

**ĐẠI HỌC BÁCH KHOA HÀ NỘI**  
**VIỆN TOÁN ỨNG DỤNG VÀ TIN HỌC**

—o0o—



**ĐỒ ÁN I**

**Xây dựng hệ thống phân tích kho hàng**

*Chuyên ngành:* **Hệ thống thông tin quản lý**

Giảng viên hướng dẫn: **TS. Lê Quang Hòa**

Sinh viên: **Bùi Hồng Giang**

Lớp: **MI2-02 K65**

**HÀ NỘI, 5/2023**

# Nhận xét của giảng viên hướng dẫn

Mục tiêu và nội dung của đề án

Kết quả đạt được

Ý thức làm việc của sinh viên

Hà Nội, ngày      tháng      năm  
Chữ kí của giảng viên

TS Lê Quang Hòa

# Lời nói 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 người sử dụng trở nên đúng đắn hơn. Trong đề án này, tôi sẽ trình bày về hệ thống phân tích dữ liệu được thiết kế và xây dựng để giúp tổ chức của bạn đáp ứng nhu cầu phân tích dữ liệu một cách nhanh chóng và hiệu quả. Hệ thống này sẽ được xây dựng trên nền tảng các công nghệ phần mềm và phần cứng tiên tiến nhất để đảm bảo tính tin cậy và hiệu suất cao. Nội dung của đề án này bao gồm 4 chương chính với mục tiêu giúp chỉ ra các hướng phân tích của một dữ liệu, nêu ra các cơ sở lý thuyết cần có để có thể xây dựng một hệ thống phân tích dữ liệu. Sau là các bước làm sạch dữ liệu, xác định chiều sâu của dữ liệu từ đó chuyển đổi dữ liệu qua các cơ sở dữ liệu OLTP, OLAP. Từ đó sử dụng các công cụ phân tích xây dựng các báo cáo, phân tích về dữ liệu.

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, tôi sẽ giải thích các thành phần chính của hệ thống, cách thức hoạt động và các ứng dụng của nó. Ngoài ra, tôi cũng sẽ trình bày về quá trình thiết kế, triển khai và kiểm thử của hệ thống. Mong rằng đồ án này sẽ giúp ích cho các bạn đọc 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 nhưng do khả năng và thời gian còn hạn chế 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.

# Mục lục

<b>1</b>	<b>Cơ sở lý thuyết</b>	<b>1</b>
1.1	Tổng quan về kho dữ liệu (Data Warehouse)	1
1.1.1	<i>Khái niệm</i>	1
1.1.2	<i>Mô hình dữ liệu OLAP</i>	2
1.1.3	<i>So sánh OLAP và OLTP</i>	5
1.2	Tổng quan về kinh doanh thông minh (Business Intelligence)	6
1.2.1	<i>Khái niệm</i>	6
1.2.2	<i>Thành phần cơ bản của hệ thống kinh doanh thông minh</i>	6
1.2.3	<i>Tiện ích từ các ứng dụng BI</i>	8
<b>2</b>	<b>Khảo sát hệ thống</b>	<b>10</b>
2.1	Hệ thống hiện tại	10
2.2	Quy trình nghiệp vụ	11
2.2.1	<i>Quy trình hoạt động của kho hàng</i>	11
2.2.2	<i>Luồng dữ liệu</i>	12
2.3	Kiến trúc mô hình	13
2.4	Yêu cầu hệ thống mới	14
2.4.1	<b>Chức năng cơ bản</b>	14
2.4.2	<i>Chức năng phân tích báo cáo</i>	14
<b>3</b>	<b>Phân tích và thiết kế hệ thống</b>	<b>16</b>
3.1	Kiến trúc của hệ thống phân tích	16
3.2	Thiết kế vùng xử lý dữ liệu - Staging	17
3.3	Xây dựng Dimension-Fact	23
3.3.1	<i>Dimension</i>	23
3.3.2	<i>Fact</i>	23
3.3.3	<i>Mô hình dữ liệu quan hệ</i>	24
3.4	Thiết kế cơ sở dữ liệu hệ thống OLTP	24
3.5	Thiết kế cơ sở dữ liệu hệ thống OLAP	25

3.6	Khai phá dữ liệu . . . . .	27
3.6.1	<i>Theo thời gian</i> . . . . .	27
3.6.2	<i>Theo doanh thu</i> . . . . .	28
3.6.3	<i>Theo lợi nhuận</i> . . . . .	29
<b>4</b>	<b>Xây dựng hệ thống</b>	<b>32</b>
4.1	Truyền dữ liệu . . . . .	32
4.2	Xây dựng hệ thống Power BI . . . . .	32
	<b>Kết luận</b>	<b>34</b>
	<b>Tài liệu tham khảo</b>	<b>35</b>

# Chương 1

## Cơ sở lý thuyết

### 1.1 Tổng quan về kho dữ liệu (Data Warehouse)

#### 1.1.1 *Khái niệm*

Data warehouse (DW) hay 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 đầy đủ:
- Data Sources: Là nơi thu thập, tích hợp dữ liệu từ nhiều nguồn khác nhau.
- Staging Area: Thực hiện chuẩn hóa dữ liệu, làm sạch, đưa dữ liệu về chung một định dạng và lưu dữ liệu đã tổng hợp.
- Warehouse: Nơi lưu trữ dữ liệu đã được xử lý
- Data Marts: Chia dữ liệu thành các chủ đề nhỏ hơn
- Users: Nơi thực hiện các thao tác truy vấn, báo cáo, phân tích để tìm ra xu hướng, trung bình, tổng hợp, ...

### 1.1.2 *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.

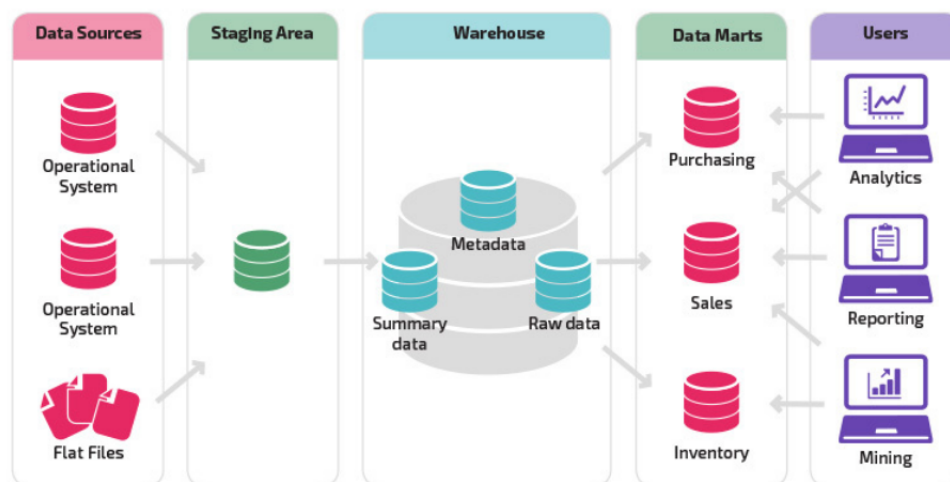
Để thực hiện điều đó ta cần dữ liệu được thu thập từ nhiều nguồn dữ liệu và được lưu trữ trong kho dữ liệu, sau đó được "làm sạch" và sắp xếp thành các khối dữ liệu.

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





## Data Warehouse Models



Hình 1.1: Kiến trúc kho dữ liệu

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:

1. **MOLAP**: 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.
2. **ROLAP**: 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 OPAP 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.
3. **HOLAP**: 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:

1. *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.
2. *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:

1. *Lược đồ bông tuyết (Snow flake schema)*:
  - 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 xấu cần sự kết nối giữa nhiều bảng gây giảm hiệu năng

2. *Lược đồ hình sao (Star schema)*:

- 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

### 1.1.3 So sánh OLAP và OLTP

Yếu tố	OLAP	OLTP
Người dùng	Người ra quyết định Nhóm người quản lý được chỉ định	Nhóm nhân viên, khách hàng, có thể truy cập bởi rất nhiều người
Chức năng	Phân tích đa chiều, khai thác dữ liệu, Phân tích phức tạp, báo cáo kinh doanh	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

## 1.2 Tổng quan về kinh doanh thông minh (Business Intelligence)

### 1.2.1 *Khái niệm*

Hệ thống kinh doanh thông minh (BI) 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.

### 1.2.2 *Thành phần cơ bản của hệ thống kinh doanh thông minh*

Trong đó :

*Data Sources:* 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:* 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



Hình 1.2: Quy trình kinh doanh thông minh

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:* 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:* 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:* 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.

### 1.2.3 *Tiện 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

## Chương 2

# Khảo sát hệ thống

### 2.1 Hệ thống hiện tại

- Kho 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 doanh thu, nhập xuất dưới dạng bảng. Chưa có sự thể hiện trên biểu đồ hay thể hiện rõ ràng được các xu thế của mặt hàng trong kho hàng, dẫn đến chưa có cái nhìn tổng quan về hoạt động của kho 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 kho 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.



## 2.2 Quy trình nghiệp vụ

### 2.2.1 Quy trình hoạt động của kho hàng

Quy trình hoạt động của một kho hàng có 5 bộ phận chính:

*Quản lý kho hàng:* Người quản lý chịu trách nhiệm tổ chức và điều hành mọi hoạt động trong kho hàng. Nhiệm vụ của người này bao gồm lập kế hoạch, quản lý hàng hóa, đảm bảo tuân thủ các quy định và tiêu chuẩn, và giám sát hoạt động hàng ngày của kho.

*Nhân viên kho:* Đây là nhóm nhân viên làm việc trực tiếp trong kho hàng. Công việc của họ bao gồm nhận hàng, kiểm tra chất lượng, đóng gói, lưu trữ và xuất hàng. Họ phải đảm bảo rằng quá trình nhập xuất hàng diễn ra một cách chính xác và hiệu quả.

*Nhân viên vận chuyển:* Đối với các kho hàng lớn hoặc có hệ thống phân phối rộng, có thể có một bộ phận riêng biệt chịu trách nhiệm vận chuyển hàng hóa từ kho đến địa điểm khác. Nhân viên vận chuyển quản lý lộ trình vận chuyển, đóng gói hàng hóa, và đảm bảo an toàn và chính xác trong quá trình vận chuyển.

*Nhân viên kiểm soát kho:* Nhóm nhân viên này có nhiệm vụ kiểm soát và ghi nhận số lượng hàng hóa trong kho hàng. Họ thường sử dụng các công cụ như máy quét mã vạch hoặc hệ thống quản lý kho để theo dõi việc nhập xuất hàng và đảm bảo sự khớp nhau giữa số lượng thực tế và số lượng được ghi nhận.

*Nhân viên quản lý chất lượng:* Trong một số kho hàng, có thể có một bộ phận riêng chịu trách nhiệm đảm bảo chất lượng của hàng hóa. Nhân viên quản lý chất lượng kiểm tra hàng hóa khi nhập kho, xác định và xử lý hàng hóa bị hỏng, và đảm bảo rằng các tiêu chuẩn chất lượng được tuân thủ.

Sơ đồ quy trình hoạt động:

Chi tiết quy trình:

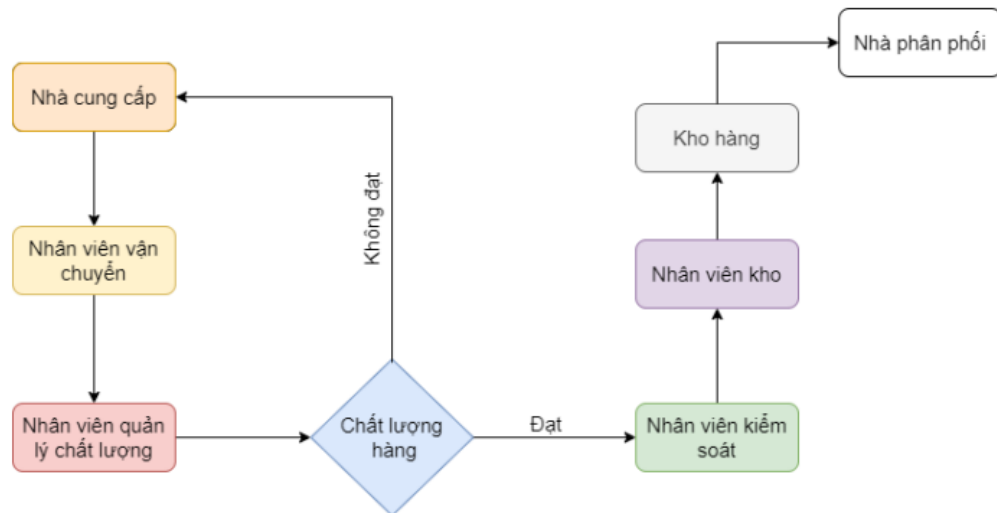
**B1:** hàng hóa nhận từ các nhà cung cấp sẽ được nhân viên vận chuyển chuyển về kho hàng.

**B2:** nhân viên quản lý chất lượng sẽ kiểm tra chất lượng của lô hàng vừa nhập và ghi kết quả.

**B3:** nhân viên kiểm soát kho sẽ ghi lại số lượng hàng trong kho hàng sau khi hàng đã thông qua quản lý chất lượng, sau đó tính toán doanh thu tho ngày, tháng, năm. Nhân viên kho sẽ sắp xếp, lưu trữ hàng vào trong kho.

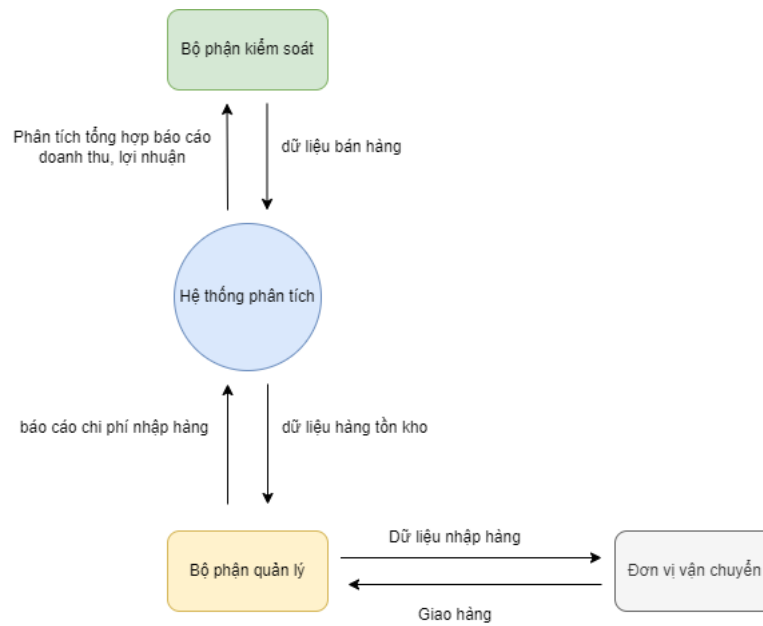
**B4:** nhân viên vận chuyển sẽ chuyển trả lại những lô hàng không đạt chất lượng về cho nhà cung cấp. Tất cả những quy trình trên đều phải thông qua quản lý kho hàng.

**B5:** nhân viên vận chuyển sẽ chuyển những lô hàng đến với các nhà phân phối, cửa hàng, người buôn hàng,...



Hình 2.1: Sơ đồ quy trình hoạt động

### 2.2.2 Luồng dữ liệu



Hình 2.2: Sơ đồ luồng dữ liệu

## 2.3 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 nhầm, trung lập
- Khả năng bảo mật kém

## 2.4 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 hoạt động của cửa hàng. Ngoài ra, hệ thống còn phải được thiết kế có thẩm mỹ, cách vận hành đơn giản.

### 2.4.1 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. . .
- 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 khách hàng, trạng thái của đơn hàng. . .

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.

### 2.4.2 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:

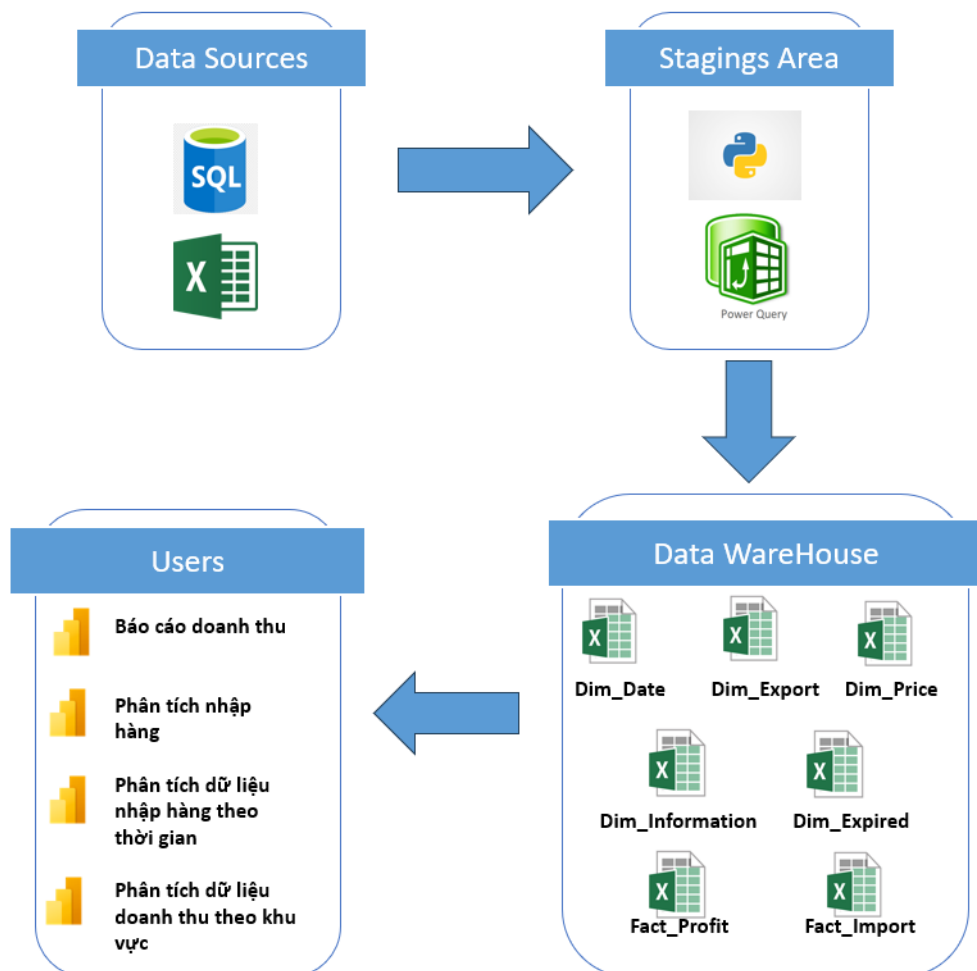
- Báo cáo doanh thu theo tháng, theo sản phẩm
- Báo cáo tỷ lệ lợi nhuận thu về
- Báo cáo sản lượng hàng hóa
- Phân tích doanh số theo quý

- Phân tích thời gian theo đơn vị vận chuyển, loại sản phẩm
- ...

## Chương 3

# Phân tích và thiết kế hệ thống

### 3.1 Kiến trúc của hệ thống phân tích



Hình 3.1: Kiến trúc 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

### 3.2 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 loại bỏ các giá trị null :

	$A_C^B$ ID đơn hàng	$A_C^B$ Thời gian	$A_C^B$ Mã kho hàng	$A_C^B$ Mã sản phẩm	$A_C^B$ Xuất kho	$A_C^B$ Giá bán	$A_C^B$ Doanh thu
1	01012023KH 0009SP 10	1/1/2023	KH 0009	SP 10	48	185000	8880000
2	01012023KH 0001SP 05	1/1/2023	KH 0001	SP 05	170	189000	32130000
3	01012023KH 0001SP 06	1/1/2023	KH 0001	SP 06	574	174000	99876000
4	01012023KH 0017SP 20	1/1/2023	KH 0017	SP 20	480	207000	99360000
5	01012023KH 0001SP 09	1/1/2023	KH 0001	SP 09	134	63000	8442000
6	01012023KH 0001SP 10	1/1/2023	KH 0001	SP 10	595	185000	110075000
7	01012023KH 0001SP 11	1/1/2023	KH 0001	SP 11	257	91000	23387000
8	01012023KH 0001SP 12	1/1/2023	KH 0001	SP 12	471	69000	32499000
9	01012023KH 0001SP 13	1/1/2023	KH 0001	SP 13	97	72000	6984000
10	01012023KH 0001SP 14	1/1/2023	KH 0001	SP 14	433	93000	40269000
11	01012023KH 0100SP 30	1/1/2023	KH 0100	SP 30	185	160000	29600000
12	01012023KH 0001SP 17	1/1/2023	KH 0001	SP 17	144	69000	9936000
13	01012023KH 0001SP 18	1/1/2023	KH 0001	SP 18	369	228000	84132000
14	01012023KH 0001SP 19	1/1/2023	KH 0001	SP 19	48	96000	4608000
15	01012023KH 0001SP 20	1/1/2023	KH 0001	SP 20	288	207000	59616000
16	01012023KH 0001SP 21	1/1/2023	KH 0001	SP 21	29	79000	2291000
17	01012023KH 0001SP 22	1/1/2023	KH 0001	SP 22	101	188000	18988000
18	01012023KH 0001SP 23	1/1/2023	KH 0001	SP 23	423	154000	65142000
19	01012023KH 0001SP 24	1/1/2023	KH 0001	SP 24	33	203000	6699000
20	01012023KH 0001SP 25	1/1/2023	KH 0001	SP 25	145	90000	13050000
21	01012023KH 0001SP 26	1/1/2023	KH 0001	SP 26	289	220000	63580000
22	01012023KH 0001SP 27	1/1/2023	KH 0001	SP 27	519	120000	62280000
23	01012023KH 0001SP 28	1/1/2023	KH 0001	SP 28	317	121000	38357000
24	01012023KH 0001SP 29	1/1/2023	KH 0001	SP 29	114	157000	17898000
25	01012023KH 0001SP 30	1/1/2023	KH 0001	SP 30	505	160000	80800000
26	01012023KH 0100SP 29	1/1/2023	KH 0100	SP 29	549	157000	86193000
27	01012023KH 0002SP 03	1/1/2023	KH 0002	SP 03	418	218000	91124000
28	01012023KH 0002SP 04	1/1/2023	KH 0002	SP 04	486	41000	19926000

Hình 3.2: File dữ liệu sau khi xóa bỏ null





Xóa bỏ các cột không dùng đến:

		$A^B_C$ ID đơn hàng	$A^B_C$ Thời gian	$A^B_C$ Mã kho hàng	$A^B_C$ Mã sản phẩm	$A^B_C$ Xuất kho
1		01012023KH 0009SP 10	1/1/2023	KH 0009	SP 10	48
2		01012023KH 0001SP 05	1/1/2023	KH 0001	SP 05	170
3		01012023KH 0001SP 06	1/1/2023	KH 0001	SP 06	574
4		01012023KH 0017SP 20	1/1/2023	KH 0017	SP 20	480
5		01012023KH 0001SP 09	1/1/2023	KH 0001	SP 09	134
6		01012023KH 0001SP 10	1/1/2023	KH 0001	SP 10	595
7		01012023KH 0001SP 11	1/1/2023	KH 0001	SP 11	257
8		01012023KH 0001SP 12	1/1/2023	KH 0001	SP 12	471
9		01012023KH 0001SP 13	1/1/2023	KH 0001	SP 13	97
10		01012023KH 0001SP 14	1/1/2023	KH 0001	SP 14	433
11		01012023KH 0100SP 30	1/1/2023	KH 0100	SP 30	185
12		01012023KH 0001SP 17	1/1/2023	KH 0001	SP 17	144
13		01012023KH 0001SP 18	1/1/2023	KH 0001	SP 18	369
14		01012023KH 0001SP 19	1/1/2023	KH 0001	SP 19	48
15		01012023KH 0001SP 20	1/1/2023	KH 0001	SP 20	288
16		01012023KH 0001SP 21	1/1/2023	KH 0001	SP 21	29
17		01012023KH 0001SP 22	1/1/2023	KH 0001	SP 22	101
18		01012023KH 0001SP 23	1/1/2023	KH 0001	SP 23	423
19		01012023KH 0001SP 24	1/1/2023	KH 0001	SP 24	33
20		01012023KH 0001SP 25	1/1/2023	KH 0001	SP 25	145
21		01012023KH 0001SP 26	1/1/2023	KH 0001	SP 26	289
22		01012023KH 0001SP 27	1/1/2023	KH 0001	SP 27	519
23		01012023KH 0001SP 28	1/1/2023	KH 0001	SP 28	317
24		01012023KH 0001SP 29	1/1/2023	KH 0001	SP 29	114
25		01012023KH 0001SP 30	1/1/2023	KH 0001	SP 30	505
26		01012023KH 0100SP 29	1/1/2023	KH 0100	SP 29	549
27		01012023KH 0002SP 03	1/1/2023	KH 0002	SP 03	418
28		01012023KH 0002SP 04	1/1/2023	KH 0002	SP 04	486

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

Định dạng lại kiểu dữ liệu cho phù hợp:

	 $A^B_C$ ID đơn hàng	 Thời gian	$A^B_C$ Mã kho hàng	$A^B_C$ Mã sản phẩm	$t^2_3$ Xuất kho
1	01012023KH 0009SP 10	1/1/2023	KH 0009	SP 10	48
2	01012023KH 0001SP 05	1/1/2023	KH 0001	SP 05	170
3	01012023KH 0001SP 06	1/1/2023	KH 0001	SP 06	574
4	01012023KH 0017SP 20	1/1/2023	KH 0017	SP 20	480
5	01012023KH 0001SP 09	1/1/2023	KH 0001	SP 09	134
6	01012023KH 0001SP 10	1/1/2023	KH 0001	SP 10	595
7	01012023KH 0001SP 11	1/1/2023	KH 0001	SP 11	257
8	01012023KH 0001SP 12	1/1/2023	KH 0001	SP 12	471
9	01012023KH 0001SP 13	1/1/2023	KH 0001	SP 13	97
10	01012023KH 0001SP 14	1/1/2023	KH 0001	SP 14	433
11	01012023KH 0100SP 30	1/1/2023	KH 0100	SP 30	185
12	01012023KH 0001SP 17	1/1/2023	KH 0001	SP 17	144
13	01012023KH 0001SP 18	1/1/2023	KH 0001	SP 18	369
14	01012023KH 0001SP 19	1/1/2023	KH 0001	SP 19	48
15	01012023KH 0001SP 20	1/1/2023	KH 0001	SP 20	288
16	01012023KH 0001SP 21	1/1/2023	KH 0001	SP 21	29
17	01012023KH 0001SP 22	1/1/2023	KH 0001	SP 22	101
18	01012023KH 0001SP 23	1/1/2023	KH 0001	SP 23	423
19	01012023KH 0001SP 24	1/1/2023	KH 0001	SP 24	33
20	01012023KH 0001SP 25	1/1/2023	KH 0001	SP 25	145
21	01012023KH 0001SP 26	1/1/2023	KH 0001	SP 26	289
22	01012023KH 0001SP 27	1/1/2023	KH 0001	SP 27	519
23	01012023KH 0001SP 28	1/1/2023	KH 0001	SP 28	317
24	01012023KH 0001SP 29	1/1/2023	KH 0001	SP 29	114
25	01012023KH 0001SP 30	1/1/2023	KH 0001	SP 30	505
26	01012023KH 0100SP 29	1/1/2023	KH 0100	SP 29	549
27	01012023KH 0002SP 03	1/1/2023	KH 0002	SP 03	418
28	01012023KH 0002SP 04	1/1/2023	KH 0002	SP 04	486

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

File dữ liệu sau khi được sắp xếp lại

	A <sup>B</sup> <sub>C</sub> ID đơn hàng	Thời gian	A <sup>B</sup> <sub>C</sub> Mã kho hàng	A <sup>B</sup> <sub>C</sub> Mã sản phẩm	1 <sup>2</sup> <sub>3</sub> Xuất kho
1	01012023KH 0001SP 01	1/1/2023	KH 0001	SP 01	81
2	01012023KH 0001SP 02	1/1/2023	KH 0001	SP 02	597
3	01012023KH 0001SP 03	1/1/2023	KH 0001	SP 03	354
4	01012023KH 0001SP 04	1/1/2023	KH 0001	SP 04	262
5	01012023KH 0001SP 05	1/1/2023	KH 0001	SP 05	170
6	01012023KH 0001SP 06	1/1/2023	KH 0001	SP 06	574
7	01012023KH 0001SP 07	1/1/2023	KH 0001	SP 07	110
8	01012023KH 0001SP 08	1/1/2023	KH 0001	SP 08	488
9	01012023KH 0001SP 09	1/1/2023	KH 0001	SP 09	134
10	01012023KH 0001SP 10	1/1/2023	KH 0001	SP 10	595
11	01012023KH 0001SP 11	1/1/2023	KH 0001	SP 11	257
12	01012023KH 0001SP 12	1/1/2023	KH 0001	SP 12	471
13	01012023KH 0001SP 13	1/1/2023	KH 0001	SP 13	97
14	01012023KH 0001SP 14	1/1/2023	KH 0001	SP 14	433
15	01012023KH 0001SP 15	1/1/2023	KH 0001	SP 15	293
16	01012023KH 0001SP 16	1/1/2023	KH 0001	SP 16	152
17	01012023KH 0001SP 17	1/1/2023	KH 0001	SP 17	144
18	01012023KH 0001SP 18	1/1/2023	KH 0001	SP 18	369
19	01012023KH 0001SP 19	1/1/2023	KH 0001	SP 19	48
20	01012023KH 0001SP 20	1/1/2023	KH 0001	SP 20	288
21	01012023KH 0001SP 21	1/1/2023	KH 0001	SP 21	29
22	01012023KH 0001SP 22	1/1/2023	KH 0001	SP 22	101
23	01012023KH 0001SP 23	1/1/2023	KH 0001	SP 23	423
24	01012023KH 0001SP 24	1/1/2023	KH 0001	SP 24	33
25	01012023KH 0001SP 25	1/1/2023	KH 0001	SP 25	145
26	01012023KH 0001SP 26	1/1/2023	KH 0001	SP 26	289
27	01012023KH 0001SP 27	1/1/2023	KH 0001	SP 27	519
28	01012023KH 0001SP 28	1/1/2023	KH 0001	SP 28	317

Hình 3.5: File dữ liệu sau được sắp xếp lại

Tạo thêm một số cột mới:

- Tạo thêm cột lợi nhuận:

	Thời gian	A <sub>C</sub> Mã kho hàng	A <sub>C</sub> Mã sản phẩm	1 <sub>3</sub> Chi Phí	1 <sub>3</sub> Doanh Thu	1 <sub>3</sub> Lợi Nhuận
1	1/1/2023	KH 0001	SP 01	2754000	3321000	567000
2	1/1/2023	KH 0001	SP 02	94923000	119400000	24477000
3	1/1/2023	KH 0001	SP 03	64428000	77172000	12744000
4	1/1/2023	KH 0001	SP 04	9432000	10742000	1310000
5	1/1/2023	KH 0001	SP 05	25840000	32130000	6290000
6	1/1/2023	KH 0001	SP 06	84952000	99876000	14924000
7	1/1/2023	KH 0001	SP 07	17710000	21560000	3850000
8	1/1/2023	KH 0001	SP 08	57584000	71248000	13664000
9	1/1/2023	KH 0001	SP 09	6700000	8442000	1742000
10	1/1/2023	KH 0001	SP 10	89250000	110075000	20825000
11	1/1/2023	KH 0001	SP 11	18761000	23387000	4626000
12	1/1/2023	KH 0001	SP 12	27789000	32499000	4710000
13	1/1/2023	KH 0001	SP 13	5917000	6984000	1067000
14	1/1/2023	KH 0001	SP 14	34640000	40269000	5629000
15	1/1/2023	KH 0001	SP 15	30179000	38383000	8204000
16	1/1/2023	KH 0001	SP 16	23560000	28728000	5168000
17	1/1/2023	KH 0001	SP 17	8064000	9936000	1872000
18	1/1/2023	KH 0001	SP 18	71586000	84132000	12546000
19	1/1/2023	KH 0001	SP 19	3888000	4608000	720000
20	1/1/2023	KH 0001	SP 20	47808000	59616000	11808000
21	1/1/2023	KH 0001	SP 21	1885000	2291000	406000
22	1/1/2023	KH 0001	SP 22	14645000	18988000	4343000
23	1/1/2023	KH 0001	SP 23	56682000	65142000	8460000
24	1/1/2023	KH 0001	SP 24	5610000	6699000	1089000
25	1/1/2023	KH 0001	SP 25	10295000	13050000	2755000
26	1/1/2023	KH 0001	SP 26	54621000	63580000	8959000
27	1/1/2023	KH 0001	SP 27	48786000	62280000	13494000
28	1/1/2023	KH 0001	SP 28	30749000	38357000	7608000

Hình 3.6: Tạo thêm cột lợi nhuận

- Tạo thêm cột phần trăm lợi nhuận:

	Thời gian	A <sup>B</sup> <sub>C</sub> Mã kho hàng	A <sup>B</sup> <sub>C</sub> Mã sản phẩm	1 <sup>2</sup> <sub>3</sub> Chi Phí	1 <sup>2</sup> <sub>3</sub> Doanh Thu	1 <sup>2</sup> <sub>3</sub> Lợi Nhuận	% Phần trăm lợi nhuận
1	1/1/2023	KH 0001	SP 01	2754000	3321000	567000	20.59%
2	1/1/2023	KH 0001	SP 02	94923000	119400000	24477000	25.79%
3	1/1/2023	KH 0001	SP 03	64428000	77172000	12744000	19.78%
4	1/1/2023	KH 0001	SP 04	9432000	10742000	1310000	13.89%
5	1/1/2023	KH 0001	SP 05	25840000	32130000	6290000	24.34%
6	1/1/2023	KH 0001	SP 06	84952000	99876000	14924000	17.57%
7	1/1/2023	KH 0001	SP 07	17710000	21560000	3850000	21.74%
8	1/1/2023	KH 0001	SP 08	57584000	71248000	13664000	23.73%
9	1/1/2023	KH 0001	SP 09	6700000	8442000	1742000	26.00%
10	1/1/2023	KH 0001	SP 10	89250000	110075000	20825000	23.33%
11	1/1/2023	KH 0001	SP 11	18761000	23387000	4626000	24.66%
12	1/1/2023	KH 0001	SP 12	27789000	32499000	4710000	16.95%
13	1/1/2023	KH 0001	SP 13	5917000	6984000	1067000	18.03%
14	1/1/2023	KH 0001	SP 14	34640000	40269000	5629000	16.25%
15	1/1/2023	KH 0001	SP 15	30179000	38383000	8204000	27.18%
16	1/1/2023	KH 0001	SP 16	23560000	28728000	5168000	21.94%
17	1/1/2023	KH 0001	SP 17	8064000	9936000	1872000	23.21%
18	1/1/2023	KH 0001	SP 18	71586000	84132000	12546000	17.53%
19	1/1/2023	KH 0001	SP 19	3888000	4608000	720000	18.52%
20	1/1/2023	KH 0001	SP 20	47808000	59616000	11808000	24.70%
21	1/1/2023	KH 0001	SP 21	1885000	2291000	406000	21.54%
22	1/1/2023	KH 0001	SP 22	14645000	18988000	4343000	29.66%
23	1/1/2023	KH 0001	SP 23	56682000	65142000	8460000	14.93%
24	1/1/2023	KH 0001	SP 24	5610000	6699000	1089000	19.41%
25	1/1/2023	KH 0001	SP 25	10295000	13050000	2755000	26.76%
26	1/1/2023	KH 0001	SP 26	54621000	63580000	8959000	16.40%
27	1/1/2023	KH 0001	SP 27	48786000	62280000	13494000	27.66%
28	1/1/2023	KH 0001	SP 28	30749000	38357000	7608000	24.74%

Hình 3.7: Tạo thêm cột phần trăm lợi nhuận

### 3.3 Xây dựng Dimension-Fact

#### 3.3.1 Dimension

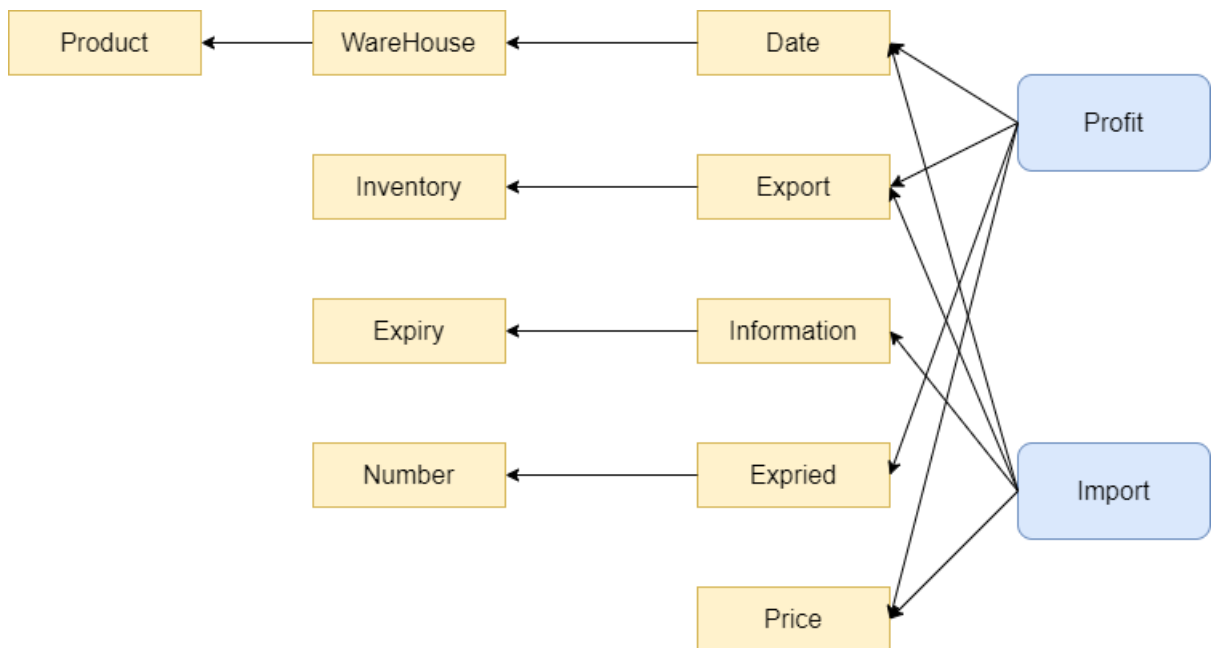
Tên bảng	Mô tả
Dim Date	Chứa thông tin về thời gian nhập, xuất của các loại hàng theo từng ngày, tháng, năm
Dim Export	Dữ liệu về hàng nhập kho
Dim Price	Dữ liệu về giá nhập, xuất của các loại hàng
Dim Information	Chứa thông tin về các loại hàng
Dim Expried	dữ liệu về sản phẩm đã hết hạn

#### 3.3.2 Fact

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

- Fact Profit: Phân tích về lợi nhuận kho hàng qua từng ngày, tháng, năm
- Fact Import: Phân tích số lượng hàng xuất ra và tồn kho để tính toán số lượng hàng nhập.

### 3.3.3 Mô hình dữ liệu quan hệ



Hình 3.8: Mô hình dữ liệu quan hệ

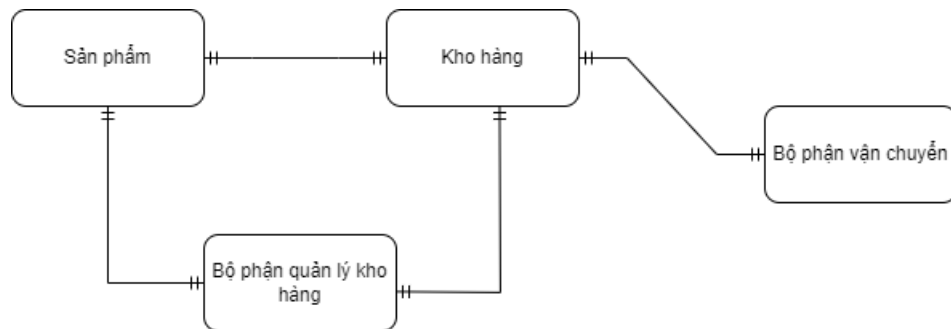
## 3.4 Thiết kế cơ sở dữ liệu hệ thống OLTP

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 3.9: Mô hình OLTP

Mô tả các bảng trong cơ sở dữ liệu OLTP

- Sản phẩm: Chứa thông tin về các sản phẩm
- Bộ phận quản lý kho hàng: Chứa thông tin về số lượng của từng sản phẩm của tất cả kho hàng.
- Kho hàng: Chứa thông tin về các kho hàng và số lượng hàng còn lại trong kho của từng sản phẩm.
- Bộ phận vận chuyển: Chứa thông tin về số lượng hàng nhập, xuất của từng sản phẩm.

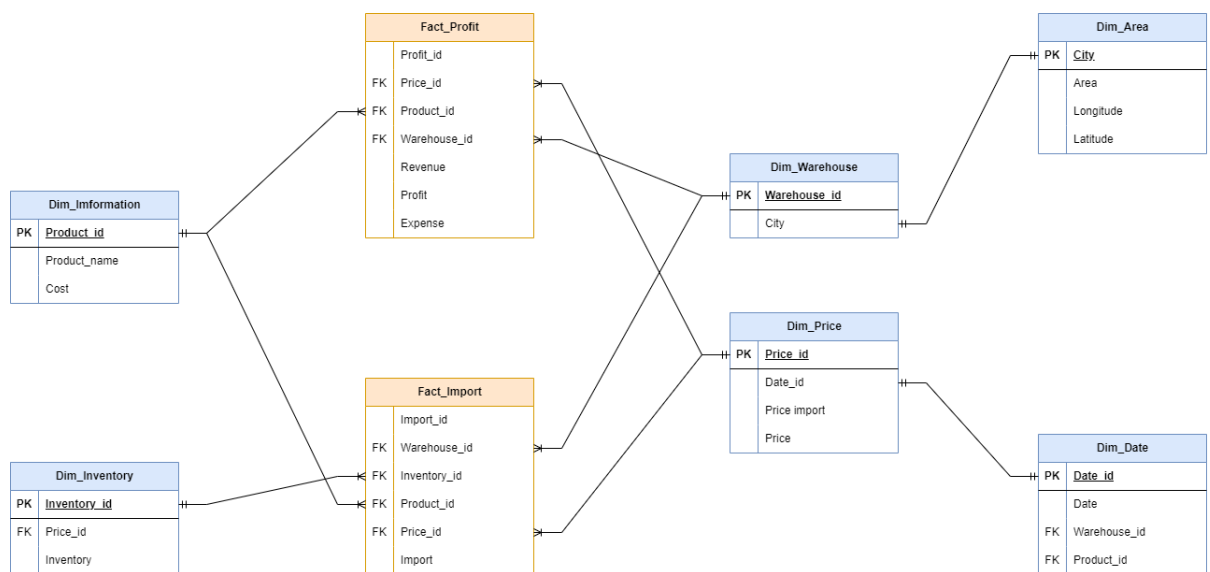
### 3.5 Thiết kế cơ sở dữ liệu hệ thống OLAP

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 được, 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, khu vực, 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.

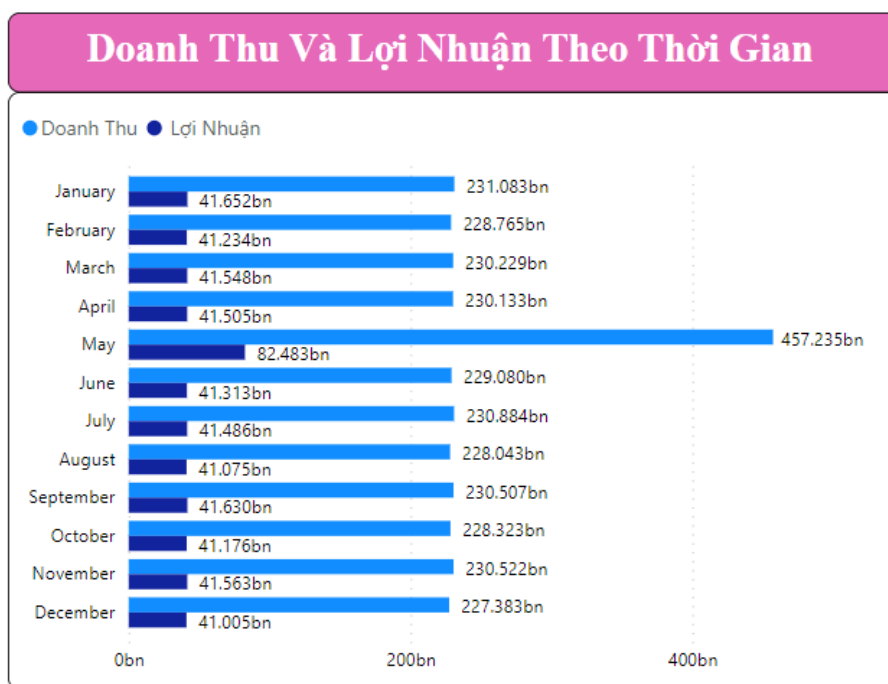


Hình 3.10: Mô hình OLAP



## 3.6 Khai phá dữ liệu

### 3.6.1 Theo thời gian

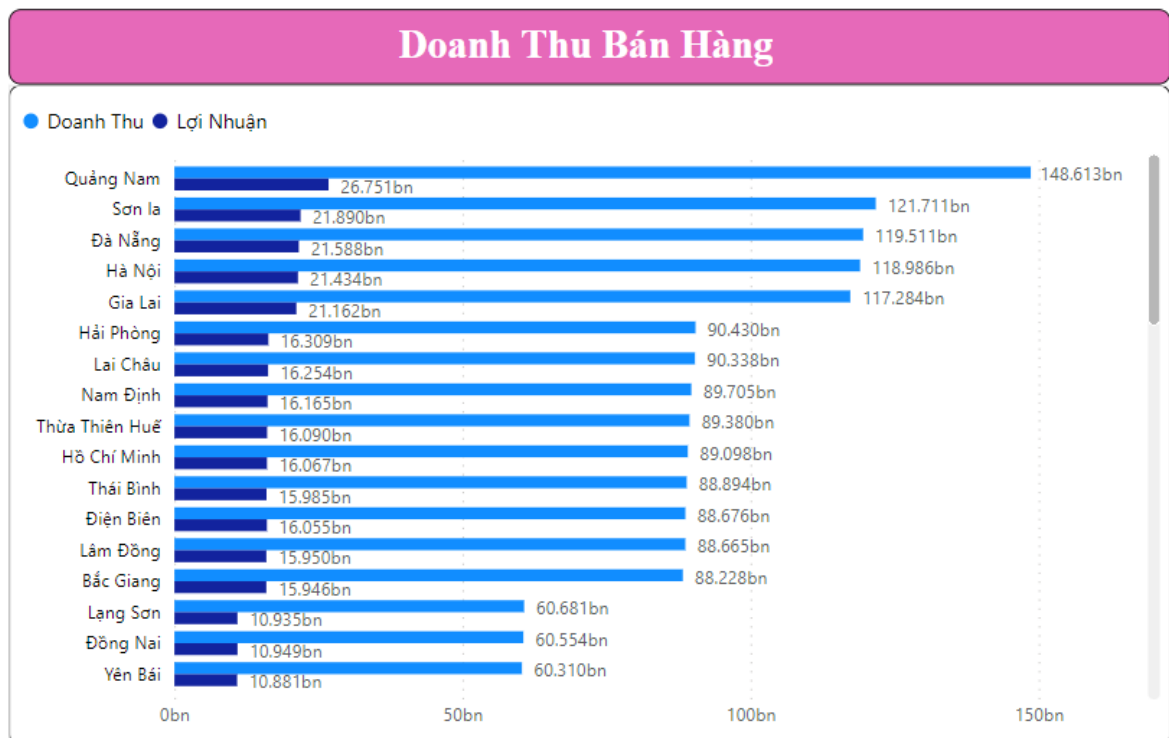


Hình 3.11: Doanh thu và lợi nhuận theo thời gian

Lý do vì sao tháng 5 lại có doanh thu và lợi nhuận vượt trội hơn các tháng còn lại :

- Tháng 5 là tháng bắt đầu sang hè, mọi người có nhu cầu mua sắm đồ để chuẩn bị cho một mùa hè mới.
- Các em học sinh được nghỉ hè vào tháng 5 này, phụ huynh có xu hướng cho các con đi du lịch hè sau một năm học dài, đây cũng là tháng mọi người chọn đi nghỉ mát, đi chơi vì thế sẽ mua nhiều đồ để chuẩn bị cho một chuyến du lịch thật vui vẻ.

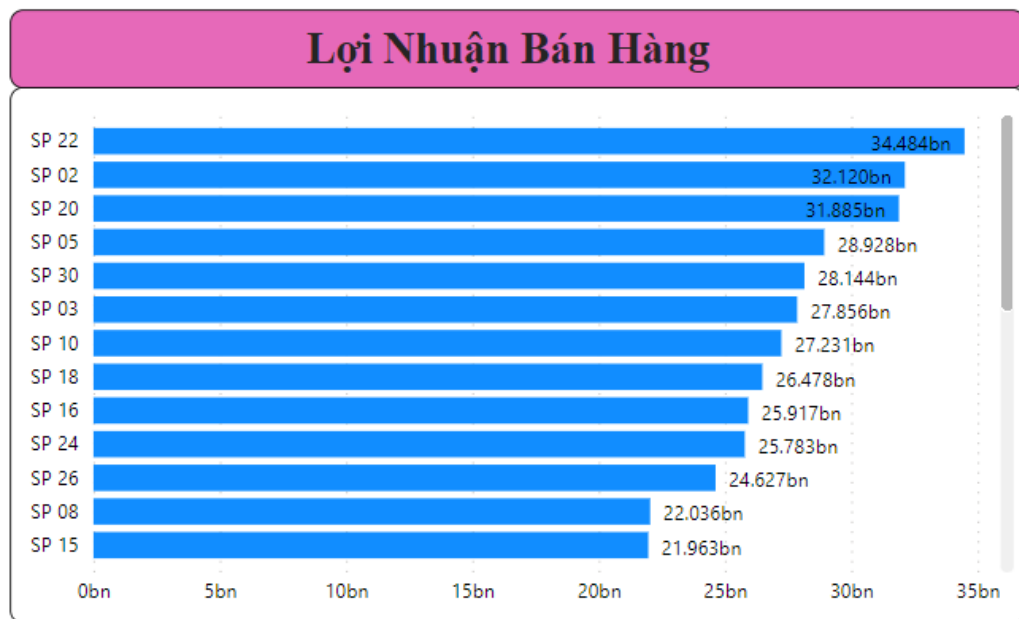
### 3.6.2 Theo doanh thu



Hình 3.12: Doanh thu bán hàng

- Các tỉnh, thành phố như Quảng Nam, Đà Nẵng có nhiệt độ trung bình năm cao hơn những tỉnh ngoài miền Bắc nên nhu cầu về quần áo mùa hè sẽ cao hơn và nhịp sống của con người nơi đây rất phát triển về cả đời sống, xã hội nên sẽ có doanh thu cao ở những khu vực này.
- Các thành phố lớn như Hà Nội, Hải Phòng, Đà Nẵng ... thuộc nơi dân cư đông đúc nên doanh thu khu vực này cũng sẽ cao hơn.

### 3.6.3 Theo lợi nhuận



Hình 3.13: Lợi nhuận bán hàng

- Giới trẻ hiện nay thích những bộ đồ thoải mái nhưng nổi bật và chất liệu từ bò như SP02 (áo phong họa tiết), SP10 (áo phong đen họa tiết), SP22 (quần đùi bò), SP20 (quần bò rách gối) và cũng có thể là những kiểu dáng rườm rà nhưng vẫn năng động như SP05 (áo polo).
- Bên cạnh đó, phụ kiện cũng là một món không thể thiếu khi phối đồ, đa số hiện nay mọi người đều khá thích đi giày khi ra đường, đặc biệt là đi chơi hay đi chụp ảnh,... Vì thế những sản phẩm về tất không thể không đi kèm như SP30 (tất hình hoa),...

Thông Tin Sản Phẩm			
Mã sản phẩm	Tên sản phẩm	Giá nhập	Giá bán
SP 01	Áo sơ mi trắng	159000	200000
SP 02	Áo phông họa tiết	182000	218000
SP 03	Áo croptop	152000	189000
SP 04	Áo sơ mi kẻ sọc	148000	174000
SP 05	Áo polo	161000	196000
SP 06	Áo phông trơn	150000	185000
SP 07	Áo ba lỗ	73000	91000
SP 08	Áo phông mặt cười	103000	131000
SP 09	Áo polo tay lửng	155000	155000
SP 10	Áo phông đen họa tiết	166000	207000
SP 11	Áo phông tay lửng hình mèo	134000	154000
SP 12	Áo ba lỗ vải lanh	71000	90000
SP 13	Áo sơ mi caro	124000	160000
SP 14	Quần đùi bò xanh	118000	146000
SP 15	Quần đùi vải	80000	93000

Hình 3.14: Thông tin sản phẩm

Thông Tin Sản Phẩm			
Mã sản phẩm	Tên sản phẩm	Giá nhập	Giá bán
SP 15	Quần đùi vải	93000	93000
SP 16	Quần bò dài	194000	228000
SP 17	Quần sooc vải	81000	96000
SP 18	Quần vải lạnh	65000	79000
SP 19	Quần đùi hoa	145000	188000
SP 20	Quần bò rách gối	170000	203000
SP 21	Quần đùi nhung tằm	94000	120000
SP 22	Quần đùi bò đen	97000	121000
SP 23	Bộ đồ ngủ đùi	189000	220000
SP 24	Bộ đồ ngủ vải lạnh	132000	157000
SP 25	Mũ lưỡi trai	50000	63000
SP 26	Găng tay	36000	41000
SP 27	Khẩu trang vải	56000	69000
SP 28	Tất trơn	34000	41000
SP 29	Tất hình mặt cười	59000	69000
SP 30	Tất hình hoa	61000	72000

Hình 3.15: Thông tin sản phẩm

## Chương 4

# Xây dựng hệ thống

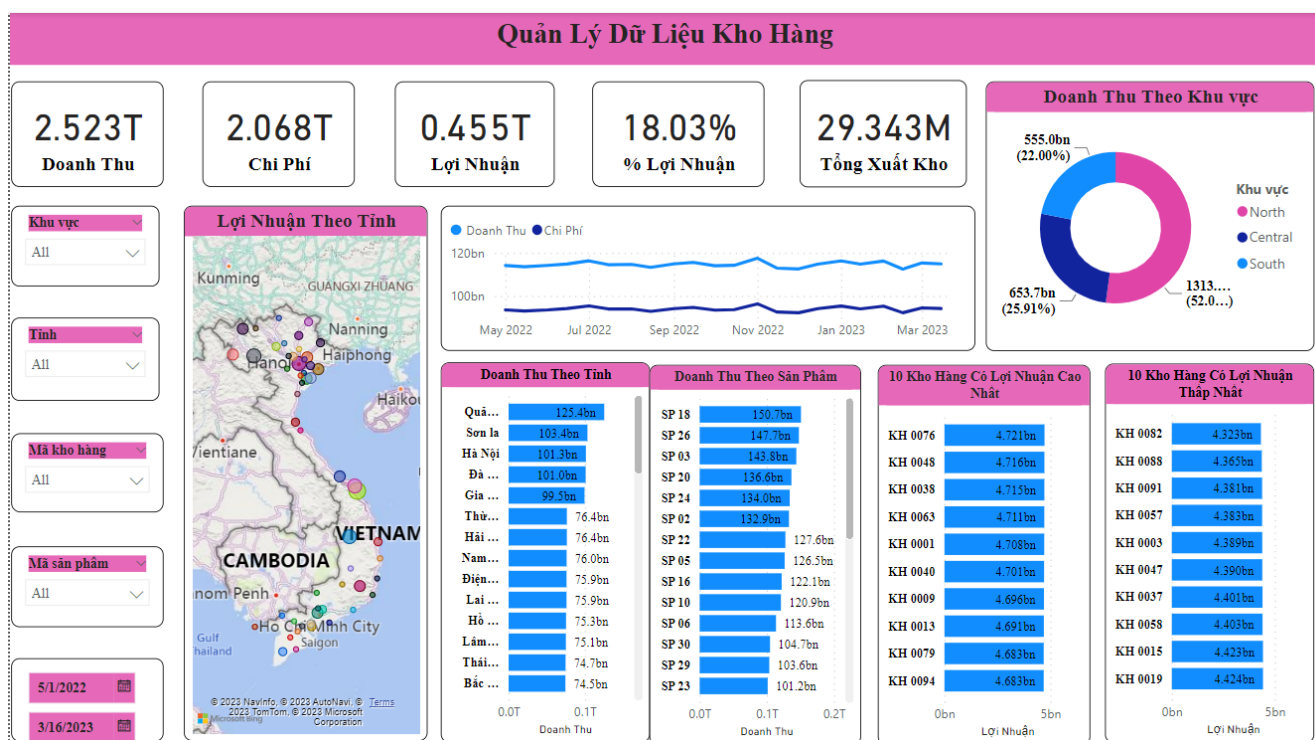
### 4.1 Truyền dữ liệu

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 và Fact
- Sau đó, dữ liệu được truyền vào cơ sở dữ liệu OLAP và tải lên hệ thống Power BI

### 4.2 Xây dựng hệ thống Power BI



Hình 4.1: Dashboard quản lý kho hàng

# Kết luận

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 thầy và bạn đọc để đề tài này được hoàn thiện hơn.

Em xin chân thành cảm ơn !



# 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