

BỘ CÔNG THƯƠNG
TRƯỜNG ĐẠI HỌC CÔNG THƯƠNG TP.HCM
KHOA CÔNG NGHỆ THÔNG TIN



BÁO CÁO BÀI TẬP NHÓM
THỰC HÀNH KHO DỮ LIỆU VÀ OLAP
Đề tài: XÂY DỰNG KHO DỮ LIỆU PHÂN TÍCH KINH
DOANH CỦA CỬA HÀNG BÁN QUẦN ÁO

.....

Giảng viên hướng dẫn : Đinh Thị Mận

TP. HỒ CHÍ MINH – 2024

MỤC LỤC

CHƯƠNG 1 TỔNG QUAN ĐỀ TÀI.....	6
1.1. GIỚI THIỆU	6
1.2. MỤC TIÊU VÀ PHẠM VI ĐỀ TÀI	7
1.2.1. Mục tiêu tổng quát	7
1.2.2. Mục tiêu cụ thể	7
1.2.3. Phạm vi về nội dung	8
1.2.4. Phạm vi về không gian.....	8
1.3. KHẢO SÁT HỆ THỐNG.....	9
1.3.1. Khảo sát thực trạng	9
1.4. ĐỐI TƯỢNG NGHIÊN CỨU	10
1.4.1. Đối tượng nghiên cứu chính	10
1.4.2. Đối tượng người dùng.....	11
CHƯƠNG 2 CƠ SỞ LÝ THUYẾT	12
2.1. ĐỐI TƯỢNG NGHIÊN CỨU	12
2.1.1. Định nghĩa kho dữ liệu (Data Warehouse)	12
2.1.2. Vai trò của kho dữ liệu trong kinh doanh	13
2.2. OLAP TRONG KHO DỮ LIỆU	14
2.2.1. Định nghĩa OLAP (Online Analytical Processing).....	14
2.2.2. Các mô hình OLAP.....	15
2.3. QUY TRÌNH ETL TRONG XÂY DỰNG KHO DỮ LIỆU OLAP	16
2.3.1. Extract (Trích xuất dữ liệu).....	16
2.3.2. Transform (Chuyển đổi dữ liệu)	16
2.3.3. Load (Nạp dữ liệu vào kho dữ liệu).....	17
2.4. CÔNG CỤ XÂY DỰNG KHO DỮ LIỆU OLAP.....	17
2.4.1. Công cụ ETL (Extract, Transform, Load)	17
2.4.2. Công cụ lưu trữ dữ liệu OLAP	18
2.4.3. Công cụ trực quan hóa và phân tích dữ liệu	18
CHƯƠNG 3 THIẾT KẾ KHO DỮ LIỆU.....	20
3.1. PHÂN TÍCH ĐỀ TÀI	20
3.1.1. Phân tích yêu cầu	20

3.1.2. Yêu cầu hệ thống	21
3.2. THIẾT KẾ HỆ THỐNG	23
3.2.1. Cơ sở dữ liệu.....	23
3.2.2. Kho dữ liệu	28
3.2.3. Lược đồ quan hệ	31
3.2.4. SSIS	33
3.2.4. SSAS.....	39
CHƯƠNG 4 CÀI ĐẶT ỨNG DỤNG	41
CHƯƠNG 5 KẾT LUẬN	52
5.1. KẾT QUẢ.....	54
5.1.1. Về mặt chức năng	54
5.1.2. Về mặt kỹ thuật.....	54
5.2. HẠN CHẾ.....	55
5.2.1. Hạn chế về chức năng.....	55
5.2.2. Hạn chế về kỹ thuật	55
5.4. KẾT LUẬN.....	55
TÀI LIỆU THAM KHẢO.....	56

DANH MỤC HÌNH ẢNH

Hình 3.1: Database Diagram.....	31
Hình 3.2: Kho dữ liệu mô hình chòm sao.....	32
Hình 3.3: SSIS	33
Hình 3.4: File Excel	33
Hình 3.5: SSIS Dim_Time.....	34
Hình 3.6: SSIS Dim_Coupon.....	35
Hình 3.7: SSIS Dim_Customer.....	35
Hình 3.8: SSIS Dim_Product.....	36
Hình 3.9: SSIS Dim_Geography	36
Hình 3.10: SSIS Dim_Warehouse	37
Hình 3.11: SSIS Fact	38
Hình 3.12: SSAS.....	39
Hình 3.13: SSAS Cube	40
Hình 4.1: Giao diện trang chủ.....	41
Hình 4.2: Giao diện truy vấn MDX	42
Hình 4.3: Giao diện thực hiện xuất PDF	43
Hình 4.4: File PDF xuất ra.....	43
Hình 4.5: Thống kê 1	44
Hình 4.6: Thống kê 2	45
Hình 4.7: Thống kê 3	46
Hình 4.8: Thống kê 4	47
Hình 4.9: Thống kê 5	48
Hình 4.10: Giao diện Tableau.....	49
Hình 4.11: Truy vấn tổng doanh thu theo năm và lọc theo doanh thu > 0	50
Hình 4.12: Truy vấn tổng doanh thu theo thành phố và lọc theo doanh thu > 1000	50
Hình 4.13: Truy vấn tổng doanh thu theo sản phẩm và khách hàng, để tính doanh thu trung bình	50
Hình 4.14: Truy vấn lấy top 5 sản phẩm theo doanh thu trong năm hiện tại và sắp xếp theo thứ tự giảm dần	51
Hình 4.15: Truy vấn tổng doanh thu theo quý và sắp xếp theo thứ tự giảm dần.....	51
Hình 4.16: Truy vấn doanh thu trung bình theo dòng sản phẩm và năm.....	52

Hình 4.17: Truy vấn tổng doanh thu và số lượng đơn hàng theo lãnh thổ	52
Hình 4.18: Truy vấn tổng số lượng đơn hàng theo khách hàng.....	52
Hình 4.20: Truy vấn tổng doanh thu theo quý	53
Hình 4.19: Truy vấn tổng số lượng đơn hàng theo sản phẩm.....	53

CHƯƠNG 1

TỔNG QUAN ĐỀ TÀI

1.1. GIỚI THIỆU

Đề tài xây dựng kho dữ liệu phân tích kinh doanh cửa hàng bán quần áo với việc sử dụng dữ liệu để đưa ra các chiến lược kinh doanh, thống kê một cách chính xác và kịp thời đang là một yếu tố quan trọng trong lĩnh vực kinh doanh thời trang. Kho dữ liệu này không chỉ hỗ trợ các cửa hàng quần áo và lưu trữ dữ liệu từ nhiều nguồn khác nhau mà còn lập báo cáo và giám sát chi tiết để cải thiện phương pháp tối ưu hóa bán hàng và kiểm soát hàng tồn kho, đồng thời nâng cao dịch vụ khách hàng.

1.2. MỤC TIÊU VÀ PHẠM VI ĐỀ TÀI

1.2.1. Mục tiêu tổng quát

Xây dựng một hệ thống kho dữ liệu phân tích kinh doanh hiện đại và hiệu quả cho cửa hàng bán quần áo, giúp cải thiện khả năng quản lý, tối ưu hóa hoạt động kinh doanh và hỗ trợ ra quyết định dựa trên dữ liệu.

1.2.2. Mục tiêu cụ thể

Kho lưu trữ dữ liệu được kiến trúc và tạo ra để đáp ứng đầy đủ các yêu cầu, thao tác và trích xuất chi tiết từ nhiều nguồn khác nhau như hệ thống quản lý bán lẻ, hồ sơ khách hàng và dữ liệu sản phẩm. Hệ thống còn đưa ra các báo cáo thể hiện cách thức hoạt động của doanh nghiệp. Các chỉ số như thu nhập, lượng hàng tồn kho và tác động tiếp thị được hiển thị một cách rõ ràng và dễ hiểu, mang lại sự hỗ trợ quyết định mạnh mẽ. Hiệu suất của hệ thống được đánh giá bằng các số liệu kinh doanh thực tế trong hoạt động hàng ngày của cửa hàng.

1.2.3. Phạm vi về nội dung

Đề tài tập trung vào việc xây dựng hệ thống kho dữ liệu phục vụ phân tích kinh doanh cho cửa hàng bán quần áo, bao gồm các nội dung:

- Thu thập dữ liệu: Thu thập từ excel của trang web bán quần áo và database của cửa hàng bán quần áo
- Xử lý dữ liệu: Triển khai quy trình ETL (Extract, Transform, Load) để làm sạch và chuẩn hóa dữ liệu từ nhiều nguồn khác nhau.
- Lưu trữ dữ liệu: Chuyển từ các quy trình xử lý dữ liệu ra mô hình DDS
- Phân tích dữ liệu: Phân tích từ SSAS

1.2.4. Phạm vi về không gian

Đề tài được triển khai trong phạm vi hoạt động của một cửa hàng bán quần áo cụ thể. Phạm vi không gian tập trung vào dữ liệu nội bộ của cửa hàng, bao gồm thông tin bán hàng, quản lý hàng hóa, và thông tin khách hàng trong hệ thống hiện tại.

1.3. KHẢO SÁT HỆ THỐNG

1.3.1. Khảo sát thực trạng

Thực trạng quản lý dữ liệu:

Hiện tại, công việc quản lý dữ liệu tại cửa hàng bán quần áo vẫn còn khá lộn xộn và thiếu tính kết nối. Mọi thứ chủ yếu được thực hiện qua các hệ thống rời rạc và công cụ thủ công như file Excel. Dữ liệu từ các bộ phận như bán hàng, tồn kho và thông tin khách hàng không đồng bộ và không có sự liên kết chặt chẽ. Điều này khiến dữ liệu bị phân tán, khó kiểm soát, thậm chí còn bị trùng lặp hoặc sai lệch.

Thực trạng phân tích và ra quyết định kinh doanh:

Cửa hàng vẫn chưa áp dụng các công cụ phân tích hiện đại để hỗ trợ quá trình ra quyết định kinh doanh. Hiện tại, các báo cáo chỉ dừng lại ở mức tổng hợp đơn giản, chưa thể phản ánh toàn diện mọi khía cạnh hoạt động của cửa hàng.

Cửa hàng vẫn còn thiếu khả năng phân tích chuyên sâu, như xác định nhóm khách hàng tiềm năng, nắm bắt xu hướng mua sắm hay đo lường hiệu quả của các chương trình khuyến mãi. Thay vì dựa vào dữ liệu cụ thể, các quyết định phần lớn vẫn dựa vào kinh nghiệm cá nhân, khiến hiệu quả kinh doanh chưa thực sự ổn định.

Thách thức lớn:

Hiện tại, hoạt động quản lý dữ liệu tại cửa hàng bán quần áo vẫn còn phụ thuộc vào các hệ thống rời rạc và công cụ thủ công như file Excel. Dữ liệu từ các bộ phận như bán hàng, tồn kho và thông tin khách hàng vẫn đang "đứng riêng lẻ", chưa được kết nối một cách chặt chẽ. Điều này khiến dữ liệu trở nên phân tán, khó kiểm soát và không ít lần rơi vào tình trạng trùng lặp hoặc sai lệch.

1.4. ĐỐI TƯỢNG NGHIÊN CỨU

1.4.1. Đối tượng nghiên cứu chính

Đối tượng nghiên cứu chính của đề tài là dữ liệu kinh doanh liên quan đến hoạt động của cửa hàng bán quần áo. Các loại dữ liệu này bao gồm:

- Dữ liệu bán hàng: Ghi nhận doanh thu, số lượng sản phẩm bán ra, thời điểm giao dịch, và các chương trình khuyến mãi đi kèm.
- Dữ liệu hàng tồn kho: Thông tin về số lượng hàng hóa còn lại, tình trạng nhập - xuất hàng, và các vấn đề liên quan đến quản lý tồn kho.
- Dữ liệu khách hàng: Hồ sơ khách hàng, lịch sử mua sắm, các phân nhóm khách hàng (khách hàng tiềm năng, khách hàng trung thành).
- Dữ liệu từ thị trường: Xu hướng thời trang, hành vi tiêu dùng, và các yếu tố bên ngoài ảnh hưởng đến nhu cầu mua sắm.
- Dữ liệu từ vị trí địa lý: Thông tin về địa điểm mua hàng, từ mức quốc gia, tiểu bang, thành phố, quận.
- Dữ liệu từ coupon: Lấy dữ liệu thống kê về các mã giảm giá được sử dụng

1.4.2. Đối tượng người dùng

Chủ cửa hàng:

Chủ cửa hàng đóng vai trò là người đưa ra các quyết định chiến lược, sử dụng hệ thống để nắm bắt bức tranh tổng thể về hoạt động kinh doanh. Dựa trên các báo cáo chi tiết và phân tích xu hướng, chủ cửa hàng có thể xây dựng các kế hoạch dài hạn nhằm tối ưu hóa doanh thu và lợi nhuận.

Quản lý cửa hàng:

Quản lý cửa hàng chịu trách nhiệm điều hành các hoạt động vận hành hàng ngày. Họ sử dụng hệ thống để kiểm soát tình trạng tồn kho, điều phối hoạt động nhập hàng và đánh giá hiệu suất làm việc của nhân viên, đảm bảo mọi hoạt động diễn ra trơn tru và hiệu quả.

CHƯƠNG 2

CƠ SỞ LÝ THUYẾT

2.1. ĐỐI TƯỢNG NGHIÊN CỨU

2.1.1. Định nghĩa kho dữ liệu (Data Warehouse)

Kho dữ liệu (Data Warehouse) là một hệ thống tập trung lưu trữ và tổ chức dữ liệu từ nhiều nguồn khác nhau nhằm phục vụ các mục đích phân tích và hỗ trợ ra quyết định. Đây là một nền tảng quan trọng trong việc triển khai các hệ thống thông tin hỗ trợ quyết định (DSS - Decision Support Systems), đặc biệt là các ứng dụng xử lý phân tích trực tuyến (OLAP - Online Analytical Processing).

Khác với cơ sở dữ liệu giao dịch (OLTP - Online Transaction Processing), nơi dữ liệu được xử lý để phục vụ các giao dịch hàng ngày, kho dữ liệu tập trung vào việc lưu trữ và phân tích dữ liệu lịch sử. Dữ liệu trong kho thường được tổ chức theo các chủ đề cụ thể (subject-oriented), không thay đổi (non-volatile), và mang tính thời gian (time-variant).

Các đặc điểm chính của kho dữ liệu

- **Tích hợp (Integration):**

Kho dữ liệu tích hợp dữ liệu từ nhiều nguồn khác nhau, bao gồm các cơ sở dữ liệu giao dịch, file Excel, và thậm chí cả dữ liệu từ mạng xã hội. Quy trình này đòi hỏi xử lý và chuẩn hóa dữ liệu để đảm bảo tính nhất quán.

- **Định hướng chủ đề (Subject-oriented):**

Dữ liệu trong kho được tổ chức xung quanh các lĩnh vực cụ thể trong doanh nghiệp, như bán hàng, khách hàng, sản phẩm, hoặc tài kho. Điều này giúp

doanh nghiệp dễ dàng truy cập và phân tích dữ liệu liên quan đến một chủ đề cụ thể mà không bị phân tán bởi các thông tin không cần thiết.

- **Không thay đổi (Non-volatile):**

Một khi dữ liệu đã được nạp vào kho, nó không bị chỉnh sửa hoặc xóa đi. Điều này giúp duy trì lịch sử dữ liệu và cho phép thực hiện các phân tích dài hạn dựa trên dữ liệu chính xác.

- **Dữ liệu thời gian (Time-variant):**

Dữ liệu trong kho thường chứa thông tin liên quan đến các mốc thời gian cụ thể, như ngày, tuần, tháng, hoặc năm. Điều này cho phép phân tích xu hướng, so sánh các giai đoạn khác nhau, và hỗ trợ dự đoán tương lai.

2.1.2. Vai trò của kho dữ liệu trong kinh doanh

Kho dữ liệu đóng vai trò trung tâm trong việc hỗ trợ các hoạt động kinh doanh hiện đại, đặc biệt là trong các doanh nghiệp bán lẻ và thời trang. Dưới đây là những lợi ích chính mà kho dữ liệu mang lại:

- **Hỗ trợ ra quyết định chiến lược:**

Dữ liệu được tổ chức và trình bày một cách rõ ràng, giúp nhà quản lý có được cái nhìn toàn diện về hoạt động kinh doanh và đưa ra các quyết định chiến lược dựa trên thông tin thực tế.

- **Phân tích xu hướng:**

Kho dữ liệu cho phép theo dõi và phân tích các xu hướng kinh doanh, chẳng hạn như xu hướng tiêu dùng, hành vi mua sắm của khách hàng, hoặc hiệu quả của các chương trình khuyến mãi.

- **Tăng hiệu quả hoạt động:**

Việc tích hợp và tự động hóa dữ liệu từ nhiều nguồn khác nhau giúp giảm thiểu thời gian và chi phí liên quan đến việc xử lý dữ liệu thủ công.

- **Nâng cao khả năng cạnh tranh:**

Với khả năng phân tích nhanh chóng và chính xác, doanh nghiệp có thể nắm bắt các cơ hội thị trường, cải thiện chiến lược bán hàng, và tối ưu hóa các quy trình kinh doanh.

- **Cải thiện dịch vụ khách hàng:**

Phân tích dữ liệu khách hàng giúp doanh nghiệp hiểu rõ hơn về nhu cầu và mong muốn của họ, từ đó cung cấp dịch vụ và sản phẩm phù hợp hơn.

2.2. OLAP TRONG KHO DỮ LIỆU

2.2.1. Định nghĩa OLAP (Online Analytical Processing)

OLAP (Online Analytical Processing) là một công nghệ quan trọng trong hệ thống kho dữ liệu, cho phép người dùng phân tích dữ liệu đa chiều để đưa ra các quyết định kinh doanh. Khác với OLTP (Online Transaction Processing) tập trung vào việc xử lý giao dịch, OLAP tập trung vào việc phân tích dữ liệu để hỗ trợ các truy vấn phức tạp.

Các đặc điểm của OLAP

- **Hỗ trợ phân tích đa chiều (Multidimensional Analysis):**

OLAP tổ chức dữ liệu dưới dạng các khối (cubes), với mỗi chiều (dimension) đại diện cho một khía cạnh phân tích, chẳng hạn như thời gian, sản phẩm, địa điểm, và doanh số.

- **Truy vấn phức tạp (Complex Querying):**

OLAP cho phép thực hiện các truy vấn phức tạp như tổng hợp, phân tích xu hướng, hoặc so sánh dữ liệu giữa các giai đoạn khác nhau.

- **Hiệu suất cao:**

Dữ liệu trong OLAP thường được tổ chức và lập chỉ mục trước, giúp tăng tốc độ truy vấn ngay cả khi làm việc với khối lượng lớn dữ liệu.

2.2.2. Các mô hình OLAP

- **MOLAP (Multidimensional OLAP):**

MOLAP lưu trữ dữ liệu trong các khối đa chiều và tối ưu hóa cho việc truy vấn nhanh. Đây là mô hình OLAP truyền thống với hiệu suất cao trong các phân tích phức tạp.

- **ROLAP (Relational OLAP):**

ROLAP lưu trữ dữ liệu trong cơ sở dữ liệu quan hệ và sử dụng SQL để thực hiện truy vấn. Phù hợp với các hệ thống có khối lượng dữ liệu lớn và thường xuyên thay đổi.

- **HOLAP (Hybrid OLAP):**

HOLAP kết hợp ưu điểm của MOLAP và ROLAP, cho phép lưu trữ dữ liệu tổng hợp trong khối đa chiều (MOLAP) và dữ liệu chi tiết trong cơ sở dữ liệu quan hệ (ROLAP).

2.3. QUY TRÌNH ETL TRONG XÂY DỰNG KHO DỮ LIỆU OLAP

2.3.1. Extract (Trích xuất dữ liệu)

Quy trình trích xuất dữ liệu là bước đầu tiên trong ETL, nhằm thu thập dữ liệu từ nhiều nguồn khác nhau. Các nguồn dữ liệu phổ biến bao gồm:

- Hệ thống POS (Point of Sale): Dữ liệu giao dịch bán hàng.
- Hệ thống ERP (Enterprise Resource Planning): Dữ liệu quản lý hàng tồn kho, nhà cung cấp.
- CRM (Customer Relationship Management): Dữ liệu khách hàng và hành vi tiêu dùng.

2.3.2. Transform (Chuyển đổi dữ liệu)

Dữ liệu thô sau khi được trích xuất sẽ trải qua quá trình làm sạch và chuẩn hóa, bao gồm:

- Loại bỏ dữ liệu bị lỗi, không hợp lệ hoặc trùng lặp.
- Chuẩn hóa định dạng dữ liệu, ví dụ: chuyển đổi tất cả ngày tháng về cùng một định dạng chuẩn.
- Tích hợp dữ liệu từ nhiều nguồn để tạo thành một tập dữ liệu duy nhất, nhất quán.

2.3.3. Load (Nạp dữ liệu vào kho dữ liệu)

Sau khi chuyển đổi, dữ liệu sẽ được nạp vào kho dữ liệu OLAP. Quá trình này có thể thực hiện theo hai phương pháp chính:

- **Nạp toàn bộ (Full Load):** Áp dụng khi xây dựng kho dữ liệu lần đầu tiên.
- **Nạp gia tăng (Incremental Load):** Chỉ nạp các thay đổi mới để cập nhật dữ liệu hiện có, giúp tiết kiệm thời gian và tài nguyên.

2.4. CÔNG CỤ XÂY DỰNG KHO DỮ LIỆU OLAP

Để xây dựng và vận hành kho dữ liệu OLAP hiệu quả, doanh nghiệp cần sử dụng các công cụ phù hợp trong từng giai đoạn. Dưới đây là các công cụ phổ biến trong việc triển khai kho dữ liệu OLAP.

2.4.1. Công cụ ETL (Extract, Transform, Load)

- **Informatica PowerCenter:**
 - Là một công cụ ETL hàng đầu, hỗ trợ tích hợp dữ liệu từ nhiều nguồn khác nhau.
 - Có khả năng xử lý các quy trình phức tạp và đảm bảo tính nhất quán của dữ liệu.
- **Talend Open Studio:**
 - Một công cụ mã nguồn mở phổ biến, dễ dàng sử dụng và hỗ trợ các định dạng dữ liệu đa dạng.
 - Phù hợp với các doanh nghiệp vừa và nhỏ nhờ chi phí thấp và tính linh hoạt cao.

- **Microsoft SQL Server Integration Services (SSIS):**
 - Công cụ tích hợp mạnh mẽ trong hệ sinh thái Microsoft, hỗ trợ xử lý ETL tự động.

2.4.2. Công cụ lưu trữ dữ liệu OLAP

- **Microsoft SQL Server Analysis Services (SSAS):**
 - Hỗ trợ xây dựng các khối OLAP đa chiều, tối ưu hóa cho việc truy vấn.
 - Tích hợp chặt chẽ với các công cụ phân tích như Power BI.
- **Oracle Essbase:**
 - o Một trong những giải pháp OLAP hàng đầu, cung cấp hiệu suất cao cho các phân tích phức tạp.
- **SAP BW/4HANA:**
 - o Nền tảng kho dữ liệu tiên tiến, phù hợp với các doanh nghiệp lớn, hỗ trợ phân tích thời gian thực.

2.4.3. Công cụ trực quan hóa và phân tích dữ liệu

- **Tableau:**
 - Công cụ trực quan hóa mạnh mẽ, hỗ trợ các phân tích tương tác.
 - Phù hợp với người dùng không chuyên nhờ giao diện thân thiện.
- **Power BI:**

- Cung cấp các báo cáo động và trực quan, tích hợp tốt với các hệ thống của Microsoft.
- **QlikView:**
 - Công cụ trực quan hóa và phân tích mạnh mẽ, giúp doanh nghiệp tìm kiếm các thông tin ẩn giấu trong dữ liệu.

CHƯƠNG 3

THIẾT KẾ KHO DỮ LIỆU

3.1. PHÂN TÍCH ĐỀ TÀI

3.1.1. Phân tích yêu cầu

Mục tiêu chính:

Đề tài hướng đến việc thiết kế và triển khai một hệ thống kho dữ liệu (Data Warehouse) nhằm lưu trữ và phân tích toàn diện dữ liệu kinh doanh của cửa hàng bán quần áo. Hệ thống sẽ hỗ trợ ban quản lý ra quyết định chiến lược, tối ưu hóa vận hành và nâng cao hiệu quả kinh doanh.

Yêu cầu cụ thể:

Thu thập dữ liệu:

Tích hợp dữ liệu từ nhiều nguồn khác nhau, bao gồm:

- Bán hàng: Dữ liệu về doanh thu, sản phẩm, thời gian, khu vực bán hàng.
- Khách hàng: Thông tin cá nhân, lịch sử mua hàng, hành vi tiêu dùng.
- Kho hàng: Số lượng hàng tồn, sản phẩm bán chạy, sản phẩm sắp hết hàng.
- Khuyến mãi: Thông tin các chương trình ưu đãi, giảm giá, hiệu quả chiến dịch quảng cáo.
- Địa lý: Thông tin về các địa điểm mua hàng, địa điểm tiêu dùng

Xử lý dữ liệu:

Làm sạch dữ liệu để đảm bảo tính chính xác và đồng nhất.

Chuyển đổi dữ liệu từ các định dạng khác nhau sang một cấu trúc thống nhất.

Áp dụng các kỹ thuật chuẩn hóa nhằm giảm thiểu trùng lặp và tối ưu hóa lưu trữ.

Lưu trữ dữ liệu:

Xây dựng kho dữ liệu tập trung, được thiết kế với cấu trúc hỗ trợ phân tích và báo cáo.

Đảm bảo kho dữ liệu có khả năng mở rộng khi khối lượng dữ liệu tăng lên.

Phân tích và báo cáo:

Xây dựng hệ thống báo cáo linh hoạt, tập trung vào các chỉ số kinh doanh cốt lõi như:

- Doanh thu theo ngày, tháng, năm.
- Sản phẩm bán chạy nhất và kém nhất.
- Phân tích hành vi mua sắm của khách hàng.

Triển khai các công cụ trực quan hóa dữ liệu (dashboards, biểu đồ) để hỗ trợ ra quyết định.

3.1.2. Yêu cầu hệ thống**Chức năng chính:**

Hệ thống kho dữ liệu phải đảm bảo cung cấp đầy đủ các chức năng sau:

Tự động hóa quy trình ETL:

- Trích xuất (Extract) dữ liệu từ các nguồn đầu vào.
- Chuyển đổi (Transform) dữ liệu theo yêu cầu phân tích.
- Tải (Load) dữ liệu vào kho lưu trữ.

Lưu trữ và quản lý dữ liệu:

Hệ thống cần đảm bảo tính toàn vẹn dữ liệu và hỗ trợ truy vấn nhanh.

Thiết kế kho dữ liệu theo mô hình Star Schema hoặc Snowflake Schema để tối ưu hiệu suất truy xuất và phân tích.

Phân tích và báo cáo:

- Cung cấp giao diện trực quan cho người dùng cuối để xem các báo cáo và biểu đồ.
- Hỗ trợ các truy vấn động (dynamic queries) để phân tích dữ liệu theo nhiều tiêu chí khác nhau.

Quản lý bảo mật và phân quyền:

- Xây dựng hệ thống phân quyền người dùng để đảm bảo chỉ những người có thẩm quyền mới được truy cập dữ liệu nhạy cảm.
- Áp dụng mã hóa dữ liệu nhằm ngăn chặn rủi ro bảo mật.

Yêu cầu kỹ thuật:**Phần mềm:**

- Sử dụng hệ quản trị cơ sở dữ liệu mạnh mẽ như MySQL, SQL Server, hoặc MongoDB.
- Triển khai công cụ ETL chuyên dụng như Talend, Apache Nifi hoặc SSIS.
- Công cụ trực quan hóa dữ liệu như Power BI, Tableau hoặc Google Data Studio.

Phần cứng:

- Máy chủ lưu trữ cần đảm bảo hiệu năng cao, với khả năng xử lý hàng triệu bản ghi dữ liệu mỗi ngày.
- Sử dụng hệ thống lưu trữ có khả năng sao lưu và phục hồi dữ liệu (backup & recovery).

Tích hợp giao diện người dùng:

- Giao diện phải thân thiện, dễ sử dụng, với khả năng tương thích trên các thiết bị như máy tính và điện thoại.
- Hiện thị các báo cáo theo thời gian thực (real-time) để hỗ trợ ra quyết định tức thời.

3.2. THIẾT KẾ HỆ THỐNG

3.2.1. Cơ sở dữ liệu

- **Users**(User_id, First_name, Last_name, adr, phone, email);

TÊN TRƯỜNG	KIỂU DỮ LIỆU	MÔ TẢ	KHÓA CHÍNH
<u>User_id</u>	int	Mã user	✓
First_name	Nvarchar(50)	Họ	
Last_name	Nvarchar(50)	Tên	
adr	Nvarchar(200)	Địa chỉ	
phone	varchar(50)	Số điện thoại	
email	varchar(50)	Email	

- **Account**(Account_name, Account_password, role, User_id);

TÊN TRƯỜNG	KIỂU DỮ LIỆU	MÔ TẢ	KHÓA CHÍNH
<u>Account_name</u>	varchar(50)	Tên tài khoản	✓
Account_password	Varchar(255)	Mật khẩu	
role	varchar(20)	Vai trò	
User_id	int	Mã user	

- **Coupon** (Coupon_code, discount_percent, limit);

TÊN TRƯỜNG	KIỂU DỮ LIỆU	MÔ TẢ	KHÓA CHÍNH
<u>Coupon_code</u>	varchar(20)	Mã coupon	✓
discount_percent	int	Phần trăm giảm	
limit	int	Giới hạn dùng	

- **Invoice**(Invoice_id, Account_name, Invoice_date, total_amount, status, Coupon_code, Discount_amount, FinalTotal);

TÊN TRƯỜNG	KIỂU DỮ LIỆU	MÔ TẢ	KHÓA CHÍNH
<u>Invoice_id</u>	varchar(20)	Mã hóa đơn	✓
Account_name	varchar(50)	Tên tài khoản	
Invoice_date	date	Ngày tạo	
total_amount	money	Tổng tiền	
status	Nvarchar(30)	Trạng thái	
Coupon_code	varchar(20)	Mã coupon	
Discount_amount	money	Tiền giảm	
FinalTotal	money	Tổng tiền sau cùng	

- **Categories** (Category_id, category_name);

TÊN TRƯỜNG	KIỂU DỮ LIỆU	MÔ TẢ	KHÓA CHÍNH
<u>Category_id</u>	int	Mã loại	✓
category_name	Nvarchar(50)	Tên loại	

- **Product**(Product_id, product_name, description, price, image_url);

TÊN TRƯỜNG	KIỂU DỮ LIỆU	MÔ TẢ	KHÓA CHÍNH
<u>Product_id</u>	int	Mã sản phẩm	✓
product_name	nvarchar(50)	Tên sản phẩm	
description	Nvarchar(255)	Mô tả	
price	money	Đơn giá	
image_url	varchar (100)	Hình ảnh	

- **Categories_Product**(Product_id, category_id);

TÊN TRƯỜNG	KIỂU DỮ LIỆU	MÔ TẢ	KHÓA CHÍNH
<u>Product_id</u>	int	Mã sản phẩm	✓
category_id	Nvarchar(50)	Mã Loại	

- **Cart** (Cart_id, product_id, account_name, quantity, color, size, total_price);

TÊN TRƯỜNG	KIỂU DỮ LIỆU	MÔ TẢ	KHÓA CHÍNH
<u>Cart_id</u>	int	Mã giỏ hàng	✓
product_id	int	Mã sản phẩm	
account_name	varchar(50)	Tên tài khoản	
quantity	int	Số lượng	
color	nvarchar(50)	Màu	
size	Char(10)	Kích thước	
total_price	money	Tổng giá	

- **Warehouse**(Product_id, color, size, quantity);

TÊN TRƯỜNG	KIỂU DỮ LIỆU	MÔ TẢ	KHÓA CHÍNH
<u>Product_id</u>	int	Mã sản phẩm	✓
color	nvarchar(20)	Màu	
size	Char(10)	Kích thước	
quantity	int	Số lượng	

- **InvoiceDetails** (invoicedetails_id, invoice_id, product_id, quantity, color, size, total_price);

TÊN TRƯỜNG	KIỂU DỮ LIỆU	MÔ TẢ	KHÓA CHÍNH
<u>invoicedetails_id</u>	int	Mã chi tiết hóa đơn	✓
invoice_id	Varchar(20)	Mã hóa đơn	
product_id	int	Mã sản phẩm	
quantity	int	Số lượng	
color	nvarchar(50)	Màu	
size	Char(10)	Kích thước	
total_price	money	Tổng giá	

- **Review** (**review_id**, product_id, account_name, rating, comment, review_date);

TÊN TRƯỜNG	Kiểu dữ liệu	Mô tả	Khóa chính
<u>review_id</u>	int	Mã đánh giá	✓
product_id	int	Mã sản phẩm	
account_name	Varchar(50)	Tên tài khoản	
rating	int	Đánh giá sao	
comment	nvarchar(max)	Bình luận	
review_date	datetime	Ngày đánh giá	

3.2.2. Kho dữ liệu

- **Geography_Dimension** (geography_key, country, state, city, postal_code, territory);

TÊN TRƯỜNG	KIỂU DỮ LIỆU	MÔ TẢ	KHÓA CHÍNH
<u>geography_key</u>	int	Mã địa điểm	✓
country	Nvarchar(50)	Quốc gia	
state	Nvarchar(50)	Tiểu bang	
city	Nvarchar(50)	Thành phố	
postal_code	Nvarchar(20)	Mã bưu chính	
territory	Nvarchar(50)	Lãnh thổ	

- **Customer_Dimension** (customer_key, customer_name, phone, address);

TÊN TRƯỜNG	KIỂU DỮ LIỆU	MÔ TẢ	KHÓA CHÍNH
<u>customer_key</u>	int	Mã khách hàng	✓
customer_name	Nvarchar(100)	Tên khách hàng	
phone	Nvarchar(50)	Số điện thoại	
address	Nvarchar(255)	Địa chỉ	

-

- **Product_Dimension** (product_code, product_name, product_line, msrp);

TÊN TRƯỜNG	KIỂU DỮ LIỆU	MÔ TẢ	KHÓA CHÍNH
<u>product_code</u>	Nvarchar(50)	Mã sản phẩm	✓
product_name	Nvarchar(50)	Tên sản phẩm	
product_line	Nvarchar(255)	Dòng sản phẩm	
msrp	money	Giá bán lẻ đề xuất	

- **Time_Dimension** (full_date, day_of_month, month, quarter, year);

TÊN TRƯỜNG	KIỂU DỮ LIỆU	MÔ TẢ	KHÓA CHÍNH
<u>full_date</u>	Date	Thời gian	✓
day_of_month	Int	Ngày trong tháng	
month	Int	Tháng	
quarter	Int	Quý	
year	Int	Năm	

- **Coupon_Dimension** (coupon_key, coupon_code);

TÊN TRƯỜNG	KIỂU DỮ LIỆU	MÔ TẢ	KHÓA CHÍNH
<u>coupon_key</u>	int	Khóa coupon	✓
coupon_code	Nvarchar(20)	Coupon	

- **Warehouse_Dimension** (warehouse_key, product_id, stock_quantity);

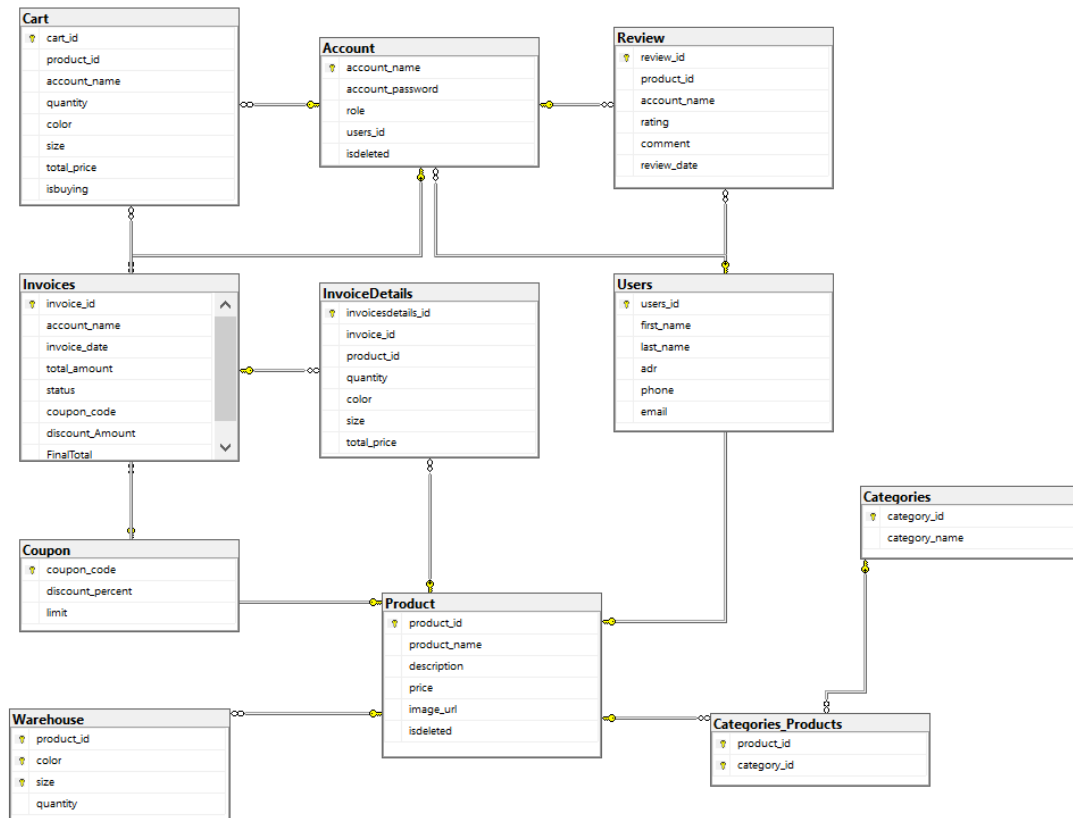
TÊN TRƯỜNG	KIỂU DỮ LIỆU	MÔ TẢ	KHÓA CHÍNH
<u>warehouse_key</u>	int	Mã kho	✓
product_id	Nvarchar(50)	Mã sản phẩm	
stock_quantity	int	Số lượng tồn	

- **Sales_Fact** (product_code, product_name, product_line, msrp);

TÊN TRƯỜNG	KIỂU DỮ LIỆU	MÔ TẢ	KHÓA CHÍNH
<u>order_number</u>	Int	Mã đơn hàng	✓
product_code	Nvarchar(50)	Mã sản phẩm	
customer_key	Int	Mã khách hàng	
time_key	Date	Thời gian	
coupon_key	Int	Mã coupon	
warehouse_key	Int	Mã kho	
geography_key	Int	Mã địa điểm	
quantity_ordered	Int	Số lượng đặt hàng	
total_sales	Money	Tổng doanh thu	
order_status	Nvarchar(50)	Trạng thái đơn hàng	
deal_size	Nvarchar(20)	Kích thước giao dịch	

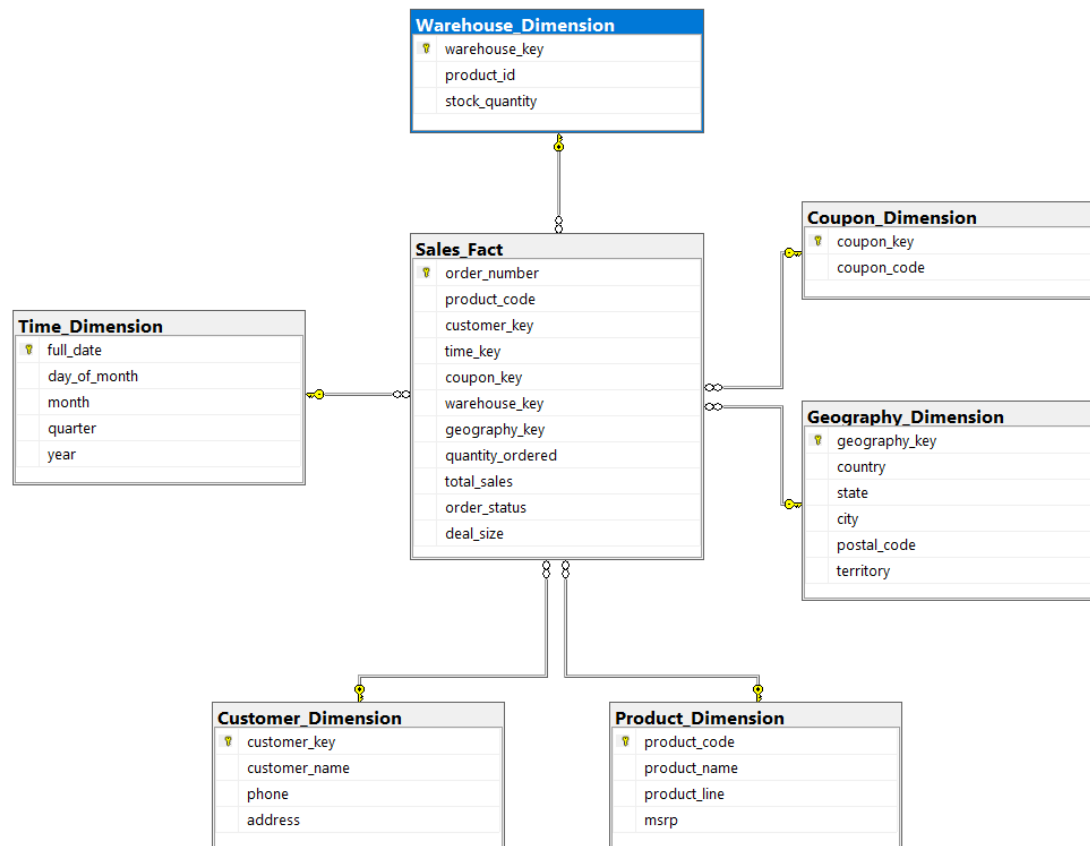
3.2.3. Lược đồ quan hệ

Cơ sở dữ liệu: Từ dữ liệu từ SQL.



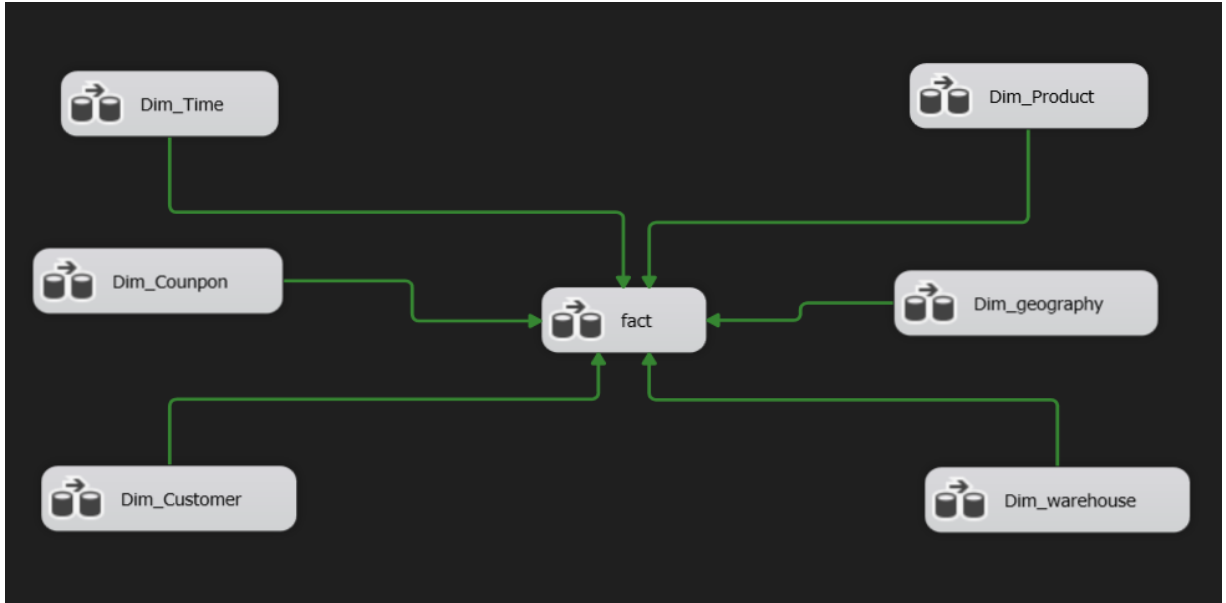
Hình 3.1: Database Diagram

Kho dữ liệu: Phân tích được kho dữ liệu DDS mô hình chòm sao với các bảng Sales_Fact, Warehouse_Dimension, Coupon_Dimension, Geography_Dimension, Product_Dimension, Customer_Dimension, Time_Dimension.



Hình 3.2: Kho dữ liệu mô hình chòm sao

3.2.4. SSIS



Hình 3.3: SSIS

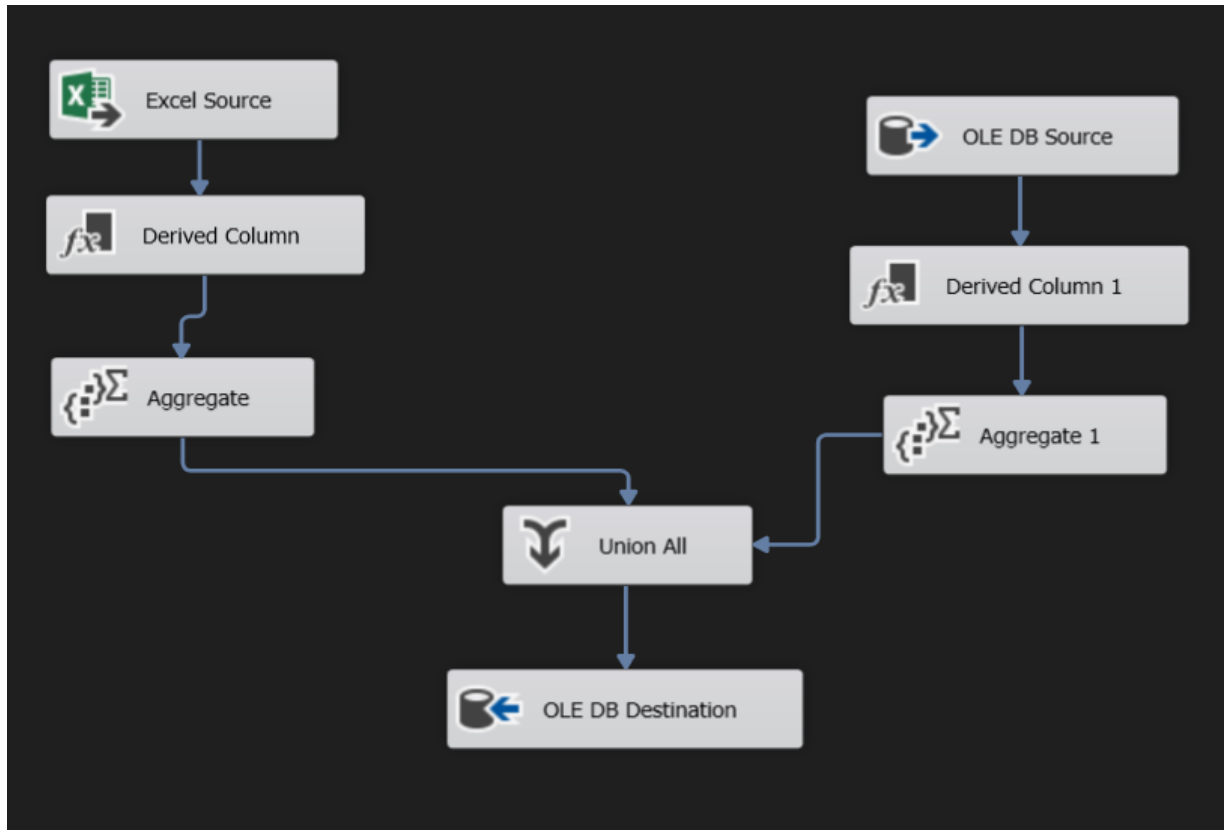
Từ dữ liệu từ SQL trên và file Excel:

PRICE	ORDERLINE	SALES	ORDERDATE	STATUS	PRODUCTLINE	MSRP	PRODUCTCODE	CUSTOMERNAME	PHONE	ADDRESSLINE	CITY	STATE	POSTALCODE	COUNTRY	TERRITORY	DEALSIZE	STOCK	COUPONCODE	
95.7	2	2871	2/24/2003 0:00	Shipped	Harry-Dawson Leather Jackets	95.510.1675		Land of Toys Inc.	2125557118	887 Long Airport Avenue	NYC	NY	10022 USA	NA	S			5 CP82724	
51.35	1	2785	5/7/2003 0:00	Shipped	Harry-Dawson Leather jackets	95.510.1675		Reims Collectables	26 47 1555	59 rue de l'abbaye	Reims		51100 France	EMEA	S			76 CP82637	
94.74	2	3684	34 7/1/2003 0:00	Shipped	Harry-Dawson Leather jackets	95.510.1675		Lyon Souvenirs	+33 1 46 62 7555	27 rue du Colonel Pierre Asia	Paris		75008 France	EMEA	M			42 CP84190	
83.26	6	3746	7/25/2003 0:00	Shipped	Harry-Dawson Leather Jackets	95.510.1675		Toyko Souvenirs	6260707250	7834 Holside Dr	Pasadena	CA	90003 USA	NA	M			49 CP84036	
100	14	5205	27 10/15/2003 0:00	Shipped	Harry-Dawson Leather jackets	95.510.1675		Corporate Gift Ideas Co.	6050501386	1724 Strong St.	San Francisco	CA	94104 USA	NA	M			21 CP22596	
96.66	1	3479	76 10/28/2003 0:00	Shipped	Harry-Dawson Leather jackets	95.510.1675		Technics Stores Inc.	605050989	9408 Futh Circle	Burlington	CA	94217 USA	NA	M			87 CP81887	
95.13	9	2487	77 11/11/2003 0:00	Shipped	Harry-Dawson Leather jackets	95.510.1675		Dandelion Design Imports	20 16 1005	154, chaussee de Tournai	Lille		59000 France	EMEA	S			78 CP84855	
100	1	5512	32 11/18/2003 0:00	Shipped	Harry-Dawson Leather Jackets	95.510.1675		Henku Gifts	+47 2287 3215	Drammen 121, PR 744 Sentrum	Bergen		Norway	EMEA	M			53 CP82915	
98.37	2	2168	54 12/12/2003 0:00	Shipped	Harry-Dawson Leather Jackets	95.510.1675		Mini Vreem Co.	6050503587	5557 North Pendse Street	San Francisco	CA	94134 USA	NA	S			24 CP23257	
14	1	4708	44 11/5/2003 0:00	Shipped	Harry-Dawson Leather jackets	95.510.1675		Auto Canal-Petit	(1) 47 55 6555	25, rue Lauriston	Paris		75016 France	EMEA	M			82 CP80655	
100	1	3686	68 2/20/2004 0:00	Shipped	Harry-Dawson Leather jackets	95.510.1675		Australian Collectors, Co.	03 9523 4555	636 St Kilda Road	Melbourne	Victoria	3004 Australia	APAC	M			74 CP87468	
100	7	2333	12 4/5/2004 1:00	Shipped	Harry-Dawson Leather jackets	95.510.1675		Vitacore Inc.	2125555300	3575 Kingston Rd.	NYC	NY	10022 USA	NA	S			87 CP80831	
100	2	3188	64 5/18/2004 0:00	Shipped	Harry-Dawson Leather jackets	95.510.1675		Telen Collectables Inc.	2010509000	7476 Moss Rd.	Newark	NJ	04019 USA	NA	M			82 CP82743	
100	2	3678	76 6/25/2004 0:00	Shipped	Harry-Dawson Leather jackets	95.510.1675		Gill Dept Inc.	2050502700	22933 South Bay Ln.	Bridgeville	CT	07162 USA	NA	M			18 CP81989	
92.83	1	4177	35 7/23/2004 0:00	Shipped	Harry-Dawson Leather Jackets	95.510.1675		La Rochelle Gifts	40 67 8555	67, rue des Cinquante Otages	Nantes		44000 France	EMEA	M			44 CP81884	
100	6	4098	68 6/27/2004 0:00	Shipped	Harry-Dawson Leather jackets	95.510.1675		Martin's Replicas Co.	6171059555	59233 Sprinaker Dr	Cambridge	MA	01147 USA	NA	M			35 CP81751	
100	9	2587	38 9/30/2004 0:00	Shipped	Harry-Dawson Leather jackets	95.510.1675		Top of Finland, Co.	90 224 8000	Helsinki 45	Helsinki		00100 Finland	EMEA	S			45 CP84849	
100	5	4384	38 10/15/2004 0:00	Shipped	Harry-Dawson Leather jackets	95.510.1675		Baane Mini Imports	07-68 9555	Erving Skates gate 76	Stavern		4110 Norway	EMEA	M			99 CP81929	
94.14	1	4358	94 11/2/2004 0:00	Shipped	Harry-Dawson Leather jackets	95.510.1675		Daniel's Classics Inc.	2105011655	17566 Pungston St.	Altemus	PA	71027 USA	NA	M			85 CP87445	
100	1	4386	14 11/15/2004 0:00	Shipped	Harry-Dawson Leather Jackets	95.510.1675		Land of Toys Inc.	2125557118	887 Long Airport Avenue	NYC	NY	10022 USA	NA	M			46 CP80863	
100	9	7777	18 11/24/2004 0:00	Shipped	Harry-Dawson Leather jackets	95.510.1675		Midway Collectables	60502 9555	Corning 14	Saltford		9200 Austria	EMEA	L			11 CP87936	
72.55	13	1481	12/17/2004 0:00	Shipped	Harry-Dawson Leather jackets	95.510.1675		Southern And Tingo Co.	+61 2 8855 8855	Monitor Money Building, 515 Pacific Hwy	Chatswood	NSW	2057 Australia	APAC	S			34 CP80670	
34.91	12	731	11 2/20/2005 0:00	Shipped	Harry-Dawson Leather jackets	95.510.1675		La Rochelle Gifts	40 67 8555	67, rue des Cinquante Otages	Nantes		44000 France	EMEA	S			11 CP73293	
76.36	4	3207	12 3/3/2005 1:00	Shipped	Harry-Dawson Leather jackets	95.510.1675		Funclibooks.com	9880502055	1783 First Street	New Bedford	MA	01903 USA	NA	M			52 CP80516	
100	7	2434	56 4/5/2005 0:00	Shipped	Harry-Dawson Leather jackets	95.510.1675		UK Collectables, Ltd	(171) 555-2282	Berkley Gardens 12, Brewery	Liverpool		W01 6L	UK	EMEA	S			64 CP83561
100	2	7548	68 5/13/2005 0:00	Shipped	Harry-Dawson Leather jackets	95.510.1675		Euro Shopping Channel	(91) 855 54 44	C/ Morazan 96	Madrid		28034 Spain	EMEA	L			13 CP82960	
100	11	5404	62 12/28/2005 0:00	Shipped	Harry-Dawson Leather jackets	214.910.1949		Baane Mini Imports	07-68 9555	Erving Skates gate 76	Stavern		4110 Norway	EMEA	M			48 CP80983	
100	1	7209	11 3/24/2006 0:00	Shipped	Harry-Dawson Leather jackets	214.910.1949		Voxto Model Replicas, Co	0621-12 3555	Bergsmyr 9	Lulea		S-955 22	Sweden	EMEA	L			86 CP84928
100	11	7209	66 5/28/2006 0:00	Shipped	Harry-Dawson Leather jackets	214.910.1949		Comda Auto Replicas, Ltd	(91) 855 52 82	C/ Ansel 27	Madrid		28023 Spain	EMEA	L			53 CP82618	
100	11	7314	1 7/24/2006 0:00	Shipped	Harry-Dawson Leather jackets	214.910.1949		Technics Stores Inc.	605050989	9408 Futh Circle	Burlington	CA	94217 USA	NA	L			96 CP80513	
100	9	10965	5/18/2006 0:00	Shipped	Harry-Dawson Leather jackets	214.910.1949		Dragon Souvenirs, Ltd	+65 207 7555	Brown Sub, Brown Apt 3/6 Teckway	Singapore		79603 Singapore	Japan	L			25 CP86177	
100	1	4868	24 10/29/2006 0:00	Shipped	Harry-Dawson Leather jackets	214.910.1949		Casac, Legends Inc.	2125554893	5965 Pungston St.	NYC	NY	10022 USA	NA	M			7 CP86996	
100	4	8014	48 11/6/2006 0:00	Shipped	Harry-Dawson Leather jackets	214.910.1949		Australian Gift Centre, Co.	61-73-844-8555	31 Duncan St. West End	South Brisbane	Queensland	4101 Australia	APAC	L			2 CP87296	
100	8	5372	27 11/15/2006 0:00	Shipped	Harry-Dawson Leather jackets	214.910.1949		Casac, Gift Ideas, Inc.	2125554893	782 First Street	Phoenicia	PA	71270 USA	NA	M			81 CP80524	
100	11	7209	36 11/25/2006 0:00	Shipped	Harry-Dawson Leather jackets	214.910.1949		Sawley & Henriot, Co.	70 32 5555	2, rue du Commerce	Lyon		69004 France	EMEA	L			41 CP82430	
100	6	9064	88 12/3/2006 0:00	Shipped	Harry-Dawson Leather jackets	214.910.1949		Canadian Gift Exchange Network	(848) 655-3382	1800 Oak St.	Vancouver	BC	V3F 2M1	Canada	NA	L			27 CP86284
100	3	6073	5 12/29/2006 0:00	Shipped	Harry-Dawson Leather jackets	214.910.1949		West Coast Collectables Co.	2125553232	3675 Futh Circle	Burlington	CA	94019 USA	NA	M			73 CP82240	
100	2	6482	23 3/12/2006 0:00	Shipped	Harry-Dawson Leather jackets	214.910.1949		Cambridge Collectables Co.	6171055565	4558 Eastern Av.	Cambridge	MA	01247 USA	NA	M			14 CP89622	
100	9	6120	34 5/4/2006 0:00	Shipped	Harry-Dawson Leather jackets	214.910.1949		Super Scale Inc.	3030509645	567 North Pendse Street	New Haven	CT	07823 USA	NA	M			18 CP84889	
100	6	7680	84 6/15/2006 0:00	Shipped	Harry-Dawson Leather jackets	214.910.1949		Toyko Collectables, Ltd	+61 3 2884 0055	2-8 Rippings	Melboure	Victoria	3004 Australia	APAC	M			84 CP81893	
100	9	4905	36 7/18/2006 0:00	Shipped	Harry-Dawson Leather jackets	214.910.1949		Southern And Tingo Co.	+61 2 8855 8855	Monitor Money Building, 515 Pacific Hwy	Chatswood	NSW	2057 Australia	APAC	M			4 CP84553	
100	2	8014	42 8/17/2006 0:00	Shipped	Harry-Dawson Leather jackets	214.910.1949		Amica Models & Co.	011-4895555	Via Monte Bianco 34	Torino		10100 Italy	EMEA	L			89 CP85480	
100	11	7156	18 9/8/2006 0:00	Shipped	Harry-Dawson Leather jackets	214.910.1949		Scandinavian Gift Ideas	0620-34 4305	Thengsten 34	Bonn		5344 67	Sweden	EMEA	L			87 CP86171
100	6	10172	1 10/11/2006 0:00	Shipped	Harry-Dawson Leather jackets	214.910.1949		Auto Assoc. & Cie	30 59 8555	67, avenue de l'Europe	Venables		78600 France	EMEA	L			96 CP84863	
100	3	11623	1 10/21/2006 0:00	Shipped	Harry-Dawson Leather jackets	214.910.1949		Online Decast Creations Ltd	6050509847	2304 Long Airport Avenue	Nashua	NH	03065 USA	NA	M			4 CP81996	
100	3	3093	11/18/2006 0:00	Shipped	Harry-Dawson Leather jackets	214.910.1949		Mini Vreem Co.	6050503587	5557 North Pendse Street	San Francisco	CA	94134 USA	NA	M			97 CP83832	
100	1	3647	11/28/2006 0:00	Shipped	Harry-Dawson Leather jackets	214.910.1949		Australian Collectors, Co.	03 9523 4555	636 St Kilda Road	Melbourne	Victoria	3004 Australia	APAC	M			100 CP84230	
100	10	5691	84 12/19/2006 0:00	Shipped	Harry-Dawson Leather jackets	214.910.1949		Mini Gifts Distributors Ltd	4150501400	9677 Strong St.	San Rafael	CA	97562 USA	NA	M			96 CP80831	
100	2	4514	92 12/28/2006 0:00	Shipped	Harry-Dawson Leather jackets	214.910.1949		Collectables For Less Inc.	6171055565	7320 Douglas Av.	Brookline	MA	01903 USA	NA	M			82 CP81995	
100	3	6254	8 2/17/2006 0:00	Shipped	Harry-Dawson Leather jackets	214.910.1949		Corporate Gift Ideas Co.	6050501386	1724 Strong St.	San Francisco	CA	94104 USA	NA	L			8 CP80734	
100	4	2416	56 3/29/2006 0:00	Shipped	Harry-Dawson Leather jackets	214.910.1949		Amara Decorations, Ltd	53 9886 8555	201 Miller Street	North Sydney	NSW	2060 Australia	APAC	S			17 CP89524	
100	6	4142	23 5/12/2006 0:00	Shipped	Harry-Dawson Leather jackets	214.910.1949		Quebec Home Shopping Network	6141 555-8554	43 rue St. Laurent	Montreal	Quebec	H3J 1C3	Canada	NA	M			5 CP89611

Hình 3.4: File Excel

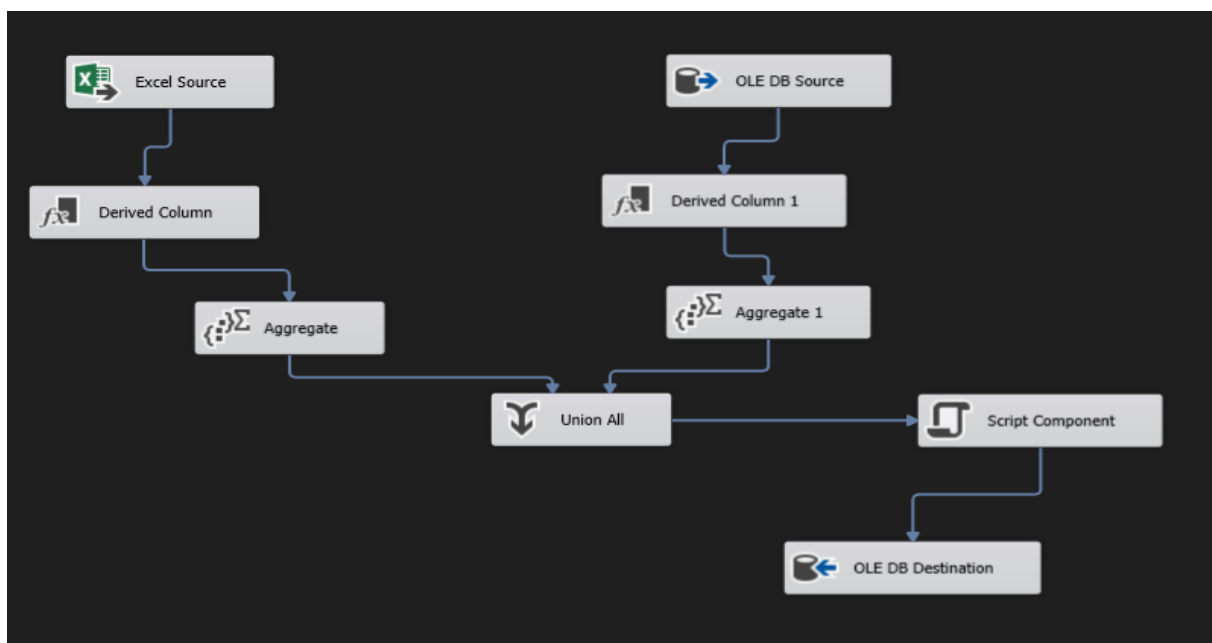
Phân tích được các Dimension như sau:

Dim_Time: Dùng để phân tích dữ liệu theo thời gian.



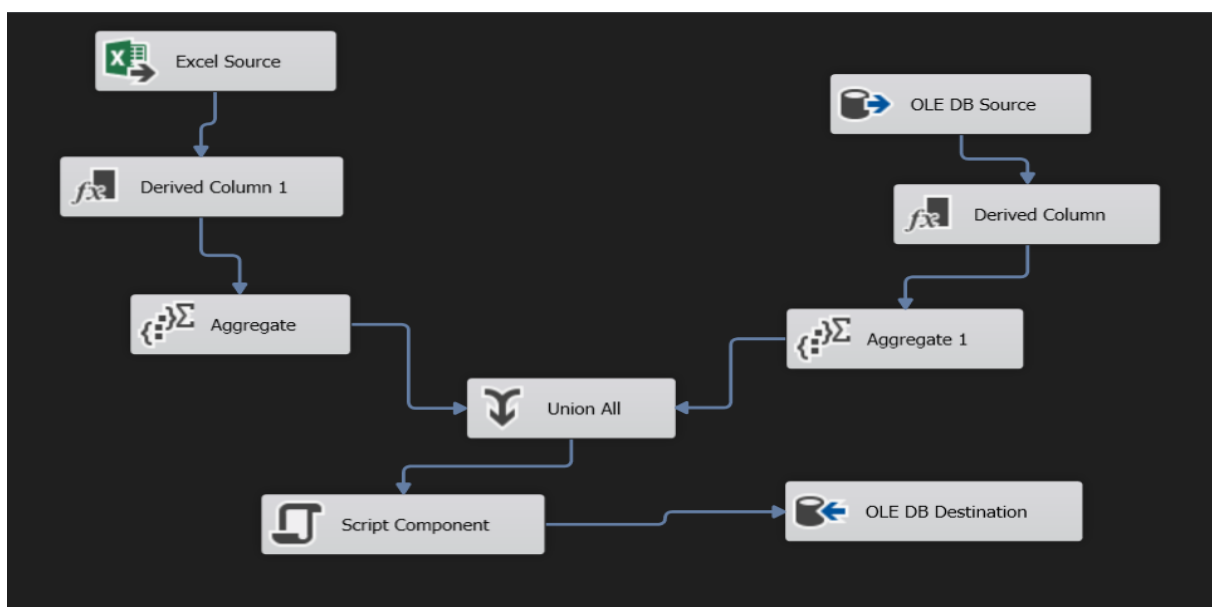
Hình 3.5: SSIS Dim_Time

Dim_Coupon: Dùng để phân tích dữ liệu theo coupon



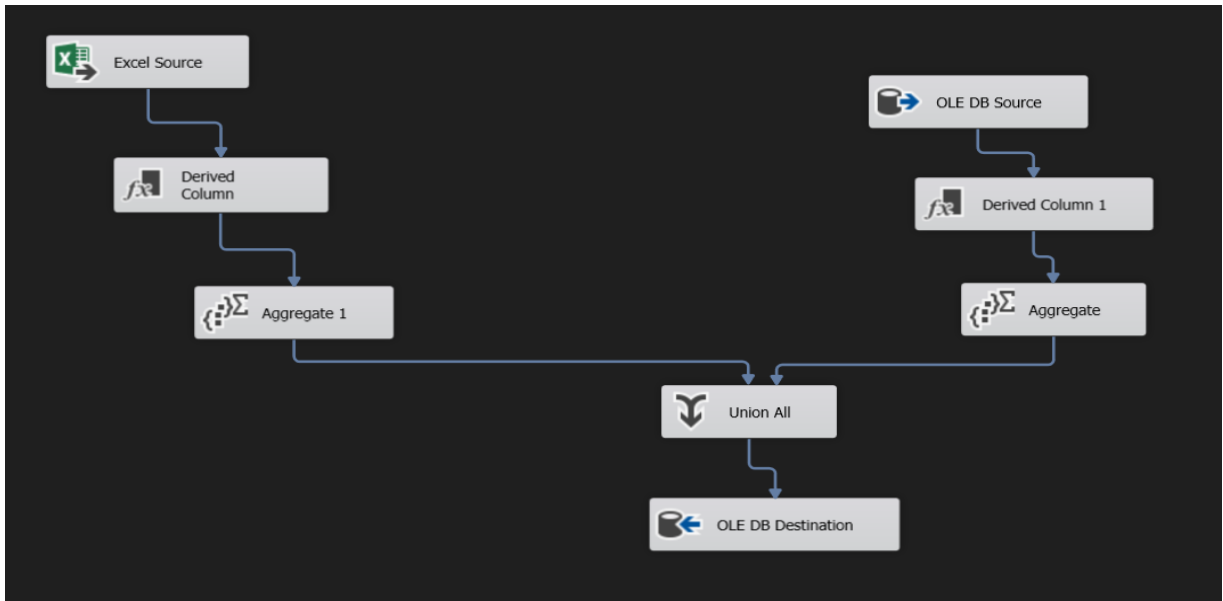
Hình 3.6: SSIS Dim_Coupon

Dim_Customer: Dùng để phân tích dữ liệu theo khách hàng.



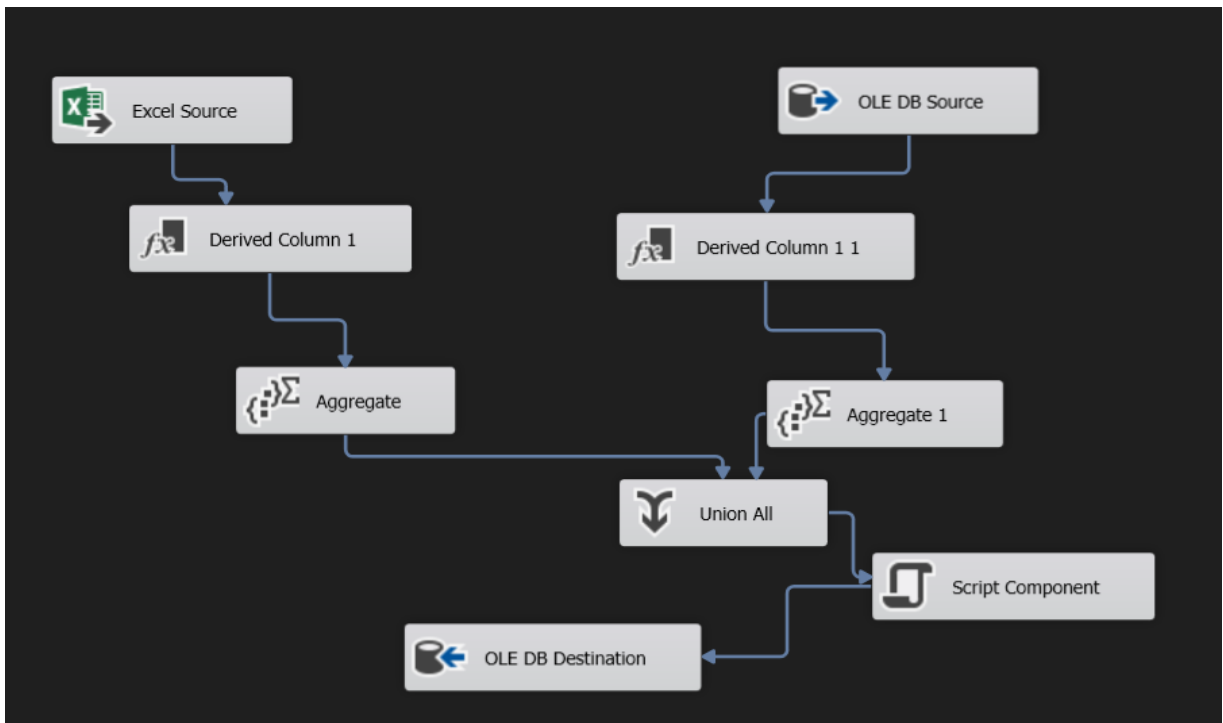
Hình 3.7: SSIS Dim_Customer

Dim_Product: Dùng để phân tích dữ liệu theo sản phẩm



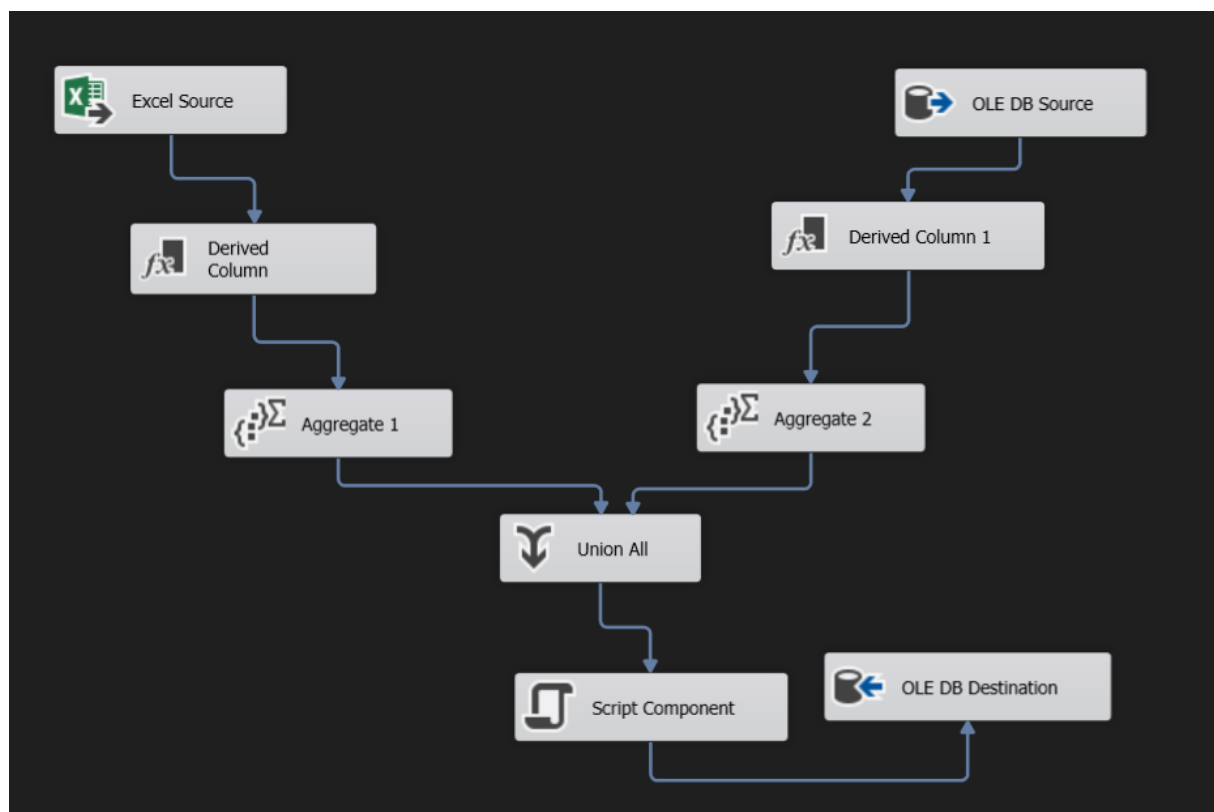
Hình 3.8: SSIS Dim_Product

Dim_Geography: Dùng để phân tích dữ liệu theo địa lý



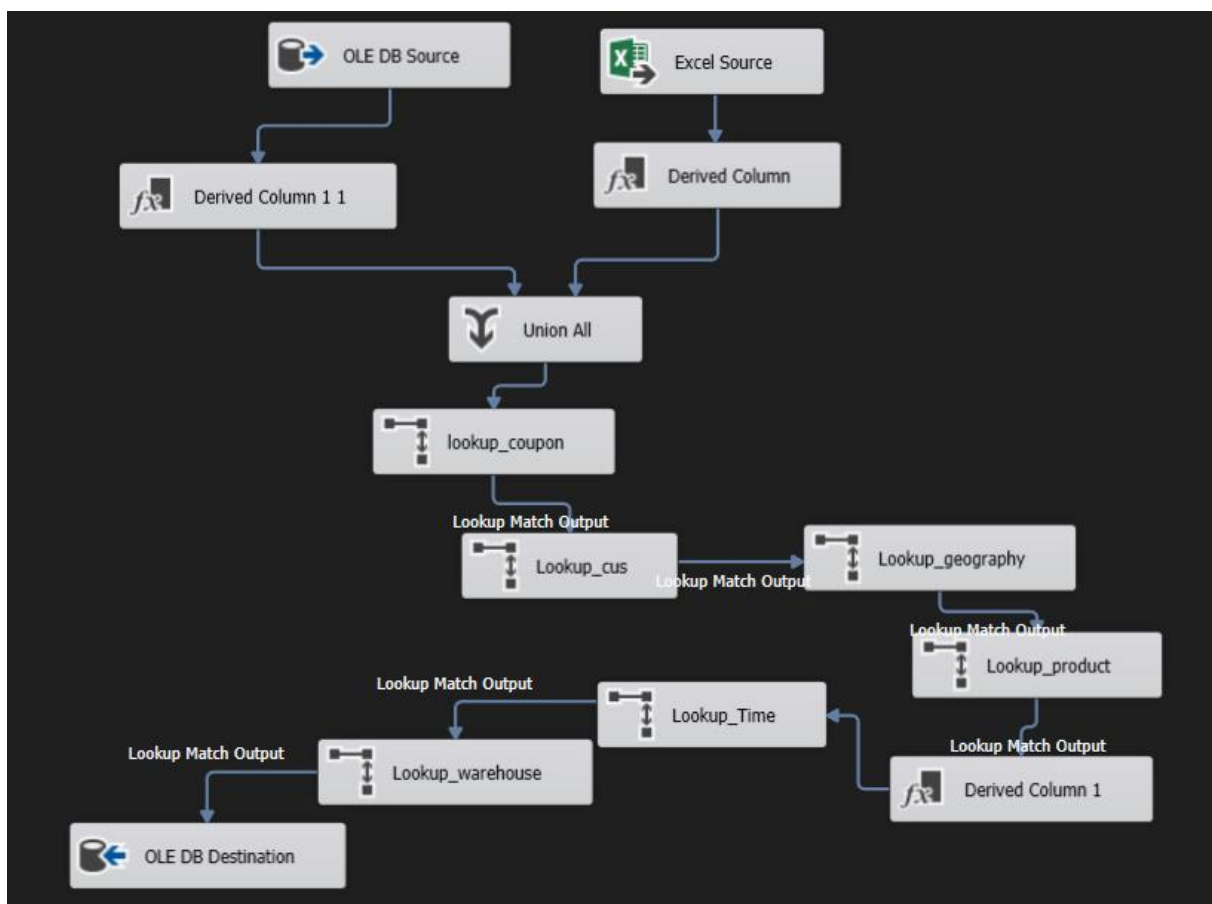
Hình 3.9: SSIS Dim_Geography

Dim_Warehouse: Dùng để phân tích dữ liệu theo kho.



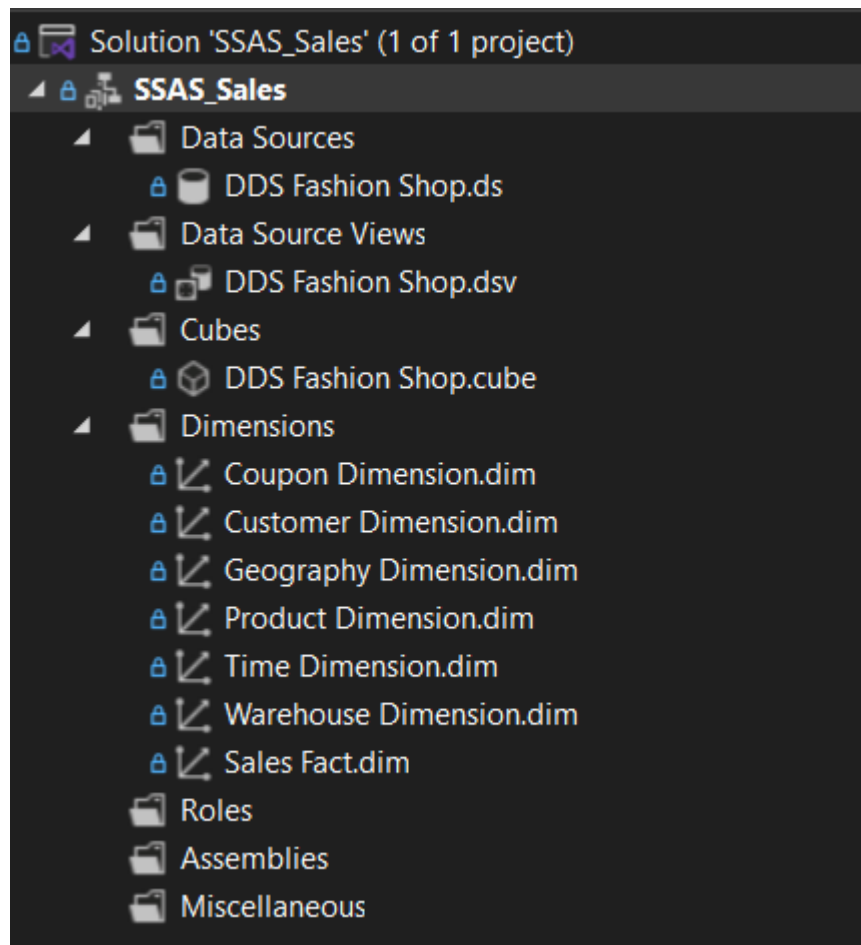
Hình 3.10: SSIS Dim_Warehouse

Fact: Để phân tích dữ liệu bán hàng, tổng hợp và đo lường hiệu suất



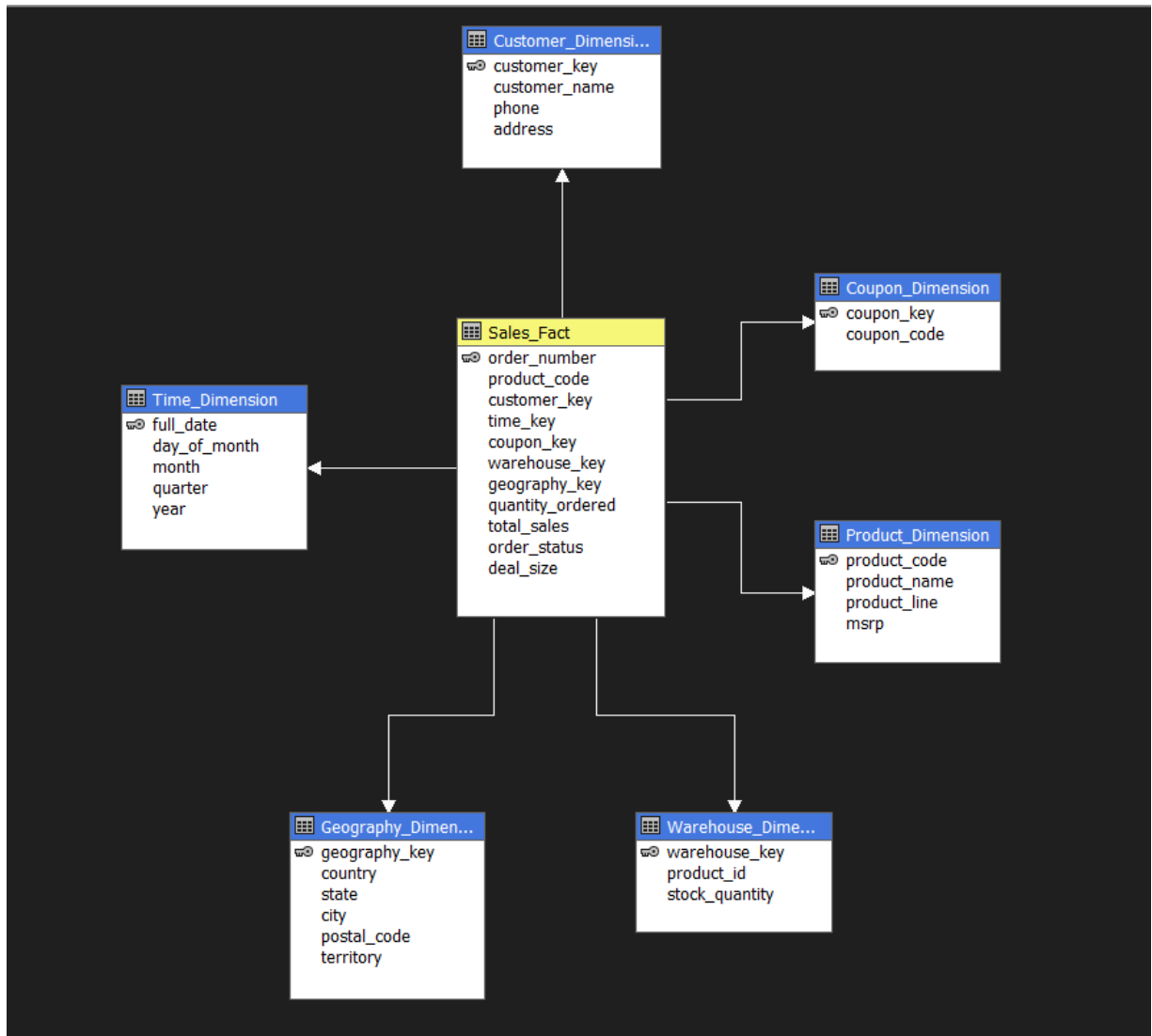
Hình 3.11: SSIS Fact

3.2.4. SSAS



Hình 3.12: SSAS

Cube:



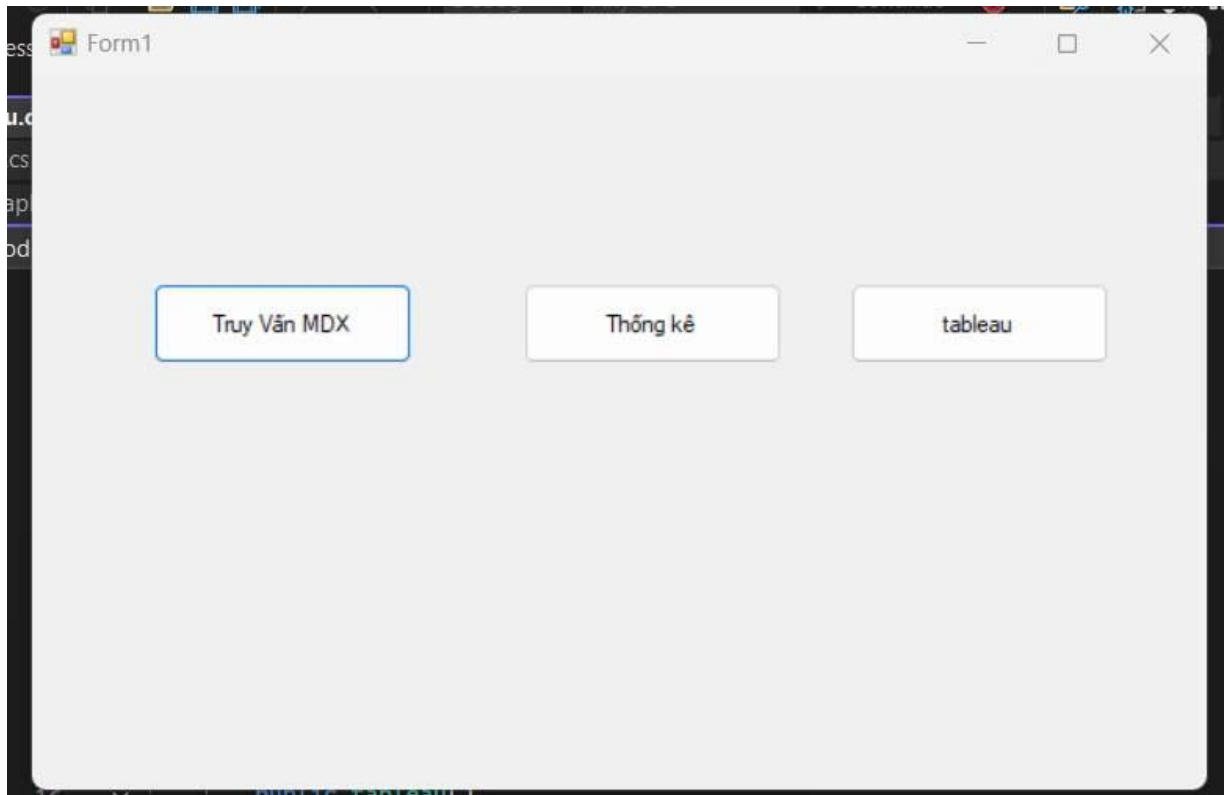
Hình 3.13: SSAS Cube

CHƯƠNG 4

CÀI ĐẶT ỨNG DỤNG

4.1. GIAO DIỆN VÀ CHỨC NĂNG CỦA ỨNG DỤNG

Giao diện trang chủ: Với 3 chức năng Truy vấn MDX, Thống kê và Tableau



Hình 4.1: Giao diện trang chủ

- Nút Truy vấn MDX sẽ hiện ra Giao diện truy vấn MDX để truy vấn.
- Nút Thống kê sẽ hiện Giao diện thống kê.
- Nút Tableau hiển thị trang phân tích Tableau.

Giao diện truy vấn MDX:

The screenshot shows the 'TruyVanMDX' application window. It features a table of product sales data and an MDX query editor on the right.

	[Product Dimension].[Product Line].[Product Line].[MEMBER_C...	[Time Dimension].[Year].[Sales]	[Time Dimension].[Year]& Sales]	[...
►		40457128.85	3516979.54	4
	Áo cưới nam pho...	1820000		
	Áo dài nam truyề...			
	Áo đồng phục ch...	775000		
	Áo hoodie nữ ấm ...			
	Áo khoác bombe...			
	Áo khoác da na...	2032500		
	Áo khoác nhẹ th...	322500		
	Áo len ấm áp, ph...	2032500		
	Áo len nam kiểu ...			
	Áo Oxford lịch lã...	1649000		

MDX Query:

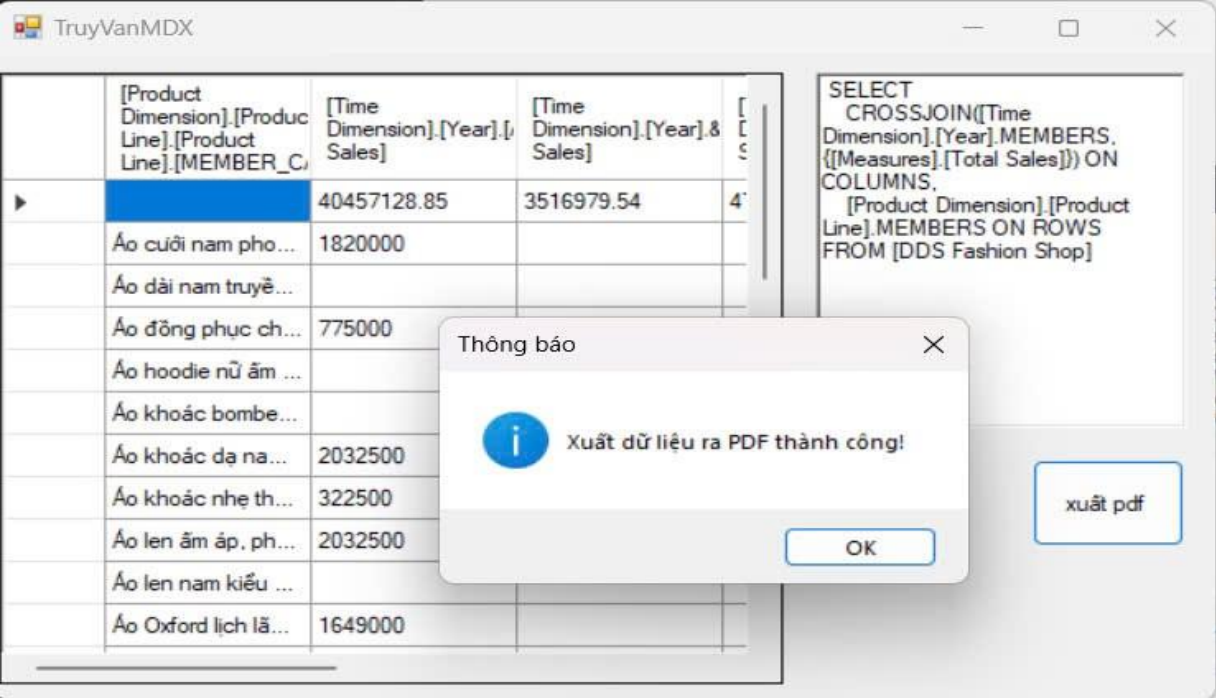
```
SELECT  
  CROSSJOIN([Time  
    Dimension].[Year].MEMBERS,  
    {[Measures].[Total Sales]}) ON  
  COLUMNS,  
  [Product Dimension].[Product  
    Line].MEMBERS ON ROWS  
FROM [DDS Fashion Shop]
```

Buttons: thực hiện, xuất pdf

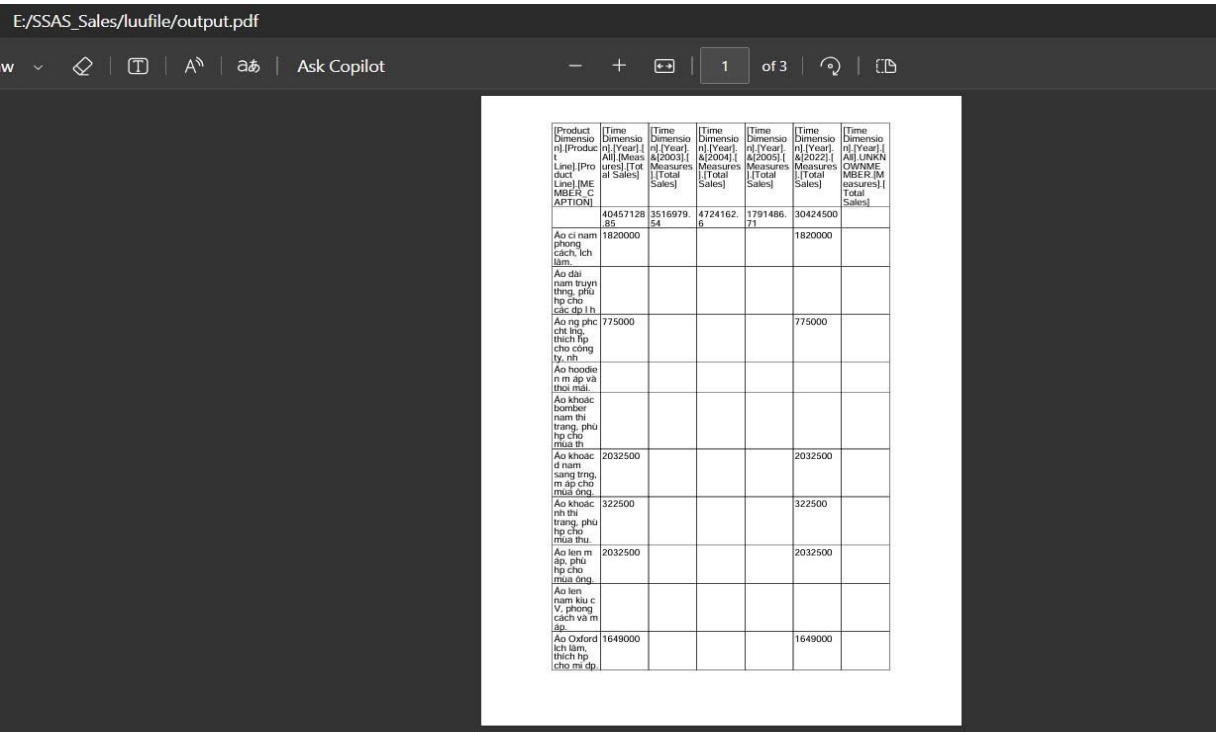
Hình 4.2: Giao diện truy vấn MDX

- Nút thực hiện sẽ thực hiện câu truy vấn MDX ở textbox trên.
- Nút Xuất PDF sẽ xuất ra file PDF dựa trên thống kê từ truy vấn MDX.

File PDF được thống kê và xuất ra:

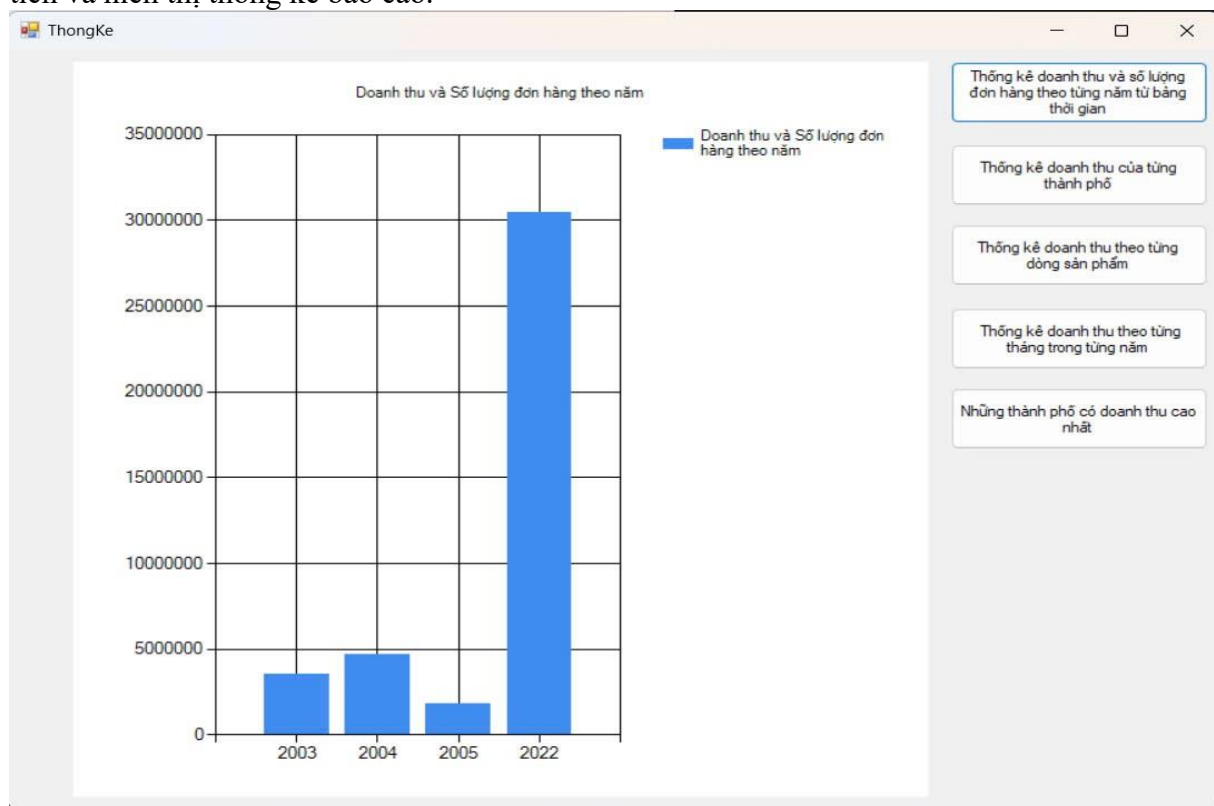


Hình 4.3: Giao diện thực hiện xuất PDF

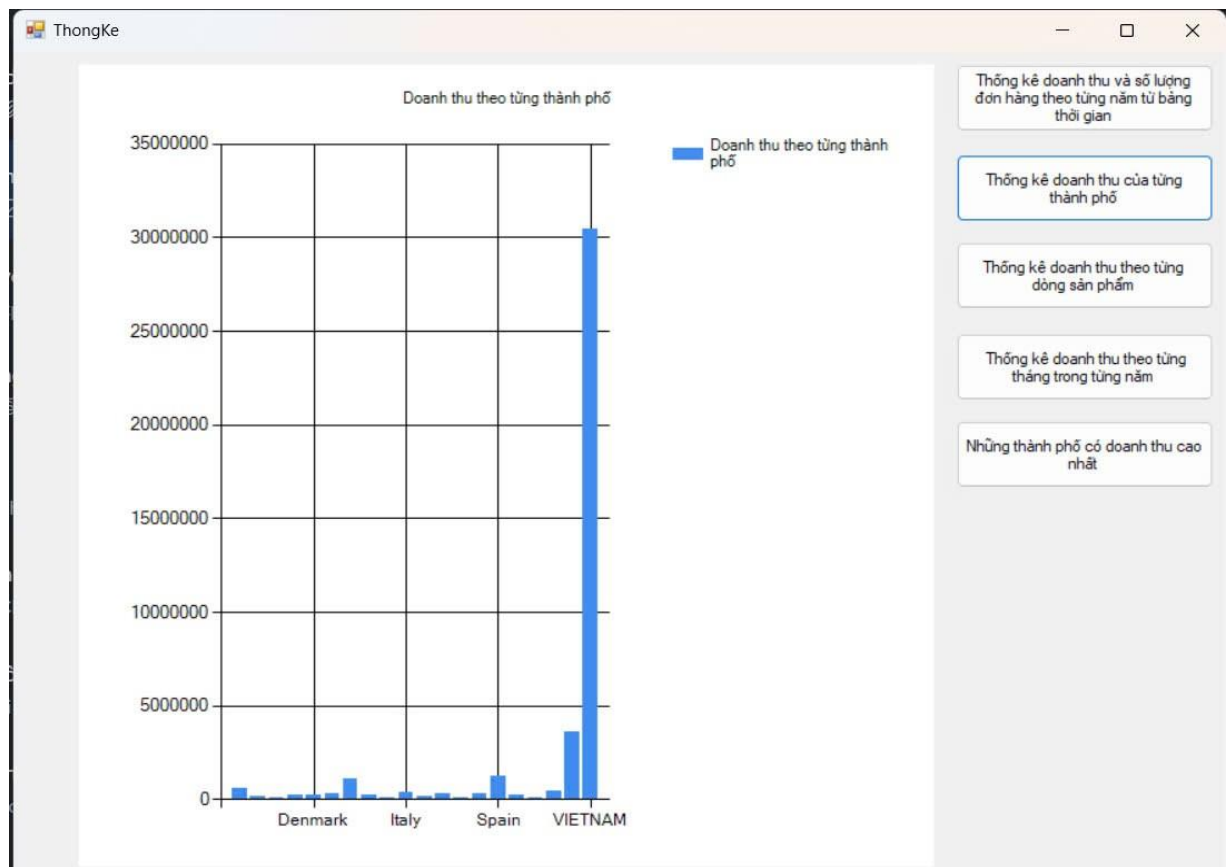


Hình 4.4: File PDF xuất ra

Chức năng thống kê với các lựa chọn thống kê: Dựa trên từng lựa chọn, hệ thống sẽ phân tích và hiển thị thống kê báo cáo.



Hình 4.5: Thống kê 1



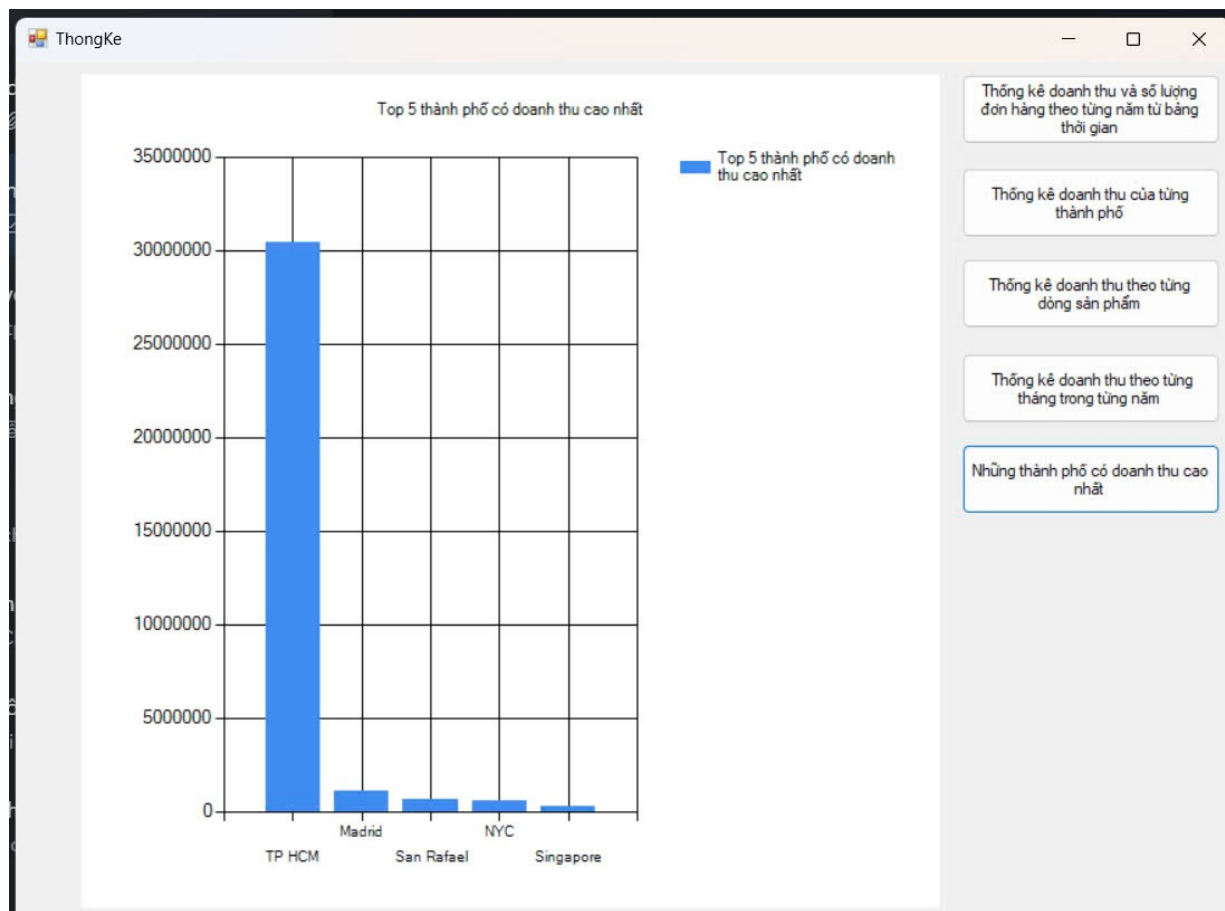
Hình 4.6: Thống kê 2



Hình 4.7: Thống kê 3

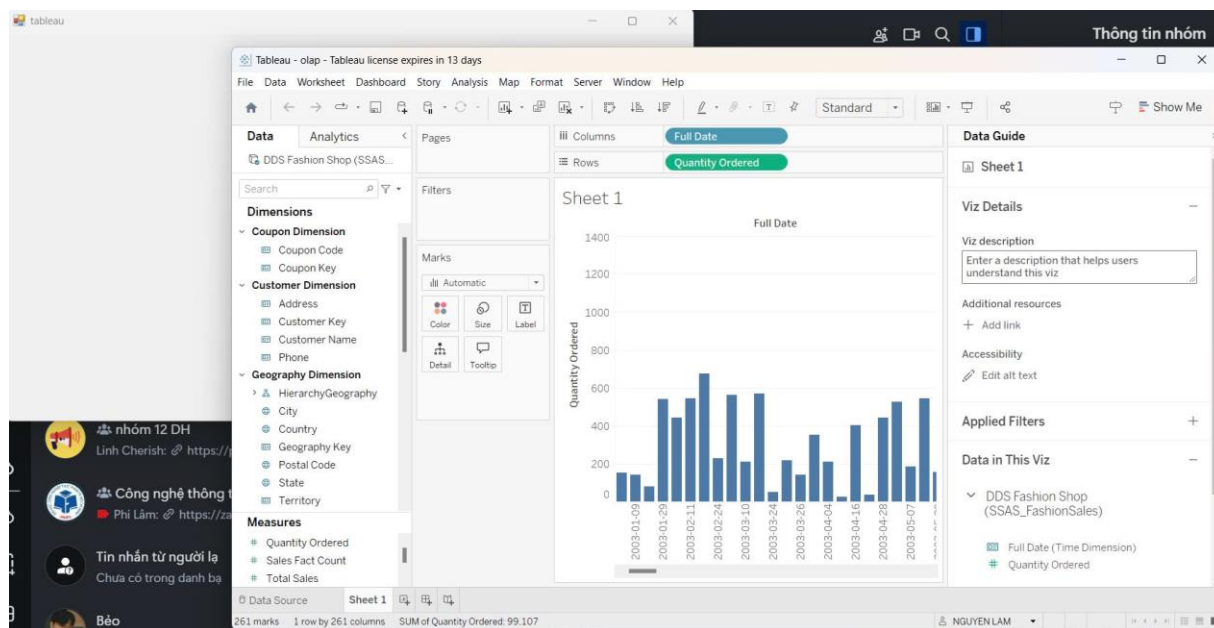


Hình 4.8: Thống kê 4



Hình 4.9: Thống kê 5

Giao diện tableau: Dùng để phân tích dữ liệu bằng Tableau



Hình 4.10: Giao diện Tableau

4.2. TRUY VẤN MDX

- Truy vấn tổng doanh thu theo năm và lọc theo doanh thu > 0.

```
-- 1. Tổng doanh thu theo năm và lọc theo doanh thu > 0
SELECT
    FILTER([Time Dimension].[Year].MEMBERS, |
        [Time Dimension].[Year].CURRENTMEMBER.Name <> "All" AND [Measures].[Total Sales] > 0) ON ROWS,
    [Measures].[Total Sales] ON COLUMNS
FROM [DDS Fashion Shop]
```

Hình 4.11: Truy vấn tổng doanh thu theo năm và lọc theo doanh thu > 0

- Truy vấn tổng doanh thu theo thành phố và lọc theo doanh thu > 1000.

```
-- 2. Tổng doanh thu theo thành phố và lọc theo doanh thu > 1000
SELECT
    FILTER([Geography Dimension].[City].MEMBERS,
        [Geography Dimension].[City].CURRENTMEMBER.Name <> "All" AND [Measures].[Total Sales] > 1000) ON ROWS,
    [Measures].[Total Sales] ON COLUMNS
FROM [DDS Fashion Shop]
```

Hình 4.12: Truy vấn tổng doanh thu theo thành phố và lọc theo doanh thu > 1000

- Truy vấn tổng doanh thu theo sản phẩm và khách hàng, để tính doanh thu trung bình.

```
-- 3. Tổng doanh thu theo sản phẩm và khách hàng, để tính doanh thu trung bình
WITH
    MEMBER [Measures].[Avg Sales] AS
        AVG([Product Dimension].[Product Name].MEMBERS, [Measures].[Total Sales])
SELECT
    FILTER([Product Dimension].[Product Name].MEMBERS,
        [Product Dimension].[Product Name].CURRENTMEMBER.Name <> "All") ON ROWS,
    { [Measures].[Total Sales], [Measures].[Avg Sales] } ON COLUMNS
FROM [DDS Fashion Shop]
```

Hình 4.13: Truy vấn tổng doanh thu theo sản phẩm và khách hàng, để tính doanh thu trung bình

- Truy vấn lấy top 5 sản phẩm theo doanh thu trong năm hiện tại và sắp xếp theo thứ tự giảm dần.

```
-- 4. Lấy Top 5 sản phẩm theo doanh thu trong năm hiện tại và sắp xếp theo thứ tự giảm dần
SELECT
    TOPCOUNT(
        FILTER([Product Dimension].[Product Name].MEMBERS,
            [Product Dimension].[Product Name].CURRENTMEMBER.Name <> "All"),
        5,
        [Measures].[Total Sales]
    ) ON ROWS,
    [Measures].[Total Sales] ON COLUMNS
FROM [DDS Fashion Shop]
WHERE
    ([Time Dimension].[Year].[2022])
```

Hình 4.14: Truy vấn lấy top 5 sản phẩm theo doanh thu trong năm hiện tại và sắp xếp theo thứ tự giảm dần

- Truy vấn tổng doanh thu theo quý và sắp xếp theo thứ tự giảm dần.

```
-- 5. Tổng doanh thu theo quý và sắp xếp theo thứ tự giảm dần
SELECT
    ORDER(
        FILTER([Time Dimension].[Quarter].MEMBERS,
            [Time Dimension].[Quarter].CURRENTMEMBER.Name <> "All" AND
            [Measures].[Total Sales] > 0 AND
            [Time Dimension].[Quarter].CURRENTMEMBER.Name <> "Unknown"),
        [Measures].[Total Sales],
        BDESC
    ) ON ROWS,
    [Measures].[Total Sales] ON COLUMNS
FROM [DDS Fashion Shop]
```

Hình 4.15: Truy vấn tổng doanh thu theo quý và sắp xếp theo thứ tự giảm dần

-

- Truy vấn doanh thu trung bình theo dòng sản phẩm và năm.

```
--6. Doanh thu trung bình theo dòng sản phẩm và năm
WITH
MEMBER [Measures].[Avg Sales] AS
    AVG(
        FILTER(
            [Time Dimension].[Year].MEMBERS,
            NOT [Time Dimension].[Year].CURRENTMEMBER IS [Time Dimension].[Year].[ALL]
            AND NOT IsEmpty([Measures].[Total Sales])
        ),
        [Measures].[Total Sales]
    )
SELECT
    FILTER(
        [Product Dimension].[Product Line].MEMBERS,
        NOT [Product Dimension].[Product Line].CURRENTMEMBER IS [Product Dimension].[Product Line].[ALL]
        AND NOT IsEmpty([Measures].[Avg Sales])
    ) ON ROWS,
    {[Measures].[Avg Sales]} ON COLUMNS
FROM [DDS Fashion Shop]
```

Hình 4.16: Truy vấn doanh thu trung bình theo dòng sản phẩm và năm

- Truy vấn tổng doanh thu và số lượng đơn hàng theo lãnh thổ.

```
--7. Tổng doanh thu và số lượng đơn hàng theo lãnh thổ
SELECT
    FILTER(
        [Geography Dimension].[Territory].MEMBERS,
        NOT [Geography Dimension].[Territory].CURRENTMEMBER IS [Geography Dimension].[Territory].[ALL]
        AND NOT IsEmpty([Measures].[Total Sales])
        AND NOT IsEmpty([Measures].[Sales Fact Count])
    ) ON ROWS,
    {[Measures].[Total Sales], [Measures].[Sales Fact Count]} ON COLUMNS
FROM [DDS Fashion Shop]
```

Hình 4.17: Truy vấn tổng doanh thu và số lượng đơn hàng theo lãnh thổ

- Tổng số lượng đơn hàng theo khách hàng.

```
-- 8. Tổng số lượng đơn hàng theo khách hàng
SELECT
    FILTER(
        [Customer Dimension].[Customer Name].MEMBERS,
        NOT IsEmpty([Measures].[Quantity Ordered])
        AND NOT [Customer Dimension].[Customer Name].CURRENTMEMBER IS [Customer Dimension].[Customer Name].[ALL]
    ) ON ROWS,
    {[Measures].[Quantity Ordered]} ON COLUMNS
FROM [DDS Fashion Shop]
```

Hình 4.18: Truy vấn tổng số lượng đơn hàng theo khách hàng

- Tổng số lượng đơn hàng theo sản phẩm.

```
-- 9. Tổng số lượng đơn hàng theo sản phẩm
SELECT
  FILTER(
    [Product Dimension].[Product Name].MEMBERS,
    NOT IsEmpty([Measures].[Quantity Ordered])
    AND NOT [Product Dimension].[Product Name].CURRENTMEMBER IS [Product Dimension].[Product Name].[ALL]
  ) ON ROWS,
  {[Measures].[Quantity Ordered]} ON COLUMNS
FROM [DDS Fashion Shop]
```

Hình 4.19: Truy vấn tổng số lượng đơn hàng theo sản phẩm

- Tổng doanh thu theo quý.

```
-- 10. Tổng doanh thu theo quý
SELECT
  FILTER(
    [Time Dimension].[Quarter].MEMBERS,
    NOT IsEmpty([Measures].[Total Sales]) |
    AND NOT [Time Dimension].[Quarter].CURRENTMEMBER IS [Time Dimension].[Quarter].[ALL]
  ) ON ROWS,
  {[Measures].[Total Sales]} ON COLUMNS
FROM [DDS Fashion Shop]
```

Hình 4.20: Truy vấn tổng doanh thu theo quý

CHƯƠNG 5

KẾT LUẬN

5.1. KẾT QUẢ

5.1.1. Về mặt chức năng

Hệ thống kho dữ liệu đã được xây dựng để đáp ứng nhu cầu phân tích kinh doanh của cửa hàng bán quần áo. Hệ thống tích hợp các chức năng quan trọng như quản lý sản phẩm, thông tin khách hàng, doanh thu và chi phí, giúp việc quản lý trở nên dễ dàng và hiệu quả hơn.

Ngoài ra, hệ thống còn cung cấp các báo cáo phân tích trực quan, hỗ trợ chủ cửa hàng đưa ra các quyết định chiến lược chính xác và kịp thời. Với khả năng truy xuất dữ liệu nhanh chóng và chính xác, hệ thống giúp tối ưu hóa quy trình kinh doanh, tiết kiệm thời gian và nâng cao hiệu quả hoạt động.

5.1.2. Về mặt kỹ thuật

Hệ thống kho dữ liệu được xây dựng theo kiến trúc Data Warehouse, đảm bảo sự ổn định còn có khả năng mở rộng để đáp ứng nhu cầu kinh doanh trong tương lai. Công nghệ ETL được áp dụng để trích xuất, chuyển đổi và tải dữ liệu từ nhiều nguồn khác nhau, giúp dữ liệu được tập trung và xử lý một cách chính xác, nhất quán.

5.2. HẠN CHẾ

5.2.1. Hạn chế về chức năng

Hệ thống hiện tại vẫn còn một số hạn chế trong việc triển khai các chức năng phân tích nâng cao. Chẳng hạn, các tính năng như dự báo doanh thu hay phân tích chuyên sâu về hành vi khách hàng vẫn chưa được tích hợp. Bên cạnh đó, hệ thống chưa hỗ trợ tùy chỉnh báo cáo theo nhu cầu của từng người dùng, gây hạn chế trong việc cá nhân hóa trải nghiệm và đáp ứng các yêu cầu đặc thù.

5.2.2. Hạn chế về kỹ thuật

Về mặt kỹ thuật, hệ thống vẫn phụ thuộc vào một số công cụ của bên thứ ba, điều này có thể tiềm ẩn rủi ro về chi phí phát sinh hoặc ảnh hưởng đến tính ổn định lâu dài. Hiệu suất xử lý dữ liệu cũng có thể bị giảm nếu lượng dữ liệu tăng trưởng đột biến mà chưa có các giải pháp tối ưu hóa kịp thời.

Ngoài ra, vấn đề bảo mật dữ liệu chưa được hoàn thiện toàn diện. Hệ thống cần được bổ sung thêm các lớp bảo mật nâng cao để bảo vệ tốt hơn thông tin nhạy cảm và đảm bảo an toàn dữ liệu trong mọi tình huống.

5.4. KẾT LUẬN

Dự án xây dựng kho dữ liệu phân tích kinh doanh của cửa hàng bán quần áo đã đạt được những thành tựu đáng kể, đáp ứng phần lớn yêu cầu ban đầu. Hệ thống không chỉ hỗ trợ quản lý và phân tích dữ liệu hiệu quả mà còn mở ra tiềm năng phát triển trong tương lai. Tuy nhiên, để đạt được sự hoàn thiện tối ưu, cần tiếp tục cải thiện cả về chức năng và kỹ thuật, đồng thời hướng đến việc mở rộng quy mô và nâng cao trải nghiệm người dùng.

TÀI LIỆU THAM KHẢO

- Các tài liệu từ internet/website:
 - Tài liệu SSAS connect windowform c#
 - Tài liệu Tableau connect windowform c#