

TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI
VIỆN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG



ĐỒ ÁN TỐT NGHIỆP

Đề tài:

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

Sinh viên thực hiện: **Tạ Quang Tùng**

Giảng viên hướng dẫn: **TS. Phạm Quang Dũng**

Hà Nội, 6/2020

Tóm tắt

Trong ngày nay, việc xây dựng, phát triển các ứng dụng áp dụng công nghệ blockchain đều cần thực hiện hai tương tác với nó chính là ghi dữ liệu lên các mạng và đọc dữ liệu từ chúng. Hai công việc này đòi hỏi các nhà phát triển phải có một hiểu biết sâu rộng về công nghệ blockchain, và sẽ là một thách thức lớn hơn khi các nhà lập trình muốn sử dụng nhiều nền tảng blockchain khác nhau cho ứng dụng của mình. Cho đến nay vẫn chưa có bất kỳ một cơ chế nhất quán nào thực hiện đọc và ghi dữ liệu lên những nền tảng blockchain hiện có. Để giải quyết vấn đề này, tôi đề xuất trong đề án một ngôn ngữ ánh xạ dữ liệu được đặt tên BML. BML cho phép nhà phát triển định nghĩa một cách nhất quán ánh xạ cho việc chuyển đổi dữ liệu từ các dạng lưu trữ dữ liệu truyền thống lên trên mạng blockchain. Hơn thế nữa, BML cũng hỗ trợ người dùng đọc những dữ liệu đã được lưu trữ trên mạng. Hiện tại, BML hỗ trợ chuyển đổi dữ liệu với năm loại dữ liệu đầu vào phổ biến, bao gồm XML, JSON, XLSX, SQL (cơ sở dữ liệu quan hệ), NoSQL (cơ sở dữ liệu phi quan hệ) đến hai nền tảng blockchain khác nhau, bao gồm Ethereum và Hyperledger Sawtooth. Trong đề án này, tôi giới thiệu thiết kế và triển khai của BML, sau đó thực hiện các đánh giá hiệu quả của ngôn ngữ trên các khía cạnh về hiệu năng, tính mở rộng, độ bảo mật và phi tập trung.

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 Nguyễn Bình Minh. Thầy là người đã chỉ dẫn tận tình, cho em những kinh nghiệm quý báu để 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.

AFBDF

Mục lục

Abstract	2
Acknowledgements	3
1 Giới thiệu đề tài	9
1.1 Đặt vấn đề	9
1.2 Mục tiêu và phạm vi đề tài	10
1.3 Định hướng giải pháp	10
1.4 Bố cục đồ án	11
2 Cơ sở lý thuyết và công nghệ sử dụng	12
2.1 Giải pháp quản lý phân phối	12
2.1.1 Các khái niệm trong quản lý phân phối	12
2.1.2 Lợi ích của hệ thống quản lý phân phối	13
2.1.3 Khảo sát thực tế việc sử dụng hệ thống quản lý phân phối	14
2.2 Cơ sở lý thuyết	17
2.2.1 Role-Based Access Control	17

2.2.2 Thuật toán phân cụm dữ liệu	17
---	----

Danh sách bảng

Danh sách hình vẽ

Listings

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

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:
- Đầu ra:

Tài liệu tham khảo

- [1] Kelly Olson, Mic Bowman, James Mitchell, Shawn Amundson, Dan Middleton, and Cian Montgomery. PostgreSQL. *The Linux Foundation, Jan*, 2018.