

TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI

ĐỒ ÁN TỐT NGHIỆP

Thiết kế hệ thống quản lý phân phối
trong chuỗi cung ứng

TẠ QUANG TÙNG

tung.tq154280@sis.hust.edu.vn

Ngành Công nghệ thông tin

Lớp Kỹ sư tài năng - Công nghệ thông tin - K60

Giảng viên hướng dẫn: TS. Phạm Quang Dũng

Chữ ký của GVHD

Bộ môn: Khoa học máy tính

Viện: Công nghệ thông tin và truyền thông

HÀ NỘI, 6/2020

Lời cảm ơn

Đầu tiên, em xin được gửi lời cảm ơn chân thành đến các thầy giáo, cô giáo thuộc trường đại học Bách Khoa Hà Nội, đặc biệt là các thầy giáo, cô giáo thuộc Viện Công nghệ Thông tin và Truyền Thông đã tận tình dạy dỗ và trang bị cho em những kiến thức bổ ích trong năm năm vừa qua.

Đồng thời em cũng xin được gửi lời cảm ơn đặc biệt đến TS. Phạm Quang Dũng. Thầy đã cho em những kiến thức và kinh nghiệm quý báu giúp em có thể hoàn thành đồ án này.

Em xin gửi lời cảm ơn tới gia đình và bạn bè đã sát cánh, động viên và hỗ trợ em trong suốt những năm qua.

Tóm tắt

Trong ngày nay, xã hội ngày càng hiện đại hóa, nhu cầu con người cũng cao lên. Cùng với đó kéo theo sự phát triển của các công ty, cửa hàng bán lẻ và chợ thương mại điện tử. Nhu cầu về quản lý hàng hóa bán lẻ cũng ngày càng tăng lên. Trên thị trường đã có một số ứng dụng cho quản lý bán lẻ phổ biến như KiotViet hay Sapo, tuy nhiên những sản phẩm này còn sử dụng công nghệ cũ và còn một số hạn chế. Vì vậy em muốn xây dựng một phần mềm quản lý bán lẻ trực tuyến cung cấp các giải pháp quản lý phân phối hàng hóa.

Trong những năm gần đây, công nghệ Web đang càng trở nên phổ biến trong xây dựng các ứng dụng quản lý doanh nghiệp. Đồng thời nhờ sự phát triển của trình duyệt và ngôn ngữ JavaScript đã đem đến nhiều các Framework hỗ trợ lập trình viên xây dựng những ứng dụng phức tạp, tương tác cao, đơn trang (Single Page Application) mà không thể hoặc khó có thể thực hiện được dựa vào công nghệ làm Web cũ. Trong đồ án này, em đã lựa chọn những công nghệ được coi là phổ biến được sử dụng nhiều hiện nay như ReactJS, Redux, Redux-Saga, Material-UI, Go làm back-end, Redis, PostgreSQL.

Đến nay, đồ án đã hoàn thành được các vấn đề đặt ra và em có mong muốn tiếp tục duy trì và phát triển thêm tính năng cho ứng dụng đồng thời đưa hệ thống vào chạy trong thực tế.

Mục lục

Lời cảm ơn	2
Tóm tắt	2
1 Giới thiệu đề tài	11
1.1 Đặt vấn đề	11
1.2 Mục tiêu và phạm vi đề tài	12
1.3 Định hướng giải pháp	12
1.4 Bố cục đồ án	13
2 Cơ sở lý thuyết và công nghệ sử dụng	14
2.1 Giải pháp quản lý phân phối	14
2.1.1 Các khái niệm trong quản lý phân phối	14
2.1.2 Lợi ích của hệ thống quản lý phân phối	15
2.1.3 Khảo sát thực tế việc sử dụng hệ thống quản lý phân phối	16
2.2 Cơ sở lý thuyết	19
2.2.1 Role-Based Access Control	19

2.2.2	Thuật toán phân cụm dữ liệu	21
2.3	Các công nghệ sử dụng	23
2.3.1	Công nghệ lưu trữ - Redis	23
2.3.2	Công nghệ lưu trữ - PostgreSQL	26
2.3.3	Công nghệ back-end	32
2.3.4	Docker	34
2.3.5	Kubernetes	36
2.3.6	Microsoft Azure	38
3	Phân tích thiết kế hệ thống	40
3.1	Tổng quan các chức năng	40
3.1.1	Biểu đồ use case tổng quan	40
3.1.2	Biểu đồ use case phân rã các chức năng của hệ thống	41
3.2	Thiết kế cơ sở dữ liệu	52
4	Các giải pháp, đóng góp nổi bật và minh họa các chức năng của hệ thống	57
4.1	Vấn đề lựa chọn công nghệ	57
4.2	Chức năng phân quyền động	58
4.3	Chức năng phân cụm cửa hàng	59
4.4	Minh họa các chức năng của hệ thống	60
5	Thử nghiệm và đánh giá	67
5.1	Triển khai chạy thử nghiệm trên cloud	67

5.2	Kiểm thử hiệu năng (performance testing)	68
5.2.1	Kiểm thử hiệu năng cơ sở dữ liệu	69
5.2.2	Kiểm thử hiệu năng Rest API	71
5.3	Kiểm thử sức chịu tải (stress testing)	72
6	Kết luận và hướng phát triển	74
6.1	Kết luận	74
6.2	Hướng phát triển của đồ án trong tương lai	75

Danh sách bảng

2.1	So sánh Redis và MySQL	26
3.1	Thông tin các bảng nhóm quản lý phân quyền, người dùng	53
3.2	Thông tin các bảng nhóm quản lý phân quyền, người dùng	54
3.3	Thông tin các bảng nhóm quản lý sản phẩm	55
3.4	Thông tin các bảng nhóm quản lý tuyến bán hàng	56
4.1	Security_group và permission tương ứng	58

Danh sách hình vẽ

2.1	Các mô hình RBAC	21
2.2	Biểu đồ khối mô tả thuật toán K-Means	22
2.3	Redis server	24
2.4	Ứng dụng Redis trong bảng xếp hạng game	25
2.5	Cấu trúc của GIN	31
3.1	Biểu đồ use case tổng quan	40
3.2	Use-case quản lý người dùng	41
3.3	Biểu đồ hoạt động ca sử dụng quản lý người dùng	42
3.4	Biểu đồ hoạt động cho thao tác phân quyền	43
3.5	Use case quản lý kho	44
3.6	Biểu đồ hoạt động cho thao tác thêm cửa hàng bán lẻ	45
3.7	Use case quản lý xuất – nhập hàng	46
3.8	Use-case quản lý sản phẩm	47
3.9	Biểu đồ hoạt động cho thao tác gán giá của sản phẩm	48
3.10	Biểu đồ hoạt động cho thao tác thêm đơn hàng	49

3.11 Use case quản lý tuyển bán hàng	50
3.12 Biểu đồ hoạt động cho thao tác thêm tuyển bán hàng cho nhân viên bán hàng .	51
3.13 Use-case nhân viên bán hàng check-in	52
3.14 Cơ sở dữ liệu quản lý phân quyền, người dùng	53
3.15 Cơ sở dữ liệu quản lý đơn hàng	54
3.16 Cơ sở dữ liệu quản lý đơn hàng	55
3.17 Cơ sở dữ liệu quản lý tuyển bán hàng	56
4.1 Chức năng đăng nhập	60
4.2 Giao diện trang chủ hệ thống	60
4.3 Quản lý người dùng	61
4.4 Thêm người dùng / khách hàng	61
4.5 Quản lý các nhóm quyền	62
4.6 Gán quyền cho tài khoản người dùng	62
4.7 Gán quyền cho tài khoản người dùng	63
4.8 Tạo đơn hàng	63
4.9 Quản lý sản phẩm và giá	64
4.10 Thêm mới một cửa hàng bán lẻ	65
4.11 Tạo lịch trình tuyển bán hàng	66
4.12 Salesman check-in	66
5.1 Kiến trúc tổng quan của hệ thống	67

5.2	Triển khai lên Azure Kubernetes Service	68
5.3	Truy vấn count trên máy local	69
5.4	Truy vấn count trên máy cloud	69
5.5	Truy vấn select trên máy local	70
5.6	Truy vấn select trên máy cloud	70
5.7	Hiệu năng Rest API trên máy local (no aggregation)	71
5.8	Hiệu năng Rest API trên máy local (aggregated)	71
5.9	Hiệu năng Rest API trên máy cloud (no aggregation)	71
5.10	Hiệu năng Rest API trên máy cloud (aggregated)	72
5.11	Stress test thêm inventory item trên máy local	72
5.12	Stress test thêm inventory item trên máy cloud	73

Listings

- 2.1 Tạo index sử dụng Hash 28
- 2.2 Tạo index sử dụng GIN 30
- 2.3 Tạo index sử dụng GiST 30
- 2.4 Xây dựng image của back-end server 36

Chương 1

Giới thiệu đề tài

1.1 Đặt vấn đề

Hiện nay với sự bùng nổ của thương mại điện tử, việc mua bán và thanh toán trực tuyến đã trở nên phổ biến rộng rãi trong cộng đồng, vì vậy lượng hàng hóa cung và cầu mỗi ngày luôn rất lớn. Điều này rất có lợi cho những người tiêu dùng, bởi sẽ có rất nhiều mặt hàng với đủ mọi chủng loại, càng nhiều lựa chọn mua hàng thì giá cả càng rẻ, những người tiêu dùng thông thái sẽ có thể mua được những sản phẩm tốt với mức giá phải chăng. Đối với những người tiêu dùng nghi ngờ về hình ảnh, chất lượng của các sản phẩm trực tuyến, họ hoàn toàn có thể đến các cửa hàng bán lẻ, nơi mà hàng hóa vẫn luôn dồi dào, được nhìn tận mắt, được cầm tận tay.

Các cửa hàng bán lẻ là điểm cuối trong mạng lưới vận chuyển hàng hóa trước khi sản phẩm đến được tay người tiêu dùng. Khi hàng hóa nhiều lên và các phương pháp quản lý kho cũ không còn đáp ứng được thì mỗi công ty hay tập đoàn bán lẻ đều cần giải pháp mới để quản lý hàng hóa.

Hệ thống quản lý phân phối (Distribution Management System) là một ứng dụng cung cấp giải pháp cho vấn đề này. Với hệ thống quản lý phân phối, các công ty hay tập đoàn bán lẻ có thể nắm được chính xác biến động thị trường, hàng hóa đang ở vị trí nào trong chuỗi cung ứng, hàng tồn kho còn bao nhiêu hay hiệu suất làm việc của các nhân viên bán hàng, ...

1.2 Mục tiêu và phạm vi đề tài

Từ sự cần thiết phải chuyển đổi phương pháp quản lý phân phối này, nhiều phần mềm nghiệp vụ đã được triển khai. Ở Việt Nam hiện nay, các ứng dụng quản lý phân phối trực tuyến phổ biến có thể kể đến như KiotViet, Sapo, Suno, ..., các ứng dụng quản lý không cần trực tuyến như Adaline, BS Silver, Perfect Warehouse, ... Điểm chung của các ứng dụng này là đều cung cấp các tính năng phục vụ về quản lý kho, nhập – xuất sản phẩm, các ứng dụng trực tuyến thì đưa ra nhiều giải pháp hơn bao gồm cả quản lý nhân viên, quản lý lương, thậm chí có cả kênh bán hàng riêng. KiotViet hay Sapo là những cái tên tiêu biểu khi mà sản phẩm của hai tập đoàn này khá phổ biến.

Trong quá trình sử dụng thử phần mềm của KiotViet, em thấy các tính năng của họ đưa ra là rất tốt và khá đầy đủ cho nhiều loại sản phẩm, mặt hàng. Tuy nhiên, em muốn bổ sung thêm tính năng quản lý tuyển bán hàng thông qua bản đồ, quản lý hiệu quả làm việc của nhân viên bán hàng qua thông tin check-in, cải thiện giao diện cũng như tăng hiệu năng của ứng dụng.

Ứng dụng quản lý phân phối của em sẽ tập trung vào 7 tính năng chính, đó là phân quyền động, quản lý tài khoản, quản lý sản phẩm, quản lý kho, quản lý xuất nhập, quản lý tuyển bán hàng và quản lý nhân viên bán hàng.

1.3 Định hướng giải pháp

Go là một ngôn ngữ lập trình mới do Google phát triển, được sinh ra để giúp ngành công nghiệp phần mềm khai thác nền tảng đa lõi của bộ vi xử lý và hoạt động đa nhiệm tốt hơn. Vì vậy em sử dụng Go để viết back-end server nhằm tăng hiệu năng xử lý yêu cầu trên một thời điểm.

Còn về phần giao diện, giao tiếp với người dùng (front-end) thì hiện nay các Framework Javascript như ReactJS, Angular hay VueJS đang là xu thế bởi khả năng xây dựng giao diện nhanh, bảo trì và mở rộng code dễ dàng. Ứng dụng của em sử dụng ReactJS, kết hợp Redux và Material-UI để xây dựng giao diện đảm bảo thiết kế chuẩn Material Design của Google.

1.4 Bố cục đề án

Phần còn lại của báo cáo đề án tốt nghiệp được tổ chức thành các chương như sau.

Chương 2 trình bày về cơ sở lý thuyết của đề án. Trong đó bao gồm các công nghệ cốt lõi đã tìm hiểu để xây dựng ứng dụng, so sánh giữa các công nghệ được sử dụng với các công nghệ khác hiện nay. Bên cạnh đó là các thuật toán cơ sở được sử dụng trong xử lý logic cho cả phần front-end, back-end và cơ sở dữ liệu như Index, Role-Based Access Control, thuật toán phân cụm.

Chương 3 trình bày chi tiết về thiết kế hệ thống. Từ sự cần thiết của giải pháp đã nêu ở trên, em xác định ra những tính năng cần thiết nhất và xây dựng các ca sử dụng xung quanh những tính năng này. Phần này sẽ trình bày các biểu đồ ca sử dụng cho các chức năng, biểu đồ hoạt động thể hiện cách thức tương tác với hệ thống của người dùng. Cùng với đó là thiết kế cơ sở dữ liệu, xây dựng dữ liệu mẫu.

Chương 4 đưa ra đóng góp chính của em trong đề án tốt nghiệp này, so sánh chúng với những giải pháp hiện tại, cơ hội áp dụng trong tương lai và khả năng mở rộng của đóng góp này. Chương này cũng đưa ra những vấn đề và hướng giải quyết của những vấn đề này trong quá trình thực hiện đề án.

Chương 5 minh họa các chức năng của hệ thống, mô tả quy trình triển khai ứng dụng lên server. Đánh giá hiệu năng hệ thống bằng stress test và performance test. Đưa ra các vấn đề, thách thức khi triển khai hệ thống với nhiều người sử dụng và với rất nhiều người sử dụng.

Chương 6 là chương cuối cùng đưa ra kết luận, những vấn đề còn chưa giải quyết được và hướng phát triển của ứng dụng trong tương lai.

Chương 2

Cơ sở lý thuyết và công nghệ sử dụng

2.1 Giải pháp quản lý phân phối

2.1.1 Các khái niệm trong quản lý phân phối

Trước tiên, ta cần tìm biết về một từ khóa đã trở nên rất phổ biến trong thời gian gần đây, đó là Logistics. Hiểu một cách đơn giản thì Logistics là quá trình lên kế hoạch, áp dụng và kiểm soát các luồng dịch chuyển của hàng hóa hay thông tin liên quan tới nguyên nhiên liệu vật tư (đầu vào) và sản phẩm cuối cùng (đầu ra) từ thời điểm xuất phát tới điểm tiêu thụ. Đi kèm với Logistics là khái niệm về Quản trị chuỗi cung ứng (Supply Chain Management), Quản trị chuỗi cung ứng bao gồm tất cả những hoạt động quản trị logistics cũng như những hoạt động sản xuất và thúc đẩy sự phối hợp về quy trình và hoạt động của các bộ phận marketing, kinh doanh, thiết kế sản phẩm, tài chính, công nghệ thông tin. Khái niệm chuỗi cung ứng rộng hơn, bao gồm cả logistics và quá trình sản xuất. Ngoài ra chuỗi cung ứng chú trọng hơn đến hoạt động mua hàng (procurement) trong khi logistics giải quyết các vấn đề chiến lược, phối hợp giữa marketing và sản xuất.

Khi đã có một hình dung cơ bản về chuỗi cung ứng, ta đến với phần chính là **hệ thống quản lý phân phối** (Distribution Management System). Hệ thống quản lý phân phối (Distribution Management System) là phần mềm quản lý chuỗi cung ứng hàng hóa của doanh nghiệp, giúp họ quản lý các hoạt động phân phối hàng hóa ra thị trường, kiểm soát các kênh phân phối,

quản lý nhân viên, người bán hàng, kiểm soát hàng hóa trong kho, hàng tồn kho, kế hoạch vận chuyển hàng hóa đến địa chỉ mua hàng,...

Ví dụ, Masan Group là một công ty lớn trong lĩnh vực kinh tế tư nhân Việt Nam, tập trung hoạt động trong ngành hàng tiêu dùng và tài nguyên của Việt Nam. Masan quản lý các nền tảng kinh doanh có quy mô lớn nhằm phát triển và khai thác các tiềm năng trong lĩnh vực tiêu dùng và tài nguyên. Còn rất nhiều những tập đoàn bán lẻ khác có cùng cơ chế quản lý kinh doanh như Masan, mỗi tập đoàn như vậy đều cần có một hệ thống riêng để quản lý chuỗi cung ứng của mình.

Vì vậy Distribution Management System là một trong số những phần mềm quản lý doanh nghiệp có tính ứng dụng cao, phù hợp với mọi doanh nghiệp sản xuất và phân phối. Đối với các doanh nghiệp lớn (ví dụ Masan, VinGroup) có đội ngũ nhân viên bán hàng đông đảo và các kênh phân phối phức tạp, phần mềm DMS càng quan trọng và là công cụ không thể thiếu. Các nhà quản lý của các tập đoàn lớn này luôn đau đầu vì những câu hỏi như “Làm thế nào để nắm được nhanh nhất xu thế, biến động của thị trường?”, “Làm thế nào để kiểm soát phân phối tốt, duy trì tồn kho ở mức tối ưu, tiết kiệm thời gian?”, “Tự động hóa bán hàng, tăng hiệu quả bán hàng cho đội ngũ nhân viên bán hàng như thế nào?”. Với hệ thống quản lý kênh phân phối (DMS), các doanh nghiệp lớn với vài trăm, hoặc vài nghìn nhân viên bán hàng, hàng chục nghìn điểm bán sẽ dễ dàng làm được việc này.

2.1.2 Lợi ích của hệ thống quản lý phân phối

Phần trên đã trình bày về lý do tại sao các doanh nghiệp và tập đoàn bán lẻ phải sử dụng giải pháp quản lý phân phối và sau đây là những lợi ích chính, thấy rõ nhất theo như tìm hiểu của em.

Thứ nhất Distribution Management System là công cụ tự động hóa bán hàng, giúp nhân viên bán hàng (salesman) tiết kiệm thời gian, tăng chất lượng chăm sóc khách hàng, tối ưu doanh thu. Mọi thông tin nhân viên bán hàng cần khi ghé thăm một điểm bán lẻ (rental outlet) bao gồm thông tin về khách hàng (customer), lịch sử mua hàng, các báo cáo bán hàng, thông tin sản phẩm, chương trình khuyến mãi,... có sẵn trong ứng dụng di động DMS để nhân viên

có thể dễ dàng theo dõi tại điểm bán.

Thứ hai Distribution Management System có thể quản lý hiệu quả làm việc của nhân viên bán hàng, nhà quản lý có thể nắm được lộ trình ghé thăm khách hàng của nhân viên trên bản đồ lịch sử checkin. Thông qua đó đánh giá được nhân viên bán hàng có đang tích cực ngoài thị trường hay không, khách hàng có được chăm sóc tốt hay không.

Thứ ba, đặc biệt hơn là qua hệ thống nhà quản trị sẽ biết được những cửa hàng nào đang liên tục phát sinh doanh số, những cửa hàng nào lâu rồi chưa phát sinh đơn hàng mới. Từ đó có thể sắp xếp, phân chia hợp lý các nhân viên bán hàng vào các tuyến bán hàng, thay đổi tần suất viếng thăm khách hàng cho phù hợp với thực tế, tránh lãng phí nguồn lực đồng thời có thể chăm sóc được kỹ hơn các cửa hàng trọng tâm.

Thứ tư, cập nhật thị trường là điểm nổi bật của phần mềm Distribution Management System. Giúp các doanh nghiệp phân phối có thể kiểm soát được hàng tồn tại từng điểm bán lẻ, đại lý nhằm đưa ra kế hoạch sản xuất, điều phối hàng hóa phù hợp.

2.1.3 Khảo sát thực tế việc sử dụng hệ thống quản lý phân phối

Khi chưa sử dụng Distribution Management System, có rất nhiều lỗ hổng mà các doanh nghiệp phân phối phải đối mặt như sau đây:

- **Hàng tồn kho:** Không kiểm soát được hàng tồn kho của nhà phân phối, cửa hàng để tối ưu hóa chuỗi cung ứng và các chiến lược tiêu thụ hàng. Công ty sẽ không có sự chủ động trong việc cung ứng và tiêu thụ hàng trên thị trường, đặc biệt là loại mặt hàng có thời hạn sử dụng ngắn, sẽ không kiểm soát được hàng đang ở đâu trong chuỗi cung ứng của mình.
- **Bao phủ:** Việc quản lý và tối ưu tuyến bán hàng theo từng ngành hàng, khu vực, đội bán hàng để gia tăng độ bao phủ cũng là một vấn đề.
- **Doanh số:** Doanh số luôn là vấn đề đau đầu của các nhà quản lý, họ sẽ không có đầy đủ thông tin để triển khai các chương trình bán hàng một cách có hiệu quả.

- **Hiệu suất:** Làm thế nào để biết đội ngũ bán hàng của công ty có hoạt động với hiệu suất 100% hoặc hơn không? Làm sao để thiết kế các KPI và kiểm soát hiệu quả giúp họ có động lực bán hàng, tăng hiệu suất làm việc? Đội ngũ nhân viên bán hàng có được theo sát và huấn luyện các kỹ năng bán hàng hay không?
- **Số ảo:** Tình trạng số ảo khá phổ biến trong bán hàng và phân phối, các số liệu ảo về điểm bán, doanh số, khuyến mãi, tồn kho, ... luôn là nỗi lo lắng của các nhà quản lý vì điều này ảnh hưởng lớn đến việc ra quyết định của họ.
- **Tích hợp:** Trong mô hình phân phối hiện đại và đa kênh, các công ty cung cấp, phân phối hàng hóa ra thị trường cũng sẽ đối diện với vấn đề tiếp xúc và xử lý thông tin với nhiều hệ thống, hình thức dữ liệu khác nhau. Làm thế nào để thiết kế mô hình dữ liệu chung đồng nhất nhưng vẫn đảm bảo các quy trình, điểm đặc thù của mỗi kênh, mỗi hệ thống, đảm bảo tính xuyên suốt, đồng nhất và kịp thời của thông tin.

Còn sau khi áp dụng giải pháp quản lý hệ thống phân phối (DMS) thì các nhà quản lý thu được:

- **Dữ liệu thật:** Với các quy trình và chức năng được thiết kế chặt chẽ và hỗ trợ hiệu quả các đối tượng sử dụng trong hệ thống sẽ giúp công ty có được bộ dữ liệu đầy đủ và thật từ dữ liệu thị trường, điểm bán, tuyến bán hàng đến tồn kho, sell-in, sell-out, khuyến mãi, trưng bày, ... hạn chế và ngăn chặn tối đa các trường hợp thay đổi dữ liệu bán hàng. Dữ liệu thật này sẽ giúp công ty ra quyết định và triển khai các chiến lược bán hàng hiệu quả và chính xác hơn.
- **Chủ động và làm chủ thị trường:** Doanh nghiệp hoàn toàn chủ động trong triển khai chiến lược bán hàng, chủ động khi có sự thay đổi, biến động về nhân sự trong hệ thống phân phối và dễ dàng thiết lập các nhà phân phối mới, khu vực mới cho các hệ thống bán hàng mới.
- **Gia tăng giá trị thương hiệu:** Việc triển khai và quản lý hiệu quả các chương trình bán hàng giúp thu hút và thúc đẩy tiêu thụ từ người tiêu dùng làm gia tăng sự trung thành từ người tiêu dùng cũng như sự hợp tác từ nhà phân phối vì họ thấy được lợi ích thật sự

Hiện nay trên thị trường có một số ít sản phẩm phần mềm quản lý hệ thống phân phối cho doanh nghiệp (DMS). Sử dụng phần mềm này thường là những người quản lý của doanh nghiệp, công ty, hãng phân phối. Do tính chất đặc thù như vậy mà những người tiếp cận và sử dụng phần mềm kiểu này khá ít, và phần mềm quản lý hệ thống phân phối DMS cũng không phổ biến rộng rãi cho mọi đối tượng. Một số phần mềm DMS có thể kể đến như Adaline, BS Silver, SSE, Perfect Warehouse, GM Sales, KiotViet, Suno, Sapo, ... Mỗi phần mềm sẽ cung cấp những tính năng đặc biệt riêng, tính phí hoặc không tính phí, dùng offline hoặc online, từng nền tảng (Windows, Android, iOS) và ưu nhược điểm khác nhau. Các phần mềm offline như Adaline, BS Silver, Perfect Warehouse, ... sẽ có ưu điểm là miễn phí, tốc độ nhanh, bảo mật cao. Tuy nhiên chỉ phù hợp với những cửa hàng bán lẻ, vừa và nhỏ, tính năng hạn chế, không có khả năng mở rộng. Sapo hay KiotViet là các phần mềm quản lý phân phối phổ biến hơn khi cung cấp cho người dùng nhiều tính năng với mức giá rẻ, giao diện làm việc trực tuyến, nhanh, dễ sử dụng, ... Các phần mềm kiểu này phù hợp với những công ty, doanh nghiệp phân phối lớn hơn khi khả năng nắm bắt biến động thị trường là cần thiết để đưa ra những chiến lược kinh doanh kịp thời.

Ở Việt Nam hiện nay, KiotViet là phần mềm quản lý bán hàng phổ biến nhất với hơn 100,000 cửa hàng đang sử dụng và hơn 5,000 cửa hàng mới mỗi tháng. Đơn giản, dễ dùng, tiết kiệm chi phí và phù hợp với hơn 15 ngành hàng khác nhau. KiotViet sẽ cung cấp cho người dùng một khung tính năng cố định ứng với từng ngành hàng khác nhau. Một phần mềm quản lý hoàn chỉnh có hai phần, một là trang quản lý, hai là trang bán hàng. Trang quản lý của KiotViet có các tính năng về quản lý hàng hóa (quản lý danh mục, thiết lập giá, ...), quản lý giao dịch (lên hóa đơn, đặt hàng, xuất hàng, ...), quản lý nhân viên, quản lý khách hàng và nhà cung cấp, xuất báo cáo (cuối ngày, theo tuần, theo tháng, ...). Trang bán hàng của KiotViet cung cấp một giao diện đơn giản tương tự như giao diện thanh toán bán hàng ở các cửa hàng bán lẻ hay siêu thị nhỏ. Đơn giản chỉ chọn sản phẩm, tính tiền, thanh toán và xuất hóa đơn cho khách.

Trong ứng dụng quản lý phân phối của chúng em, ngoài những tính năng cơ bản, cần thiết như quản lý sản phẩm, quản lý hàng trong kho, quản lý giao dịch, quản lý nhân viên, quản lý phân quyền động thì ứng dụng một tính năng mới. Đó là ứng dụng cung cấp chức năng

gợi ý, lên lịch trình viếng thăm cho các nhân viên bán hàng (salesman), lưu lại lịch sử check-in của nhân viên bán hàng. Nhân viên bán hàng là người hàng ngày sẽ thăm một số cửa hàng bán lẻ (rental outlet), kiểm tra lượng hàng hóa bán ra, từ đó lên hóa đơn để nhập hàng hóa mới về. Để các nhân viên bán hàng này hoạt động hiệu quả nhất, cần phân công họ vào các tuyến bán hàng phù hợp, để họ vừa có thể di chuyển một cách dễ dàng giữa các cửa hàng bán lẻ, vừa đến được nhiều cửa hàng nhất có thể. Chi tiết về tính năng mới này sẽ được trình bày trong chương sau của đề án.

2.2 Cơ sở lý thuyết

2.2.1 Role-Based Access Control

Trong an toàn thông tin, role-based access control (RBAC) là một cách tiếp cận cho vấn đề hạn chế quyền truy cập của người dùng. RBAC được sử dụng phổ biến tại các tổ chức lớn với nhiều hơn 500 nhân viên.

RBAC có nhiều mô hình khác nhau. Mô hình cơ bản là $RBAC_0$, mô hình có chứa quan hệ phân cấp là $RBAC_1$, có chứa ràng buộc $RBAC_2$ và mô hình tăng cường (consolidated) $RBAC_3$ [3].

Các thành phần chính của RBAC bao gồm user (người dùng), role (vai trò), permission (quyền) và session (phiên truy cập) cùng với quan hệ giữa chúng nhằm đơn giản hóa quá trình gán quyền cho người dùng.

User hay người dùng trong mô hình RBAC là con người, tuy nhiên khái niệm này có thể được tổng quát hóa bằng việc bao gồm cả những tác nhân tự động như robot hay phần mềm khác. Một role là một chức năng công việc hay chức vụ trong tổ chức đi kèm với các thẩm quyền mà nhân viên trong chức vụ đó nắm giữ. Còn một permission là sự một chấp nhận cho một phương thức truy cập vào một hay nhiều các đối tượng trong hệ thống. Việc một permission tương ứng với gì thì tùy thuộc vào hệ thống implement ra sao mô hình RBAC.

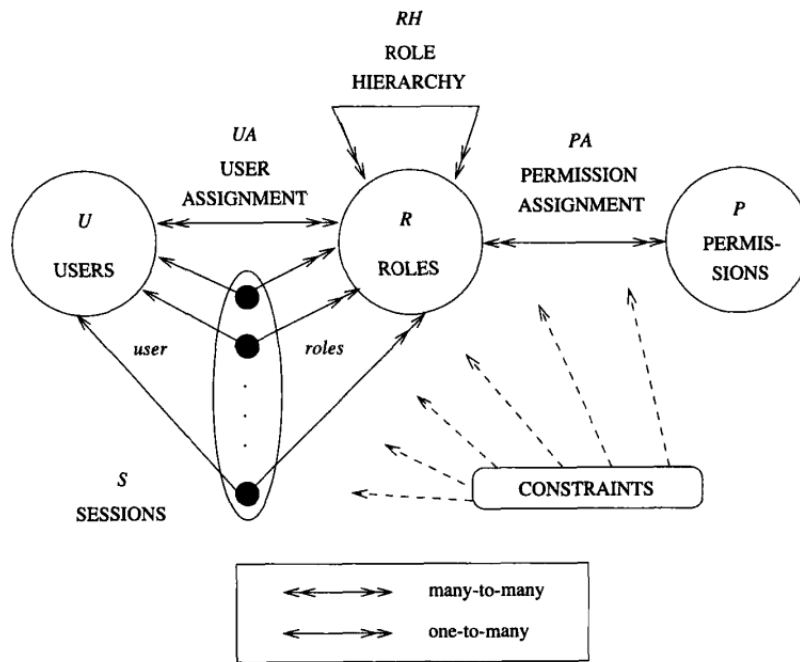
Mỗi một session là một ánh xạ từ mỗi phần tử của tập user sang tập con của tập các role. Session thiết lập một phiên truy cập với tập các role được kích hoạt (là một tập con của

tập các role mà người dùng đó được gán). Mỗi một phiên truy cập tương ứng cho một user, trong RBAC₀ sự tương ứng này là không thay đổi theo thời gian.

Trong RBAC, permission được gán với một hoặc nhiều role, người dùng hay nhân viên được gán với một hay nhiều role. Người dùng trong RBAC không được gán trực tiếp các permission mà phải thông qua role, từ đó đơn giản hóa quá trình gán permission hay việc chuyển phòng ban công tác của nhân viên.

Trong đồ án này chỉ sử dụng mô hình điều khiển truy cập dựa trên RBAC₀, biểu diễn toán học của RBAC₀ như sau:

- Tập U , R , P và S tương ứng với tập hợp của các user, role, permission và session.
- Tập $PA \subseteq P \times R$ biểu diễn quan hệ nhiều nhiều giữa permission và role.
- Tập $UA \subseteq U \times R$ biểu diễn quan hệ nhiều nhiều giữa user và role.
- Ánh xạ $user : S \rightarrow U$ gán mỗi một session s cho một user $user(s)$.
- Ánh xạ $roles : S \rightarrow \mathcal{P}(R)$ gán mỗi một session s cho tập con của các role: $roles(s) \subseteq \{r \mid (user(s), r) \in UA\}$ (có thể thay đổi theo thời gian) và session sẽ có các permission là $\bigcup_{r \in roles(s)} \{p \mid (p, r) \in PA\}$



Hình 2.1: Các mô hình RBAC

2.2.2 Thuật toán phân cụm dữ liệu

2.2.2.1 Giới thiệu về bài toán phân cụm dữ liệu

Phân cụm là kỹ thuật quan trọng trong xử lý dữ liệu, nó thuộc lớp các phương pháp học không giám sát (Unsupervised Learning) trong học máy (Machine Learning). Giải thích dễ hiểu hơn thì phân cụm là các quy trình tìm cách nhóm các đối tượng đã cho vào các cụm (cluster), sao cho các đối tượng trong cùng một cụm tương tự nhau (similar) và các đối tượng khác cụm thì không tương tự nhau (dissimilar).

Mục đích của phân cụm là tìm ra bản chất bên trong các nhóm của dữ liệu và có thể áp dụng trong rất nhiều lĩnh vực như:

- Trong marketing, xác định các nhóm khách hàng (khách hàng tiềm năng, khách hàng giá trị, phân loại và dự đoán hành vi khách hàng, ...) sử dụng sản phẩm hay dịch vụ của công ty để giúp công ty có chiến lược kinh doanh hiệu quả hơn.
- Trong sinh học, phân nhóm động vật và thực vật dựa vào các thuộc tính của chúng.
- Trong bảo hiểm, tài chính: phân nhóm các đối tượng sử dụng bảo hiểm và các dịch vụ tài chính, dự đoán xu hướng của khách hàng, phát hiện gian lận tài chính, ...

- Trong phân tích dữ liệu web: phân loại tài liệu, phân loại người dùng web, ...

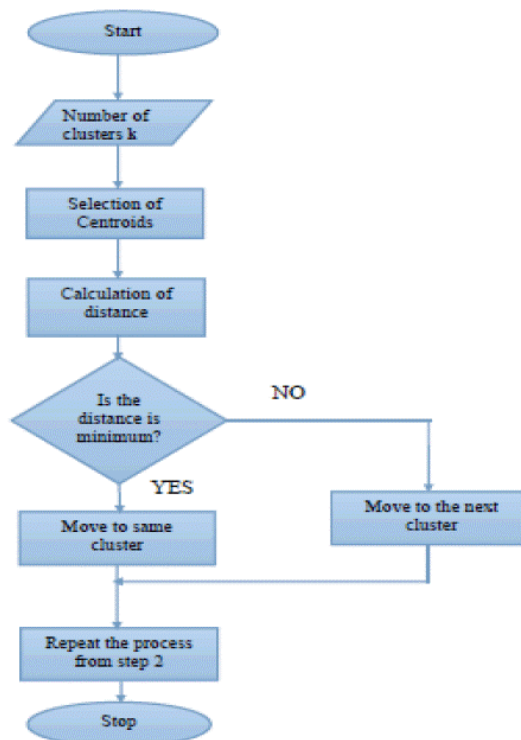
2.2.2.2 Thuật toán K-Means

K-Means là thuật toán quan trọng và được sử dụng phổ biến trong kỹ thuật phân cụm. Tư tưởng chính của thuật toán K-Means là tìm cách phân nhóm các đối tượng đã cho vào K cụm (K là số cụm được xác định trước, K nguyên dương) sao cho khoảng cách từ các đối tượng đến tâm nhóm (centroid) là nhỏ nhất.

Phát biểu bài toán: Cho N điểm trên không gian d chiều. Làm thế nào để phân chia thành K nhóm mà các điểm trong một nhóm có khoảng cách gần trọng tâm của nhóm hơn so với khoảng cách đến trọng tâm của bất kỳ 1 nhóm nào khác.

Đầu vào, đầu ra của thuật toán:

- Đầu vào: Cho N điểm, mỗi điểm có dạng (x, y) . K là số nhóm (cụm) ($K \leq N$).
- Đầu ra: Danh sách K nhóm là các điểm của mỗi nhóm.



Hình 2.2: Biểu đồ khối mô tả thuật toán K-Means

Chi tiết các bước thực hiện thuật toán K-Means:

- Bước 1: Chọn ngẫu nhiên K tâm (centroid) cho K cụm (cluster). Mỗi cụm được đại diện bằng các tâm của cụm.
- Bước 2: Tính khoảng cách giữa các điểm đến tâm (c_x, c_y) của từng nhóm. Dùng khoảng cách Euclidean: $d(a, b) = \sqrt{(a_x - b_x)^2 + (a_y - b_y)^2}$
- Bước 3: Nhóm các điểm vào nhóm có khoảng cách từ tâm đến điểm đó nhỏ nhất.
- Bước 4: Xác định lại tâm mới cho các nhóm, sử dụng công thức:

$$c_x = \frac{a_{1x} + a_{2x} + \dots + a_{mx}}{m}$$

$$c_y = \frac{a_{1y} + a_{2y} + \dots + a_{my}}{m}$$

với m số phần tử của nhóm.

- Bước 5: Nếu (c_x, c_y) không đổi thì đó là tâm cần tìm. Nếu không thì lặp lại bước 2.

Thuật toán này được áp dụng vào chức năng gợi ý phân tuyến bán hàng cho nhân viên bán hàng của người quản lý tuyến. Đầu vào của bài toán là N điểm (tương ứng với N cửa hàng bán lẻ) và K cụm (do người quản lý thiết lập). Cần chia N điểm này cho K cụm một cách cân bằng và co cụm nhất. Cụ thể về chức năng này sẽ trình bày ở chương 4 của đề án.

2.3 Các công nghệ sử dụng

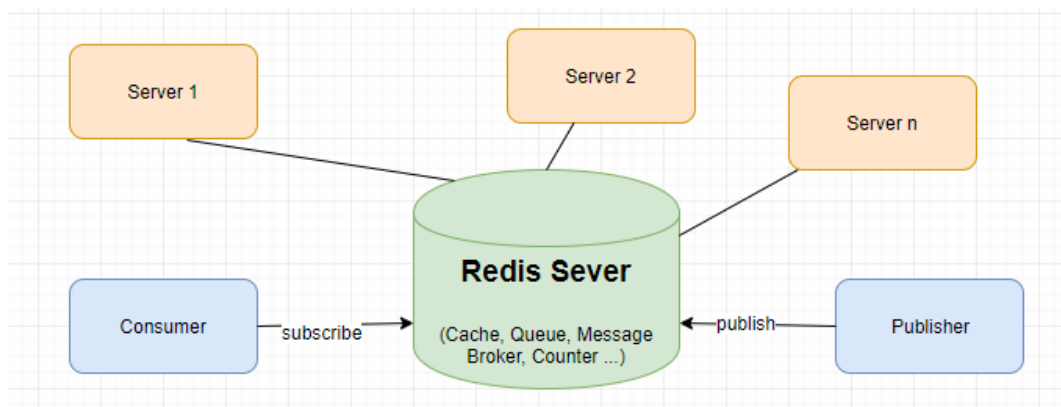
2.3.1 Công nghệ lưu trữ - Redis

Redis là viết tắt của Remote Dictionary Server (máy chủ từ điển từ xa), lưu trữ dữ liệu dưới dạng KEY-VALUE trong bộ nhớ. Là phần mềm mã nguồn mở có tốc độ truy cập nhanh để dùng làm cơ sở dữ liệu đơn giản, bộ nhớ đệm (cache), trình chuyển tiếp (broker) tin nhắn hoặc được sử dụng làm danh sách tác vụ chờ xử lý (queue).

Redis hiện cung cấp thời gian phản hồi ở tốc độ chưa đến một mili giây, giúp thực hiện hàng triệu yêu cầu mỗi giây cho các ứng dụng thời gian thực trong lĩnh vực Trò chơi, Quảng cáo, Dịch vụ tài chính, Chăm sóc sức khỏe, IoT, ... Cụ thể, Redis thường được chọn sử dụng

cho các hoạt động lưu trữ bộ nhớ đệm, quản lý phiên, trò chơi, bảng xếp hạng, phân tích thời gian thực, dữ liệu không gian địa lý, ứng dụng đặt xe, trò chuyện / nhắn tin, phát trực tiếp nội dung đa phương tiện, ...

Redis là một cơ sở dữ liệu NoSQL. NoSQL là một dạng cơ sở dữ liệu phi quan hệ, sử dụng nhiều loại mô hình dữ liệu đa dạng để truy cập và quản lý dữ liệu trong bộ nhớ và tìm kiếm, NoSQL được tối ưu hóa dành riêng cho các ứng dụng yêu cầu mô hình dữ liệu linh hoạt có lượng dữ liệu lớn và độ trễ thấp, đạt được bằng cách giảm bớt một số hạn chế về tính nhất quán của dữ liệu.



Hình 2.3: Redis server

Cách thức hoạt động của Redis:


Toàn bộ dữ liệu của Redis nằm trong bộ nhớ (RAM), khác với những loại cơ sở dữ liệu thông thường khác lưu dữ liệu trên ổ đĩa hoặc ổ SSD. Phần lớn các tác vụ trên cơ sở dữ liệu truyền thống đều yêu cầu truy cập qua lại tới ổ đĩa, do vậy sẽ tốn thời gian tìm kiếm, dữ liệu của Redis nằm trên RAM nên sẽ không mất thời gian này. Do đó Redis có thể hỗ trợ thêm khá nhiều tác vụ và có thời gian phản hồi nhanh hơn. Hiệu suất của Redis rất tốt với các tác vụ đọc ghi thông thường mất chưa đầy một mili giây và hỗ trợ hàng triệu tác vụ mỗi giây.

Khác với các cơ sở dữ liệu quan hệ như MySQL hay PostgreSQL, Redis không có bảng. Redis lưu dữ liệu dưới dạng KEY-VALUE và hỗ trợ nhiều cấu trúc dữ liệu cơ bản như hash, list, set, sorted set, string, ... Bên cạnh đó Redis có 2 background threads chuyên làm nhiệm vụ định kỳ ghi dữ liệu trên đĩa cứng, cơ chế backup này giúp cho Redis có độ bảo mật và sửa

lỗi cao.

Một số ứng dụng phổ biến của Redis:

- Lưu trữ bộ nhớ đệm (caching): Sử dụng bộ nhớ đệm để giảm độ trễ khi truy cập dữ liệu, tăng năng suất và giảm tải cho cơ sở dữ liệu và ứng dụng của bạn. Redis có thể phục vụ những dữ liệu thường xuyên được yêu cầu với thời gian phản hồi chưa đến một mili giây và cho phép dễ dàng thay đổi quy mô nhằm đáp ứng mức tải cao hơn mà không cần tốn kém chi phí vào back-end.
- Bảng xếp hạng game: Redis là giải pháp hay được các nhà phát triển game dùng để xây dựng bảng xếp hạng theo thời gian thực (real-time leaderboard). Sử dụng cấu trúc dữ liệu Sorted Set của Redis, cấu trúc dữ liệu này đảm bảo tính duy nhất của các thành phần trong khi vẫn duy trì danh sách được sắp xếp theo điểm số của người dùng. Danh sách cập nhật mỗi khi điểm số người dùng thay đổi.



POSITION	RACER	TIME
#1	Isaac Clark	00:14.2
#2	Nick Serfassanti	00:15.0
#3	Shubham Kacker	00:17.0
#4	NguyenXV	00:17.8
#5	Madhu Cheluvajulu	00:19.0
#6	Hemk	00:19.0
#7	Prathig2	00:19.0
#8	Saikat Sengupta	00:19.1
#9	Freddie	00:19.4
#10	Si Chen	00:20.0
#11	Jeremy T	00:20.2
#12	J Hilar	00:20.5
#13	Srinath	00:21.0

Hình 2.4: Ứng dụng Redis trong bảng xếp hạng game

- Lưu trữ phiên (session): Các nhà phát triển ứng dụng thường sử dụng Redis để lưu trữ, quản lý phiên cho các ứng dụng quy mô Internet. Quản lý dữ liệu phiên chẳng hạn như hồ sơ người dùng, thông tin xác thực đăng nhập (token), trạng thái phiên, ...
- Trò chuyện, nhắn tin, hàng chờ xử lý tác vụ: Redis hỗ trợ Pub/Sub (là cấu trúc gửi – nhận tin nhắn mà người gửi và người nhận không biết nhau) với nhiều cấu trúc dữ liệu

như list, sorted set, hash. Điều này cho phép Redis hỗ trợ những chat rooms hiệu năng cao, luồng tin nhắn theo thời gian thực.

So sánh Redis và một loại cơ sở dữ liệu quan hệ thông thường (MySQL)

	Redis	MySQL
Cấu trúc cơ sở dữ liệu	Lưu trữ dạng KEY-VALUE Lưu dữ liệu trong RAM, là máy chủ cấu trúc dữ liệu vì các key có thể chứa string, hash, list, set và sorted set	Lưu trữ dạng bảng MySQL cung cấp một máy chủ cơ sở dữ liệu quan hệ SQL rất nhanh, đa luồng, đa người dùng, mạnh mẽ
Ưu điểm	<ul style="list-style-type: none"> • Dễ cài đặt, sử dụng, deploy, maintain, ... • Lưu trữ dữ liệu trong bộ nhớ nên cho hiệu năng cao và tốc độ nhanh • Mã nguồn mở, ổn định, chi phí hiệu quả • Khả năng mở rộng cao, hỗ trợ sao lưu vào đĩa cứng • Cấu trúc dữ liệu đa dạng 	<ul style="list-style-type: none"> • Cơ sở dữ liệu quan hệ mã nguồn mở được sử dụng rộng rãi • Dễ sử dụng, khả năng tương thích cao, hỗ trợ đa nền tảng, cộng đồng phát triển mạnh mẽ • Hỗ trợ index và full-text searching
Trường hợp nên sử dụng	<ul style="list-style-type: none"> • Có dữ liệu dạng KEY-VALUE • Cần lưu cache • Cần hiệu năng cao • Kích thước dữ liệu ổn định 	<ul style="list-style-type: none"> • Cần cơ sở dữ liệu quan hệ • Khi có các hoạt động phân tán • Cần bảo mật cao và hoạt động đơn giản • Không sử dụng khi dữ liệu lớn dần và không thể cache hết lên bộ nhớ
Nhược điểm	<ul style="list-style-type: none"> • Redis sử dụng RAM nên khi lượng file cache lớn sẽ dẫn đến thiếu RAM cho server • Không thể truy vấn trực tiếp các object 	<ul style="list-style-type: none"> • Nhiều vấn đề về lưu trữ procedure và trigger

Bảng 2.1: So sánh Redis và MySQL

2.3.2 Công nghệ lưu trữ - PostgreSQL

2.3.2.1 Tổng quan về PostgreSQL

PostgreSQL, còn được gọi là Postgres, là một hệ thống quản lý cơ sở dữ liệu quan hệ miễn phí mã nguồn mở chú trọng vào tính mở rộng và tính tương thích với chuẩn SQL. Ban đầu có tên là POSTGRES, có nguồn gốc từ cơ sở dữ liệu Ingress được phát triển tại

Đại học California, Berkeley. Vào năm 1996, dự án được đổi tên thành PostgreSQL. PostgreSQL được viết bằng ngôn ngữ C và có thể chạy trên nhiều hệ điều hành như Linux, macOS, Windows, ... Đồng thời hỗ trợ các tính năng tương tự như nhiều các cơ sở dữ liệu quan hệ khác. Phiên bản PostgreSQL sử dụng trong đề án này là 11.8.

2.3.2.2 Tổng quan về Index

Index trong cơ sở dữ liệu là một cấu trúc dữ liệu giúp cải thiện tốc độ truy vấn dữ liệu trên các bảng với đánh đổi là sử dụng thêm không gian lưu trữ nhằm duy trì cấu trúc dữ liệu index cũng như làm tăng thời gian ghi dữ liệu. Index được sử dụng để nhanh chóng định vị dữ liệu mà không phải tìm duyệt qua toàn bộ các hàng trong bảng của cơ sở dữ liệu mỗi khi bảng đó được truy cập. Index có thể được tạo sử dụng một hay nhiều cột trong cùng một bảng, cung cấp cơ chế tăng tốc quá trình tra cứu ngẫu nhiên và truy cập có thứ tự các bản ghi trong cơ sở dữ liệu.

Index được chia làm hai loại kiến trúc: Clustered Index và Non-clustered Index.

Với clustered index, các hàng trong cơ sở dữ liệu được lưu trữ trên đĩa theo cùng thứ tự với index. Do đó mỗi bảng chỉ tồn tại duy nhất một clustered index. Với non-clustered index thì cấu trúc dữ liệu index là cấu trúc dữ liệu nằm ngoài bảng, các hàng trong bảng không được đảm bảo thứ tự giống như thứ tự của index. Một bảng có thể có rất nhiều non-clustered index. Thông thường với truy cập trên clustered index sẽ nhanh hơn trên non-clustered index vì dữ liệu sẽ không bị phân mảnh ở nhiều vị trí trong ổ cứng.

2.3.2.3 Index trong PostgreSQL

Trong PostgreSQL, clustered index không được hỗ trợ. Tuy nhiên PostgreSQL lại hỗ trợ nhiều kiểu index như B-Tree, Hash, GiST, SP-GiST, GIN và BRIN. Mỗi một kiểu index sử dụng một thuật toán riêng biệt để tối ưu cho nhiều loại câu truy vấn khác nhau. Mặc định khi sử dụng lệnh CREATE INDEX thì index B-Tree sẽ được tạo, index này cũng phù hợp với hầu hết các trường hợp truy vấn.

Index kiểu B-Tree có thể được sử dụng cho các câu truy vấn liên quan đến tìm kiếm bằng, tìm kiếm khoảng hoặc trong trường hợp dữ liệu được sắp xếp theo một trường nào đó.

Trình lập kế hoạch cho các truy vấn của PostgreSQL sẽ xem xét sử dụng index B-Tree bất kì khi nào trường mà có index nằm trong một trong các phép toán so sánh như $<$ $<=$ $=$ $>=$ $>$.

Các toán tử tương đương hoặc tổ hợp của chúng, như là phép toán BETWEEN và IN, cũng có thể sử dụng được index B-Tree. Ngoài ra các phép toán liên quan đến NULL như IS NULL hay IS NOT NULL cũng có thể được tăng tốc bằng index B-Tree.

Trình tối ưu cũng có thể sử dụng index B-Tree cho các câu truy vấn sử dụng toán tử pattern matching như LIKE và nếu mà pattern được sử dụng bắt đầu là một xâu hằng số như 'foo%' nhưng không phải là '%bar'.

B-Tree có thể cũng được sử dụng để lấy dữ liệu được sắp xếp theo thứ tự của cột được index.

Index kiểu Hash chỉ có thể được sử dụng trong các câu truy vấn sử dụng toán tử so sánh bằng. Để tạo index Hash trong PostgreSQL, ta sử dụng:

```
1 CREATE INDEX name ON table USING HASH (column);
```

Listing 2.1: Tạo index sử dụng Hash

2.3.2.4 Full Text Search trên PostgreSQL

2.3.2.4.1 Cơ bản về Full Text Search

Các toán tử tìm kiếm text trên các CSDL quan hệ đã tồn tại từ rất lâu như LIKE hay ILIKE. Nhưng những toán tử đó thiếu đi những tính chất mà cần có trong các hệ thống thông tin hiện đại như:

- Hỗ trợ ngôn ngữ, thậm chí với cả tiếng Anh. Các toán tử trên không phù hợp cho việc xử lý những từ gần giống nghĩa hoặc được suy diễn, như satisfy và satisfies trong tiếng Anh. Có thể dùng toán tử logic OR để tìm kiếm cả 2 nhưng phải liệt kê rất nhiều trường hợp và một số từ có thể có rất nhiều từ được suy diễn hoặc gần nghĩa.
- Không hỗ trợ xếp hạng (ranking) trên tập kết quả, dẫn đến tìm kiếm không hiệu quả khi mà có thể có tới rất nhiều document được tìm thấy trong một câu truy vấn.

- Các phép toán đó thường chậm vì không hỗ trợ đánh chỉ số, mỗi một câu truy vấn phải duyệt qua toàn bộ các document.

PostgreSQL hỗ trợ tiền xử lý các document và đánh chỉ số để tăng tốc các câu truy vấn.

Tiền xử lý document bao gồm việc:

- Phân tích document thành các token.
- Chuyển các token thành các lexeme.
- Lưu trữ các dữ liệu sau khi tiền xử lý document để tối ưu cho truy vấn.

Một document trong PostgreSQL thông thường là một cột chứa xâu của bảng CSDL, hoặc có thể là một tổ hợp (thông qua phép ghép nối xâu) của những trường đó trong một bảng hay trong nhiều bảng hoặc có thể được truyền từ ngoài vào.

Một token là một xâu biểu diễn một từ được phân tích ra từ document.

Một lexeme cũng là một xâu, tương tự như token, nhưng đã được chuẩn hóa (normalize) sao cho những dạng khác nhau của một từ trở thành giống nhau. Việc chuẩn hóa luôn luôn thực hiện chuyển tất cả các kí tự thành lower-case, và thông thường có cắt bỏ một vài các hậu tố như e hoặc es trong tiếng Anh. Đồng thời ở bước chuyển từ token thành càn lexeme cũng thông thường loại bỏ những stop word, là những từ rất phổ biến trở thành vô nghĩa khi tìm kiếm, PostgreSQL sử dụng Dictionary để thực thi bước này.

Dictionary trong PostgreSQL cho phép hiệu chỉnh quá trình chuẩn hóa thành lexeme, bao gồm:

- Định nghĩa các stop word được loại bỏ.
- Ánh xạ các từ đồng nghĩa thành một từ.
- Ánh xạ cụm từ thành một từ.
- Ánh xạ các biến thể của một từ thành dạng chuẩn.

PostgreSQL sử dụng kiểu dữ liệu `tsvector` để lưu trữ dữ liệu sau khi tiền xử lý document, đồng thời sử dụng kiểu `tsquery` để biểu diễn các câu truy vấn.

Để chuyển từ một `xâu` sang kiểu `tsvector` ta sử dụng hàm `to_tsvector`. Còn để chuyển một `xâu` bất kì sang một `tsquery`, ta sử dụng `plainto_tsquery`.

Để thực hiện Full Text Search một tsvector với một tsquery, ta sử dụng toán tử @@.

2.3.2.4.2 Full Text Search cho tiếng Việt

Để thực hiện Full Text Search với tiếng Việt, trong đề án này sử dụng một hàm được định nghĩa bằng SQL như sau:

```
CREATE OR REPLACE FUNCTION vn_unaccent(TEXT)
RETURNS TEXT AS $$
    SELECT LOWER(TRANSLATE($1,
        ' 1 2 3 AAAA AAAA AAAA AA Eaaa aaaaaaa aa aa aa Aa AA AAAA aa aa aa aa Aa C'
        ' 123AAAAAAAAAAAAAAaaaaaaaaaaaaaaAaAAAAAaaaaaaaAaC'
    ));
$$ LANGUAGE sql IMMUTABLE;
```

Khi đó, thay vì sử dụng trực tiếp `to_tsvector` và `plainto_tsquery`, ta sử dụng `to_tsvector(vn_unaccent(text))` và `plainto_tsquery(vn_unaccent(text))`.

2.3.2.4.3 Đánh chỉ số cho Full Text Search

Để sử dụng Index cho Full Text Search, trước tiên ta cần tạo một cột trong CSDL lưu trữ kiểu dữ liệu tsvector. Có hai loại Index hay được sử dụng cho Full Text Search là GIN và GiST.

Để Index sử dụng GIN, cú pháp như sau:

```
1 CREATE INDEX idx_textsearch ON sometable USING GIN (tsvector_column);
```

Listing 2.2: Tao index sử dụng GIN

Còn nếu sử dụng GiST:

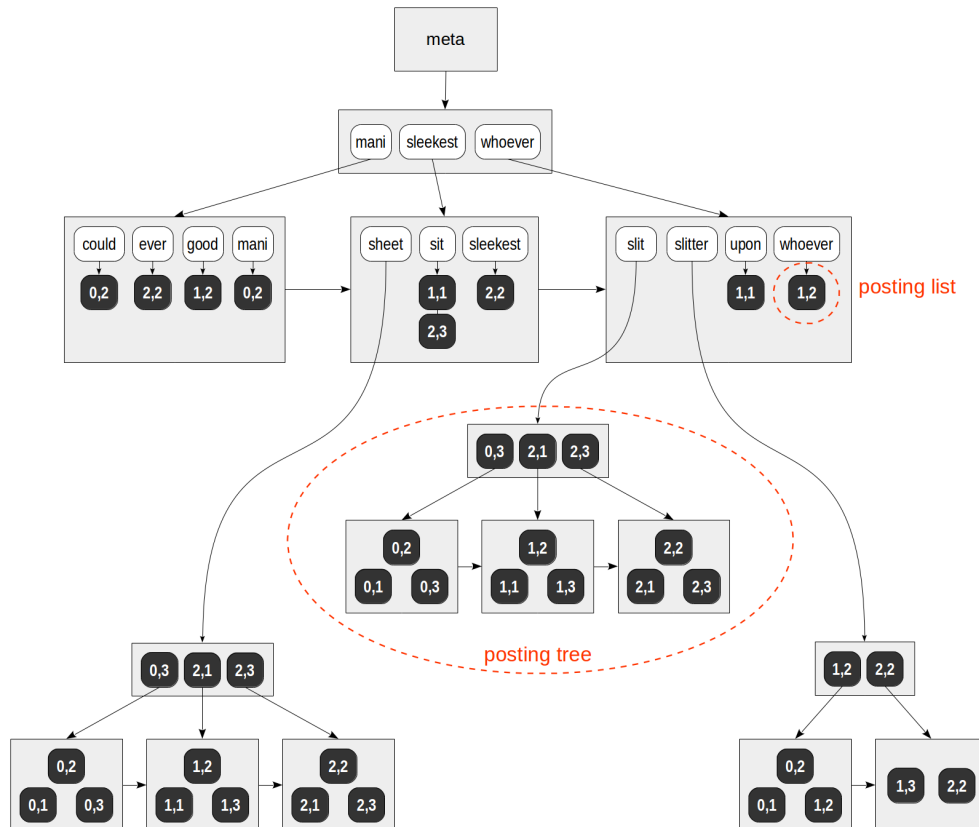
```
1 CREATE INDEX idx_textsearch ON sometable USING GIST (tsvector_column);
```

Listing 2.3: Tao index sử dụng GiST

Khi đó toán tử @@ có thể sử dụng Index để tăng tốc câu truy vấn. Trong PostgreSQL, Index Full Text Search thông thường sử dụng GIN hơn sử dụng GiST.

Thông thường sẽ sử dụng trigger để cập nhật các trường tsvector trong CSDL khi chèn hoặc chỉnh sửa.

2.3.2.4.4 GIN



Hình 2.5: Cấu trúc của GIN

GIN là viết tắt của Generalized Inverted Index. GIN được thiết kế cho việc đánh chỉ số những giá trị tổng hợp (composite value), và cho phép tìm kiếm cho các phần tử xuất hiện trong các composite value.

Nếu quy ước một composite value là một item, còn phần tử trong composite value là một key thì GIN luôn luôn lưu trữ và tìm kiếm trên key chứ không phải trên item. Một GIN index lưu trữ tập các cặp (key, posting list) trong đó posting list là tập các ID của các hàng trong bảng của CSDL. Một ID có thể được nằm trong nhiều các posting list vì một item có thể có

nhiều các key. Mỗi một key được lưu trữ một lần duy nhất, do đó nên GIN rất phù hợp trong trường hợp một key xuất hiện nhiều lần. posting list được lưu với thứ tự từ nhỏ đến lớn các ID của các hàng. Nếu posting list quá lớn sẽ được chuyển thành posting tree, cấu trúc tương tự như B-Tree.

2.3.3 Công nghệ back-end

2.3.3.1 Giới thiệu về ngôn ngữ Go

Go là ngôn ngữ biên dịch kiểu tĩnh được tạo ra tại Google. Cú pháp của Go tương tự như của ngôn ngữ C, đồng thời cũng bị ảnh hưởng lớn bởi ngôn ngữ này. Ngôn ngữ Go cũng thông thường được gọi với tên là Golang. Sau đây là một số đặc điểm của ngôn ngữ Go:

- Cú pháp và môi trường lập trình của Go có tính tương đồng với các ngôn ngữ lập trình kiểu động (dynamic language) như:
 - Rút gọn khai báo biến, không cần phải khai báo kiểu thông qua cơ chế type inference.
 - Biên dịch nhanh.
 - Quản lý các gói thư viện từ xa và các package có document truy cập online.
- Sử dụng cách tiếp cận đặc biệt với một số vấn đề:
 - Xây dựng sẵn trong ngôn ngữ một số cơ chế cho lập trình đồng bộ như: tiến trình nhẹ (light-weight process, hay còn gọi là goroutine), channel và câu lệnh select.
 - Toolchain mặc định khi biên dịch sinh ra file thực thi sẽ không chứa phụ thuộc ngoài.
- Ngôn ngữ được thiết kế đủ đơn giản để lập trình viên có thể dễ dàng hiểu và ghi nhớ.

2.3.3.2 Goroutine, Chanel và khối lệnh Select

Ngôn ngữ Go được xây dựng sẵn trong ngôn ngữ những cơ chế hỗ trợ việc lập trình đồng bộ. Đồng bộ ở đây không chỉ là song song ở mức CPU mà đồng thời cả việc sử dụng Asynchronous I/O cho phép các câu lệnh như truy cập database hay đọc gói tin từ mạng chạy trong khi tiến trình vẫn được sử dụng cho công việc khác. Kỹ thuật này phổ biến trong các server mà sử dụng event-based.

Trung tâm của xây dựng chương trình đồng bộ trong Go là tiến trình nhẹ (light-weight process) gọi là Goroutine. Một lời gọi hàm mà được đặt sau từ khóa go sẽ bắt đầu hàm đó trong một Goroutine mới. Trong đặc tả ngôn ngữ (Language Specification) không chỉ định Goroutine được cài đặt như nào nhưng với cài đặt hiện tại thì các Goroutine sẽ được phân bổ vào một tập nhỏ các luồng ở mức hệ điều hành, tương tự như cơ chế phân phát tiến trình trong ngôn ngữ Erlang. Khi chương trình Go mới bắt đầu sẽ chứa duy nhất một goroutine gọi là main goroutine.

Chương trình viết bằng Go sẽ thường sử dụng Channel, một cơ chế để gửi các thông điệp (Message) giữa các goroutine. Channel trong Go có thể có chứa Buffer hoặc không. Nếu không chứa Buffer thì Channel mỗi một thời điểm chỉ chứa tối đa một Message, các lời gọi gửi message vào channel tiếp theo đó sẽ bị block. Nếu chứa Buffer thì kích thước của buffer cũng bị giới hạn, các message được lưu trữ và truy xuất theo cơ chế FIFO. Khi đọc từ một channel mà channel chưa chứa message nào cả thì goroutine đọc từ channel đó sẽ bị block.

Channel trong Go có chứa kiểu tĩnh, nghĩa là một channel có kiểu là chan T sẽ chỉ có thể gửi và nhận message thuộc kiểu T. Go có chứa cú pháp đặc biệt để tương tác với channel:

- `x <- ch` để đọc từ channel ch vào biến x.
- `ch <- x` để gửi message x vào channel ch.

Trong trường hợp tương tác với nhiều lệnh nhận hoặc gửi message với các channel, chương trình Go có thể sử dụng khối lệnh select để lựa chọn lệnh nhận/gửi đầu tiên mà kết thúc block trên các channel. Cú pháp của khối lệnh select trong Go tương tự với switch nhưng khác là mỗi một nhãn case trong khối lệnh select sẽ là một lệnh tương tác với channel.

2.3.3.3 Các công cụ tích hợp

Bản cài đặt bộ công cụ của Go bao gồm nhiều các công cụ liên quan đến xây dựng, kiểm thử và phân tích code, gồm:

- **go build**, để xuất file nhị phân từ các file mã nguồn.
- **go test**, dùng để chạy các unit test và các benchmark.

- **go fmt**, dùng để format code.
- **go get**, tải xuống và cài đặt các thư viện và các file thực thi đi kèm.
- **go vet**, một static analyzer cho việc tìm kiếm các vị trí có thể có lỗi trong code.
- **go run**, một shortcut cho phép biên dịch và chạy chương trình.
- **go doc**, dùng để hiển thị document của thư viện.
- **go mod**, dùng để quản lý module - xuất hiện từ phiên bản Go 1.11.
- **go generate**, dùng để gọi code generator.

2.3.4 Docker

2.3.4.1 Công nghệ container

Công nghệ container được sinh ra có mục đích tương tự như công nghệ máy ảo (Virtual Machine) [2]. Điểm khác biệt lớn nhất là mỗi container sẽ không yêu cầu phải chạy một hệ điều hành đầy đủ và riêng biệt. Tất cả các container trên một máy chủ đều chia sẻ chung một hệ điều hành. Chính điểm khác biệt này làm cho container sử dụng ít tài nguyên CPU, RAM và bộ nhớ lưu trữ hơn VM rất nhiều dẫn đến tiết kiệm chi phí cho việc chạy server, và lỗi hệ điều hành và các vấn đề bảo trì hệ thống khác.

Công nghệ Container hiện đại được bắt đầu từ hệ điều hành Linux. Một số các công nghệ của nhân Linux giúp tạo nên sự phát triển của công nghệ container bao gồm: kernel namespace, control group, union filesystem. Dù cho các công nghệ này đã tồn tại từ lâu nhưng để sử dụng nó thì rất phức tạp và ngoài tầm với của phần lớn các công ty công nghệ. Sự thay đổi bắt đầu khi Docker được tạo ra giúp đơn giản hóa rất nhiều quá trình tạo và quản lý các container.

Docker ban đầu được tạo nên chỉ để chạy trên môi trường Linux. Tuy nhiên, trong một vài năm gần đây, Microsoft đã phát triển công nghệ container trên hệ điều hành Windows 10 và Windows Server từ đó giúp việc sử dụng Docker trên Windows tương tự như trên Linux. Nhưng bởi vì container sẽ chia sẻ hệ điều hành với hệ điều hành mà Docker chạy trên (được gọi

là host), từ đó dẫn đến một container dùng Linux sẽ không thể chạy trên Windows và ngược lại.

2.3.4.2 Sơ lược về Docker

Docker được bắt đầu phát triển bởi Docker, Inc vào năm 2013, bao gồm các thành phần:

- Software: gồm một chương trình chạy ngầm (daemon) gọi là dockerd, là chương trình quản lý các container và xử lý các container object. Daemon dockerd lắng nghe các yêu cầu từ bên ngoài gửi đến thông qua Docker Engine API. Đi kèm là một chương trình client, gọi là docker, cung cấp một giao diện dòng lệnh giúp tương tác với daemon dockerd.
- Object: các Docker Object là các thực thể cấu thành lên một ứng dụng chạy trên Docker. Các Docker object được chia thành ba loại là image, container và service.
 - Một Docker image là một tập các file chỉ đọc dùng làm template để tạo nên các container. Các image được dùng để lưu trữ và triển khai các ứng dụng.
 - Một Docker container là một môi trường được đóng gói và chuẩn hóa để chạy ứng dụng.
 - Một Docker service là một thực thể dùng để chạy nhiều container trên nhiều máy trong một cluster. Tập các máy chạy Docker trong một cluster gọi là một Swarm.

2.3.4.3 Dockerfile

Để xây dựng các image cho việc đóng gói ứng dụng, ta có thể chạy container, cài đặt các file cần thiết và sử dụng lệnh **docker commit** để lưu container thành image. Tuy nhiên thông thường việc tạo image thông qua việc sử dụng Dockerfile.

Dockerfile là một file văn bản có chứa những instruction(chỉ thị) để xây dựng một image. Trong đó có các lệnh như: chỉ định image gốc nào mà sẽ được xây dựng lên, chỉ định lệnh nào được chạy khi bắt đầu một container, chỉ định những file hay thư mục nào được sao chép từ host vào container, Để xây dựng lên image bằng Dockerfile, ta sử dụng docker build.

Ví dụ trong đồ án này sử dụng Dockerfile sau để xây dựng lên Webserver viết bằng Go:

```
1 FROM golang:1.14.3-alpine3.11 as builder
2 WORKDIR /go/baseweb/
3 COPY . .
4 RUN go build
5
6 FROM alpine:3.11
7 WORKDIR /go/
8 COPY --from=builder /go/baseweb/baseweb .
9 EXPOSE 8080
10 CMD ["/baseweb"]
```

Listing 2.4: Xây dựng image của back-end server

2.3.4.4 Registry & Docker Hub

Registry là một kho lưu trữ các image sử dụng để tải xuống image và tạo container từ đó. Registry có thể là một private server trong một cluster, hay một dịch vụ trên cloud (như Azure Container Registry) hay phổ biến và mặc định của Docker là Docker Hub.

Để đăng nhập vào Docker Hub bằng Docker client, ta sử dụng lệnh **docker login** và để đẩy image lên registry, ta sử dụng **docker push**.

Docker Hub hỗ trợ cả public repository và private tương tự như GitHub.

2.3.5 Kubernetes

2.3.5.1 Sơ lược về Kubernetes

Kubernetes (viết tắt là K8s) là một phần mềm mã nguồn mở cho việc tự động hóa quá trình triển khai (deploy), mở rộng (scale) và quản lý các ứng dụng đã được đóng gói để chạy trên container. Ban đầu được thiết kế bởi Google và hiện tại được quản lý bởi Cloud Native Computing Foundation. K8s hoạt động với nhiều công cụ quản lý container, phổ biến nhất là với Docker. Rất nhiều các nhà cung cấp dịch vụ Cloud như AWS, Google Cloud và Azure cho phép các ứng dụng triển khai bằng Kubernetes có thể chạy trực tiếp từ dịch vụ Kubernetes mà không phải tự thiết lập cluster sao cho phù hợp.

2.3.5.2 Các object trong Kubernetes

Kubernetes định nghĩa một tập các khối căn bản khi hoạt động cùng nhau sẽ cung cấp cơ chế giúp triển khai, duy trì và mở rộng ứng dụng dựa trên CPU, bộ nhớ hoặc là một metric bất kì. Các object chính trong Kubernetes bao gồm:

- Pod: là một tập các container mà luôn được đảm bảo sẽ chạy trên cùng một host và cho phép chia sẻ tài nguyên với nhau [1]. Pod là đơn vị cho thuật toán phân bổ container trong Kubernetes. Mỗi pod được gán một địa chỉ IP duy nhất trong cluster. Các container trong cùng một pod có thể tham chiếu lẫn nhau sử dụng địa chỉ localhost, những các container ở khác pod sẽ không thể truy cập trực tiếp với nhau thông qua localhost, khi đó phải sử dụng địa chỉ IP mà mỗi Pod được gán. Tuy nhiên, ứng dụng chạy trên container nằm trong pod không nên sử dụng địa chỉ IP trực tiếp của Pod bởi vì địa chỉ IP của Pod là không tĩnh, có thể bị thay đổi khi một Pod khởi động lại.
- Replica Set: là một cơ chế với mục tiêu đảm bảo số lượng ổn định của các Pod trong cluster tại bất kì thời điểm nào. Thông thường được sử dụng để chỉ định số lượng Pod giống nhau sẽ được chạy trong cluster.
- Service: Một Service trong Kubernetes là cơ chế nhóm các pod hoạt động cùng với nhau lại, cung cấp cơ chế Load Balance theo round-robin cho phép container nằm trong Pod ở bên ngoài service có thể truy cập các Pod ở bên trong service thông qua địa chỉ IP của Service thay vì địa chỉ IP trực tiếp của các Pod. Địa chỉ IP này có thể được lấy thông qua các biến môi trường hoặc thông qua một server DNS trong cluster. Service mặc định là public ở trong cluster. Tuy nhiên nếu muốn truy cập service từ bên ngoài cluster ta phải thay đổi kiểu của Service thành **type: LoadBalancer**. Khi đó Service sẽ chứa địa chỉ IP cho phép truy cập từ bên ngoài cluster.
- Deployment: là một cơ chế được xây dựng trên Replica Set cho phép thay đổi cấu hình của các Pod hoặc của Replica Set. Những thay đổi này được mô tả bởi người dùng bởi việc chỉ định cấu hình mong muốn, Deployment sẽ cập nhật các Pod và các Replica Set tương ứng để trạng thái hệ thống đạt được cấu hình mong muốn đó.

Ngoài ra còn các object khác như ReplicationController, StatefulSet, Volume, Endpoint, ...

2.3.6 Microsoft Azure

2.3.6.1 Sơ lược về Microsoft Azure

Microsoft Azure, thông thường được gọi là Azure, là một nền tảng điện toán đám mây (cloud computing) được tạo bởi Microsoft cho việc xây dựng, kiểm thử và quản lý các ứng dụng và dịch vụ chạy trên các trung tâm dữ liệu do Microsoft quản lý. Azure cung cấp nhiều các dịch vụ điện toán đám mây ở nhiều mức độ như Software as a Service (SaaS), Platform as a Service (PaaS) và Infrastructure as a Service (IaaS).

Azure được phát hành vào năm 2010 với tên là Windows Azure sau đó được đổi tên thành Microsoft Azure vào năm 2014.

2.3.6.2 Các dịch vụ của Microsoft Azure

Azure cung cấp rất nhiều các dịch vụ, trong đó phổ biến là:

- Máy ảo (virtual machine): là một Infrastructure as a Service (IaaS) cho phép người dùng có thể chạy bất kỳ một máy ảo Windows hay Linux đi kèm với các gói phần mềm phổ biến.
- App Service: là một Platform as a Service (PaaS) cho phép người dùng dễ dàng xuất bản và quản lý website.
- Các cơ sở dữ liệu như SQL Server, PostgreSQL, Redis, MySQL, MariaDB,...
- Container Instance: cho phép chạy và quản lý các container Docker độc lập mà không cần cơ chế điều phối như Kubernetes.
- Kubernetes Service: cho phép chạy và quản lý một Kubernetes cluster.

Và nhiều các dịch vụ khác.

2.3.6.3 Azure Kubernetes Service

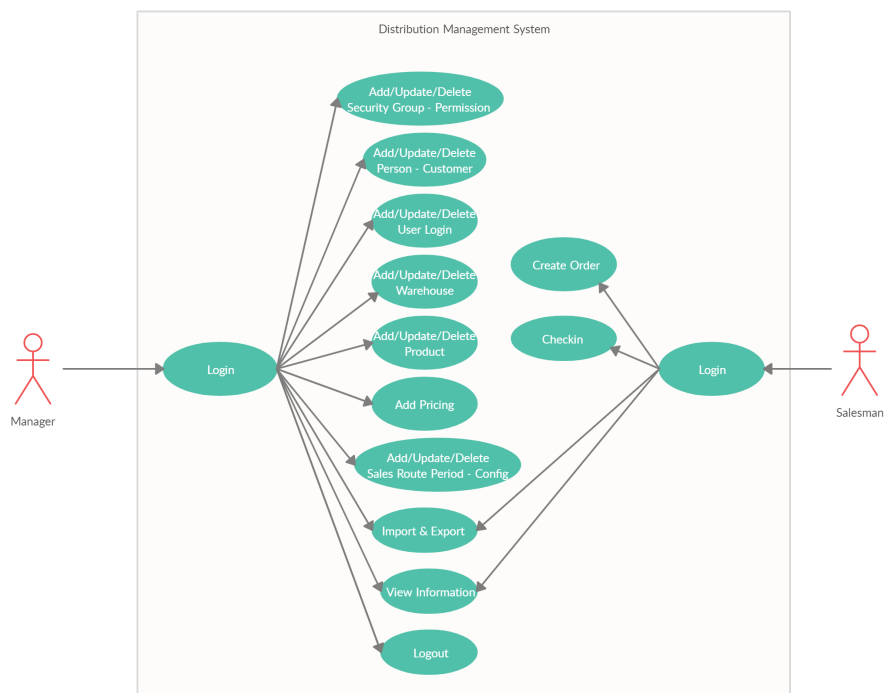
Azure Kubernetes Service (AKS) là một dịch vụ của Azure cho phép triển khai một cluster Kubernetes mà không phải quan tâm đến việc cài đặt một cluster Kubernetes ra sao. Để tương tác với AKS, ta có thể sử dụng Azure portal - một giao diện web, hoặc sử dụng Azure CLI với chương trình có tên là az để cài đặt và chạy chương trình dòng lệnh kubectl (chương trình quản lý cluster Kubernetes) cho phép quản lý AKS từ xa.

Chương 3

Phân tích thiết kế hệ thống

3.1 Tổng quan các chức năng

3.1.1 Biểu đồ use case tổng quan



Hình 3.1: Biểu đồ use case tổng quan

Hai tác nhân chính trong biểu đồ use case tổng quan là người quản lý (manager) và nhân viên bán hàng (salesman). Người quản lý ở đây là từ dùng chung, đại diện cho những người

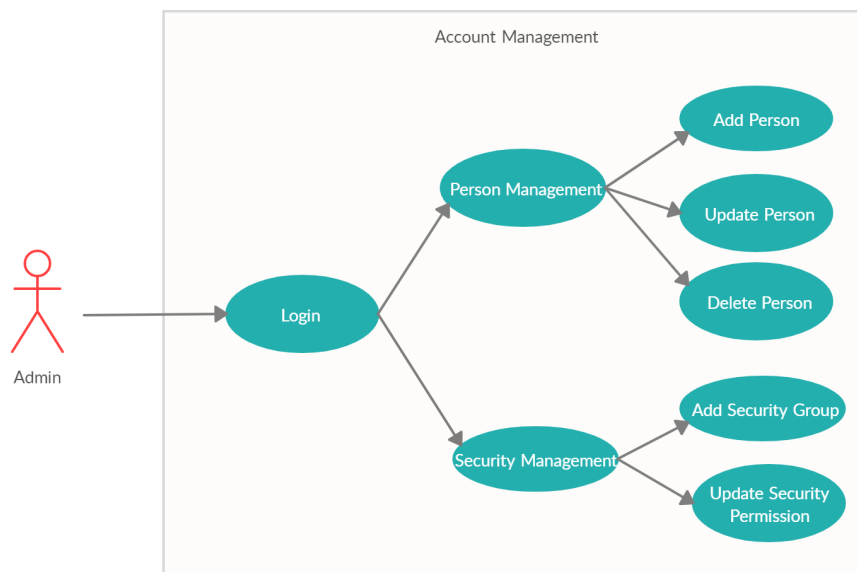
quản lý nghiệp vụ riêng (quản lý sản phẩm, quản lý hàng tồn kho, quản lý kho, quản lý giao dịch, quản lý tuyển bán hàng). Người quản lý sau khi đăng nhập vào hệ thống có thể thực hiện các nghiệp vụ quản lý như thêm / sửa / xóa sản phẩm, người dùng, ... Nhân viên bán hàng sau khi đăng nhập có thể xem được lịch trình di chuyển của mình, lên hóa đơn mua hàng, check-in.

3.1.2 Biểu đồ use case phân rã các chức năng của hệ thống

Trong phần này, em sẽ trình bày các ca sử dụng chính trong hệ thống, đưa ra biểu đồ use case và biểu đồ hoạt động chỉ cách thức tương tác của người dùng với giao diện hệ thống. Do nhiều thao tác có tính chất tương đồng nên em sẽ đưa ra các biểu đồ hoạt động điển hình, các thao tác khác được thực hiện tương tự.

3.1.2.1 Quản lý người dùng và phân quyền

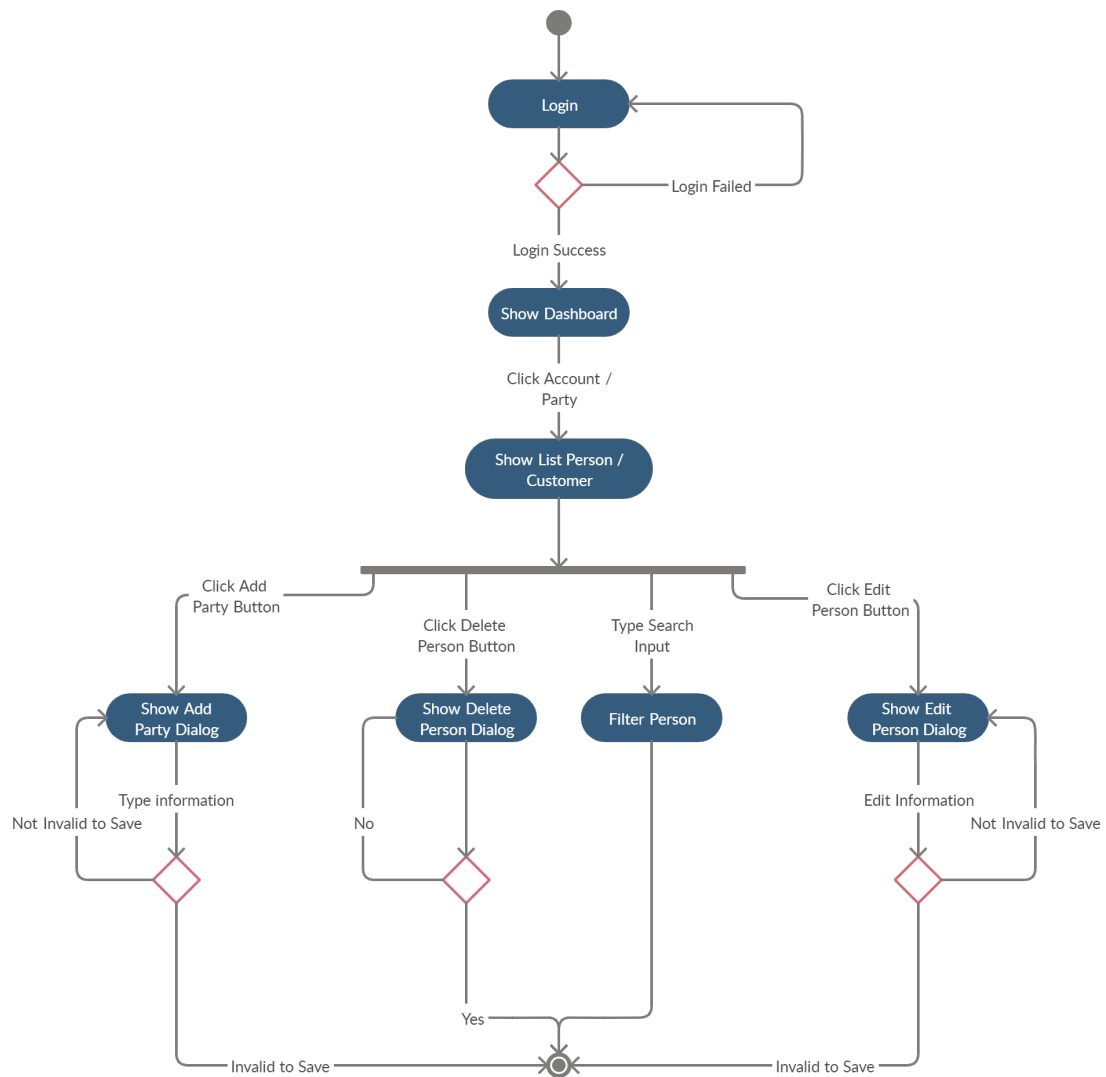
Biểu đồ use case cho ca sử dụng quản lý người dùng



Hình 3.2: Use-case quản lý người dùng

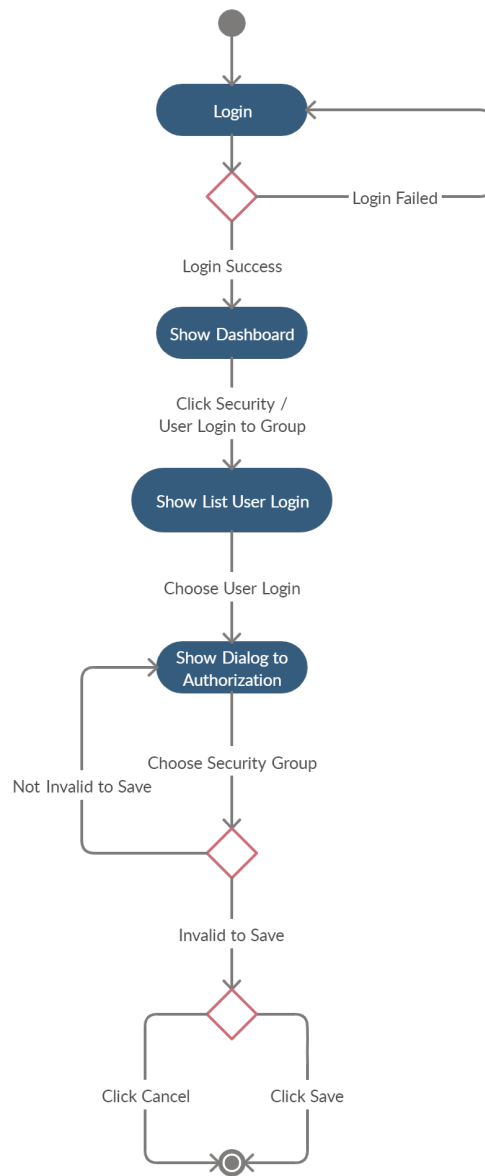
Người quản trị (admin) có thể quản lý các người dùng trong hệ thống, quản lý quyền hạn của người dùng. Có thể thêm / sửa / xóa người dùng, khách hàng, thêm / sửa quyền cho các người quản lý cấp dưới.

Biểu đồ hoạt động cho các thao tác trong quản lý người dùng:



Hình 3.3: Biểu đồ hoạt động ca sử dụng quản lý người dùng

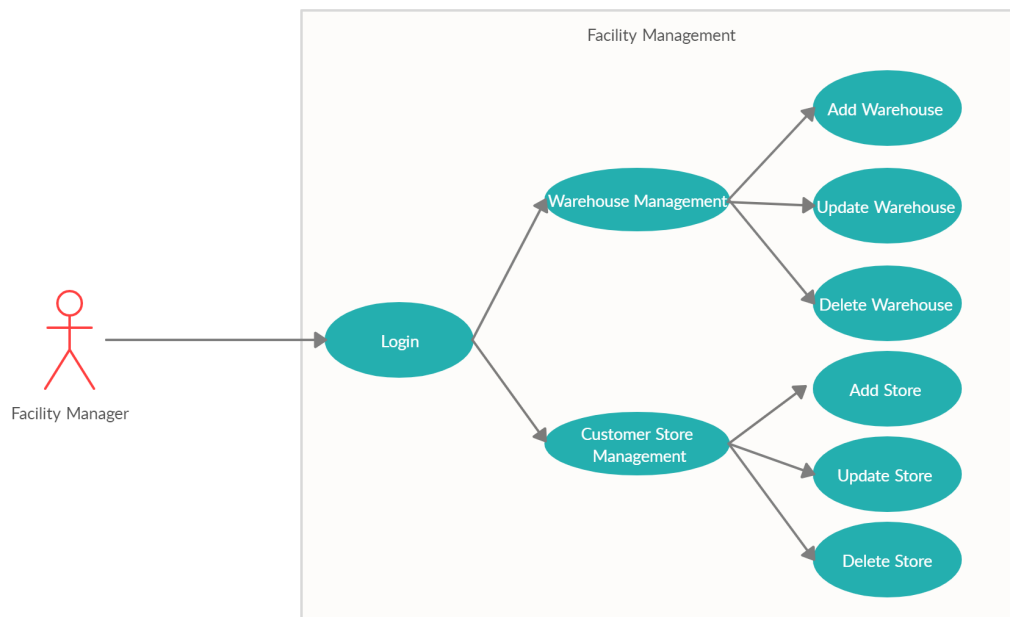
Biểu đồ hoạt động cho thao tác phân quyền:



Hình 3.4: Biểu đồ hoạt động cho thao tác phân quyền

3.1.2.2 Quản lý kho

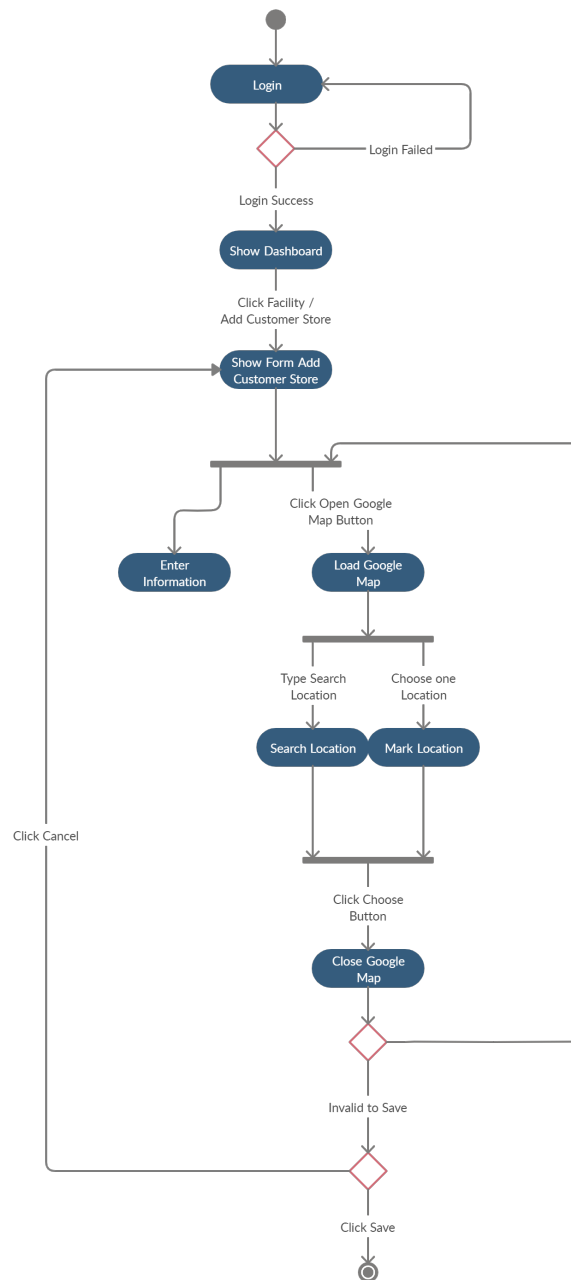
Biểu đồ use case cho ca sử dụng quản lý kho:



Hình 3.5: Use case quản lý kho

Người quản lý kho (Facility Manager) nắm được các thông tin về kho của doanh nghiệp, tập đoàn và kho của cửa hàng bán lẻ, có thể thêm / sửa / xóa các kho này.

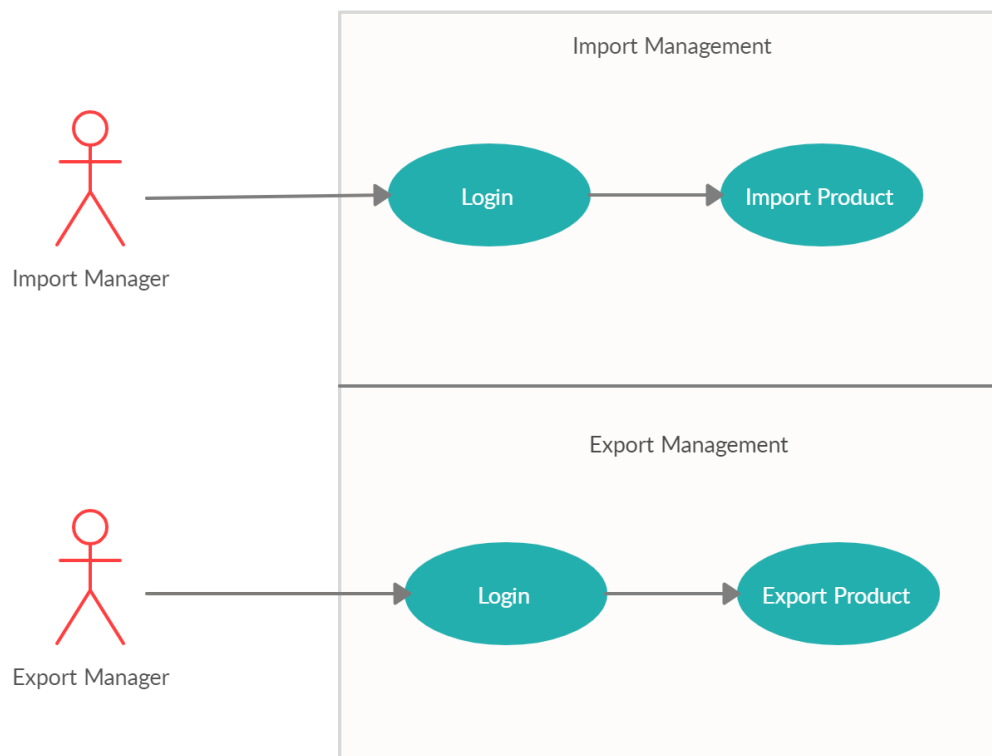
Biểu đồ hoạt động cho thao tác thêm cửa hàng bán lẻ:



Hình 3.6: Biểu đồ hoạt động cho thao tác thêm cửa hàng bán lẻ

3.1.2.3 Quản lý nhập – xuất hàng

Biểu đồ use case cho ca sử dụng quản lý nhập – xuất hàng:



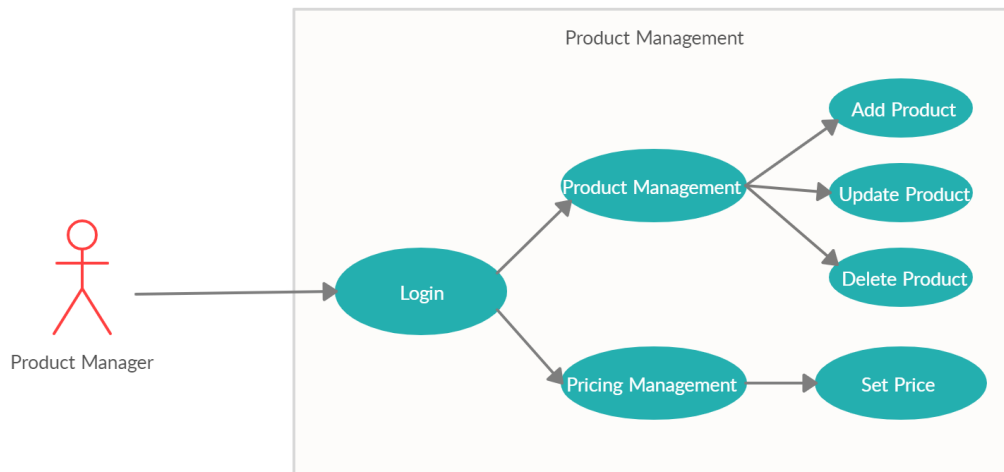
Hình 3.7: Use case quản lý xuất – nhập hàng

Người quản lý nhập hàng (Import Manager) có nhiệm vụ quản lý quy trình nhập hàng hóa từ nguồn bên ngoài (một doanh nghiệp sản xuất nào đó hoặc một tập đoàn bán lẻ khác, ...) vào kho của doanh nghiệp, tập đoàn mình.

Người quản lý xuất hàng (Export Manager) có nhiệm vụ xuất hàng từ kho của doanh nghiệp, tập đoàn sang các cửa hàng bán lẻ sau khi hóa đơn xuất hàng đã được duyệt.

3.1.2.4 Quản lý sản phẩm

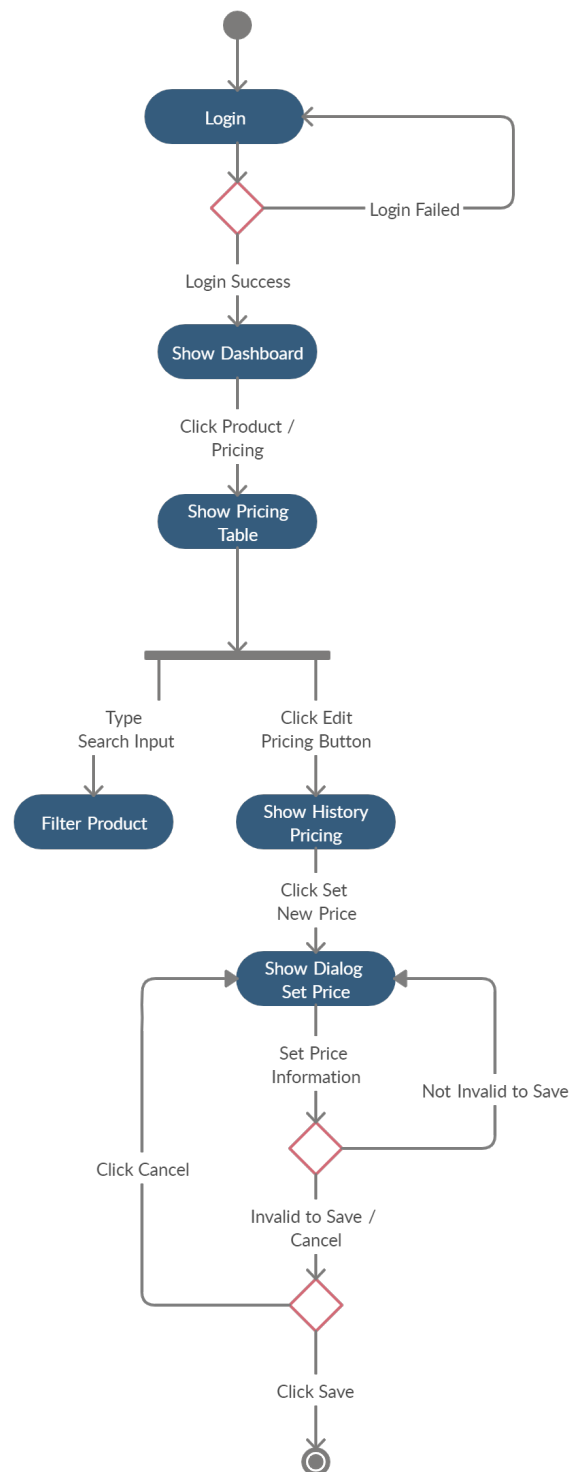
Biểu đồ use case cho ca sử dụng quản lý sản phẩm:



Hình 3.8: Use-case quản lý sản phẩm

Người quản lý sản phẩm (Product Manager) quản lý các sản phẩm đã được nhập vào và quản lý giá của sản phẩm.

Biểu đồ hoạt động cho thao tác gán giá cho sản phẩm:



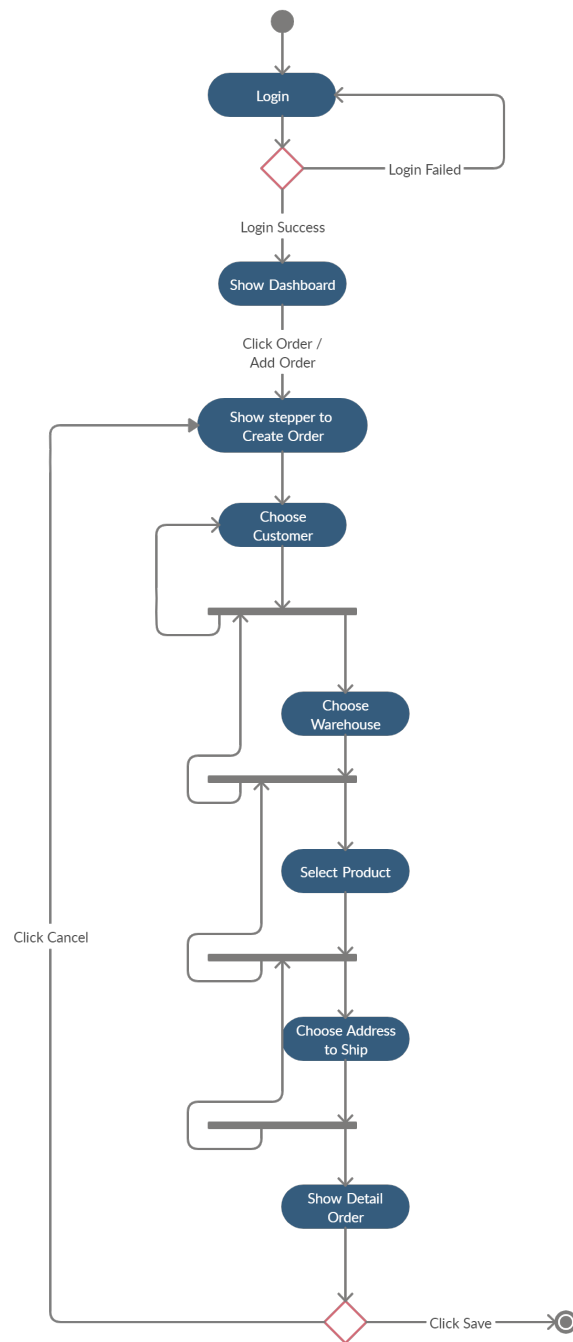
Hình 3.9: Biểu đồ hoạt động cho thao tác gán giá của sản phẩm

3.1.2.5 Quản lý đơn hàng

Biểu đồ use case cho ca sử dụng quản lý đơn hàng:

(chưa có)

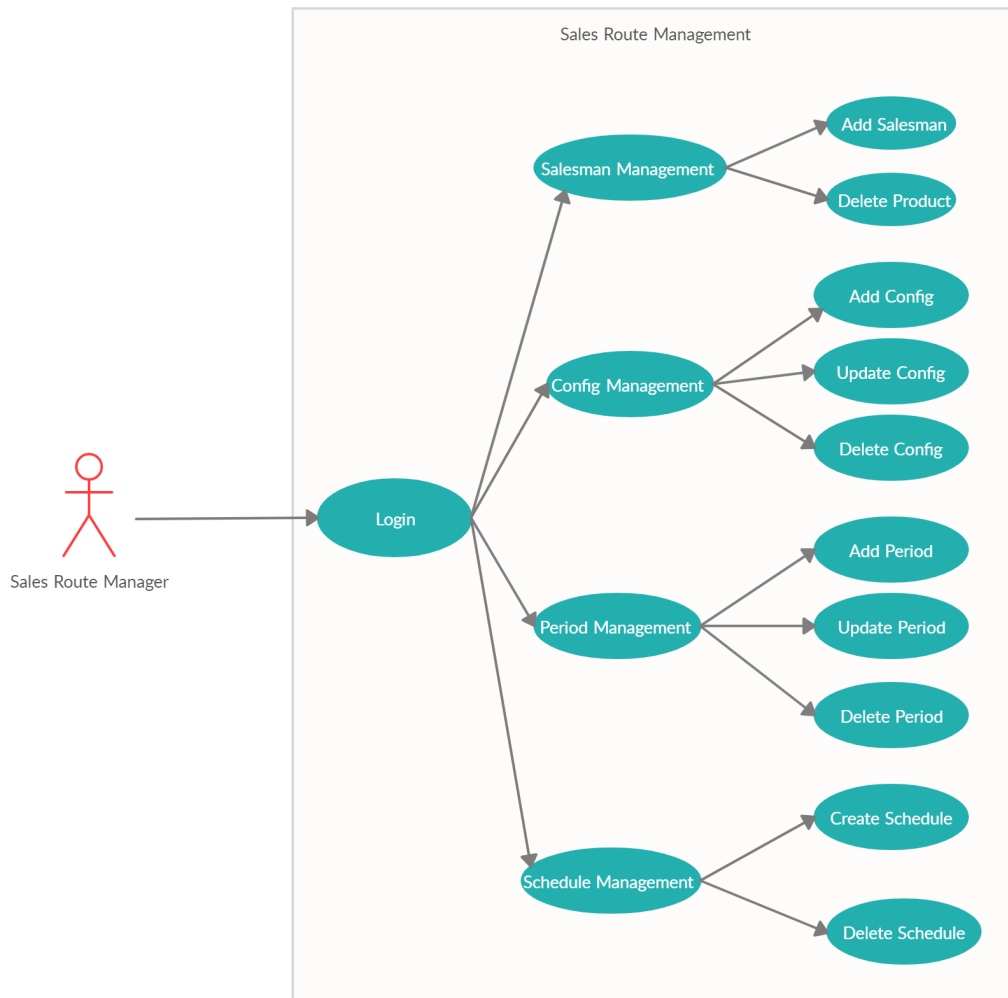
Biểu đồ hoạt động cho thao tác thêm đơn hàng:



Hình 3.10: Biểu đồ hoạt động cho thao tác thêm đơn hàng

3.1.2.6 Quản lý tuyến bán hàng

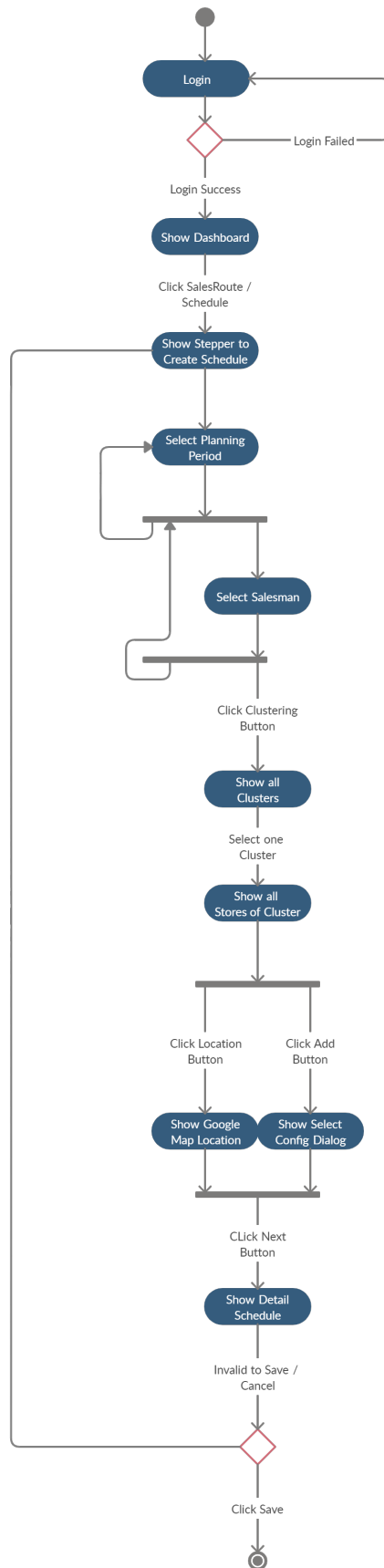
Biểu đồ use case cho ca sử dụng quản lý tuyến bán hàng:



Hình 3.11: Use case quản lý tuyến bán hàng

Người quản lý tuyến bán hàng (Sales Route Manager) quản lý danh sách tuyến, cấu hình các tuyến, lên lịch trình tuyến cho nhân viên bán hàng.

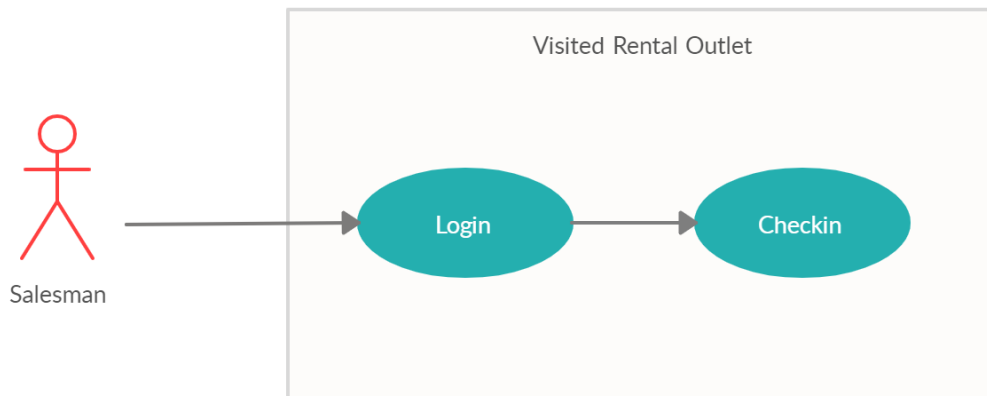
Biểu đồ hoạt động cho thao tác thêm tuyến bán hàng cho nhân viên bán hàng:



Hình 3.12: Biểu đồ hoạt động cho thao tác thêm tuyến bán hàng cho nhân viên bán hàng

3.1.2.7 Nhân viên bán hàng check-in

Biểu đồ use case cho ca sử dụng nhân viên bán hàng check-in:



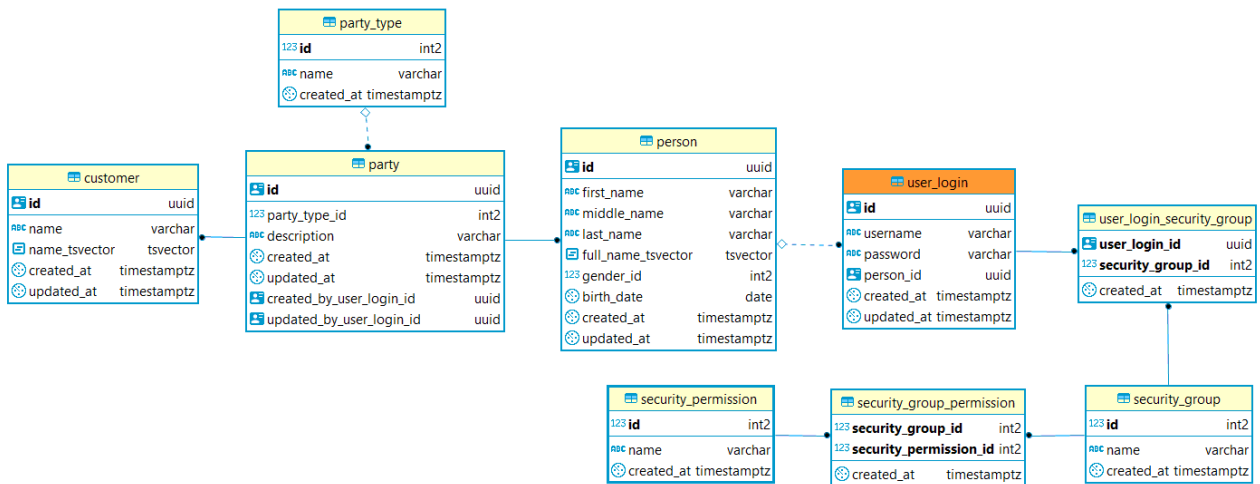
Hình 3.13: Use-case nhân viên bán hàng check-in

Khi nhân viên bán hàng đến check-in tại cửa hàng bán lẻ, hệ thống sẽ lưu lại thông tin về vị trí, thời điểm nhân viên bán hàng check-in.

3.2 Thiết kế cơ sở dữ liệu

Phần này trình bày chi tiết thiết kế các bảng của cơ sở dữ liệu sử dụng PostgreSQL phục vụ cho việc triển khai các use case đã trình bày ở trên.

Nhóm các bảng phục vụ chức năng quản lý phân quyền, quản lý người dùng được trình bày qua sơ đồ thực thể liên kết (Hình 3.14) và bảng thông tin (Bảng 3.1).

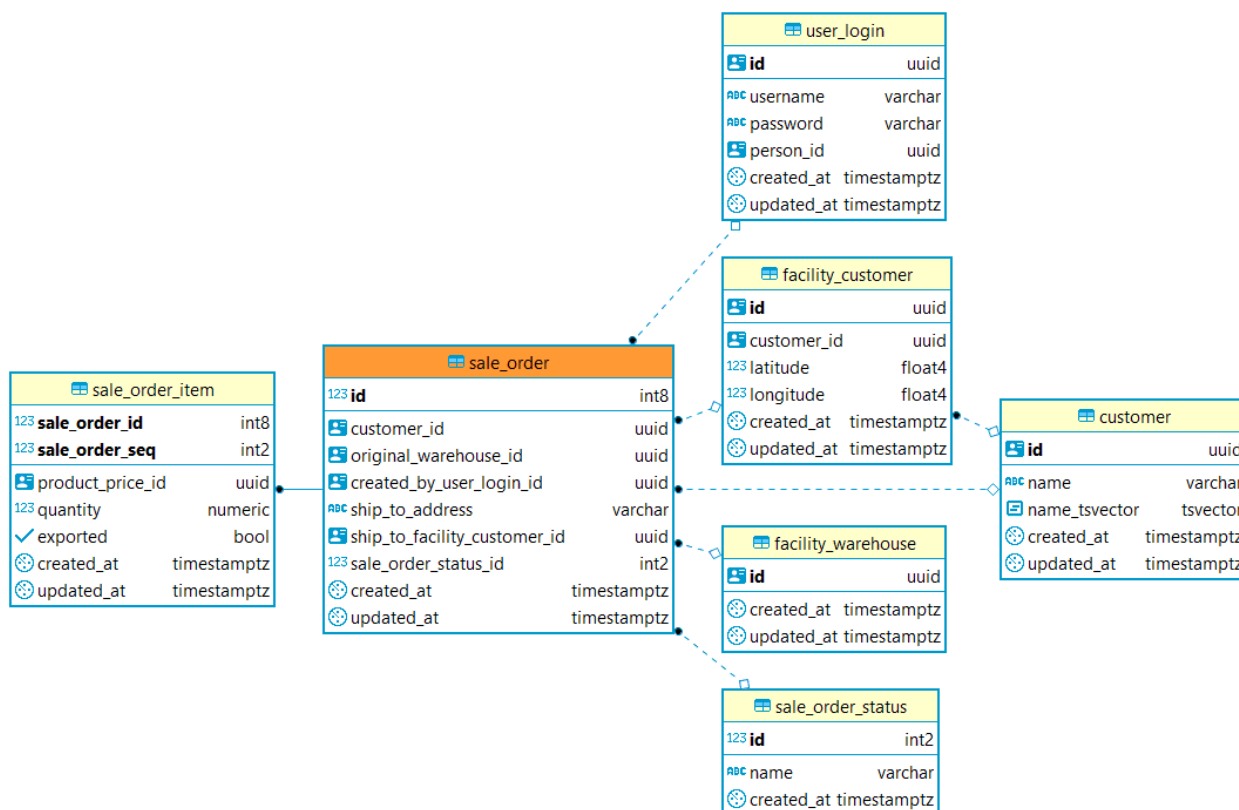


Hình 3.14: Cơ sở dữ liệu quản lý phân quyền, người dùng

Tên bảng	Thông tin chính được thể hiện
party_type	Loại người dùng (nhân viên, khách hàng)
party	Danh sách thông tin người dùng
customer	Danh sách thông tin khách hàng
person	Danh sách thông tin một nhân viên
user_login	Danh sách thông tin một tài khoản
security_group	Danh sách các role (account manager, product manager, ...)
security_permission	Danh sách các quyền (ADD_USER, EDIT_SALESMAN, ...)

Bảng 3.1: Thông tin các bảng nhóm quản lý phân quyền, người dùng

Nhóm các bảng phục vụ chức năng quản lý đơn hàng được trình bày qua sơ đồ thực thể liên kết (Hình 3.15) và bảng thông tin (Bảng 3.2).

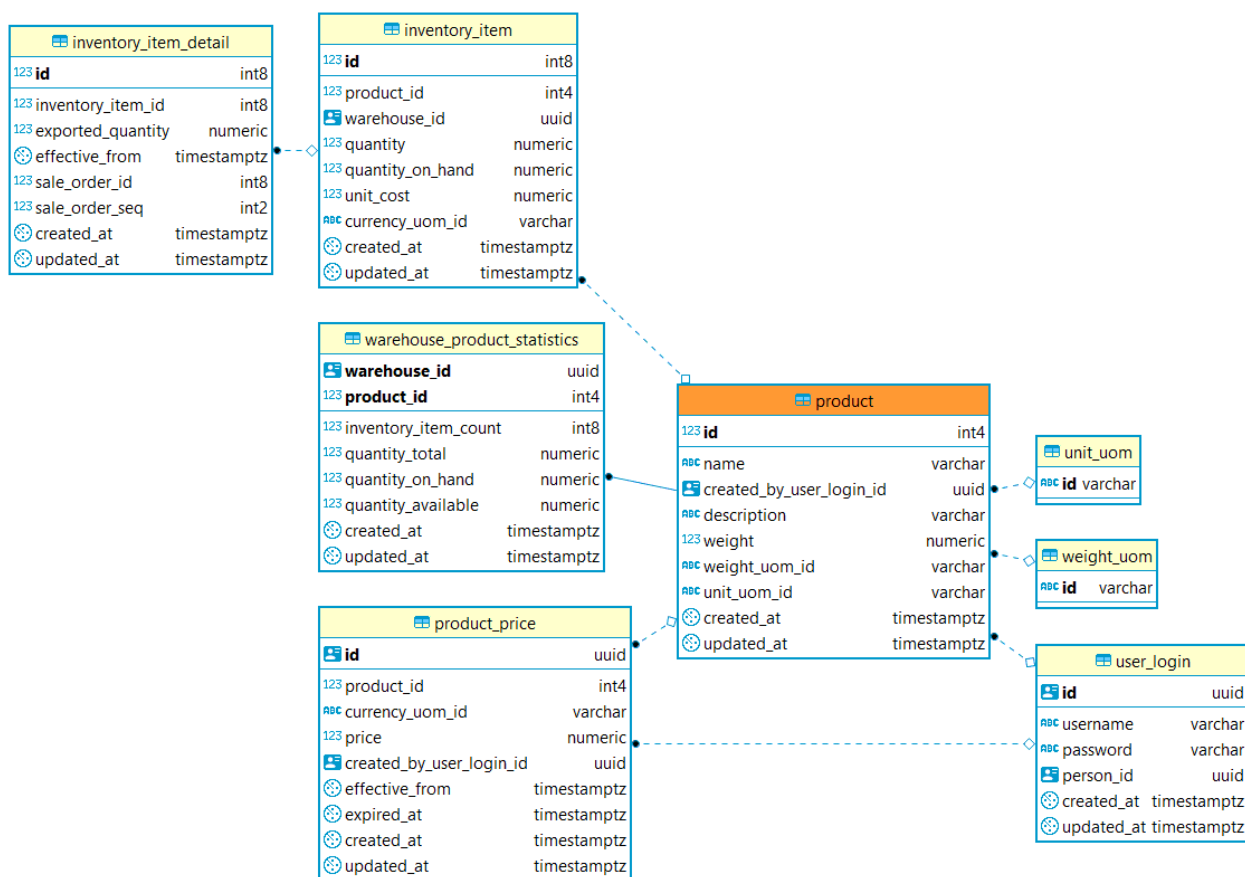


Hình 3.15: Cơ sở dữ liệu quản lý đơn hàng

Tên bảng	Thông tin chính được thể hiện
sale_order	Danh sách thông tin đơn hàng
sale_order_item	Danh sách các item trong đơn hàng
facility_customer	Danh sách thông tin cửa hàng bán lẻ
facility_warehouse	Danh sách thông tin kho
sale_order_status	Trạng thái của đơn hàng

Bảng 3.2: Thông tin các bảng nhóm quản lý phân quyền, người dùng

Nhóm các bảng phục vụ chức năng quản lý sản phẩm được trình bày qua sơ đồ thực thể liên kết (Hình 3.16) và bảng thông tin (Bảng 3.3).

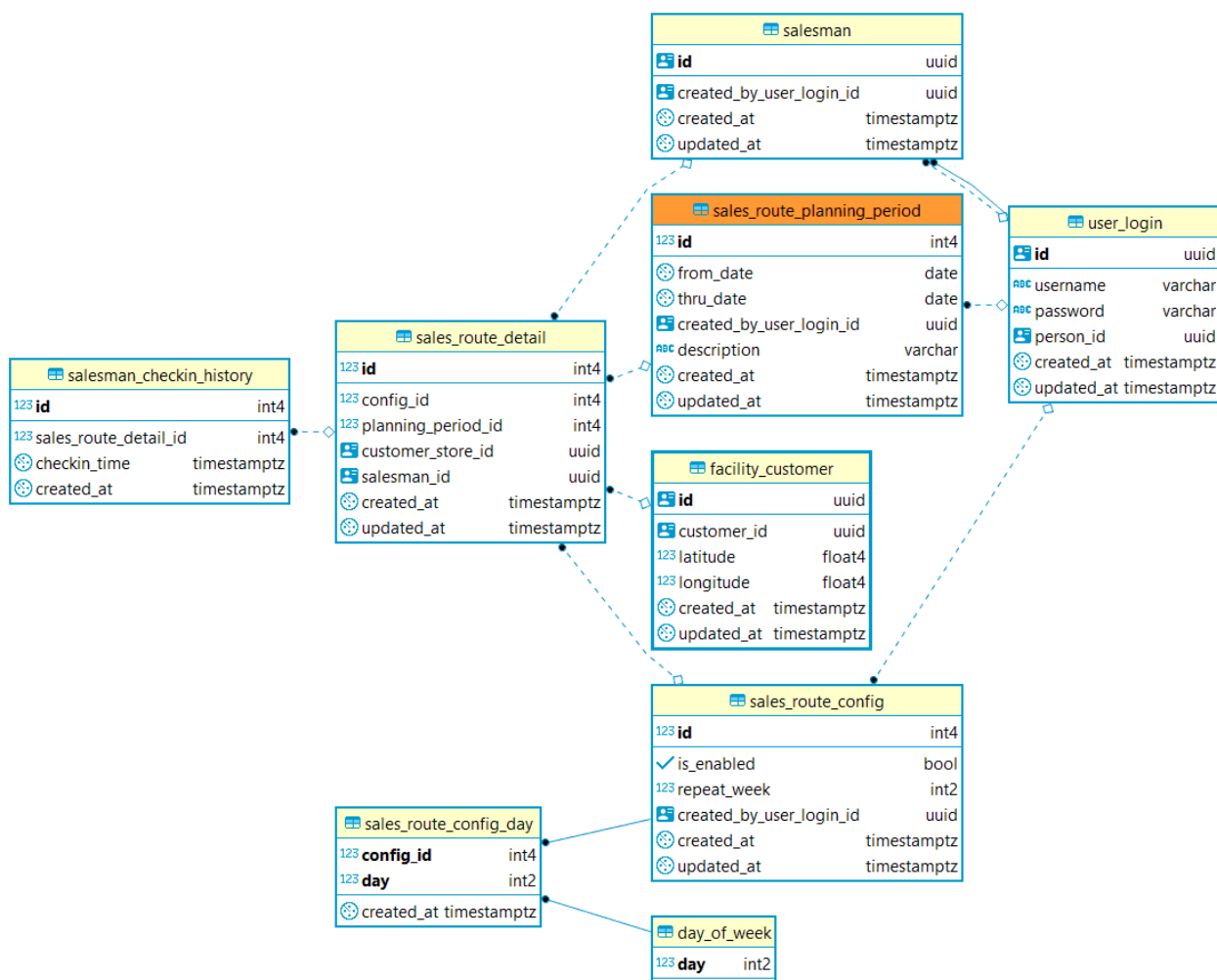


Hình 3.16: Cơ sở dữ liệu quản lý đơn hàng

Tên bảng	Thông tin chính được thể hiện
product	Danh sách thông tin sản phẩm (product nói chung, chưa nhập kho)
product_price	Thông tin giá ứng với sản phẩm
warehouse_product_statistics	Thống kê các thông tin của sản phẩm (số lượng tổng, số lượng đặt, số lượng có sẵn, ...)
inventory_item	Thông tin sản phẩm trong kho
inventory_item_detail	Thông tin chi tiết đối với từng sản phẩm trong kho
unit_uom	Đơn vị hình dạng (hộp, chai, gói)
weight_uom	Đơn vị khối lượng (kg, g, mg)

Bảng 3.3: Thông tin các bảng nhóm quản lý sản phẩm

Nhóm các bảng phục vụ chức năng quản lý tuyến bán hàng được trình bày qua sơ đồ thực thể liên kết (Hình 3.17) và bảng thông tin (Bảng 3.4).



Hình 3.17: Cơ sở dữ liệu quản lý tuyến bán hàng

Tên bảng	Thông tin chính được thể hiện
salesman	Danh sách thông tin nhân viên bán hàng
sales_route_planning_period	Danh sách thông tin tuyến bán hàng (thời gian bắt đầu – thời gian kết thúc)
sales_route_config	Danh sách thông tin cấu hình tuyến bán hàng (thăm cửa hàng vào thứ nào, mấy tuần một lần)
sales_route_detail	Chi tiết tuyến bán hàng
salesman_checkin_history	Thông tin checkin của nhân viên bán hàng
day_of_week	Danh sách các ngày trong tuần

Bảng 3.4: Thông tin các bảng nhóm quản lý tuyến bán hàng

Chương 4

Các giải pháp, đóng góp nổi bật và minh họa các chức năng của hệ thống

4.1 Vấn đề lựa chọn công nghệ

Có rất nhiều công nghệ bao gồm các thư viện, ngôn ngữ, frame work, tool giúp các lập trình viên dựng lên một trang web. Mỗi một công nghệ mới ra đời đều đáp ứng tốt nhu cầu thị trường tại thời điểm đó và khắc phục được những nhược điểm của các công nghệ cũ. Ví dụ trước đây, một lập trình viên front-end cơ bản chỉ cần biết về HTML, CSS sao cho dựng lên một trang web thật đẹp, responsive đầy đủ thì nay mọi thứ đã khác. Các website hiện nay phải động, phải linh hoạt, phải thay đổi, phải được cập nhật để không đi lùi, vì vậy các lập trình viên front-end hiện nay cần phải biết nhiều hơn về công nghệ, phải làm nhiều hơn.

Dựa theo xu thế công nghệ hiện nay, em chọn ReactJS – Redux để xây dựng phần giao diện khi mà thư viện này có một cộng đồng lớn và được Facebook phát triển. React-Router để định tuyến, xây dựng Single Page Application, i18next để triển khai đa ngôn ngữ, Redux-Saga để xử lý middleware, gửi request và nhận api từ back-end server.

Về phần back-end, ban đầu em chọn Java Spring để xây dựng. Java Spring với Maven hay Gradle hỗ trợ quản lý các gói thư viện thêm vào (dependency), tạo sẵn dự án theo chuẩn MVC (Model-View-Controller). Tuy nhiên trong quá trình code với Java Spring, có nhiều vấn đề xảy

ra như biên dịch Java chậm, máy tính cá nhân chạy dự án Java Spring bị chậm, Hibernate của Spring khó dùng các câu truy vấn SQL thông thường để giao tiếp với cơ sở dữ liệu, ... Vì vậy em chọn Go để xây dựng lại back-end server. Về cơ bản thì Go đã khắc phục được các nhược điểm ở trên, ngoài ra Go còn hỗ trợ tốt xử lý đa luồng với Goroutine và Chanel đã trình bày ở phần công nghệ back-end nên em nghĩ Go sẽ tăng hiệu năng xử lý request trong một thời điểm (kiểm tra bằng test hiệu năng).

4.2 Chức năng phân quyền động

Phân quyền là tính năng không thể thiếu được trong các trang mang tính chất quản trị, mỗi người cần phải được giới hạn trong truy cập. Ví dụ không thể để một nhân viên bán hàng sử dụng được tính năng quản lý sản phẩm hay giá sản phẩm. Do đó dựa trên các tính năng đã phân tích như quản lý phân quyền, quản lý sản phẩm, quản lý tài khoản, quản lý kho, quản lý tuyến bán hàng, ... em phân chia người dùng hệ thống vào các security_group, mỗi security_group lại có một tập quyền (permissions). Hệ thống sẽ dựa vào các permission mà người dùng có để xác định xem họ có thể truy cập được vào tính năng đó hay không.

security_group	security_permission
ADMIN	VIEW_EDIT_PARTY, VIEW_EDIT_USER_LOGIN, VIEW_EDIT_SECURITY_GROUP, VIEW_EDIT_SECURITY_PERMISSION
PRODUCT_MANAGER	VIEW_EDIT_PRODUCT
SALES_MANAGER	VIEW_EDIT_ORDER
FACILITY_MANAGER	VIEW_EDIT_FACILITY
INVENTORY_MANAGER	IMPORT
EXPORT_MANAGER	EXPORT
SALESMAN_MANAGER	VIEW_EDIT_SALESMAN
SALESMAN	SALESMAN_CHECKIN

Bảng 4.1: Security_group và permission tương ứng

Khi một người dùng được thêm vào hệ thống, họ sẽ được cấp một user_login (hiểu như là username), giả sử người này là quản lý sản phẩm thì sẽ được cấp một quyền VIEW_EDIT_PRODUCT, nếu người này có khả năng quản lý rộng hơn làm được cả nhập kho thì sẽ cấp thêm quyền IMPORT, ... Đường dẫn (url) sẽ được gán theo permission để kiểm soát truy nhập. Ví dụ PRODUCT_MANAGER muốn xem danh sách các sản phẩm thì phải truy cập url có dạng

.../view-product, url này chỉ được truy cập khi có quyền VIEW_EDIT_PRODUCT.

4.3 Chức năng phân cụm cửa hàng

Đây là một tính năng hoàn toàn mới, theo tìm hiểu và khảo sát của em thì chưa có trên các phần mềm quản lý phân phối hiện nay. Xuất phát từ ý tưởng các nhân viên bán hàng hàng ngày phải thăm các cửa hàng bán lẻ. Vậy làm thế nào có thể tối ưu được quãng đường di chuyển của nhân viên, để họ có thể di chuyển một cách dễ dàng nhất. Và em đã đưa ra tính năng này, với mục đích là đưa ra gợi ý về các cụm cửa hàng gần nhau, giúp người quản lý tuyển bán hàng chọn ra các tuyến bán hàng phù hợp.

Sử dụng thuật toán phân cụm K-Means với đầu vào là N cửa hàng bán lẻ trong hệ thống, và K cụm (do người quản lý tuyển chỉ định). Đầu ra của thuật toán là K cụm, trong đó là vị trí các cửa hàng gần nhau. Tuy nhiên khoảng cách sử dụng không phải là khoảng cách Euclidean, mà là khoảng cách Haversine – công thức tính khoảng cách giữa hai điểm trên tọa độ thực tế (dựa trên kinh độ và vĩ độ).

Công thức Haversine tính khoảng cách giữa 2 điểm (x_1, y_1) , (x_2, y_2) trong không gian sử dụng kinh độ và vĩ độ:

$$\Delta x = x_1 - x_2$$

$$\Delta y = y_1 - y_2$$

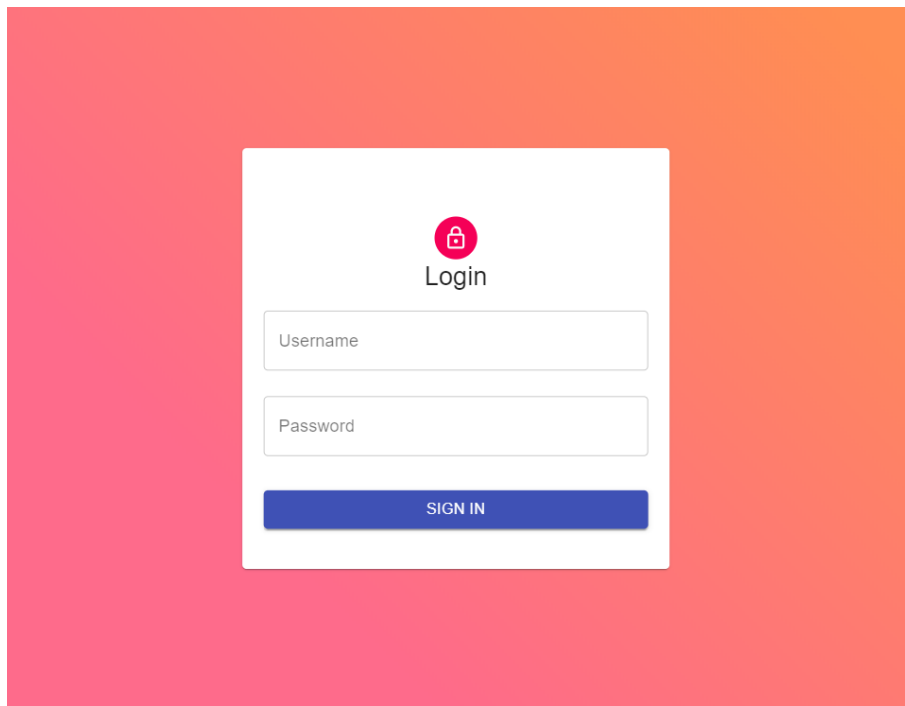
$$a = \sin^2\left(\frac{\Delta x}{2}\right) + \cos(x_1) \times \cos(x_2) \times \sin^2\left(\frac{\Delta y}{2}\right)$$

$$c = 2 \times \arcsin(\sqrt{a})$$

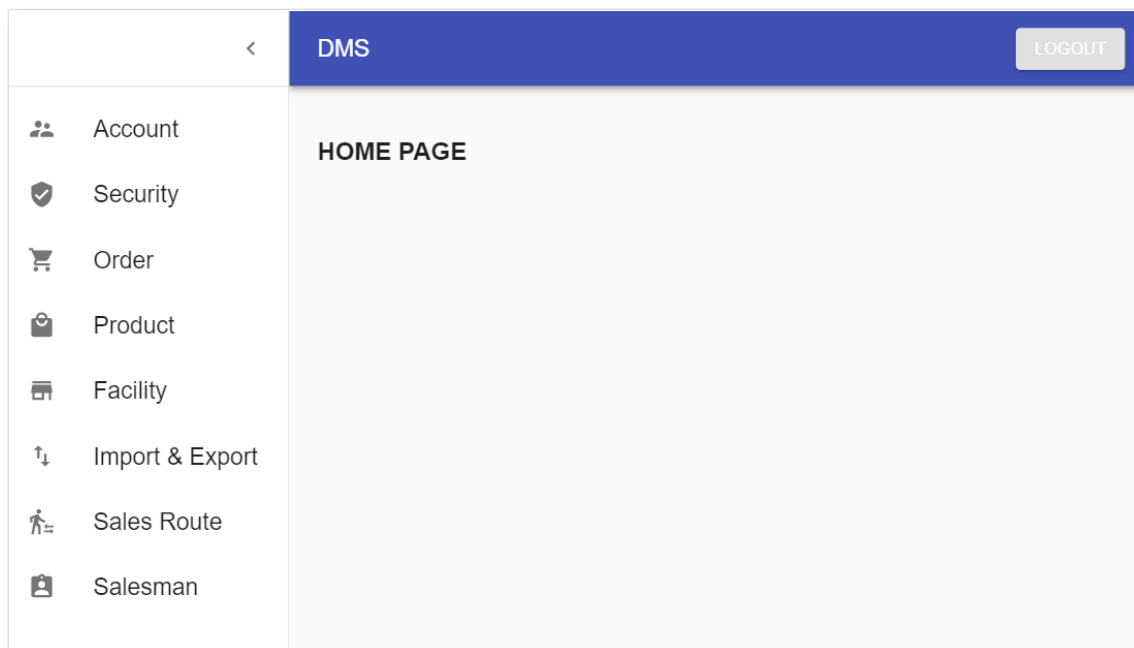
$$d = R \times c$$

Trong đó: x_1, x_2 là kinh độ, y_1, y_2 là vĩ độ, tính bằng radian. R là bán kính trái đất (xấp xỉ 6,371 km). d là khoảng cách cần tìm.

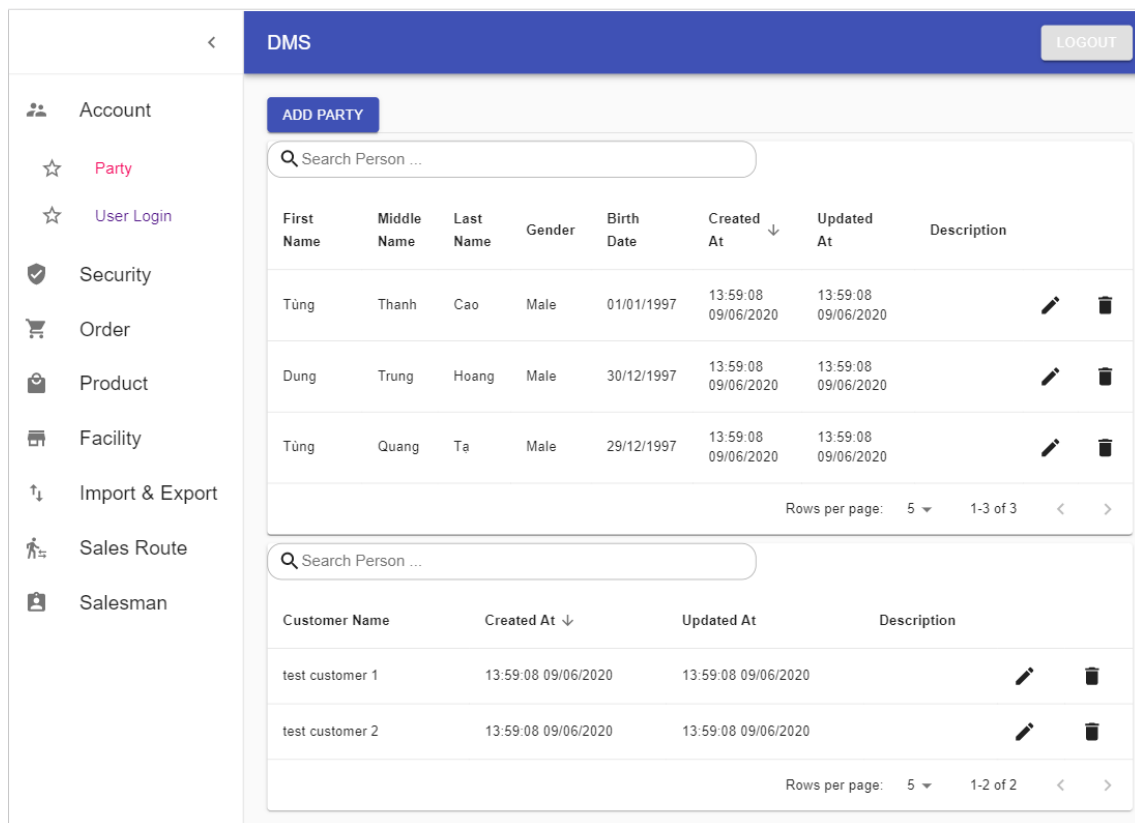
4.4 Minh họa các chức năng của hệ thống



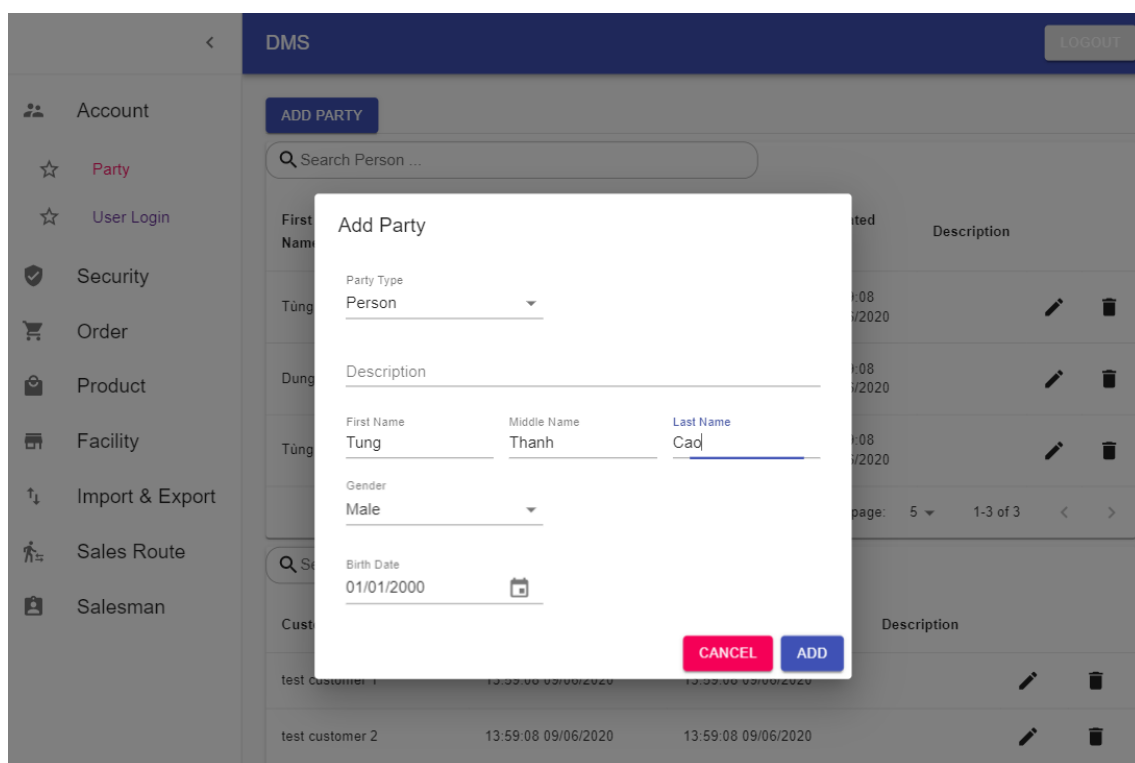
Hình 4.1: Chức năng đăng nhập



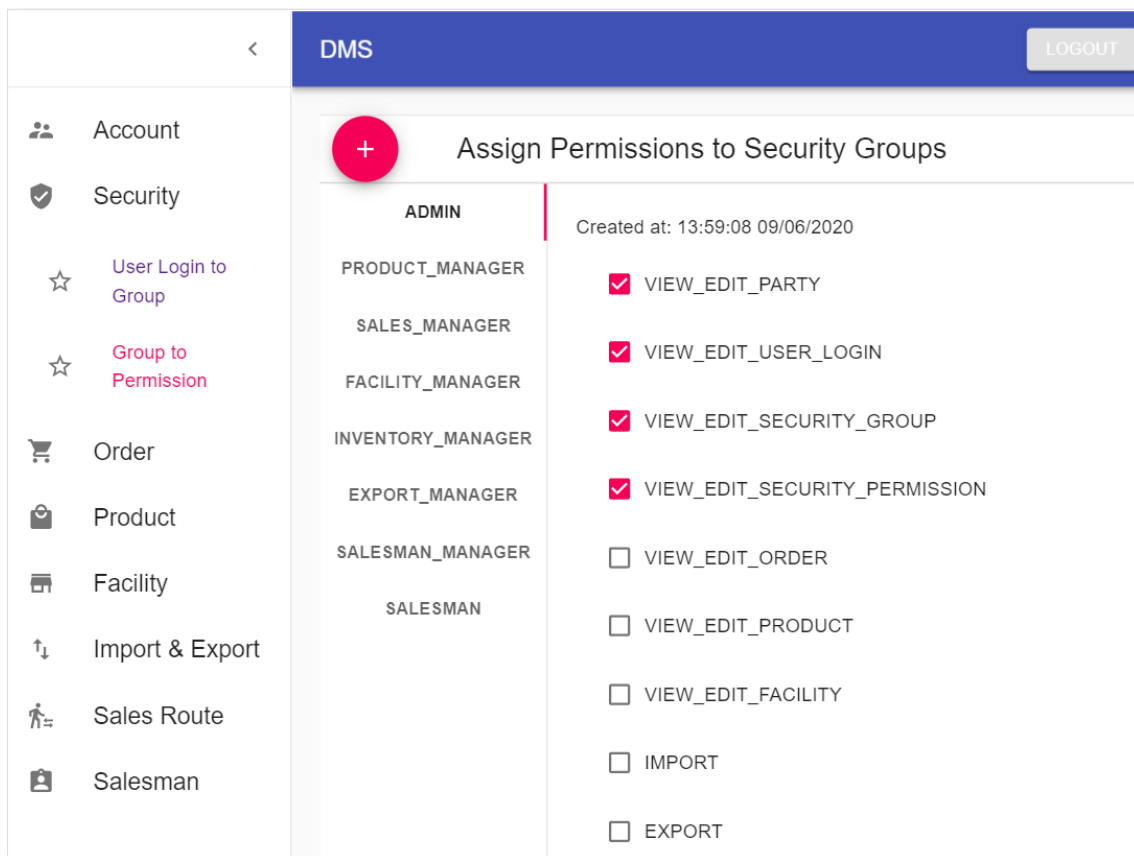
Hình 4.2: Giao diện trang chủ hệ thống



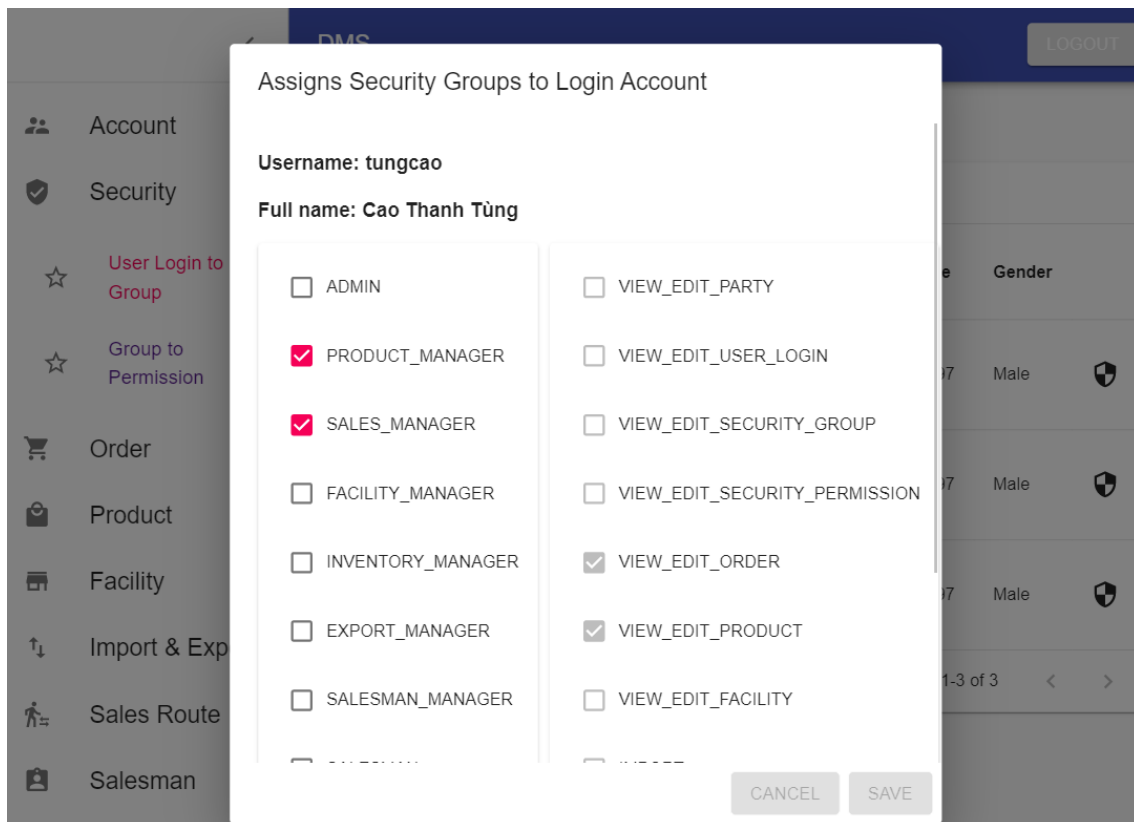
Hình 4.3: Quản lý người dùng



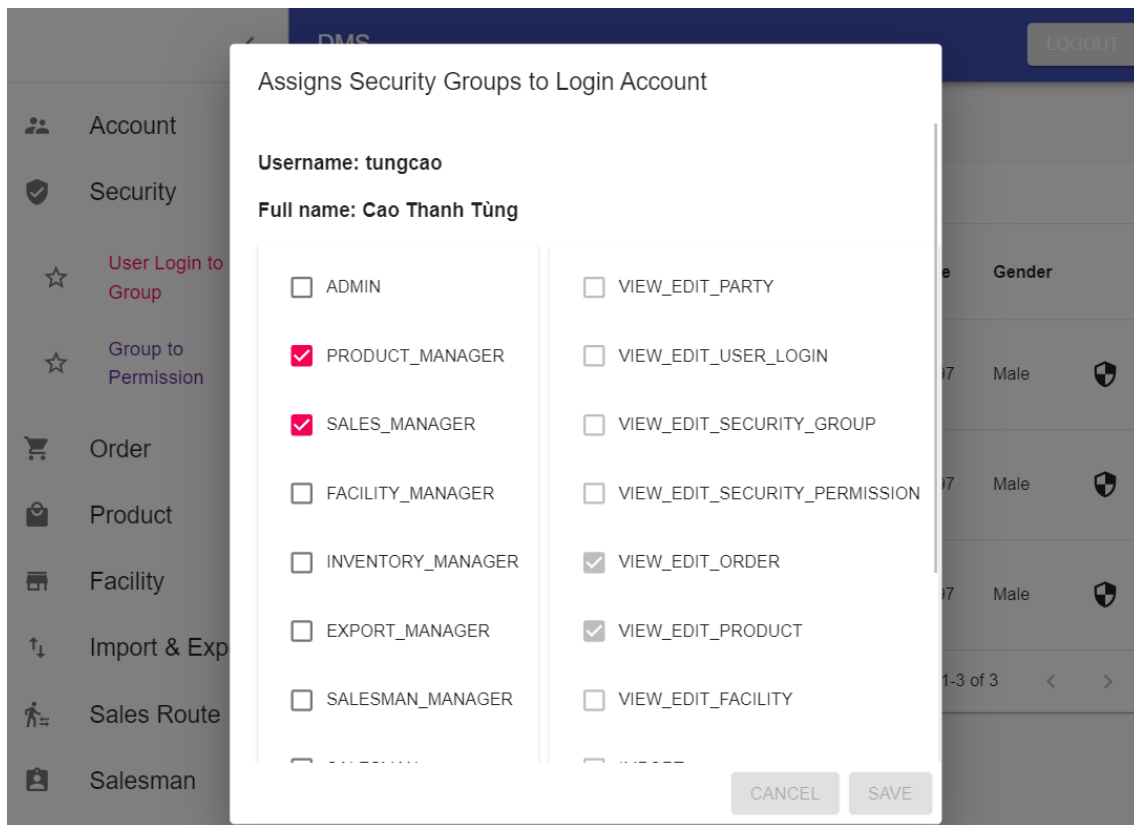
Hình 4.4: Thêm người dùng / khách hàng



Hình 4.5: Quản lý các nhóm quyền

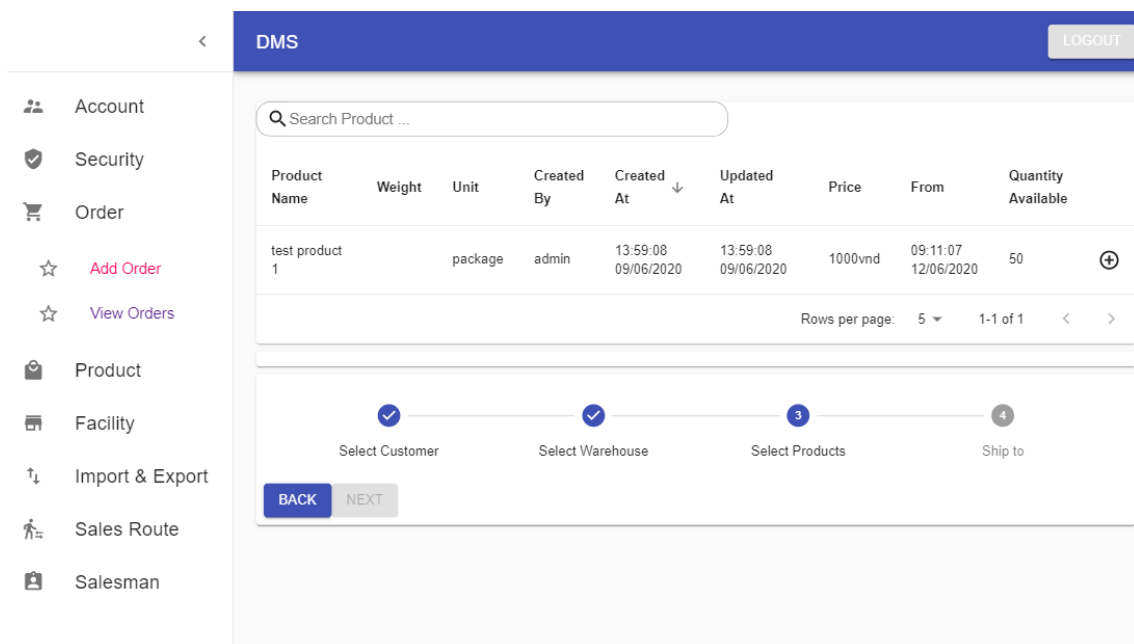


Hình 4.6: Gán quyền cho tài khoản người dùng



Hình 4.7: Gán quyền cho tài khoản người dùng

Tạo đơn hàng mới phải qua các bước như chọn khách hàng (để giao hàng đến), chọn kho chứa hàng hóa, chọn hàng hóa, nhập địa chỉ (nếu có hoặc chọn kho của cửa hàng).



Hình 4.8: Tạo đơn hàng

<

DMS

LOGOUT

Account

Security

Order

Product

View & Edit

Pricing

Facility

Import & Export

Sales Route

Salesman

Search Product ...

Product Name	Weight	Unit	Created By	Created At	Updated At	Price	From
test product 1		package	admin	13:59:08 09/06/2020	13:59:08 09/06/2020	1000vnd	09:11:07 12/06/2020
test product 2		box	admin	13:59:08 09/06/2020	13:59:08 09/06/2020		

Rows per page: 5 1-2 of 2

Product Name: test product 1

Weight: None

Price: 1000vnd

Price Set By: admin

Effective From: 09:11:07 12/06/2020

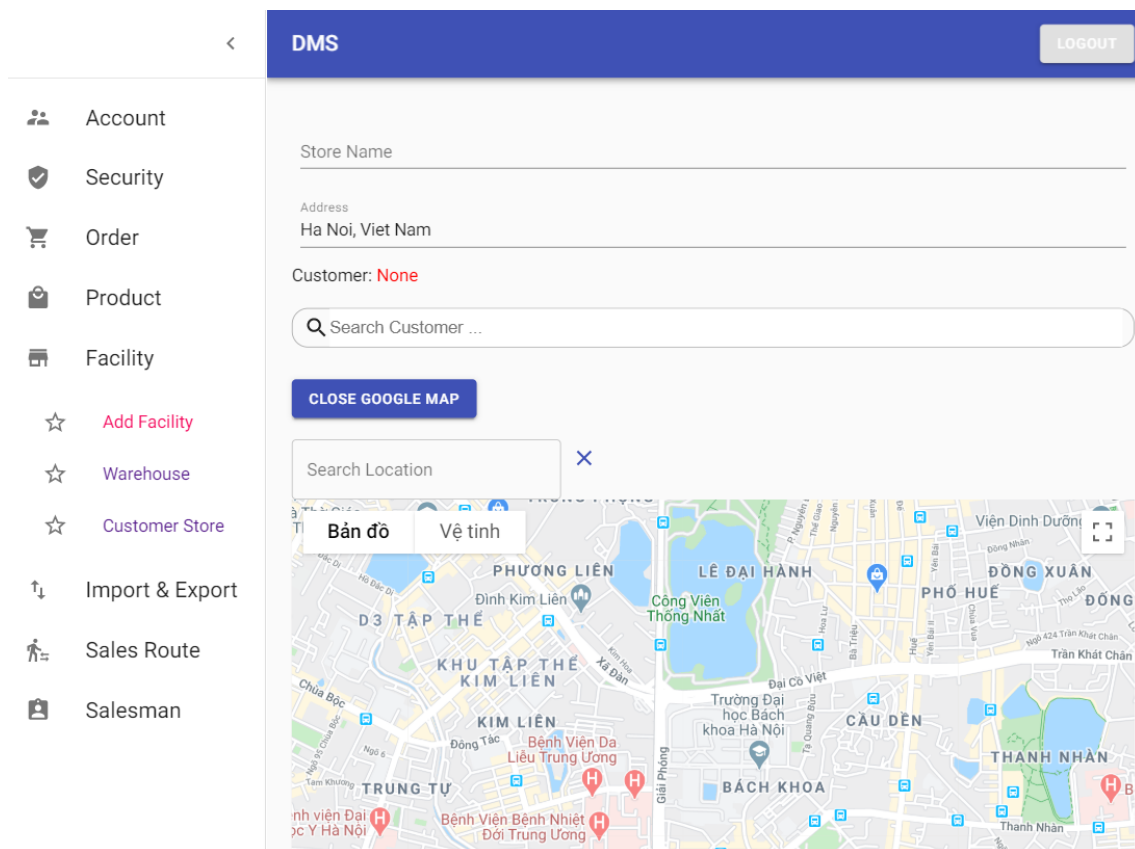
Expired At: None

SET NEW PRICE

Price	Set By	Effective From	Expired At	Created At	Updated At
-------	--------	----------------	------------	------------	------------

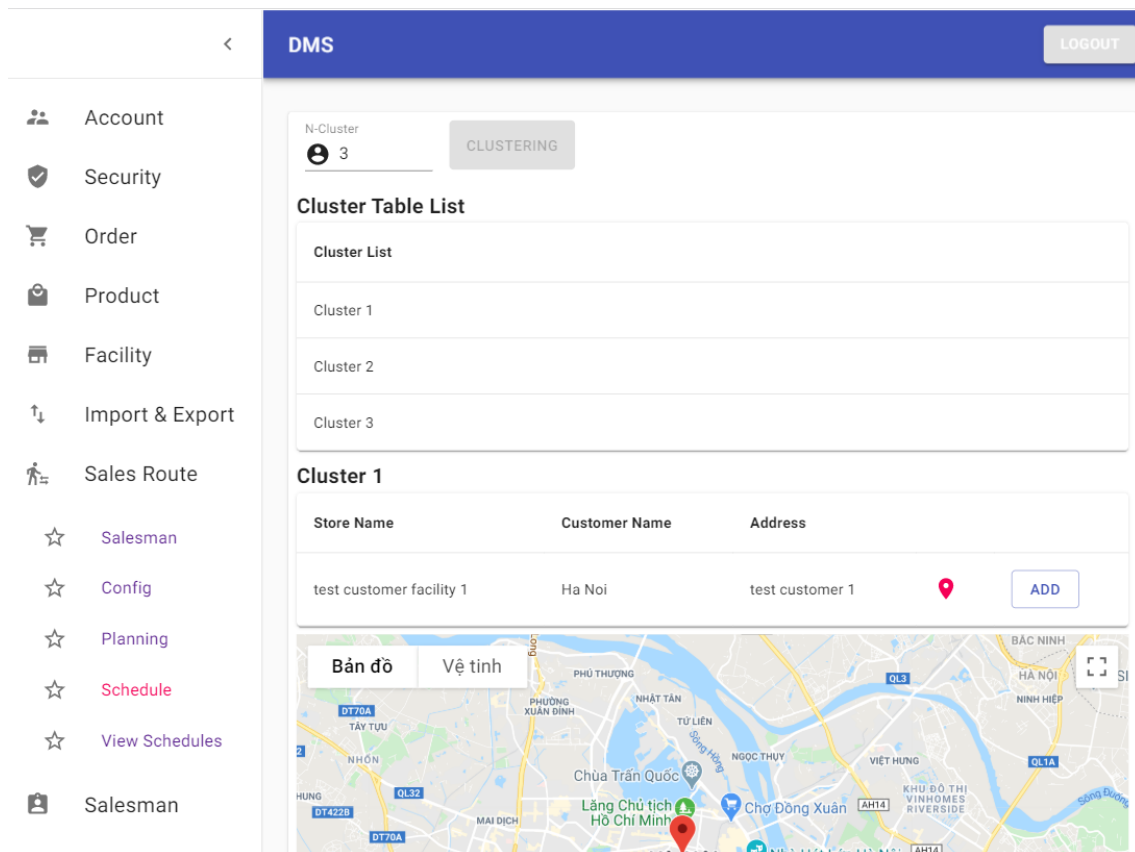
Hình 4.9: Quản lý sản phẩm và giá

Chức năng thêm mới cửa hàng bán lẻ, hỗ trợ bản đồ Google Map để xác định vị trí.

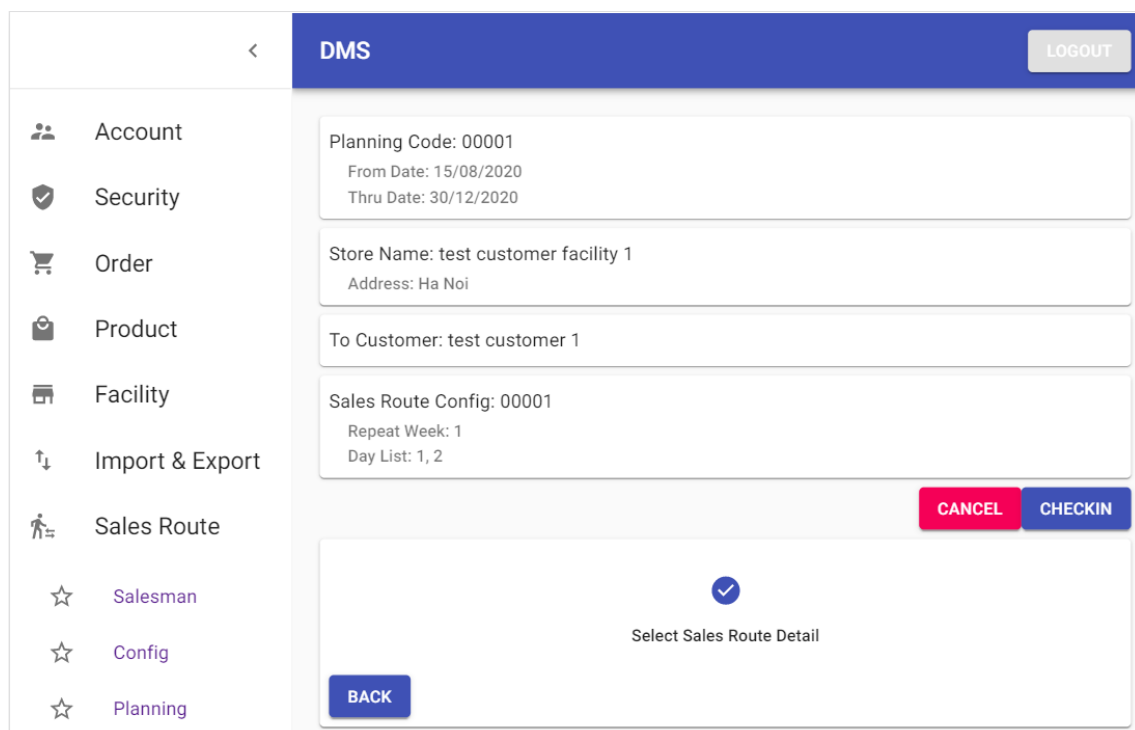


Hình 4.10: Thêm mới một cửa hàng bán lẻ

Chức năng tạo lịch trình tuyến bán hàng, sử dụng thuật toán K-Means phân cụm các cửa hàng bán lẻ, từ đó người quản lý chọn ra tuyến bán hàng phù hợp nhất cho nhân viên bán hàng.



Hình 4.11: Tạo lịch trình tuyến bán hàng



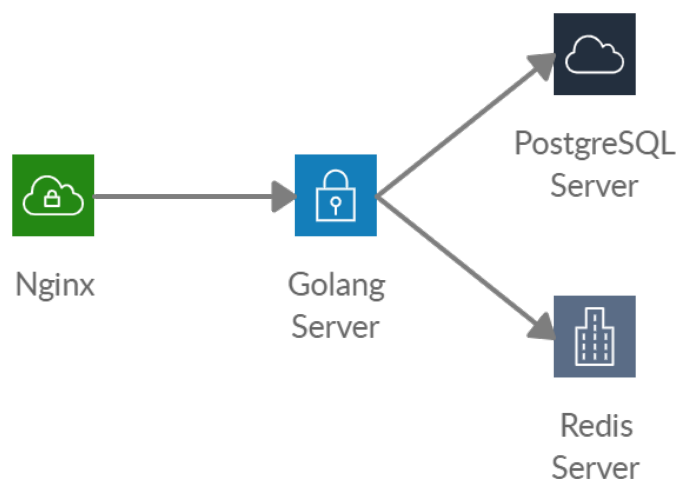
Hình 4.12: Salesman check-in

Chương 5

Thử nghiệm và đánh giá

5.1 Triển khai chạy thử nghiệm trên cloud

Trong đồ án này, em sử dụng Microsoft Azure làm nhà cung cấp dịch vụ cloud và sử dụng Azure Kubernetes Service để triển khai các server. Kiến trúc tổng quan của hệ thống như hình sau:



Hình 5.1: Kiến trúc tổng quan của hệ thống

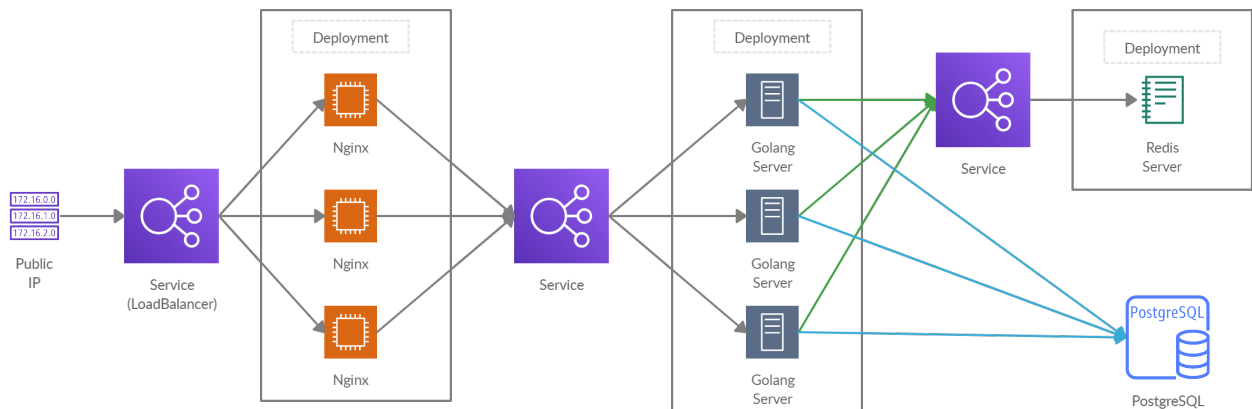
Trong đó:

- Nginx: sử dụng để làm web server trả về các file tĩnh html và js khi ứng dụng được mở

lên lần đầu, đồng thời được sử dụng làm Reverse Proxy đến Golang Server cho các Rest API.

- Golang Server: web server cho các Rest API từ trình duyệt gửi lên thông qua AJAX.
- PostgreSQL Server: cơ sở dữ liệu quan hệ của hệ thống.
- Redis Server: lưu trữ dữ liệu của các phiên làm việc (session).

Còn để triển khai lên AKS, em đã sử dụng những thành phần sau:



Hình 5.2: Triển khai lên Azure Kubernetes Service

Kết quả sau khi triển khai là hệ thống có thể truy cập được từ địa chỉ public IP.

5.2 Kiểm thử hiệu năng (performance testing)

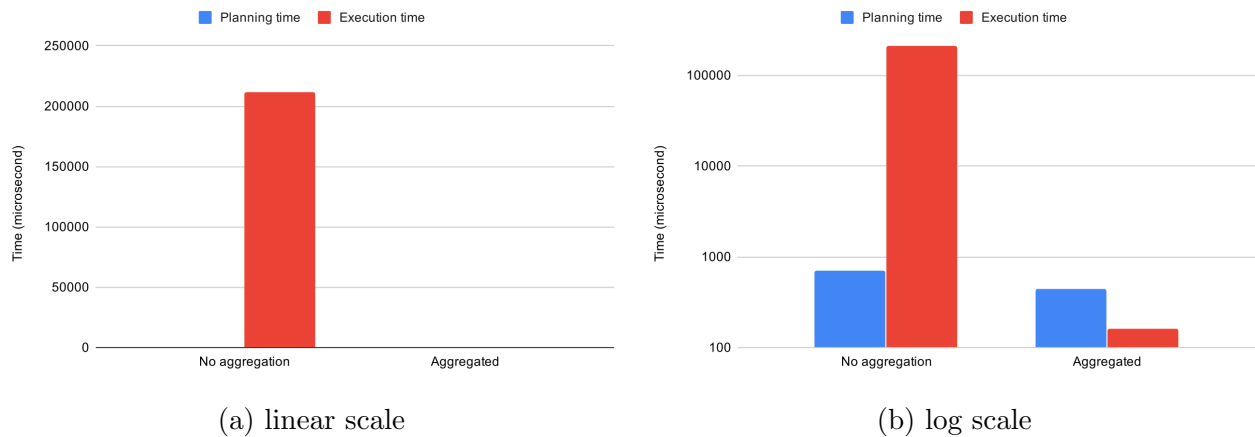
Dưới đây là kết quả chạy thử nghiệm với số lượng gần một triệu bản ghi trong bảng. Do cách mà PostgreSQL hoạt động, câu lệnh select count sẽ rất chậm trong bảng kích thước lớn do PostgreSQL phải duyệt qua toàn bộ phần tử trong bảng mới xác định được số lượng phần tử tương ứng. Vì vậy, trong bài kiểm thử hiệu năng này, em sử dụng hai cách để đếm số lượng phần tử: sử dụng select count (no aggregation) và sử dụng một biến đếm riêng biệt (aggregated). Đồng thời cũng chia các trường hợp có sử dụng hoặc không sử dụng Index.

Bài kiểm thử hiệu năng này được thử nghiệm trên cả việc chạy trên máy tính cá nhân (máy local) và cả sau khi đã triển khai trên cloud (máy cloud). PostgreSQL trên máy local sử

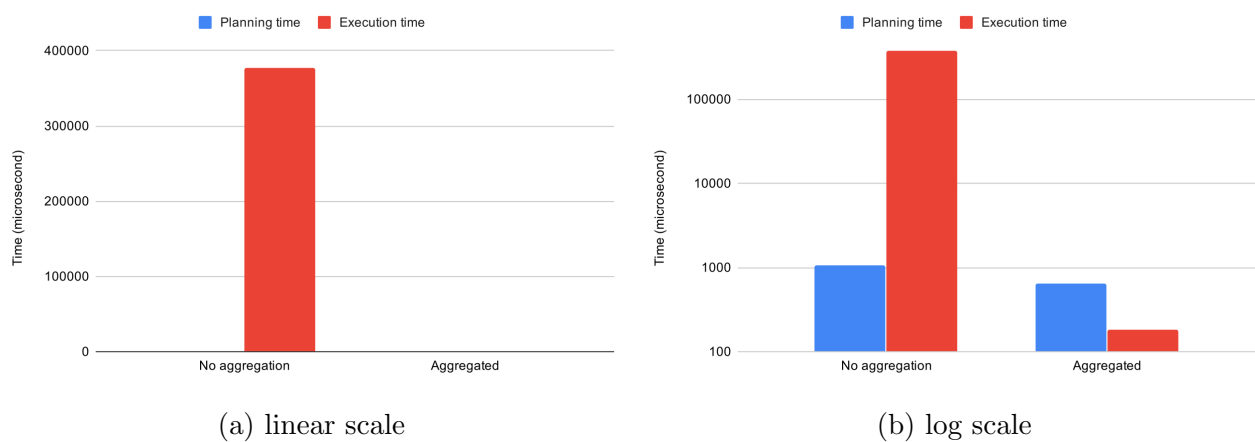
dùng Intel core i3 3217U còn máy trên cloud sử dụng PostgreSQL của Microsoft Azure với 2 core.

5.2.1 Kiểm thử hiệu năng cơ sở dữ liệu

5.2.1.1 Truy vấn count



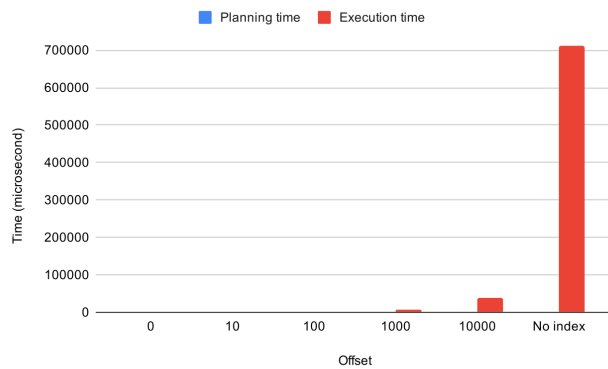
Hình 5.3: Truy vấn count trên máy local



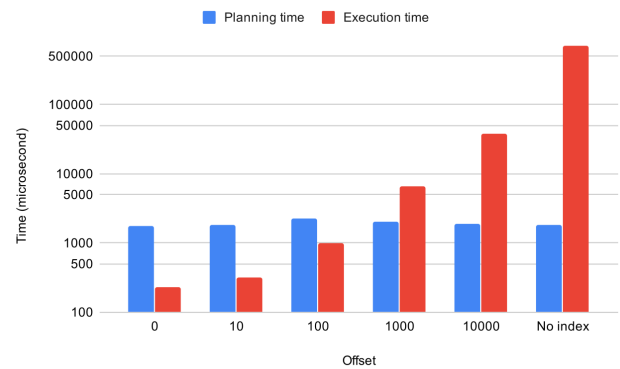
Hình 5.4: Truy vấn count trên máy cloud

Đánh giá kết quả: Có thể thấy được rằng với một triệu phần tử thì việc sử dụng trực tiếp select count là rất tốn kém, gấp rất nhiều lần so với việc sử dụng một biến đếm riêng biệt.

5.2.1.2 Truy vấn select

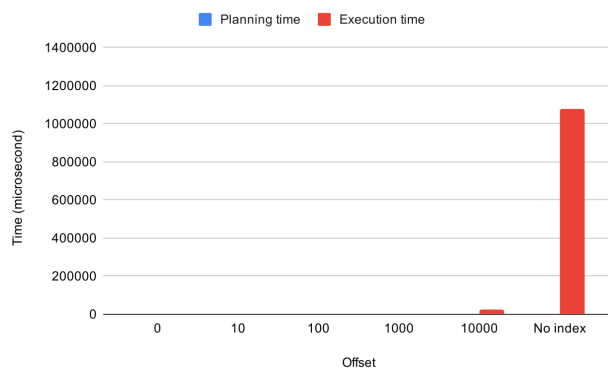


(a) linear scale

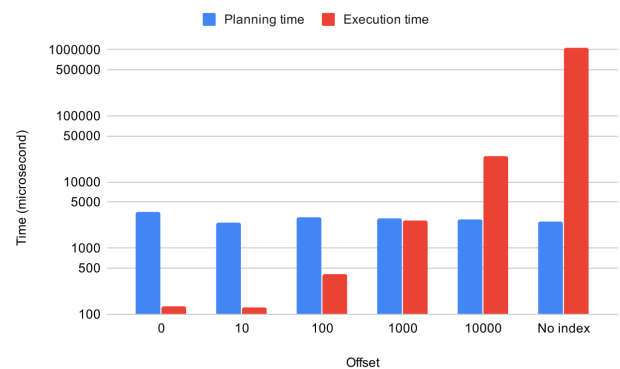


(b) log scale

Hình 5.5: Truy vấn select trên máy local



(a) linear scale

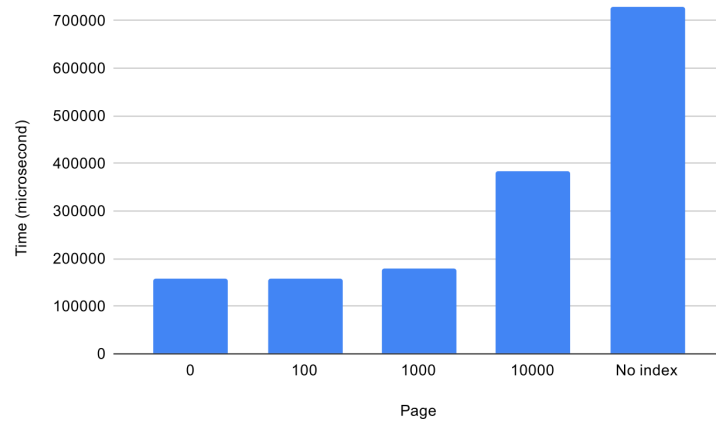


(b) log scale

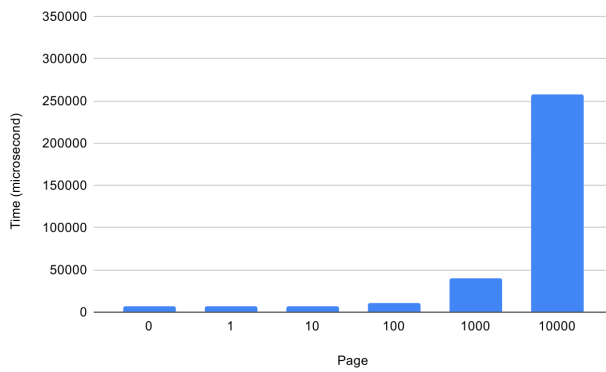
Hình 5.6: Truy vấn select trên máy cloud

Đánh giá kết quả: Có thể thấy được việc sử dụng index giúp câu lệnh select có sắp xếp được cải thiện hiệu năng hơn rất nhiều với bằng một triệu phần tử. Đồng thời việc tăng offset của lệnh select cũng làm tăng dần thời gian thực thi. Việc không sử dụng index có thể làm cho thời gian một câu lệnh select lên đến 1s.

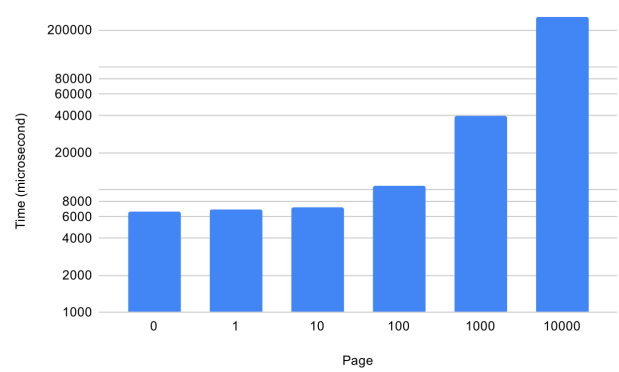
5.2.2 Kiểm thử hiệu năng Rest API



Hình 5.7: Hiệu năng Rest API trên máy local (no aggregation)

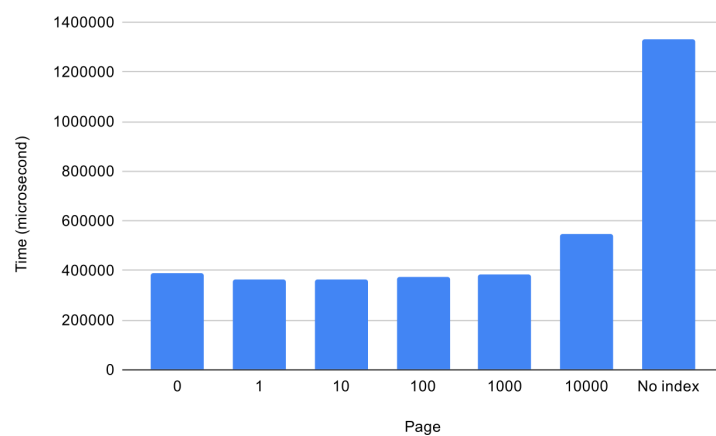


(a) linear scale

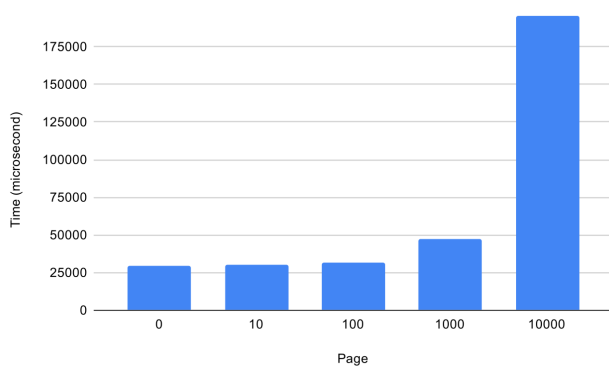


(b) log scale

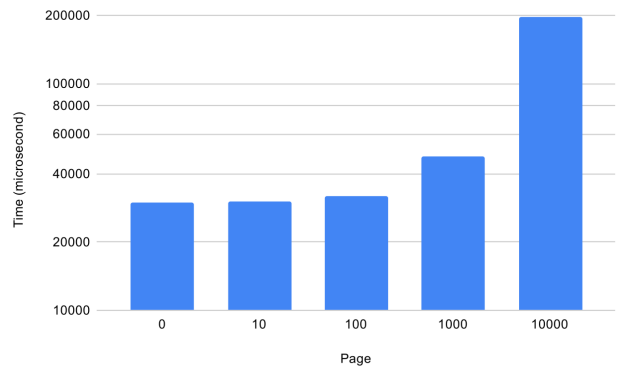
Hình 5.8: Hiệu năng Rest API trên máy local (aggregated)



Hình 5.9: Hiệu năng Rest API trên máy cloud (no aggregation)



(a) linear scale



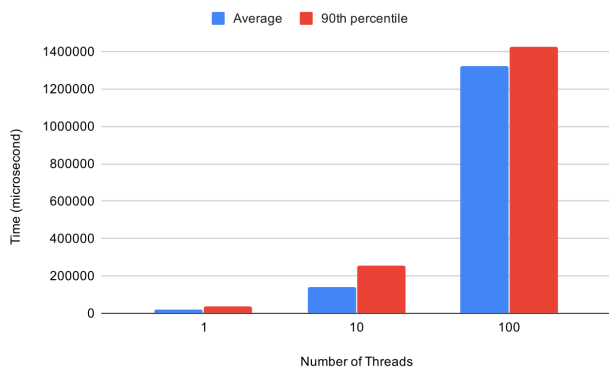
(b) log scale

Hình 5.10: Hiệu năng Rest API trên máy cloud (aggregated)

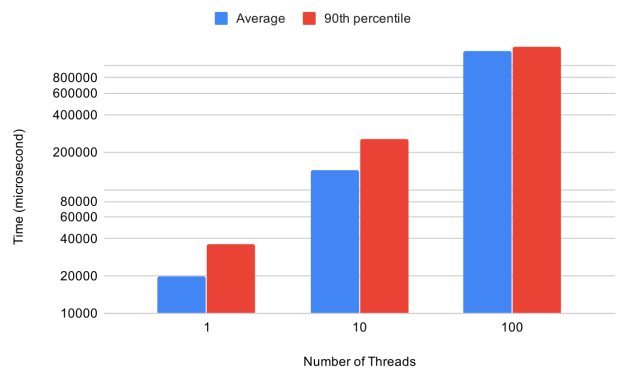
Đánh giá kết quả: Có thể thấy ảnh hưởng của việc sử dụng trực tiếp lệnh select count lên tốc độ của Rest API là rất lớn. Tốc độ chậm của lệnh select count khiến cho Rest API không thể kết thúc dưới 100ms.

5.3 Kiểm thử sức chịu tải (stress testing)

Với bài toán kiểm thử sức chịu tải, em sẽ cho đồng thời một số lượng luồng chạy và gửi request lên server và tính thời gian hoàn thành. Số lượng luồng tương ứng là 1, 10 và 100.

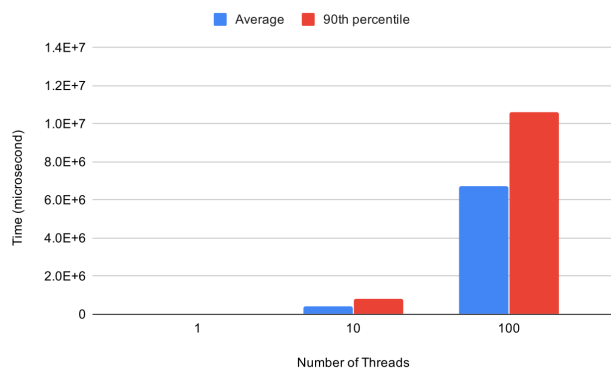


(a) linear scale

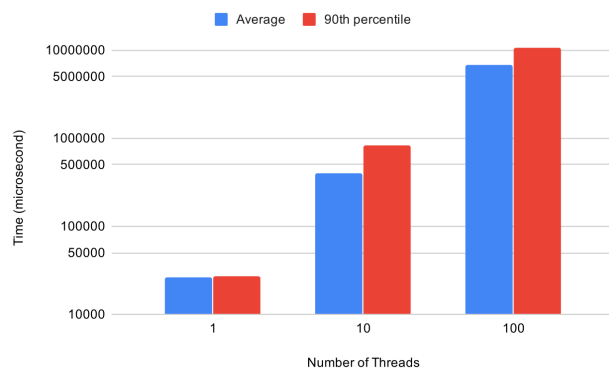


(b) log scale

Hình 5.11: Stress test thêm inventory item trên máy local



(a) linear scale



(b) log scale

Hình 5.12: Stress test thêm inventory item trên máy cloud

Đánh giá kết quả: Khi tăng số request chạy đồng thời thì thời gian chạy của các request cũng tăng lên do server chỉ có thể xử lý đồng thời được một số lượng hữu hạn các request, các request vượt giới hạn đó sẽ phải đợi request trước đó hoàn thành.

Chương 6

Kết luận và hướng phát triển

6.1 Kết luận

Qua quá trình tìm hiểu và thực hiện hệ thống quản lý phân phối, nghiên cứu quy trình và các chức năng nghiệp vụ, thiết kế giao diện, cơ sở dữ liệu thì phần mềm nhóm em tạo ra đã cơ bản hoàn thiện, đáp ứng đủ những ca sử dụng như thiết kế. Phần giao diện viết bằng ReactJS theo chuẩn Single Page Application, thiết kế theo phong cách Material Design.

Để trở thành một dự án khả thi trong thực tế thì ứng dụng cần phải khắc phục nhiều hạn chế. Như đã trình bày, hiện nay KiotViet hay Sapo là những công ty cung cấp các ứng dụng quản lý phân phối rất lớn, việc cạnh tranh một chỗ đứng với sản phẩm của họ là hạn chế đầu tiên. Phần giao diện còn tương đối sơ sài vì chưa có nhiều thời gian chỉnh sửa, việc này đòi hỏi nhiều thứ như nên thiết kế trang chủ thế nào, thiết kế logo, vị trí đầu trang và chân trang cần những thông tin như thế nào, hiệu ứng trượt dọc, dropdown, ... Ở các công ty thì họ có kỹ sư thiết kế UX/UI, một đội photoshop, một đội dựng giao diện và hiệu ứng, do đó đây là hạn chế thứ hai của nhóm em. Hạn chế tiếp theo là, dù đã sử dụng webpack để nén mã nguồn, Single Page để tối ưu tốc độ tải của trang web nhưng do dùng nhiều thư viện vẫn chưa thực sự kiểm soát được hiệu năng của front-end. Ví dụ như trang thegioididong.com có tốc độ tải trang vô cùng nhanh, vì họ có nhiều giải pháp tối ưu. Ảnh của thegioididong rất nhỏ, chỉ tầm 20-40kb, điều này cho thấy các ảnh được tối ưu rất kỹ, họ cũng không dùng bootstrap; ngoài ra trang

còn áp dụng lazy load tức khi cuộn trang xuống mới thấy các ảnh bên dưới. Thứ hai, họ xử lý CSS và JS đúng cách, không phải tải bất kì file CSS nào mà bỏ toàn bộ CSS vào head (giúp thời gian render giảm từ 378ms xuống còn 225ms). Thứ ba, họ cache mọi thứ, cụ thể các tài nguyên như ảnh, CSS, JS được cache trong vòng 1 năm, do vậy khi load lại trang trình duyệt không cần tải lại ảnh, CSS hay JS này nữa. Một hạn chế nữa, ứng dụng chưa có giải pháp xử lý trong tình huống có nhiều request lên server cùng một lúc, ví dụ như có nhiều yêu cầu lên đơn hàng đồng thời trong khi số lượng hàng trong kho lại không đủ.

Sau khi thực hiện đồ án tốt nghiệp, bản thân em nhận thấy đã có nhiều tiến bộ về cả kiến thức và kỹ năng. Em đã được thực hành nhiều công nghệ mới về web front-end, web back-end, cơ sở dữ liệu; biết được những công nghệ này là lợi thế lớn khi đi làm sau này. Học được cách xử lý tình huống khi gặp lỗi (lỗi cú pháp lập trình, lỗi logic, ...), tỉ mỉ hơn khi thiết kế giao diện, học được cách tối ưu hiệu năng front-end, thuật toán phân cụm dữ liệu. Về kỹ năng thì em đã học được về kỹ năng làm việc nhóm khi mà mỗi người làm một nhiệm vụ khác nhau, quản lý code qua các branch trên GitHub; kỹ năng tìm kiếm tài liệu trên Google, tìm đúng từ khóa, đúng nội dung cần tìm; ... Và hơn nữa sau khi làm đồ án, em đã có một sản phẩm thật, chạy được đáp ứng đủ yêu cầu về nghiệp vụ cho hệ thống quản lý phân phối trong chuỗi cung ứng.

6.2 Hướng phát triển của đồ án trong tương lai

Là một phần mềm mang tính thương mại nên em rất muốn có thể triển khai ứng dụng trong thực tế. Để làm được việc này thì trước tiên em cần hoàn thiện phần giao diện của mình, sau đó là chỉnh sửa lại các tính năng đã có cho dễ sử dụng hơn, đẹp hơn; khắc phục các hạn chế đã nêu ở trên; thiết kế thêm các tính năng mới.

Tài liệu tham khảo

- [1] Marko Luksa. *Kubernetes in Action*. Manning Publications, 1 edition, 1 2018.
- [2] Nigel Poulton. *Docker Deep Dive*. 7 2017.
- [3] Ravi S.Sandhu. Role-based access control. *Advances in Computers*, 46:237–286, 1998.