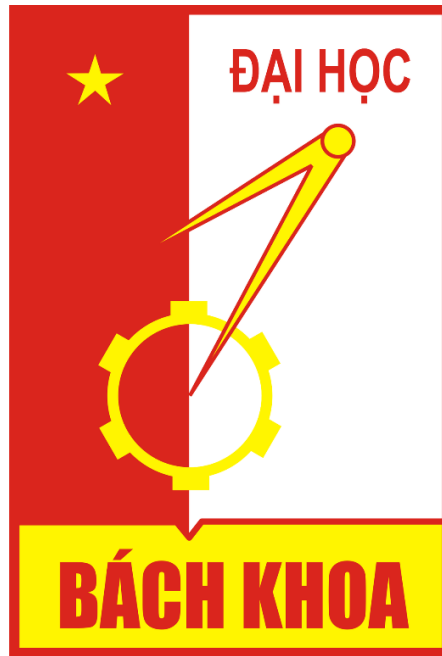


ĐẠI HỌC BÁCH KHOA HÀ NỘI

VIỆN TOÁN ỨNG DỤNG VÀ TIN HỌC



Đồ án 1

Chuyên ngành: Hệ thống thông tin quản lý
Xây dựng hệ thống báo cáo kho hàng

Giáo viên hướng dẫn: Ths. Lê Kim Thư

Sinh viên thực hiện: Lê Trọng Hoàng

Mã số sinh viên: 20206281

Lớp: Hệ thống thông tin quản lý – 02 – K65

Hà Nội - 2023

Nhận xét của giảng viên hướng dẫn

1. Mục tiêu và nội dung của đồ án

a. Mục tiêu

.....

.....

.....

b. Nội dung

.....

.....

.....

2. Kết quả đạt được

.....

.....

.....

3. Ý thức làm việc của sinh viên

.....

.....

.....

Hà nội, tháng ... năm 2023

Giảng viên

Ths. Lê Kim Thư

MỤC LỤC

LỜI CẢM ƠN	4
LỜI MỞ ĐẦU	5
Danh mục hình ảnh	6
Danh mục cụm từ viết tắt	7
I. Cơ sở lý thuyết	8
1. Tổng quan về kho dữ liệu (Data Warehouse)	8
1.1. Khái niệm	8
1.2. Mô hình dữ liệu OLAP	9
1.3. So sánh OLAP và OLTP	12
2. Tổng quan về kinh doanh thông minh (Business Intelligence)	13
2.1. Khái niệm	13
2.2. Thành phần cơ bản của hệ thống kinh doanh thông minh	14
2.3. Tiềm ích từ các ứng dụng BI	15
II. Khảo sát hệ thống	16
1. Hệ thống hiện tại	16
2. Quy trình nghiệp vụ	16
2.1. Quy trình hoạt động của kho hàng	16
2.2. Luồng dữ liệu	18
3. Kiến trúc mô hình	18
4. Yêu cầu hệ thống mới	18
4.1. Chức năng cơ bản	19
4.2. Chức năng phân tích báo cáo	19

III.	Phân tích và thiết kế hệ thống.....	20
1.	Kiến trúc của hệ thống phân tích.....	20
2.	Thiết kế vùng xử lý dữ liệu – Staging	20
3.	Xây dựng Dimension & Fact.....	24
3.1.	Dimension.....	24
3.2.	Fact	24
4.	Thiết kế cơ sở dữ liệu hệ thống OLTP	25
5.	Thiết kế cơ sở dữ liệu hệ thống OLAP.....	26
6.	Khai phá dữ liệu.....	29
6.1.	Theo khu vực	29
6.2.	Theo doanh thu	30
6.3.	Theo lợi nhuận.....	31
6.4.	Theo thời gian.....	32
IV.	Xây dựng hệ thống.....	33
1.	Truyền dữ liệu.....	33
2.	Xây dựng hệ thống Power BI	34
3.	Tổng kết	34
V.	Tài liệu tham khảo	35

LỜI CẢM ƠN

Trước khi đi vào bài báo cáo, lời đầu tiên cho em xin chân thành cảm ơn sự hỗ trợ và hướng dẫn của cô Lê Kim Thư, giảng viên Bộ môn Toán Tin tại viện Toán Ứng Dụng và Tin Học, Đại học Bách Khoa Hà Nội. Cảm ơn cô đã luôn tận tâm tư vấn, hướng dẫn và luôn sẵn sàng hỗ trợ em trong quá trình thực hiện đồ án này. Qua quá trình làm đồ án với cô em đã nhận ra được nhiều điều mới mẻ và rất nhiều kỹ năng có ích cho hướng đi của em sau này. Tất cả những điều đó đã góp phần giúp đỡ em hoàn thành được đồ án theo ý tưởng ban đầu và đáp ứng được những yêu cầu đặt ra.

LỜI MỞ ĐẦU

Trong thời đại kỹ thuật số hiện nay, dữ liệu là tài nguyên quan trọng và có giá trị lớn đối với các doanh nghiệp, tổ chức và cá nhân. Với số lượng dữ liệu khổng lồ được tạo ra hàng ngày, việc phân tích và trích xuất thông tin từ dữ liệu đã trở thành một thách thức lớn, điều đó đòi hỏi sự ra đời của các hệ thống hỗ trợ nghiệp vụ phân tích, trích xuất và tạo báo cáo. Từ đó cho nhiều góc nhìn hơn về dữ liệu giúp cho những đánh giá, quyết định của chúng ta trở nên đúng đắn hơn.

Tóm tắt nội dung:

Chương 1: Cơ sở lý thuyết

- Giới thiệu chung về dữ liệu, các khái niệm về kho dữ liệu, kinh doanh thông minh, cơ sở dữ liệu OLTP-OLAP. So sánh ưu nhược của từng loại cơ sở.

Chương 2: Khảo sát hệ thống

- Phân tích các nghiệp vụ của hệ thống cũ, các kiến trúc, luồng dữ liệu cũ từ đó đưa ra gợi ý cho hệ thống mới.

Chương 3: Phân tích và thiết kế hệ thống

- Phân tích, đưa ra các kiến trúc của hệ thống mới, các bước xử lý dữ liệu nguồn, phân tích dữ liệu theo các chiều và đưa ra phương hướng thiết kế.

Chương 4: Xây dựng hệ thống

- Truyền dữ liệu vào mô hình, xây dựng báo cáo trên công cụ Power BI

Trong phần tiếp theo của đồ án, em sẽ giải thích các thành phần chính của hệ thống, cách thức hoạt động và các ứng dụng của nó. Ngoài ra, em cũng sẽ trình bày về quá trình thiết kế, triển khai và kiểm thử của hệ thống. Mong rằng đồ án này của em sẽ giúp ích trong việc tìm hiểu về phân tích dữ liệu và các công nghệ liên quan.

Trong quá trình làm đồ án, em đã cố gắng, kiểm tra cẩn thận, hoàn thiện mọi chi tiết nhất có thể nhưng trong quá trình làm đồ án có thể xảy ra những sai sót không đáng có mà em không thể nhận ra. Do đó, đồ án không thể tránh khỏi có những sai sót. Hy vọng quý thầy cô và các bạn đọc thông cảm và lượng thứ.

Cuối cùng em rất mong nhận được sự góp ý, nhận xét của thầy và các bạn, để giúp đồ án của em ngày càng đầy đủ và hoàn thiện hơn.

Em xin chân thành cảm ơn.

Danh mục hình ảnh

Hình 1.1 Kiến trúc kho dữ liệu.....	9
Hình 1.2 Lược đồ bông tuyết	11
Hình 1.3 Lược đồ ngôi sao	12
Hình 1.4 Quy trình kinh doanh thông minh	14
Hình 2.1 Sơ đồ quy trình hoạt động	17
Hình 2.2 Sơ đồ luồng dữ liệu	18
Hình 3.1 Kiến trúc của hệ thống phân tích	20
Hình 3.2 File dữ liệu sau khi xử lý bỏ null	21
Hình 3.3 File dữ liệu sau khi xóa bỏ những cột không dùng đến	21
Hình 3.4 File sau khi được định dạng lại kiểu dữ liệu	22
Hình 3.5 File sau khi dữ liệu được sắp xếp lại	22
Hình 3.6 Tạo thêm trường lợi nhuận	23
Hình 3.7 Tạo thêm trường phần trăm lợi nhuận	23
Hình 3.8 Sơ đồ hệ thống OLTP	25
Hình 3.9 Sơ đồ hệ thống OLAP	26
Hình 3.10 Doanh thu và lợi nhuận theo khu vực	29
Hình 3.11 Doanh thu bán hàng	30
Hình 3.12 Lợi nhuận bán hàng	31
Hình 3.13 Thông tin sản phẩm	31
Hình 3.14 Doanh thu và lợi nhuận theo thời gian	32
Hình 4.1 Dashboard quản lý dữ liệu kho hàng	34

Danh mục cụm từ viết tắt

Số thứ tự	Cụm từ	Viết tắt
1	Data warehouse	DW
2	Online analytical processing	OLAP
3	Multidimensional online analytical processing	MOLAP
4	Relational online analytical processing	ROLAP
5	Hybrid online analytical processing	HOLAP
6	Online transactional processing	OLTP
7	Extract – Load – Transform	ELT
8	Business Intelligence	BI
9	Structured Query Language	SQL

I. Cơ sở lý thuyết

1. Tổng quan về kho dữ liệu (Data Warehouse)

1.1. Khái niệm

Data warehouse (Kho dữ liệu) là một hệ thống lưu trữ dữ liệu từ nhiều nguồn, nhiều môi trường khác nhau như: phần mềm bán hàng, kế toán, nhân sự hay hệ thống lõi ngân hàng,... giúp tăng cường hiệu suất của các truy vấn cho báo cáo và phân tích.

Data Warehouse hoạt động như một kho lưu trữ trung tâm. Dữ liệu đi vào kho dữ liệu từ hệ thống giao dịch và các cơ sở dữ liệu liên quan khác. Sau đó, dữ liệu được xử lý, chuyển đổi để người dùng có thể truy cập những dữ liệu này thông qua công cụ Business Intelligence, SQL client hay bảng tính.

Một kho dữ liệu thường bao gồm các yếu tố như:

- Một cơ sở dữ liệu quan hệ để lưu trữ và quản lý dữ liệu.
- Giải pháp trích xuất, tải và biến đổi ETL để chuẩn bị dữ liệu cho phân tích.
- Khả năng phân tích thống kê, báo cáo và khai thác dữ liệu.
- Các công cụ phân tích khách hàng để trực quan hóa và trình bày dữ liệu cho người dùng doanh nghiệp.
- Các ứng dụng phân tích khác, phức tạp hơn tạo ra thông tin có thể hành động bằng cách áp dụng khoa học dữ liệu và thuật toán trí tuệ nhân tạo AI hoặc các tính năng đồ thị và không gian cho phép nhiều loại phân tích dữ liệu hơn trên quy mô lớn.

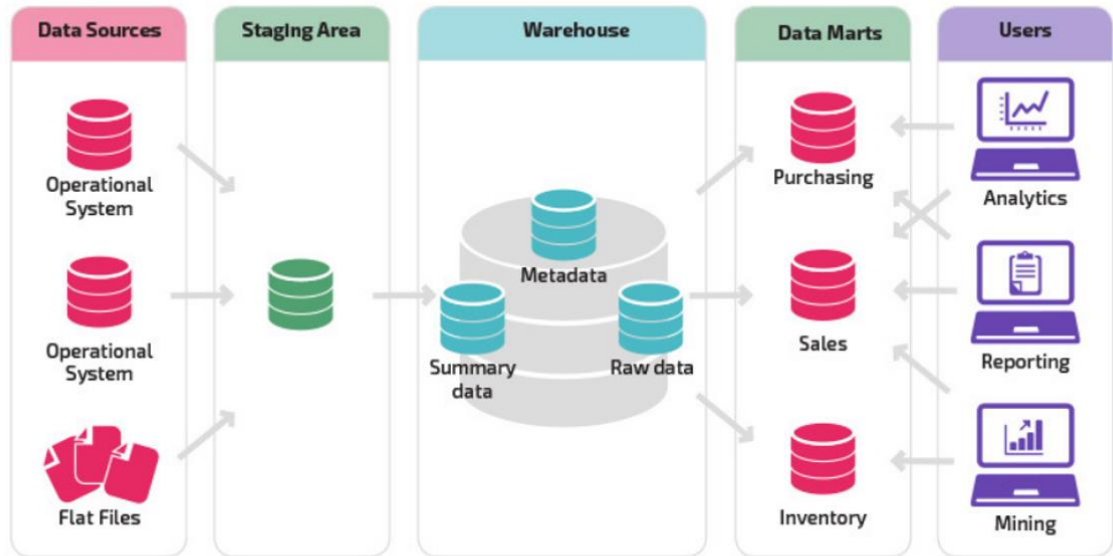
Kho dữ liệu không phải là một sản phẩm mà là một môi trường, cung cấp một mô hình kiến trúc cho luồng thông tin hỗ trợ từ hệ thống vận hành đến các môi trường hỗ trợ quyết định. Do đó kho dữ liệu sẽ có những đặc tính sau:

- Dữ liệu toàn vẹn (Integration)
- Dữ liệu gắn thời gian và có tính lịch sử (Isolation)
- Dữ liệu vận hành ổn định (Nonvolatility)
- Dữ liệu cố định (Durable)
- Dữ liệu được hướng theo chủ đề (Subject orientation)

Kiến trúc đầy đủ của một kho dữ liệu đầy đủ bao gồm:

- Data Sources (Nguồn dữ liệu): Là nơi thu thập, tích hợp dữ liệu từ nhiều nguồn khác nhau.
- Staging Area (Khu vực xử lý): Thực hiện chuẩn hóa dữ liệu, làm sạch, đưa dữ liệu về chung một định dạng và lưu dữ liệu đã tổng hợp.

- Warehouse (Kho): Nơi lưu trữ dữ liệu đã được xử lý.
- Data Marts (Kho dữ liệu): Chia dữ liệu thành các chủ đề nhỏ hơn.
- Users (Người dùng): Nơi thực hiện các thao tác truy vấn, báo cáo, phân tích để tìm ra xu hướng, trung bình, tổng hợp...



Hình 1.1 Kiến trúc kho dữ liệu

1.2. Mô hình dữ liệu OLAP

OLAP - Online analytical processing (Xử lý phân tích trực tuyến) là một phần mềm tính toán cho phép người dùng trích xuất và truy vấn dữ liệu một cách chọn lọc và dễ dàng. OLAP là dữ liệu đa chiều, điều này có nghĩa là tất cả thông tin có thể được so sánh theo nhiều cách thức khác nhau.

Các truy vấn thông minh kinh doanh OLAP thường hỗ trợ trong báo cáo tài chính, phân tích xu hướng, dự báo bán hàng, lập ngân sách và các mục đích lập kế hoạch khác. Để thực hiện điều đó ta cần dữ liệu được thu thập từ nhiều nguồn dữ liệu và được lưu trữ trong kho dữ liệu, sau đó được "làm sạch" và sắp xếp thành các khối dữ liệu.

Một khối OLAP sẽ chứa dữ liệu được phân loại theo các thứ nguyên (khách hàng, khu vực bán hàng địa lý và khoảng thời gian) được dẫn xuất bởi các bảng thứ nguyên có sẵn trong kho dữ liệu. Các thứ nguyên sau đó được điền bởi các thành phần như tên khách hàng, quốc gia và tháng) sắp xếp theo thứ tự nhất định. Thông thường các khối OLAP sẽ được tóm tắt trước theo thứ nguyên để cải thiện đáng kể thời gian truy vấn trên cơ sở dữ liệu quan hệ.

Hệ thống OLAP được phân thành 3 loại chính:

- **MOLAP (Multidimensional OLAP):** OLAP hoạt động trực tiếp với khối OLAP đa chiều được gọi là OLAP đa chiều hay MOLAP. Đối với hầu hết mọi mục đích sử dụng, MOLAP là loại phân tích dữ liệu đa chiều nhanh nhất và thiết thực nhất.
- **ROLAP (Relational OLAP):** ROLAP hay OLAP quan hệ, là loại phân tích trực tuyến (OLAP) các mô hình dữ liệu đa chiều. Điểm khác biệt giữa các ROLAP và OLAP là nó truy cập dữ liệu lưu trữ ngay trong cơ sở dữ liệu quan hệ thay vì dữ liệu cơ sở đa chiều (cơ sở dữ liệu được sử dụng phổ biến trong các OLAP). Thêm vào đó, nó cũng có thể tạo ra các truy vấn SQL với mục đích thực hiện các phép tính khi người dùng cuối muốn như vậy.
- **HOLAP (Hybrid OLAP):** HOLAP hay OLAP kết hợp là sự kết hợp giữa ROLAP (xử lý phân tích trực tuyến quan hệ) và MOLAP (xử lý phân tích trực tuyến đa chiều). HOLAP cung cấp lợi thế từ cả hai quy trình ROLAP và MOLAP do nó hỗ trợ cho cả hai định dạng lưu trữ. HOLAP để giải quyết câu hỏi : " cái nào tốt hơn?" bằng việc kết hợp khả năng xử lý của MOLAP và dung lượng dữ liệu của MOLAP. Ngoài ra, do có kiến trúc phức tạp và phải lưu trữ, xử lý tất cả cơ sở dữ liệu từ MOLAP và ROLAP nên HOLAP sẽ yêu cầu cập nhật và bảo trì thường xuyên hơn.

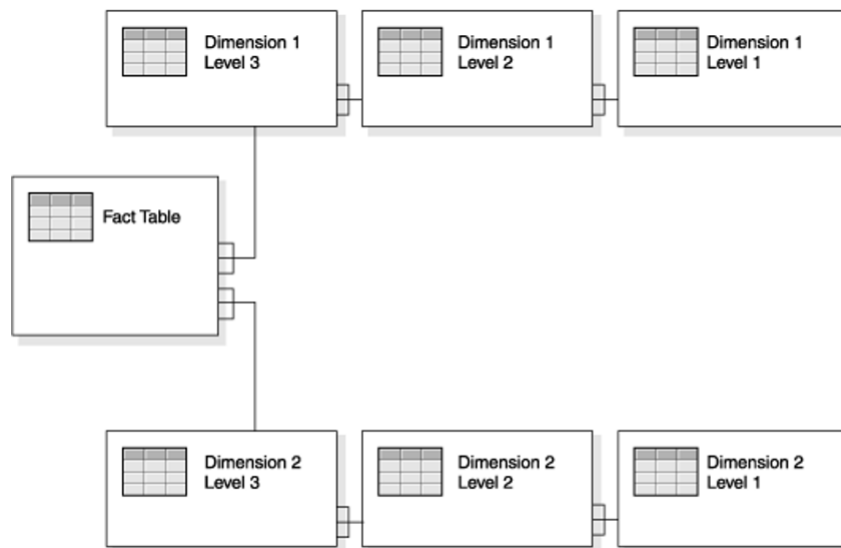
Mô hình chiều dữ liệu:

- **Dimension:** Các bảng dimension được sử dụng để mô tả dữ liệu mà chúng ta muốn lưu trữ. Ví dụ: một nhà bán lẻ muốn lưu trữ thời gian, cửa hàng, và nhân viên tham gia vào một hoá đơn. Mỗi một bảng dimension là một danh mục của chính nó (ngày tháng, nhân viên, cửa hàng) và có thể có một hoặc nhiều thuộc tính (attributes). Với mỗi một cửa hàng, chúng ta lưu chúng các thông tin như vị trí trong thành phố, vùng miền, tỉnh thành và quốc gia. Mỗi một ngày tháng chúng ta lưu năm, tháng, ngày trong tháng, ngày trong tuần...Điều này liên quan đến sự phân cấp của các thuộc tính trong bảng dimension.
- **Fact:** Bảng Fact chứa dữ liệu mà chúng ta muốn thêm vào reports, tổng hợp trên các giá trị trong các bảng dimension. Một bảng fact chỉ có các cột lưu giá trị và các cột khoá ngoại tham chiếu đến bảng dimensions. Kết hợp tất cả các khoá ngoại và khoá chính trong bảng fact. Ví dụ, một bảng fact

có thể lưu trữ một số lượng các hợp đồng và số lượng các nhân viên bán hàng từ các danh sách hợp đồng.

Một số mô hình đơn giản:

- **Lược đồ bông tuyết (Snow flake schema):** là một bảng dữ kiện được liên kết với nhiều bảng dimension, trong đó có một hoặc nhiều bảng dimension không kết nối trực tiếp với bảng dữ kiện mà phải kết nối thông qua các bảng dimension khác. Các bảng trong Snowflake Schema thường được chuẩn hóa thành dạng chuẩn thứ ba. Mỗi bảng dimension thực hiện chính xác một cấp trong hệ thống phân cấp.
 - **Ưu điểm:**
Kích thước các bảng Dimension giảm, nâng cao tốc độ truy vấn Cho phép thực hiện các truy vấn phức tạp theo các chiều Dễ dàng thiết lập và bảo trì.
 - **Nhược điểm:**
Số lượng các bảng cần quản lý tăng Truy vấn xâu cần sự kết nối giữa nhiều bảng gây giảm hiệu năng.

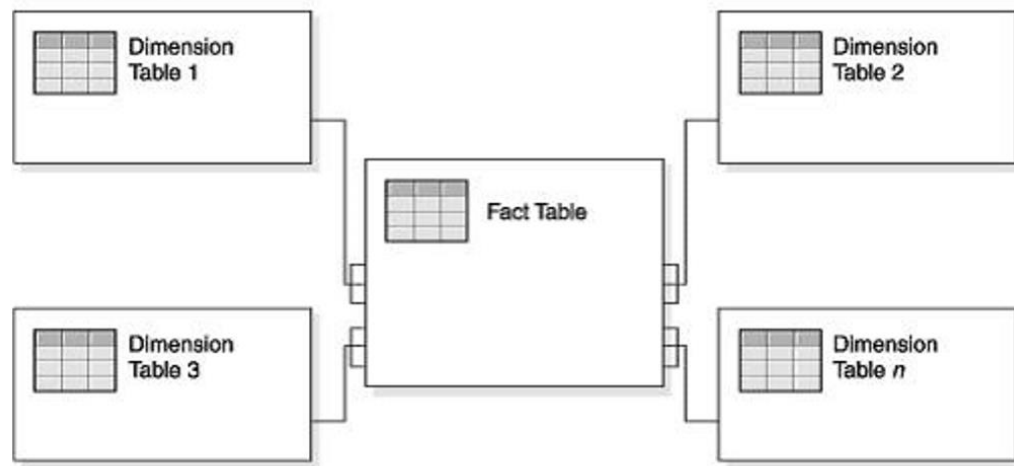


Hình 1.2 Lược đồ bông tuyết

- **Lược đồ hình sao (Star schema):** là một kiểu mô hình cơ sở dữ liệu quan hệ, trong đó dữ liệu được tổ chức theo dạng sao, tức là có một bảng trung tâm (fact table) và các bảng chi tiết (dimension tables) được kết nối với bảng trung tâm. Bảng trung tâm: Được thiết kế để lưu trữ các số liệu cơ bản của hệ thống, bảng này thường chứa các khóa ngoại kết nối tới các

bảng chi tiết. Bảng chi tiết: Chứa các thông tin mô tả về các khía cạnh của dữ liệu trong bảng trung tâm.

- **Ưu điểm:**
Cải tiến hiệu năng, thời gian truy vấn nhanh Có ít bảng và cấu trúc đơn giản.
- **Nhược điểm:**
Do có ít bảng nên lượng thông tin trong bảng là lớn, gây dư thừa
Các mối quan hệ không được thể hiện quá rõ ràng.



Hình 1.3 Lược đồ ngôi sao

1.3. So sánh OLAP và OLTP

Yếu tố	OLAP	OLTP
Người dùng	<ul style="list-style-type: none"> - Người ra quyết định - Nhóm người quản lý được chỉ định 	<ul style="list-style-type: none"> - Nhóm nhân viên, khách hàng - Có thể truy cập bởi rất nhiều người
Chức năng	<ul style="list-style-type: none"> - Phân tích đa chiều, khai thác dữ liệu - Phân tích phức tạp, báo cáo kinh doanh 	<ul style="list-style-type: none"> - Xử lý giao dịch gần đây - Nhanh chóng và chính xác
Bản chất	Cung cấp thông tin tóm tắt, tổng hợp cho người truy vấn dữ liệu	Ghi lại các bước thao tác

Thiết kế	Theo hướng chủ đề và xem thông tin dưới dạng đa chiều	Theo hướng ứng dụng và xem bản ghi dưới dạng tập hợp bảng
Dữ liệu	Cần thông tin dữ liệu trong vài năm	Cần trạng thái hiện tại của dữ liệu
Loại sử dụng	Thường ít được cập nhật dữ liệu	Thao tác đọc và ghi
Chế độ xem	Tổng hợp dữ liệu từ nhiều nguồn, trong quá khứ	Tập trung vào dữ liệu hiện tại của tổ chức
Các mẫu truy cập	Chủ yếu là các hoạt động đọc, nghiên cứu	Các giao dịch ngắn

2. Tổng quan về kinh doanh thông minh (Business Intelligence)

2.1. Khái niệm

Hệ thống kinh doanh thông minh (Business Intelligence) là một quy trình dựa trên công nghệ để phân tích dữ liệu và cung cấp thông tin có thể hành động giúp các giám đốc điều hành, người quản lý và người lao động đưa ra các quyết định kinh doanh sáng suốt. Là một phần của quy trình BI, các tổ chức thu thập dữ liệu từ các hệ thống công nghệ thông tin nội bộ và các nguồn bên ngoài, chuẩn bị cho phân tích, chạy các truy vấn dựa trên dữ liệu và tạo trực quan hóa dữ liệu, bảng điều khiển BI và báo cáo để cung cấp kết quả phân tích cho người dùng doanh nghiệp để đưa ra quyết định hoạt động- lập và hoạch định chiến lược

Mục tiêu cuối cùng của các sáng kiến BI là thúc đẩy các quyết định kinh doanh tốt hơn cho phép các tổ chức tăng doanh thu, cải thiện hiệu quả hoạt động và đạt được lợi thế cạnh tranh so với các đối thủ kinh doanh. Để đạt được mục tiêu đó, BI kết hợp sự kết hợp của các công cụ phân tích, quản lý dữ liệu và báo cáo, cùng với nhiều phương pháp khác nhau để quản lý và phân tích dữ liệu.

Các chương trình BI thường kết hợp các hình thức phân tích nâng cao, chẳng hạn như khai thác dữ liệu, phân tích dự đoán, khai thác văn bản, phân tích thống kê và phân tích dữ liệu lớn. Một ví dụ phổ biến là mô hình dự đoán cho phép phân tích điều gì xảy ra nếu xảy ra các tình huống kinh doanh khác nhau. Tuy nhiên, trong hầu hết các trường hợp, các dự án phân tích nâng cao được thực hiện bởi các nhóm riêng biệt gồm các nhà khoa học dữ liệu, nhà thống kê, nhà

lập mô hình dự đoán và các chuyên gia phân tích lãnh nghề khác, trong khi nhóm BI giám sát việc truy vấn và phân tích dữ liệu kinh doanh đơn giản hơn.

2.2. Thành phần cơ bản của hệ thống kinh doanh thông minh



Hình 1.4 Quy trình kinh doanh thông minh

- **Data Sources (Nguồn dữ liệu):** Trong tầng đầu tiên của thành phần kiến trúc hệ thống BI, cần phải tập hợp và tích hợp các dữ liệu được chứa trong nhiều nguồn trực tiếp và nguồn gián tiếp không đồng nhất về xuất xứ và loại.
- **Data Warehouse (Kho dữ liệu):** Chỗ chứa trước tiên nhất cho việc phát triển kiến trúc của hệ BI. Khối dữ liệu là các hệ thống thu thập tất cả các dữ liệu yêu cầu bởi một phòng ban nào đó của công ty như tiếp thị, đánh giá, cho mục đích phân tích một vài chức năng của hệ thống BI.
- **Data Mining (Khai thác dữ liệu):** Đây là phần rất quan trọng trong hệ thống BI, là các phần sẽ biến đổi từ dữ liệu thô, khai thác những thông tin cần thiết để đưa ra và hỗ trợ trong việc ra quyết định. Bao gồm các kỹ thuật trích xuất thông tin, tri thức từ tập dữ liệu, gồm cả các mô hình toán học cho việc nhận dạng mẫu, học máy và các kỹ thuật của khai phá dữ liệu.
- **Data Analysis Reporting (Báo cáo phân tích dữ liệu):** Thành phần tối ưu hóa cho phép xác định giải pháp tốt nhất từ tập hợp các hành động liên quan. Tập các hành động này có thể rất rộng và đôi khi không xác định. Từ đó tạo nên các báo cáo phân tích theo các phía cạnh.

- **Decision (Quyết định):** Việc lựa chọn và thực thi phương thức quyết định nào đó dựa trên sự tính toán, so sánh đối chiếu của các phương thức toán học. Tuy nhiên, mặc dù cách thức lựa chọn được thông qua do cách thức toán học, việc quyết định theo hướng nào đó lại phụ thuộc vào người ra quyết định.

2.3. Tiềm ích từ các ứng dụng BI

Business Intelligence giúp cho các doanh nghiệp sử dụng thông tin một cách hiệu quả, chính xác để thích ứng với môi trường thay đổi liên tục và cạnh tranh khốc liệt trong kinh doanh:

- Giúp đánh giá, cải thiện và tối ưu hóa quy trình của tổ chức
- Xác định các vị thế và khả năng cạnh tranh của doanh nghiệp
- Phân tích hành vi khách hàng, xác định các mục tiêu chiến lược
- Tìm ra các vấn đề phát sinh ảnh hưởng tới doanh thu
- Đưa ra gợi ý, thông tin, cải thiện quá trình nghiên cứu, ra quyết định
- Khả năng trực quan hóa dữ liệu theo các góc nhìn khác nhau
- Hỗ trợ tối đa công tác quản lý, tiết kiệm thời gian, quản trị nhân lực

Một số công cụ BI hiện hành:

- Power BI
- Tableau
- Datapine
- Sisense
- Yellowfin BI

Khả năng mô hình hóa dữ liệu (Data Modeling):

Mô hình hóa dữ liệu là quá trình tạo ra một biểu diễn trực quan của toàn bộ hệ thống thông tin hoặc các bộ phận của nó để giao tiếp các kết nối giữa các điểm và cấu trúc dữ liệu. Mục đích là minh họa các loại dữ liệu được sử dụng và lưu trữ trong hệ thống, mối quan hệ giữa các loại dữ liệu này, cách dữ liệu có thể được nhóm và tổ chức cũng như các định dạng và thuộc tính của nó.

Gồm các bước:

- Xác định các thực thể, thuộc tính thực thể
- Xác định mối quan hệ giữa các thực thể
- Xác định các khóa
- Xác thực và hoàn thiện mô hình

II. Khảo sát hệ thống

1. Hệ thống hiện tại

Cửa hàng vẫn đang còn dùng các công cụ google sheet, excel...để lưu trữ các dữ liệu, việc tổng hợp dữ liệu gặp nhiều bất cập do còn chia thành nhiều trang để thu thập, lưu trữ. Do vậy, có rất nhiều dữ liệu bị trùng lặp, lưu ở nhiều nơi khiến việc quản lý trở nên khó khăn.

Các mẫu báo cáo, thống kê mới chỉ dùng ở mức độ hiển thị chi phí, doanh thu, nhập, bán dưới dạng bảng. Chưa có sự thể hiện trên biểu đồ hay thể hiện rõ ràng được xu thế của cửa hàng, dẫn đến chưa có cái nhìn tổng quan về hoạt động kinh doanh của cửa hàng.

Chưa hỗ trợ truy xuất sâu, hiển thị thống kê theo các chiều, các nguồn sản phẩm. Dẫn đến sự thiếu thông tin về các điểm mạnh, điểm yếu của cửa hàng hoặc thông tin mới ở mức thấp, chưa xác định rõ được các yếu tố.

Thao tác cập nhật, chỉnh sửa thông tin còn gặp nhiều hạn chế do hệ thống chưa hỗ trợ phân quyền mà chỉ dùng ở mức dùng chung một tài khoản chủ. Điều đó tiềm ẩn nguy cơ bị chỉnh sửa, xóa nhầm thông tin hoặc tệ hơn là gian lận thông tin.

Vấn đề bảo mật của hệ thống còn thấp, dễ bị các đối tượng xấu thâm nhập, chỉnh sửa, sao lưu dữ liệu hoặc tệ hơn là phá hoại hệ thống. Gây ảnh hưởng trực tiếp đến việc vận hành, quản lý của cửa hàng.

2. Quy trình nghiệp vụ

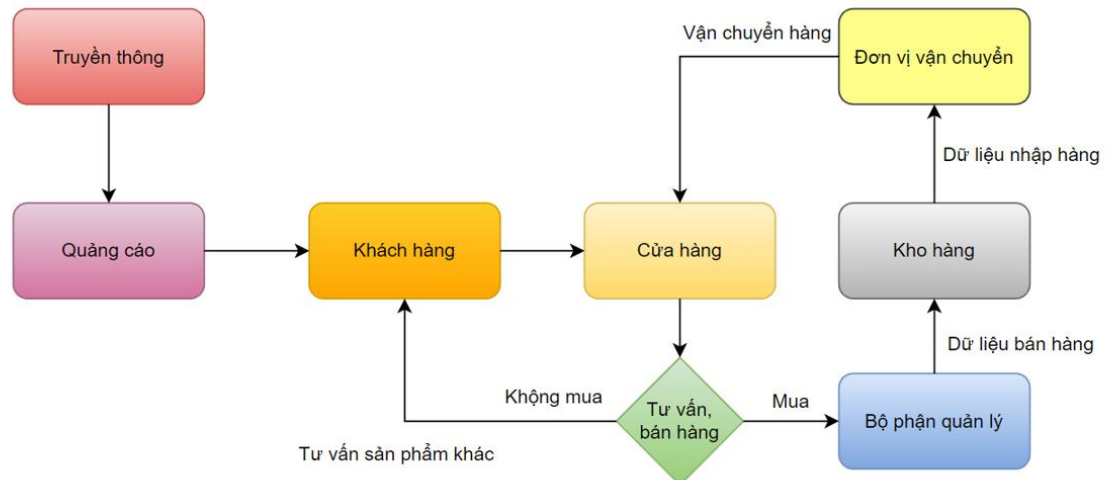
2.1. Quy trình hoạt động của kho hàng

Quy trình hoạt động của một cửa hàng có 3 bộ phận chính:

- Bộ phận truyền thông: Có nhiệm vụ lên các kế hoạch, chiến dịch quảng cáo, quảng bá các sản phẩm trên các nền tảng nhằm thu về sự chú ý của khách hàng. Từ đó tạo ra số lượng khách hàng mới cho cửa hàng.
- Bộ phận bán hàng: Có nhiệm vụ liên hệ, trả lời câu hỏi của khách hàng, cung cấp thông tin sản phẩm mà khách hàng yêu cầu. Xin thông tin của khách hàng để tạo đơn mua và chốt số lượng khách mua.
- Bộ phận quản lý: Có nhiệm vụ giám sát, quản lý các quy trình bán hàng. Theo dõi chi phí, doanh thu bán hàng. Đánh giá sản phẩm nào bán chạy hoặc ít được thị trường tiêu thụ. Đưa ra các đề xuất phương án cải thiện chất lượng sản phẩm và lên kế hoạch thúc đẩy doanh số cho sản phẩm.

Báo cáo số lượng hàng hóa bán được và tồn kho. Kiểm soát hàng hóa, tài sản của cửa hàng tránh gây thâm hụt ngân sách.

Sơ đồ quy trình hoạt động:



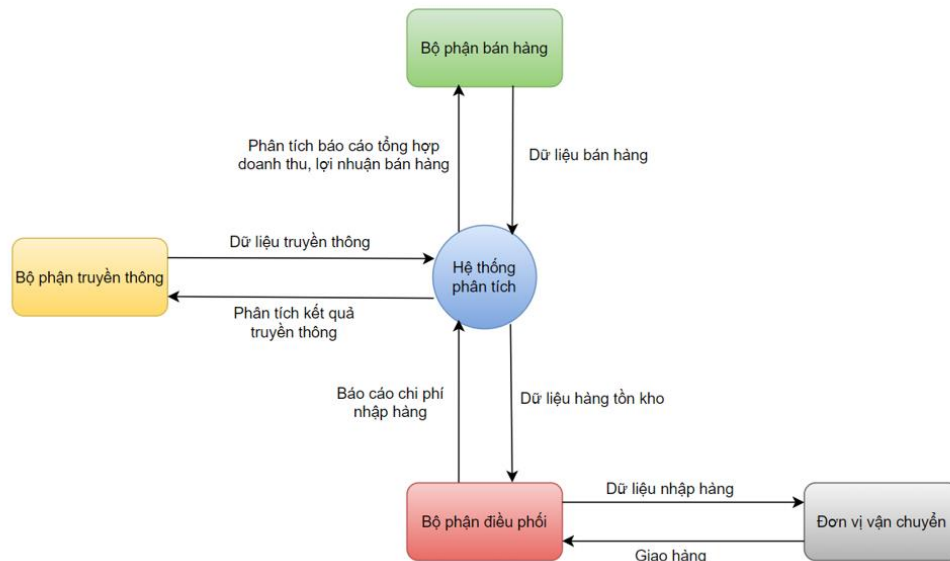
Hình 2.1 Sơ đồ quy trình hoạt động

Chi tiết quy trình:

- B1: Bộ phận truyền thông lên kế hoạch quảng bá các sản phẩm tạo ra số lượng khách hàng cho cửa hàng.
- B2: Bộ phận bán hàng sẽ trao đổi, liên hệ, tư vấn cho khách hàng và chốt đơn.
 - Không mua: Trao đổi với khách hàng về mặt hàng khác rồi thực hiện lại
 - Mua: Điền thông tin đơn hàng vào file excel, chuyển đến bước tiếp theo
- B3: Bộ phận quản lý tổng hợp chi phí, doanh thu, lợi nhuận của từng cửa hàng theo ngày, tháng, năm.
 - Đánh giá ra sản phẩm nào bán chạy
 - Đưa ra phương án thúc đẩy bán hàng và cải thiện sản phẩm ít được tiêu thụ
 - Kiểm soát số lượng hàng bán được và tồn kho để nhập hàng mới

2.2. Luồng dữ liệu

Sơ đồ luồng dữ liệu:



Hình 2.2 Sơ đồ luồng dữ liệu

3. Kiến trúc mô hình

Hệ thống đang sử dụng kiến trúc hai tầng dữ liệu được xây dựng dựa trên mô hình ứng dụng Client – Sever. Tồn tại giao tiếp trực tiếp giữa máy Client và Data Source do vậy dữ liệu được tổng hợp bằng cách truy vấn trực tiếp trên các bảng từ kho dữ liệu.

Điều này khiến kiến trúc có một số ưu điểm như:

- Dễ dàng xây dựng, bảo trì ...
- Có thể trực tiếp chỉnh sửa dữ liệu lưu trữ trong kho

Tuy nhiên, kiến trúc này cũng tồn tại nhiều nhược điểm như:

- Thời gian tạo, xuất kết quả là lớn thậm chí lỗi nếu lượng dữ liệu trong kho là quá lớn
- Tiềm ẩn nguy cơ dữ liệu lưu trữ bị chỉnh sửa cho thao tác nhằm, trung lập
- Khả năng bảo mật kém

4. Yêu cầu hệ thống mới

Một hệ thống mới phải đáp ứng được những yêu cầu cơ bản của các chức năng quản lý, khả năng tạo báo cáo, thống kê doanh thu cũng như truy xuất

thông tin số lượng hàng hóa của cửa hàng. Ngoài ra, hệ thống còn phải được thiết kế có thẩm mỹ, cách vận hành đơn giản.

4.1. Chức năng cơ bản

Hệ thống phục vụ nhóm đối tượng chính là quản lý và nhân viên bán hàng, các đối tượng đều được cấp username và password để đăng nhập vào hệ thống.

- Quản lý: Có quyền cao nhất, kiểm soát mọi hoạt động của hệ thống. Ngoài ra còn có thể thực hiện các chức năng:
 - Chức năng xem thêm xóa sửa danh mục, sản phẩm, chi phí, giá thành...
 - Phân quyền cho các tài khoản khác
 - Truy xuất dữ liệu, đọc báo cáo
- Nhân viên: Có quyền được chỉnh sửa, thêm mới các thông tin của số lượng sản phẩm được bán...

Mặt khác, hệ thống phải đáp ứng được việc đảm bảo an toàn cho thông tin của khách hàng đồng thời dễ dàng nâng cấp, bảo trì và sửa chữa khi cần thiết.

4.2. Chức năng phân tích báo cáo

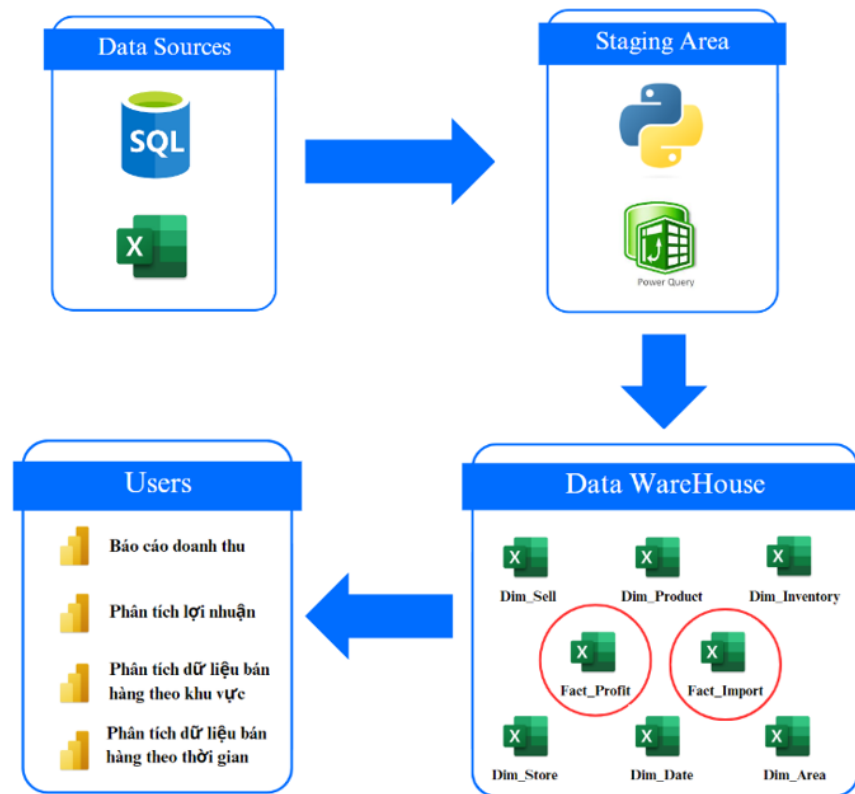
Chức năng phân tích và tạo báo cáo của hệ thống là dựa trên data được cập nhật trước nên chỉ có chức năng tổng hợp, truy vấn... không có chức năng thêm, xóa hay chỉnh sửa dữ liệu.

Các phân tích, báo cáo phải có chiều sâu, theo các chủ đề và phù hợp với từng nhóm đối tượng. Ví dụ:

- Báo cáo doanh thu theo tháng, năm, theo sản phẩm
- Báo cáo doanh thu theo sản phẩm
- Phân tích doanh số theo quý
- Báo cáo lợi nhuận theo tháng, năm, theo sản phẩm
- Phân tích tỷ lệ lợi nhuận thu về
- Báo cáo số lượng hàng hóa
- Báo cáo số lượng hàng tồn kho
- Báo cáo số lượng hàng nhập

III. Phân tích và thiết kế hệ thống

1. Kiến trúc của hệ thống phân tích



Hình 3.1 Kiến trúc của hệ thống phân tích

Trong đó:

- Data Sources: Là nơi lưu trữ, thu thập các bộ dữ liệu
- Staging Area: Là nơi thực hiện các thao tác làm sạch, chuẩn dữ liệu
- Data Warehouse: Nơi lưu trữ các tệp dữ liệu đã qua Staging
- Users: Sử dụng các dữ liệu sạch, đưa ra các phân tích, báo cáo

2. Thiết kế vùng xử lý dữ liệu – Staging

Quá trình này cần có sự hỗ trợ của công cụ ETL(Extract, Transform, Load) nhằm lấy dữ liệu từ các nguồn khác nhau(excel, google sheet...) sau đó chuyển đổi dữ liệu thành các dạng theo yêu cầu rồi tải vào Data Warehouse.

Công cụ ETL: Excel, Power query

Tiến trình xử lý:

- Xử lý thô các file: Xóa null, dịch ngôn ngữ...

- Đưa các file vào Power query để chỉnh sửa các định dạng, loại bỏ các trường không cần thiết...
- Đóng và tải các file

File sau khi được loại bỏ các giá trị null:

	A _C Mã bán hàng	A _C Thời gian	A _C Mã cửa hàng	A _C Mã sản phẩm	A _C Giá bán	A _C Số hàng bán được	A _C Doanh thu
1	SELL01272023CH478SP03	1/27/2023	CH 478	SP 03	305000	86	26230000
2	SELL01232023CH318SP03	1/23/2023	CH 318	SP 03	305000	22	6710000
3	SELL02062023CH413SP10	2/6/2023	CH 413	SP 10	323000	9	2907000
4	SELL02252023CH130SP09	2/25/2023	CH 130	SP 09	238000	30	7140000
5	SELL01112023CH103SP10	1/11/2023	CH 103	SP 10	323000	27	8721000
6	SELL02072023CH457SP02	2/7/2023	CH 457	SP 02	201000	51	10251000
7	SELL04192023CH125SP07	4/19/2023	CH 125	SP 07	124000	51	6324000
8	SELL03162023CH248SP09	3/16/2023	CH 248	SP 09	238000	77	18326000
9	SELL03062023CH300SP09	3/6/2023	CH 300	SP 09	238000	21	4998000
10	SELL03292023CH093SP04	3/29/2023	CH 093	SP 04	247000	69	17043000
11	SELL01292023CH025SP09	1/29/2023	CH 025	SP 09	238000	36	8568000
12	SELL03292023CH041SP05	3/29/2023	CH 041	SP 05	128000	76	9728000
13	SELL03132023CH313SP04	3/13/2023	CH 313	SP 04	247000	5	1235000
14	SELL01302023CH277SP10	1/30/2023	CH 277	SP 10	323000	34	10982000
15	SELL03142023CH364SP01	3/14/2023	CH 364	SP 01	230000	60	13800000
16	SELL02102023CH007SP02	2/10/2023	CH 007	SP 02	201000	72	14472000
17	SELL01102023CH352SP08	1/10/2023	CH 352	SP 08	78000	52	4056000
18	SELL01122023CH362SP07	1/12/2023	CH 362	SP 07	124000	33	4092000
19	SELL02222023CH138SP10	2/22/2023	CH 138	SP 10	323000	36	11628000
20	SELL04192023CH145SP01	4/19/2023	CH 145	SP 01	230000	87	20010000
21	SELL03122023CH277SP03	3/12/2023	CH 277	SP 03	305000	42	12810000
22	SELL02242023CH223SP06	2/24/2023	CH 223	SP 06	215000	16	3440000
23	SELL01062023CH113SP05	1/6/2023	CH 113	SP 05	128000	82	10496000
24	SELL02062023CH067SP10	2/6/2023	CH 067	SP 10	323000	65	20995000
25	SELL04202023CH408SP10	4/20/2023	CH 408	SP 10	323000	51	16473000
26	SELL02072023CH425SP08	2/7/2023	CH 425	SP 08	78000	30	2340000
27	SELL01042023CH147SP01	1/4/2023	CH 147	SP 01	230000	81	18630000
28	SELL01132023CH493SP10	1/13/2023	CH 493	SP 10	323000	32	10336000

Hình 3.2 File dữ liệu sau khi xử lý bỏ null

Xóa bỏ các cột dữ liệu không dùng đến.

	A _C Mã bán hàng	A _C Thời gian	A _C Mã cửa hàng	A _C Mã sản phẩm	A _C Số hàng bán được
1	SELL01272023CH478SP03	1/27/2023	CH 478	SP 03	86
2	SELL01232023CH318SP03	1/23/2023	CH 318	SP 03	22
3	SELL02062023CH413SP10	2/6/2023	CH 413	SP 10	9
4	SELL02252023CH130SP09	2/25/2023	CH 130	SP 09	30
5	SELL01112023CH103SP10	1/11/2023	CH 103	SP 10	27
6	SELL02072023CH457SP02	2/7/2023	CH 457	SP 02	51
7	SELL04192023CH125SP07	4/19/2023	CH 125	SP 07	51
8	SELL03162023CH248SP09	3/16/2023	CH 248	SP 09	77
9	SELL03062023CH300SP09	3/6/2023	CH 300	SP 09	21
10	SELL03292023CH093SP04	3/29/2023	CH 093	SP 04	69
11	SELL01292023CH025SP09	1/29/2023	CH 025	SP 09	36
12	SELL03292023CH041SP05	3/29/2023	CH 041	SP 05	76
13	SELL03132023CH313SP04	3/13/2023	CH 313	SP 04	5
14	SELL01302023CH277SP10	1/30/2023	CH 277	SP 10	34
15	SELL03142023CH364SP01	3/14/2023	CH 364	SP 01	60
16	SELL02102023CH007SP02	2/10/2023	CH 007	SP 02	72
17	SELL01102023CH352SP08	1/10/2023	CH 352	SP 08	52
18	SELL01122023CH362SP07	1/12/2023	CH 362	SP 07	33
19	SELL02222023CH138SP10	2/22/2023	CH 138	SP 10	36
20	SELL04192023CH145SP01	4/19/2023	CH 145	SP 01	87
21	SELL03122023CH277SP03	3/12/2023	CH 277	SP 03	42
22	SELL02242023CH223SP06	2/24/2023	CH 223	SP 06	16
23	SELL01062023CH113SP05	1/6/2023	CH 113	SP 05	82
24	SELL02062023CH067SP10	2/6/2023	CH 067	SP 10	65
25	SELL04202023CH408SP10	4/20/2023	CH 408	SP 10	51
26	SELL02072023CH425SP08	2/7/2023	CH 425	SP 08	30
27	SELL01042023CH147SP01	1/4/2023	CH 147	SP 01	81
28	SELL01132023CH493SP10	1/13/2023	CH 493	SP 10	32

Hình 3.3 File dữ liệu sau khi xóa bỏ những cột không dùng đến

Định dạng lại kiểu dữ liệu cho phù hợp.

	A _C Mã bán hàng	Thời gian	A _C Mã cửa hàng	A _C Mã sản phẩm	I ₃ Số hàng bán được
1	SELL01272023CH478SP03	1/27/2023	CH 478	SP 03	86
2	SELL01232023CH318SP03	1/23/2023	CH 318	SP 03	22
3	SELL02062023CH413SP10	2/6/2023	CH 413	SP 10	9
4	SELL02252023CH130SP09	2/25/2023	CH 130	SP 09	30
5	SELL01112023CH103SP10	1/11/2023	CH 103	SP 10	27
6	SELL02072023CH457SP02	2/7/2023	CH 457	SP 02	51
7	SELL04192023CH125SP07	4/19/2023	CH 125	SP 07	51
8	SELL03162023CH248SP09	3/16/2023	CH 248	SP 09	77
9	SELL03062023CH300SP09	3/6/2023	CH 300	SP 09	21
10	SELL03292023CH093SP04	3/29/2023	CH 093	SP 04	69
11	SELL01292023CH025SP09	1/29/2023	CH 025	SP 09	36
12	SELL03292023CH041SP05	3/29/2023	CH 041	SP 05	76
13	SELL03132023CH313SP04	3/13/2023	CH 313	SP 04	5
14	SELL01302023CH277SP10	1/30/2023	CH 277	SP 10	34
15	SELL03142023CH364SP01	3/14/2023	CH 364	SP 01	60
16	SELL02102023CH007SP02	2/10/2023	CH 007	SP 02	72
17	SELL01102023CH352SP08	1/10/2023	CH 352	SP 08	52
18	SELL01122023CH362SP07	1/12/2023	CH 362	SP 07	33
19	SELL02222023CH138SP10	2/22/2023	CH 138	SP 10	36
20	SELL04192023CH145SP01	4/19/2023	CH 145	SP 01	87
21	SELL03122023CH277SP03	3/12/2023	CH 277	SP 03	42
22	SELL02242023CH223SP06	2/24/2023	CH 223	SP 06	16
23	SELL01062023CH113SP05	1/6/2023	CH 113	SP 05	82
24	SELL02062023CH067SP10	2/6/2023	CH 067	SP 10	65
25	SELL04202023CH408SP10	4/20/2023	CH 408	SP 10	51
26	SELL02072023CH425SP08	2/7/2023	CH 425	SP 08	30
27	SELL01042023CH147SP01	1/4/2023	CH 147	SP 01	81
28	SELL01132023CH493SP10	1/13/2023	CH 493	SP 10	32

Hình 3.4 File sau khi được định dạng lại kiểu dữ liệu

Thực hiện sắp xếp lại thứ tự dữ liệu theo ngày, tháng, năm, mã cửa hàng, mã sản phẩm của bộ dữ liệu.

	A _C Mã bán hàng	Thời gian	A _C Mã cửa hàng	A _C Mã sản phẩm	I ₃ Số hàng bán được
1	SELL01012023CH001SP01	1/1/2023	CH 001	SP 01	42
2	SELL01012023CH001SP02	1/1/2023	CH 001	SP 02	80
3	SELL01012023CH001SP03	1/1/2023	CH 001	SP 03	61
4	SELL01012023CH001SP04	1/1/2023	CH 001	SP 04	54
5	SELL01012023CH001SP05	1/1/2023	CH 001	SP 05	27
6	SELL01012023CH001SP06	1/1/2023	CH 001	SP 06	44
7	SELL01012023CH001SP07	1/1/2023	CH 001	SP 07	13
8	SELL01012023CH001SP08	1/1/2023	CH 001	SP 08	9
9	SELL01012023CH001SP09	1/1/2023	CH 001	SP 09	34
10	SELL01012023CH001SP10	1/1/2023	CH 001	SP 10	53
11	SELL01012023CH002SP01	1/1/2023	CH 002	SP 01	30
12	SELL01012023CH002SP02	1/1/2023	CH 002	SP 02	40
13	SELL01012023CH002SP03	1/1/2023	CH 002	SP 03	19
14	SELL01012023CH002SP04	1/1/2023	CH 002	SP 04	86
15	SELL01012023CH002SP05	1/1/2023	CH 002	SP 05	69
16	SELL01012023CH002SP06	1/1/2023	CH 002	SP 06	31
17	SELL01012023CH002SP07	1/1/2023	CH 002	SP 07	74
18	SELL01012023CH002SP08	1/1/2023	CH 002	SP 08	23
19	SELL01012023CH002SP09	1/1/2023	CH 002	SP 09	70
20	SELL01012023CH002SP10	1/1/2023	CH 002	SP 10	86
21	SELL01012023CH003SP01	1/1/2023	CH 003	SP 01	84
22	SELL01012023CH003SP02	1/1/2023	CH 003	SP 02	47
23	SELL01012023CH003SP03	1/1/2023	CH 003	SP 03	78
24	SELL01012023CH003SP04	1/1/2023	CH 003	SP 04	36
25	SELL01012023CH003SP05	1/1/2023	CH 003	SP 05	68
26	SELL01012023CH003SP06	1/1/2023	CH 003	SP 06	74
27	SELL01012023CH003SP07	1/1/2023	CH 003	SP 07	57
28	SELL01012023CH003SP08	1/1/2023	CH 003	SP 08	47

Hình 3.5 File sau khi dữ liệu được sắp xếp lại

Tạo thêm một số trường mới:

- Tạo thêm trường lợi nhuận

	A _C Mã lợi nhuận	Thời gian	A _C Mã của hàng	A _C Mã sản phẩm	1 ₂ Chi phí	1 ₂ Doanh thu	1 ₂ Lợi nhuận
1	PROFIT01012023CH001SP01	1/1/2023	CH 001	SP 01	14878000	19780000	4902000
2	PROFIT01012023CH001SP02	1/1/2023	CH 001	SP 02	3278000	4422000	1144000
3	PROFIT01012023CH001SP03	1/1/2023	CH 001	SP 03	2106000	2745000	639000
4	PROFIT01012023CH001SP04	1/1/2023	CH 001	SP 04	4980000	7410000	2430000
5	PROFIT01012023CH001SP05	1/1/2023	CH 001	SP 05	2322000	3456000	1134000
6	PROFIT01012023CH001SP06	1/1/2023	CH 001	SP 06	8160000	10965000	2805000
7	PROFIT01012023CH001SP07	1/1/2023	CH 001	SP 07	4743000	6324000	1581000
8	PROFIT01012023CH001SP08	1/1/2023	CH 001	SP 08	4158000	6006000	1848000
9	PROFIT01012023CH001SP09	1/1/2023	CH 001	SP 09	3423000	4998000	1575000
10	PROFIT01012023CH001SP10	1/1/2023	CH 001	SP 10	16905000	22287000	5382000
11	PROFIT01012023CH002SP01	1/1/2023	CH 002	SP 01	6228000	8280000	2052000
12	PROFIT01012023CH002SP02	1/1/2023	CH 002	SP 02	11324000	15276000	3952000
13	PROFIT01012023CH002SP03	1/1/2023	CH 002	SP 03	1170000	1525000	355000
14	PROFIT01012023CH002SP04	1/1/2023	CH 002	SP 04	5644000	8398000	2754000
15	PROFIT01012023CH002SP05	1/1/2023	CH 002	SP 05	5160000	7680000	2520000
16	PROFIT01012023CH002SP06	1/1/2023	CH 002	SP 06	11520000	15480000	3960000
17	PROFIT01012023CH002SP07	1/1/2023	CH 002	SP 07	4836000	6448000	1612000
18	PROFIT01012023CH002SP08	1/1/2023	CH 002	SP 08	1782000	2574000	792000
19	PROFIT01012023CH002SP09	1/1/2023	CH 002	SP 09	5868000	8568000	2700000
20	PROFIT01012023CH002SP10	1/1/2023	CH 002	SP 10	21315000	28101000	6786000
21	PROFIT01012023CH003SP01	1/1/2023	CH 003	SP 01	7266000	9660000	2394000
22	PROFIT01012023CH003SP02	1/1/2023	CH 003	SP 02	2384000	3216000	832000
23	PROFIT01012023CH003SP03	1/1/2023	CH 003	SP 03	19188000	25010000	5822000
24	PROFIT01012023CH003SP04	1/1/2023	CH 003	SP 04	10790000	16055000	5265000
25	PROFIT01012023CH003SP05	1/1/2023	CH 003	SP 05	4386000	6528000	2142000
26	PROFIT01012023CH003SP06	1/1/2023	CH 003	SP 06	4800000	6450000	1650000
27	PROFIT01012023CH003SP07	1/1/2023	CH 003	SP 07	7533000	10044000	2511000
28	PROFIT01012023CH003SP08	1/1/2023	CH 003	SP 08	1728000	2496000	768000

Hình 3.6 Tạo thêm trường lợi nhuận

- Tạo thêm trường phần trăm lợi nhuận

	Thời gian	A _C Mã của hàng	A _C Mã sản phẩm	1 ₂ Chi phí	1 ₂ Doanh thu	1 ₂ Lợi nhuận	% Phần trăm lợi nhuận
1	1/1/2023	CH 001	SP 01	14878000	19780000	4902000	24.78%
2	1/1/2023	CH 001	SP 02	3278000	4422000	1144000	25.87%
3	1/1/2023	CH 001	SP 03	2106000	2745000	639000	23.28%
4	1/1/2023	CH 001	SP 04	4980000	7410000	2430000	32.79%
5	1/1/2023	CH 001	SP 05	2322000	3456000	1134000	32.81%
6	1/1/2023	CH 001	SP 06	8160000	10965000	2805000	25.58%
7	1/1/2023	CH 001	SP 07	4743000	6324000	1581000	25.00%
8	1/1/2023	CH 001	SP 08	4158000	6006000	1848000	30.77%
9	1/1/2023	CH 001	SP 09	3423000	4998000	1575000	31.51%
10	1/1/2023	CH 001	SP 10	16905000	22287000	5382000	24.15%
11	1/1/2023	CH 002	SP 01	6228000	8280000	2052000	24.78%
12	1/1/2023	CH 002	SP 02	11324000	15276000	3952000	25.87%
13	1/1/2023	CH 002	SP 03	1170000	1525000	355000	23.28%
14	1/1/2023	CH 002	SP 04	5644000	8398000	2754000	32.79%
15	1/1/2023	CH 002	SP 05	5160000	7680000	2520000	32.81%
16	1/1/2023	CH 002	SP 06	11520000	15480000	3960000	25.58%
17	1/1/2023	CH 002	SP 07	4836000	6448000	1612000	25.00%
18	1/1/2023	CH 002	SP 08	1782000	2574000	792000	30.77%
19	1/1/2023	CH 002	SP 09	5868000	8568000	2700000	31.51%
20	1/1/2023	CH 002	SP 10	21315000	28101000	6786000	24.15%
21	1/1/2023	CH 003	SP 01	7266000	9660000	2394000	24.78%
22	1/1/2023	CH 003	SP 02	2384000	3216000	832000	25.87%
23	1/1/2023	CH 003	SP 03	19188000	25010000	5822000	23.28%
24	1/1/2023	CH 003	SP 04	10790000	16055000	5265000	32.79%
25	1/1/2023	CH 003	SP 05	4386000	6528000	2142000	32.81%
26	1/1/2023	CH 003	SP 06	4800000	6450000	1650000	25.58%
27	1/1/2023	CH 003	SP 07	7533000	10044000	2511000	25.00%

Hình 3.7 Tạo thêm trường phần trăm lợi nhuận

3. Xây dựng Dimension & Fact

3.1. Dimension

Đối với bộ dữ liệu sử dụng trong báo cáo, em đã phân tích theo các Dim sau:

Tên bảng	Mô tả
Dim_Sell	Gồm dữ liệu về số lượng hàng bán ra của tất cả sản phẩm
Dim_Inventory	Gồm dữ liệu về số lượng hàng còn lại tại các cửa hàng của tất cả sản phẩm
Dim_Product	Thông tin về sản phẩm, giá nhập, giá bán
Dim_Store	Thông tin về cửa hàng, địa chỉ của cửa hàng
Dim_Date	Chứa thông tin về thời gian bán hàng của thùng của hàng và từng sản phẩm theo ngày, tháng, năm
Dim_Area	Chứa thông tin về khu vực, các tỉnh, kinh độ, vĩ độ của tỉnh đó

3.2. Fact

Các chủ điểm phân tích trong bộ dữ liệu này mà em hướng tới đó là:

- Fact_Profit: Phân tích về lợi nhuận của cửa hàng qua từng ngày, tháng, năm.
- Fact_Import: Phân tích số lượng hàng bán ra và tồn kho để đưa ra số lượng hàng cần nhập.

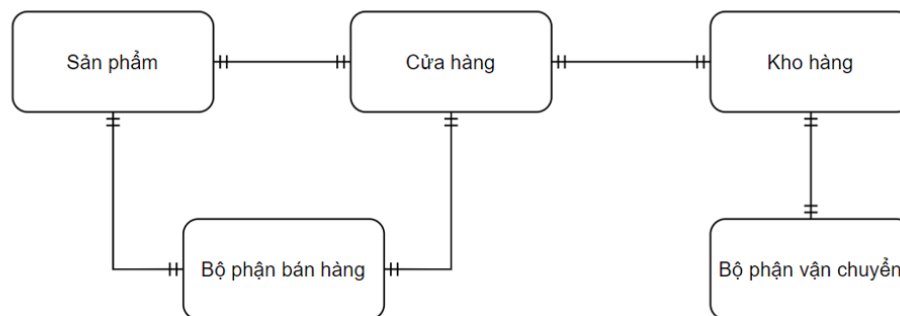
4. Thiết kế cơ sở dữ liệu hệ thống OLTP

Cơ sở dữ liệu OLTP là một cơ sở dữ liệu được thiết kế để hỗ trợ xử lý các hoạt động kinh doanh hàng ngày và cung cấp thông tin thời gian thực về quản lý kho hàng.

Cơ sở dữ liệu này được sử dụng để lưu trữ, truy xuất và cập nhật dữ liệu liên quan đến số số lượng bán hàng, hàng tồn, chi phí, doanh thu và các hoạt động khác trong quá trình vận hành kho hàng.

Cơ sở dữ liệu OLTP trong hệ thống quản lý kho hàng đảm bảo tính toàn vẹn và đáng tin cậy của dữ liệu, cho phép thực hiện các giao dịch hàng hóa một cách hiệu quả và quản lý thông tin kho hàng một cách chính xác.

Sơ đồ hệ thống OLTP:



Hình 3.8 Sơ đồ hệ thống OLTP

Mô tả các bảng trong cơ sở dữ liệu OLTP

- Sản phẩm: Chứa thông tin về các sản phẩm
- Cửa hàng: Chứa thông tin về các cửa hàng
- Bộ phận bán hàng: Chứa thông tin về số lượng bán hàng của từng sản phẩm của tất cả cửa hàng.
- Kho hàng: Chứa thông tin về số lượng hàng còn lại trong kho của từng sản phẩm.
- Bộ phận vận chuyển: Chứa thông tin về số lượng hàng nhập của từng sản phẩm.

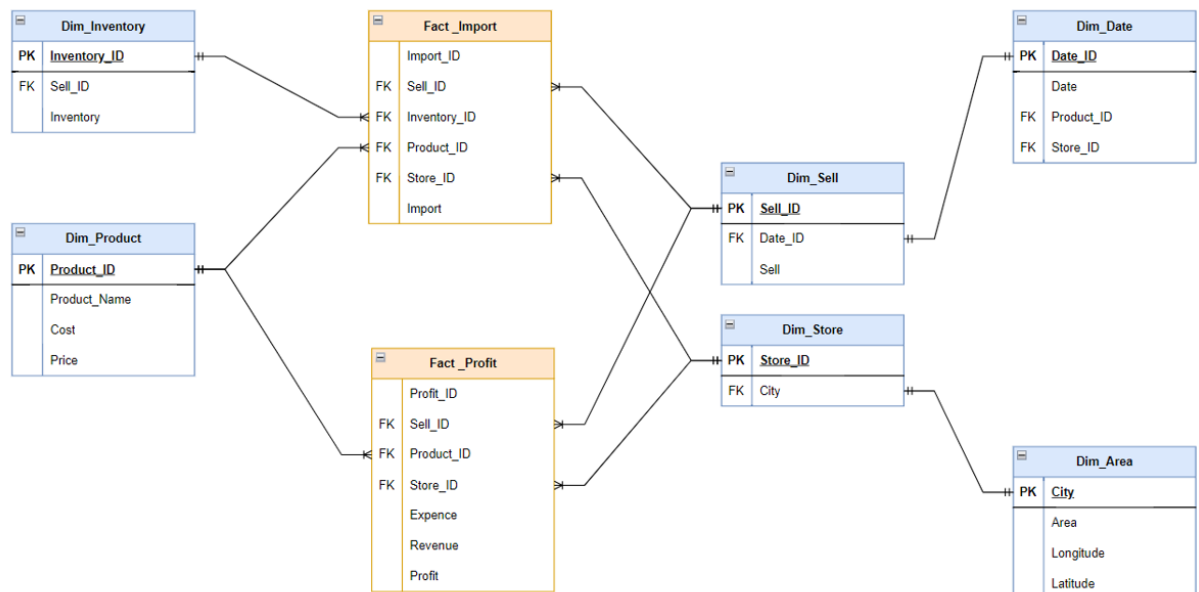
5. Thiết kế cơ sở dữ liệu hệ thống OLAP

Cơ sở dữ liệu OLAP trong hệ thống quản lý kho hàng được sử dụng để phân tích và tổng hợp dữ liệu từ cơ sở dữ liệu OLTP nhằm cung cấp thông tin chi tiết và chính xác về hoạt động và hiệu suất quản lý kho hàng.

Cơ sở dữ liệu OLAP dùng để thực hiện các tác vụ phân tích, báo cáo, và truy vấn dữ liệu theo các phương pháp đa chiều như:

- Phân tích hàng tồn như tồn kho hiện tại, tồn kho trong quá khứ, biến động hàng tồn kho theo thời gian, tổng giá trị hàng tồn kho, các mặt hàng bán chạy, và các mặt hàng tồn kho không bán được.
- Tổng hợp doanh số bán hàng và so sánh doanh số bán hàng theo từng đơn hàng, khách hàng, sản phẩm, hoặc thời gian để hiểu rõ hơn về hiệu suất kinh doanh và xu hướng tiêu thụ.
- Giúp quản lý kho hàng đưa ra các dự báo và kế hoạch về việc nhập hàng, xuất hàng, tồn kho, và tối ưu hóa các quy trình quản lý kho.
- Cung cấp khả năng tạo các báo cáo đa dạng và thống kê tổng hợp về tình trạng hàng tồn kho, doanh số bán hàng, hiệu suất vận chuyển và các chỉ số quản lý kho hàng khác.

Sơ đồ hệ thống OLAP:



Hình 3.9 Sơ đồ hệ thống OLAP

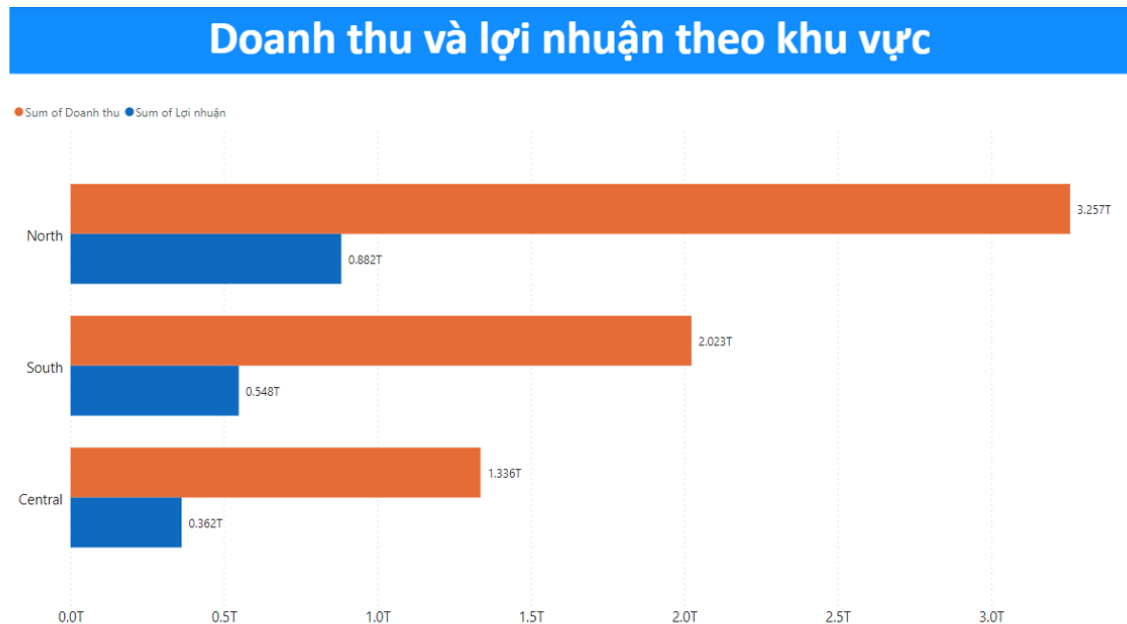
Giới thiệu về dữ liệu trong sơ đồ hệ thống OLAP:

Tên Dim&Fact	Tên cột	Giới thiệu cột
Dim_Date	Date_ID	Mã thời gian
	Date	Thời gian bán sản phẩm
	Product_ID	Mã sản phẩm
	Store_ID	Mã cửa hàng
Dim_Sell	Sell_ID	Mã bán hàng
	Date_ID	Mã thời gian
	Sell	Số lượng hàng bán được
Dim_Inventory	Inventory_ID	Mã hàng còn lại
	Sell_ID	Mã bán hàng
	Inventory	Số lượng hàng còn lại trong kho
Dim_Product	Product_ID	Mã sản phẩm
	Product_Name	Tên sản phẩm
	Cost	Giá nhập
	Price	Giá bán
Dim_Store	Store_ID	Mã cửa hàng
	City	Tên thành phố
Dim_Area	City	Tên thành phố
	Area	Tên miền
	Longitude	Kinh độ
	Latitude	Vĩ độ

Fact_Profit	Profit_ID	Mã lợi nhuận
	Sell_ID	Mã bán hàng
	Product_ID	Mã sản phẩm
	Store_ID	Mã cửa hàng
	Expencc	Chi phí
	Revenue	Doanh thu
	Profit	Lợi nhuận
Fact_Import	Import_ID	Mã nhập hàng
	Sell_ID	Mã bán hàng
	Inventory_ID	Mã hàng còn lại
	Product_ID	Mã sản phẩm
	Store_ID	Mã cửa hàng
	Import	Số lượng hàng nhập

6. Khai phá dữ liệu

6.1. Theo khu vực

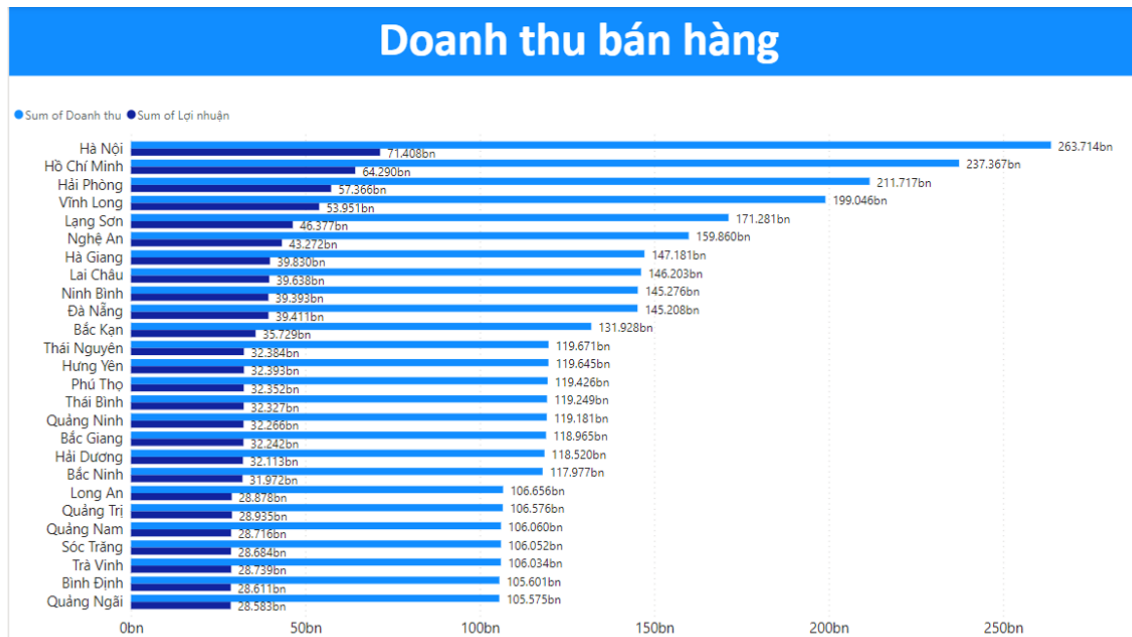


Hình 3.10 Doanh thu và lợi nhuận theo khu vực

Có một số lý do có thể giải thích tại sao miền Bắc thường bán được nhiều đồ chơi hơn. Dưới đây là một số lý do phổ biến:

- Dân số đông: Khu vực miền Bắc có dân số lớn hơn so với miền Nam. Với dân số đông, tỷ lệ trẻ em và gia đình có trẻ em cũng tăng. Điều này tạo ra một thị trường tiềm năng cho đồ chơi, vì có nhiều gia đình cần mua đồ chơi cho con cái của mình.
- Môi trường đô thị phát triển: Miền Bắc có nhiều thành phố lớn như Hà Nội, Hải Phòng, Hải Dương... Những thành phố này có môi trường đô thị phát triển, với nhiều trung tâm mua sắm, siêu thị và cửa hàng chuyên dụng. Sự phát triển này tạo điều kiện thuận lợi cho việc tiếp cận và mua sắm đồ chơi.
- Nhu cầu giáo dục và giải trí: Gia đình ở miền Bắc thường có mức độ quan tâm cao đối với việc giáo dục và giải trí cho trẻ em. Đồ chơi có thể được xem là công cụ giáo dục và giải trí quan trọng, giúp trẻ em phát triển kỹ năng và sáng tạo. Do đó, việc mua đồ chơi cho trẻ em được coi là một nhu cầu thiết yếu trong gia đình.

6.2. Theo doanh thu



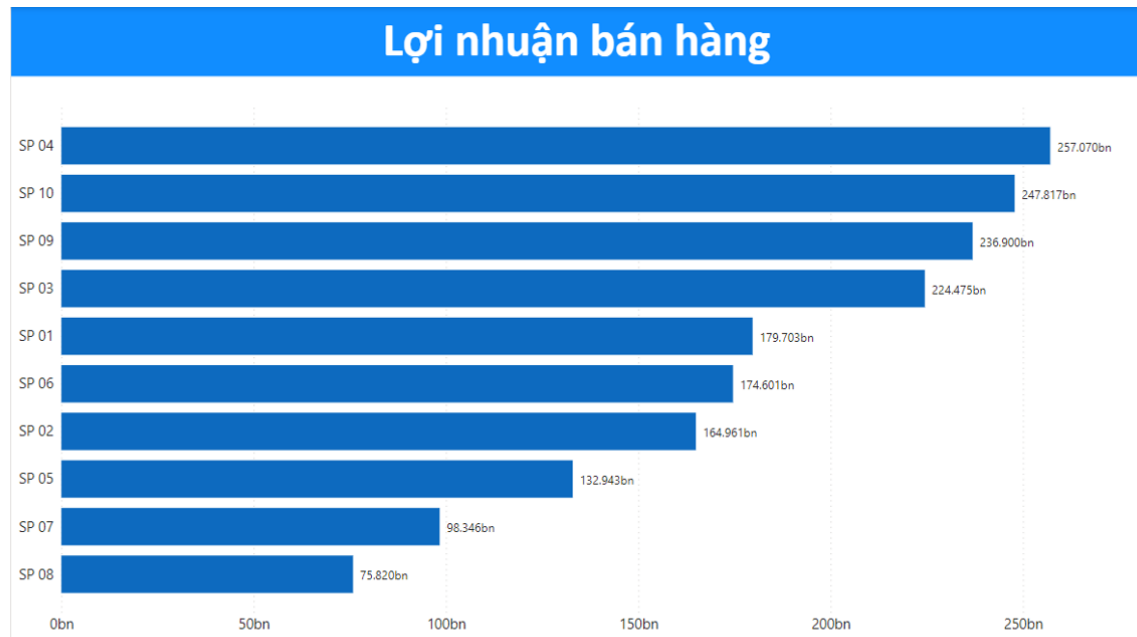
Hình 3.11 Doanh thu bán hàng

Những thành phố lớn như Hà Nội, Hồ Chí Minh và Hải Phòng có dân số đông đúc và môi trường kinh tế phát triển. Điều này tạo điều kiện thuận lợi cho việc mua sắm đồ chơi. Dân số đông cùng với thu nhập trung bình tương đối cao trong các thành phố này tạo ra nhu cầu mua sắm đồ chơi cao.

Các thành phố lớn này có hạ tầng thương mại phát triển, với nhiều trung tâm mua sắm, siêu thị và cửa hàng đồ chơi. Sự hiện diện của các trung tâm mua sắm lớn như Lotte Mart, AEON Mall, Vincom Center và các cửa hàng chuyên dụng đồ chơi tạo điều kiện thuận lợi cho người dân tiếp cận và mua sắm đồ chơi dễ dàng.

Các thành phố lớn thường có nhịp sống nhanh, với nhiều hoạt động giải trí và giáo dục cho trẻ em. Gia đình có xu hướng quan tâm đến việc giáo dục và phát triển kỹ năng cho con cái thông qua việc mua đồ chơi. Đồ chơi cũng giúp tạo ra niềm vui và giải trí cho trẻ em trong môi trường đô thị sôi động.

6.3. Theo lợi nhuận



Hình 3.12 Lợi nhuận bán hàng

Thông tin về từng sản phẩm của tất cả các cửa hàng:

Thông tin sản phẩm			
Mã sản phẩm	Tên sản phẩm	Giá nhập	Giá bán
SP 01	Lego	173000	230000
SP 02	Búp bê Barbie	149000	201000
SP 03	Xe ô tô điều khiển từ xa	234000	305000
SP 04	Đồ chơi xếp hình Rubik	166000	247000
SP 05	Đồ chơi mô hình xe ô tô	86000	128000
SP 06	Đồ chơi robot lắp ráp	160000	215000
SP 07	Đồ chơi đàn piano nhỏ	93000	124000
SP 08	Đồ chơi mô hình bàn trang điểm	54000	78000
SP 09	Đồ chơi cờ cá ngựa	163000	238000
SP 10	Đồ chơi cờ vua	245000	323000

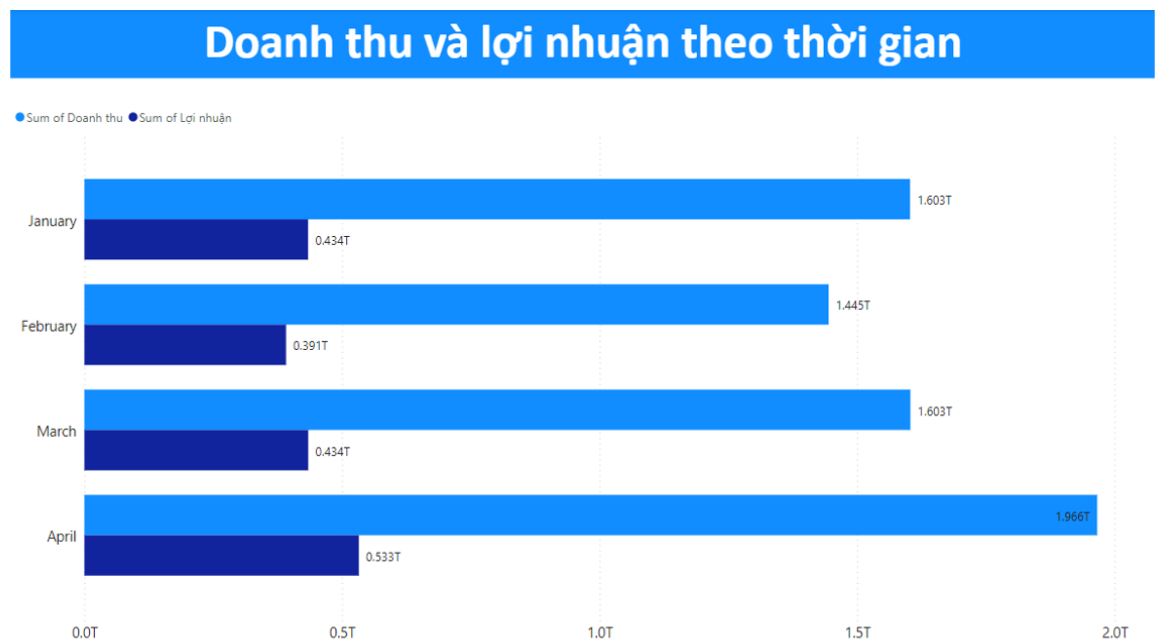
Hình 3.13 Thông tin sản phẩm

Các đồ chơi như rubik, cờ vua, cờ cá ngựa có khả năng giúp phát triển kỹ năng tư duy logic, tư duy chiến lược, sáng tạo và tăng cường trí tuệ. Chơi Rubik có thể cải thiện khả năng giải quyết vấn đề và tư duy không gian. Cờ vua đòi hỏi người chơi suy nghĩ chiến lược, lập kế hoạch và đánh giá tình huống. Cờ cá ngựa có thể cải thiện khả năng đọc hiểu và tính toán. Nên rất được sự ưu chuộng của các bậc phụ huynh mua cho trẻ em sử dụng.

Rubik, cờ vua và cờ cá ngựa đều có lịch sử và truyền thống lâu đời. Chúng đã trở thành các đồ chơi phổ biến và được biết đến rộng rãi trên toàn thế giới. Điều này tạo ra sự quen thuộc và lòng tin trong người mua hàng, và có thể làm tăng nhu cầu mua sắm cho các sản phẩm này.

Rubik, cờ vua và cờ cá ngựa thường được chơi trong nhóm hoặc với đối thủ. Những trò chơi này tạo ra cơ hội cho người chơi tương tác xã hội, giao tiếp và cạnh tranh với nhau. Điều này có thể thu hút sự quan tâm của những người thích giao lưu và chia sẻ niềm đam mê chơi trò chơi với người khác. Đồng thời cũng lan tỏa nhu cầu về đồ chơi này đến những người khác giúp tăng nhu cầu về các đồ chơi này.

6.4. Theo thời gian



Hình 3.14 Doanh thu và lợi nhuận theo thời gian

Tháng 4 và tháng 1 có một số ngày lễ và kỳ nghỉ quan trọng, như Tết Nguyên Đán (tháng 1 theo lịch âm) và Lễ Phục Sinh (tháng 4). Những dịp này thường là thời điểm mọi người tặng quà và mua sắm đồ chơi cho trẻ em. Gia đình có thể muốn chia sẻ niềm vui và hạnh phúc với con cái bằng cách mua đồ chơi mới.

Tháng 4 thường là mùa xuân và tháng 1 là mùa đông. Trong thời tiết mát mẻ của mùa xuân và mùa đông, trẻ em có xu hướng chơi ngoài trời ít hơn và chơi trong nhà nhiều hơn. Do đó, việc mua sắm đồ chơi để giải trí và giữ trẻ em vui vẻ trong những tháng này có thể tăng lên.

Tháng 4 là thời gian các bạn học sinh được nghỉ hè nên các gia đình thường có xu hướng đi nghỉ mát. Khi đó nhu cầu mua đồ chơi của các gia đình sẽ tăng lên. Và các gia đình không đi nghỉ mát cũng sẽ có nhu cầu mua đồ chơi cho trẻ em trong dịp nghỉ hè ở nhà.

Tháng 2 chỉ là sau dịp nghỉ Tết Nguyên Đán. Học sinh thường đi học và việc mua đồ chơi cho trẻ em đã được thực hiện trước Tết Nguyên Đán. Vậy nên nhu cầu mua đồ chơi tại tháng 2 có xu hướng giảm.

IV. Xây dựng hệ thống

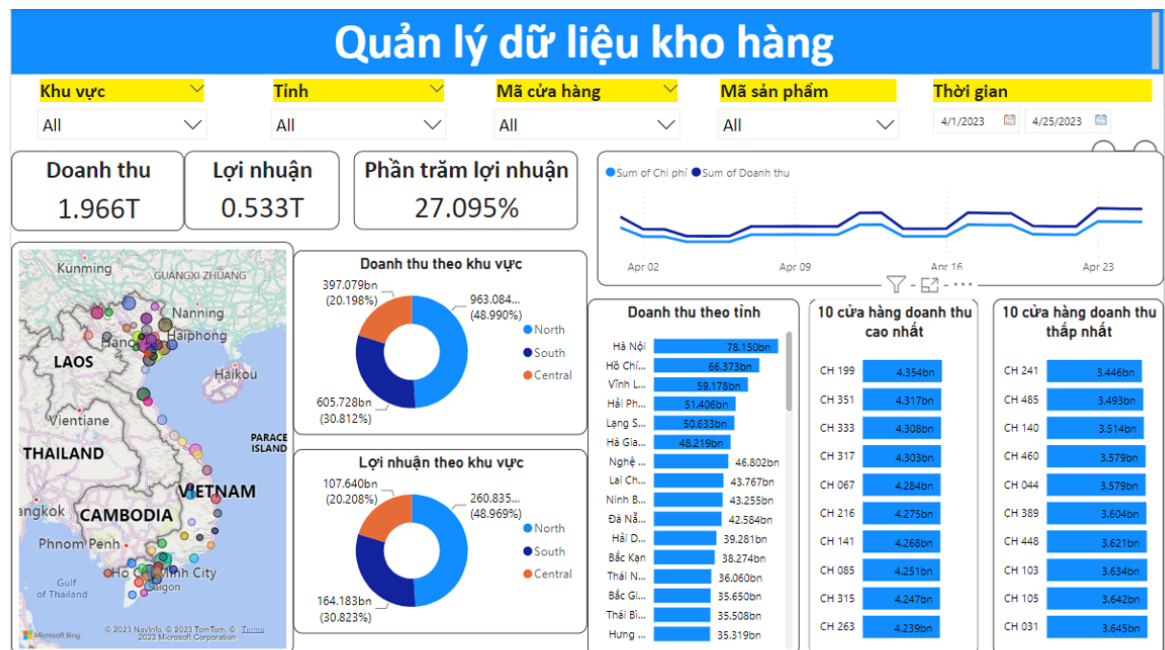
1. Truyền dữ liệu

Để một hệ thống có thể vận hành được thì ngoài việc thiết kế chúng ta còn cần phải truyền dữ liệu và trong hệ thống. Từ đó hệ thống mới có thể vận hành và đưa ra các phân tích hỗ trợ quá trình hoạt động của kho hàng.

Quy trình truyền dữ liệu:

- Dữ liệu được thu thập và truyền vào cơ sở dữ liệu OLTP
- Từ cơ sở dữ liệu OLTP dữ liệu được làm sạch bởi quá trình ELT ở tầng Staging.
- Sử dụng các procedure để truyền dữ liệu tương ứng của các bảng vào Dim & Fact
- Sau đó, dữ liệu được truyền vào cơ sở dữ liệu OLAP và tải lên hệ thống Power BI

2. Xây dựng hệ thống Power BI



Hình 4.1 Dashboard quản lý dữ liệu kho hàng

3. Tổng kết

Trong quá trình thực hiện đồ án em đã đạt được các kết quả sau:

- Hiểu được những kiến thức cơ bản của kho dữ liệu, kinh doanh thông minh, biết cách phân tích dữ liệu và phân tích kinh doanh.
- Đã có thể sử dụng được các công cụ ETL, công cụ trực quan hóa như Excel, Power query, Power BI, SQL Sever, Python.
- Biết được cách phân tích, thiết kế một hệ thống, biết một hệ thống cần có những mô hình gì, như thế nào.
- Phần nào đó đã hoàn thành được các mục tiêu đề ra.

Tuy nhiên, do thời gian và kiến thức còn hạn chế nên hệ thống phân tích của em có lẽ vẫn còn nhiều thiếu sót. Vì vậy, em rất mong nhận được sự góp ý của cô và bạn đọc để đề tài này được hoàn thiện hơn.

Em xin chân thành cảm ơn !

V. Tài liệu tham khảo

1. Nguyễn Danh Tú, Viện Toán Ứng Dụng và Tin Học, Slide bài giảng môn kho dữ liệu và kinh doanh thông minh.
2. [Power BI Documentation](#)
3. Website: easystudyonline.com