

HỌC VIỆN NGÂN HÀNG
KHOA HỆ THỐNG THÔNG TIN QUẢN LÝ



BÁO CÁO BÀI TẬP LỚN

Học phần:

**HỆ HỖ TRỢ RA QUYẾT ĐỊNH
VÀ KINH DOANH THÔNG MINH**

Đề tài:

**XÂY DỰNG KHO DỮ LIỆU VÀ HỆ THỐNG BÁO CÁO
TÌNH HÌNH KINH DOANH CHO HỆ THỐNG SIÊU THỊ
BIG C THĂNG LONG**

Nhóm sinh viên thực hiện : Nhóm 11
Lớp học phần : 231IS31A02
Giảng viên hướng dẫn : TS. Đinh Trọng Hiếu

Hà Nội, 12/2023

HỌC VIỆN NGÂN HÀNG
KHOA HỆ THỐNG THÔNG TIN QUẢN LÝ



BÁO CÁO BÀI TẬP LỚN

Học phần:

**HỆ HỖ TRỢ RA QUYẾT ĐỊNH
VÀ KINH DOANH THÔNG MINH**

Đề tài:

**XÂY DỰNG KHO DỮ LIỆU VÀ HỆ THỐNG BÁO CÁO
TÌNH HÌNH KINH DOANH CHO HỆ THỐNG SIÊU THỊ
BIG C THĂNG LONG**

Nhóm sinh viên thực hiện : Nhóm 11
Lớp học phần : 231IS31A02
Giảng viên hướng dẫn : TS. Đinh Trọng Hiếu

Hà Nội, 12/2023

LỜI CẢM ƠN

Lời đầu tiên, chúng em xin gửi lời cảm ơn chân thành nhất đến TS. Đinh Trọng Hiếu, là giảng viên của chúng em trong học phần Hệ hỗ trợ ra quyết định và kinh doanh thông minh. Cảm ơn thầy đã luôn tận tình trong công việc và giúp đỡ chúng em hoàn thành bài làm của nhóm.

Đồng thời, chúng em xin cảm ơn on khoa Hệ thống thông tin quản lý đã đưa môn học Hệ hỗ trợ ra quyết định và kinh doanh thông minh vào chương trình đào tạo. Đây là môn học rất bổ ích với nhiều kiến thức và rất gần gũi với thực tiễn. Thông qua môn học này, chúng em đã biết những nguyên tắc và những bước cơ bản để xây dựng kho dữ liệu và hệ thống báo cáo kết quả kinh doanh cho doanh nghiệp. Đây là những công cụ cần thiết với rất nhiều doanh nghiệp, đặc biệt là những doanh nghiệp có quy mô lớn trên thị trường hiện nay. Việc được tiếp cận những nội dung này trong chương trình đào tạo ngành Hệ thống thông tin quản lý sẽ là nền tảng để sinh viên chúng em làm quen được với những công việc liên quan tới kho dữ liệu và các báo cáo trong công việc sau này.

Trong quá trình thực hiện bài làm, với sự hướng dẫn của giảng viên, chúng em đã cố gắng vận dụng những kiến thức đã học kết hợp những sự tìm hiểu của cá nhân các thành viên với mong muốn hoàn thành bài làm một cách tốt nhất. Song, do lượng kiến thức là rất lớn, khả năng tìm hiểu còn hạn chế và còn chưa có nhiều kinh nghiệm thực tế nên bài làm của nhóm có thể còn tồn tại những thiếu sót. Chúng em rất mong sẽ nhận được ý kiến đóng góp của thầy để bài làm được hoàn thiện hơn!

Chúng em xin chân thành cảm ơn!

Hà Nội, ngày 22 tháng 12 năm 2023

NHÓM SINH VIÊN THỰC HIỆN

DANH SÁCH THÀNH VIÊN NHÓM

Họ và tên	Mã sinh viên	Tỷ lệ đóng góp
Lưu Quang Hà	22A4050260	21%
Lại Thị Thúy Nga	23A4040093	21%
Trịnh Thị Hương	23A4040056	20%
Nguyễn Trung Kiên	23A4040059	19%
Nguyễn Minh Phụng	23A4040119	19%

BẢNG PHÂN CHIA CÔNG VIỆC

Họ và tên	Nội dung thực hiện
Lưu Quang Hà (Nhóm trưởng)	<ul style="list-style-type: none"> - Chương 1 - Chương 2: Tổng quan doanh nghiệp - Chương 3: Đưa dữ liệu vào kho dữ liệu. - Chương 4: Xây dựng dashboard; Đánh giá hệ thống báo cáo
Lại Thị Thúy Nga	<ul style="list-style-type: none"> - Chương 2: Xác định quy trình nhập kho; Xác định BI User, các quyết định và dữ liệu cần thiết - Chương 3: Thiết kế cơ sở dữ liệu; Đưa dữ liệu vào kho dữ liệu - Chương 4: Xây dựng và trình bày nội dung báo cáo
Trịnh Thị Hương	<ul style="list-style-type: none"> - Chương 2: Cơ cấu tổ chức; Xác định BI User, các quyết định và dữ liệu cần thiết - Chương 3: Thiết kế cơ sở dữ liệu - Chương 4: Xây dựng và trình bày nội dung báo cáo
Nguyễn Trung Kiên	<ul style="list-style-type: none"> - Chương 2: Cơ cấu tổ chức; Xác định BI User, các quyết định và dữ liệu cần thiết - Chương 3: Lựa chọn công cụ - Chương 4: Lựa chọn công cụ; Xây dựng và trình bày nội dung báo cáo
Nguyễn Minh Phụng	<ul style="list-style-type: none"> - Chương 2: Cơ cấu tổ chức; Xác định BI User, các quyết định và dữ liệu cần thiết - Chương 3: Lựa chọn công cụ - Chương 4: Lựa chọn công cụ; Xây dựng và trình bày nội dung báo cáo

MỤC LỤC

LỜI CẢM ƠN	i
DANH SÁCH THÀNH VIÊN NHÓM.....	ii
BẢNG PHÂN CHIA CÔNG VIỆC	ii
MỤC LỤC.....	iii
DANH MỤC HÌNH ẢNH	v
DANH MỤC BẢNG.....	vii
LỜI MỞ ĐẦU	1
1. Lý do chọn đề tài	1
2. Mục tiêu đề tài	2
3. Kết cấu đề tài	2
CHƯƠNG 1: TỔNG QUAN VỀ KHO DỮ LIỆU VÀ BÁO CÁO TÌNH HÌNH KINH DOANH	4
1.1. Hệ thống thông tin hỗ trợ ra quyết định	4
1.2. Kho dữ liệu	5
1.3. Báo cáo tình hình kinh doanh.....	9
CHƯƠNG 2: PHÁT BIỂU BÀI TOÁN	12
2.1. Tổng quan về doanh nghiệp.....	12
2.2. Cơ cấu tổ chức	14
2.3. Mô tả hoạt động nghiệp vụ	15
2.3.1. Nghiệp vụ bán hàng.....	15
2.3.2. Nghiệp vụ nhập kho	15
2.4. Báo cáo tình hình kinh doanh	17
2.4.1. Người sử dụng (BI Users)	17
2.4.2. Các quyết định của BI Users và dữ liệu cần và phân tích	19
CHƯƠNG 3: XÂY DỰNG KHO DỮ LIỆU CHO DOANH NGHIỆP	22
3.1. Thiết kế cơ sở dữ liệu	22
3.1.1. Thiết kế các bảng Dim.....	22
3.1.2. Thiết kế các bảng Fact.....	23
3.2. Xây dựng lược đồ dữ liệu	25
3.3. Công cụ triển khai kho dữ liệu	25
3.3.1. Microsoft SQL Server	25
3.3.2. Microsoft Visual Studio	25

3.3.3.	SQL Server Integration Services (SSIS)	26
3.4.	Đưa dữ liệu vào kho dữ liệu	26
3.4.1.	Bảng Dim_NhanVien	28
3.4.2.	Bảng Dim_LoaiHang	29
3.4.3.	Bảng Dim_NCC	30
3.4.4.	Bảng Dim_SanPham	32
3.4.5.	Bảng Dim_KhachHang	33
3.4.6.	Bảng Fact_ChiTietBan	35
3.4.7.	Bảng Fact_ChiTietNhap	40
CHƯƠNG 4: XÂY DỰNG BÁO CÁO TÌNH HÌNH KINH DOANH.....		43
4.1.	Giới thiệu công cụ Tableau.....	43
4.2.	Xây dựng báo cáo tình hình kinh doanh.....	43
4.2.1.	Báo cáo doanh thu	43
4.2.2.	Báo cáo số lượng hàng bán	45
4.2.3.	Báo cáo số lượng hàng nhập vào và bán ra	47
4.3.	Đánh giá hệ thống báo cáo	47
KẾT LUẬN		49
TÀI LIỆU THAM KHẢO.....		50

DANH MỤC HÌNH ẢNH

Hình 1-1: Mô hình kiến trúc và thành phần của hệ hỗ trợ ra quyết định.....	5
Hình 1-2: Mô hình kiến trúc kho dữ liệu	6
Hình 1-3: Sơ đồ hình sao.....	8
Hình 1-4: Sơ đồ hình bông tuyết.....	8
Hình 1-5: Sơ đồ chòm sao.....	9
Hình 1-6: Quá trình phát triển của BI	10
Hình 1-7: Mô hình kiến trúc mức cao của BI	10
Hình 2-1: Quy trình bán hàng tại siêu thị.....	15
Hình 2-2: Biểu đồ hoạt động quy trình nhập kho	16
Hình 3-1: Lược đồ dữ liệu.....	25
Hình 3-2: Tạo connection với đầu vào là file csv	27
Hình 3-3: Tạo connect với đầu ra là database trong Microsoft SQL Server	28
Hình 3-4: Cấu trúc bảng Dim_NhanVien	28
Hình 3-5: Data flow của Dim_NhanVien	29
Hình 3-6: Kết quả đổ dữ liệu vào bảng Dim_NhanVien trên SSDT	29
Hình 3-7: Dữ liệu trên bảng Dim_NhanVien.....	29
Hình 3-8: Cấu trúc bảng Dim_LoaiHang.....	29
Hình 3-9: Data flow của Dim_LoaiHang.....	30
Hình 3-10: Kết quả đổ dữ liệu vào bảng Dim_LoaiHang trên SSDT.....	30
Hình 3-11: Dữ liệu trên bảng Dim_LoaiHang	30
Hình 3-12: Cấu trúc bảng Dim_NCC	31
Hình 3-13: Data flow của Dim_NCC	31
Hình 3-14: Kết quả đổ dữ liệu vào bảng Dim_NCC trên SSDT.....	31
Hình 3-15: Dữ liệu trên bảng Dim_NCC.....	32
Hình 3-16: Cấu trúc bảng Dim_SanPham.....	32
Hình 3-17: Cấu trúc bảng Dim_SanPham.....	33
Hình 3-18: Kết quả đổ dữ liệu vào bảng Dim_SanPham trên SSDT.....	33
Hình 3-19: Dữ liệu trên bảng Dim_SanPham.....	33
Hình 3-20: Cấu trúc bảng Dim_KhachHang.....	34

Hình 3-21: Data flow của Dim_KhachHang.....	34
Hình 3-22: Kết quả đổ dữ liệu vào bảng Dim_KhachHang trên SSDT.....	34
Hình 3-23: Dữ liệu trên bảng Dim_KhachHang	35
Hình 3-24: Cấu trúc bảng Fact_ChiTietBan	35
Hình 3-25: Data flow của Fact_ChiTietBan	36
Hình 3-26: Sort bảng Hóa Đơn Bán.....	36
Hình 3-27: Sort bảng Chi Tiết Hóa Đơn Bán	36
Hình 3-28: Ghép bảng Hóa Đơn Bán và Chi Tiết Hóa Đơn Bán bằng cột orderID .	37
Hình 3-29: Chuyển đổi kiểu dữ liệu đầu vào cho bảng Fact_ChiTietBan.....	38
Hình 3-30: Nối các cột đầu vào và cột đích của bảng Fact_ChiTietBan.....	39
Hình 3-31: Kết quả đổ dữ liệu vào bảng Fact_ChiTietBan trên SSDT	39
Hình 3-32: Dữ liệu bảng Fact_ChiTietBan.....	40
Hình 3-33: Cấu trúc bảng Fact_ChiTietNhap.....	40
Hình 3-34: Data flow của Fact_ChiTietNhap.....	41
Hình 3-35: Kết quả đổ dữ liệu vào bảng Fact_ChiTietNhap trên SSDT	41
Hình 3-36: Dữ liệu trên bảng Fact_ChiTietNhap	42
Hình 4-1: Báo cáo doanh thu	44
Hình 4-2: Báo cáo số lượng hàng bán.....	46
Hình 4-3: Báo cáo số lượng hàng nhập/xuất - Giai đoạn 2016-2018	47

DANH MỤC BẢNG

Bảng 2-1: Một số thông tin về Big C Thăng Long	13
Bảng 2-2: Cơ cấu tổ chức Siêu thị Big C Thăng Long	14
Bảng 2-3: Danh sách người sử dụng	19
Bảng 2-4: Các quyết định của người sử dụng và dữ liệu cần cung cấp	21
Bảng 3-1: Thiết kế bảng Dim_LoaiHang	22
Bảng 3-2: Thiết kế bảng Dim_KhachHang	22
Bảng 3-3: Thiết kế bảng Dim_NhanVien	23
Bảng 3-4: Thiết kế bảng Dim_SanPham	23
Bảng 3-5: Thiết kế bảng Dim_NCC	23
Bảng 3-6: Thiết kế bảng Fact_ChiTietNhap	24
Bảng 3-7: Thiết kế bảng Fact_ChiTietBan	24

LỜI MỞ ĐẦU

1. Lý do chọn đề tài

Trong lĩnh vực bán lẻ hiện nay, nhiều hệ thống cửa hàng đã triển khai các phần mềm hỗ trợ bán hàng. Không chỉ những hệ thống lớn mà những cửa hàng tạp hóa nhỏ lẻ cũng dần chuyển đổi từ hình thức bán hàng truyền thống sang việc quét mã vạch để ghi lại thông tin đối với mỗi sản phẩm được bán ra. Việc sử dụng các phần mềm bán hàng sẽ giúp các cửa hàng dễ dàng ghi lại được số lượng sản phẩm được bán ra kèm theo doanh thu và lợi nhuận. Các hệ thống siêu thị hay cửa hàng bán lẻ cũng thông qua đó có thể lưu trữ những thông tin về khách hàng để làm tốt hơn hoạt động chăm sóc khách hàng.

Phần lớn các doanh nghiệp hiện nay chưa có hệ thống thông tin tích hợp, làm nảy sinh tình trạng bất cân xứng thông tin giữa các bộ phận. Chẳng hạn, các doanh nghiệp thường mua phần mềm bán hàng và phần mềm quản lý kho của hai nhà cung cấp dịch vụ khác nhau, dẫn đến thông tin không đồng bộ. Thông tin về số lượng hàng hóa bán ra và nhập vào không được cập nhật kịp thời hoặc cập nhật sẽ làm cho việc theo dõi hoạt động kinh doanh trở nên khó khăn và đôi khi sẽ không kịp thời đặt thêm hàng hóa từ nhà cung cấp. Các doanh nghiệp cũng cần có sự phân tích, đánh giá hoạt động kinh doanh sau một khoảng thời gian nhất định để phát huy điểm mạnh và khắc phục điểm yếu. Ngoài ra, các quyết định kinh doanh của các nhà quản lý có thể thiếu chính xác nếu các dữ liệu thu về là không đầy đủ, không được phân tích kỹ càng và khó theo dõi. Trong vấn đề này, các báo cáo tình hình kinh doanh là công cụ hỗ trợ đắc lực cho các nhà quản lý.

Tuy nhiên, việc xây dựng các báo cáo sẽ đòi hỏi sự tổng hợp dữ liệu từ nhiều nguồn và phải là dữ liệu trong một khoảng thời gian đủ lâu để mang lại tính chính xác cho các thông tin được rút ra. Khi đó, ***kho dữ liệu*** trở thành sự lựa chọn của nhiều doanh nghiệp bán lẻ, nhất là đối với các hệ thống có nhiều chi nhánh, có nhu cầu tổng hợp dữ liệu của các chi nhánh để phân tích, dự báo, đáp ứng tốt nhu cầu của khách hàng cũng như hỗ trợ cho hoạt động quản lý, ra quyết định của các cấp lãnh đạo.

Nhận thức được các vấn đề nêu trên, Nhóm 11 đã quyết định lựa chọn đề tài ***“Xây dựng kho dữ liệu và hệ thống báo cáo tình hình kinh doanh cho hệ thống siêu thị***

Big C Thăng Long” để thực hiện bài tập lớn trong học phần này.

2. Mục tiêu đề tài

Bài làm sẽ tiến hành xác định các hoạt động nghiệp vụ của doanh nghiệp và phân tích các quyết định kinh doanh. Từ đó, xác định các dữ liệu cần thu thập, xây dựng kho dữ liệu để hỗ trợ người sử dụng trong việc đưa ra các quyết định kinh doanh. Các hoạt động trên cũng là cơ sở cho việc xây dựng hệ thống báo cáo để cung cấp góc nhìn trực quan cho người sử dụng, hỗ trợ tốt hơn cho việc ra quyết định.

Bài làm được thực hiện bằng việc trả lời lần lượt 03 câu hỏi nghiên cứu:

- *Câu hỏi số 1:* Doanh nghiệp có những nghiệp vụ nào?
- *Câu hỏi số 2:* Xác định và phân tích các quyết định kinh doanh của doanh nghiệp?
- *Câu hỏi số 3:* Xây dựng kho dữ liệu cho doanh nghiệp
- *Câu hỏi số 4:* Các nhà quản lý của doanh nghiệp sử dụng những loại báo cáo nào?

3. Kết cấu đề tài

Ngoài phần mở đầu, kết luận và tài liệu tham khảo, bài làm được chia ra thành 03 chương chính:

Chương 1: Tổng quan về kho dữ liệu và báo cáo tình hình kinh doanh

Chương 1 trình bày cơ sở lý thuyết về hệ hỗ trợ ra quyết định, báo cáo tình hình kinh doanh và nhấn mạnh việc trình bày về kho dữ liệu, bao gồm việc trình bày khái niệm và các yếu tố liên quan và mối quan hệ giữa ba đối tượng trên. Đây cũng là cơ sở cho việc xây dựng kho dữ liệu.

Chương 2: Phát biểu bài toán

Chương 2 trình bày các thông tin tổng quan về doanh nghiệp, các hoạt động nghiệp vụ, xác định các yếu tố có liên quan và làm rõ các yêu cầu cho việc xây dựng kho dữ liệu, phục vụ cho hệ thống báo cáo tình hình kinh doanh. Mục tiêu quan trọng của chương này là xác định các yếu tố có liên quan đến hệ thống báo cáo tình hình kinh doanh: người sử dụng, các quyết định cần hỗ trợ. Từ đó, xác định những dữ liệu nào cần thu thập để cung cấp cho người sử dụng, hỗ trợ cho hoạt động ra quyết định.

Chương 3: Xây dựng kho dữ liệu cho doanh nghiệp

Chương 3 trình bày quá trình xây dựng kho dữ liệu với các bước: xây dựng cơ sở dữ liệu, xây dựng các lược đồ dữ liệu và đưa dữ liệu vào kho dữ liệu.

Chương 4: Xây dựng hệ thống báo cáo tình hình kinh doanh

Chương 4 trình bày về những báo cáo cho hoạt động kinh doanh của doanh nghiệp, sử dụng dữ liệu trong kho dữ liệu được xây dựng từ chương 3 thông qua công cụ Tableau.

CHƯƠNG 1: TỔNG QUAN VỀ KHO DỮ LIỆU VÀ BÁO CÁO TÌNH HÌNH KINH DOANH

1.1. Hệ thống thông tin hỗ trợ ra quyết định

1.1.1. Định nghĩa hệ hỗ trợ ra quyết định

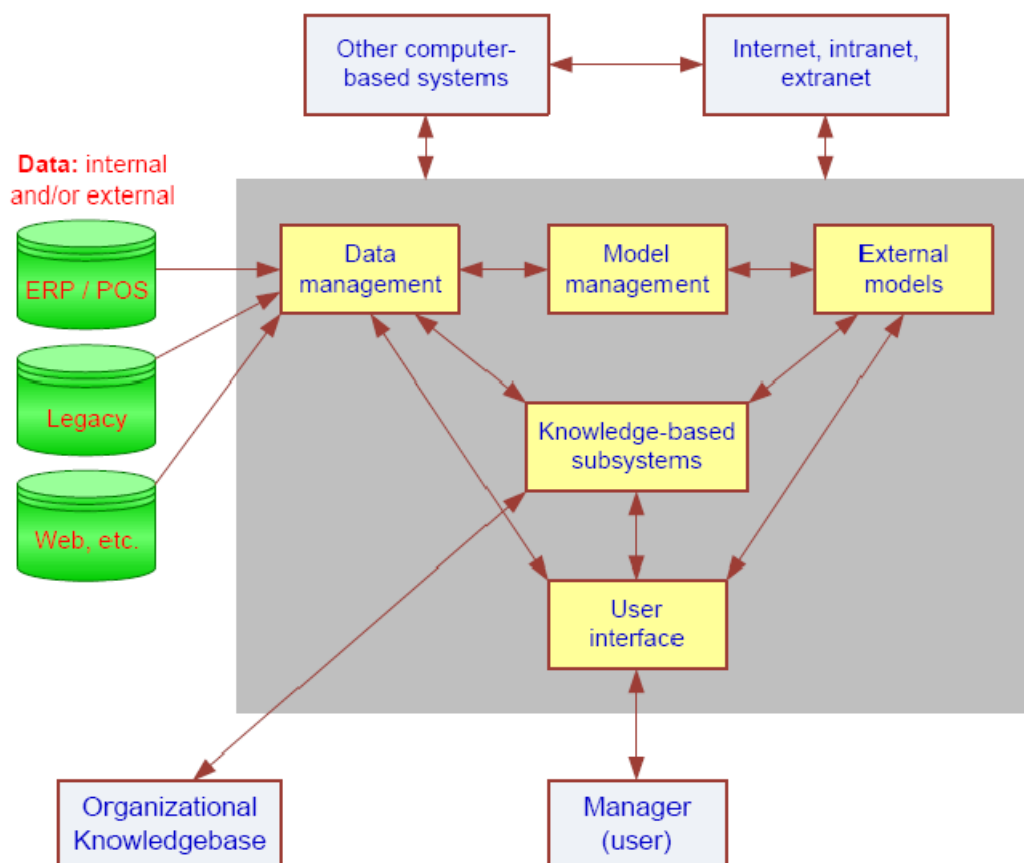
Khái niệm về hệ hỗ trợ ra quyết định (*Decision Support System – viết tắt: DSS*) được Scott Morton công bố vào đầu những năm 1970 với nội dung là các hệ hỗ trợ ra quyết định trong quản lý. Hệ hỗ trợ ra quyết định được định nghĩa như là “các hệ thống tương tác với máy tính, nó giúp các nhà quản lý ra quyết định sử dụng dữ liệu và các mô hình để giải quyết bài toán phi cấu trúc”.

Một định nghĩa hệ hỗ trợ ra quyết định được sử dụng để làm việc có thể được phát biểu như sau: *một hệ hỗ trợ ra quyết định là một hệ thống tương tác, mềm dẻo, tương thích dựa trên máy tính, đặc biệt được phát triển để hỗ trợ giải quyết các bài toán quản lý phi cấu trúc nhằm hoàn thiện việc ra quyết định. Nó tận dụng dữ liệu, cung cấp giao diện người dùng tiện dụng, và cho phép người ra các quyết định hiểu thấu đáo hoạt động của mình* (Trần Thị Thu Hà, 2020).

1.1.2. Kiến trúc thành phần của hệ hỗ trợ ra quyết định

Hệ hỗ trợ ra quyết định được cấu thành từ các hệ con sau:

- *Quản trị dữ liệu*: bao gồm một hoặc nhiều cơ sở dữ liệu (viết tắt: CSDL), chứa dữ liệu có liên quan đến các tình huống được xem xét và được một phần mềm gọi là hệ quản trị CSDL quản lý.
- *Quản trị mô hình*: là một gói bao gồm các phần mềm tài chính, thống kê, khoa học quản lý hoặc các mô hình định lượng khác cung cấp các khả năng phân tích mọi hệ thống, và một hệ quản trị phần mềm thích hợp.
- *Quản trị hội thoại/Truyền thông*: cung cấp một giao diện tiện dụng để người dùng tương tác với máy tính.
- *Quản trị tri thức*: là hệ con tùy chọn có thể hỗ trợ bất cứ hệ con hoặc một hành động nào khác và như một thành phần độc lập.



Hình 1-1: Mô hình kiến trúc và thành phần của hệ hỗ trợ ra quyết định

Nguồn: (Khoa Hệ thống thông tin quản lý)

1.2. Kho dữ liệu

1.2.1. Định nghĩa Kho dữ liệu

Bill Inmon, người được coi là cha đẻ của kho dữ liệu, đưa ra định nghĩa về kho dữ liệu là “*tập hợp dữ liệu theo định hướng chủ đề, tích hợp, không biến đổi và được thu thập theo thời gian để hỗ trợ các quyết định của ban quản lý, chứa dữ liệu chi tiết của công ty, được sử dụng cho nhiều mục đích khác nhau*” (Bill Inmon, 2005).

1.2.2. Các đặc trưng của Kho dữ liệu

- *Hướng chủ đề (Subject-Oriented)*: Dữ liệu được phân lớp, tập hợp, xử lý và lưu trữ theo từng chủ đề khác nhau. Những dữ liệu của từng chủ đề chính sẽ được liên kết bởi các khóa chính và sau đó thì đưa vào cùng vị trí.
- *Tích hợp (Integrated)*: Trong một tổ chức, có rất nhiều dữ liệu từ nhiều hệ thống khác nhau. Việc hợp nhất các dữ liệu này trở thành một nhóm các dữ liệu có nghĩa cho việc phân tích là rất khó khăn. Dữ liệu mang đi tích hợp phải được đảm

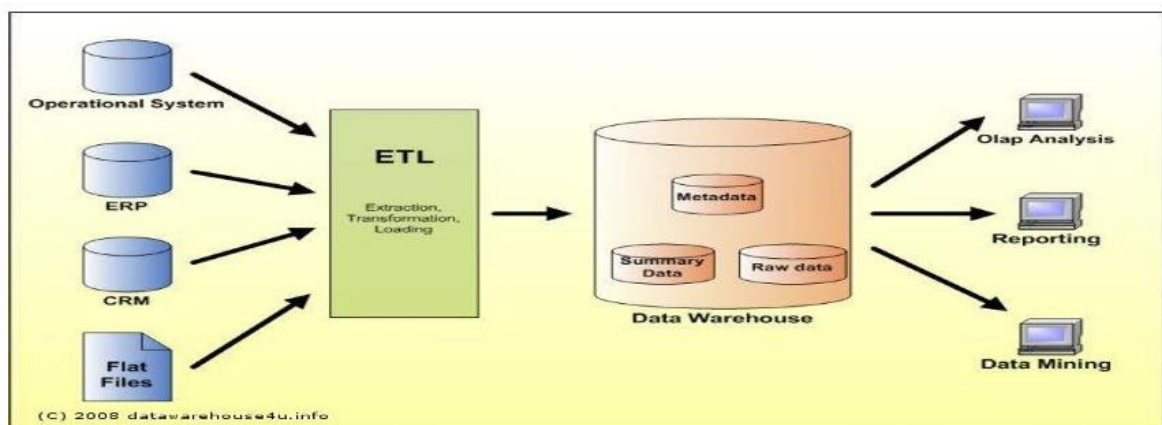
bảo tính nhất quán, đôi khi chấp nhận tính dư thừa để tăng tính hiệu quả của các truy vấn.

- *Bất biến (Non-volatile)*: Dữ liệu được tải (Load) vào kho dữ liệu ở lần đầu tiên dưới dạng bản sao tại một thời điểm (Snapshot) và sau đó thường xuyên được làm tươi (Refresh). Trước khi dữ liệu trong hệ thống điều hành tác nghiệp được xác định thì trong kho dữ liệu, thông tin cần phải được tải vào đó. Mặc dù những dữ liệu cũ đã có từ rất lâu trong kho dữ liệu, nhưng khi dữ liệu mới được nhập vào thì những dữ liệu cũ đó vẫn còn nguyên vẹn, không bị xóa hay biến mất.

- *Giá trị lịch sử (Time-variant)*: Kho dữ liệu có thể lưu trữ được một lượng lớn dữ liệu dù những dữ liệu cũ đã có từ rất lâu trong kho dữ liệu được lưu trong data warehouse dưới dạng bản sao dữ liệu, phản ánh lại những giá trị của dữ liệu tại những thời điểm khác nhau, cho phép so sánh dữ liệu theo từng khoảng thời gian khác nhau. Yếu tố đóng vai trò quan trọng nhất ở đặc điểm này là thời gian vì nó cung cấp những đặc điểm về thời gian đồng thời đảm bảo tính đơn nhất của dữ liệu (Đặng Văn Ninh, 2004).

1.2.3. Kiến trúc kho dữ liệu

Một kho dữ liệu hoàn chỉnh có thể chia thành nhiều phần tùy theo nhu cầu của công ty nhưng sẽ có một số phần chính như sau:



Hình 1-2: Mô hình kiến trúc kho dữ liệu

Nguồn: (Ayad Hameed Mousa, 2014)

- *Nguồn dữ liệu*: dữ liệu được thu thập được từ các hệ thống tác nghiệp. Dữ liệu có thể được cấu trúc, bán cấu trúc và không cấu trúc.

- *Phần ETL (Extract – Transform – Load)*: bao gồm những công cụ giúp kết hợp nhiều loại dữ liệu đến từ nhiều nguồn khác nhau của các hệ thống tác nghiệp (Extract), sau đó thực hiện chuẩn hóa, đưa về định dạng phù hợp (Transform) và nạp những dữ liệu đó vào kho dữ liệu (Load).

- *Kho dữ liệu*: Đây là nơi lưu trữ dữ liệu để tạo nên data warehouse hoàn chỉnh.

- *Khai thác dữ liệu*: công cụ giúp thực hiện khai thác dữ liệu sau khi đã thực hiện xây dựng hoàn chỉnh một data warehouse, như công cụ thực hiện xử lý phân tích trực tuyến (OLAP), hệ thống khai phá dữ liệu, ...

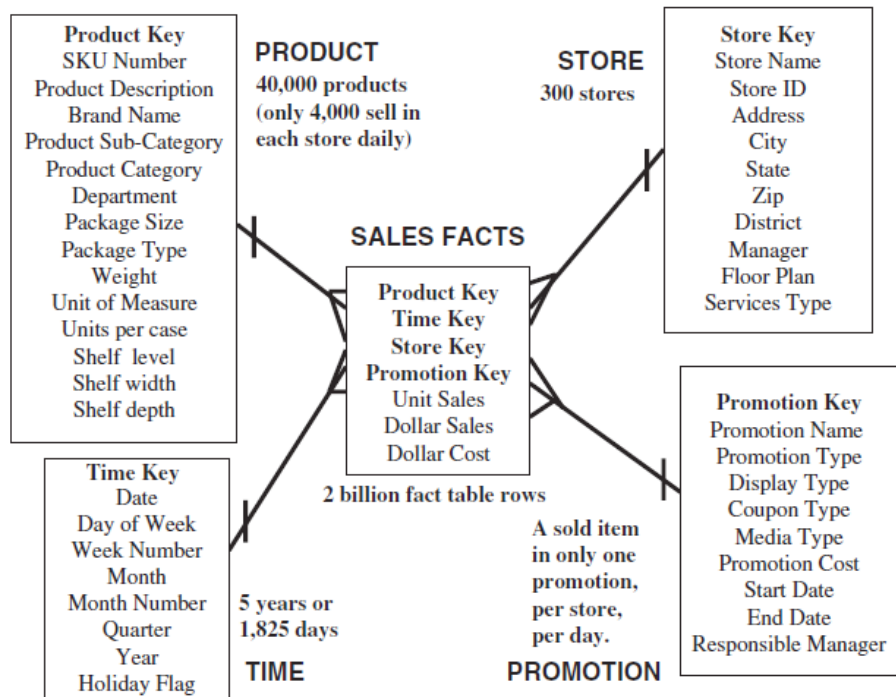
1.2.4. Các thành phần chính của mô hình dữ liệu trong kho dữ liệu

- *Bảng Dimension*: Các bảng Dimension chứa các thuộc tính mô tả các thực thể trong tổ chức và được tham chiếu tới bảng Fact. Dữ liệu trong bảng Dimension là dữ liệu tĩnh và giải thích cho các khóa chính tương ứng trong bảng Fact.

- *Bảng Fact*: Bảng Fact là bảng chứa lượng dữ liệu lớn trong kho dữ liệu và có hai loại cột chính: Khóa chính của các bảng Dimension liên quan đến bảng Fact và dữ liệu số (measurement). Bảng Fact có thể chứa dữ liệu ở cấp chi tiết nhất của một hoạt động xảy ra trong một tổ chức hoặc ở cấp tổng hợp dữ liệu. Mỗi bản ghi trong bảng Fact thể hiện một sự kiện xảy ra trong một khoảng thời gian xác định.

1.2.5. Các dạng sơ đồ của Kho dữ liệu

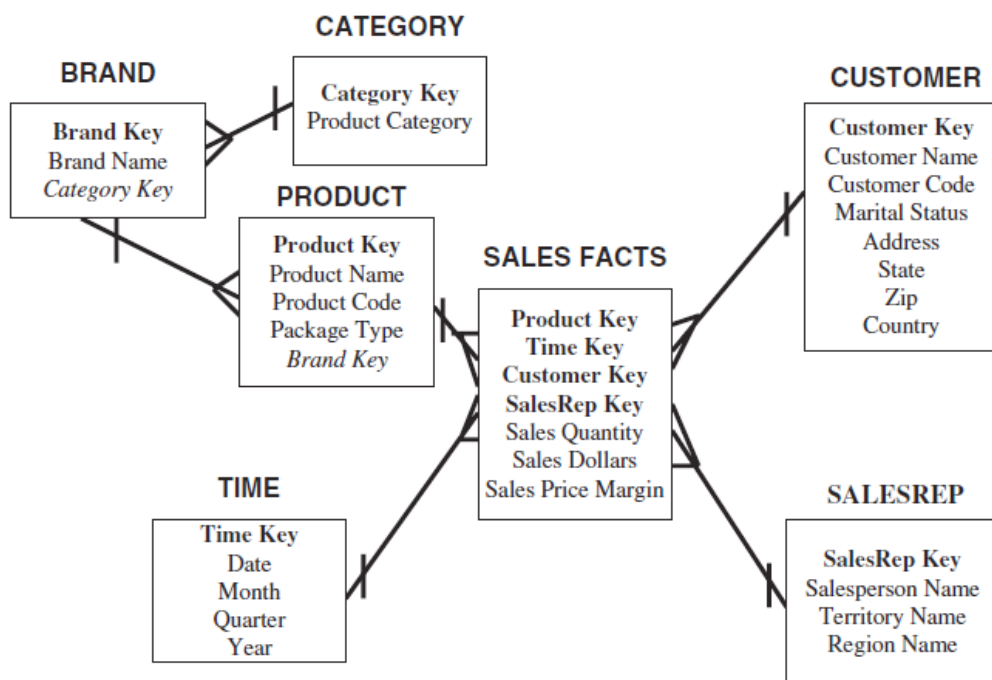
- *Sơ đồ hình sao (star schema)*: Một bảng sự kiện ở trung tâm được kết nối với một tập các bảng chiều.



Hình 1-3: Sơ đồ hình sao

Nguồn: (Paulraj Ponniah, 2010)

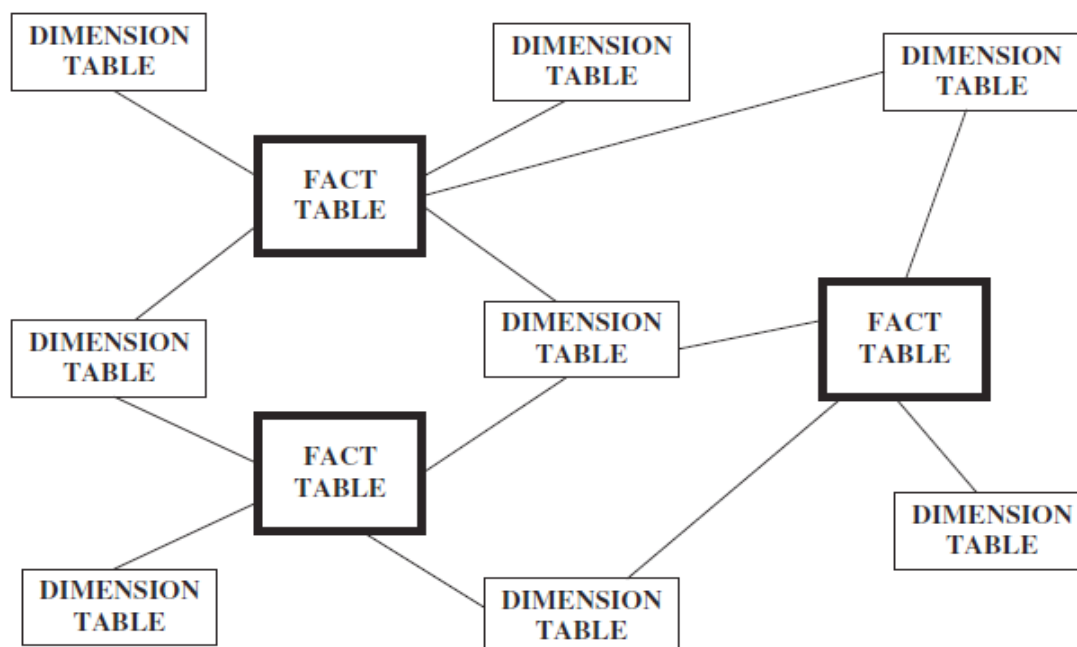
- *Sơ đồ bông tuyết (Snowflake schema)*: Một mở rộng của sơ đồ hình sao trong đó một vài cấu trúc chiều được chuẩn hóa thành một tập các bảng chiều nhỏ hơn, hình thức tương tự như bông tuyết.



Hình 1-4: Sơ đồ hình bông tuyết

Nguồn: (Paulraj Ponniah, 2010)

- Sơ đồ chòm sao sự kiện (Fact constellations schema): Bảng sự kiện phức chia sẻ các bảng chiều, tạo khung nhìn một tập các “ngôi sao”, nên còn được gọi “sơ đồ ngân hà” (galaxy schema) hoặc “chòm sao sự kiện”.



Hình 1-5: Sơ đồ chòm sao

Nguồn: (Paulraj Ponniah, 2010)

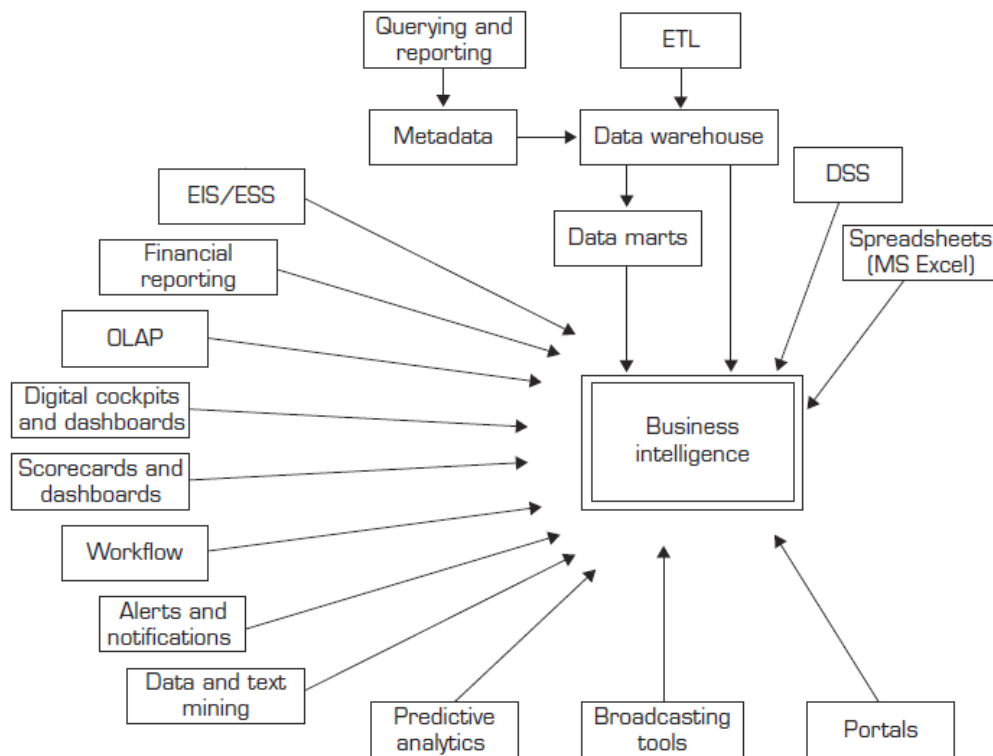
1.3. Báo cáo tình hình kinh doanh

1.3.1. Định nghĩa báo cáo tình hình kinh doanh

Báo cáo tình hình kinh doanh (Business Intelligence – viết tắt: BI) biểu thị một nội dung tùy ý, có nghĩa là gồm những thứ khác nhau cho những người khác nhau, nên thường có những sự nhầm lẫn trong khái niệm. BI cho phép truy cập tương tác (đôi khi trong thời gian thực) vào dữ liệu, để thao tác dữ liệu và cung cấp cho người quản lý và những người phân tích về khả năng tiến hàng ác phân tích thích hợp. Khi phân tích các trạng thái lịch sử và hiện tại, các nhà ra quyết định thu được thông tin có giá trị giúp họ đưa ra các quyết định tốt hơn. BI giúp chuyển đổi dữ liệu thành tri thức, sau đó chuyển thành quyết định và cuối cùng là hành động (Trần Thị Thu Hà, 2020).

Dựa vào Hình 2-6, có thể thấy rằng BI được phát triển dựa trên DSS. BI là kết quả

của quá trình phát triển liên tục và DSS là một trong những yếu tố gốc của BI.

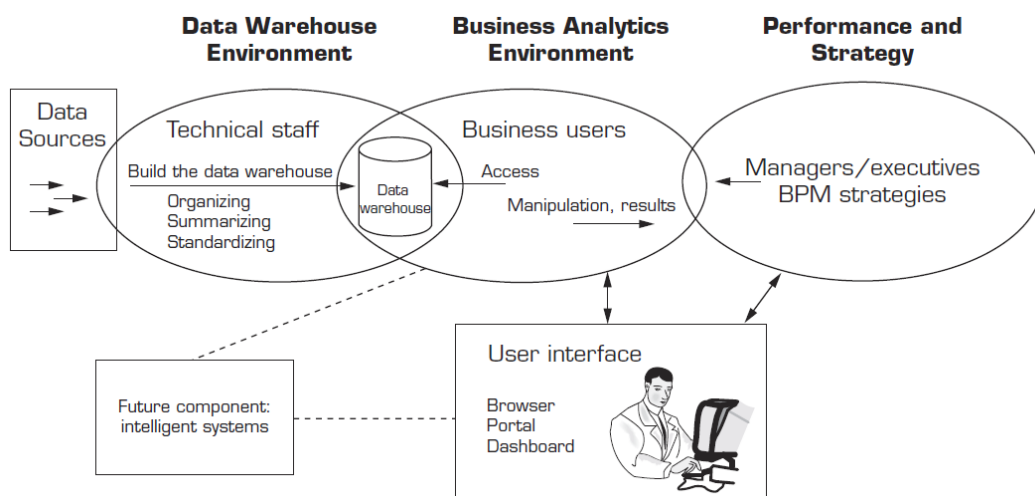


Hình 1-6: Quá trình phát triển của BI

Nguồn: (Efraim Turban, 2008)

1.3.2. Kiến trúc BI

Một hệ thống BI gồm bốn thành phần chính:



Hình 1-7: Mô hình kiến trúc mức cao của BI

Nguồn: (Efraim Turban, 2008)

- *Kho dữ liệu (Data Warehouse – viết tắt: DW)* cùng các nguồn dữ liệu của nó. Người dùng có thể xem hoặc thao tác dữ liệu và thông tin bên trong DW. Một số DW hiện nay còn bao gồm cả dữ liệu hiện tại, hỗ trợ việc theo dõi và ra quyết định trong thời gian thực.

- *Các công cụ phân tích (Business Analytics)*, bao gồm các công cụ và kỹ thuật thuộc hai loại chính:

- Báo cáo và truy vấn
- Khai phá dữ liệu, khai phá văn bản, khai phá web, các công cụ thống kê và toán học khác

- *Bộ quản lý hiệu năng kinh doanh (Business Performance Strategy – viết tắt: BPM)* để theo dõi và phân tích hiệu suất. BPM giúp giám sát, đo lường và so sánh doanh thu, lợi nhuận, chi phí.

- *Giao diện người dùng (User Interface)*, gồm hai thành phần:

- *Bảng điều khiển (Dashboard)*, cung cấp góc nhìn trực quan về hiệu quả doanh nghiệp được biểu diễn dưới dạng đồ thị.
- *Các công cụ phân phối thông tin khác* như các cổng thông tin của doanh nghiệp, buồng điều khiển số và các công cụ trực quan khác.

CHƯƠNG 2: PHÁT BIỂU BÀI TOÁN

2.1. Tổng quan về doanh nghiệp

Big C Thăng Long được thành lập năm 1999, là liên doanh giữa công ty TNHH Nhà nước Một thành viên Thăng Long GTC của Việt Nam và Công ty VINDEMIASAS (thuộc tập đoàn Casino của Pháp) với số vốn điều lệ là 30 triệu USD. Trong đó Công ty TNHH Nhà nước một thành viên Thăng Long GTC góp 35% và VINDEMIASAS góp 65%.

Tên đầy đủ	Công ty TNHH Thương Mại Dịch Vụ Quốc Tế Big C Thăng Long
Trụ sở chính	Số 222 Trần Duy Hưng, Trung Hòa, Cầu Giấy, Hà Nội
Điện thoại	(024)3 784 6900
Webiste	www.bigc.vn
Mã số doanh nghiệp	0100956381
Các cơ sở kinh doanh trực thuộc	<i>Cơ sở 1: Siêu thị Big C Thăng Long</i> Địa chỉ: Số 222 Trần Duy Hưng, Trung Hòa, Cầu Giấy, Hà Nội
	<i>Cơ sở 2: Siêu thị Big C The Garden</i> Địa chỉ: Tầng hầm thứ nhất Trung tâm thương mại The Garden, đường Mễ Trì, Mỹ Đình 1, Nam Từ Liêm, Hà Nội
	<i>Cơ sở 3: Siêu thị Big C Long Biên</i> Địa chỉ: Tầng hầm, Trung tâm thương mại Gia Thụy, Khu X2, Gia Thụy, Long Biên, Hà Nội

	<p><i>Cơ sở 4: Siêu thị Big C Mê Linh</i></p> <p>Địa chỉ: Tầng 1, Trung tâm thương mại VLXD và trang thiết bị nội thất Mê Linh Plaza, Km8 đường cao tốc Thăng Long – Nội Bài, thị trấn Quang Minh, Mê Linh, Hà Nội</p>
	<p><i>Cơ sở 5: Siêu thị Big C Hồ Gươm</i></p> <p>Địa chỉ: Tầng 1 & 2 Trung tâm thương mại, văn phòng và nhà ở cao cấp Hồ Gươm Plaza, Mỗ Lao, Hà Đông, Hà Nội</p>
	<p><i>Cơ sở 6: Tops Market Lê Trọng Tấn</i></p> <p>Địa chỉ: Tầng hầm 1 & 2, Tòa nhà Artemis, số 3 Lê Trọng Tấn, Khương Mai, Thanh Xuân, Hà Nội</p>
	<p><i>Cơ sở 7: Siêu thị Go! Market Nguyễn Xiển</i></p> <p>Địa chỉ: Tầng hầm, Tổ hợp thương mại, siêu thị, văn phòng, nhà ở Eco Green City, Nguyễn Xiển, Tân Triều, Thanh Trì, Hà Nội</p>

Bảng 2-1: Một số thông tin về Big C Thăng Long

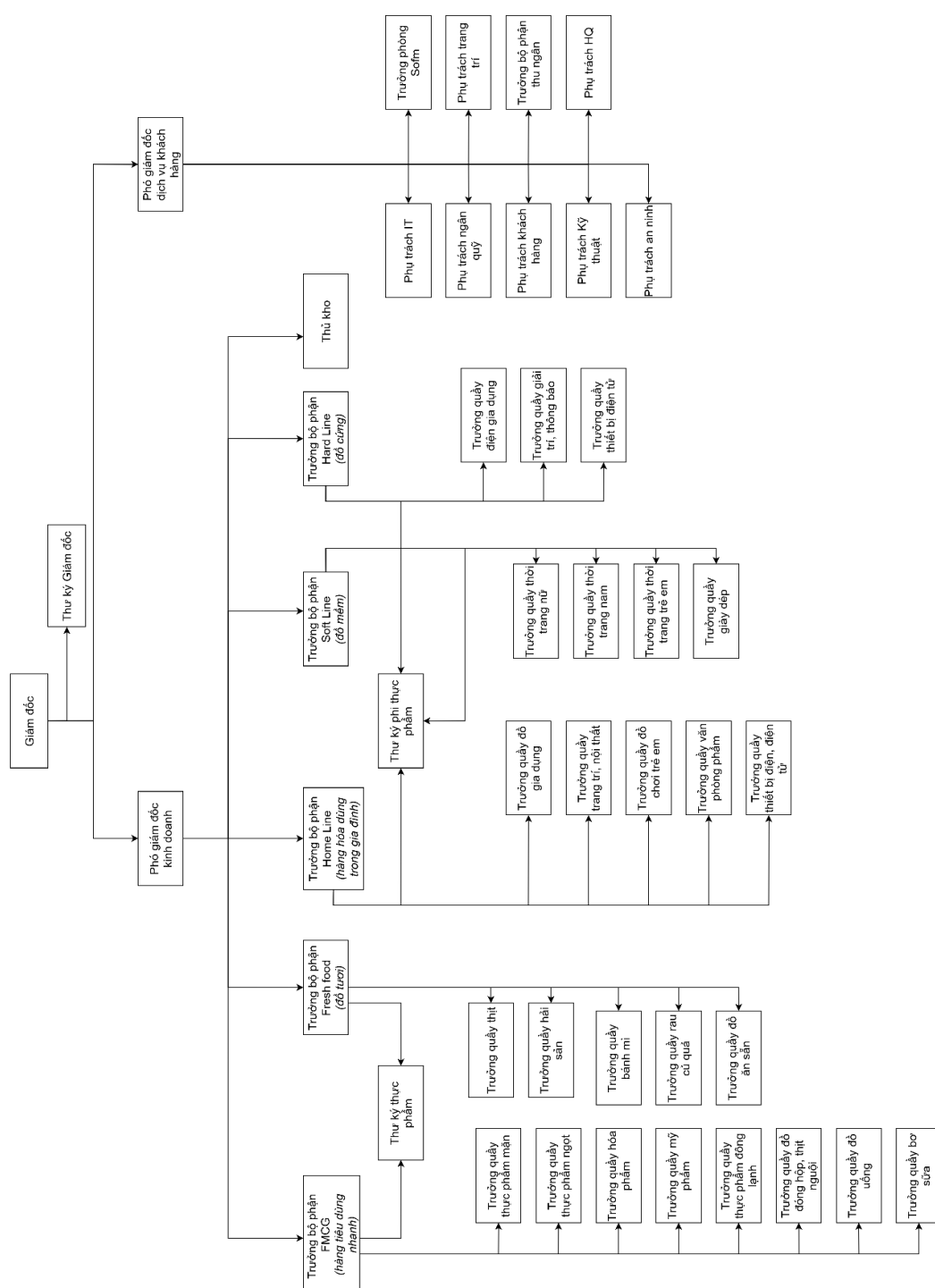
Nguồn: (Công ty TNHH Thương Mại Dịch Vụ Quốc Tế Big C Thăng Long, 2019)

Sau khi Tập đoàn Central Group mua lại Big C từ tay Casino, hàng loạt đại siêu thị đã được đổi tên thành Go! và các siêu thị nằm trong các tòa nhà được đổi tên thành Tops Market. Big C Thăng Long hiện vẫn được giữ nguyên tên gọi ban đầu. Với khẩu hiệu “*Giá luôn luôn thấp*”, Big C cung cấp cho người tiêu dùng hơn 50.000 mặt hàng, từ thực phẩm, sữa, bánh kẹo đến các mặt hàng thời trang, đồ gia dụng, ... với giá thấp hơn so với khi mua ở những nơi khác.

Sản phẩm được bày bán trong hệ thống Big C Thăng Long rất đa dạng về chủng loại. Đây là cơ hội để người tiêu dùng được lựa chọn sản phẩm mình mong muốn từ những nhà sản xuất khác nhau mà vẫn được mua sản phẩm đó với mức giá thấp. Có lẽ cũng vì được nhiều người tiêu dùng lựa chọn mà doanh thu của Big C Thăng Long luôn ở mức cao và có sự tăng nhẹ trong một số năm gần đây. Trong các năm 2016,

2017, 2018 và 2019, doanh thu của siêu thị lần lượt đạt 2.695,8 tỷ đồng; 2.698,5 tỷ đồng; 3.169,1 tỷ đồng và 3.639,3 tỷ đồng. Tuy mức chênh lệch giữa các năm không quá lớn nhưng doanh thu luôn có sự tăng trưởng qua các năm. Điều này phần nào phản ánh được sự tin tưởng của người dân với việc mua hàng tại siêu thị này.

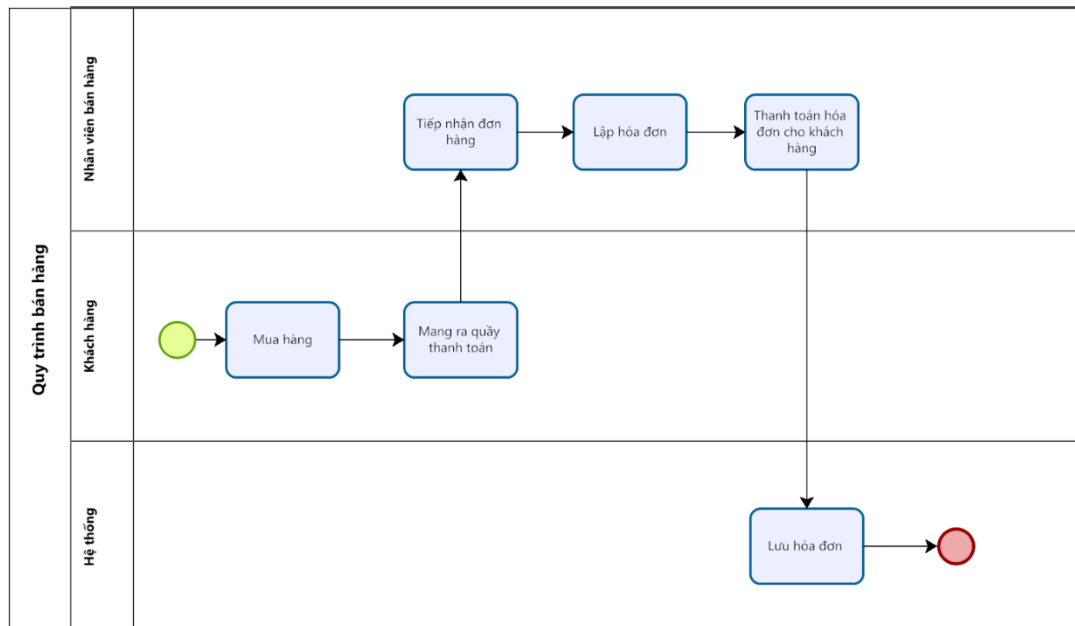
2.2. Cơ cấu tổ chức



Bảng 2-2: Cơ cấu tổ chức Siêu thị Big C Thăng Long

2.3. Mô tả hoạt động nghiệp vụ

2.3.1. Nghiệp vụ bán hàng

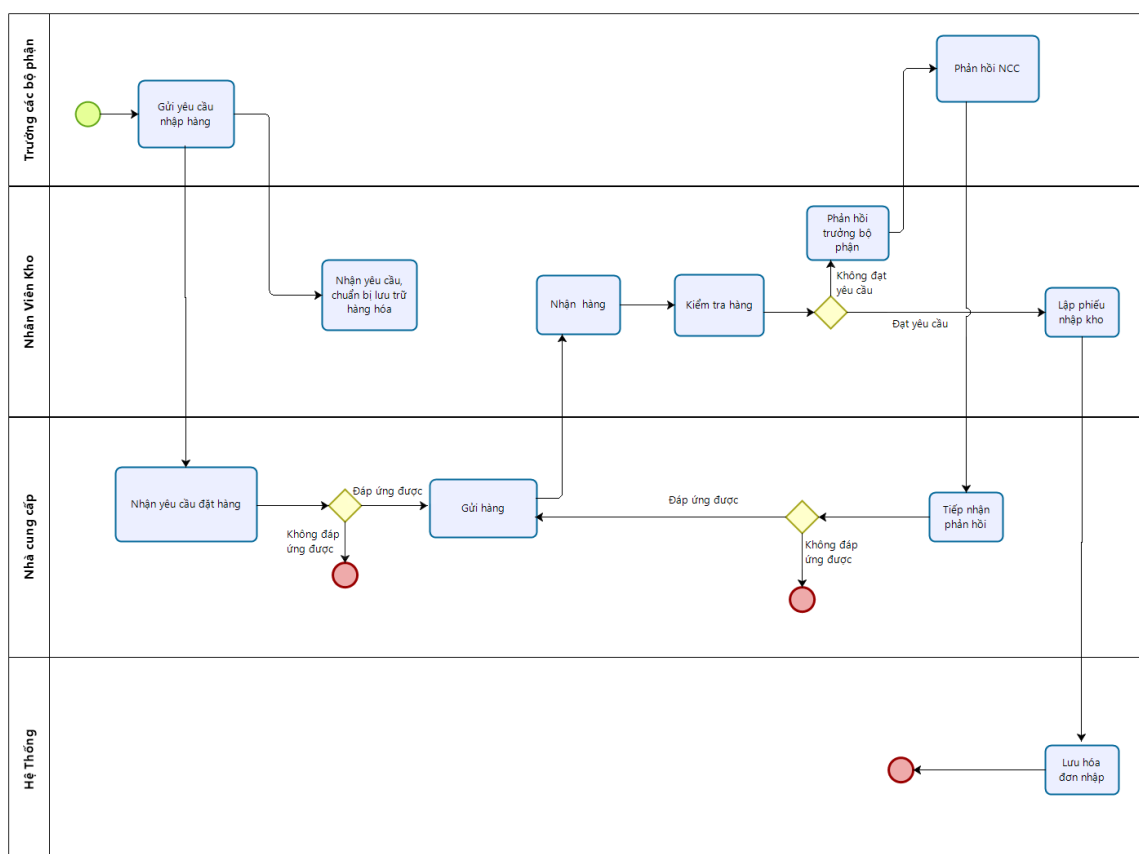


Hình 2-1: Quy trình bán hàng tại siêu thị

Mô tả quy trình bán hàng

Khách hàng sẽ lựa chọn các sản phẩm cần mua tại các quầy hàng sau đó mang ra quầy thanh toán. Tại đó nhân viên bán hàng sẽ tiếp nhận các sản phẩm của đơn hàng, sau đó lập hóa đơn và thanh toán hóa đơn đó cho khách hàng. Đồng thời, sau khi thanh toán xong hệ thống sẽ lưu lại hóa đơn của khách hàng một cách tự động và kết thúc quy trình tại đây.

2.3.2. Nghiệp vụ nhập kho



Hình 2-2: Biểu đồ hoạt động quy trình nhập kho

Mô tả quy trình nhập kho

Bước 1: Trưởng các bộ phận gửi yêu cầu nhập hàng

Bước 2: Nhà cung cấp tiếp nhận yêu cầu đặt hàng

- Nếu đáp ứng được thì chuyển sang bước 3
- Không đáp ứng được thì kết thúc quy trình

Bước 3: Nhà cung cấp tiến hành gửi hàng

Bước 4: Nhân viên kho nhận hàng sau đó kiểm tra hàng

- Nếu đạt yêu cầu thì chuyển sang bước 5
- Không đạt yêu cầu thì chuyển sang bước 7

Bước 5: Nhân viên kho lập phiếu nhập kho

Bước 6: Hóa đơn được lưu vào hệ thống và kết thúc quy trình

Bước 7: Trưởng bộ phận phản hồi NCC

Bước 8: Nhà cung cấp tiếp nhận phản hồi

- Nếu đáp ứng được thì quay lại bước 3
- Không đáp ứng được thì kết thúc quy trình

2.4. Báo cáo tình hình kinh doanh**2.4.1. Người sử dụng (BI Users)**

Chức vụ	Nghịệp vụ	Quản lý	Điểm chưa đạt hiệu quả
CẤP CHIẾN LƯỢC			
Giám đốc	Đưa ra các quyết định chung cấp chiến lược, các quyết định chiến lược chung, mục tiêu phương hướng phát triển.	Toàn bộ các bộ phận gồm 2 khối chính: kinh doanh và dịch vụ.	Thiếu các công cụ có khả năng phân tích, so sánh và xác định xu hướng, cập nhật các dòng dữ liệu theo sát hoạt động.
Phó giám đốc kinh doanh	Đưa ra các quyết định về loại hình kinh doanh cho doanh nghiệp.	Quản lý phòng kinh doanh hàng hóa cho các loại hàng Tiêu dùng nhanh - đồ tươi - dùng trong gia đình - đồ mềm - đồ cứng.	Thiếu các số liệu dự báo loại hình kinh doanh, xu hướng marketing
Phó giám đốc dịch vụ khách hàng	Đưa ra các quyết định, chính sách an toàn sản phẩm-thực phẩm, chăm	Thực hiện chức năng điều hành phòng IT - Sofm - ngân quỹ - trang trí	Thiếu các dữ liệu về khảo sát, tình hình và nhu cầu sử dụng của khách

	sóc khách hàng, quản lý nhân sự và thông tin khách hàng	- khách hàng - thu ngân - kĩ thuật - HQ - an ninh.	hàng; khó đưa ra các chính sách chăm sóc và an toàn sản phẩm-thực phẩm
CẤP CHIẾN THUẬT			
Trưởng bộ phận FMCG (hàng tiêu dùng nhanh)	Kiểm soát nguồn cung cấp hàng tiêu dùng nhanh, tiêu chuẩn chất lượng của các sản phẩm trong kho hàng.	Quản lý các trường quầy Mặn - ngọt - hóa phẩm - mỹ phẩm - đông lạnh - đồ đóng hộp, thịt nguội - đồ uống - bơ sữa.	Đảm bảo rằng sự cung ứng hàng hóa được thực hiện đúng theo yêu cầu và tiêu chuẩn chất lượng của BigC.
Trưởng bộ phận Fresh food (đồ tươi)	Đảm bảo nguồn cung cấp sản phẩm thực phẩm tươi sống đáp ứng đủ và chất lượng. Thiết lập và duy trì mối quan hệ với các nhà cung cấp, đảm bảo quá trình đặt hàng và giao nhận diễn ra đúng tiến độ và đảm bảo chất lượng của hàng hóa.	Quản lý các trường quầy Thịt - hải sản - bánh mì - rau củ quả - đồ ăn sẵn.	Chưa xử lý triệt để phản hồi từ khách hàng và giải quyết các vấn đề phát sinh để đảm bảo sự hài lòng của khách hàng với sản phẩm.
Trưởng bộ phận Home Line (hàng	Kiểm soát và đảm bảo nguồn cung	Quản lý các quầy dụng cụ bếp, đồ	Thiếu các dữ liệu về hiệu năng, công

hóa dùng trong gia đình)	cấp sản phẩm đầy đủ và đáp ứng nhu cầu khách hàng về các dụng cụ bếp, các vật dụng trang trí,... nhu yếu phẩm về hàng hóa trong gia đình	nhựa, vật dụng trang trí, ...	suất của các dụng cụ bếp để khách hàng lựa chọn sản phẩm phù hợp
Trưởng bộ phận Soft Line (hàng hóa mềm)	Kiểm tra và rà soát chặt chẽ về chất lượng sản phẩm, đảm bảo số lượng đáp ứng đủ nhu cầu về mọi mặt.	Quản lý các quầy chăn, ga gối đệm, thời trang (quần, áo, giày dép,...)	Chưa đủ đa dạng mẫu mã để khách hàng có thêm nhiều lựa chọn
Trưởng bộ phận Hard Line (hàng hóa cứng)	Đảm bảo không bị hư hỏng, thường xuyên kiểm tra, cập nhật tình trạng của sản phẩm để kịp thời thay thế bằng sản phẩm khác	Quản lý các quầy đồ nội thất, thiết bị, dụng cụ, điện tử, trang sức và đồ thể thao.	Giá cả chưa hợp lý, có sự chênh lệch nhiều giữa các sản phẩm

Bảng 2-3: Danh sách người sử dụng

2.4.2. Các quyết định của BI Users và dữ liệu cần và phân tích

Chức vụ	Câu hỏi	Thông tin cần có trong báo cáo
Tổng giám đốc	Dữ liệu được trực quan hóa, thể hiện được góc nhìn bao quát	1. Số lượng đơn hàng 2. Doanh thu

	nhất về tình hình phát triển, xu hướng của cả hệ thống, từ đó có thể đưa ra các quyết định tốt nhất về chiến lược.	<p>3. Lợi nhuận</p> <p>4. Thông tin nhà cung cấp</p>
Phó giám đốc kinh doanh	<ul style="list-style-type: none"> - Doanh thu theo thời gian (tháng, quý, năm) của toàn hệ thống. - Dòng sản phẩm nào đang bán chạy tại siêu thị? - Dòng sản phẩm nào mang lại doanh thu cao tại siêu thị? - Top các sản phẩm bán chạy. - Doanh thu theo khu vực. - Nên đẩy mạnh kinh doanh loại mặt hàng hóa nào? - Có nên cập nhật lại xu thế hàng hóa hiện nay? 	<p>1. Số lượt mua mỗi sản phẩm.</p> <p>2. Doanh thu trên mỗi sản phẩm/dòng sản phẩm.</p> <p>3. Doanh thu theo thời gian.</p> <p>4. Doanh thu theo khu vực.</p> <p>5. Số lượng hàng hóa được bán ra theo từng nhóm sản phẩm.</p> <p>6. Số lượng hàng hóa được bán ra theo các nhà cung cấp.</p>
Phó tổng giám đốc dịch vụ khách hàng	<ul style="list-style-type: none"> - Lượng khách hàng tại mỗi khu vực. - Đây là những khách hàng mang lại nhiều doanh thu? 	<p>1. Thống kê số lượng khách hàng của siêu thị.</p> <p>2. Những khách hàng đem lại nhiều doanh thu nhất.</p> <p>3. Nhu cầu của khách hàng với các loại mặt hàng.</p>
Trưởng bộ phận FMCG (hàng hóa tiêu dùng nhanh)	<ul style="list-style-type: none"> - Hàng hóa nào được ưa chuộng nhất? - Hàng hóa kinh doanh nào 	<p>1. Số lượng mỗi sản phẩm được bán ra</p> <p>2. Doanh thu, lợi nhuận</p>

Trưởng bộ phận Fresh food (đồ tươi)	<p>đem lại doanh thu cao nhất?</p> <ul style="list-style-type: none"> - Nhu cầu về sự đa dạng các mẫu mã của khách hàng về hàng hóa hiện nay như thế nào? - Cần xem xét lại giá cả hàng hóa không? - Các chính sách ưu đãi để đảm bảo sự hài lòng của khách hàng? 	<p>trên mỗi sản phẩm</p> <p>3. Các sản phẩm bán chạy</p> <p>4. Số lượng hàng nhập/ bán từ mỗi nhà cung cấp</p> <p>5. Số lượng hàng tồn kho</p>
Trưởng bộ phận Home Line (hàng hóa dùng trong gia đình)		
Trưởng bộ phận Soft Line (hàng hóa mềm)		
Trưởng bộ phận Hard Line (hàng hóa cứng)		
Thủ kho	<ul style="list-style-type: none"> - Hàng hóa của nhà cung cấp nào có doanh số lớn nhất? - Số lượng hàng hóa nhập vào có phù hợp với số lượng hàng bán ra hay không? 	<p>1. Số lượt/ giá trị đặt mua các mặt hàng</p> <p>2. Số lượt/ giá trị đặt mua từ các nhà cung cấp</p> <p>3. Số lượng bán các mặt hàng</p> <p>4. Số lượng tồn kho của mỗi mặt hàng.</p>

Bảng 2-4: Các quyết định của người sử dụng và dữ liệu cần cung cấp

CHƯƠNG 3: XÂY DỰNG KHO DỮ LIỆU CHO DOANH NGHIỆP

3.1. Thiết kế cơ sở dữ liệu

3.1.1. Thiết kế các bảng Dim

Bảng Dim_LoaiHang		
Tên cột	Kiểu dữ liệu	Ý nghĩa
<u>MaTheLoai</u>	char(10)	Mã thẻ loại
TenTheLoai	nvarchar(50)	Tên thẻ loại
MoTa	nvarchar(100)	Mô tả loại hàng

Bảng 3-1: Thiết kế bảng Dim_LoaiHang

Bảng Dim_KhachHang		
Tên cột	Kiểu dữ liệu	Ý nghĩa
<u>MaKhachHang</u>	char(10)	Mã khách hàng
TenKhachHang	nvarchar(50)	Tên khách hàng
DiaChi	nvarchar(50)	Địa chỉ
SDT	char(30)	Số điện thoại

Bảng 3-2: Thiết kế bảng Dim_KhachHang

Bảng Dim_NhanVien		
Tên cột	Kiểu dữ liệu	Ý nghĩa
<u>MaNhanVien</u>	char(10)	Mã nhân viên
TenNhanVien	nvarchar(50)	Tên nhân viên
ChucVu	nvarchar(50)	Chức vụ
GioiTinh	nvarchar(10)	Giới tính
NgaySinh	date	Ngày sinh
NgayBatDauLam	date	Ngày bắt đầu làm việc

DiaChi	nvarchar(10)	Địa chỉ
SDT	char(30)	Số điện thoại

Bảng 3-3: Thiết kế bảng Dim_NhanVien

Bảng Dim_SanPham		
Tên cột	Kiểu dữ liệu	Ý nghĩa
<u>MaSanPham</u>	char(10)	Mã sản phẩm
TenSanPham	nvarchar(50)	Tên sản phẩm
MaNCC	char(10)	Mã nhà cung cấp
MaTheLoai	char(10)	Mã loại hàng
SoLuongTheoDV	char(50)	Số lượng theo đơn vị
DonGia	float	Đơn giá
TonKho	int	Tồn kho
ReOrder	int	Đặt lại
TrangThai	nvarchar(10)	Trạng thái

Bảng 3-4: Thiết kế bảng Dim_SanPham

Bảng Dim_NCC		
Tên cột	Kiểu dữ liệu	Ý nghĩa
<u>MaNCC</u>	char(10)	Mã nhà cung cấp
TenNCC	nvarchar(50)	Tên nhà cung cấp
ContactName	nvarchar(50)	Tên người liên hệ
ChucDanh	nvarchar(50)	Chức danh
DiaChi	nvarchar(50)	Địa chỉ
SDT	char(30)	Số điện thoại

Bảng 3-5: Thiết kế bảng Dim_NCC

3.1.2. Thiết kế các bảng Fact

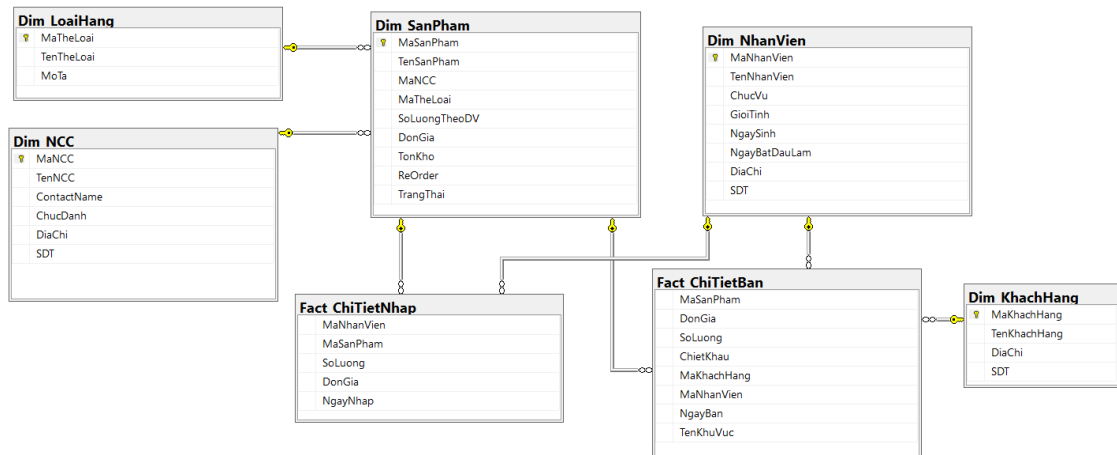
Bảng Fact_ChiTietNhap		
Tên cột	Kiểu dữ liệu	Ý nghĩa
MaNhanVien	char(10)	Mã nhân viên
NgayNhan	date	Ngày nhận hàng
MaSanPham	char(10)	Mã sản phẩm
SoLuong	int	Số lượng hàng nhập
DonGia	float	Đơn giá nhập

Bảng 3-6: Thiết kế bảng Fact_ChiTietNhap

Bảng Fact_ChiTietBan		
Tên cột	Kiểu dữ liệu	Ý nghĩa
MaKhachHang	char(10)	Mã khách hàng
MaNhanVien	char(10)	Mã nhân viên
NgayBan	date	Ngày bán hàng
MaSanPham	char(10)	Mã sản phẩm
SoLuong	int	Số lượng hàng bán
DonGia	float	Đơn giá bán
ChietKhau	float	Chiết khấu
TenKhuVuc	char(50)	Khu vực

Bảng 3-7: Thiết kế bảng Fact_ChiTietBan

3.2. Xây dựng lược đồ dữ liệu



Hình 3-1: Lược đồ dữ liệu

3.3. Công cụ triển khai kho dữ liệu

3.3.1. Microsoft SQL Server

Microsoft SQL Server là một hệ quản trị cơ sở dữ liệu quan hệ được phát triển bởi Microsoft. Đây là một máy chủ cơ sở dữ liệu, là một phần mềm có chức năng chính là lưu trữ và truy xuất dữ liệu theo yêu cầu của các ứng dụng phần mềm khác.

Một số ưu điểm

- Tích hợp các ứng dụng Microsoft: SQL Server hoàn thiện tích hợp với các sản phẩm và dịch vụ khác giúp dàng tạo và quản lý ứng dụng dựa trên dữ liệu.
- Quản lý dữ liệu và ETL (Trích xuất, Chuyển đổi, Tải): SQL Server cung cấp dịch vụ tích hợp quản lý dữ liệu và ETL (Dịch vụ tích hợp máy chủ SQL – SSIS), giúp trích xuất, biến đổi và tải dữ liệu một cách hiệu quả.
- Quản lý cơ sở dữ liệu dịch vụ: SQL Server có một loạt cơ sở dữ liệu quản lý dịch vụ, bao gồm SQL Server Management Studio (SSMS) để quản lý và cơ sở dữ liệu tối ưu hóa.
- Truy vấn và xử lý dữ liệu mạnh mẽ: SQL Server cung cấp khả năng thực hiện truy vấn phức tạp và xử lý dữ liệu bằng ngôn ngữ truy vấn T-SQL.

3.3.2. Microsoft Visual Studio

Microsoft Visual Studio là một phần mềm hỗ trợ đắc lực hỗ trợ công việc lập

trình. Công cụ này được tạo lên và thuộc quyền sở hữu của ông lớn công nghệ Microsoft. Đây là hệ thống tập hợp tất cả những gì liên quan tới phát triển ứng dụng, bao gồm trình chỉnh sửa mã, trình thiết kế, gỡ lỗi.

Một số ưu điểm

- Visual Studio là môi trường phát triển hợp pháp phát triển (IDE) cho nhiều ứng dụng và lập trình ngôn ngữ, không chỉ trong lĩnh vực dữ liệu mà còn cả các loại ứng dụng khác.
- Visual Studio cho phép tổ chức và quản lý nhiều dự án khác nhau trong một giải pháp. Điều này sẽ hữu ích khi cần phát triển nhiều thành phần ứng dụng cùng lúc.
- Visual Studio cung cấp các công cụ mạnh mẽ để kiểm tra và gỡ lỗi mã nguồn và ứng dụng.

3.3.3. SQL Server Integration Services (SSIS)

SQL Server Integration Services là dịch vụ tích hợp máy chủ SQL của Microsoft được dùng để quản lý và thực hiện các quá trình tích hợp dữ liệu trong môi trường SQL Server và hệ thống dữ liệu khác. SSIS được dùng để trích xuất dữ liệu từ các nguồn khác nhau, biến đổi dữ liệu, nạp dữ liệu vào cơ sở dữ liệu SQL Server hoặc hệ thống lưu trữ dữ liệu khác, là công cụ lưu trữ dữ liệu nhanh chóng và linh hoạt.

Một số ưu điểm

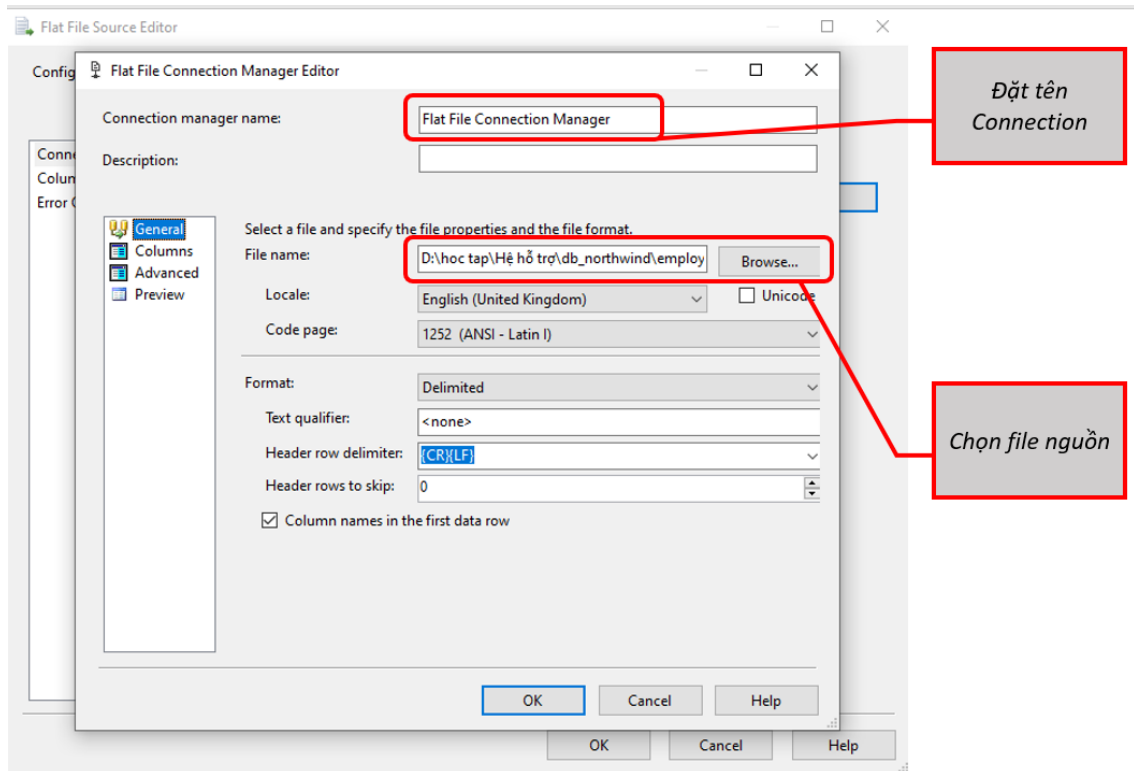
- SSIS cung cấp giao diện trực quan, cho phép kéo và giải phóng ETL hoạt động để xây dựng quy trình ETL một cách dễ dàng.
- SSIS được tích hợp mạnh mẽ với SQL Server, cho phép trích xuất và cập nhật dữ liệu trực tiếp từ cơ sở dữ liệu SQL Server.
- SSIS cho phép lưu các gói ETL và quy trình dữ liệu trong thư mục thư viện, dễ dàng quản lý và phát triển chúng.
- Người dùng có thể lên lịch chạy quy trình ETL theo khoảng thời gian cố định hoặc theo nhu cầu.

3.4. Đưa dữ liệu vào kho dữ liệu

Với nguồn dữ liệu là các file csv, ta thực hiện đổ dữ liệu vào kho dữ liệu qua các

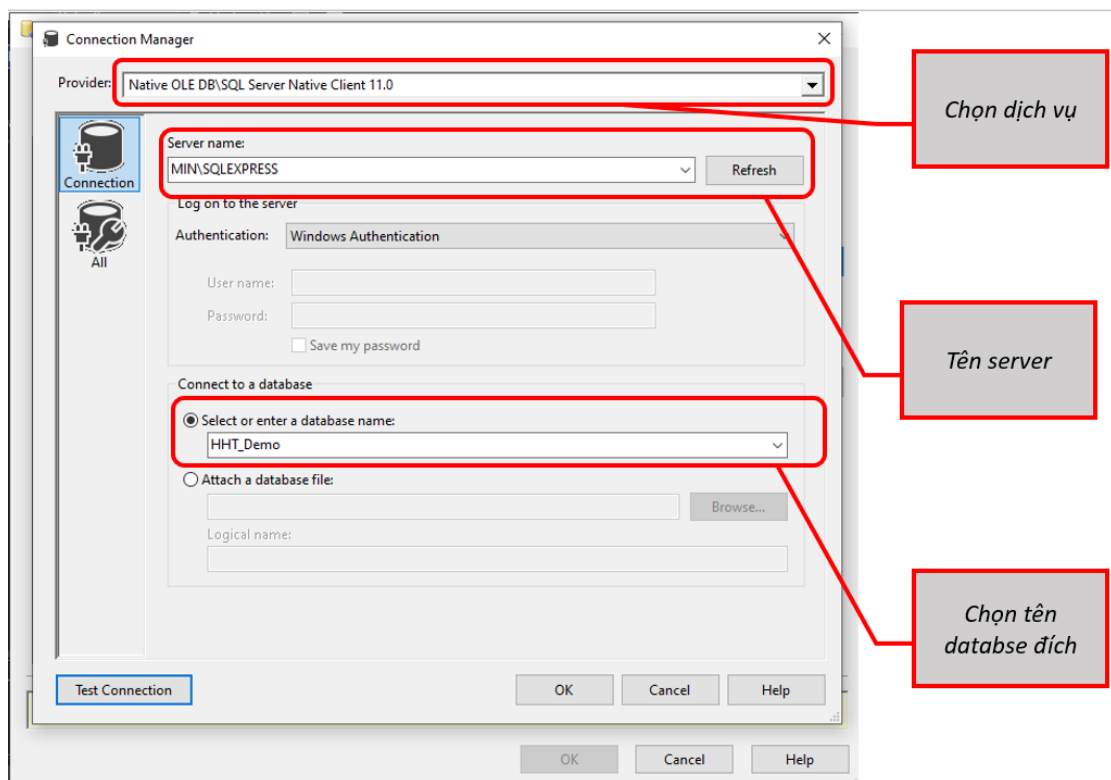
bước sau:

- **Bước 1:** Tạo đầu vào với file csv: Chọn đối tượng Flat File Source trong Toolbox và thiết lập các thông số như hình dưới:



Hình 3-2: Tạo connection với đầu vào là file csv

- **Bước 2:** Tạo connect với database trong Microsoft SQL Server




Hình 3-3: Tạo connect với đầu ra là database trong Microsoft SQL Server

- **Bước 3:** Đổ dữ liệu cho từng bảng

3.4.1. Bảng Dim_NhanVien

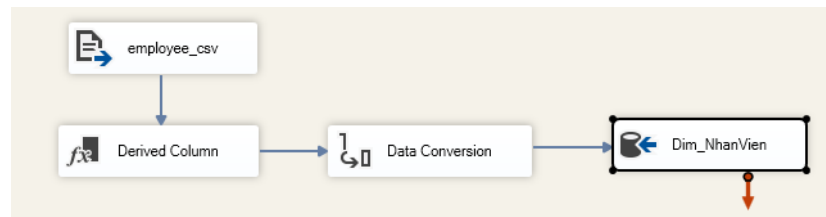
Bảng Dim_NhanVien mô tả các thông tin của nhân viên cần lưu trữ là Mã nhân viên, Tên nhân viên, chức vụ, giới tính, ngày sinh, ngày bắt đầu làm, địa chỉ, số điện thoại nhằm thuận tiện trong việc quản lý nhân viên.

- Cấu trúc bảng Dim_NhanVien như sau:

Dim_NhanVien	
	MaNhanVien
	TenNhanVien
	ChucVu
	GioiTinh
	NgaySinh
	NgayBatDauLam
	DiaChi
	SDT

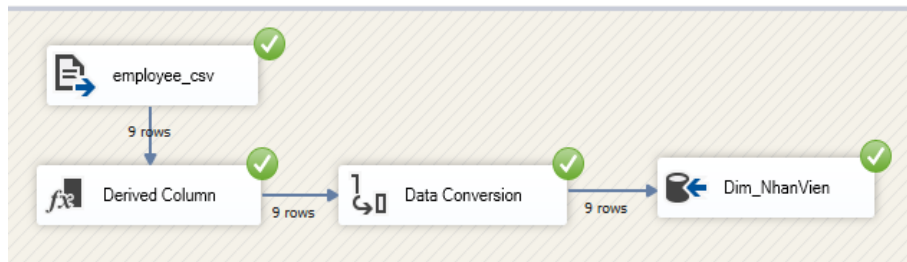
Hình 3-4: Cấu trúc bảng Dim_NhanVien

- Data flow của Dim_NhanVien



Hình 3-5: Data flow của Dim_NhanVien

- Kết quả chạy thành công trên SSDT



Hình 3-6: Kết quả đổ dữ liệu vào bảng Dim_NhanVien trên SSDT

- Import thành công dữ liệu bảng Dim_NhanVien

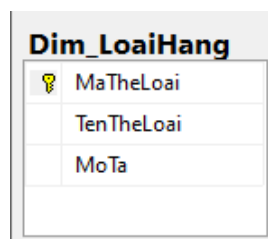
	MaNhanVien	TenNhanVien	ChucVu	GioiTinh	NgaySinh	NgayBatDauLam	DiaChi	SDT
1	1	Davolio Nancy	Sales Representative	Nu	1993-10-16	2013-06-25	507 20th Ave. E. Apt. 2A	(206) 555-9857
2	2	Fuller Andrew	Vice President Sales	Nam.	1994-08-15	2013-06-16	908 W. Capital Way	(206) 555-9482
3	3	Leverling Janet	Sales Representative	Nu	1995-02-11	2014-02-18	722 Moss Bay Blvd.	(206) 555-3412
4	4	Peacock Margaret	Sales Representative	Nu	1997-08-15	2013-09-11	4110 Old Redmond Rd.	(206) 555-8122
5	5	Buchanan Steven	Sales Manager	Nam	1993-01-27	2013-09-28	14 Garrett Hill	(71) 555-4848
6	6	Suyama Michael	Sales Representative	Nam	1996-07-22	2013-01-15	Coventry House Miner Rd.	(71) 555-7773
7	7	King Robert	Sales Representative	Nam	1990-08-21	2014-07-22	Edgeham Hollow Winchester Way	(71) 555-5598
8	8	Callahan Laura	Inside Sales Coordinator	Nu	1990-12-20	2015-05-12	4726 11th Ave. N.E.	(206) 555-1189
9	9	Dodsworth Anne	Sales Representative	Nu	1993-01-05	2015-09-10	7 Houndstooth Rd.	(71) 555-4444

Hình 3-7: Dữ liệu trên bảng Dim_NhanVien

3.4.2. Bảng Dim_LoaiHang

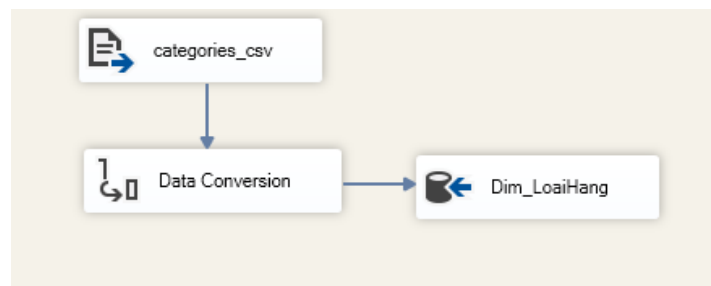
Bảng Dim_LoaiHang mô tả các thông tin của các loại hàng cần lưu trữ như Mã thẻ loại, Tên thẻ loại và Mô tả.

- Cấu trúc bảng Dim_LoaiHang như sau:



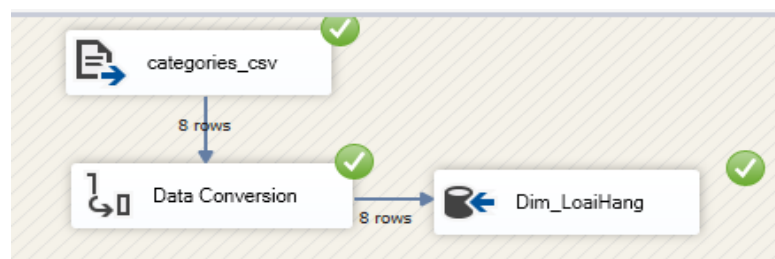
Hình 3-8: Cấu trúc bảng Dim_LoaiHang

- Data flow của Dim_LoaiHang:



Hình 3-9: Data flow của Dim_LoaiHang

- Kết quả chạy thành công trên SSDT



Hình 3-10: Kết quả đổ dữ liệu vào bảng Dim_LoaiHang trên SSDT

- Import thành công dữ liệu bảng Dim_LoaiHang


	MaTheLoai	TenTheLoai	MoTa
1	1	Beverages	Soft drinks coffees teas beers and ales
2	2	Condiments	Sweet and savory sauces relishes spreads
3	3	Confections	Desserts candies and sweet breads
4	4	Dairy Products	Cheeses
5	5	Grains Cereals	Breads crackers pasta and cereal
6	6	Meat Poultry	Prepared meats
7	7	Produce	Dried fruit and bean curd
8	8	Seafood	Seaweed and fish

Hình 3-11: Dữ liệu trên bảng Dim_LoaiHang

3.4.3. Bảng Dim_NCC

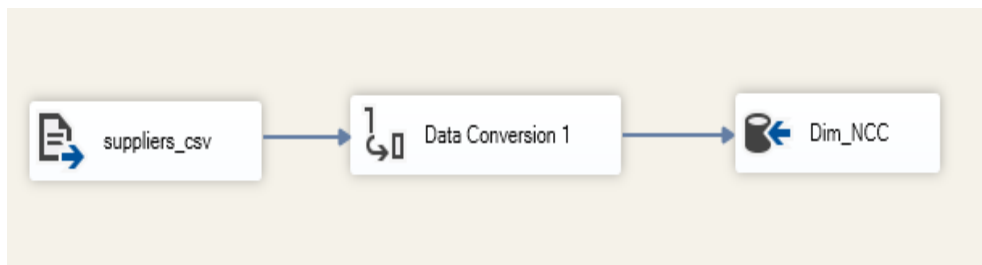
Bảng Dim_NCC mô tả các thông tin của nhà cung cấp được lưu trữ như mã nhà cung cấp, tên nhà cung cấp, contact name, chức danh, địa chỉ, số điện thoại.

- Cấu trúc bảng Dim_NCC như sau:

Dim_NCC	
	MaNCC
	TenNCC
	ContactName
	ChucDanh
	DiaChi
	SDT

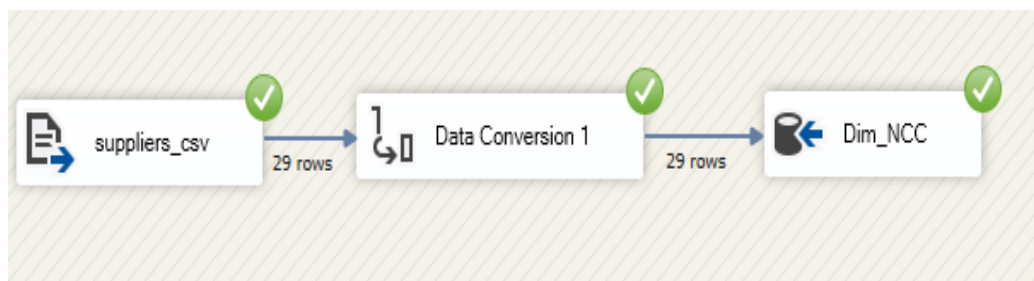
Hình 3-12: Cấu trúc bảng Dim_NCC

- Data flow của Dim_NCC



Hình 3-13: Data flow của Dim_NCC

- Kết quả chạy thành công trên SSDT



Hình 3-14: Kết quả đổ dữ liệu vào bảng Dim_NCC trên SSDT

- Import thành công dữ liệu bảng Dim_NCC


	MaNCC	TenNCC	ContactName	ChucDanh	DiaChi	SDT
1	1	Exotic Liquids	Charlotte Cooper	Purchasing Manager	49 Gilbert St.	(171) 555-2222
2	10	Refrescos Americanas LTDA	Carlos Diaz	Marketing Manager	Av. das Americanas 12.890	(11) 555 4640
3	11	Heil SÄÄYwaren GmbH & Co. KG	Petra Winkler	Sales Manager	Tiergartenstraße 5	(010) 9984510
4	12	Plutzer LebensmittelgroÄYmärkte AG	Martin Bein	International Marketing Mgr.	Bogenallee 51	(069) 992755
5	13	Nord-Ost-Fisch Handelsgesellschaft mbH	Sven Petersen	Coordinator Foreign Markets	Frahmredder 112a	(04721) 8713
6	14	Fornaggi Fortini s.r.l.	Elio Rossi	Sales Representative	Viale Dante 75	(0544) 60323
7	15	Norske Meierier	Beate Vileid	Marketing Manager	Hatlevegen 5	(0)2-953010
8	16	Bigfoot Breweries	Cheryl Saylor	Regional Account Rep.	3400 - 8th Avenue Suite 210	(503) 555-9931
9	17	Svensk SÄÄtÅda AB	Michael BjÄÄrm	Sales Representative	BrovallavÄngen 231	08-123 45 67
10	18	Aux joyeux ecclÄsiastiques	GuyLÄne Nodier	Sales Manager	203 Rue des Francs-Bourgeois	(1) 03.83.00.68
11	19	New England Seafood Cannery	Robb Merchant	Wholesale Account Agent	Order Processing Dept. 2100 Paul Revere Blvd.	(617) 555-3267
12	2	New Orleans Cajun Delights	Shelley Burke	Order Administrator	P.O. Box 78934	(100) 555-4822
13	20	Leka Trading	Chandra Leka	Owner	471 Serangoon Loop Suite #402	555-8787
14	21	Lyngbysild	Niels Petersen	Sales Manager	Lyngbysild Fiskebakken 10	(12345) 1212
15	22	Zaanse Snoepfabriek	Dirk Luchte	Accounting Manager	Verkoop Rijnweg 22	(12345) 1212
16	23	Karkki Oy	Anne Heikonen	Product Manager	Valtakatu 12	(953) 10956
17	24	G'day Mate	Wendy Mackenzie	Sales Representative	170 Prince Edward Parade Hunter's Hill	(02) 555-5914
18	25	Ma Maison	Jean-Guy Lauzon	Marketing Manager	2960 Rue St. Laurent	(514) 555-9022
19	26	Pasta Buttini s.r.l.	Giovanni Giudici	Order Administrator	Via dei Gelsomini 153	(089) 6547665
20	27	Escargots Nouveaux	Marie Delamare	Sales Manager	22 rue H. Voiron	85.57.00.07
21	28	Gai pÄcturage	Eliane Noz	Sales Representative	Bat. B 3 rue des Alpes	38.76.98.06
22	29	ForÄts dÄrÅrables	Chantal Goulet	Accounting Manager	148 rue Chasseur	(514) 555-2955
23	3	Grandma Kelly's Homestead	Regina Murphy	Sales Representative	707 Oxford Rd.	(313) 555-5735
24	4	Tokyo Traders	Yoshi Nagase	Marketing Manager	9-8 Sekimai Musashino-shi	(03) 3555-5011
25	5	Compania de Queros "Los Caberos"	Antonio del Valle	Export Administrator	Calle del Canal 4	(98) 588 76 54

Hình 3-15: Dữ liệu trên bảng Dim_NCC

3.4.4. Bảng Dim_SanPham

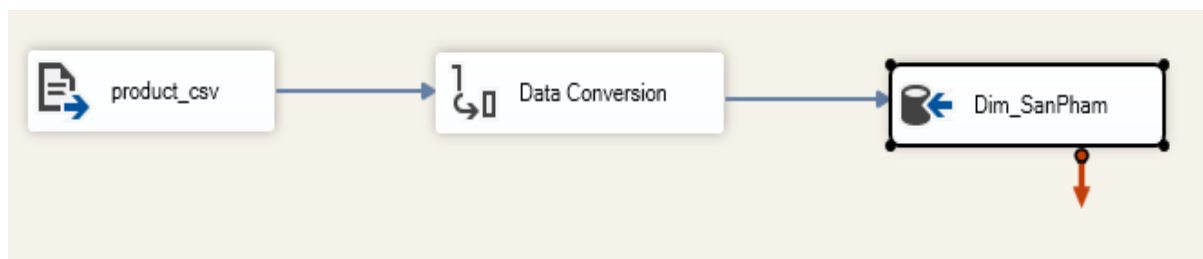
Bảng Dim_SanPham mô tả các thông tin của các sản phẩm cần lưu trữ như Mã sản phẩm, Tên sản phẩm, Mã nhà cung cấp, Mã thể loại, Số lượng theo đơn vị, Đơn giá, Tồn kho, ReOrder, Trạng thái sản phẩm.

- Cấu trúc bảng Dim_SanPham như sau:

Dim_SanPham	
	MaSanPham
	TenSanPham
	MaNCC
	MaTheLoai
	SoLuongTheoDV
	DonGia
	TonKho
	ReOrder
	TrangThai

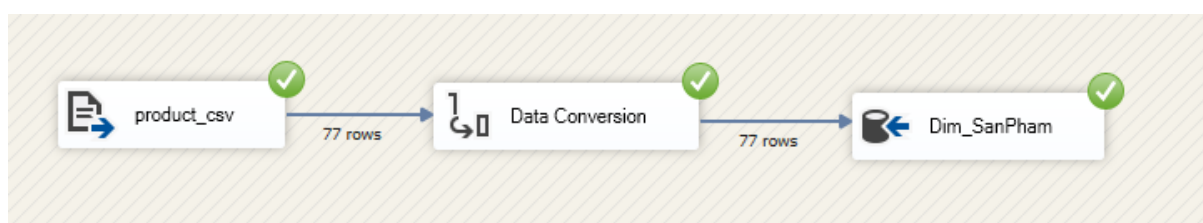
Hình 3-16: Cấu trúc bảng Dim_SanPham

- Data flow của Dim_SanPham:



Hình 3-17: Cấu trúc bảng Dim_SanPham

- Kết quả chạy thành công trên SSDT:



Hình 3-18: Kết quả đổ dữ liệu vào bảng Dim_SanPham trên SSDT

- Import thành công dữ liệu bảng Dim_SanPham

	MaSanPham	TenSanPham	MaNCC	MaTheLoai	SoLuongTheoDV	DonGia	TonKho	ReOrder	TrangThai
1	1	Chai	1	1	10 boxes x 20 bags	18	39	10	0
2	10	Ikura	4	8	12 - 200 ml jars	31	31	0	0
3	11	Queso Cabrales	5	4	1 kg pkg.	21	22	30	0
4	12	Queso Manchego La Pastora	5	4	10 - 500 g pkgs.	38	86	0	0
5	13	Konbu	6	8	2 kg box	6	24	5	0
6	14	Tofu	6	7	40 - 100 g pkgs.	23	35	0	0
7	15	Genen Shouyu	6	2	24 - 250 ml bottles	16	39	5	0
8	16	Pavlova	7	3	32 - 500 g boxes	17	29	10	0
9	17	Alice Mutton	7	6	20 - 1 kg tins	39	0	0	1
10	18	Camarvon Tigers	7	8	16 kg pkg.	62	42	0	0
11	19	Teatime Chocolate Biscuits	8	3	10 boxes x 12 pieces	9	25	5	0
12	2	Chang	1	1	24 - 12 oz bottles	19	17	25	0
13	20	Sir Rodney's Marmalade	8	3	30 gift boxes	81	40	0	0
14	21	Sir Rodney's Scones	8	3	24 pkgs. x 4 pieces	10	3	5	0
15	22	Gustaf's Knäckebröd	9	5	24 - 500 g pkgs.	21	104	25	0
16	23	Tunnbröd	9	5	12 - 250 g pkgs.	9	61	25	0
17	24	Guaraní Fantástica	10	1	12 - 355 ml cans	4	20	0	1
18	25	NuNuCa NuY-Nougat-Cre...	11	3	20 - 450 g glasses	14	76	30	0
19	26	Gumbi Gummibärchen	11	3	100 - 250 g bags	31	15	0	0
20	27	Schoggi Schokolade	11	3	100 - 100 g pieces	44	49	30	0
21	28	Rössle Sauerkraut	12	7	25 - 825 g cans	46	26	0	1
22	29	Thüringer Rostbratwurst	12	6	50 bags x 30 sausgs.	124	0	0	1
23	3	Aniseed Syrup	1	2	12 - 550 ml bottles	10	13	25	0
24	30	Nord-Ost Matjeshering	13	8	10 - 200 g glasses	26	10	15	0
25	31	Gorgonzola Telino	14	4	12 - 100 g pkgs.	12	0	20	0

Hình 3-19: Dữ liệu trên bảng Dim_SanPham

3.4.5. Bảng Dim_KhachHang

Bảng Dim_KhachHang mô tả các thông tin của khách hàng cần lưu trữ như Mã

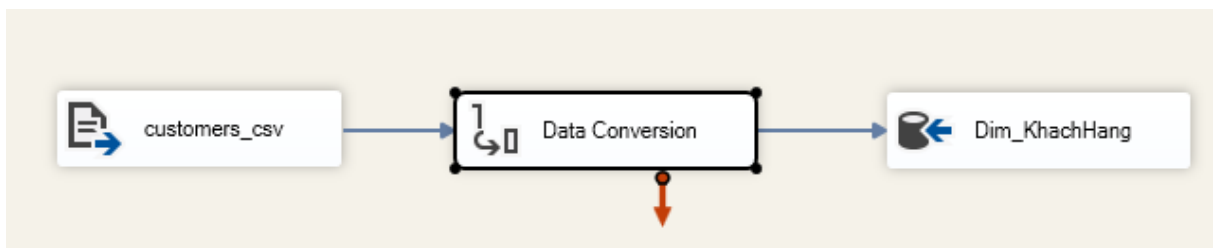
Khách hàng, Tên Khách Hàng, Địa Chỉ, Số điện thoại.

- Cấu trúc bảng Dim_KhachHang như sau:

Dim_KhachHang	
	MaKhachHang
	TenKhachHang
	DiaChi
	SDT

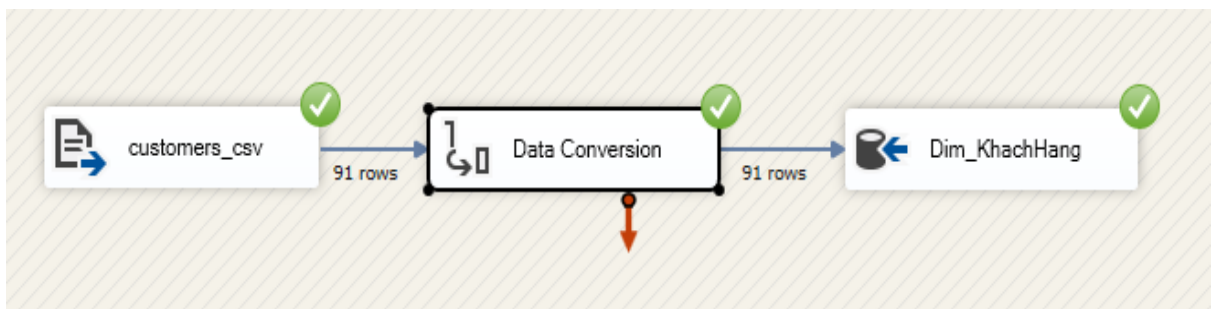
Hình 3-20: Cấu trúc bảng Dim_KhachHang

- Data flow của Dim_KhachHang:



Hình 3-21: Data flow của Dim_KhachHang

- Kết quả chạy thành công trên SSDT:



Hình 3-22: Kết quả đổ dữ liệu vào bảng Dim_KhachHang trên SSDT

- Import thành công dữ liệu bảng Dim_KhachHang

	MaKhachHang	TenKhachHang	DiaChi	SDT
1	ALFKI	Maria Anders	Obere Str. 57	030-0074321
2	ANATR	Ana Trujillo	Avda. de la Constitución 2222	(5) 555-4729
3	ANTON	Antonio Moreno	Mataderos 2312	(5) 555-3932
4	AROUT	Thomas Hardy	120 Hanover Sq.	(171) 555-7788
5	BERGS	Christina Berglund	Berguvsvägen 8	0921-12 34 65
6	BLAUS	Hanna Moos	Forsterstr. 57	0621-08460
7	BLONP	Frédérique Citeaux	24 place Kléber	88.60.15.31
8	BOLID	Martin Sommer	67C Araquil	(91) 555 22 82
9	BONAP	Laurence Leblanc	12 rue des Bouchers	91.24.45.40
10	BOTTM	Elizabeth Lincoln	23 Tsawassen Blvd.	(604) 555-4729
11	BSBEV	Victoria Ashworth	Fauntleroy Circus	(171) 555-1212
12	CACTU	Patricio Simpson	Cemito 333	(1) 135-5555
13	CENTC	Francisco Chang	Sierras de Granada 9993	(5) 555-3392
14	CHOPS	Yang Wang	Hauptstr. 29	0452-076545
15	COMMI	Pedro Afonso	23 Av. dos Lusíadas	(11) 555-7647
16	CONSH	Elizabeth Brown	Berkeley Gardens 12 Brewery	(171) 555-2282
17	DRACD	Sven Ottlieb	Walserweg 21	0241-039123
18	DUMON	Janine Labrune	67 rue des Cinquante Otages	40.67.88.88
19	EASTC	Ann Devon	35 King George	(171) 555-0297
20	ERNSH	Roland Mendel	Kirchgasse 6	7675-3425

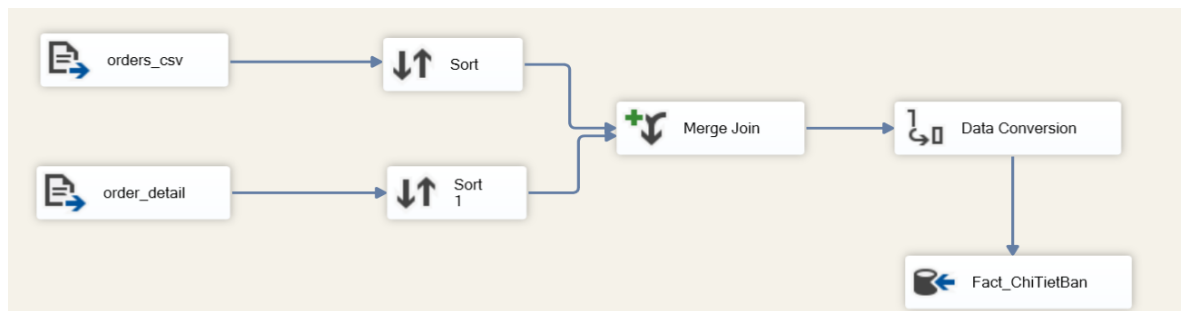
Hình 3-23: Dữ liệu trên bảng Dim_KhachHang

3.4.6. Bảng Fact_ChiTietBan

Fact_ChiTietBan	
	MaSanPham
	DonGia
	SoLuong
	ChietKhau
	MaKhachHang
	MaNhanVien
	NgayBan
	TenKhuVuc

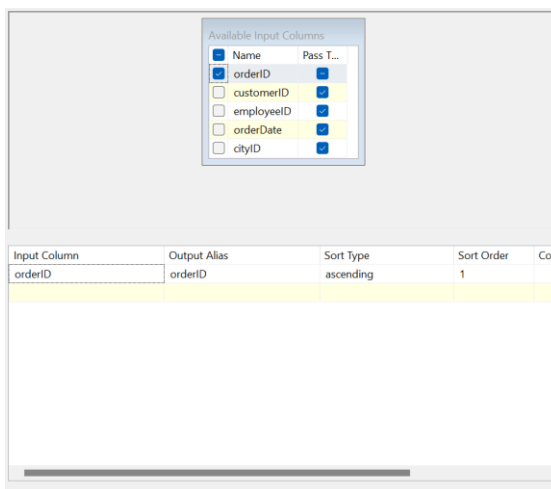
Hình 3-24: Cấu trúc bảng Fact_ChiTietBan

- Nguồn dữ liệu:
 - Bảng Hóa Đơn Bán
 - Bảng Chi Tiết Hóa Đơn Bán
- Data flow của Fact_ChiTietBan

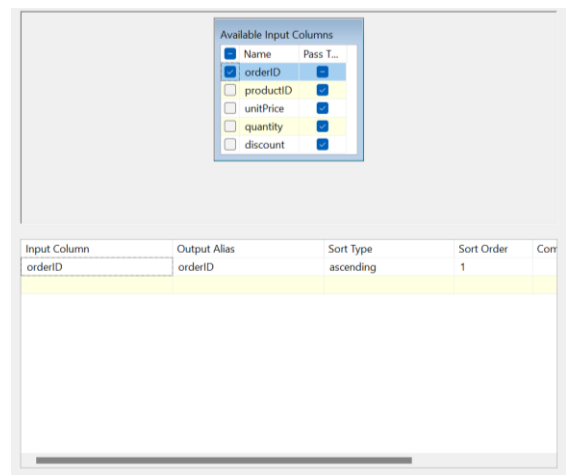


Hình 3-25: Data flow của Fact_ChiTietBan

- Sử dụng 02 đối tượng sort để sắp xếp hai bảng theo cùng mã hóa đơn.



Hình 3-26: Sort bảng Hóa Đơn Bán



Hình 3-27: Sort bảng Chi Tiết Hóa Đơn Bán

- Chèn tất cả các cột của hai bảng bằng phương thức Full Outer Join trong đối tượng Merge Join.

Join type: Full outer join Swap Inputs

Sort

<input checked="" type="checkbox"/>	Name	Or...	Joi...
<input checked="" type="checkbox"/>	orderID	1	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	customerID	0	<input type="checkbox"/>
<input checked="" type="checkbox"/>	employeeID	0	<input type="checkbox"/>
<input checked="" type="checkbox"/>	orderDate	0	<input type="checkbox"/>
<input checked="" type="checkbox"/>	cityID	0	<input type="checkbox"/>

Sort 1

<input checked="" type="checkbox"/>	Name	Or...	Joi...
<input checked="" type="checkbox"/>	orderID	1	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	productID	0	<input type="checkbox"/>
<input checked="" type="checkbox"/>	unitPrice	0	<input type="checkbox"/>
<input checked="" type="checkbox"/>	quantity	0	<input type="checkbox"/>
<input checked="" type="checkbox"/>	discount	0	<input type="checkbox"/>

Input	Input Column	Output Alias
Sort	orderID	orderID
Sort	customerID	customerID
Sort	employeeID	employeeID
Sort	orderDate	orderDate
Sort	cityID	cityID
Sort 1	orderID	orderID (1)
Sort 1	productID	productID
Sort 1	unitPrice	unitPrice
Sort 1	quantity	quantity
Sort 1	discount	discount

Hình 3-28: Ghép bảng Hóa Đơn Bán và Chi Tiết Hóa Đơn Bán bằng cột orderID

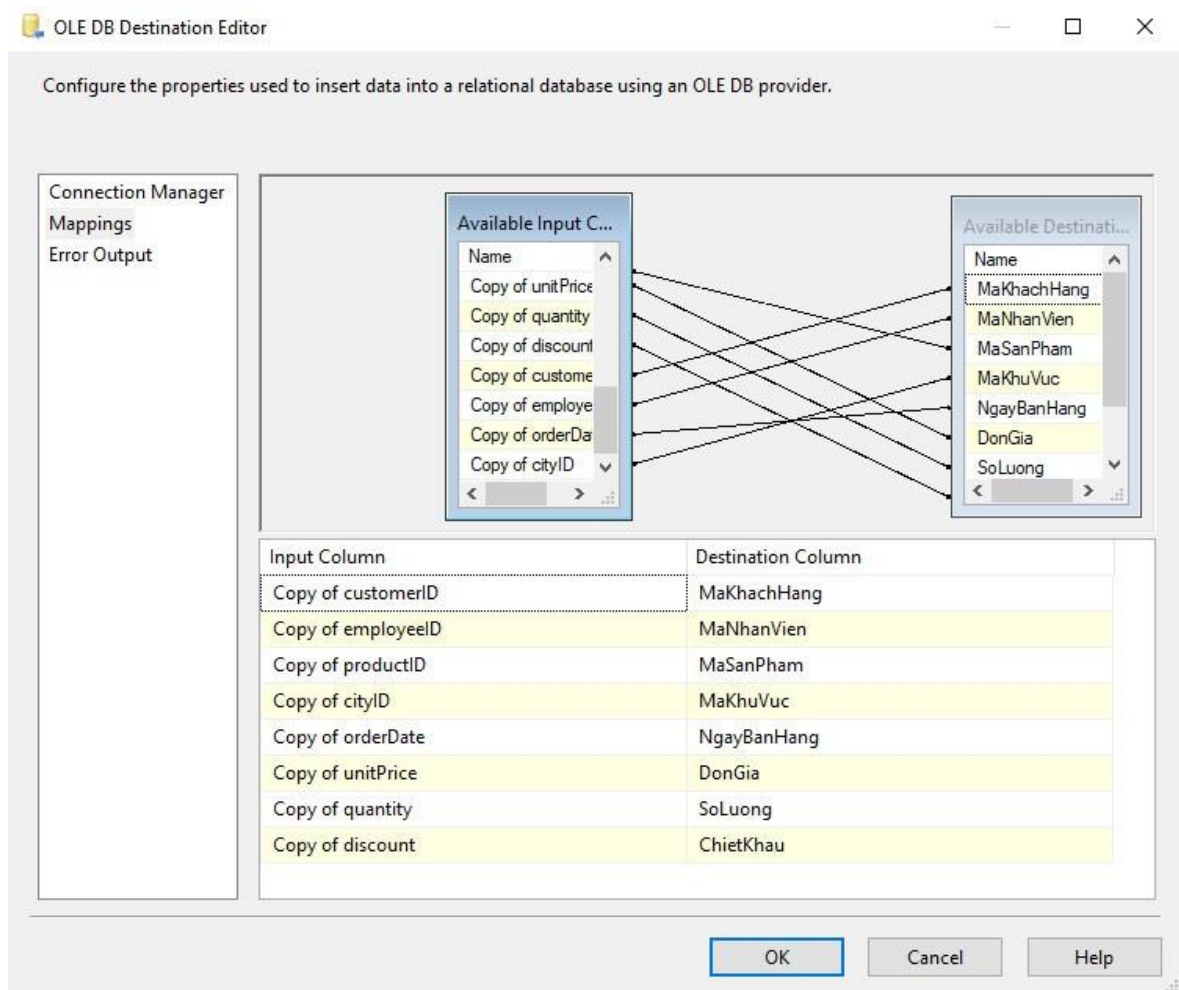
- Sử dụng Data Conversion để chuyển đổi kiểu dữ liệu và độ dài dữ liệu cho khớp với định dạng bảng đích.

<div> <div>Available Input Col...</div> <div> <input checked="" type="checkbox"/> Name <input checked="" type="checkbox"/> orderID <input checked="" type="checkbox"/> customerID <input checked="" type="checkbox"/> employeeID <input checked="" type="checkbox"/> orderDate <input checked="" type="checkbox"/> cityID <input checked="" type="checkbox"/> orderID (1) <input checked="" type="checkbox"/> productID </div> </div>						
Input Column	Output Alias	Data Type	Length	Precision	Scale	Code Page
orderID	Copy of orderID	string [DT_STR]	10			1252 (ANSI)
customerID	Copy of customerID	string [DT_STR]	10			1252 (ANSI)
employeeID	Copy of employeeID	string [DT_STR]	10			1252 (ANSI)
orderDate	Copy of orderDate	Unicode string [DT_WST...]	10			
cityID	Copy of cityID	string [DT_STR]	10			1252 (ANSI)
orderID (1)	Copy of orderID (1)	string [DT_STR]	50			1252 (ANSI)
productID	Copy of productID	string [DT_STR]	10			1252 (ANSI)
unitPrice	Copy of unitPrice	string [DT_STR]	50			1252 (ANSI)
quantity	Copy of quantity	string [DT_STR]	50			1252 (ANSI)
discount	Copy of discount	string [DT_STR]	50			1252 (ANSI)

Hình 3-29: Chuyển đổi kiểu dữ liệu đầu vào cho bảng Fact_ChiTietBan

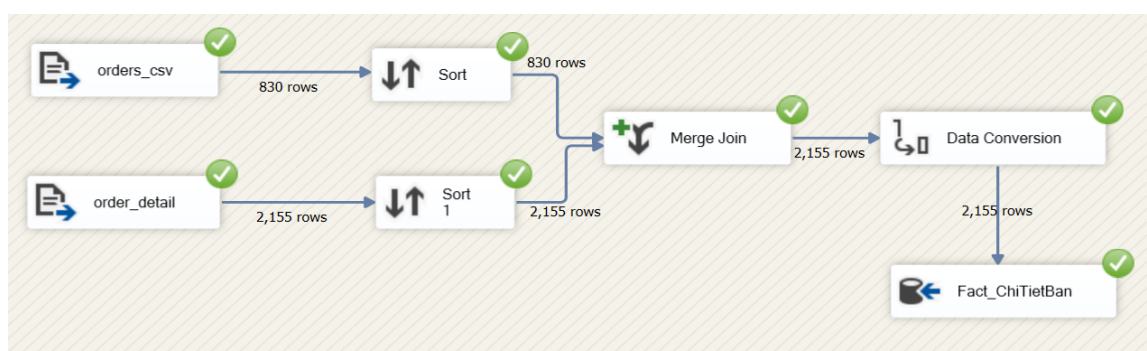
- Sử dụng đối tượng OLE DB Destination để kết nối đến bảng

Fact_ChiTietBan trong kho dữ liệu tại SQL Server; thực hiện Mapping các cột từ nguồn đến đích.



Hình 3-30: Nối các cột đầu vào và cột đích của bảng Fact_ChiTietBan

- Kết quả quá trình ETL:



Hình 3-31: Kết quả đổ dữ liệu vào bảng Fact_ChiTietBan trên SSDT

- Bảng Fact_ChiTietBan trong kho dữ liệu

	MaSanPham	DonGia	SoLuong	ChietKhau	MaKhachHang	MaNhanVien	NgayBan	TenKhuVuc
1	59	55	25	0	KOENE	9	2017-12-26	Brandenburg
2	13	6	20	0.15	KOENE	9	2017-12-26	Brandenburg
3	24	4.5	20	0.15	KOENE	9	2017-12-26	Brandenburg
4	54	7.45	7	0.1	SEVES	1	2017-12-26	London
5	11	21	50	0.1	SEVES	1	2017-12-26	London
6	51	53	10	0.1	SEVES	1	2017-12-26	London
7	17	39	40	0.25	BOLID	4	2017-12-29	Madrid
8	29	123.79	20	0.25	BOLID	4	2017-12-29	Madrid
9	30	25.89	25	0.25	SIMOB	4	2017-12-29	Kobenhavn
10	51	53	30	0.25	SIMOB	4	2017-12-29	Kobenhavn
11	55	24	60	0.25	SIMOB	4	2017-12-29	Kobenhavn
12	62	49.3	5	0.25	SIMOB	4	2017-12-29	Kobenhavn
13	59	55	15	0.05	WELLI	4	2017-12-30	Resende
14	19	9.2	24	0.05	WELLI	4	2017-12-30	Resende
15	25	14	15	0.05	WELLI	4	2017-12-30	Resende
16	10	31	36	0	SEVES	6	2017-12-30	London
17	28	45.6	24	0	SEVES	6	2017-12-30	London
18	49	20	4	0.15	SEVES	6	2017-12-30	London
19	34	14	10	0	THEBI	2	2017-12-30	Portland
20	38	263.5	10	0	THEBI	2	2017-12-30	Portland
21	74	10	15	0.25	VICTE	3	2017-12-31	Lyon
22	2	19	20	0.25	VICTE	3	2017-12-31	Lyon
23	65	21.05	2	0	VICTE	3	2017-12-31	Lyon
24	40	18.4	1	0	FRANS	4	2017-12-31	Torino
25	56	38	20	0.15	OLDWO	2	2018-01-01	Anchorage

Hình 3-32: Dữ liệu bảng Fact_ChiTietBan

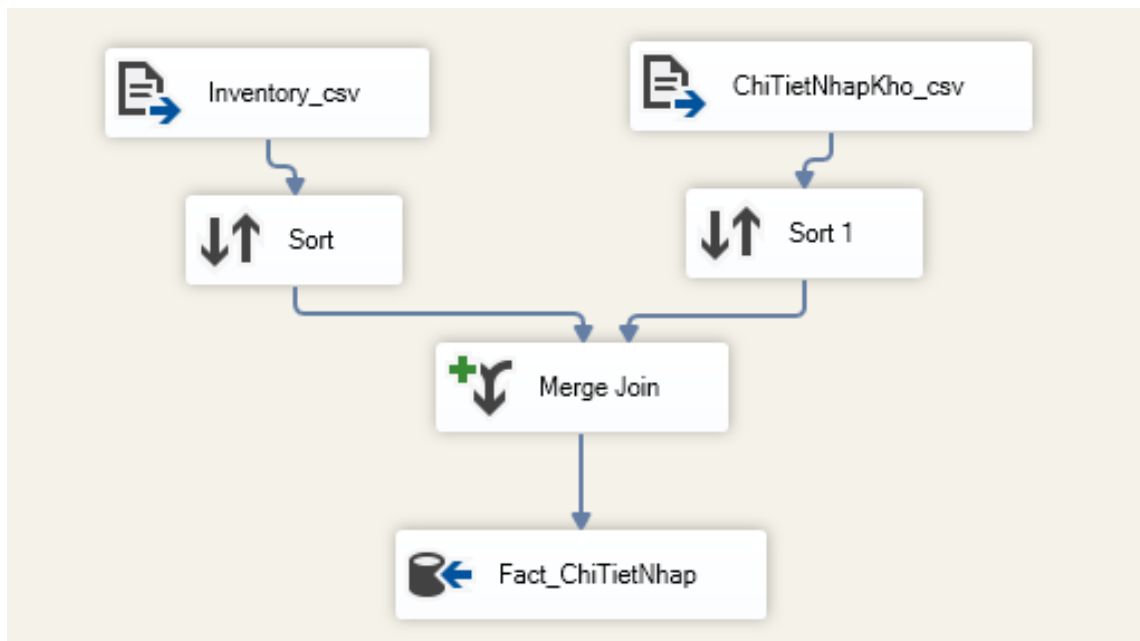
3.4.7. Bảng Fact_ChiTietNhap

- Nguồn dữ liệu gồm:
 - Bảng Hóa Đơn Nhập
 - Bảng Chi Tiết Nhập Kho
- Cấu trúc bảng Fact_ChiTietNhap

Fact ChiTietNhap	
MaNhanVien	
MaSanPham	
SoLuong	
DonGia	
NgayNhap	

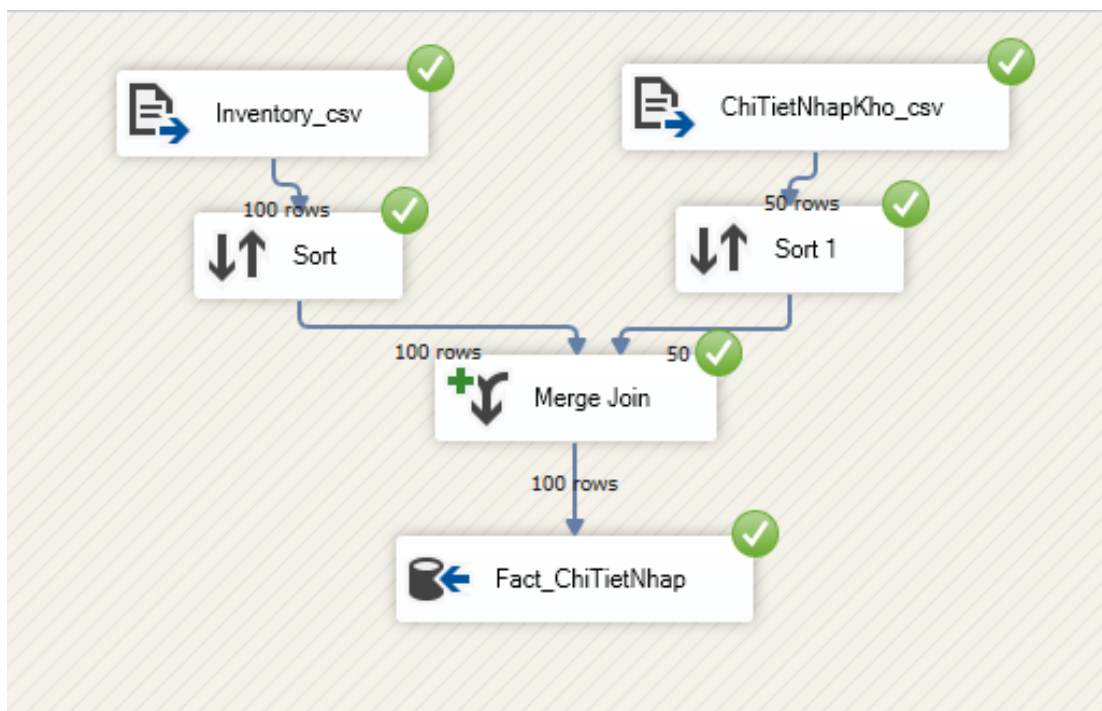
Hình 3-33: Cấu trúc bảng Fact_ChiTietNhap

- Data flow của Fact_ChiTietNhap



Hình 3-34: Data flow của Fact_ChiTietNhap

- Kết quả chạy thành công bảng Fact_ChiTietNhap trên SSDT:



Hình 3-35: Kết quả đổ dữ liệu vào bảng Fact_ChiTietNhap trên SSDT

- Import thành công dữ liệu bảng Fact_ChiTietNhap

	MaSanPham	MaNhanVien	SoLuong	DonGia	NgayNhap
1	70	2	328	15	10/22/2016
2	1	2	525	18	10/22/2016
3	60	2	658	34	2/19/2017
4	17	2	819	39	2/19/2017
5	75	5	858	7.75	12/8/2017
6	57	5	560	19.5	12/8/2017
7	22	5	737	21	1/4/2017
8	36	5	539	19	1/4/2017
9	35	6	155	18	1/28/2017
10	55	6	420	24	1/28/2017
11	54	7	980	7.45	5/4/2017
12	16	7	275	17.45	5/4/2017
13	70	7	151	15	3/16/2017
14	33	7	908	2.5	3/16/2017
15	19	3	609	9.2	2/4/2017
16	70	3	109	15	2/4/2017
17	12	9	991	38	3/12/2017
18	34	9	469	14	3/12/2017
19	69	8	451	36	7/18/2016
20	65	8	217	21.05	7/18/2016
21	77	9	692	13	8/19/2016
22	75	9	650	7.75	8/19/2016

Hình 3-36: Dữ liệu trên bảng Fact_ChiTietNhap

CHƯƠNG 4: XÂY DỰNG BÁO CÁO TÌNH HÌNH KINH DOANH

4.1. Giới thiệu công cụ Tableau

Tableau là phần mềm hỗ trợ phân tích (Data Analyst) và trực quan hóa dữ liệu (Data Visualization), biểu diễn dữ liệu bằng đồ họa có sử dụng các yếu tố trực quan như biểu đồ, đồ thị, xây dựng các dashboard để phát hiện xu hướng, mẫu hình và các giá trị ngoại lệ. Từ đó, đúc kết nhanh chóng thông tin chuyên sâu và hỗ trợ ra quyết định tức thời.

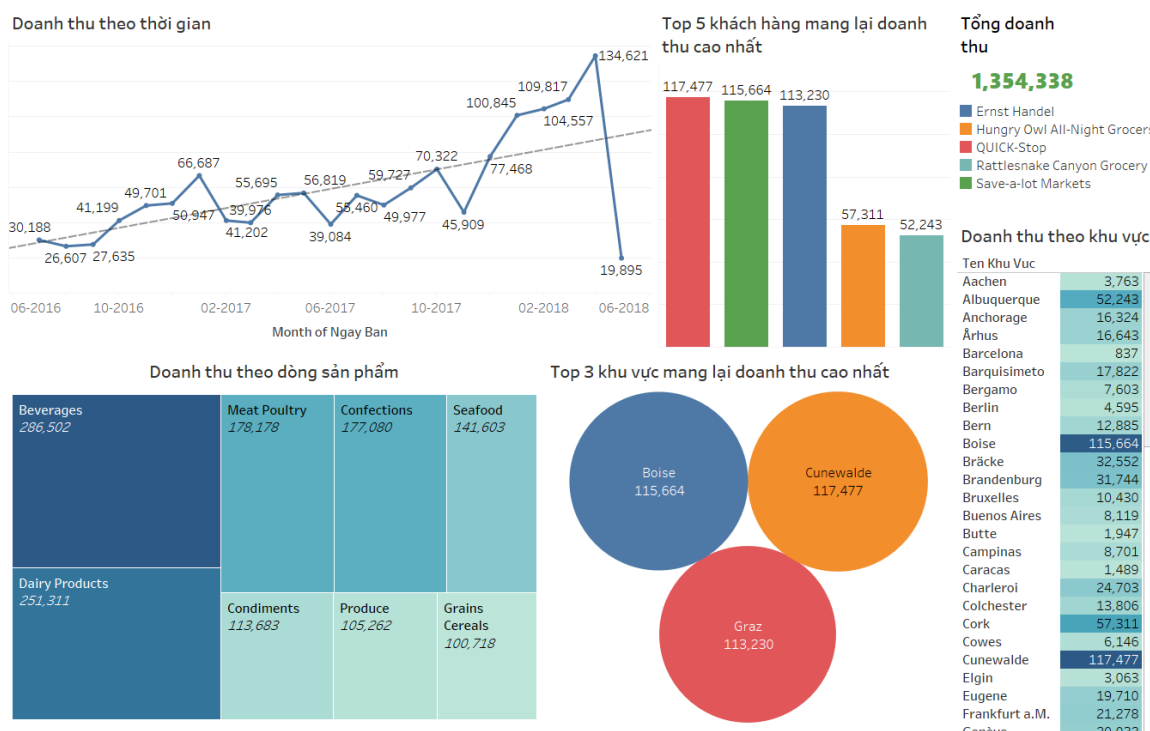
Một số ưu điểm

- Dễ sử dụng: Tableau được thiết kế để dễ sử dụng, ngay cả với những người không có kinh nghiệm về DSS hoặc BI. Phần mềm này cung cấp một giao diện trực quan, dễ hiểu, với các công cụ và tính năng được tổ chức một cách hợp lý. Điều này giúp người dùng, ngay cả những người không có chuyên môn về công nghệ, cũng có thể dễ dàng sử dụng Tableau để thu thập, phân tích và trực quan hóa dữ liệu.
- Mạnh mẽ: Tableau cung cấp một loạt các tính năng và công cụ mạnh mẽ, cho phép người dùng thu thập, phân tích và trực quan hóa dữ liệu từ nhiều nguồn khác nhau. Tableau cũng hỗ trợ nhiều loại phân tích dữ liệu, bao gồm phân tích định lượng, phân tích định tính và phân tích dự đoán. Điều này giúp người dùng có thể thực hiện nhiều loại phân tích dữ liệu phức tạp, giúp họ đưa ra quyết định kinh doanh sáng suốt hơn.
- Tùy biến: Tableau cho phép người dùng tùy chỉnh các báo cáo và biểu đồ của mình theo nhiều cách khác nhau, bao gồm thay đổi màu sắc, kiểu dáng, kích thước, vị trí và các yếu tố khác. Điều này giúp người dùng tạo ra các báo cáo và biểu đồ phù hợp với nhu cầu và sở thích của họ.
- Tích hợp: Tableau tích hợp với nhiều ứng dụng và công cụ khác, bao gồm SQL Server, Microsoft Office, Salesforce và các ứng dụng khác. Điều này giúp người dùng dễ dàng chia sẻ và phân phối các báo cáo và biểu đồ của mình với những người khác.

4.2. Xây dựng báo cáo tình hình kinh doanh

4.2.1. Báo cáo doanh thu

Giám đốc và các Phó Giám đốc là người sử dụng báo cáo để theo dõi tình hình kinh doanh mà không cần trực tiếp quản lý, giám sát. Báo cáo về doanh thu thể hiện doanh thu theo thời gian, theo khu vực, khách hàng, dòng sản phẩm. Đồng thời cũng chỉ ra những khách hàng, khu vực mang lại nhiều doanh thu để có phương án mở rộng quy mô cho hoạt động kinh doanh của siêu thị và phát triển mối quan hệ nhằm giữ chân những khách hàng trung thành.



Hình 4-1: Báo cáo doanh thu

- **Biểu đồ doanh thu theo thời gian:** Thể hiện sự tăng trưởng trong doanh thu của hệ thống siêu thị qua các năm. Tuy doanh thu của siêu thị giữa các tháng vẫn có những sự chênh lệch đáng kể nhưng vẫn cho thấy được xu hướng tăng qua các năm. Điều này được thể hiện qua *đường xu hướng (nét đứt)*.
- **Biểu đồ doanh thu theo dòng sản phẩm:** thể hiện sự ưa chuộng của khách hàng cho các dòng sản phẩm khác nhau được bày bán tại siêu thị. Trong đó, doanh thu cao nhất là nhóm sản phẩm **Beverages** (đồ uống giải khát). Tiếp theo đó là **Dairy Products** (các sản phẩm sữa), **Meat Poultry** (các sản phẩm thịt) và **Confections** (bánh kẹo). Doanh thu của những nhóm này vượt trội hơn so với doanh thu các nhóm sản phẩm khác. Đối với những nhóm sản phẩm mang lại nhiều doanh thu, siêu thị có thể có những phương án nhằm tối ưu chi phí để gia tăng lợi nhuận.

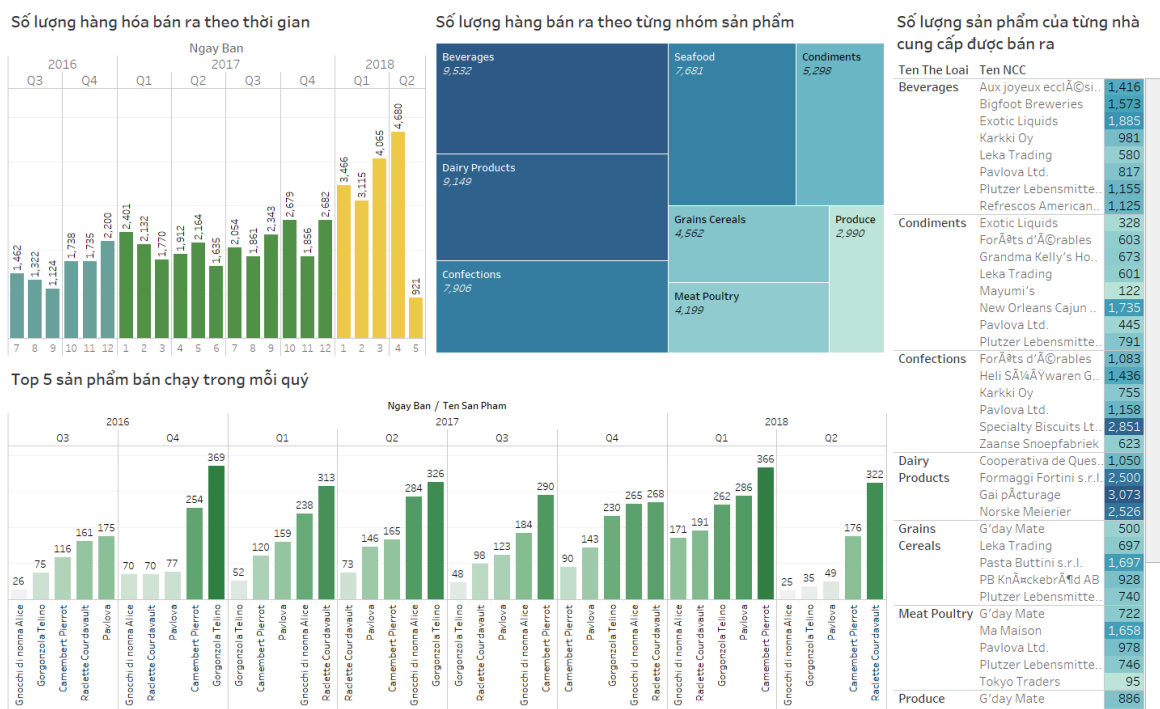
- **Biểu đồ doanh thu theo khu vực:** Thể hiện năng lực của cửa hàng tại mỗi khu vực trong việc phân phối sản phẩm đến tay khách hàng và sự sẵn sàng của khách hàng tại khu vực đó trong việc lựa chọn các sản phẩm của siêu thị.

- **Các biểu đồ Top 03 khu vực mang lại doanh thu lớn nhất, Top 05 khách hàng mang lại doanh thu cao nhất:** Thể hiện những nguồn doanh thu lớn mà siêu thị cần đặc biệt quan tâm.

- Ba khu vực mang lại doanh thu cao nhất là **Cunewalde** (doanh thu: 117.477), **Boise** (doanh thu: 115.664) và **Graz** (doanh thu: 113.230). Với các khu vực này, cần tập trung thỏa mãn tốt nhu cầu của khách hàng. Đồng thời cũng cần nghiên cứu để mở rộng quy mô cho hoạt động kinh doanh tại các khu vực này.
- Năm khách hàng mang lại nhiều doanh thu nhất là **QUICK-Stop** (doanh thu: 117.477), **Save-a-lot Markets** (doanh thu: 115.664), **Ernst Handel** (doanh thu: 113.230), **Hungry Owl All-Night-Grocers** (doanh thu: 57.311) và **Rattlesnake Canyon Grocery** (doanh thu: 52.243). Đây đều là các khách hàng tổ chức, cụ thể là cửa hàng tạp hóa, cửa hàng tiện lợi và Big C là nơi cung cấp nguồn hàng cho họ. Việc duy trì mối quan hệ với những khách hàng lớn là cần thiết vì đây là nguồn doanh thu lớn và cũng là nguồn tiêu thụ cho những sản phẩm được bày bán tại Big C.

4.2.2. Báo cáo số lượng hàng bán

Báo cáo về lượng hàng bán ra thể hiện mức độ tiêu thụ của các sản phẩm, nhóm sản phẩm. Số lượng bán ra của một mặt hàng cũng thể hiện sự ưa chuộng của khách hàng và sự sẵn sàng trong việc chi tiêu của khách hàng đối với hàng hóa đó.



mỗi dòng sản phẩm sẽ có nhiều đơn vị cung cấp những sản phẩm khác nhau và sự lựa chọn của khách hàng đối với các sản phẩm thể hiện sự tin tưởng vào chất lượng sản phẩm và uy tín của nhà cung cấp.

4.2.3. Báo cáo số lượng hàng nhập vào và bán ra

Báo cáo về số lượng hàng nhập vào và bán ra của mỗi dòng sản phẩm sẽ được sử dụng cho đối tượng là thủ kho và các trưởng bộ phận. Mục đích của việc này là nhằm theo dõi tình hình đặt hàng và bán hàng của mỗi dòng sản phẩm. Từ đó có thể kịp thời đáp ứng hàng hóa cho khách hàng và cũng hạn chế tình trạng hàng hóa để trong kho quá lâu không được tiêu thụ.

Số lượng hàng nhập và bán của từng nhóm sản phẩm

Ten The Loai	Số lượng hàng nhập	Số lượng hàng bán
Beverages	24,452	9,532
Condiments	24,024	5,298
Confections	24,298	7,906
Dairy Products	13,734	9,149
Grains Cereals	8,658	4,562
Meat Poultry	8,906	4,199
Produce	7,424	2,990
Seafood	13,042	7,681

Hình 4-3: Báo cáo số lượng hàng nhập/xuất - Giai đoạn 2016-2018

Bảng thống kê trên cho thấy số lượng hàng hóa của mỗi nhóm sản phẩm được nhập vào và bán ra, phản ánh sự chênh lệch giữa nhu cầu của khách hàng và số lượng hàng đặt của siêu thị. Những sự chênh lệch lớn giữa số lượng nhập vào và bán ra sẽ gây ra sự dư thừa hàng hóa, tăng số lượng hàng tồn kéo theo tăng chi phí lưu kho, tăng tổng chi phí hoạt động của toàn bộ hệ thống.

4.3. Đánh giá hệ thống báo cáo

Hệ thống báo cáo được xây dựng từ dữ liệu trong hệ thống data warehouse, thông qua nền tảng Tableau đã giải quyết được một số vấn đề sau:

- Báo cáo sinh động, trực quan. Dữ liệu được tổng hợp biểu diễn dưới dạng hình ảnh, biểu đồ, giúp người xem dễ dàng nhận biết được những thông tin ẩn trong dữ liệu và đưa ra quyết định dựa trên những hiểu biết mới. Các báo cáo được xây dựng thông qua nền tảng Tableau, là nền tảng dễ dàng sử dụng cho mọi đối tượng người dùng, kể cả những người dùng không có kiến thức về lập trình.

- Hệ thống gồm nhiều báo cáo đa dạng khác nhau, được xây dựng dựa trên nhiều tiêu chí, thể hiện được mối quan hệ với nhau giữa các đối tượng, chỉ tiêu so sánh. Mỗi dashboard thể hiện được mối liên quan giữa các báo cáo, để từ đó người xem dễ dàng so sánh và đưa ra những đánh giá.

- Những biểu đồ như doanh thu theo thời gian, lượng hàng bán theo thời gian, đã biểu diễn dữ liệu theo các khoảng thời gian như tháng, quý, năm. Người dùng có thể dễ dàng xem thông tin chi tiết theo mức thời gian khác nhau. Hơn nữa, báo cáo thể hiện thông tin theo trường thời gian nên người xem có thể nhận thấy xu hướng tăng hay giảm của dữ liệu để có những đánh giá và đưa ra các quyết định phù hợp.

- Hệ thống gồm nhiều báo cáo phân tích, thống kê. Người dùng có thể dựa vào số liệu này để đưa ra các chiến lược phát triển mới cho doanh nghiệp.

- Nhờ việc xây dựng hệ thống data warehouse, dữ liệu lấy trực tiếp từ nguồn là kho dữ liệu với dữ liệu đã được xử lý, làm sạch và tổng hợp theo các chỉ tiêu so sánh nên loại bỏ được những thông tin không cần thiết, góp phần cải thiện đáng kể tốc độ xây dựng báo cáo.

- Dữ liệu trong data warehouse là dữ liệu độc lập, không liên quan đến hệ thống cơ sở dữ liệu tác nghiệp đang vận hành trong quá trình hoạt động của toàn hệ thống siêu thị. Vì vậy, thực hiện báo cáo trên kho dữ liệu sẽ không ảnh hưởng đến việc xử lý các giao dịch đang thực hiện hàng ngày.

- Dữ liệu sử dụng để xây dựng báo cáo được lấy trực tiếp từ nguồn cần khai thác. Vì thế, nếu thực hiện kết nối giữa công cụ BI và cơ sở dữ liệu tác nghiệp thì mỗi lần xuất hiện một giao dịch mới, báo cáo cần phải được làm mới lại một lần. Mà với những loại hình kinh doanh có số lượng giao dịch mỗi ngày lớn thì đây là vấn đề mà không người dùng nào muốn gặp phải.

KẾT LUẬN

Toàn bộ nội dung của bài làm đã trình bày những kiến thức tổng quan về kho dữ liệu và hệ thống báo cáo tình hình kinh doanh. Đây là những công cụ cần thiết với các doanh nghiệp hiện nay, nhất là đối với những doanh nghiệp có quy mô lớn, nhiều chi nhánh, đã sở hữu những hệ thống thông tin tác nghiệp, ghi lại được những dữ liệu phát sinh thông qua các quy trình nghiệp vụ. Qua đó, những nhà quản lý sẽ có được cái nhìn tổng thể về toàn bộ hoạt động của doanh nghiệp và kịp thời đưa ra những quyết định kinh doanh.

Chương 3 của bài làm đã thực hiện mô phỏng việc xây dựng kho dữ liệu, từ việc xây dựng các bảng, xây dựng lược đồ kho dữ liệu và thực hiện quá trình ETL. Mặc dù việc xây dựng kho dữ liệu là cần thiết và là cơ sở cho việc xây dựng hệ thống báo cáo tình hình kinh doanh, nhưng quá trình trích xuất, chuyển đổi và đổ dữ liệu vào kho dữ liệu lại gặp rất nhiều khó khăn, xảy ra nhiều hiện tượng lỗi, đòi hỏi người thực hiện phải linh hoạt xử lý. Điều này làm cho toàn bộ quá trình phải diễn ra trong thời gian dài và những người thực hiện buộc phải có sự kiên trì.

Trong chương 4, toàn bộ dữ liệu trong kho dữ liệu được xây dựng từ chương 3 đã được sử dụng để xây dựng hệ thống báo cáo tình hình kinh doanh. Những báo cáo được xây dựng trên cơ sở dữ liệu thu thập trong khoảng thời gian đủ dài để rút ra những quy luật, xu hướng mà Big C Thăng Long có thể nhận biết để nâng cao hiệu quả cho hoạt động kinh doanh trên toàn hệ thống.

TÀI LIỆU THAM KHẢO

1. Ayad Hameed Mousa, M. S. (2014). Virtual Data Mart for Measuring Organizational Achivement Using Data Visualization Technique (KPIVDM).
2. Công ty TNHH Thương Mại Dịch Vụ Quốc Tế Big C Thăng Long. (2019). Hồ sơ tự công bố.
3. Đặng Văn Ninh. (2004). Retrieved 10 31, 2023, from Xây dựng kho dữ liệu: <https://123docz.net/document/2596907-xay-dung-kho-du-lieu-data-warehousing.htm>
4. Efraim Turban, R. S. (2008). Business Inteligence: A Managerial Approach. Prentice Hall.
5. Fon Silvers. (2007). Building and Maintaining a Data Warehouse. CRC Press.
6. Khoa Hệ thống thông tin quản lý, H. v. (n.d.). Bài giảng: Hệ hỗ trợ ra quyết định và kinh doanh thông minh.
7. Paulraj Ponniah. (2010). Data Warehousing: Fundamentals for IT professionals.
8. Trần Thị Thu Hà. (2020). Giáo trình Hệ thống thông tin hỗ trợ ra quyết định. Nhà xuất bản Đại học Kinh tế Quốc dân.
9. William H. Inmon. (2005). Building the Data Warehouse.