chuyển chúng đến các phân hệ dịch vụ phù hợp, chẳng hạn như Service Handler hoặc Database Handler. Database Handler đóng vai trò là cầu nối giữa các dịch vụ nghiệp vụ và hồ dữ liệu vệ tinh, thực hiện truy vấn và truy xuất dữ liệu ảnh vệ tinh liên quan dựa trên yêu cầu truy vấn. Sau khi truy vấn hoàn tất, dữ liệu được trả về cho Service Handler. Cuối cùng, trình duyệt của người dùng nhận phản hồi này và hiển thị dữ liệu ảnh vệ tinh trên giao diện. Người dùng có thể xem thông tin ảnh viễn thám hoặc tiếp tục tương tác với bản đồ như phóng to, thu nhỏ, di chuyển, hoặc thay đổi tiêu chí tìm kiếm để truy xuất dữ liệu mới. Quy trình này được lặp lại với mỗi yêu cầu mới (Hình 3).



Hình 3. Biểu đồ tuần tự của dịch vụ web truy vấn siêu dữ liệu ảnh.

Mặc dù các cơ sở dữ liệu NoSQL có hiệu quả trong việc quản lý dữ liệu lớn, cho phép truy cập và truy vấn dữ liệu nhanh, nhưng chúng không hỗ trợ lập chỉ mục không gian cho dữ liệu vector. Trong trường hợp này, cơ sở dữ liệu không gian PostGIS là giải pháp tối ưu.

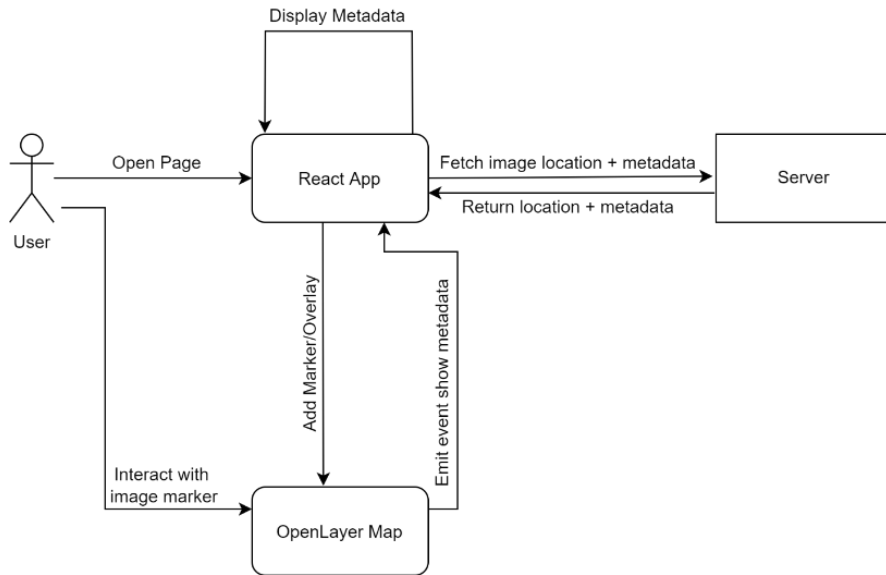
Quy trình truy vấn siêu dữ liệu ảnh theo không gian được thực hiện qua các bước sau: (i) Người dùng tạo truy vấn không gian theo các tùy chọn trên giao diện Web dưới dạng điểm, đường gấp khúc, ranh giới hoặc đa giác. Ứng dụng phía client gửi yêu cầu truy vấn kèm theo phạm vi không gian đến máy chủ web; (ii) Một API phía máy chủ tìm kiếm tất cả các ảnh vệ tinh nằm trong phạm vi không gian yêu cầu và trả về các mã ảnh vệ tinh hợp lệ cho Service Handler; (iii) Service Handler tiếp tục gửi truy vấn thông tin siêu dữ liệu của các mã ảnh vệ tinh trong cụm hệ thống (Redis + HBase); (iv) Một API tìm kiếm các thông tin phù hợp với yêu cầu truy vấn và trả kết quả về cho phía client.



Hình 4. Sơ đồ trình tự của truy vấn không gian.

2.3. Truy vấn và hiển thị thông tin ảnh viễn thám trên phía Client

Với sự phát triển của công nghệ web, các thư viện mã nguồn mở có thể hỗ trợ người dùng định vị dữ liệu trên bản đồ và hiển thị thông tin ảnh viễn thám. Các bước định vị và hiển thị dữ liệu viễn thám được minh họa trong mô hình ở Hình 4.



Hình 5. Mô hình kết hợp của các thành phần phía Client.

Mô hình định vị và hiển thị thông tin ảnh viễn thám được truy vấn trên phía client đảm bảo khả năng xác định vị trí nhanh chóng, hiển thị siêu dữ liệu và hỗ trợ các thao tác tương tác với bản đồ như phóng to, thu nhỏ và di chuyển. Trong nghiên cứu này, chúng tôi đề xuất một kiến trúc dựa trên ứng dụng React kết hợp với OpenLayers, cho phép người dùng tương tác trực tiếp với bản đồ. Thông tin vị trí ảnh và siêu dữ liệu được truy vấn từ máy chủ và trả về theo yêu cầu của người dùng, đảm bảo hiệu quả và khả năng mở rộng khi triển khai ở quy mô lớn.

3. Kết quả và thảo luận

Trong phần này, chúng tôi trình bày việc triển khai thực nghiệm kiến trúc dịch vụ web đã được mô tả ở các phần trước. Dữ liệu thực nghiệm bao gồm một hồ dữ liệu vệ tinh chứa ảnh quang học miễn phí từ Landsat 8 và Sentinel-2, với dung lượng lưu trữ hơn 2TB, được thu thập từ các tổ chức cung cấp dữ liệu miễn phí như Copernicus Open Access Hub và Sentinel Hub. Ngoài ra, hệ thống còn vận hành đồng thời một máy chủ bản đồ đa tỷ lệ với dung lượng lưu trữ dữ liệu ảnh vượt quá 2TB. Hệ thống được triển khai với cấu hình phần cứng và phần mềm cụ thể, sẽ được mô tả trong các phần tiếp theo.

3.1. Hệ thống máy chủ dịch vụ web

Hệ thống phần cứng được triển khai trên một cụm gồm 6 máy tính, bao gồm 1 máy chủ web và 5 máy tính đa năng dùng để vận hành hệ thống phân tán Hadoop và cơ sở dữ liệu phân tán HBase. Cấu hình chi tiết của các máy được trình bày trong bảng dưới đây:

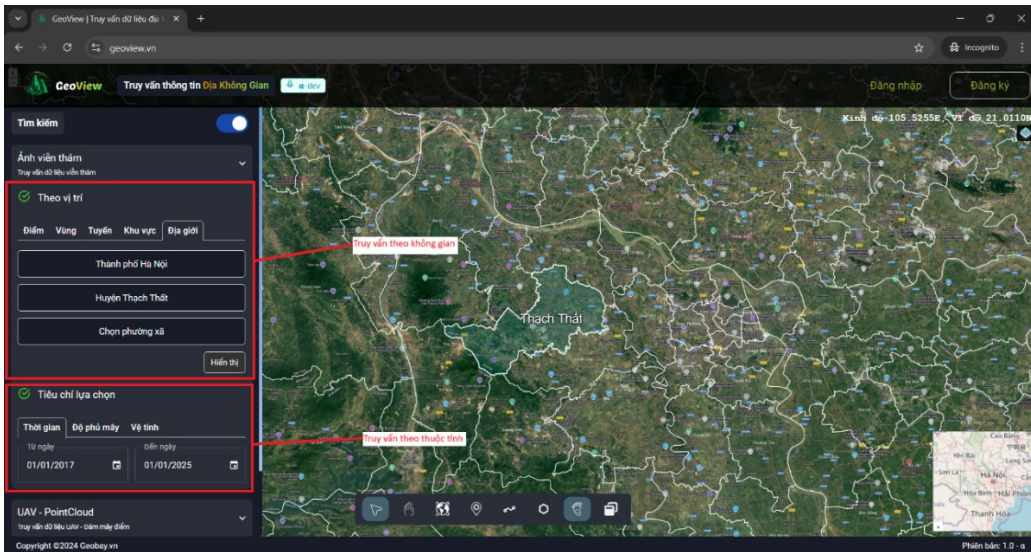
Bảng 1. Cấu hình máy tính của hệ thống.

Máy tính	Số lượng	Configuration
Máy chủ Web	01	- 2 processor: 48 nhân xử lý - 128GB RAM
Hệ thống phân tán	05	- Core i5 processor, - 32GB RAM, - 6TB ổ cứng lưu trữ
Mạng Lan		- 1Gbps

Các hệ thống phần mềm được sử dụng để triển khai mô hình dữ liệu bao gồm: hệ điều hành Ubuntu 22.04; máy chủ web Apache HTTP Server; hệ thống phân tán Apache Hadoop phiên bản 3.4.1; Apache HBase phiên bản 2.6.2; PostGIS phiên bản 3.5; và Redis phiên bản 7.4.1. Các ngôn ngữ lập trình và framework phần mềm bao gồm: Java, Python, ReactJS; các thư viện sử dụng gồm: GDAL, GeoTools và PySpark.

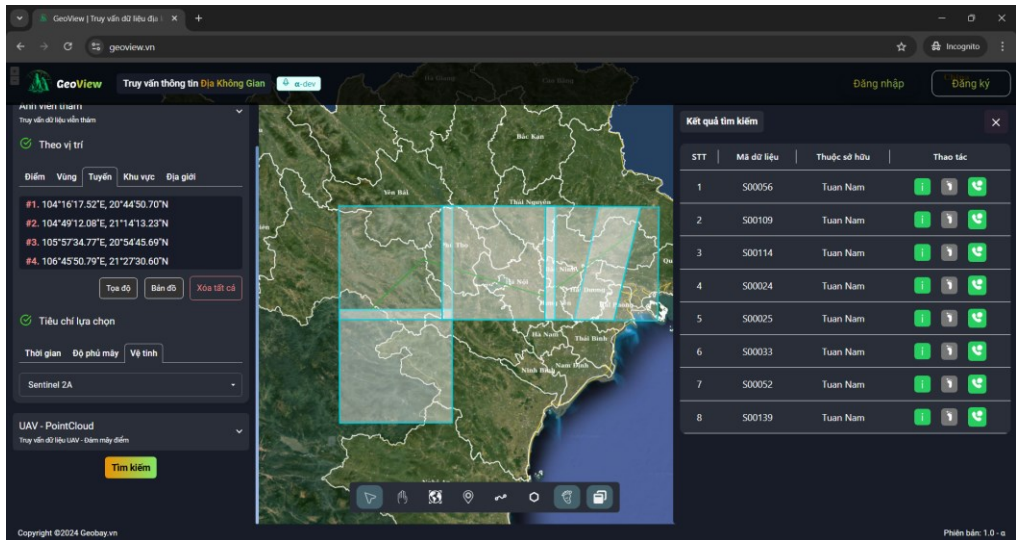
3.2. Định vị và truy vấn thông tin ảnh vệ tinh

Dựa trên kiến trúc phần cứng và các nền tảng phần mềm đã trình bày ở trên, chúng tôi đã xây dựng và triển khai một hệ thống định vị và truy vấn ảnh vệ tinh với các chức năng cụ thể, bao gồm: truy vấn dữ liệu, định vị dữ liệu và hiển thị dữ liệu. Nhiệm vụ truy vấn: Việc truy vấn dữ liệu ảnh vệ tinh được hỗ trợ linh hoạt dựa trên cả tiêu chí không gian và thuộc tính. Đối với truy vấn không gian, người dùng có thể trực tiếp vẽ các đối tượng hình học như điểm, đường hoặc đa giác trên bản đồ để xác định vùng quan tâm (AOI). Hệ thống sẽ tự động thực hiện kiểm tra giao cắt giữa AOI và phạm vi các ảnh vệ tinh trong hồ dữ liệu. Kết quả thực nghiệm cho thấy các thao tác phóng to, thu nhỏ và di chuyển bản đồ sau khi truy vấn vẫn được hệ thống duy trì mượt mà, giúp người dùng truy cập dữ liệu ảnh một cách liền mạch. Đối với truy vấn theo thuộc tính, hệ thống cung cấp các công cụ cho phép người dùng lọc ảnh vệ tinh dựa trên các tiêu chí như thời gian thu nhận, độ phân giải không gian, loại cảm biến hoặc mức độ mây che phủ. Các yêu cầu truy vấn thuộc tính được xử lý nhanh chóng. Kết quả thực nghiệm cho thấy hệ thống xử lý hiệu quả cả hai loại truy vấn, cho phép truy xuất dữ liệu ảnh vệ tinh nhanh, chính xác và phù hợp với nhu cầu định vị và tìm kiếm của người dùng.



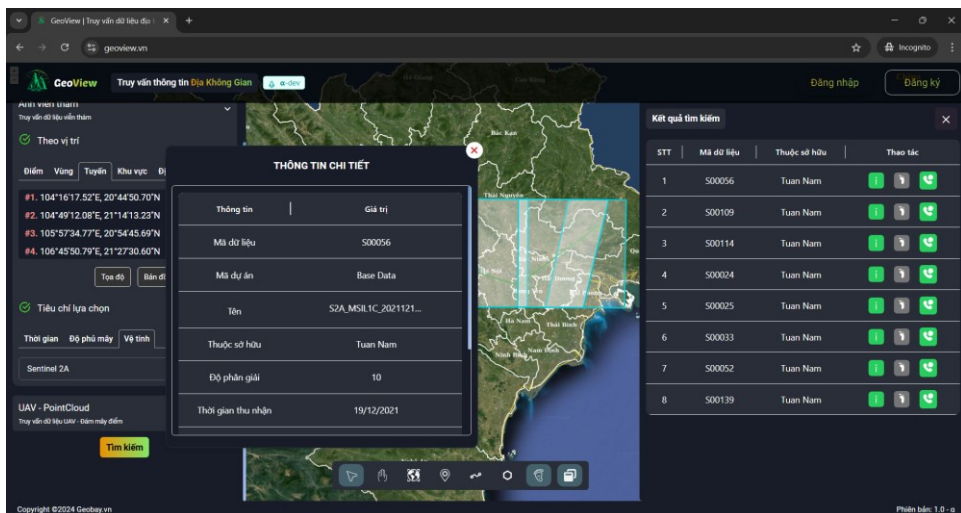
Hình 6. Công cụ truy vấn thông tin ảnh viễn thám dựa trên tiêu chí không gian và thuộc tính.

Nhiệm vụ định vị: Sau khi quá trình truy vấn dữ liệu hoàn tất, hệ thống sẽ tiến hành định vị dữ liệu trực tiếp trên bản đồ nền. Vị trí và phạm vi của ảnh vệ tinh sẽ được hiển thị và đánh dấu trên bản đồ nền OpenLayers.



Hình 7. Nhiệm vụ định vị dữ liệu ảnh viễn thám trong hồ dữ liệu lớn.

Nhiệm vụ hiển thị dữ liệu: Sau khi quá trình truy vấn dữ liệu hoàn tất, hệ thống sẽ hiển thị thông tin ảnh viễn thám thông qua các thành phần của ReactJS.



Hình 8. Nhiệm vụ hiển thị thông tin dữ liệu ảnh viễn thám trong hồ dữ liệu lớn.

Các tác vụ đã được kiểm thử đồng thời trên 20 máy khách, với giới hạn băng thông mạng là 20 Mbps. Các máy khách gửi các truy vấn không gian và thuộc tính đến máy chủ và nhận được phản hồi nhanh theo thời gian thực, mang lại trải nghiệm mượt mà trên nền tảng web.

4. Kết luận

Bài báo đã đề xuất và triển khai một ứng dụng dịch vụ dựa trên web nhằm định vị và truy vấn thông tin ảnh vệ tinh trong một hồ dữ liệu lớn. Hệ thống được thiết kế theo kiến trúc đa tầng, đảm bảo khả năng xử lý các yêu cầu truy vấn cả về không gian và thuộc tính với thời gian phản hồi nhanh và độ chính xác cao. Dịch vụ đã được kiểm thử trên một hồ dữ liệu ảnh vệ tinh lớn, nhằm cung cấp dịch vụ truy vấn từ xa hiệu quả cho ảnh viễn thám và thông tin dữ liệu thông qua giao diện web. Kết quả thực nghiệm cho thấy các tác vụ như truy vấn siêu dữ liệu, định vị ảnh và hiển thị ảnh đều đạt gần thời gian thực và mang lại thao tác thuận tiện cho người dùng.

Kiến trúc dữ liệu và dịch vụ web được đề xuất trong nghiên cứu này tuy đơn giản nhưng mang tính đặc thù, được mô tả rõ ràng và có khả năng tái hiện, giúp cộng đồng nghiên cứu địa không gian dễ dàng áp dụng và phát triển các công nghệ thông tin địa không gian mạnh mẽ hơn cho các mục đích ứng dụng cụ thể.

Lời cảm ơn

Các tác giả xin gửi lời cảm ơn tới tổ chức Copernicus Open Access Hub và Sentinel Hub đã cung cấp nguồn dữ liệu ảnh vệ tinh phục vụ cho quá trình xây dựng hệ thống. Nhóm tác giả xin chân thành cảm ơn Viện Kỹ thuật Công trình Đặc biệt/Học viện Kỹ thuật Quân sự đã tài trợ kinh phí nghiên cứu. Bài báo này được thực hiện trong khuôn khổ đề tài mã số 25.03.03.

Cam kết của các tác giả

Tất cả các tác giả có tên trong bài báo cam kết sự đồng thuận và không có xung đột lợi ích trong công bố khoa học tại bài báo này.

Tài liệu tham khảo

- [1] Ning Guo et al., “An Efficient Tile-Pyramids Building Method for Fast Visualization of Massive Geospatial Raster Datasets”, J. Advances in Electrical and Computer Engineering, Vol. 16, No. 4, pp. , (2016).
- [2] Weipeng Jing, Dongxue Tian, “An improved distributed storage and query for remote sensing data”, Proc. International Conference on Identification, Information and Knowledge in the Internet of Things, ScienceDirect, No. 129, pp. 238-247, (2018).
- [3] Hui.Wu, Kun.Fu, “A Management of Remote Sensing Big Database on Standard Metadata file and Database Management System ”, Proc. Knowledge in the Internet of Things, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XLII-3/W10, pp. 653-657, (2020).
- [4] NASA Worldview, <https://worldview.earthdata.nasa.gov/>
- [5] USGS Earth Explorer, <https://earthexplorer.usgs.gov/>
- [6] Sentinel Hub, <https://www.sentinel-hub.com/>

Article © 2026 by Magazine of Geodesy - Cartography is licensed under CC BY 4.0

