

ĐẠI HỌC BÁCH KHOA HÀ NỘI

KHOA TOÁN - TIN

oOo



VŨ VĂN HUY

huy.vv216931@sis.hust.edu.vn

XÂY DỰNG HỆ THỐNG TĂNG CƯỜNG CHIA SẺ DỮ LIỆU

ĐỒ ÁN 1

Chuyên ngành: HTTT - QL

HÀ NỘI - 2024

ĐẠI HỌC BÁCH KHOA HÀ NỘI

KHOA TOÁN - TIN

oOo



XÂY DỰNG HỆ THỐNG TĂNG CƯỜNG CHIA SẺ DỮ LIỆU

ĐỒ ÁN 1

Chuyên ngành: HTTT - QL

Giảng viên hướng dẫn: TS. Nguyễn Thị Thanh Huyền Chữ ký của GVHD

Sinh viên thực hiện: Vũ Văn Huy

Lớp: Hệ thống thông tin 01 – K66

HÀ NỘI - 2024

NHẬN XÉT CỦA GIẢNG VIÊN HƯỚNG DẪN

1. Mục tiêu và nội dung của đề án

.....

.....

.....

.....

.....

.....

2. Kết quả đạt được

.....

.....

.....

.....

.....

.....

3. Ý thức làm việc của sinh viên

.....

.....

.....

.....

.....

.....

Hà Nội, ngày ... tháng ... năm 2024

Giảng viên hướng dẫn

Mục lục

| | |
|---|-----------|
| Mở đầu | 1 |
| Lời cảm ơn | 2 |
| Chương 1 Kiến thức cơ sở | 3 |
| 1.1 Big data | 3 |
| 1.2 Tổng quan về hồ dữ liệu | 4 |
| 1.2.1 Kiến trúc hồ dữ liệu | 4 |
| 1.2.2 Lợi ích của hồ dữ liệu | 5 |
| 1.3 Công nghệ sử dụng | 7 |
| 1.3.1 Amazon S3 | 7 |
| 1.3.2 Amazon Cognito | 8 |
| 1.3.3 AWS Lambda | 8 |
| 1.3.4 Amazon Athena | 9 |
| 1.3.5 Amazon DynamoDB | 10 |
| 1.3.6 Amazon QuickSight | 10 |
| Chương 2 Khảo sát hệ thống | 12 |
| 2.1 Khảo sát thực tế | 12 |
| 2.1.1 Hiện trạng | 12 |
| 2.1.2 Quy trình chia sẻ dữ liệu | 13 |
| 2.2 Đánh giá hệ thống hiện tại | 14 |
| 2.2.1 Tiêu chí đánh giá | 14 |
| 2.2.2 Các hệ thống hiện có | 15 |
| 2.2.3 Tại sao không dùng hệ thống sẵn có? | 16 |
| 2.3 Đề xuất cho hệ thống mới | 17 |
| 2.3.1 Mục tiêu hệ thống | 17 |
| 2.3.2 Các bước triển khai hệ thống mới | 18 |
| 2.3.3 Phạm vi hệ thống | 18 |
| 2.3.4 Đối tượng sử dụng | 19 |

| | | |
|-----------------|---|-----------|
| 2.3.5 | Lợi ích của hệ thống mới | 19 |
| Chương 3 | Phân tích hệ thống | 20 |
| 3.1 | Sơ đồ phân cấp chức năng | 20 |
| 3.2 | ERD | 23 |
| 3.3 | Sơ đồ usecase | 24 |
| 3.4 | Đặc tả usecase | 26 |
| 3.4.1 | Đặc tả usecase đăng nhập, đăng ký | 26 |
| 3.4.2 | Đặc tả usecase phân quyền | 27 |
| 3.4.3 | Đặc tả usecase chia sẻ file | 27 |
| 3.4.4 | Đặc tả usecase quên mật khẩu | 28 |
| 3.4.5 | Đặc tả usecase thống kê | 29 |
| 3.5 | Sơ đồ kiến trúc phân lớp | 30 |
| 3.6 | Sơ đồ hoạt động | 31 |
| Chương 4 | Thiết kế hệ thống | 35 |
| 4.1 | Kiến trúc hệ thống | 35 |
| 4.2 | Kiến trúc vận hành | 36 |
| 4.3 | Cơ sở dữ liệu | 37 |
| 4.4 | Thiết kế giao diện | 40 |
| Chương 5 | Triển khai hệ thống | 41 |
| 5.1 | Giao diện đăng nhập, đăng ký | 41 |
| 5.2 | Quên mật khẩu | 43 |
| 5.3 | Giao diện trang chủ | 44 |
| 5.4 | Các bước chia sẻ File | 46 |
| 5.5 | Luồng hoạt động các chức năng | 48 |
| 5.6 | Dashboard phân tích | 49 |
| Chương 6 | Kết luận | 53 |
| | Tài liệu tham khảo | 55 |

Danh sách bảng

| | | |
|-----|-------------------------------|----|
| 4.1 | Mô tả bảng Account | 38 |
| 4.2 | Mô tả bảng Group | 38 |
| 4.3 | Mô tả bảng User | 38 |
| 4.4 | Mô tả bảng Role | 39 |
| 4.5 | Mô tả bảng File | 39 |
| 4.6 | Mô tả bảng Category | 39 |
| 4.7 | Mô tả bảng Action | 39 |
| 4.8 | Mô tả bảng Share | 40 |

Danh sách hình vẽ

| | | |
|------|--|----|
| 1.1 | Big Data | 3 |
| 1.2 | Data Lake | 4 |
| 1.3 | S3 | 7 |
| 1.4 | Cognito | 8 |
| 1.5 | Lambda | 9 |
| 1.6 | Athena | 9 |
| 1.7 | DynamoDB | 10 |
| 1.8 | QuickSight | 10 |
| 2.1 | Kiến trúc google drive | 15 |
| 3.1 | Sơ đồ phân cấp chức năng | 20 |
| 3.2 | Sơ đồ thực thể liên kết | 23 |
| 3.3 | Usecase đăng nhập đăng ký | 24 |
| 3.4 | Usecase phân quyền người dùng | 24 |
| 3.5 | Usecase chia sẻ file | 25 |
| 3.6 | Usecase chia sẻ file | 25 |
| 3.7 | Usecase thống kê | 26 |
| 3.8 | Phân lớp các tầng ứng dụng | 30 |
| 3.9 | Chức năng phân quyền | 31 |
| 3.10 | Chức năng quên mật khẩu | 32 |
| 3.11 | Chức năng chia sẻ file | 33 |
| 3.12 | Chức năng xem dashboard | 34 |
| 4.1 | Tổng quan kiến trúc hệ thống | 35 |
| 4.2 | Kiến trúc vận hành các service | 36 |
| 4.3 | Cơ sở dữ liệu | 37 |
| 5.1 | Giao diện đăng nhập | 41 |

| | | |
|------|---|----|
| 5.2 | Giao diện đăng ký | 42 |
| 5.3 | Xác thực mã code từ gmail | 42 |
| 5.4 | Thông tin các user sử dụng hệ thống | 43 |
| 5.5 | Chức năng quên mật khẩu | 43 |
| 5.6 | Nhập mã và mật khẩu mới | 44 |
| 5.7 | Trang chủ | 44 |
| 5.8 | Nơi chứa các file đã chia sẻ | 45 |
| 5.9 | Dữ liệu bảng file | 45 |
| 5.10 | Các file được chia sẻ | 45 |
| 5.11 | Bước 1: Chọn file | 46 |
| 5.12 | Bước 2: Nhập mail người muốn chia sẻ | 46 |
| 5.13 | Bước 3: Thêm giới hạn về ngày và lượt tải | 47 |
| 5.14 | Bước 4: Review | 47 |
| 5.15 | Dữ liệu file chia sẻ | 48 |
| 5.16 | Luồng xác thực tài khoản | 48 |
| 5.17 | Luồng tải xuống file | 49 |
| 5.18 | Dashboard chung | 49 |
| 5.19 | Dashboard phần 1 | 50 |
| 5.20 | Dashboard phần 2 | 51 |
| 5.21 | Dashboard phần 3 | 52 |

Bảng ký hiệu và chữ viết tắt

| | |
|------------|--|
| AWS | A mazon W eb S ervices (Nền tảng dịch vụ đám mây Amazon). |
| BI | B usiness I ntelligence (Kinh doanh thông minh). |
| ERD | E ntity R elationship D iagram (Sơ đồ thực thể liên kết). |
| UI | U ser I nterface (Giao diện người dùng). |
| UX | U ser E xperience (Trải nghiệm người dùng). |

Mở đầu

Bước vào kỷ nguyên số, việc chia sẻ thông tin trở nên thiết yếu hơn bao giờ hết. Dữ liệu không chỉ là một tài nguyên quý giá mà còn là **"nền tảng"** cho sự phát triển và thịnh vượng trong thế giới kỹ thuật số ngày nay. Nhu cầu kết nối, cộng tác và truy cập dữ liệu từ xa ngày càng tăng, thúc đẩy sự ra đời của các hệ thống quản lý và chia sẻ file. Để xây dựng một hệ thống quản lý chia sẻ dữ liệu an toàn, hiệu quả, đáp ứng nhu cầu của cả cá nhân và tổ chức, báo cáo này nghiên cứu hệ thống quản lý chia sẻ file trong các tổ chức nhằm đề xuất một giải pháp hiệu quả và bảo mật hơn. Nội dung báo cáo bao gồm 6 chương:

- **Chương I:** Kiến thức cơ sở.
- **Chương II:** Khảo sát hệ thống.
- **Chương III:** Phân tích hệ thống.
- **Chương IV:** Thiết kế hệ thống.
- **Chương V:** Triển khai hệ thống.
- **Chương VI:** Kết luận.

Hệ thống quản lý chia sẻ dữ liệu được thiết kế thông minh và có tính ứng dụng cao, hứa hẹn sẽ là giải pháp tối ưu cho nhu cầu chia sẻ thông tin trong thời đại công nghệ số.

Hà Nội, tháng 05 năm 2024

Sinh viên

Vũ Văn Huy

Lời cảm ơn

Đồ án này đã trở thành một hành trình đáng nhớ và đầy ấn tượng trong học kỳ vừa qua, nơi em đã dành trọn vẹn đam mê và nỗ lực không ngừng nghỉ của mình.

Trước hết, em muốn gửi lời cảm ơn chân thành tới quý thầy cô và các bạn bè tại Khoa Toán - Tin. Suốt quá trình thực hiện đồ án, em đã nhận được sự quan tâm, hỗ trợ nhiệt tình từ mọi người. Em đặc biệt trân trọng kiến thức quý báu mà các thầy cô đã không ngần ngại chia sẻ, những bài học ấy đã vô cùng quan trọng đối với sự thành công của đồ án này. Xin được bày tỏ lòng biết ơn sâu sắc tới cô TS. Nguyễn Thị Thanh Huyền, người đã tận tình hướng dẫn và đồng hành cùng em không chỉ trong đồ án này mà còn ở những môn học cốt lõi như Cơ sở dữ liệu, Kỹ thuật lập trình. Nhờ có sự hướng dẫn đầy tâm huyết của cô, em đã tích lũy được nhiều kinh nghiệm và hiểu biết sâu sắc trong ngành.

Tuy nhiên, mặc dù đã cố gắng hết sức, đồ án vẫn không tránh khỏi những thiếu sót do hạn chế về thời gian và kiến thức. Em rất mong nhận được những ý kiến đóng góp thiết thực từ quý thầy cô và các bạn sinh viên để có thể hoàn thiện đồ án này một cách tốt nhất.

Xin chân thành cảm ơn.

Chương 1

Kiến thức cơ sở

1.1 Big data



Hình 1.1: Big Data

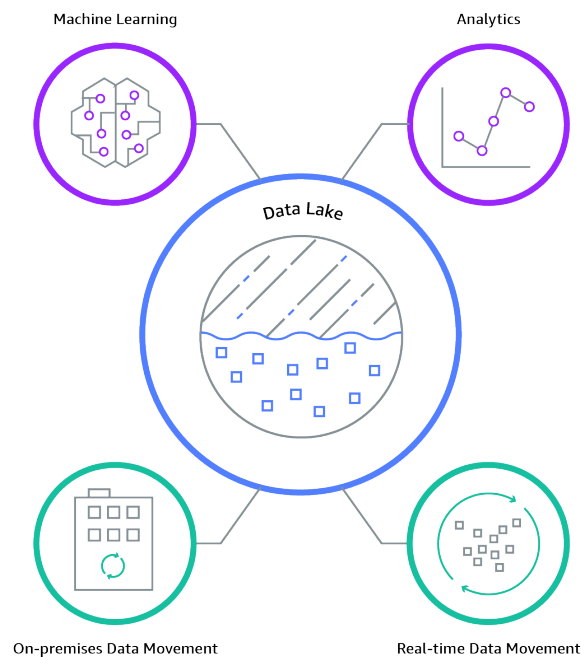
Big Data là khái niệm dùng để nói đến một lượng dữ liệu rất lớn, đa dạng, phức tạp đòi hỏi các công cụ quản lý và phân tích/xử lý trong một khoảng thời gian hợp lý. Big data không chỉ liên quan đến dữ liệu khổng lồ mà còn về tốc độ dữ liệu được sinh ra (có cấu trúc, không cấu trúc và bán cấu trúc).

Trong đề án này sẽ thực hiện xây dựng 1 hồ dữ liệu để lưu trữ dữ liệu chia sẻ của người dùng. Vậy **bao nhiêu dữ liệu mới đủ gọi là ‘BIG’?**

Không có một ngưỡng cụ thể nào xác định Big Data vì điều này phụ thuộc vào

khả năng khai thác và xử lý của công nghệ hiện tại cũng như nhu cầu riêng của từng tổ chức. Lượng dữ liệu được coi là "lớn" vào thời điểm này có thể trở nên bình thường trong vài năm tới nhờ vào sự tiến bộ của công nghệ và kỹ thuật xử lý dữ liệu. Tuy nhiên, nhìn chung, khi dữ liệu đạt đến mức petabyte trở lên và không thể xử lý hiệu quả bằng các cơ sở dữ liệu truyền thống, nó thường được xếp vào danh mục Big Data.

1.2 Tổng quan về hồ dữ liệu



Hình 1.2: Data Lake

Data Lake là một kho lưu trữ tập trung, cho phép lưu trữ tất cả dữ liệu có cấu trúc và phi cấu trúc ở bất kỳ quy mô nào. Hoạt động như một "hồ chứa", nó lưu trữ dữ liệu thô từ nhiều nguồn khác nhau, giúp đáp ứng nhu cầu xử lý và phân tích liên quan đến Big Data một cách hiệu quả.

1.2.1 Kiến trúc hồ dữ liệu

Mặc dù có nhiều biến thể, kiến trúc của Data Lake thường tuân theo ba nguyên tắc chính sau:

- Không cần chuyển đổi dữ liệu: Tất cả dữ liệu từ nhiều nguồn khác nhau có

thể được tải và lưu trữ trực tiếp vào Data Lake mà không cần qua quá trình chuyển đổi.

- Lưu trữ dữ liệu gốc: Dữ liệu khi nhập vào sẽ được giữ nguyên trạng, không có bất kỳ thay đổi hoặc xử lý nào.
- Chuyển đổi dữ liệu có cấu trúc: Dữ liệu sẽ được chuyển đổi sau đó để đáp ứng các yêu cầu truy vấn và phân tích của người dùng.

1.2.2 Lợi ích của hồ dữ liệu

Data Lake mang lại nhiều lợi ích quan trọng, đặc biệt trong lĩnh vực kinh doanh. Nó cung cấp nền tảng vững chắc cho khoa học dữ liệu và các ứng dụng phân tích tiên tiến, giúp doanh nghiệp quản lý hoạt động hiệu quả, phát hiện cơ hội mới và dự đoán xu hướng.

Bằng cách kết hợp và tích hợp các tập dữ liệu đa dạng từ nhiều nguồn khác nhau vào một kho lưu trữ tập trung duy nhất, Data Lake loại bỏ tình trạng dữ liệu bị phân mảnh và chia cắt (data silos). Điều này tạo điều kiện cho các nhóm khoa học dữ liệu có cái nhìn toàn diện về dữ liệu, đồng thời đơn giản hóa việc tìm kiếm và khai thác dữ liệu cần thiết. Dưới đây là những lợi ích thiết thực nhất của Data Lake:

- Hỗ trợ mô hình dữ liệu và phân tích nhanh chóng: Cho phép nhà khoa học dữ liệu và người dùng xây dựng mô hình dữ liệu và thực hiện các phân tích, truy vấn nhanh chóng và hiệu quả.
- Khả năng mở rộng cao: Data Lake có khả năng mở rộng lớn, giúp doanh nghiệp tiết kiệm chi phí.
- Linh hoạt trong thiết kế lược đồ và chuyển đổi dữ liệu: Thiết kế lược đồ và chuyển đổi dữ liệu có thể được trì hoãn cho đến khi làm rõ được mục đích và nhu cầu kinh doanh.
- Tương thích với nhiều phương pháp phân tích: Data Lake hỗ trợ nhiều phương pháp phân tích khác nhau bao gồm phân tích thống kê, machine learning (học máy), khai thác văn bản, mô hình dự đoán, truy vấn SQL, v.v.

Với nhiều lợi ích vượt trội, Data Lake đóng vai trò trọng yếu trong chiến lược quản trị dữ liệu và trở thành công cụ không thể thiếu để khai thác hiệu quả tài sản dữ liệu quý giá của doanh nghiệp.

1.3 Công nghệ sử dụng

1.3.1 Amazon S3



Hình 1.3: S3

AWS S3 - một kho lưu trữ cho internet. Nó được thiết kế để cung cấp lưu trữ dung lượng lớn, chi phí thấp trên nhiều khu vực địa lý. Amazon S3 cung cấp cho các nhà phát triển và nhóm CNTT bộ lưu trữ đối tượng An toàn, Bền bỉ và có khả năng mở rộng cao.

- Đạt được độ bền lên đến 11 số 9: 99,999999999 %.
- Độ sẵn sàng 4 số 9: 99,99 %
- Upload file thông qua giao thức STTP.
- Làm việc với S3 thông qua REST API(HTTP).

S3 gần như có thể lưu trữ dữ liệu không giới hạn (Theo quảng cáo của AWS). S3 có các cấp lưu trữ, tùy theo nhu cầu, theo tính chất dữ liệu, thời gian truy cập thì giá cả sẽ rẻ hơn: Standard, INT, One-Zone, S-IA, Glacier. Trong cùng 1 bucket các object có thể nằm ở các class khác nhau.

Hỗ trợ CORS và cơ chế lưu vô hạn của S3: với key = /images/photo.gif người ta sẽ lấy giá trị key của phần trước để người ta băm nó ra và người ta lấy nó làm partition key.

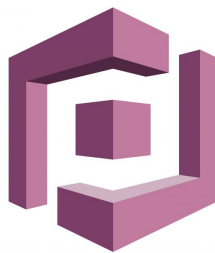
Khi chúng ta đưa số lượng object vào nó vượt quá số lượng partition thì s3 nó sẽ giải ra nó băm ra những partition khác. Chứ không thể nào nó lưu dc vô hạn được. Mỗi object trong s3 thực chất đều ngang hàng, không phân cấp và được gán 1 object key. Ví dụ: menu/sample.jpg. Chúng ta phải hiểu cái folder nó chỉ ảo thôi để mình quen đối với việc quản lý file. Cách làm như vậy để tăng hiệu suất truy cập file.

Versioning tính năng được hiểu là tạo ra 1 version mới khi thêm 1 file thì nó sẽ tạo ra 1 file khác với ID khác. Nếu xóa cái version gốc đi thì khi get sẽ không get được cái object cũ mà phải truyền vào đúng ID của object chưa xóa thì nó mới cho GET.

Versioning cho phép khôi phục đối tượng sau khi vô tình xóa hay ghi đè, có thể hỗ trợ việc tấn công ransomware/encryption attack (hiểu là khi bị tấn công thay vì bị xóa đi dữ liệu cũ thì nó sẽ khôi phục phiên bản dữ liệu trước đó).

S3 cung cấp một tính năng gọi là **PresignURL** để tạo liên kết tải xuống tạm thời, đây là cách hiệu quả và an toàn để tải lên và tải xuống dữ liệu được người dùng ủy quyền.

1.3.2 Amazon Cognito



Hình 1.4: Cognito

Việc quản lý người dùng tương đối phức tạp: tạo form đăng nhập, đăng ký, quản lý hành vi người dùng... AWS đã thử nghiệm trên hàng triệu người dùng và cho ra đời Cognito.

- Để quản quản lý người dùng sử dụng: **User Pool** người dùng đăng nhập trực tiếp bằng tên người dùng và mật khẩu thông qua một bên thứ 3 như facebook, amazon, google.
- Để cấp quyền cho người dùng sử dụng: **Identity Pool** người dùng được phép làm gì trong ứng dụng chúng ta xây dựng.

1.3.3 AWS Lambda

AWS Lambda là dịch vụ điện toán phi máy chủ (serverless) do Amazon cung cấp, cho phép bạn chạy mã mà không cần quản lý máy chủ. Với Lambda, bạn chỉ



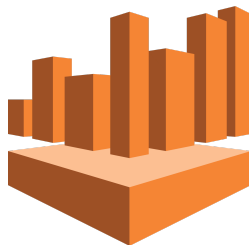
Hình 1.5: Lambda

cần tải lên mã của mình và AWS sẽ lo phần còn lại, bao gồm việc cung cấp tài nguyên máy tính và tự động scale mã của bạn dựa trên nhu cầu.

Ưu điểm của AWS Lambda:

- Dễ sử dụng: Bạn chỉ cần tải lên mã của mình và Lambda sẽ lo phần còn lại.
- Tiết kiệm chi phí: Bạn chỉ trả tiền cho tài nguyên mà bạn sử dụng.
- Tự động scale: Lambda sẽ tự động scale mã của bạn dựa trên nhu cầu, giúp bạn tránh lãng phí tài nguyên.
- Nhanh chóng: Lambda có thể khởi chạy mã của bạn trong mili giây.
- Tin cậy: Lambda là dịch vụ có độ tin cậy cao, đảm bảo mã của bạn luôn chạy.
- Hỗ trợ nhiều ngôn ngữ: Lambda hỗ trợ nhiều ngôn ngữ lập trình phổ biến như Java, Python, Node.js, C# và Go.

1.3.4 Amazon Athena



Hình 1.6: Athena

Amazon Athena là dịch vụ truy vấn dữ liệu lớn được cung cấp bởi AWS. Athena cho phép truy vấn dữ liệu được lưu trữ trong Amazon S3 một cách dễ dàng, nhanh chóng và hiệu quả bằng cách sử dụng ngôn ngữ truy vấn SQL tiêu chuẩn.

Athena cung cấp một cách hiệu quả để truy vấn S3 Access Logs nhằm trích xuất nội dung tải xuống và mức sử dụng băng thông.

1.3.5 Amazon DynamoDB



Hình 1.7: DynamoDB

DynamoDB là một dịch vụ cơ sở dữ liệu phi quan hệ NoSQL được quản lý hoàn toàn, có thể mở rộng cao và hiệu suất cao cho các ứng dụng cần tốc độ và khả năng mở rộng. Cơ sở dữ liệu này phản hồi chỉ trong vài mili giây ở mọi quy mô.

Khi quy mô của cơ sở dữ liệu tăng hoặc giảm, DynamoDB sẽ tự động thay đổi quy mô để điều chỉnh phù hợp với thay đổi về công suất, trong khi vẫn duy trì hiệu năng ổn định. Nhờ đó, đây là sự lựa chọn phù hợp với các trường hợp sử dụng yêu cầu hiệu năng cao và quy mô linh hoạt.

1.3.6 Amazon QuickSight



Hình 1.8: QuickSight

QuickSight là một dịch vụ phân tích kinh doanh (BI) được quản lý hoàn toàn, giúp bạn dễ dàng khám phá, phân tích và chia sẻ dữ liệu của mình. Sử dụng QuickSight giúp tiết kiệm thời gian để cho ra các báo cáo nhanh chóng.

Ứng dụng:

- Phân tích dữ liệu kinh doanh: QuickSight giúp bạn dễ dàng phân tích dữ liệu

kinh doanh để đưa ra quyết định sáng suốt hơn.

- Phân tích dữ liệu khoa học dữ liệu: QuickSight giúp bạn dễ dàng phân tích dữ liệu khoa học dữ liệu để tìm kiếm các mẫu và xu hướng.
- Tạo báo cáo và biểu đồ: QuickSight giúp bạn dễ dàng tạo báo cáo và biểu đồ để chia sẻ dữ liệu của bạn với người khác.

Chương 2

Khảo sát hệ thống

2.1 Khảo sát thực tế

2.1.1 Hiện trạng

Khảo sát hiện trạng và xác lập dự án là giai đoạn đầu của quá trình phân tích và thiết kế hệ thống. Việc khảo sát thường được tiến hành qua hai giai đoạn:

- Khảo sát sơ bộ nhằm xác định tính khả thi của dự án
- Khảo sát chi tiết nhằm xác định chính xác những gì sẽ thực hiện và khẳng định những lợi ích kèm theo.

Trong giai đoạn khảo sát sơ bộ, dự án đã tiến hành khảo sát các vấn đề của hệ thống quản lý chia sẻ file hiện tại tại các cá nhân thường xuyên chia sẻ file bằng cách quan sát trực tiếp, phỏng vấn người dùng cuối và các bên liên quan khác. Kết quả khảo sát cho thấy:

- Giao diện phức tạp, thiếu tính thân thiện, nhiều người dùng gặp khó khăn khi sử dụng các tính năng nâng cao.
- Quy trình thao tác phức tạp, nhiều bước để thực hiện một nhiệm vụ đơn giản. Thiếu hướng dẫn trực quan, hỗ trợ người dùng cuối.
- Không có tính năng quản lý phiên bản file, khó theo dõi các thay đổi và lịch sử truy cập file. Không ghi lại nhật ký truy cập, khó truy vết khi có sự cố xảy ra.
- Việc quản lý quyền truy cập chưa chặt chẽ, nhiều người dùng có thể truy cập vào các file mà họ không có quyền hạn.

- Tốc độ truyền tải file chậm, thường xuyên bị đứng/nghe khi nhiều người dùng cùng truy cập.
- Thời gian tải lên file 2 chậm, đặc biệt với các tệp lớn. Số lượng kết nối đồng thời bị giới hạn, gây nghẽn khi có nhiều truy cập cùng lúc.
- Khó mở rộng hệ thống khi nhu cầu người dùng và lưu trữ tăng vì kiến trúc hạn chế. Gặp hiện tượng treo/đứng máy chủ khi tải nặng, không đảm bảo khả năng hoạt động liên tục.

→ Qua đây cho thấy người dùng họ cần nhu cầu một hệ thống gia tăng về dung lượng lưu trữ, số lượng người dùng trong tương lai. Đồng thời đánh giá các lợi ích kinh tế, hiệu quả làm việc nếu triển khai hệ thống mới để xác định tính khả thi của dự án.

2.1.2 Quy trình chia sẻ dữ liệu

Chia sẻ dữ liệu là một hoạt động quan trọng nhưng cũng tiềm ẩn nhiều rủi ro nếu không được thực hiện đúng cách. Để đảm bảo an toàn và hiệu quả trong quá trình chia sẻ, cần tuân thủ một quy trình rõ ràng bao gồm các bước sau:

1. Xác định rõ ràng mục đích chia sẻ dữ liệu: Trước khi chia sẻ, cần làm rõ lý do tại sao cần chia sẻ dữ liệu, đối tượng cần chia sẻ và sử dụng dữ liệu cho mục đích gì. Ví dụ như chia sẻ dữ liệu cho mục đích nghiên cứu khoa học, hợp tác dự án, chia sẻ thông tin nội bộ, hoặc chia sẻ thông tin cá nhân cho gia đình, bạn bè.
2. Phân loại dữ liệu và đánh giá mức độ nhạy cảm: Cần xác định rõ loại dữ liệu cần chia sẻ, bao gồm dữ liệu cá nhân, dữ liệu nhạy cảm, dữ liệu kinh doanh bí mật hay chỉ đơn thuần là dữ liệu công khai. Từ đó, đánh giá mức độ bảo mật cần thiết cho từng loại dữ liệu. Các dữ liệu nhạy cảm liên quan đến thông tin cá nhân, tài chính, y tế sẽ cần có mức bảo mật cao nhất.
3. Lựa chọn phương thức chia sẻ dữ liệu phù hợp:
 - (a) Đối với cá nhân/gia đình: Có thể chia sẻ trực tiếp qua email, tin nhắn, mạng xã hội hoặc sử dụng các dịch vụ lưu trữ đám mây cá nhân phổ biến

như Google Drive, Dropbox, OneDrive. Tuy nhiên, cần đặc biệt lưu ý đến vấn đề quyền riêng tư và bảo mật khi chia sẻ trên mạng xã hội công khai.

(b) Đối với doanh nghiệp/tổ chức: Nên sử dụng các hệ thống quản lý tài liệu, cổng thông tin nội bộ, giải pháp lưu trữ đám mây doanh nghiệp với các tính năng bảo mật cao. Cần xây dựng và áp dụng các chính sách, quy trình chia sẻ dữ liệu phù hợp cho từng đối tượng, loại dữ liệu. Đối với các đối tác bên ngoài, cần ký kết các thỏa thuận bảo mật thông tin.

4. Áp dụng các biện pháp bảo mật dữ liệu: Trước khi chia sẻ, dữ liệu cần được mã hóa, đặc biệt là dữ liệu nhạy cảm. Sử dụng các phương thức xác thực và ủy quyền nghiêm ngặt để kiểm soát quyền truy cập dữ liệu chia sẻ. Theo dõi, ghi chép và giám sát tất cả các hoạt động truy cập, tải xuống dữ liệu để dễ dàng truy vết khi có sự cố xảy ra.
5. Đảm bảo tuân thủ các quy định pháp lý: Khi chia sẻ dữ liệu, cần tìm hiểu và tuân thủ các quy định pháp luật liên quan đến bảo vệ dữ liệu, quyền riêng tư thông tin cá nhân của từng quốc gia hay khu vực pháp lý. Một số quy định nghiêm ngặt như GDPR của EU, CCPA của California, PDPA của Việt Nam sẽ quy định rõ cách thức xử lý dữ liệu cá nhân.
6. Đào tạo nhận thức cho người dùng: Cuối cùng, cần tăng cường đào tạo, nâng cao nhận thức cho người dùng về tầm quan trọng của việc chia sẻ dữ liệu an toàn. Người dùng cần hiểu rõ các rủi ro có thể xảy ra, cách nhận biết và xử lý các đe dọa tấn công giả mạo, đánh cắp dữ liệu.

Việc tuân thủ nghiêm ngặt quy trình chia sẻ dữ liệu sẽ giúp tổ chức hạn chế rủi ro mất mát, rò rỉ dữ liệu quan trọng, đồng thời đáp ứng đầy đủ các yêu cầu về bảo mật, pháp lý cũng như các tiêu chuẩn, chính sách an toàn thông tin của tổ chức.

2.2 Đánh giá hệ thống hiện tại

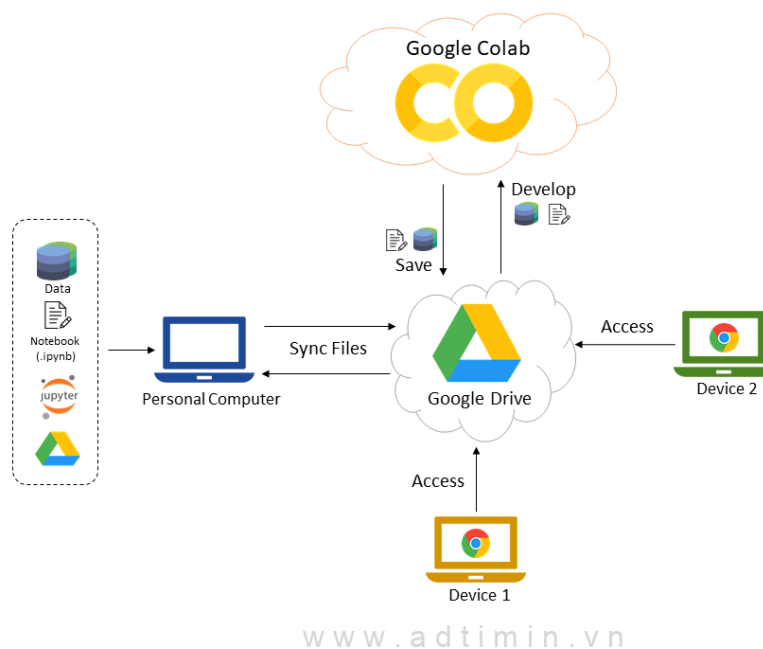
2.2.1 Tiêu chí đánh giá

- **Bảo mật:** Mức độ an toàn của dữ liệu trong quá trình truyền tải và lưu trữ.

- **Hiệu suất:** Tốc độ truyền tải và khả năng xử lý file lớn.
- **Tính dễ sử dụng:** Giao diện người dùng và trải nghiệm sử dụng.
- **Quản lý và theo dõi:** Khả năng quản lý và theo dõi các file đã chia sẻ.
- **Khả năng mở rộng:** Khả năng mở rộng hệ thống khi nhu cầu tăng lên.

2.2.2 Các hệ thống hiện có

Từ xưa đến nay, các hệ thống quản lý chia sẻ file đã trở nên quen thuộc và được sử dụng rộng rãi bởi người dùng trên toàn thế giới. Có rất nhiều giải pháp nổi bật của các ông lớn công nghệ, mỗi giải pháp đều có những ưu nhược điểm riêng:



Hình 2.1: Kiến trúc google drive

- **Google Drive:** Là dịch vụ lưu trữ đám mây phổ biến nhất hiện nay, cung cấp miễn phí 15GB dung lượng lưu trữ ban đầu. Ưu điểm là tích hợp tốt với các dịch vụ Google khác, giao diện thân thiện, hỗ trợ cộng tác tốt. Tuy nhiên, khả năng mở rộng bị giới hạn bởi dung lượng, tính bảo mật cũng chưa cao.
- **Dropbox:** Nổi tiếng về tính bảo mật cao, cung cấp nhiều tính năng cộng tác, đồng bộ hóa dữ liệu. Tuy nhiên, gói miễn phí chỉ 2GB dung lượng, các gói dung lượng cao phí khá đắt. Giao diện đơn giản nhưng thiếu tính năng nâng cao.

- OneDrive: Được tích hợp sẵn trên nền tảng Windows và Office 365, cung cấp dung lượng lưu trữ lớn lên tới 1TB. Ưu điểm là tương thích cao, dễ dàng chia sẻ và cộng tác. Nhược điểm là giao diện đơn điệu, khả năng mở rộng phụ thuộc vào gói đăng ký.
- Fshare: Dịch vụ lưu trữ đám mây phổ biến tại Việt Nam, cung cấp nhiều gói cước lưu trữ với giá khá rẻ, dung lượng có thể lên tới 5TB. Tuy nhiên, giao diện chưa thân thiện, tính năng cộng tác và đồng bộ hóa còn hạn chế.

Bên cạnh đó, còn có nhiều giải pháp khác như Box, Mega, pCloud, Amazon Drive,... Mỗi giải pháp đều có những ưu nhược điểm riêng về tính năng, dung lượng, bảo mật, giá cả và khả năng tích hợp. Để đáp ứng nhu cầu ngày càng tăng về dữ liệu số và yêu cầu phức tạp về quản lý, chia sẻ file, các doanh nghiệp thường phải tự xây dựng hệ thống nội bộ riêng cho phù hợp.

Tuy nhiên, việc xây dựng một hệ thống quản lý chia sẻ file đáp ứng đầy đủ các yêu cầu về bảo mật, tính năng, khả năng mở rộng và chi phí phù hợp là một thách thức lớn. Cần có sự đánh giá, phân tích kỹ lưỡng để lựa chọn giải pháp phù hợp nhất với nhu cầu, nguồn lực của tổ chức.

2.2.3 Tại sao không dùng hệ thống sẵn có?

Các hệ thống này đã tồn tại rất lâu và phổ biến cho người dùng hiện nay. Tuy nhiên các ứng dụng còn rất nhiều hạn chế để một doanh nghiệp có thể dùng để quản lý các tài liệu nội bộ của công ty một cách an toàn và hiệu quả.

Các hệ thống hiện tại sử dụng giao thức HTTP, FTP để upload và tải file, tuy nhiên tốc độ truy tải vẫn khá chậm, đặc biệt khi xử lý các file lớn. Hệ thống cũng gặp khó khăn trong việc quản lý và theo dõi các file đã chia sẻ, và giao diện người dùng không thân thiện, gây khó khăn trong việc sử dụng

- Nhu cầu cụ thể: hệ thống sẵn có có thể không đáp ứng đầy đủ các yêu cầu quản lý của doanh nghiệp.
- Tùy chỉnh: hệ thống tự xây dựng có thể được tùy chỉnh theo nhu cầu và đặc thù riêng. Doanh nghiệp có thể tùy chỉnh theo phong cách riêng về các tính năng, giao diện, quy trình làm việc.

- Bảo mật: hệ thống tự xây dựng có thể kiểm soát bảo mật tốt hơn, dễ dàng truy vết nếu có sai phạm.
- Chi phí: hệ thống tự xây dựng có thể tiết kiệm chi phí về lâu dài.
- Quản lý: không có trực quan qua hình ảnh, đồ thị để dễ dàng theo dõi.

Kết luận

Hệ thống chia sẻ file tự xây dựng có thể đáp ứng tốt nhu cầu cụ thể và mang lại nhiều lợi ích cho doanh nghiệp. Tuy nhiên, cần cân nhắc kỹ lưỡng trước khi quyết định xây dựng hệ thống mới và lựa chọn phương pháp phù hợp để đảm bảo hiệu quả và tính bảo mật.

2.3 Đề xuất cho hệ thống mới

Khi chia sẻ dữ liệu quan trọng hoặc nhạy cảm, các tổ chức thường có nhu cầu kiểm soát chặt chẽ hơn về cách truy cập dữ liệu. Ví dụ, họ có thể muốn giới hạn khả năng tải xuống dữ liệu dựa trên vai trò và đặc quyền của người dùng, hạn chế số lần tải xuống một đối tượng để kiểm soát chi phí truy cập dữ liệu. Ngoài ra, việc theo dõi chi tiết từng cá nhân đã truy cập và tải xuống các tệp dữ liệu cũng rất quan trọng để dễ dàng truy vết khi có sự cố xảy ra liên quan đến dữ liệu nhạy cảm.

Hệ thống mới sẽ sử dụng giao thức HTTPS và cơ sở dữ liệu phân tán tận dụng mọi khả năng quản lý của Cloud để đảm bảo bảo mật và hiệu suất. Giao thức HTTPS sẽ mã hóa dữ liệu trong quá trình truyền tải, giảm nguy cơ bị tấn công và đánh cắp thông tin. Hệ thống mới cũng sẽ cải thiện tốc độ truyền tải và giao diện người dùng, đồng thời tích hợp các công cụ quản lý và theo dõi file hiệu quả hơn

2.3.1 Mục tiêu hệ thống

Yêu cầu hệ thống mới có khả năng quản lý chia sẻ file chặt chẽ nhất có thể để giải quyết nhu cầu lưu trữ, quản lý và chia sẻ file an toàn hiệu quả cho cá nhân và tổ chức. Dễ dàng truy vết nếu có sai phạm.

Hệ thống quản lý chia sẻ file mới cần phải đáp ứng các mục tiêu sau:

- Cung cấp giải pháp lưu trữ và quản lý dữ liệu an toàn, hiệu quả với khả năng kiểm soát truy cập, quản lý phiên bản file và dữ liệu toàn diện.
- Tăng cường khả năng chia sẻ dữ liệu một cách linh hoạt, dễ dàng với người dùng trong và ngoài tổ chức.
- Đảm bảo tuân thủ các quy định về bảo mật, quyền riêng tư thông tin.
- Tiết kiệm chi phí lưu trữ, quản lý và truyền tải dữ liệu. Có khả năng tích hợp và mở rộng để đáp ứng nhu cầu trong tương lai.

2.3.2 Các bước triển khai hệ thống mới

Chuẩn bị cơ sở hạ tầng: Đảm bảo các máy chủ và thiết bị mạng đều đáp ứng yêu cầu kỹ thuật cho hệ thống mới. Cài đặt phần mềm cần thiết và cấu hình các thành phần bảo mật.

Cài đặt và cấu hình hệ thống: Thiết lập Secure HTTPS Server và cấu hình Decentralized Database. Đảm bảo tất cả các kết nối đều được mã hóa và dữ liệu được phân tán một cách hợp lý.

Đào tạo người dùng: Tổ chức các buổi đào tạo và cung cấp tài liệu hướng dẫn để người dùng làm quen với giao diện và các chức năng của hệ thống mới.

Đánh giá và cải tiến: Thực hiện các bài kiểm tra bảo mật và hiệu suất. Thu thập phản hồi từ người dùng để liên tục cải thiện và tối ưu hóa hệ thống.

2.3.3 Phạm vi hệ thống

Hệ thống được thiết kế cho người dùng trong một tổ chức để trao đổi, chia sẻ các tệp tin dữ liệu công việc với nhau. Hỗ trợ lưu trữ, quản lý các loại tệp tin văn bản, tài liệu, hình ảnh, video và các dữ liệu quan trọng khác.

2.3.4 Đối tượng sử dụng

1. Người dùng cá nhân: Nhân viên trong tổ chức có nhu cầu lưu trữ, truy cập và chia sẻ dữ liệu phục vụ công việc.
2. Nhóm dự án/phòng ban: Nhóm người dùng cần chia sẻ, cộng tác xử lý nội dung dữ liệu.
3. Quản trị viên: Đảm nhận việc quản trị hệ thống, phân quyền truy cập, thiết lập chính sách bảo mật.

2.3.5 Lợi ích của hệ thống mới

Hệ thống mới có thể dễ dàng tích hợp với hệ thống nội bộ lớn của doanh nghiệp, giúp thông tin được luân chuyển một cách nhanh chóng và an toàn.

Đồng thời hệ thống mới cũng đem lại các tính năng nổi trội:

- Dễ dàng quản lý, theo dõi và chia sẻ dữ liệu giữa các cá nhân, phòng ban trong và ngoài tổ chức.
- Hệ thống đáp ứng đầy đủ các yêu cầu về bảo mật, quản lý quyền truy cập phức tạp đáp ứng nhu cầu thực tế.
- Hiệu suất cao và khả năng mở rộng tốt. Tối ưu hóa để đáp ứng lượng truy cập lớn, cũng như khả năng chịu tải khi có sự gia tăng về số lượng file và người dùng .
- Áp dụng các công nghệ hiện đại như AI, đám mây để tối ưu hóa hiệu suất, chi phí vận hành và đảm bảo an toàn dữ liệu.

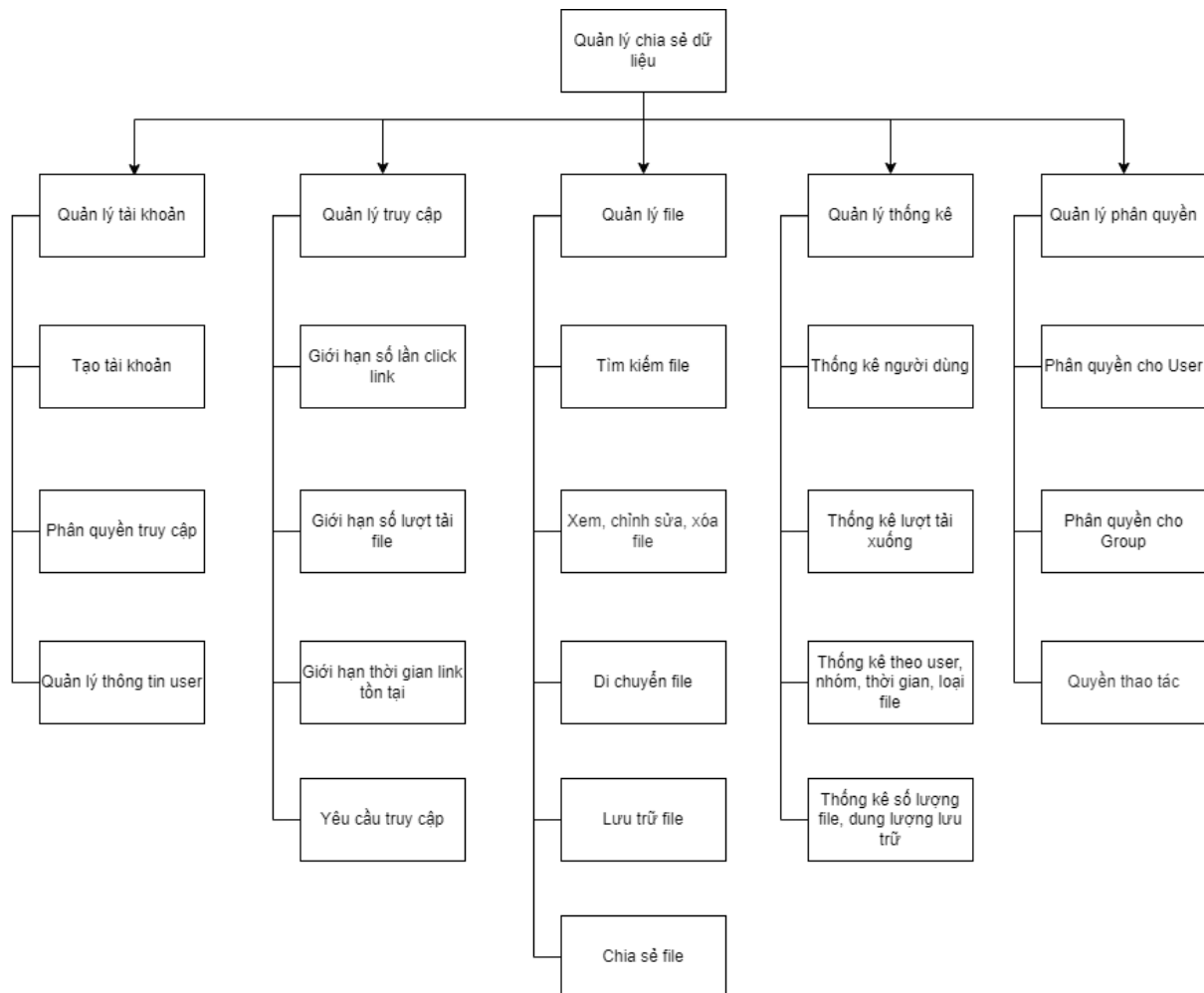
Với nhiều tính năng nổi trội, giải pháp mới sẽ giúp tối đa hóa hiệu quả làm việc, chia sẻ thông tin, đồng thời đảm bảo an toàn dữ liệu cao cho tổ chức. Đáp ứng được mọi nhu cầu về quản lý, lưu trữ và chia sẻ dữ liệu hiện đại

Chương 3

Phân tích hệ thống

3.1 Sơ đồ phân cấp chức năng

Thiết kế hệ thống với các chức năng sau:



Hình 3.1: Sơ đồ phân cấp chức năng

1. Quản lý tài khoản

- Tạo tài khoản: cho phép người dùng mới đăng ký tài khoản trên hệ thống.
- Phân quyền truy cập: quản lý quyền truy cập của người dùng đối với các tài nguyên trong hệ thống.
- Quản lý thông tin user: quản lý thông tin cá nhân và thông tin đăng nhập của người dùng.

2. Quản lý truy cập

- Giới hạn số lần click link: giới hạn số lần một liên kết có thể được nhấp.
- Giới hạn số lượt tải file: giới hạn số lần một tệp có thể được tải xuống.
- Giới hạn thời gian link tồn tại: đặt giới hạn thời gian mà một liên kết tồn tại và có thể được sử dụng.
- Yêu cầu truy cập: cho phép người dùng gửi yêu cầu truy cập vào các tệp cụ thể.

3. Quản lý file

- Tìm kiếm file: cho phép người dùng tìm kiếm các tệp trong hệ thống.
- Xem, chỉnh sửa, xóa file: cho phép người dùng xem, chỉnh sửa và xóa các tệp.
- Di chuyển file: cho phép người dùng di chuyển các tệp từ thư mục này sang thư mục khác.
- Lưu trữ file: cho phép người dùng lưu trữ các tệp trên hệ thống.
- Chia sẻ file: cho phép người dùng chia sẻ tệp với những người dùng khác.

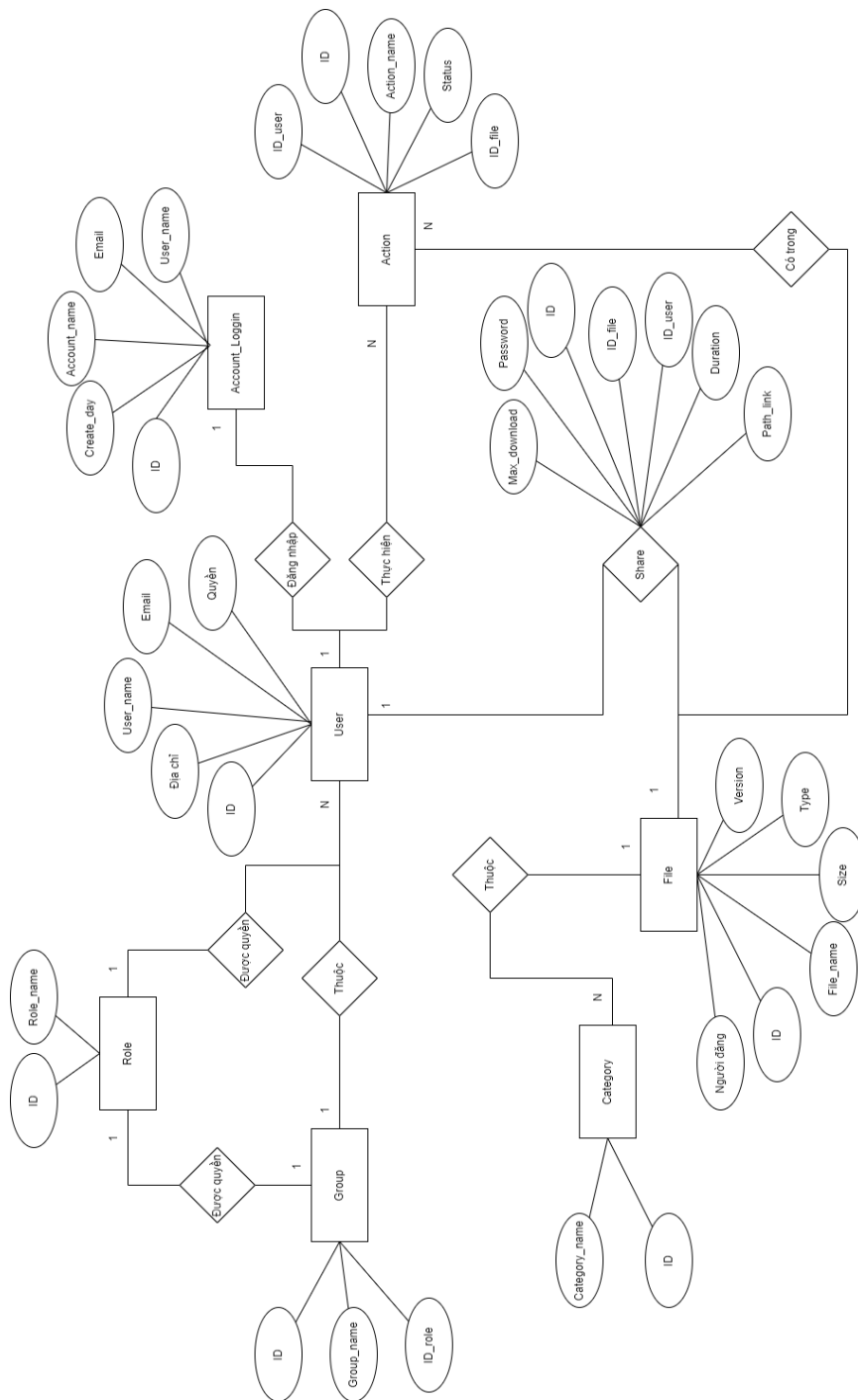
4. Quản lý thống kê

- Thống kê người dùng: cung cấp các báo cáo thống kê về người dùng của hệ thống. Thống kê lượt tải xuống: cung cấp các báo cáo thống kê về số lượt tải xuống các tệp. Thống kê theo user, nhóm, thời gian, loại file: cung cấp các báo cáo thống kê chi tiết theo người dùng, nhóm, thời gian và loại tệp. Thống kê số lượng file, dung lượng lưu trữ: cung cấp các báo cáo thống kê về số lượng tệp và dung lượng lưu trữ đã sử dụng.

5. Quản lý phân quyền

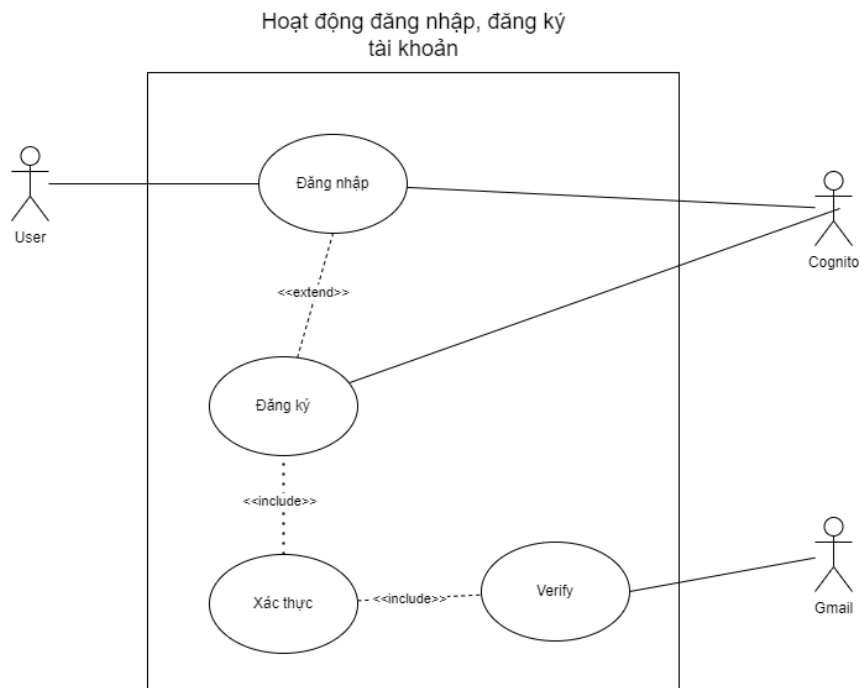
- Phân quyền cho User: cho phép phân quyền truy cập và thao tác cho từng người dùng.
- Phân quyền cho Group: cho phép phân quyền truy cập và thao tác cho từng nhóm người dùng.
- Quyền thao tác: xác định quyền cụ thể (xem, chỉnh sửa, xóa, chia sẻ, v.v.) mà người dùng hoặc nhóm có thể thực hiện trên các tệp.

3.2 ERD

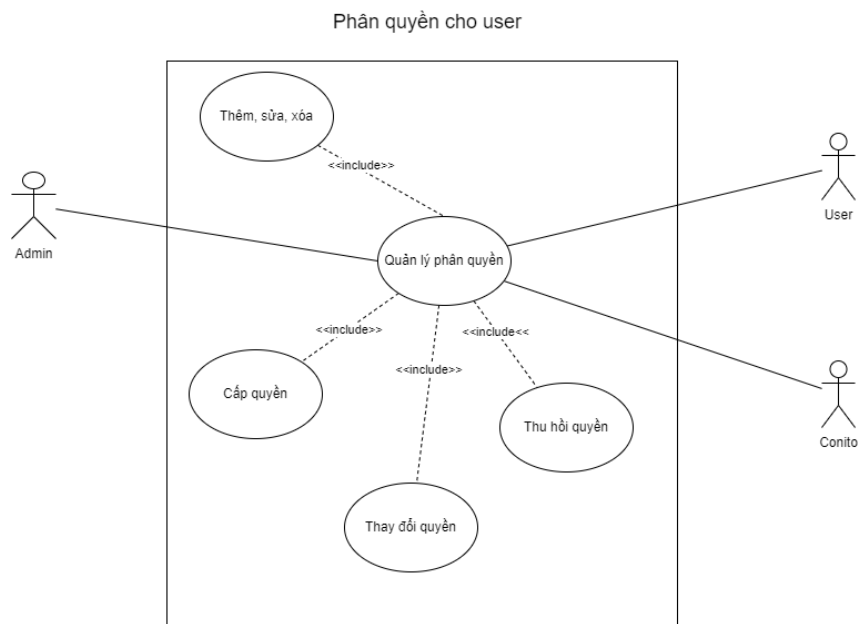


Hình 3.2: Sơ đồ thực thể liên kết

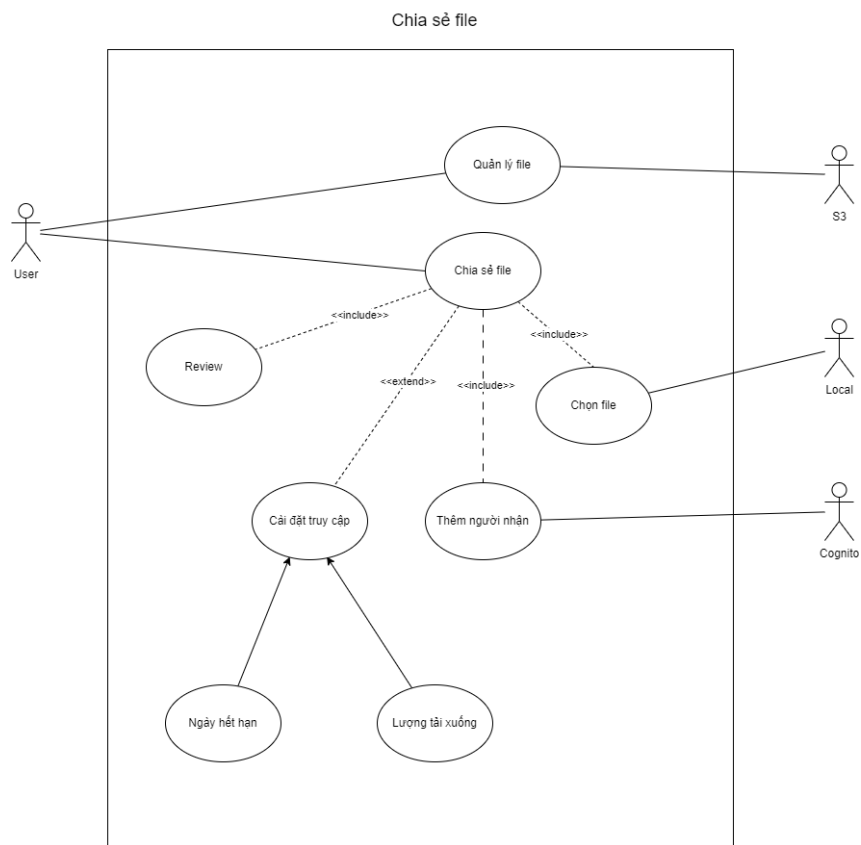
3.3 Sơ đồ usecase



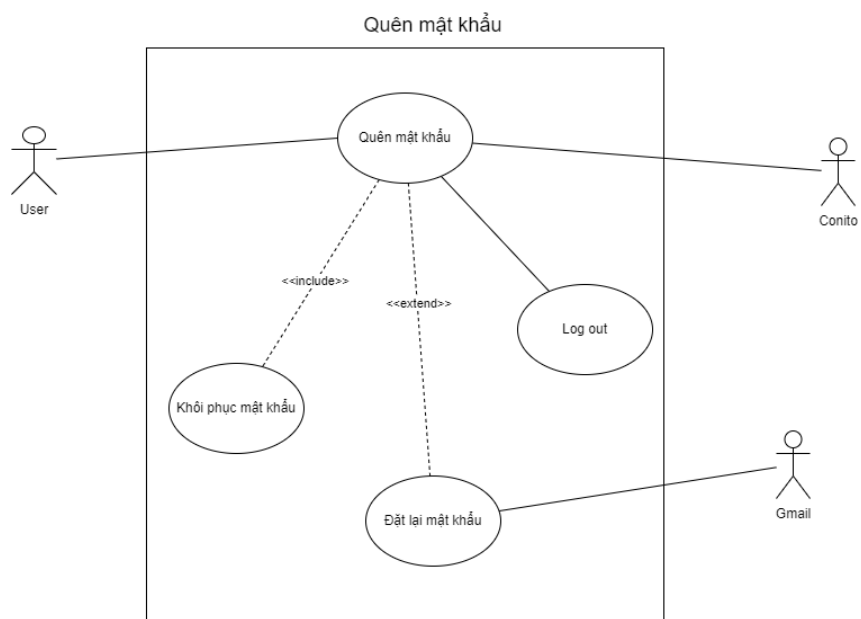
Hình 3.3: Usecase đăng nhập đăng ký



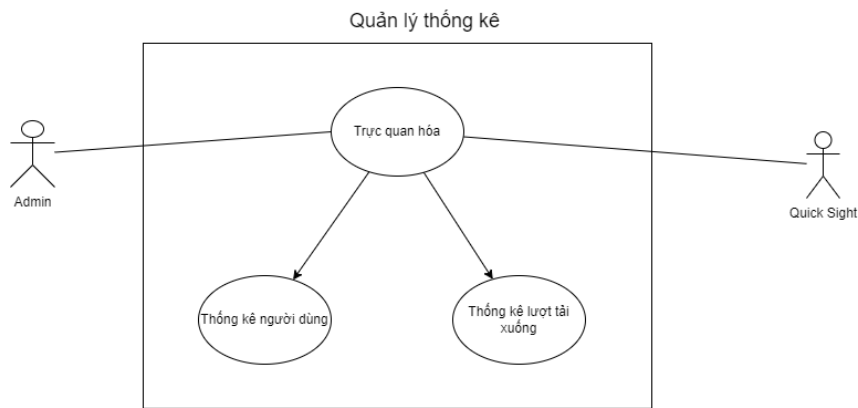
Hình 3.4: Usecase phân quyền người dùng



Hình 3.5: Usecase chia sẻ file



Hình 3.6: Usecase chia sẻ file



Hình 3.7: Usecase thống kê

3.4 Đặc tả usecase

3.4.1 Đặc tả usecase đăng nhập, đăng ký

- Mô tả: Cho phép người dùng truy cập vào hệ thống bằng cách nhập thông tin xác thực, nếu chưa có tài khoản thì người dùng thực hiện đăng ký.
- Người sử dụng: User
- Kích hoạt: Người dùng muốn sử dụng các chức năng của hệ thống.
- Tiền điều kiện: Người dùng đã có tài khoản trong hệ thống.
- Hậu điều kiện: Người dùng được phép truy cập vào hệ thống và sử dụng các chức năng tương ứng với quyền của mình.
- Các bước thực hiện:
 1. Người dùng truy cập giao diện đăng nhập của hệ thống.
 2. Hệ thống yêu cầu người dùng nhập thông tin xác thực (tên đăng nhập, mật khẩu).
 3. Người dùng nhập thông tin xác thực.
 4. Hệ thống kiểm tra tính hợp lệ của thông tin xác thực.
 5. Nếu thông tin xác thực không hợp lệ, hệ thống hiển thị thông báo lỗi và quay lại bước 2.
 6. Nếu thông tin xác thực hợp lệ, hệ thống chuyển sang giao diện chính và cho phép người dùng sử dụng các chức năng tương ứng với quyền của mình.

7. Người dùng chưa có tài khoản thì tiến hành tạo tài khoản.
8. Sau khi điền đầy đủ thông tin cần thiết hệ thống sẽ gửi mã xác nhận.
9. Người dùng nhập mã xác nhận và tạo thành công tài khoản.
10. hệ thống đăng xuất và người dùng phải nhập lại thông tin xác thực.

3.4.2 Đặc tả usecase phân quyền

- Mô tả: Cho phép quản trị viên (Admin) quản lý phân quyền cho người dùng (User) trong hệ thống.
- Người sử dụng: Admin
- Kích hoạt: Quản trị viên muốn thêm, sửa hoặc xóa quyền của người dùng trong hệ thống.
- Tiền điều kiện: Quản trị viên đã đăng nhập vào hệ thống với quyền quản trị.
- Hậu điều kiện: Quyền của người dùng trong hệ thống được cập nhật.
- Các bước thực hiện:
 1. Quản trị viên chọn chức năng quản lý phân quyền.
 2. Hệ thống hiển thị danh sách người dùng và quyền hiện tại của họ.
 3. Quản trị viên chọn người dùng cần thay đổi quyền.
 4. Quản trị viên thực hiện các thao tác
 5. Hệ thống xác thực thay đổi quyền với hệ thống quản lý người dùng bên ngoài (Cognito).
 6. Hệ thống cập nhật quyền của người dùng trong cơ sở dữ liệu.
 7. Hệ thống thông báo thay đổi quyền thành công.

3.4.3 Đặc tả usecase chia sẻ file

- Mô tả: Cho phép người dùng chia sẻ file với người khác trong hệ thống.
- Người sử dụng: User
- Kích hoạt: Người dùng muốn chia sẻ một file với người khác.

- Tiền điều kiện: Người dùng đã đăng nhập vào hệ thống và có quyền truy cập vào file cần chia sẻ.
- Hậu điều kiện: File được chia sẻ thành công với người được chỉ định.
- Các bước thực hiện:
 1. Người dùng chọn file muốn chia sẻ.
 2. Hệ thống yêu cầu người dùng nhập thông tin người nhận.
 3. Người dùng nhập thông tin người nhận (email người dùng trong hệ thống).
 4. Hệ thống kiểm tra quyền truy cập của người nhận
 5. Nếu người nhận không có quyền truy cập, hệ thống thông báo lỗi và quay lại bước 3.
 6. Nếu người nhận có quyền truy cập, hệ thống thực hiện chia sẻ file
 7. Hệ thống thông báo chia sẻ file thành công.

3.4.4 Đặc tả usecase quên mật khẩu

- Mô tả: Cho phép người dùng (User) khôi phục mật khẩu trong trường hợp quên mật khẩu.
- Người sử dụng: User
- Kích hoạt: Người dùng quên mật khẩu và muốn đặt lại mật khẩu mới.
- Tiền điều kiện: Người dùng đã đăng ký tài khoản trong hệ thống.
- Hậu điều kiện: Người dùng có mật khẩu mới để đăng nhập vào hệ thống.
- Các bước thực hiện:
 1. Người dùng chọn chức năng quên mật khẩu.
 2. Hệ thống yêu cầu người dùng nhập thông tin xác thực.
 3. Người dùng nhập thông tin xác thực (ví dụ: email đăng ký).
 4. Hệ thống kiểm tra tính hợp lệ của thông tin xác thực
 5. Nếu thông tin xác thực không hợp lệ, hệ thống thông báo lỗi và quay lại bước 2.

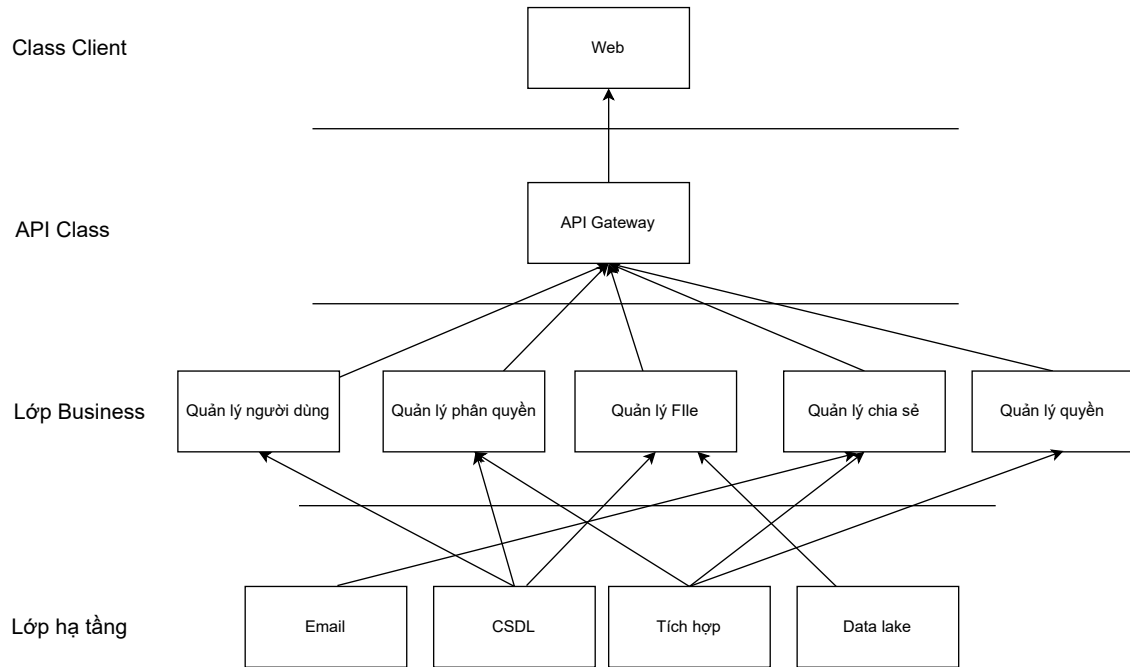
6. Nếu thông tin xác thực hợp lệ, hệ thống gửi mã xác nhận đến email của người dùng
7. Người dùng nhập mã xác nhận.
8. Nếu mã xác nhận đúng, hệ thống cho phép người dùng đặt lại mật khẩu mới.
9. Hệ thống thông báo đặt lại mật khẩu thành công.

3.4.5 Đặc tả usecase thống kê

- Mô tả: Cho phép quản trị viên (Admin) quản lý và theo dõi các thông tin thống kê về hoạt động của hệ thống.
- Người sử dụng: Admin
- Kích hoạt: Quản trị viên muốn xem và phân tích các thông tin thống kê.
- Tiền điều kiện: Quản trị viên đã đăng nhập vào hệ thống với quyền quản trị.
- Hậu điều kiện: Quản trị viên có thể xem và phân tích các thông tin thống kê.
- Các bước thực hiện:
 1. Quản trị viên chọn chức năng quản lý thông kê.
 2. Hệ thống hiển thị giao diện thống kê.
 3. Quản trị viên có thể xem các thông tin thống kê
 4. Hệ thống lấy dữ liệu thống kê từ các nguồn liên quan
 5. Hệ thống hiển thị thông tin thống kê cho quản trị viên.

Chú ý: Các trường hợp và các usecase liên quan người dùng, tài khoản, BI dù quan trọng nhưng đã được xử lý bởi các service của AWS như QuickSight, S3, Cognito nên sẽ không được đề xuất trong báo cáo này.

3.5 Sơ đồ kiến trúc phân lớp

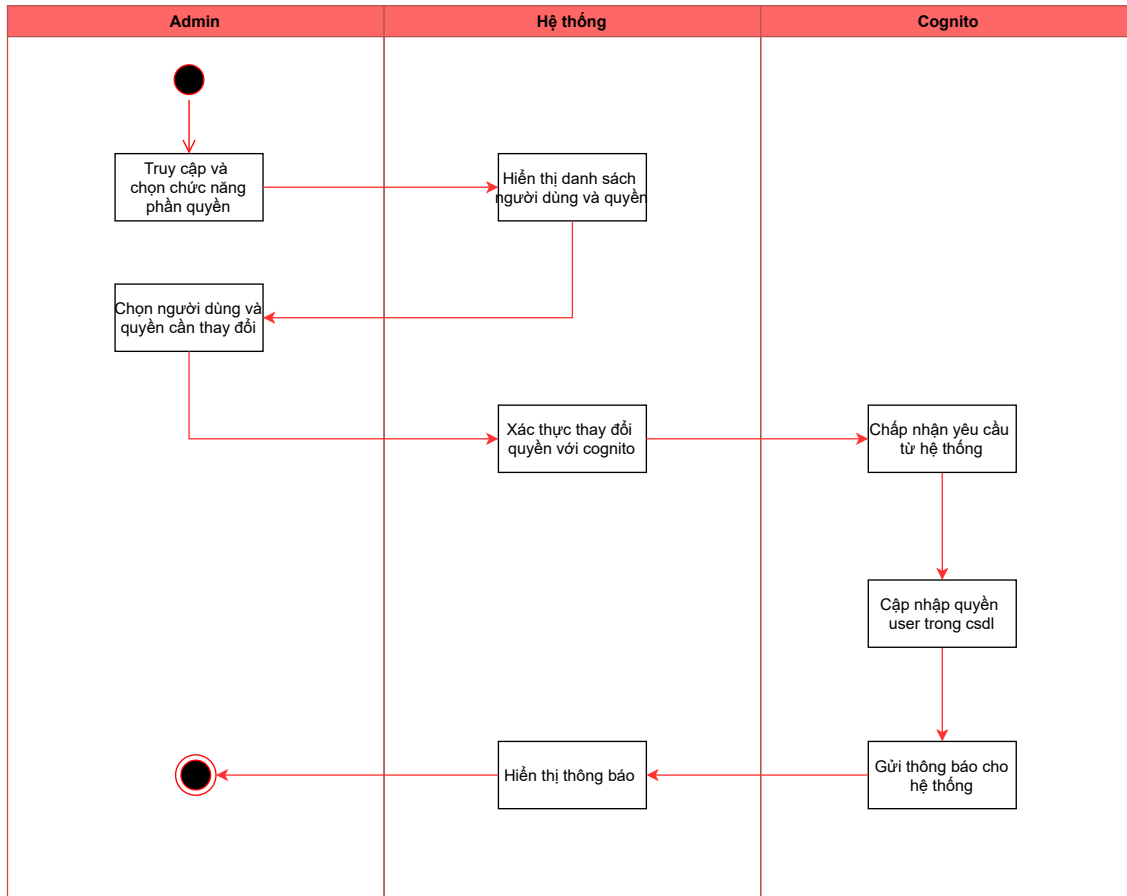


Hình 3.8: Phân lớp các tầng ứng dụng

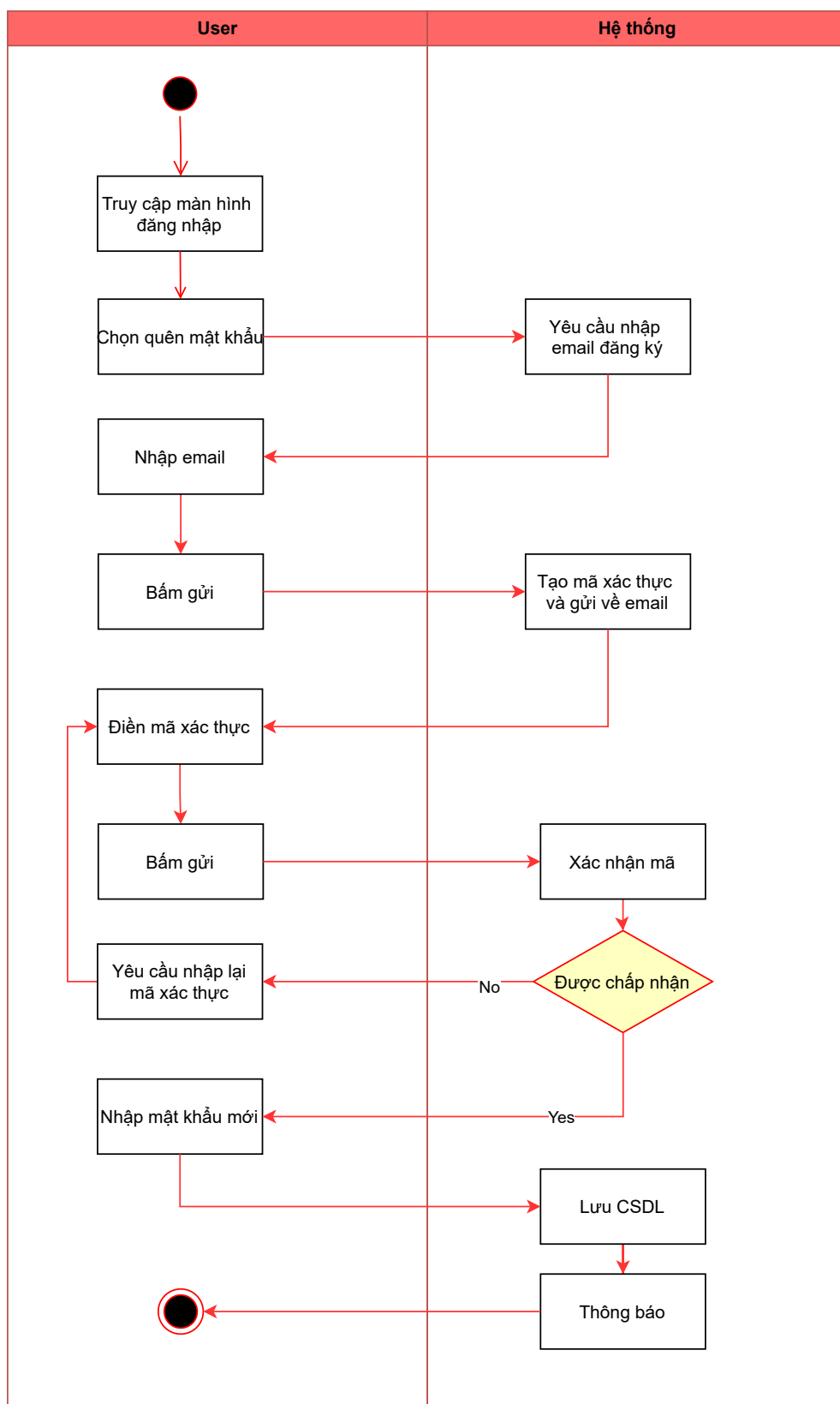
Sơ đồ mô tả các mức hạ tầng hoạt động của hệ thống.

3.6 Sơ đồ hoạt động

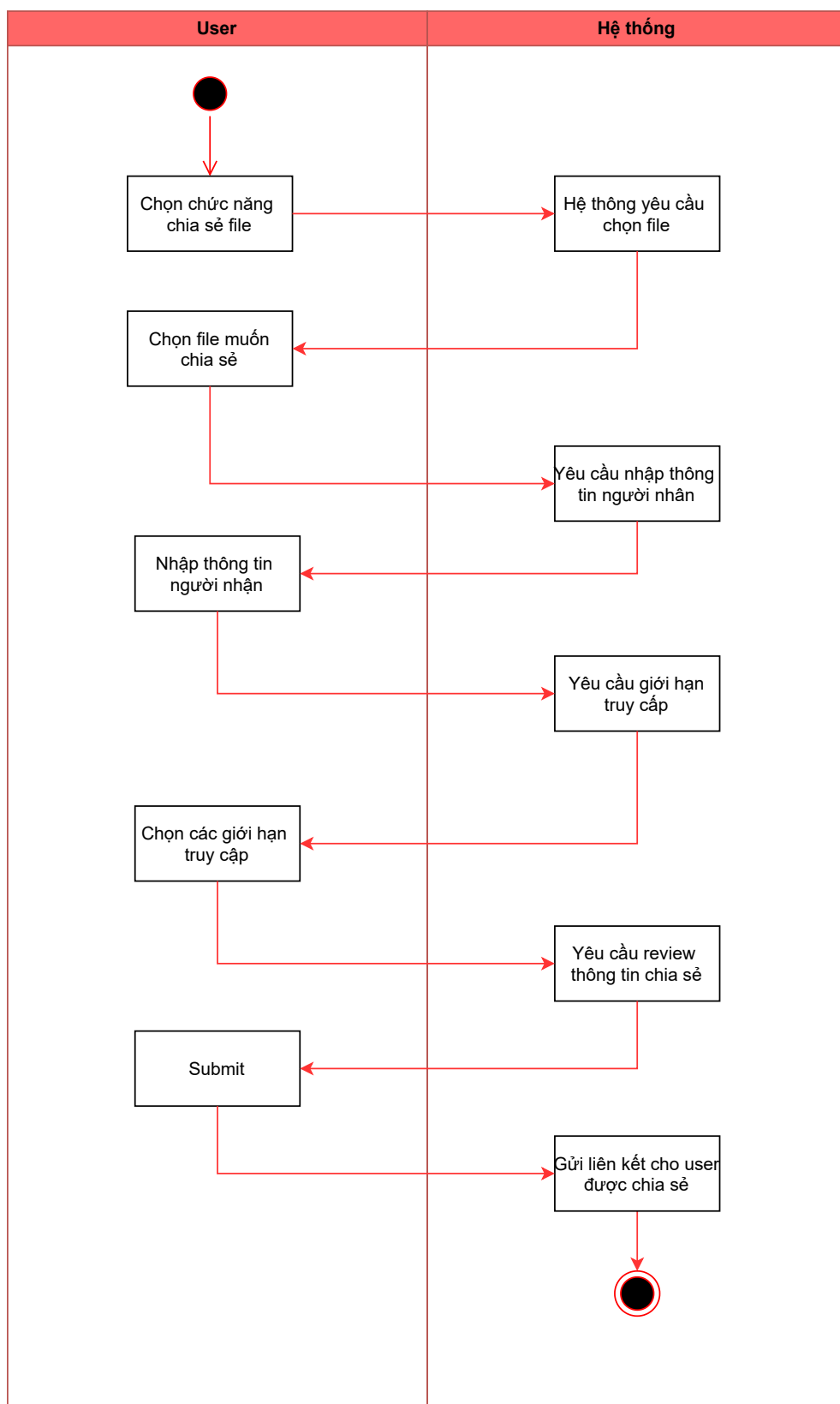
Sơ đồ mô tả hoạt động của một số chức năng chính



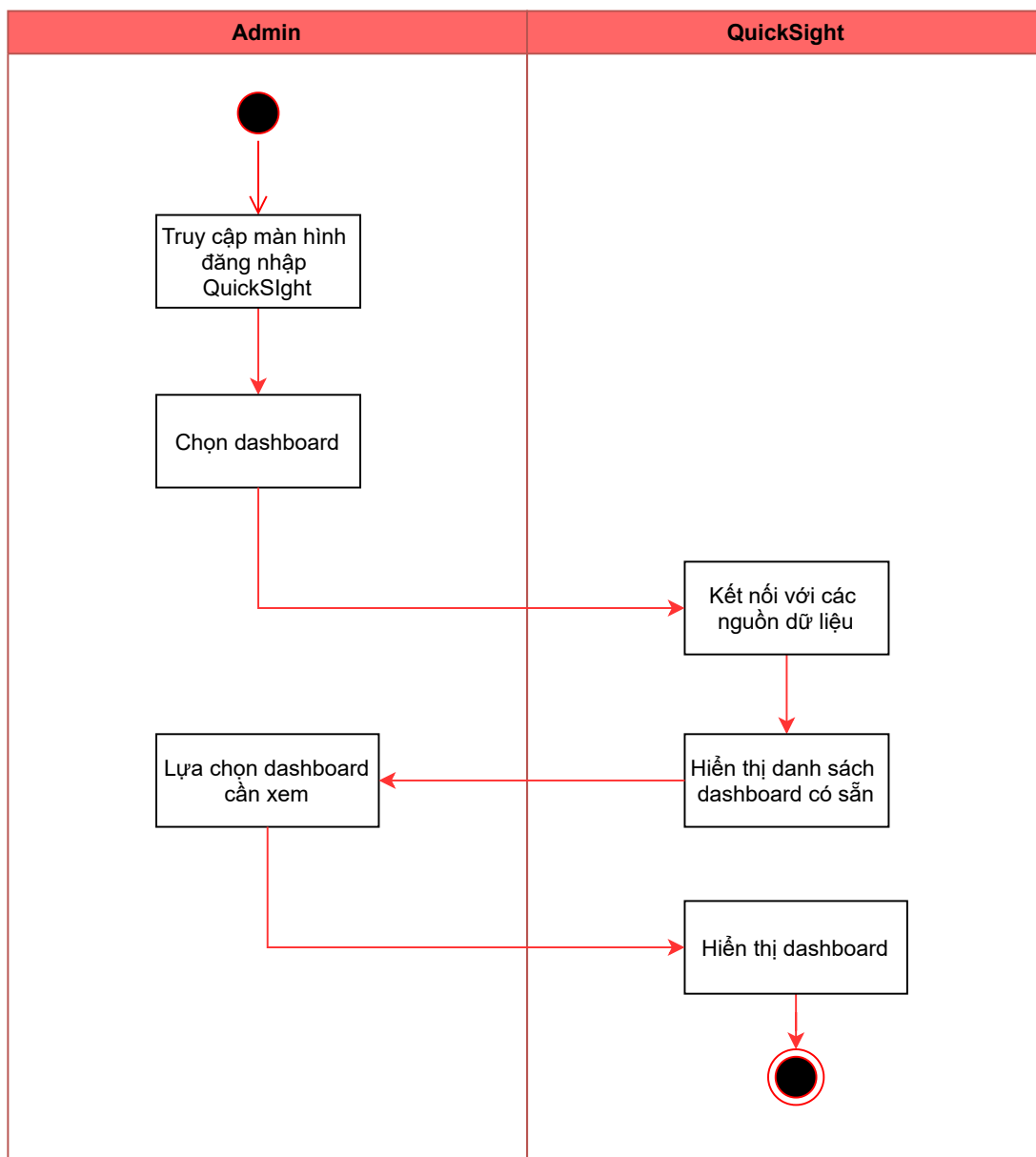
Hình 3.9: Chức năng phân quyền



Hình 3.10: Chức năng quên mật khẩu



Hình 3.11: Chức năng chia sẻ file

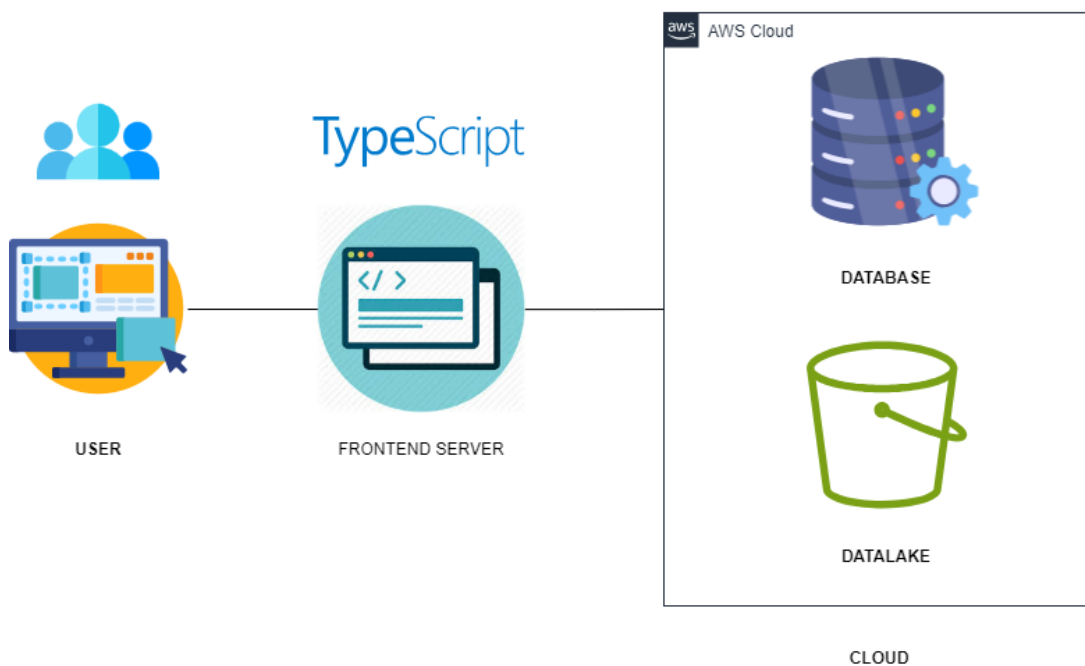


Hình 3.12: Chức năng xem dashboard

Chương 4

Thiết kế hệ thống

4.1 Kiến trúc hệ thống



Hình 4.1: Tổng quan kiến trúc hệ thống

Hệ thống bao gồm 2 thành phần chính:

- **Front-end server:** Chịu trách nhiệm cho việc xử lý và hiển thị giao diện người dùng. Trong trường hợp đăng nhập và xác minh thông tin người dùng, giao diện được thiết kế thân thiện tối ưu hóa trải nghiệm người dùng.

TypeScript: Sử dụng thư viện TypeScript, Front-end Server mang lại tính linh hoạt và khả năng tái sử dụng trong việc xây dựng các thành phần của

giao diện. Các tính năng đăng ký tài khoản, chia sẻ file được triển khai một cách hiệu quả

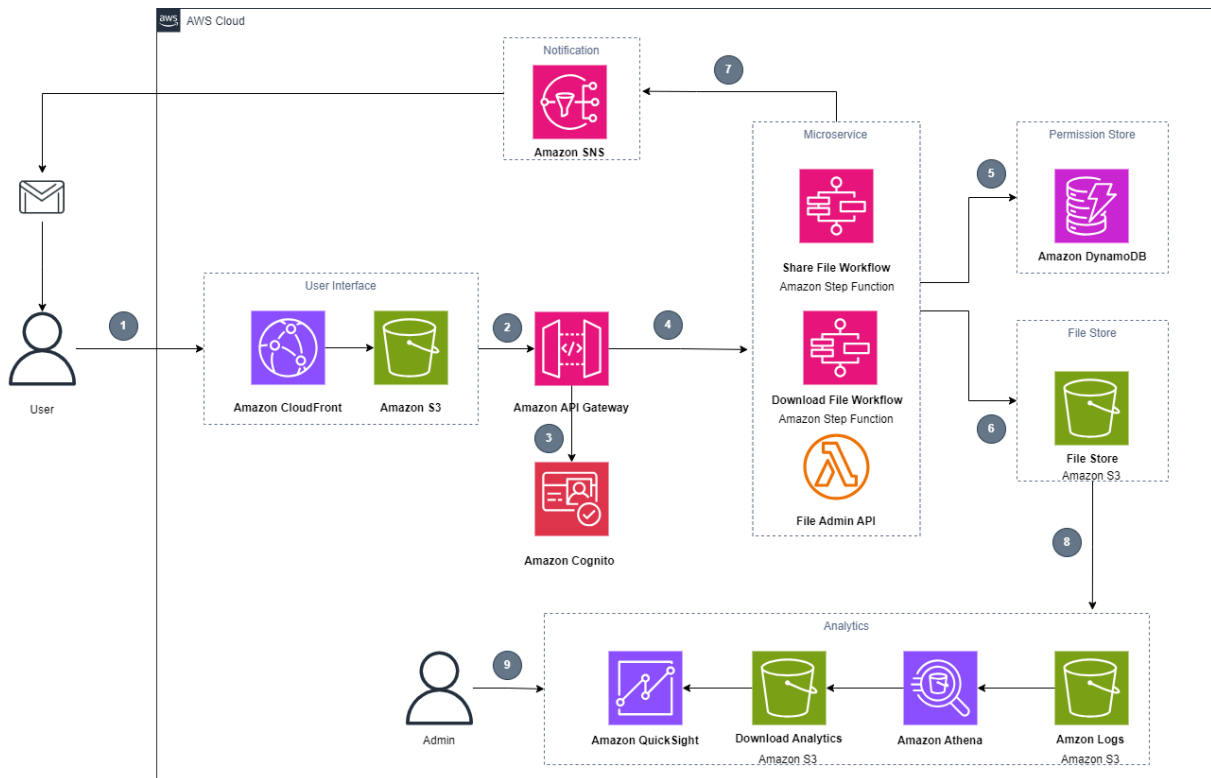
- **Cloud:** Nền tảng chịu tải gánh nặng lớn nhất của hệ thống với sự kết hợp của nhiều dịch vụ cấu phần nhằm đảm bảo tính linh hoạt, tối ưu của hệ thống.

Database: Sử dụng DynamoDB là nơi lưu trữ cơ sở dữ liệu chính của hệ thống. Cơ sở dữ liệu được thiết kế để hỗ trợ các tìm kiếm nhanh chóng và truy xuất dữ liệu một cách hiệu quả

Datalake: Sử dụng AWS S3 là nơi lưu trữ toàn bộ file được tải lên từ người dùng. Đồng thời tận dụng các tính năng sẵn có của S3 để đảm bảo tính an toàn thông tin

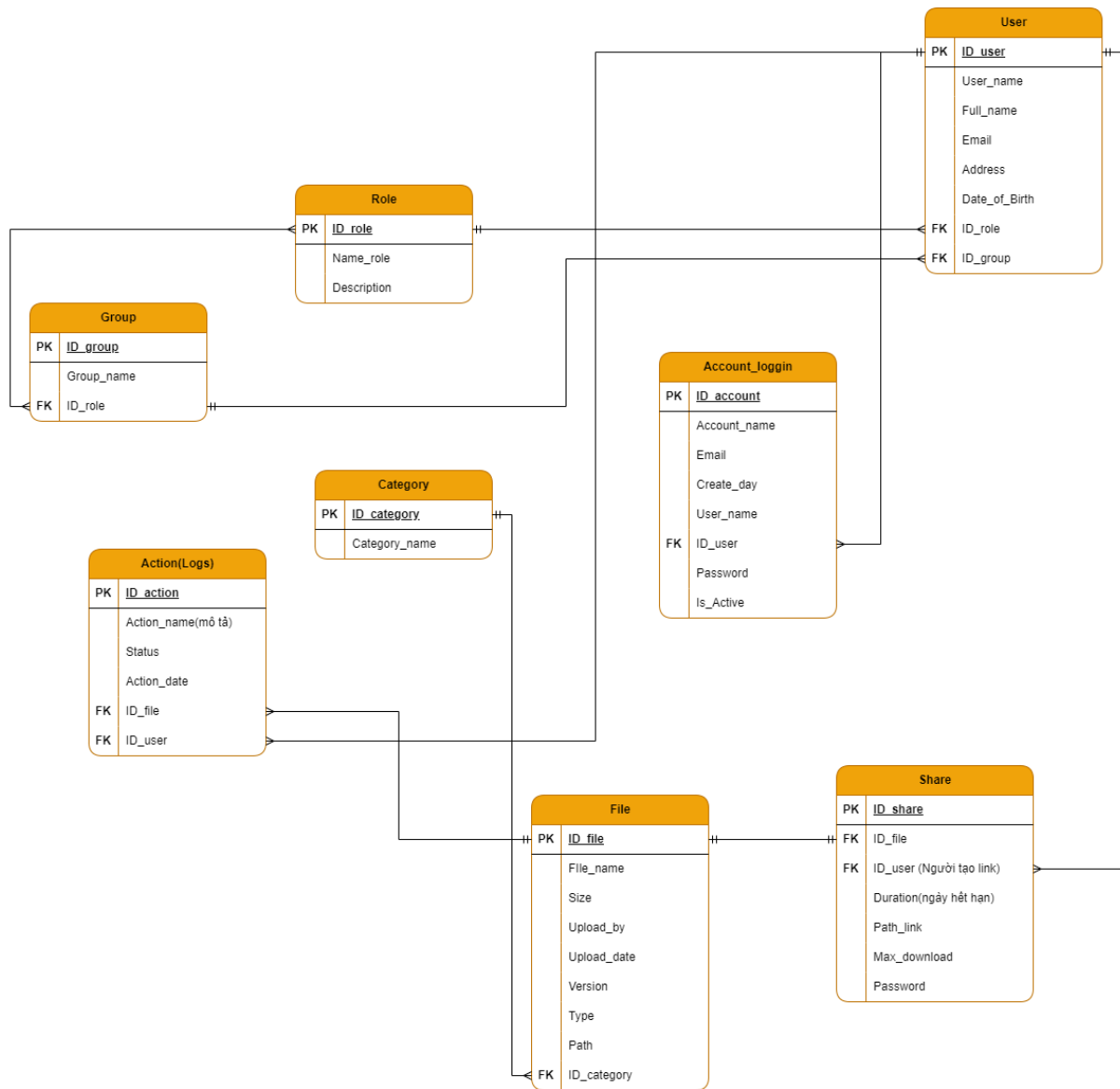
4.2 Kiến trúc vận hành

Hệ thống triển khai trên nền tảng AWS Cloud



Hình 4.2: Kiến trúc vận hành các service

4.3 Cơ sở dữ liệu



Hình 4.3: Cơ sở dữ liệu

1. Bảng Account_login - Tài khoản

| Trường | Kiểu dữ liệu | Diễn giải |
|--------------|--------------|---|
| ID | int | |
| Account_name | string | Tên tài khoản. |
| Email | string | Email đăng nhập/dăng ký tài khoản. |
| Created_day | datetime | Thời gian tạo tài khoản. |
| User_name | string | Tên User. |
| Password | string | Mật khẩu tài khoản. |
| Role | string | Phân loại người đăng nhập được phép làm gì. |
| ID_user | string | Mã người dùng. |
| Is_active | boolean | Tài khoản còn hoạt động không. |

Bảng 4.1: Mô tả bảng Account

2. Bảng Group - Nhóm

| Trường | Kiểu dữ liệu | Diễn giải |
|------------|--------------|-----------|
| ID | int | |
| Group_name | string | Tên nhóm. |
| ID_Role | string | Mã quyền. |

Bảng 4.2: Mô tả bảng Group

3. Bảng User - Người dùng

| Trường | Kiểu dữ liệu | Diễn giải |
|---------------|--------------|-------------------|
| ID | int | |
| User_name | string | Tên người dùng. |
| Full_name | string | Tên đầy đủ. |
| Email | string | Email người dùng. |
| Address | string | Địa chỉ user. |
| Date_of_Birth | datetime | Ngày sinh. |
| ID_role | string | Mã quyền. |
| ID_group | string | Mã nhóm. |

Bảng 4.3: Mô tả bảng User

4. Bảng Role - Quyền

| Trường | Kiểu dữ liệu | Diễn giải |
|-------------|--------------|------------|
| ID | int | |
| Name_Role | string | Tên quyền. |
| Description | string | Mô tả. |

Bảng 4.4: Mô tả bảng Role

5. Bảng File - Tập tin

| Trường | Kiểu dữ liệu | Diễn giải |
|-------------|--------------|--------------------------------|
| ID | int | |
| File_name | string | Tên file. |
| Size | int | Kích thước file(tính theo MB). |
| Upload_by | string | Người đăng file. |
| Upload_date | datetime | Ngày file được tải lên. |
| Version | string | Phiên bản của file. |
| Type | string | Loại file. |
| Path | string | Đường dẫn đến file. |
| ID_category | string | Mã danh mục. |

Bảng 4.5: Mô tả bảng File

6. Bảng Category - Danh mục

| Trường | Kiểu dữ liệu | Diễn giải |
|---------------|--------------|---------------|
| ID | int | |
| Category_name | string | Tên danh mục. |

Bảng 4.6: Mô tả bảng Category

7. Bảng Action - Hành động

| Trường | Kiểu dữ liệu | Diễn giải |
|---------------|--------------|----------------------------------|
| ID | int | |
| Action_name | string | Tên hành động. |
| Status | boolean | Trạng thái(đã chia sẻ hay chưa). |
| Action_date | datetime | Ngày thực hiện. |
| The_Number_of | int | Số lượng. |
| ID_file | string | Mã file. |
| ID_user | string | Mã người dùng. |

Bảng 4.7: Mô tả bảng Action

8. Bảng Share - Chia sẻ

| Trường | Kiểu dữ liệu | Diễn giải |
|--------------|--------------|---------------------------------------|
| ID | int | |
| ID_file | string | Mã file. |
| ID_user | string | Mã người dùng. |
| Duration | datetime | Thời hạn truy cập link. |
| Path_link | string | Đường dẫn truy cập. |
| Max_Download | int | Số lượng tải xuống tối đa. |
| Password | string | Mật khẩu để truy cập link. |
| Sharers | string | Người chia sẻ(người tạo đường link) . |

Bảng 4.8: Mô tả bảng Share

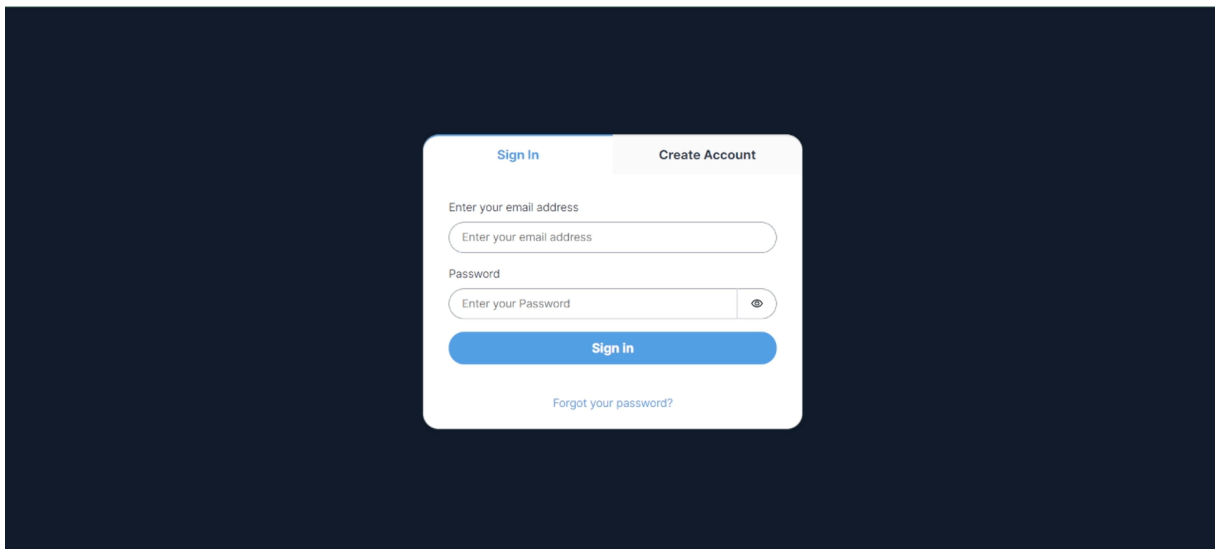
4.4 Thiết kế giao diện

Hiện chưa có cách gom nhóm các bảng đã thiết kế ở trên với hệ thống thực tế. Các bảng trên thực tế của hệ thống hiện tại đang riêng lẻ và không đổ chung vào một cơ sở dữ liệu đồng nhất.

Chương 5

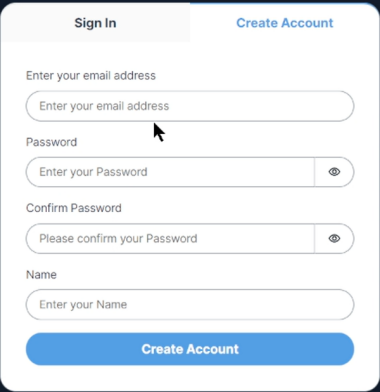
Triển khai hệ thống

5.1 Giao diện đăng nhập, đăng ký



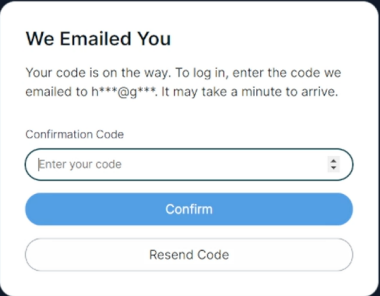
Hình 5.1: Giao diện đăng nhập

Người dùng thực hiện đăng nhập để sử dụng hệ thống.



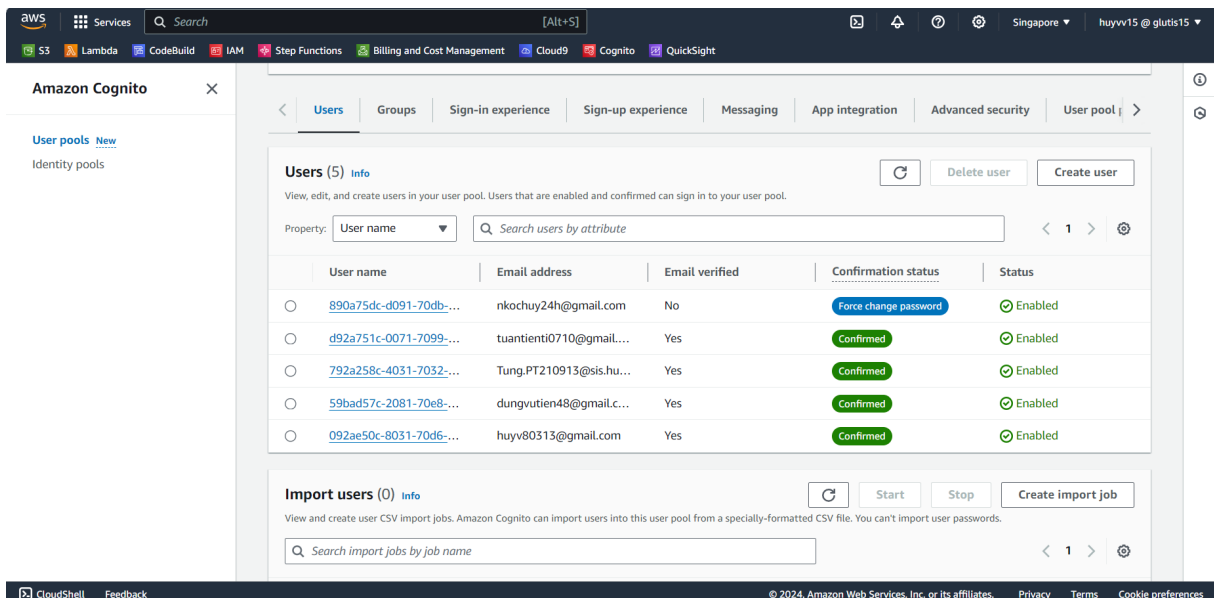
The image shows a user registration and login interface on a dark blue background. It features a white card with two tabs at the top: 'Sign In' (active) and 'Create Account'. Below the tabs are four input fields: 'Enter your email address', 'Enter your Password' (with a toggle icon), 'Please confirm your Password' (with a toggle icon), and 'Enter your Name'. A blue 'Create Account' button is positioned at the bottom of the card. A mouse cursor is hovering over the email input field.

Hình 5.2: Giao diện đăng ký



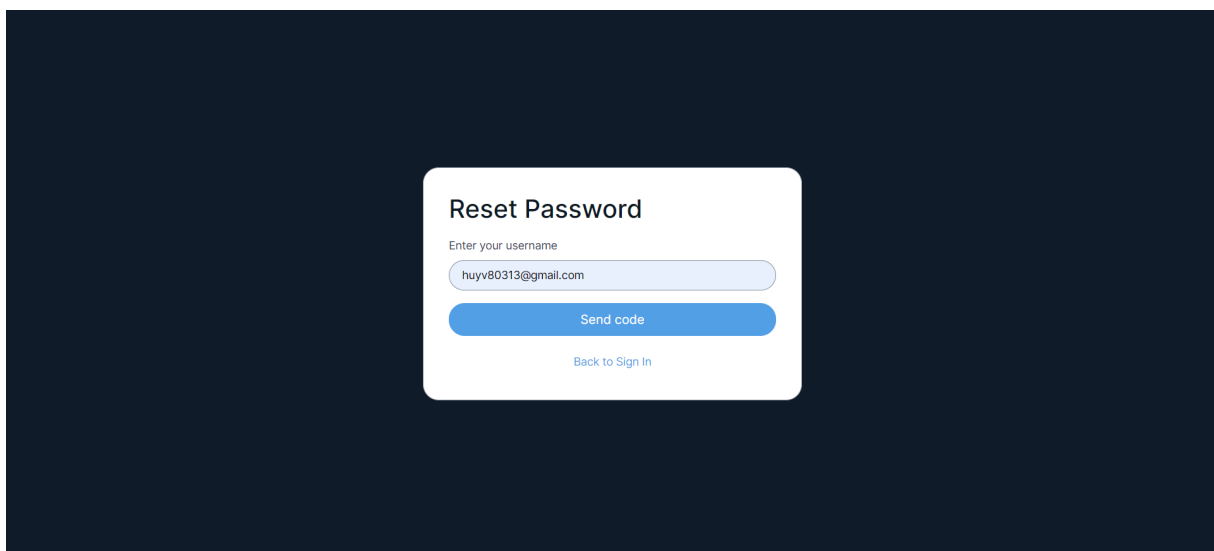
The image shows a confirmation screen on a dark blue background. It features a white card with the title 'We Emailed You'. Below the title is a message: 'Your code is on the way. To log in, enter the code we emailed to h***@g***. It may take a minute to arrive.' There is a 'Confirmation Code' input field with the placeholder 'Enter your code'. Below the input field are two buttons: a blue 'Confirm' button and a light blue 'Resend Code' button. A mouse cursor is pointing at the bottom of the card.

Hình 5.3: Xác thực mã code từ gmail

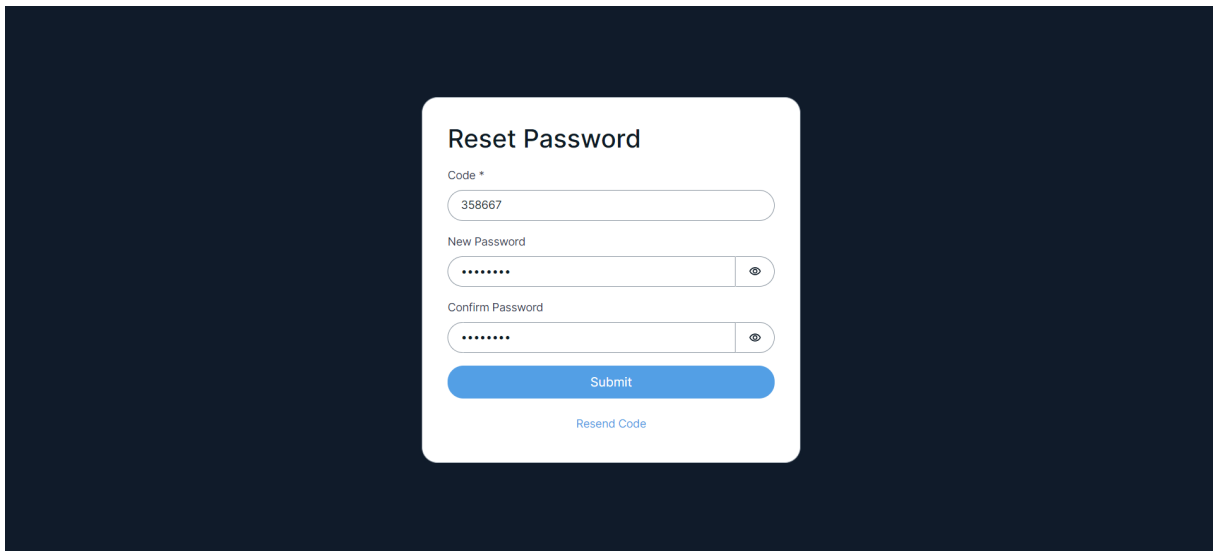


Hình 5.4: Thông tin các user sử dụng hệ thống

5.2 Quên mật khẩu

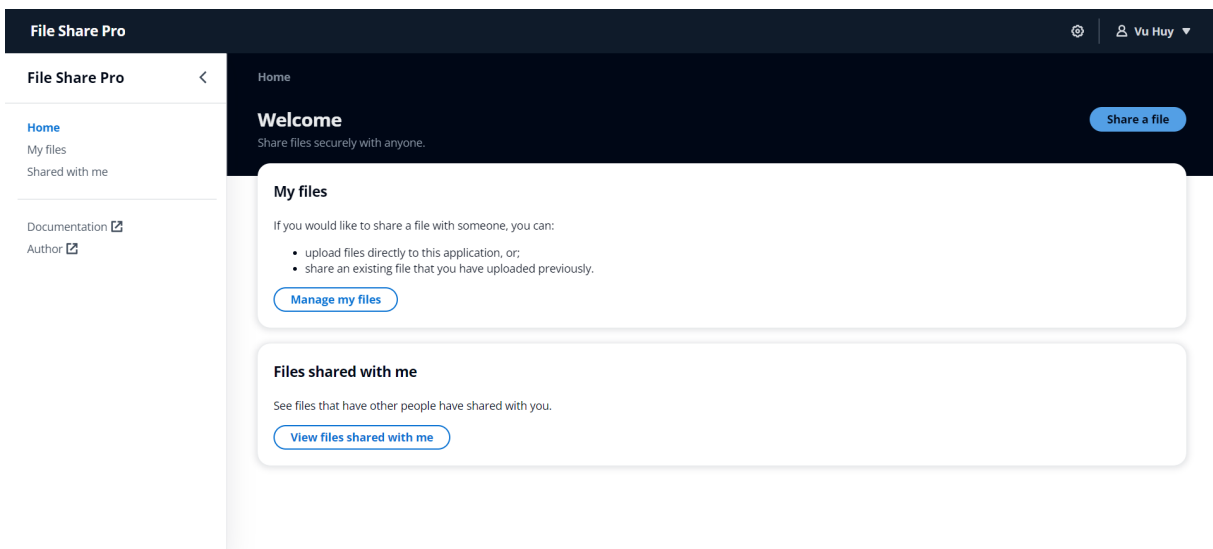


Hình 5.5: Chức năng quên mật khẩu

A screenshot of a 'Reset Password' form on a dark blue background. The form is white and contains three input fields: 'Code *' with the value '358667', 'New Password', and 'Confirm Password'. Each password field has a toggle icon for visibility. Below the fields is a blue 'Submit' button and a 'Resend Code' link.

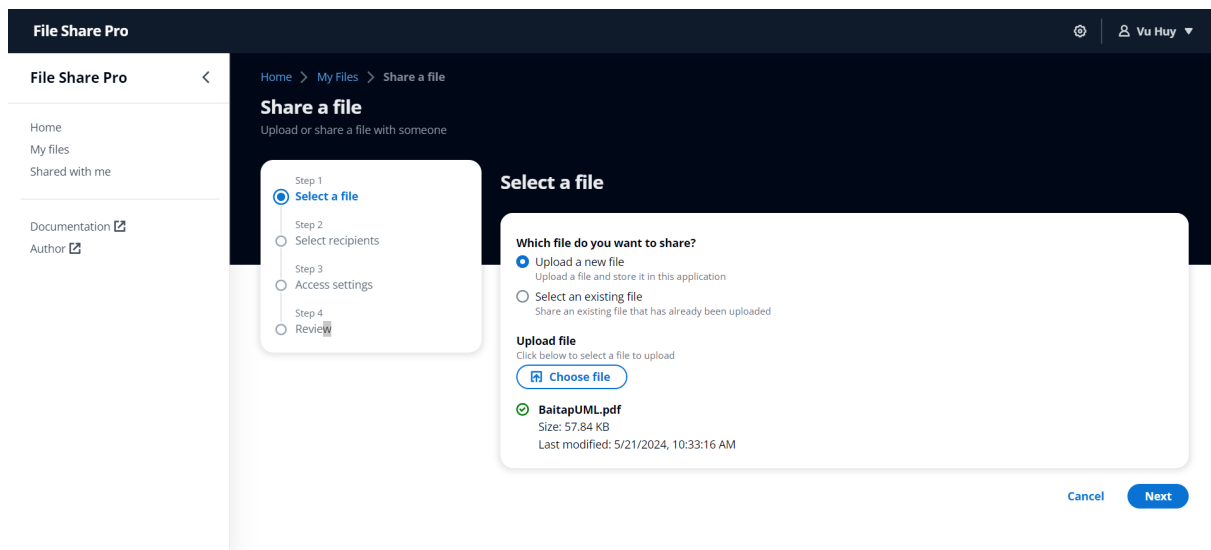
Hình 5.6: Nhập mã và mật khẩu mới

5.3 Giao diện trang chủ

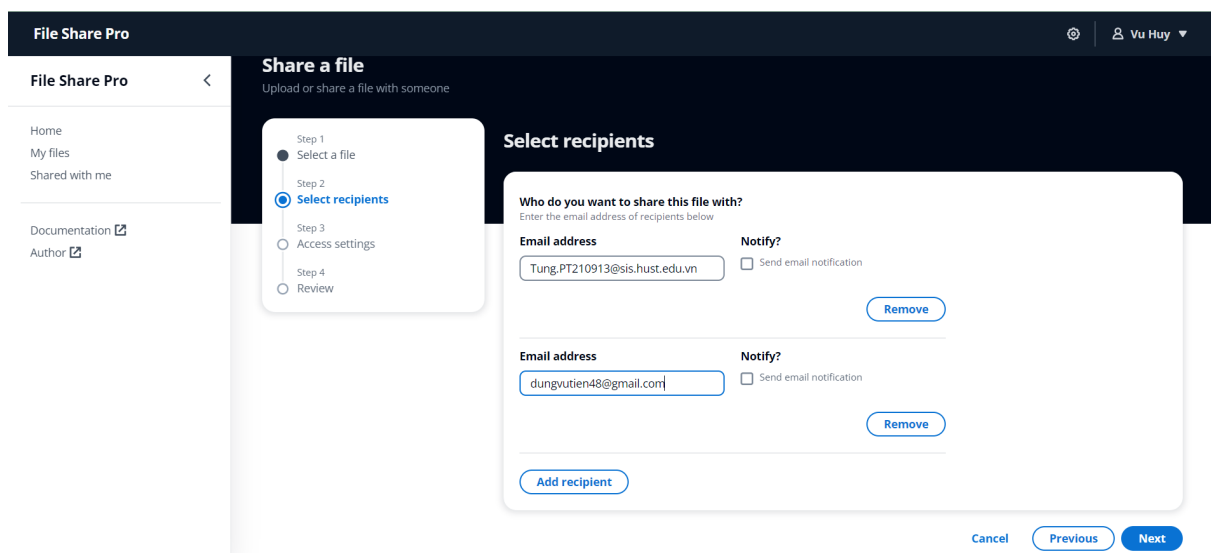


Hình 5.7: Trang chủ

5.4 Các bước chia sẻ File



Hình 5.11: Bước 1: Chọn file



Hình 5.12: Bước 2: Nhập mail người muốn chia sẻ

File Share Pro

Home
My files
Shared with me

Documentation

Author

Home > My Files > Share a file

Share a file

Upload or share a file with someone

Step 1
Select a file

Step 2
Select recipients

Step 3
Access settings

Step 4
Review

Access settings

Optional

Optional settings to restrict download access

☒ Expire link?

Expire link after specific date

Expiry date

2024/05/25

☒ Limit downloads?

Limit number of times the file can be downloaded

Enter a limit (number)

2

Cancel

Previous

Next

Hình 5.13: Bước 3: Thêm giới hạn về ngày và lượt tải

File Share Pro

Home
My files
Shared with me

Documentation

Author

Upload or share a file with someone

Review

File details

File type

Upload a new file

File

BaitapUML.pdf

Recipients

| Email | Notify |
|-------------------------------|--------|
| Tung.PT210913@sis.hust.edu.vn | No |
| dungvutien48@gmail.com | No |

Access Settings

Expiry

May 25, 2024

Download limit

2

Cancel

Previous

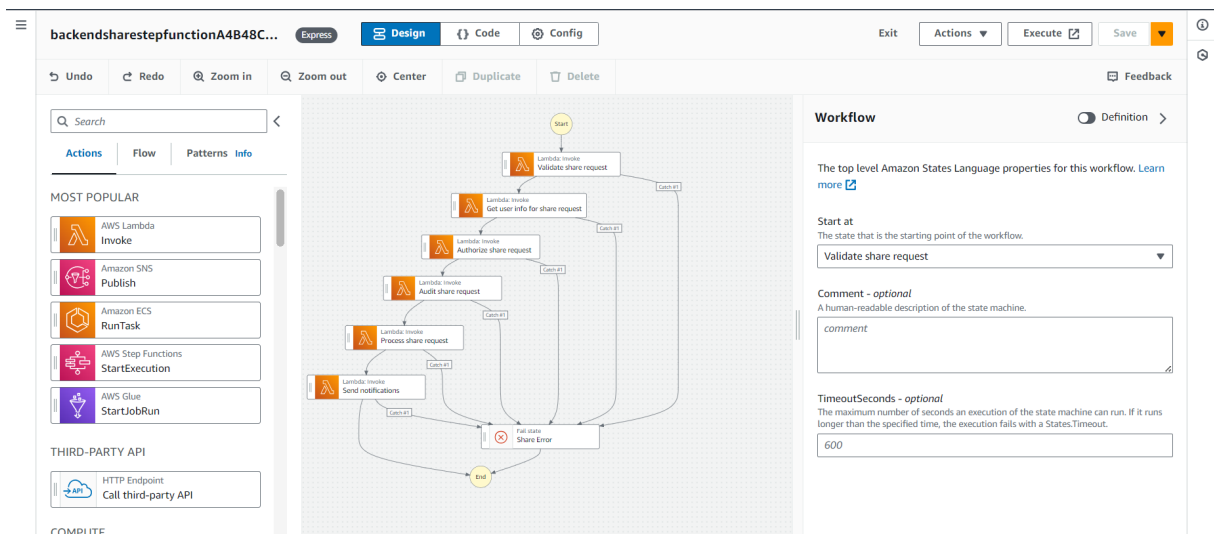
Submit

Hình 5.14: Bước 4: Review

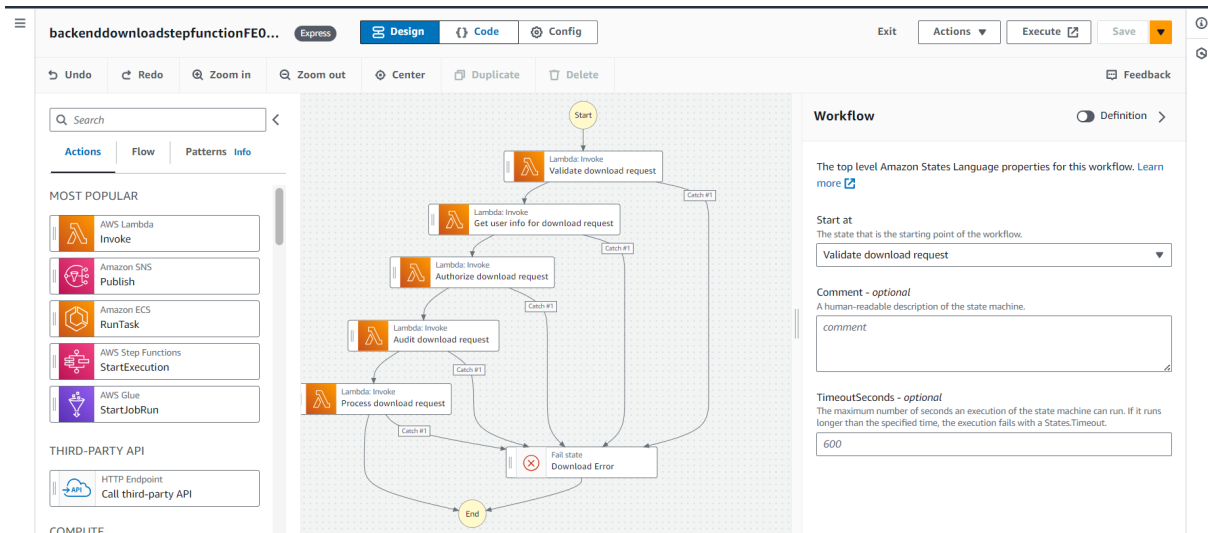
| fileid (String) | recipientEmail (String) | dateAdded | dateShared | downloadLimit | expiryDate | filename | notify |
|------------------------|------------------------------|--------------|---------------|---------------|---------------|--------------|--------|
| 3afdc9ae-ff41-4e03-... | - | 2024-05-2... | | | | BaitapUML... | |
| 3afdc9ae-ff41-4e03-... | Tung.PT210913@sis.hust.ed... | 2024-05-2... | 2024-05-24... | 2 | 2024-05-25... | BaitapUML... | |
| 3afdc9ae-ff41-4e03-... | dungvutien48@gmail.com | 2024-05-2... | 2024-05-24... | 2 | 2024-05-25... | BaitapUML... | |
| 8c921aa9-335c-4cd5-... | - | 2024-04-2... | | | | B%C3%A0i... | |
| 8c921aa9-335c-4cd5-... | tuantienti0710@gmail.com | 2024-04-2... | 2024-04-25... | | | B%C3%A0i... | |
| 49d52b3a-7c91-4936-... | - | 2024-04-2... | | | | 2024-04-0... | |
| 49d52b3a-7c91-4936-... | nkochuy24h@gmail.com | 2024-05-0... | 2024-05-09... | | | 2024-04-0... | |
| 49d52b3a-7c91-4936-... | tuantienti0710@gmail.com | 2024-04-2... | 2024-04-25... | | 2024-04-26... | 2024-04-0... | false |
| e14f8971-34d0-4952-... | - | 2024-04-2... | | | | GK-PTTKHT... | |
| e14f8971-34d0-4952-... | huyv80313@gmail.com | 2024-04-2... | 2024-04-25... | 2 | 2024-04-26... | GK-PTTKHT... | true |
| 3704b92a-f124-494b-... | - | 2024-05-0... | | | | CV_V%CS... | |
| 3704b92a-f124-494b-... | huyv80313@gmail.com | 2024-05-0... | 2024-05-07... | | | CV_V%CS... | |

Hình 5.15: Dữ liệu file chia sẻ

5.5 Luồng hoạt động các chức năng

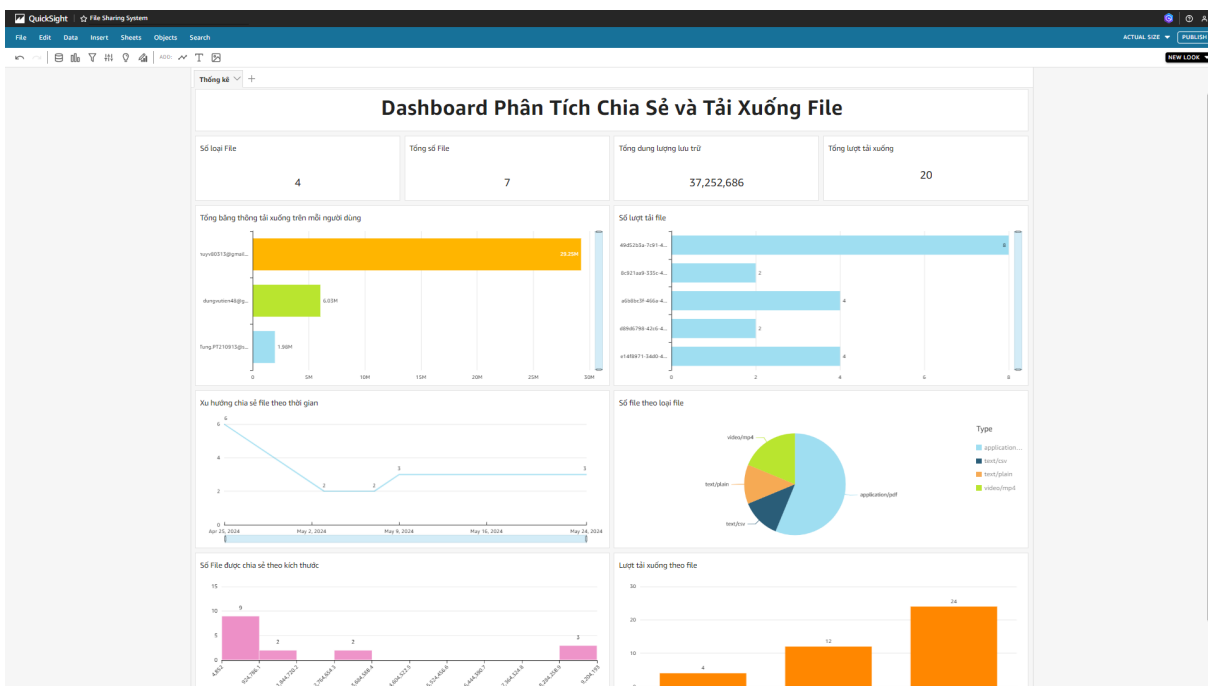


Hình 5.16: Luồng xác thực tài khoản



Hình 5.17: Luồng tải xuống file

5.6 Dashboard phân tích



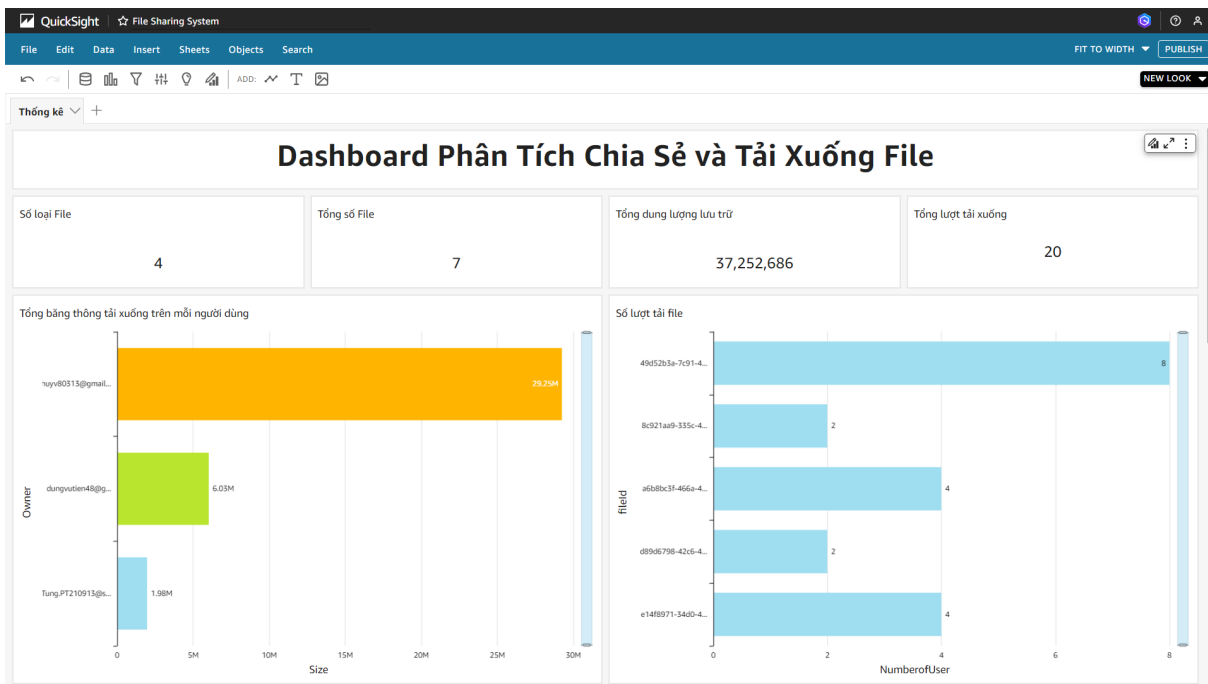
Hình 5.18: Dashboard chung

Do hệ thống chỉ mới được triển khai gần đây, khối lượng dữ liệu sinh ra từ hệ thống vẫn còn khá hạn chế. Điều này có thể hiểu được vì thời gian hoạt động của hệ thống chưa đủ dài để tích lũy một lượng dữ liệu phong phú. Hơn nữa, việc thu thập dữ liệu và phản hồi từ người dùng cũng mới chỉ ở giai đoạn khởi đầu.

Hiện tại, hệ thống đang phục vụ một nhóm gồm 5 người dùng. Đây là một bước đầu tiên quan trọng để kiểm tra tính khả dụng và hiệu quả của hệ thống. Những

người dùng này đóng vai trò quan trọng trong việc cung cấp các phản hồi quý giá, giúp nhà phát triển điều chỉnh và cải tiến hệ thống. Thông qua sự tương tác và trải nghiệm thực tế của họ, nhà phát triển có thể nhận diện các vấn đề tiềm ẩn cũng như các điểm mạnh của hệ thống, từ đó đưa ra các giải pháp tối ưu.

Mặc dù lượng dữ liệu được thu thập và phân tích trong phạm vi nghiên cứu này không quá lớn và chưa thể bao quát hết toàn bộ tình hình chia sẻ và lưu trữ dữ liệu, tuy nhiên em vẫn sẽ cố gắng khai thác triệt để các thông tin có giá trị từ tập dữ liệu đó. Thông qua việc trực quan hóa dữ liệu bằng các biểu đồ và dashboard thân thiện, hi vọng có thể phác họa được một bức tranh tổng quan về thực trạng hiện tại, từ đó phát hiện ra những vấn đề tồn tại và đưa ra các giải pháp, khuyến nghị phù hợp. Dưới đây là các phân tích đó:



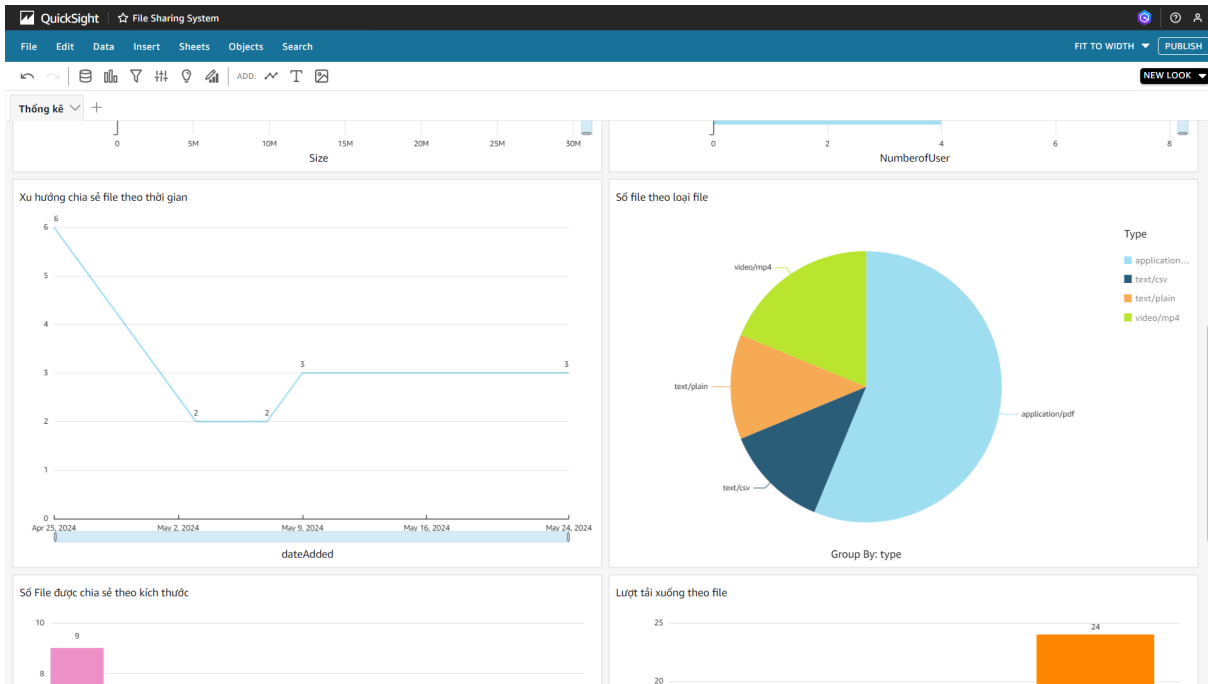
Hình 5.19: Dashboard phần 1

Số lượng file và tổng dung lượng lưu trữ: Tổng số lượng file hiện có là 7 file. Tổng dung lượng lưu trữ cho các file này là 37,252,686 bytes (khoảng 37.3MB). Số lượng file không quá lớn, nhưng dung lượng lưu trữ cũng không quá nhỏ, cho thấy các file này có kích thước khá lớn.

Phân bố kích thước file: Biểu đồ cột thể hiện phân bố kích thước của các file. Có một file chiếm tới 29,954 bytes (khoảng 29.9MB), rất lớn so với các file khác. Một file khác có kích thước 6,034 bytes (khoảng 6MB), cũng khá lớn. Các file còn

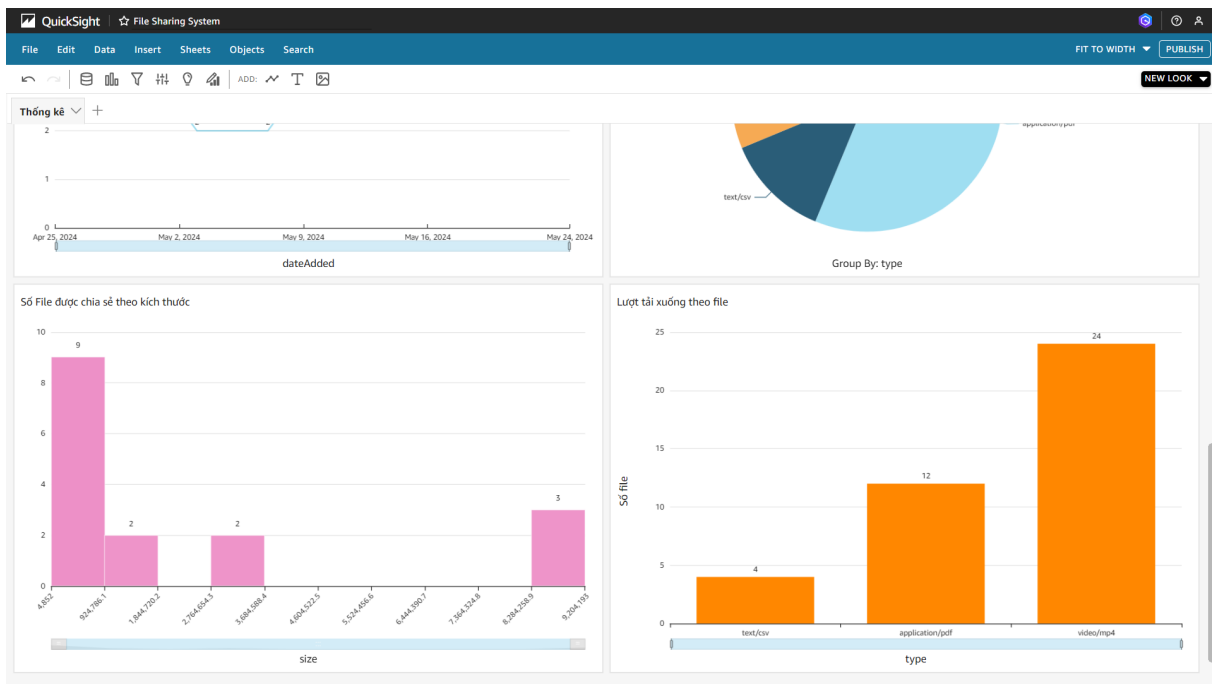
lại có kích thước nhỏ hơn, từ vài trăm KB đến vài MB.

Số lượt tải xuống theo số người dùng: Biểu đồ thanh ngang hiển thị số lượt tải xuống file theo số người dùng. Đa số chỉ có 1-2 người dùng tải xuống các file. Có một file được 8 người dùng tải xuống, là file có số lượt tải xuống cao nhất. Tổng số lượt tải xuống là 20 lượt.



Hình 5.20: Dashboard phần 2

Nhận xét chung: Số lượng file và người dùng không quá lớn, có thể là trong phạm vi một nhóm/dự án nhỏ. Việc chia sẻ và tải xuống file chưa được phổ biến rộng rãi, chủ yếu tập trung vào một số file lớn.



Hình 5.21: Dashboard phần 3

File được tải xuống hoặc chia sẻ chủ yếu thuộc định dạng video hoặc file âm thanh. Một số ít hơn là những file tài liệu, file dữ liệu như csv.

Chương 6

Kết luận

Đồ án đã hoàn thành hệ thống chia sẻ file tự xây dựng, đáp ứng tốt một phần nhu cầu cụ thể và mang lại nhiều lợi ích cho doanh nghiệp. Hệ thống này đã được tùy chỉnh theo đặc thù riêng, đảm bảo tính bảo mật cao và tiết kiệm chi phí về lâu dài. Hệ thống đã đạt được các mục tiêu quan trọng như cung cấp giải pháp lưu trữ và quản lý dữ liệu an toàn, tăng cường khả năng chia sẻ dữ liệu linh hoạt, và đảm bảo tuân thủ các quy định về bảo mật và quyền riêng tư thông tin.

Kết quả đạt được:

Sau quá trình thực hiện đồ án này em đã tiếp thu được nhiều kiến thức nhất định về kho dữ liệu phát triển trên cloud. Ngoài ra còn các kiến thức như:

- Hiểu Big Data và tầm quan trọng của nó các lĩnh vực khác nhau.
- Sử dụng công nghệ của AWS để lưu trữ dữ liệu.
- Học cách khảo sát và đánh giá các hệ thống chia sẻ file hiện có trên thị trường. Qua đó hiểu rõ hơn về những ưu và nhược điểm của từng hệ thống, cũng như các tiêu chí cần thiết để xây dựng một hệ thống hiệu quả và bảo mật.
- Thiết kế được hệ thống lưu trữ với giao diện đơn giản.
- Sử dụng QuickSight để trực quan hóa dữ liệu bằng báo cáo.

Hướng phát triển trong tương lai:

Đồ án đã cơ bản hoàn thành được mục tiêu đã đề ra. Tuy nhiên bản thân em cũng có một số định hướng để phát triển hệ thống trong tương lai như:

- Tìm hiểu và ứng dụng nhiều hơn về các dịch vụ của AWS.

- Tìm hiểu các phương pháp để tối ưu chi phí lưu trữ.
- Cải thiện giao diện người dùng và bổ sung thêm nhiều tính năng.

Tài liệu tham khảo

- [1] Nguyễn Văn Vy, *Giáo Trình Phân Tích Thiết Kế Các Hệ Thống Thông Tin*, NXB Giáo Dục, 2010.
- [2] Nguyễn Văn Ba, *Phân tích và thiết kế hệ thống thông tin*, 2003.
- [3] Raul Sidnei Wazlawick, *Object-Oriented Analysis and Design for Information Systems: Modeling with UML, OCL, and IFML*, 2014.
- [4] Nguyễn Danh Tú, *Slide bài giảng kho dữ liệu và kinh doanh thông minh*.
- [5] AWS Documentation, <https://docs.aws.amazon.com/>.
- [6] Nguyễn Thị Thanh Huyền, *Slide bài giảng cơ sở dữ liệu*.