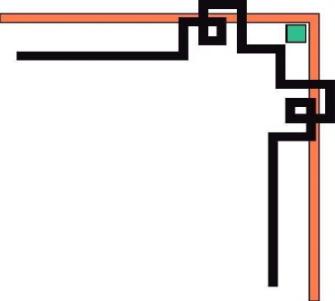
**BỘ GIÁO DỤC VÀ ĐÀO TẠO**



**TRƯỜNG ĐẠI HỌC TÀI NGUYÊN VÀ MÔI TRƯỜNG TP.HCM**

**KHOA CÔNG NGHỆ THÔNG TIN**

**A logo with a person holding a book

Description automatically generated**

***MÔN HỌC*: PHƯƠNG PHÁP NGHIÊN CỨU KHOA HỌC**

***ĐỀ TÀI*: NÉN DỮ LIỆU ẢNH VIỄN THÁM**

**BẰNG THUẬT TOÁN LZW VÀO POSTGRES/POSTGIS**

**Ở TPHCM NĂM 2016**

**Giảng viên hướng dẫn** : Ts. Dương Thị Thuý Nga

**Lớp** : 10\_ĐH\_TMĐT, 10\_ĐH\_TTMT

*Ngày 14, tháng 11, năm 2024*

**BỘ GIÁO DỤC VÀ ĐÀO TẠO**

**TRƯỜNG ĐẠI HỌC TÀI NGUYÊN VÀ MÔI TRƯỜNG TP.HCM**

**KHOA CÔNG NGHỆ THÔNG TIN**

**A logo with a person holding a book

Description automatically generated**

**THÀNH VIÊN NHÓM**

|  |  |  |
| --- | --- | --- |
| **Họ và tên** | **MSSV** | **Lớp** |
| Nguyễn Thanh Thản | 1050070027 | 10 ĐH TMĐT |
| Phạm Đình Nam | 1050070013 | 10 ĐH TMĐT |
| Lâm Vĩnh Kiện | 1050070008 | 10 ĐH TMĐT |
| Nguyễn Phước Sơn | 1050070025 | 10 ĐH TTMT |

*Ngày 14, tháng 11, năm 2024*

**Giới thiệu**

Trong thời đại dữ liệu lớn, ảnh viễn thám ngày càng trở thành một nguồn thông tin quan trọng cho nhiều lĩnh vực, bao gồm nông nghiệp, quản lý tài nguyên thiên nhiên, và quy hoạch đô thị. Tuy nhiên, kích thước của các ảnh viễn thám thường rất lớn, điều này gây khó khăn trong việc lưu trữ, quản lý và truy xuất dữ liệu trong các hệ quản trị cơ sở dữ liệu như PostgreSQL.

Việc nén ảnh viễn thám không chỉ giúp tiết kiệm không gian lưu trữ mà còn cải thiện tốc độ truy cập và xử lý dữ liệu. Tuy nhiên, việc nén ảnh cũng đặt ra nhiều thách thức, như việc lựa chọn định dạng nén phù hợp, đảm bảo chất lượng ảnh sau khi nén, và khả năng tương thích với các công cụ phân tích dữ liệu.

Chúng ta cần thảo luận về những khó khăn và lợi ích liên quan đến việc nén ảnh viễn thám vào trong PostgreSQL. Một số câu hỏi chính cần được xem xét bao gồm:

**Lời cảm ơn**

Em xin gửi lời cảm ơn chân thành đến Cô đã tận tình hướng dẫn, chỉ bảo và giúp đỡ em trong suốt quá trình thực hiện bài tiểu luận này. Nhờ sự hỗ trợ và những ý kiến quý báu của cô, em đã có thể hoàn thành bài viết một cách tốt nhất. Em rất trân trọng và biết ơn những đóng góp của Cô đối với việc học tập và phát triển kiến thức của em.

**Nhận xét và đánh giá**

….………………………………………………………………………………………

….………………………………………………………………………………………

….………………………………………………………………………………………

….………………………………………………………………………………………

….………………………………………………………………………………………

….………………………………………………………………………………………

….………………………………………………………………………………………

….………………………………………………………………………………………

….………………………………………………………………………………………

….………………………………………………………………………………………

….………………………………………………………………………………………

TP.HCM, ngày….,tháng….,năm 2024

Ký tên

TS. Dương Thị Thúy Nga

**MỤC LỤC**

[1. TỔNG QUAN VẤN ĐỀ NGHIÊN CỨU 1](#_Toc182399056)

[1.1. Tình hình nghiên cứu 1](#_Toc182399057)

[1.2. Đánh giá ưu, nhược điểm của các nghiên cứu trên 1](#_Toc182399058)

[Ưu Điểm 1](#_Toc182399059)

[Nhược Điểm 2](#_Toc182399060)

[1.3. Lý do chọn lựa đề tài 2](#_Toc182399061)

[1.4. Ý nghĩa khoa học của đề tài 3](#_Toc182399062)

[1.5. Ý nghĩa thực tiễn của đề tài 4](#_Toc182399063)

[2. MỤC TIÊU CỦA ĐỀ TÀI 4](#_Toc182399064)

[2.1. Mục tiêu chung 4](#_Toc182399065)

[2.2. Mục tiêu cụ thể 4](#_Toc182399066)

[2.2.1. Nghiên cứu và lựa chọn thuật toán nén 4](#_Toc182399067)

[2.2.1.1. Phân loại thuật toán 4](#_Toc182399068)

[2.2.1.2. Đánh giá thuật toán 4](#_Toc182399069)

[2.2.2. Xây dựng quy trình nén dữ liệu 4](#_Toc182399070)

[2.2.2.1. Chuẩn bị dữ liệu 4](#_Toc182399071)

[2.2.2.2. Quy trình nén 5](#_Toc182399072)

[2.2.2.3. Kiểm tra chất lượng 5](#_Toc182399073)

[2.2.3. Tích hợp vào PostgreSQL và PostGIS 5](#_Toc182399074)

[2.2.3.1. Thiết kế cơ sở dữ liệu 5](#_Toc182399075)

[2.2.3.2. Lưu trữ dữ liệu đã nén 5](#_Toc182399076)

[2.2.3.3. Tạo các hàm và truy vấn 5](#_Toc182399077)

[3. PHẠM VI CỦA ĐỀ TÀI 6](#_Toc182399078)

[4. ỨNG DỤNG 6](#_Toc182399079)

[4.1. Giới thiệu về LZW ( Lempel-Ziv-Welch ) 6](#_Toc182399080)

[4.1.1. Nguyên lý hoạt động của thuật toán LZW 6](#_Toc182399081)

[4.1.2. Nguyên lý hoạt động của thuật toán LZW 7](#_Toc182399082)

[4.1.3. Ứng dụng của LZW: 7](#_Toc182399083)

[4.2. Tạo table trong Postgre / Postgis 8](#_Toc182399084)

[4.3. Kết nối CSDL 9](#_Toc182399085)

[4.4. Thuật toán nén 9](#_Toc182399086)

[4.5. Thuật toán giải nén 10](#_Toc182399087)

[4.6. Thực hành 11](#_Toc182399088)

[4.6.1 Trước khi nén 11](#_Toc182399089)

[4.6.2. Kết quả sau khi nén 12](#_Toc182399090)

[4.6.3. Truy vấn từ database 12](#_Toc182399091)

[4.6.4. Kết quả 13](#_Toc182399092)

[5. KẾT LUẬN 13](#_Toc182399093)

[TÀI LIỆU THAM KHẢO 14](#_Toc182399094)

# 1. TỔNG QUAN VẤN ĐỀ NGHIÊN CỨU

## 1.1. Tình hình nghiên cứu

* Xác định đề tài và hình ảnh Viễn Thám năm 2016 để nghiên cứu
* Tải thành công PostgreSQL và PostGIS, sử dụng được những thao tác cơ bản
* Đã xác định được những bước cơ bản để nén ảnh Viễn Thám vào PostgreSQL và PostGIS

## 1.2. Đánh giá ưu, nhược điểm của các nghiên cứu trên

### Ưu Điểm

* **Tiết Kiệm Không Gian Lưu Trữ**

- Nén ảnh viễn thám giúp giảm kích thước tệp, tiết kiệm không gian trong cơ sở dữ liệu. Điều này đặc biệt quan trọng khi làm việc với lượng lớn dữ liệu viễn thám.

* **Cải Thiện Tốc Độ Truy Cập**

- Kích thước tệp nhỏ hơn giúp cải thiện tốc độ truy xuất dữ liệu, đồng thời giảm thời gian tải lên và tải xuống ảnh từ cơ sở dữ liệu.

* **Tối Ưu Hóa Hiệu Suất**

- Nén dữ liệu có thể giúp giảm tải cho hệ thống, cho phép các truy vấn và phân tích dữ liệu thực hiện nhanh hơn.

* **Bảo Quản Dữ Liệu Hiệu Quả**

- Việc nén ảnh giúp tổ chức và quản lý dữ liệu dễ dàng hơn, cho phép phát hiện và loại bỏ các tệp trùng lặp hoặc không cần thiết.

* **Dễ Dàng Tích Hợp**

- PostgreSQL hỗ trợ nhiều định dạng nén khác nhau, cho phép người dùng linh hoạt trong việc lựa chọn phương pháp nén phù hợp với nhu cầu cụ thể.

### Nhược Điểm

* **Giảm Chất Lượng Ảnh**

- Một số phương pháp nén, đặc biệt là nén mất mát (lossy compression), có thể làm giảm chất lượng của ảnh sau khi nén, ảnh hưởng đến khả năng phân tích và nhận diện thông tin.

* **Tăng Độ Phức Tạp Quản Lý**

- Việc nén và giải nén ảnh có thể làm tăng độ phức tạp trong việc quản lý cơ sở dữ liệu, yêu cầu thêm logic để xử lý dữ liệu nhị phân.

* **Thời Gian Xử Lý Tăng**

- Nén và giải nén dữ liệu có thể tiêu tốn thời gian và tài nguyên CPU, làm giảm hiệu suất tổng thể của hệ thống khi xử lý lượng lớn ảnh.

* **Khó Khăn Trong Việc Tìm Kiếm**

- Dữ liệu nén có thể làm cho việc tìm kiếm hoặc truy vấn dữ liệu trở nên khó khăn hơn, đặc biệt khi thông tin cần tìm kiếm không có trong phần metadata.

* **Tương Thích Hạn Chế**

- Không phải tất cả các định dạng nén đều được hỗ trợ tốt trong PostgreSQL, có thể gây khó khăn trong việc tích hợp với các công cụ phân tích hoặc phần mềm khác.

## 1.3. Lý do chọn lựa đề tài

Việc nén dữ liệu ảnh viễn thám trong PostgreSQL và PostGIS không chỉ mang lại lợi ích về mặt lưu trữ mà còn mở ra nhiều cơ hội ứng dụng thực tiễn trong các lĩnh vực như quản lý tài nguyên, nông nghiệp, quy hoạch đô thị và nghiên cứu môi trường. Sự kết hợp giữa công nghệ viễn thám và cơ sở dữ liệu mạnh mẽ như PostgreSQL giúp nâng cao khả năng ra quyết định và phát triển bền vững.

## 1.4. Ý nghĩa khoa học của đề tài

* **Đóng Góp Vào Khoa Học Dữ Liệu**

Đề tài này góp phần làm phong phú thêm lĩnh vực khoa học dữ liệu, đặc biệt là trong việc phát triển các phương pháp lưu trữ và quản lý dữ liệu lớn, như ảnh viễn thám, trong các hệ thống cơ sở dữ liệu hiện đại.

* **Nâng Cao Hiệu Quả Phân Tích Dữ Liệu**

Nén ảnh viễn thám không chỉ giúp tiết kiệm không gian lưu trữ mà còn cải thiện tốc độ truy cập, từ đó nâng cao khả năng phân tích và xử lý dữ liệu. Điều này giúp các nhà nghiên cứu và chuyên gia dễ dàng tiếp cận và sử dụng dữ liệu để đưa ra quyết định chính xác.

* **Khám Phá Các Phương Pháp Nén Mới**

Nghiên cứu đề tài này có thể dẫn đến việc phát triển các phương pháp nén ảnh mới, tối ưu hóa cho dữ liệu viễn thám, từ đó mở rộng các ứng dụng trong lĩnh vực viễn thám và khoa học địa lý.

* **Tích Hợp Hiệu Quả Với Công Nghệ Dữ Liệu**

Việc nghiên cứu nén ảnh trong PostgreSQL giúp khám phá khả năng tích hợp giữa các công nghệ lưu trữ dữ liệu và các ứng dụng phân tích, từ đó tạo ra các giải pháp mạnh mẽ hơn cho việc quản lý dữ liệu địa lý.

* **Hỗ Trợ Các Nghiên Cứu Ứng Dụng**

Dữ liệu viễn thám có vai trò quan trọng trong nhiều nghiên cứu ứng dụng như theo dõi biến đổi khí hậu, quản lý tài nguyên thiên nhiên và quy hoạch đô thị. Đề tài này giúp cung cấp nền tảng cho các nghiên cứu ứng dụng thực tiễn và các dự án phát triển bền vững.

* **Khuyến Khích Nghiên Cứu Liên Ngành:**

Đề tài này có thể khuyến khích sự hợp tác giữa các lĩnh vực như công nghệ thông tin, khoa học môi trường và địa lý, thúc đẩy các nghiên cứu liên ngành để giải quyết các vấn đề phức tạp trong quản lý dữ liệu viễn thám.

## 1.5. Ý nghĩa thực tiễn của đề tài

Đề tài này không chỉ đóng góp vào nền tảng lý thuyết mà còn thúc đẩy ứng dụng thực tiễn, góp phần vào việc tối ưu hóa việc sử dụng dữ liệu viễn thám trong nghiên cứu và ứng dụng, từ đó nâng cao hiệu quả quản lý và phát triển bền vững.

# 2. MỤC TIÊU CỦA ĐỀ TÀI

## 2.1. Mục tiêu chung

Nén ảnh gốc bằng thuật toán LZW vào Postgres và Postgis

## 2.2. Mục tiêu cụ thể

### 2.2.1. Nghiên cứu và lựa chọn thuật toán nén

#### 2.2.1.1. Phân loại thuật toán

**Nén tổn thất (Lossy)**: Ví dụ như JPEG2000, phù hợp cho ứng dụng yêu cầu kích thước nhỏ nhưng chấp nhận mất một phần chất lượng.

**Nén không tổn thất (Lossless)**: Ví dụ như PNG, GeoTIFF, giữ nguyên chất lượng ảnh nhưng thường có kích thước lớn hơn.

#### 2.2.1.2. Đánh giá thuật toán

**Chất lượng hình ảnh**: Sử dụng các chỉ số như PSNR (Peak Signal-to-Noise Ratio) và SSIM (Structural Similarity Index) để đánh giá chất lượng hình ảnh sau khi nén.

**Tốc độ nén**: Thử nghiệm thời gian cần thiết để nén và giải nén dữ liệu.

**Tính khả thi**: Xem xét khả năng tích hợp với PostgreSQL và PostGIS.

**Lựa chọn thuật toán**: Dựa trên các tiêu chí đánh giá, lựa chọn thuật toán nén tối ưu cho dữ liệu ảnh viễn thám.

### 2.2.2. Xây dựng quy trình nén dữ liệu

#### 2.2.2.1. Chuẩn bị dữ liệu

**Tiền xử lý**: Xử lý ảnh để loại bỏ nhiễu, điều chỉnh độ sáng và tương phản nếu cần thiết.

**Định dạng hóa**: Chuyển đổi dữ liệu ảnh về định dạng phù hợp cho nén.

#### 2.2.2.2. Quy trình nén

**Thực hiện nén**: Áp dụng thuật toán đã chọn để nén dữ liệu ảnh.

**Lưu trữ tạm thời**: Lưu trữ tệp nén trong một thư mục tạm thời trước khi chuyển vào cơ sở dữ liệu.

#### 2.2.2.3. Kiểm tra chất lượng

**Đánh giá sau nén**: So sánh chất lượng ảnh trước và sau khi nén để đảm bảo rằng dữ liệu vẫn giữ được thông tin quan trọng.

### 2.2.3. Tích hợp vào PostgreSQL và PostGIS

#### 2.2.3.1. Thiết kế cơ sở dữ liệu

**Tạo bảng**: Xây dựng cấu trúc bảng trong PostgreSQL để lưu trữ thông tin về ảnh viễn thám, bao gồm các trường như ID, tên ảnh, thông tin không gian, và dữ liệu nén.

#### 2.2.3.2. Lưu trữ dữ liệu đã nén

**Sử dụng PostGIS**: Tích hợp dữ liệu ảnh nén vào PostGIS để hỗ trợ truy vấn không gian và phân tích địa lý.

**Chuyển đổi và lưu trữ:** Chuyển đổi dữ liệu nén thành định dạng phù hợp và lưu trữ trong cơ sở dữ liệu.

#### 2.2.3.3. Tạo các hàm và truy vấn

**Hàm nén và giải nén**: Phát triển các hàm trong PostgreSQL để thực hiện nén và giải nén dữ liệu khi cần thiết.

**Truy vấn không gian**:Tạo các truy vấn để truy xuất dữ liệu ảnh theo tọa độ và các thuộc tính không gian khác.

Tóm lại, các mục tiêu cụ thể này nhằm xây dựng một hệ thống nén dữ liệu ảnh viễn thám hiệu quả, dễ dàng tích hợp và sử dụng trong môi trường PostgreSQL và PostGIS

# 3. PHẠM VI CỦA ĐỀ TÀI

Đề tàinghiên cứu nén ảnh trong PostgreSQL và PostGis trong phạm vi trong thành phố HCM năm 2016.

# ỨNG DỤNG

## Giới thiệu về LZW ( Lempel-Ziv-Welch )

Thuật toán LZW (Lempel-Ziv-Welch) là một phương pháp nén dữ liệu không mất thông tin, được phát triển bởi Abraham Lempel, Jacob Ziv và Terry Welch vào năm 1984. Thuật toán này là một sự cải tiến của các thuật toán nén trước đó như LZ78 và LZ77, với mục tiêu giảm thiểu kích thước của tệp tin mà không làm mất đi bất kỳ dữ liệu nào.

LZW được sử dụng rộng rãi trong các ứng dụng nén tệp, như trong định dạng tệp GIF (Graphics Interchange Format), TIFF (Tagged Image File Format), và trong một số giao thức nén như trong công nghệ nén dữ liệu của các máy fax. Nó còn được ứng dụng trong các hệ thống nén dữ liệu, như trong phần mềm WinZip, hoặc trên các giao thức truyền thông như HTTP.

### Nguyên lý hoạt động của thuật toán LZW

LZW là một thuật toán nén không mất thông tin, sử dụng phương pháp thay thế các chuỗi con (substrings) trong dữ liệu ban đầu bằng các mã số (code). Dưới đây là các bước cơ bản của thuật toán:

**Khởi tạo từ điển (Dictionary)**:

Thuật toán bắt đầu với một từ điển chứa tất cả các ký tự đơn giản (từ bảng mã ASCII hoặc bảng ký tự bạn đang làm việc). Ví dụ, với bảng ASCII, từ điển sẽ chứa tất cả các ký tự từ 0 đến 255.

**Xử lý chuỗi dữ liệu**:

Thuật toán quét qua chuỗi dữ liệu đầu vào và tìm kiếm các chuỗi con dài hơn mà đã có trong từ điển.

Nếu chuỗi con hiện tại đã có trong từ điển, thuật toán tiếp tục mở rộng chuỗi con đó bằng cách thêm một ký tự tiếp theo vào chuỗi.

Khi một chuỗi con không có trong từ điển, thuật toán sẽ ghi mã số (code) của chuỗi con đã tìm thấy trong từ điển và thêm chuỗi con mới vào từ điển.

**Nén và lưu trữ**:

Các chuỗi con đã được thay thế bằng mã số được ghi lại và lưu trữ dưới dạng dữ liệu nén. Điều này giúp giảm kích thước dữ liệu.

**Giải nén**:

Để giải nén, thuật toán cần phải sử dụng từ điển ban đầu, cộng với các chuỗi con được tạo ra trong quá trình nén.

Dữ liệu nén sau khi được giải mã sẽ trả lại chuỗi dữ liệu gốc.

### 4.1.2. Nguyên lý hoạt động của thuật toán LZW

**Ưu điểm của LZW:**

**Không mất dữ liệu**: LZW là một thuật toán nén không mất thông tin, có thể tái tạo lại dữ liệu gốc từ dữ liệu nén.

**Hiệu quả với dữ liệu có sự lặp lại**: Thuật toán đặc biệt hiệu quả khi xử lý các chuỗi dữ liệu có sự lặp lại, như văn bản hoặc hình ảnh có nhiều pixel giống nhau.

**Dễ dàng triển khai**: Thuật toán tương đối đơn giản và dễ triển khai trong các hệ thống nén dữ liệu.

**Nhược điểm của LZW:**

**Tăng kích thước dữ liệu với một số trường hợp**: Trong một số tình huống, đặc biệt khi dữ liệu không có nhiều sự lặp lại, thuật toán LZW có thể tạo ra một từ điển rất lớn, làm tăng kích thước của dữ liệu nén thay vì giảm.

**Hiệu suất thấp với dữ liệu không có tính lặp lại**: Nếu dữ liệu không có nhiều chuỗi con lặp lại, thuật toán không đạt hiệu quả tốt như các phương pháp nén khác.

### 4.1.3. Ứng dụng của LZW:

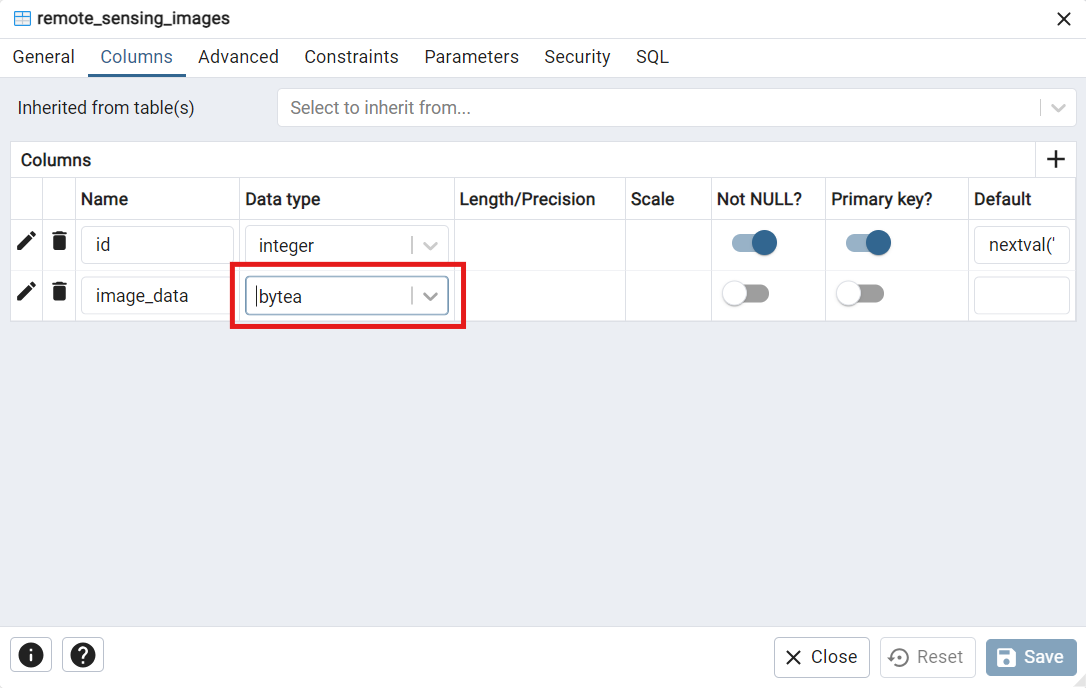
**GIF**: Định dạng hình ảnh GIF sử dụng LZW để nén ảnh, giúp giảm dung lượng mà vẫn giữ được chất lượng hình ảnh.

**TIFF**: Tương tự như GIF, TIFF cũng sử dụng LZW cho việc nén ảnh không mất dữ liệu.

**Các hệ thống nén**: LZW được sử dụng trong các phần mềm nén dữ liệu như PKZIP và WinZip.

**Máy fax**: Một số giao thức fax sử dụng thuật toán LZW để nén văn bản và hình ảnh trước khi gửi.

## 4.2. Tạo table trong Postgre / Postgis

****

- Kiểu dữ liệu binary large object

## 4.3. Kết nối CSDL

****

## 4.4. Thuật toán nén

****

## 4.5. Thuật toán giải nén

****

**-** Code sẽ thực thi theo các bước sau:

+ Bước 1: Mở ảnh viễn thám muốn nén theo đường dẫn đã lưu. (Mở ảnh TIFF)  
+ Bước 2: Tiến hành đọc file ảnh với lệnh code “Image.open”. In ra kích thước file ảnh ban đầu nếu cần.

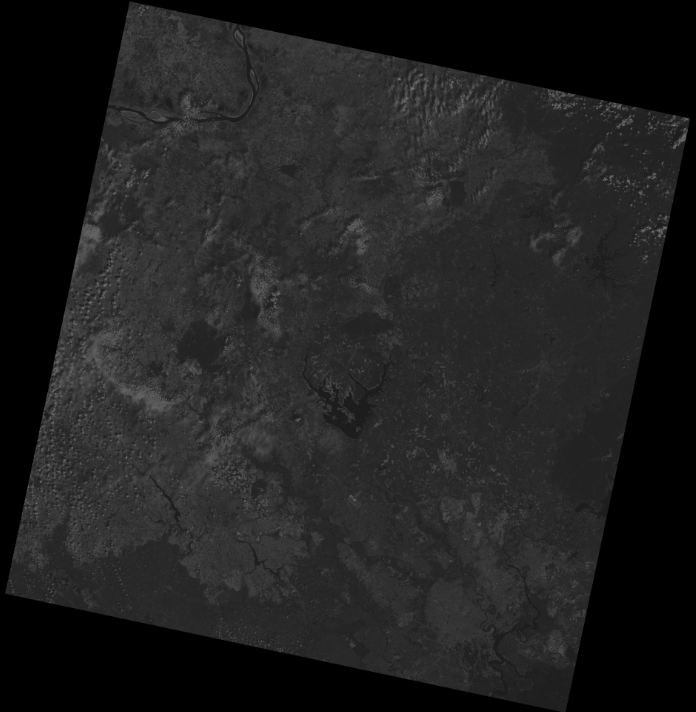
+ Bước 3: Tiến hành nén ảnh theo thuật toán LZW với lệnh code được cấu hình thuật toán nén “compression = “tiff\_lzw”.

+ Bước 4: Lấy dữ liệu ảnh nhị phân để lưu database PorgreSQL.

+ Bước 5: In kích thước file ảnh sau khi nén.  
+ Bước 6: Nén thành công hoặc không thành công sẽ in ra thông báo.

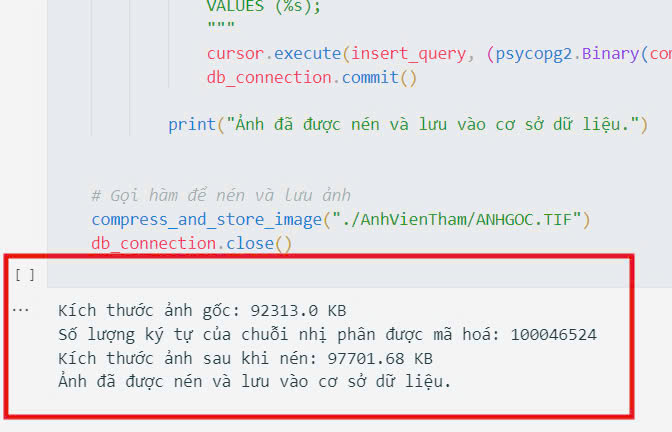
## 4.6. Thực hành

### 4.6.1 Trước khi nén

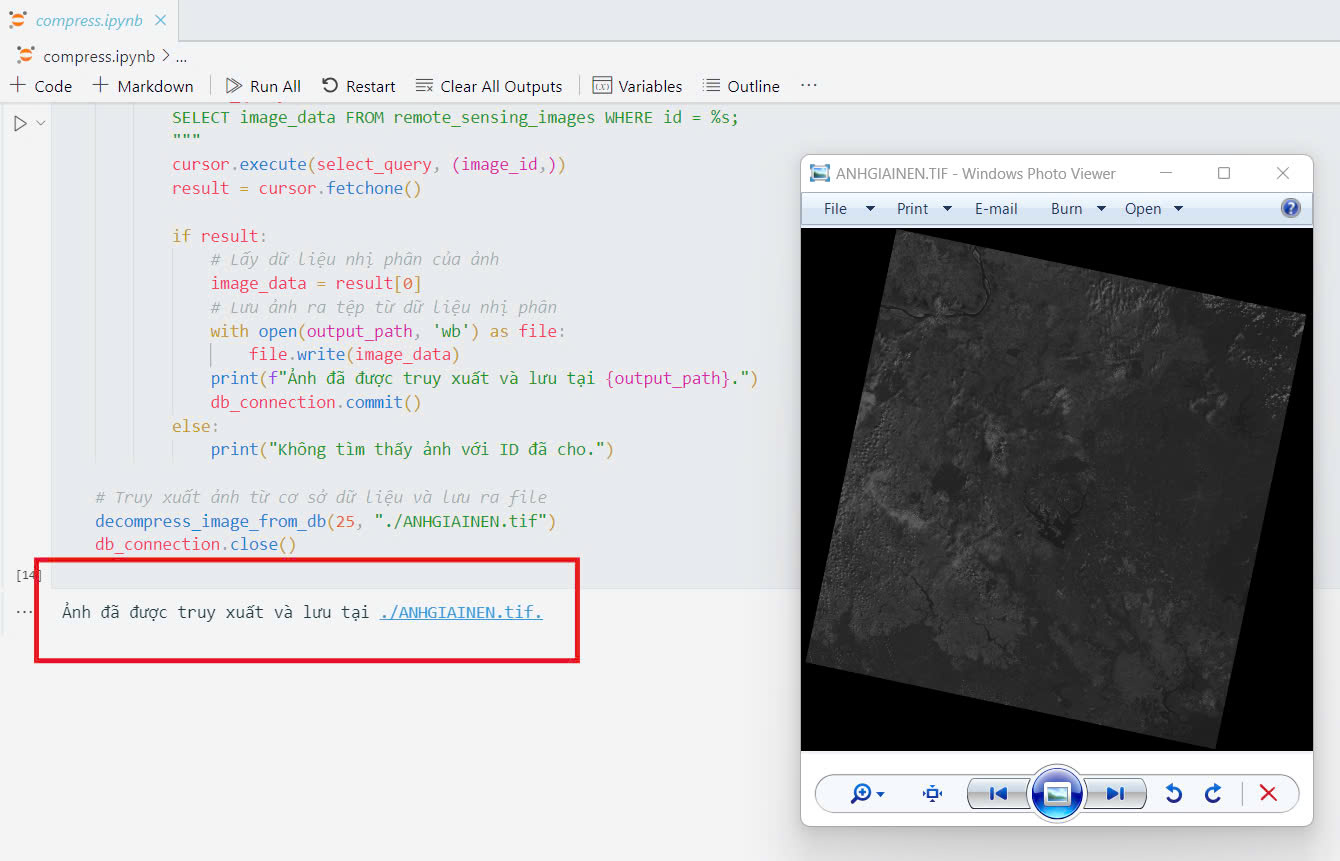
****

**-** Đây là ảnh Band 1 khu vực Tp Hồ Chí Minh năm 2016. Chúng ta sẽ tiến hành nén ảnh này.

### 4.6.2. Kết quả sau khi nén

****

### 4.6.3. Truy vấn từ database

****

### 4.6.4. Kết quả

|  |  |  |
| --- | --- | --- |
| **Ảnh** | **Trước nén** | **Sau nén** |
| **.tif** | **92MB** | **97MB** |

- Sau khi nghiên cứu nén bằng thuật toán LZW. Kết quả cho thấy thuật toán không hiệu quả. Vì thuật toán LZW tạo ra 1 file để lưu thông tin giải mã khiến cho file nén trở nên lớn thêm.

# 5. KẾT LUẬN

Tóm lại, LZW là một phương pháp nén hữu ích cho ảnh viễn thám, đặc biệt khi yêu cầu không mất dữ liệu. Tuy nhiên, cần phải cân nhắc khi làm việc với dữ liệu phức tạp hơn, nơi các phương pháp nén khác có thể hiệu quả hơn. Vì nén không mất dữ liệu nên chất lượng và kích thước ảnh vẫn giữ nguyên hoặc có thể lớn hơn khí lưu vào cơ sở dữ liệu. Thuật toán sẽ phù hợp nén ảnh viễn thám với chất lượng nhỏ.

# TÀI LIỆU THAM KHẢO

1. What is remote sensing and what is it used for? | <https://www.usgs.gov/faqs/what-remote-sensing-and-what-it-used>

2. What is Remote Sensing Image? | <https://www.sciencedirect.com/topics/computer-science/remote-sensing-image>

3. Database Postgres | <https://www.postgresql.org/docs/>

4. High-resolution image compression algorithms in remote sensing imaging | <https://www.sciencedirect.com/science/article/abs/pii/S0141938223000951>

5. Remote Sensing Image Compression: A Review | <https://ieeexplore.ieee.org/document/7153923>