**TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT TP. HCM**

**KHOA CÔNG NGHỆ THÔNG TIN**

****

**ĐỒ ÁN HỌC PHẦN**

TÊN HỌC PHẦN: **Kho dữ liệu (DWH)**

MÃ SỐ LỚP HP: **DAWH430784\_06**

Tên đề tài Nhóm: **Phát triển mô hình Kho dữ liệu về việc đặt và giao hàng khai thác sử dụng tại các đơn vị kinh doanh có liên quan tới thủ tục giao hàng đặc biệt là kinh doanh trực tuyến**

Lớp: **181330A, 181330B**

NHÓM : 4

DANH SÁCH THÀNH VIÊN NHÓM

1. Phan Thành Trung [50]\_Nhóm trưởng: Phân hệ phân tích kho dữ liệu theo
2. Huỳnh Thị Hương Ly [23]
3. Lê Đỗ Trà My [25]
4. Võ Thị Thanh Ngân [27]

**Ngày nộp: ……./……./2021**

**TP.HCM, ngày 15 Tháng 06 năm 2021**

**NHẬN XÉT CỦA GIẢNG VIÊN ĐỐI VỚI NHÓM:**

**………………………………………………………………..**

**………………………………………………………………..**

**……………………………………………………………….**

**NHÓM : 04**

**ĐIỂM ĐÁNH GIÁ**

Võ Thị Thanh Ngân [04]: **:………….(…………………)………..**

**Ngày ……../06/2021**

**Giảng viên Ký tên**

**THÀNH VIÊN:**

Võ Thị Thanh Ngân [04]: Phân tích doanh thu, doanh số theo vị trí khu vực khách hàng.

**TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT TP.HCM**

**KHOA CÔNG NGHỆ THÔNG TIN**

**ĐỀ TÀI ĐỒ ÁN HỌC PHẦN**

TÊN HỌC PHẦN: **Kho dữ liệu (DWH)**

MÃ SỐ LỚP HP: **DAWH430784\_06**

Tên đề tài Nhóm: **Phát triển mô hình Kho dữ liệu về việc đặt và giao hàng khai thác sử dụng tại các đơn vị kinh doanh có liên quan tới thủ tục giao hàng đặc biệt là kinh doanh trực tuyến**

**Giảng viên giảng dạy: VÕ XUÂN THỂ \_ Giảng viên chính**

Lớp: **181330A, 181330B (Sáng thứ 4)**

**Tên sản phẩm đề tài: 27.N4.04.VoThiThanhNgan.DAHP.DWH.rar** (Microsoft SQL Server 2019)

**DB Nguồn: 27N4VoThiThanhNganNORTHWND**

**DB DWH: 27N4VoThiThanhNganSaleDWH**

**Tên máy chủ DB SQL: .**

**Tên đăng nhập DB SQL: sa**

**Mật khẩu (password) đăng nhập SQL: 1**

**ProjectName BIDS: 27N4VoThiThanhNganSaleDWH\_PH**

**NHÓM : 04**

THÀNH VIÊN: Võ Thị Thanh Ngân [04]

**TP.HCM, ngày 15 Tháng 06 năm 2021**

# LỜI CẢM ƠN

Nhóm chúng em xin gửi lời cảm ơn chân thành đến thầy Võ Xuân Thể - giảng viên bộ môn Kho dữ liệu trong khoa Công Nghệ Thông Tin đã trang bị cho nhóm chúng em những kiến thức, kĩ năng cơ bản cần có để hoàn thành đồ án này , cảm ơn thầy đã nhiệt huyết giảng dạy truyền đạt cho chúng em .

Tuy nhiên trong quá trình nghiên cứu đề tài, do kiến thức chuyên ngành còn hạn chế nên em vẫn còn nhiều thiếu sót khi tìm hiểu, đánh giá và trình bày về đề tài. Rất mong nhận được sự quan tâm, góp ý của các thầy/ cô giảng viên bộ môn để đề tài của em được đầy đủ và hoàn chỉnh hơn.

Chúng em xin chân thành cảm ơn.

# DANH MỤC CÁC TỪ VIẾT TẮT

|  |  |
| --- | --- |
| CSDL hoặc DB | Cơ sở dữ liệu: DataBase |
| LĐQH | Lược đồ quan hệ |
| PTH | Phụ thuộc hàm = Ràng buộc tham chiếu: Reference Constraint |
| PTTK | Phân tích và thiết kế |
| RBTV | Ràng buộc toàn vẹn [dữ liệu] |
|  |  |
| BIDS | Công cụ hỗ trợ tổ chức, quản lý và khai thác kho dữ liệu (Hệ thống tổ chức dữ liệu thông minh cho các doanh nghiệp): Business Intelligence Development Studio của Microsoft |
| DWH | Kho dữ liệu (Nhà kho dữ liệu): DataWareHouse |
| ERD | Sơ đồ thực thể kết hợp: Entity Relationship Diagram |
| MS | Công ty Microsoft |
| NF | Dạng chuẩn của CSDL: Normal Form |
| SQL | Ngôn ngữ vấn tin có cấu trúc: Structured Query Language |

# DANH MỤC CÁC THUẬT NGỮ

|  |  |
| --- | --- |
| Giao diện người dùng  (User-Interface) | Là hệ thống các màn hình giao tiếp cho phép người sử dụng tương tác với các thành phần phần mềm, điều khiển phần mềm hoạt động theo yêu cầu của người dùng - tương ứng các chức năng hiện có của phần mềm. |
| Người dùng (User):  Tài khoản (Account) | Là một quyền làm việc trên hệ thống phần mềm được cấp phát cho một cá nhân thông qua tên tài khoản (username) và mật khẩu (password). |
|  |  |
| BigData | Dữ liệu lớn: là một tập hợp dữ liệu rất lớn và phức tạp, không thể xử lý dữ liệu bằng các phương pháp truyền thống. |

# MỤC LỤC

[LỜI CẢM ƠN 4](#_Toc27681)

[DANH MỤC CÁC TỪ VIẾT TẮT 5](#_Toc26229)

[DANH MỤC CÁC THUẬT NGỮ 6](#_Toc1759)

[MỤC LỤC 7](#_Toc29345)

[Chương 1. GIỚI THIỆU VỀ ĐỒ ÁN HỌC PHẦN 10](#_Toc5832)

[1.1. Tổng quan về ĐỒ ÁN HỌC PHẦN 10](#_Toc18620)

[1.2. Nội dung chuyên môn chính của ĐỒ ÁN HỌC PHẦN 10](#_Toc6553)

[1.3. Mục tiêu của ĐỒ ÁN HỌC PHẦN 11](#_Toc27911)

[1.4. Công cụ và nền tảng kỹ thuật thực hiện ĐỒ ÁN HỌC PHẦN 11](#_Toc29212)

[1.5. Sản phẩm của ĐỒ ÁN HỌC PHẦN 11](#_Toc12400)

[1.6. Bố cục của báo cáo 11](#_Toc27053)

[Chương 2. CƠ SỞ LÝ THUYẾT CỦA HỌC PHẦN 12](#_Toc2683)

[2.1. Tổng quan về kho dữ liệu (DWH) 12](#_Toc1773)

[2.1.1. Các khái niệm cơ bản về Kho dữ liệu 12](#_Toc26545)

[2.1.2. Kiến trúc cơ bản của kho dữ liệu: 12](#_Toc11018)

[2.1.3. Mô hình CSDL đa chiều (Dimensional Modeling) 12](#_Toc14401)

[2.2. Phân tích và thiết kế Kho dữ liệu 13](#_Toc12867)

[2.2.1. Phân tích và thiết kế kho dữ liệu theo hướng mô hình hóa nhiều chiều (Dimensional Modeling) 13](#_Toc10490)

[2.2.2. Các dạng (phương pháp) tích hợp dữ liệu 13](#_Toc31121)

[2.2.4. Các công cụ tích hợp dữ liệu: 14](#_Toc200)

[2.2.5. Lập các luồng công việc tích hợp dữ liệu (Data Integration Workflows) 14](#_Toc31391)

[2.2.6. Hệ thống OLTP (On-Line Transaction Processing = Xử lý giao dịch trực tuyến) 14](#_Toc12116)

[2.3. Vấn tin khai thác sử dụng Kho dữ liệu 14](#_Toc30839)

[2.3.1. Vấn tin với các phép toán tổng hợp của SQL trên kho dữ liệu: 14](#_Toc15743)

[2.3.2. Vấn tin với các hàm phân tích của SQL trên kho dữ liệu: 14](#_Toc19076)

[2.3.3. Vấn tin dùng Materialized View để phân tích trên kho dữ liệu: 14](#_Toc5833)

[2.3.4. Hệ thống OLAP : 15](#_Toc234)

[Chương 3. TỔNG QUAN VỀ KHO DỮ LIỆU VỀ VIỆC ĐẶT VÀ GIAO HÀNG KHAI THÁC SỬ DỤNG TẠI CÁC ĐƠN VỊ KINH DOANH CÓ LIÊN QUAN TỚI THỦ TỤC GIAO HÀNG ĐẶC BIỆT LÀ KINH DOANH TRỰC TUYẾN 15](#_Toc6889)

[3.1. Ý tưởng hình thành DWH về việc đặt và giao hàng khai thác sử dụng tại các đơn vị kinh doanh có liên quan tới thủ tục giao hàng đặc biệt là kinh doanh trực tuyến 15](#_Toc13366)

[3.2. Giới thiệu chung về Kho dữ liệu 16](#_Toc8476)

[3.2.1. Mô tả tổng quan về DWH 16](#_Toc9207)

[3.2.2. Các dữ liệu nguồn hình thành DWH 16](#_Toc8948)

[3.3. Xác định nhu cần tổ chức và phân tích dữ liệu của toàn bộ DWH 20](#_Toc8898)

[3.4. Các thành phần phân hệ Kho dữ liệu phân công các cá nhân phụ trách 20](#_Toc12983)

[Chương 4. PHÂN TÍCH, THIẾT KẾ VÀ KHAI THÁC SỬ DỤNG CÁC PHÂN HỆ KHO DỮ LIỆU VỀ VIỆC ĐẶT VÀ GIAO HÀNG KHAI THÁC SỬ DỤNG TẠI CÁC ĐƠN VỊ KINH DOANH CÓ LIÊN QUAN TỚI THỦ TỤC GIAO HÀNG ĐẶC BIỆT LÀ KINH DOANH TRỰC TUYẾN 21](#_Toc939)

[4.1. <27>\_ Võ Thị Thanh Ngân\_Phân tích doanh thu, doanh số theo khu vực khách hàng (quốc gia, thành phố) 21](#_Toc16775)

*[4.1.1.](#_Toc7230)* [Mô tả tổng quan về phân hệ DWH phân tích doanh thu, doanh số theo khu vực (quốc gia, thành phố) 21](#_Toc7230)

[4.1.2. Xác định nguồn dữ liệu tích hợp phân hệ Kho dữ liệu 21](#_Toc7141)

[4.1.3. Xác định ý tưởng hình thành phân hệ của Kho dữ liệu theo hệ thống Kho dl 21](#_Toc18296)

[4.1.4. Mô tả hệ thống các DB gốc liên quan phân hệ DWH 22](#_Toc30554)

[4.1.5. Phân phân tích các DB gốc xác định yêu cầu phân tích DWH của phân hệ 22](#_Toc31058)

[4.1.6. Thiết kế DB mới tổ chức phân tích Kho dữ liệu cho phân hệ 23](#_Toc577)

[4.1.7. Thực hiện các thủ tục tích hợp các DB gốc vào phân hệ DWH 23](#_Toc12251)

[4.1.8. Thiết lập các Views tính toán cần thiết để nạp dữ liệu từ DB gốc vào các Factors của phân hệ DWH 25](#_Toc12042)

[4.1.9. Triển khai phân hệ DWH lên BIDS 26](#_Toc13816)

[4.1.10. Khai thác sử dụng phân hệ DWH dùng các vấn tin SQL 33](#_Toc32367)

[4.1.11. Trực quan hóa dữ liệu DWH: 36](#_Toc3388)

[Chương 5. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN CỦA ĐỒ ÁN HỌC PHẦN 40](#_Toc26733)

[5.1. Kết luận 40](#_Toc16255)

[5.1.1. Những kết quả đạt được 40](#_Toc16905)

[5.1.2. Hạn chế 40](#_Toc7702)

[5.2. Hướng phát triển 40](#_Toc11811)

[5.2.1. Hướng khắc phục các hạn chế 40](#_Toc25558)

[5.2.2. Hướng mở rộng ĐỒ ÁN HỌC PHẦN 40](#_Toc23004)

[DANH MỤC TÀI LIỆU THAM KHẢO 41](#_Toc6231)

[ĐÁNH GIÁ KẾT QUẢ HỌC TẬP HỌC PHẦN 42](#_Toc2413)

# GIỚI THIỆU VỀ ĐỒ ÁN HỌC PHẦN

## Tổng quan về ĐỒ ÁN HỌC PHẦN

Đồ án Học phần: Phát triển mô hình Kho dữ liệu về việc đặt và giao hàng khai thác sử

dụng tại các đơn vị kinh doanh có liên quan tới thủ tục giao hàng đặc biệt là kinh doanh

trực tuyến.

## Nội dung chuyên môn chính của ĐỒ ÁN HỌC PHẦN

* Nâng cao chất lượng dữ liệu bằng các phương pháp làm sạch và tinh lọc dữ liệu theo những hướng chủ đề
* Tổng hợp và kết nối dữ liệu
* Đồng bộ hóa các nguồn dữ liệu với DWH
* Phân định và đồng nhất các hệ quản trị cơ sở dữ liệu tác nghiệp như là các công cụ chuẩn để phục vụ cho DWH
* Quản lí DWH
* Cung cấp thông tin được tích hợp, tóm tắt hoặc được liên kết, tổ chức theo các chủ đề
* Dùng trong các hệ thống hỗ trợ quyết định ( Decision support system – DSS), các hệ thống thông tin tác nghiệp hoặc hỗ trợ cho các truy vấn đặc biệt

## Mục tiêu của ĐỒ ÁN HỌC PHẦN

* Khả năng đáp ứng mọi yêu cầu về thông tin của người sử dụng
* Hỗ trợ để các tổ chức thực hiện tốt, hiệu quả công việc của mình, cũng như có những quyết định hợp lý, nhanh chóng hơn,…
* Giúp cho tổ chức, xác định và phân tích các nghiệp vụ một các hiệu quả và chính xác
* Tích hợp dữ liệu và các siêu dữ liệu từ nhiều nguồn khác nhau

## Công cụ và nền tảng kỹ thuật thực hiện ĐỒ ÁN HỌC PHẦN

Các công cụ chính sử dụng trong đồ án này là :

+ Microsoft SQL Server

+ SQL Server Business Intelligence Development Studio (BIDS)

Các nền tảng kỹ thuật thực hiện đồ án:

+ Nắm bắt và hiểu rõ các kiến thức cơ bản về kho dữ liệu cũng như cơ sở dữ liệu

+ Thành thạo sử dụng tốt các loại truy vấn

+ Khả năng phân định, tích hợp dữ liệu từ nhiều nguồn

## Sản phẩm của ĐỒ ÁN HỌC PHẦN

Nhằm theo dõi và giúp các công ty có được những thông tin cần thiết giúp ích cho việc phát triển chiến thuật và chiến lược kinh doanh. Xây dựng data warehouse là một phương pháp tối ưu các công việc và hỗ trợ đưa ra quyết định. Ví dụ tùy thuộc vào khoảng thời gian nào trong năm hoặc tùy thuộc vào khu vực nào mà ta có thể đưa ra quyết định lựa chọn những quảng cáo sản phẩm phù hợp với nhu cầu của khách hàng.

## Bố cục của báo cáo

Báo cáo gồm những nội dung như sau:

Chương 1: Giới thiệu đồ án của học phần DWH

Chương 2: Các cơ sở lý thuyết của HP phục việc thực hiện đề tài HP

Chương 3: Giới thiệu tổng quan về Kho dữ liệu (DWH) về việc đặt và giao hàng khai

thác sử dụng tại các đơn vị kinh doanh có liên quan tới thủ tục giao hàng đặc biệt là kinh

doanh trực tuyến

Chương 4: Phân tích, thiết kế và tổ chức khai thác sử dụng từng phân hệ của Kho dữ liệu

Chương 5: Tổng kết các kết quản đạt được và còn hạn chế của đồ án, đồng thời đề xuất hướng khắc phục hạn chế và phát triển Đồ án.

# CƠ SỞ LÝ THUYẾT CỦA HỌC PHẦN

## Tổng quan về kho dữ liệu (DWH)

### Các khái niệm cơ bản về Kho dữ liệu

#### Kho dữ liệu (DWH)?

#### Nguyên lý hình thành DWH

#### Các đặc điểm cơ bản của DWH

#### Phân biệt [SS] DWH với CSDL truyền thống

### Kiến trúc cơ bản của kho dữ liệu:

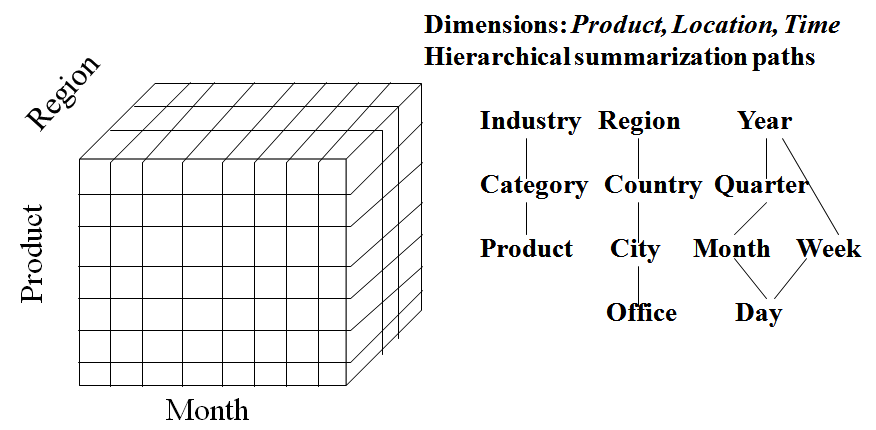
#### Tầng đáy: nạp dữ liệu

#### Tầng giữa: OLAP (OLAP server): xử lý DWH

#### Tầng trên cùng: ứng dụng = khai thác sử dụng DWH

### Mô hình CSDL đa chiều (Dimensional Modeling)

* Giới thiệu CSDL đa chiều
* Phân biệt mô hình dữ liệu quan hệ và mô hình dữ liệu nhiều chiều
  + Minh họa: Cơ sở dữ liệu dưới đây gồm 3 chiều, mỗi chiều có các mức sau
    - Chiều hàng hoá(Product), có các mức: sản phẩm (Product), loại sản phẩm (category), công nghiệp (industry)
    - Chiều thị trường có các mức: khu vực (Region), quốc gia (Country), thành phố (City), địa điểm (Office)
    - Chiều thời gian, có các mức: năm (year), quý (quarter), tháng (month), tuần (week), ngày (day)



## Phân tích và thiết kế Kho dữ liệu

### Phân tích và thiết kế kho dữ liệu theo hướng mô hình hóa nhiều chiều (Dimensional Modeling)

* Khối dữ liệu (data cube)
* Các phép toán trên khối dữ liệu

### Các dạng (phương pháp) tích hợp dữ liệu

* Hướng chủ đề (Subject Oriented)
* Hướng tích hợp (Integrated Oriented)
* Hướng biến đổi theo thời gian (Time Variant Oriented)
* Hướng ổn định (Non Volatile Oriented)
  + 1. Các bước cơ bản trong tiến trình tích hợp dữ liệu:
* B1: Trích xuất dữ liệu = thu thập dữ liệu lớn từ nhiều nguồn
  + Minh họa: Trích từ csdl bệnh nhân, các bản chép tay, word, excel
* B2: Làm sạch = tìm lỗi và sửa sai. Khắc phục các giá trị NULL, xử li các giá trị trống trong database. Việc trống quá nhiều dữ liệu dẫn đến thông tin bị lệch, dự đoán dự báo bị sai, dẫn đến hệ thống dự báo dự đoán sai
  + Minh họa: tập dữ liệu về thông tin các nạn nhân trong một trận sóng thần thì có rất nhiều thông tin thiếu xót bằng việc lấy giá trị max để thay cho giá trị thiếu, lấy hồi qui hoặc xóa bỏ bộ dữ liệu đó nếu không ảnh hưởng quá nhiều là các cách để xử lí những giá trị trống
* B3: So khớp mẫu = Lọc và chuyển đổi dữ liệu = chuyển đổi sang định dạng kho
  + Minh họa: chuyển dữ liệu từ dạng xml, json sang dạng có cấu trúc như SQL server
* B4: Lưu trữ dữ liệu theo cấu trúc = sắp xếp, hợp nhất và tổng hợp,..

### Các công cụ tích hợp dữ liệu:

### Lập các luồng công việc tích hợp dữ liệu (Data Integration Workflows)

### Hệ thống OLTP (On-Line Transaction Processing = Xử lý giao dịch trực tuyến)

## Vấn tin khai thác sử dụng Kho dữ liệu

### Vấn tin với các phép toán tổng hợp của SQL trên kho dữ liệu:

Nhắc lại mệnh đề GROUP BY

Các mở rộng của các hệ quản trị cơ sở dữ liệu để hỗ trợ cho kho dữ liệu

Mệnh đề SQL CUBE

Mệnh đề SQL ROLLUP

Mệnh đề SQL GROUPING SETS

Kết hợp các phép toán SUBTOTAL

### Vấn tin với các hàm phân tích của SQL trên kho dữ liệu:

Mô hình xử lí và cú pháp cơ bản

Cú pháp mở rộng và các hàm xếp hạng

So sánh cửa sổ (Window Comparisons)

Các hàm so sánh tỷ lệ

### Vấn tin dùng Materialized View để phân tích trên kho dữ liệu:

Phân biệt : Traditional View và Materialized View

|  |  |
| --- | --- |
| Vấn tin truyền thống | Vấn tin phân tích |
| * Chủ yếu là hỏi đáp, sử dụng các câu lệnh như: Select..from…where * Vấn tin trên cơ sở dữ liệu | * Là vấn tin trên kho dữ liệu, phân tích, rút ra kết luận nào đó từ kho dữ liệu, từ DWH: SQL Subtotal, SQL Analytic, Materialized View * Phân tích để xác định các qui luật có sẵn trong DWH, chứ không đơn giản chỉ là hỏi đáp thông thường |
| * Ví dụ : Lọc ra danh sách sinh viên có điểm trung bình lớn hơn 6.5 và điểm rèn luyện thuộc loại Giỏi . | * Ví dụ: từ DWH thông tin sinh viên đăng kí ngành học của các trường đại học dùng vấn tin phân tích cho biết qui luật đăng kí ngành học của sinh viên . * Nếu trường chuyên ngành kĩ thuật thì sẽ có xu hướng mua những dụng cụ và thiết bị nào , xác suất là bao nhiêu %. |

Viết truy vấn dùng Materialized View

Nhắc lại khái niệm khung nhìn (views)

Định nghĩa materialized view và cách thức nó được xử lí

Các quy tắc khi viết lại truy vấn (query rewrite) dùng materialized view

### Hệ thống OLAP :

Cube (khối) Khối dữ liệu + Tạo khối (cube) cho dữ liệu với

Các bảng đa chiều (dimension table) và

Bảng sự kiện (fact table)

Một số thao tác (lệnh ) cơ bản của OLAP:

Thu nhỏ (roll – up)

Mở rộng (drill- down)

Cắt lát (slice)

Rút ngắn(dice)

# **TỔNG QUAN VỀ KHO DỮ LIỆU** VỀ VIỆC ĐẶT VÀ GIAO HÀNG KHAI THÁC SỬ DỤNG TẠI CÁC ĐƠN VỊ KINH DOANH CÓ LIÊN QUAN TỚI THỦ TỤC GIAO HÀNG ĐẶC BIỆT LÀ KINH DOANH TRỰC TUYẾN

## Ý tưởng hình thành DWH về việc đặt và giao hàng khai thác sử dụng tại các đơn vị kinh doanh có liên quan tới thủ tục giao hàng đặc biệt là kinh doanh trực tuyến

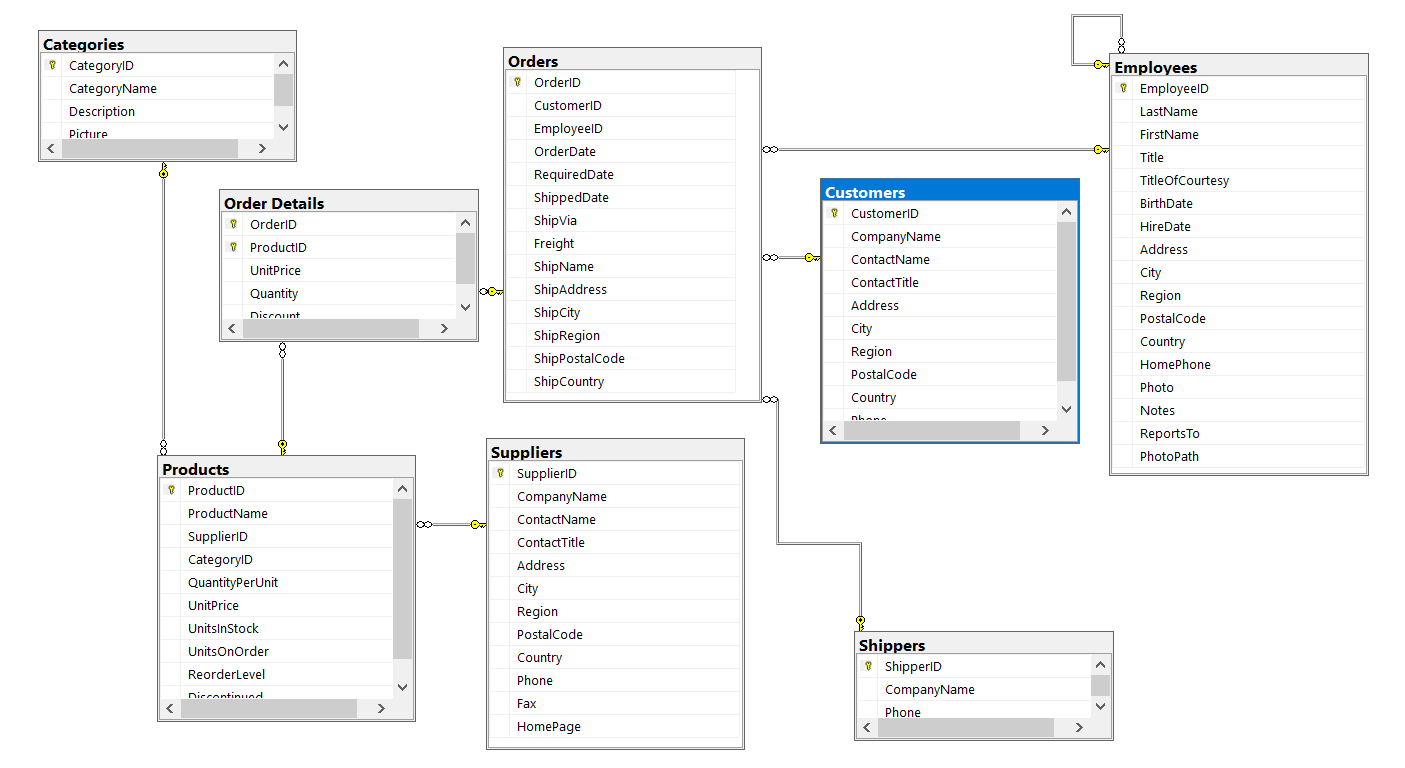
Trong vai trò là một công ty kinh doanh mà có nhiều giao dịch với các công ty cả trong lẫn ngoài nước thì việc biết được thống kê về doanh thu và doanh số là điều cực kì quan trọng. Vì chúng là cơ sở để công ty đề ra các chiến lược kinh tế và phát triển hợp lí, cũng như nắm được thị hiếu khách hàng. Theo đó, chúng em cho rằng doanh thu và doanh số sẽ chịu ảnh hưởng từ các yếu tố sản phẩm, phân loại sản phẩm và nhân viên công ty. Do vậy, chúng em muốn thực hiện các phân tích doanh thu, doanh số theo sản phẩm, nhóm sản phẩm(phân loại sản phẩm) và theo nhân viên công ty

## Giới thiệu chung về Kho dữ liệu

### Mô tả tổng quan về DWH

* DWH này tập trung thống kê các số liệu liên quan đến việc đưa ra chiến lược kinh doanh trong tương lai như là: doanh thu, doanh số,….
* Các Dim trong DWH này: nhân viên (Employee), nhà cung cấp (Supplier), danh mục sản phẩm (Category), khách hàng (Customer), sản phẩm (Product), thời gian (Date), đơn đặt hàng (Order)
* Các Fact trong DWH này: doanh thu (totalrevenue), doanh số (totalquantity)

### Các dữ liệu nguồn hình thành DWH



**Customers: Lưu trữ thông tin khách hàng**

|  |  |  |  |
| --- | --- | --- | --- |
| Tên thuộc tính | Kiểu dữ liệu | Ghi chú | Mô tả |
| CustomerID | nchar(5) | Not Null, Primary Key | Mã khách hàng |
| CompanyName | nvarchar(40) | Not Null | Tên tổ chức, tập thể mua hàng(Vì khách hàng của doanh nghiệp chủ yếu là các tập thể, tổ chức kinh tế khác ) |
| ContactName | nvarchar(30) | Null | Tên cá nhân, người đảm nhận vị trí liên lạc khi có giao dịch. |
| Contactitle | nvarchar(30) | Null | Chức danh của người đại diện liên hệ khi cần giao dịch. |
| Address | nvarchar(60) | Null | Địa chỉ khách hàng. |
| City | nvarchar(15) | Null | Thành phố mà khách hàng đang ở. |
| Region | nvarchar(15) | Null | Vùng/khu vực mà khách hàng đang ở. |
| PostalCode | nvarchar(10) | Null | Mã bưu chính của khách hàng. |
| Country | nvarchar(15) | Null | Quốc gia mà khách hàng đang ở. |
| Phone | nvarchar(24) | Null | Số điện thoại liên hệ của khách hàng. |

**Shippers: Lưu trữ thông tin các đơn vị vận chuyển hợp tác với doanh nghiệp.**

|  |  |  |  |
| --- | --- | --- | --- |
| Tên thuộc tính | Kiểu dữ liệu | Ghi chú | Mô tả |
| ShipperID | int | Not Null, Primary Key | Mã đơn vị vận chuyển. |
| CompanyName | nvarchar(40) | Not Null | Tên đơn vị vận chuyển. |
| Phone | nvarchar(24) | Null | Số điện thoại liên hệ của đơn vị vận chuyển. |

**Category : Lưu trữ thông tin các danh mục sản phẩm.**

|  |  |  |  |
| --- | --- | --- | --- |
| Tên thuộc tính | Kiểu dữ liệu | Ghi chú | Mô tả |
| CategoryID | int | Not Null, Primary Key | Mã danh mục. |
| CategoryName | nvarchar(15) | Not Null | Tên danh mục. |
| Description | ntext | Null | Mô tả danh mục |
| Picture | image | Null | Hình ảnh danh mục |

**Products : Lưu trữ thông tin các sản phẩm.**

|  |  |  |  |
| --- | --- | --- | --- |
| Tên thuộc tính | Kiểu dữ liệu | Ghi chú | Mô tả |
| ProductID | int | Not Null, Primary Key | Mã sản phẩm. |
| ProductName | nvarchar(40) | Not Null | Tên sản phẩm. |
| SupplierID | int | Null | Id nhà cung cấp |
| CategoryID | int | Null | Id danh mục |
| QuantityPerUnit | nvarchar(20) | Null | Số lượng mỗi đơn vị |
| UnitPrice | money | Null | Đơn giá |
| UnitsInStock | smallint |  | Đơn vị trong kho |
| UnitsOnOrder | smallint | Null |  |
| ReorderLevel | smallint | Null |  |
| Discontinued | bit |  |  |

**Supplier : Lưu trữ thông tin các nhà cung cấp.**

|  |  |  |  |
| --- | --- | --- | --- |
| Tên thuộc tính | Kiểu dữ liệu | Ghi chú | Mô tả |
| SupplierID | int | Not Null, Primary Key | Mã nhà cung cấp. |
| CompanyName | nvarchar(40) | Not Null | Tên công ty |
| ContactName | nvarchar(30) | Null | Tên cá nhân, người đảm nhận vị trí liên lạc khi có giao dịch. |
| Contacthuộc tínhitle | nvarchar(30) | Null | Chức danh của người đại diện liên hệ khi cần giao dịch. |
| Address | nvarchar(60) | Null | Địa chỉ nhà cung cấp |
| City | nvarchar(15) | Null | Thành phố nhà cung cấp đang ở |
| Region | nvarchar(15) |  | Khu vực nhà cung cấp |
| PostalCode | nvarchar(10) | Null | Mã số thuế |
| Country | nvarchar(15) | Null | Quốc gia của nhà cung cấp |
| Phone | nvarchar(24) |  | Số điện thoại của nhà cung cấp |

**Orders : Lưu trữ thông tin các đơn đặt hàng.**

|  |  |  |  |
| --- | --- | --- | --- |
| Tên thuộc tính | Kiểu dữ liệu | Ghi chú | Mô tả |
| OrderID | int | Not Null, Primary Key | Mã đơn đặt hàng |
| CustomerID | nchar(5) | Not Null | Mã khách hàng |
| EmployeeID | int | Null | Mã nhân viên |
| OrderDate | datetime | Null | Ngày đặt hàng |
| RequiredDate | datetime | Null | Ngày giao hàng yêu cầu |
| ShippedDate | datetime | Null | Ngày đã giao cho khách hàng |
| ShipVia | int | Null | Số lượng chuyến tàu vận chuyển hàng |
| Freight | money | Null | Chi phí vận chuyển hàng hóa |
| ShipName | nvarchar(40) | Null | Tên công ty khách hàng |
| ShipAddress | nvarchar(60) | Null | Địa chỉ khách hàng |
| ShipCity | nvarchar(15) | Null | Thành phố gia hàng tới |
| ShipRegion | nvarchar(15) | Null | Khu vực giao hàng tới |
| ShipPostalCode | nvarchar(10) | Null | Mã bưu chính khách hàng |
| ShipCountry | nvarchar(15) | Null | Quốc gia khách hàng |

**Order detail: Lưu trữ thông tin các chi tiết của từng đơn đặt hàng.**

|  |  |  |  |
| --- | --- | --- | --- |
| Tên thuộc tính | Kiểu dữ liệu | Ghi chú | Mô tả |
| OrderID | int | Not Null, Primary Key | Mã đơn đặt hàng |
| ProductID | int | Not Null | Mã sản phẩm |
| UnitPrice | money | Null | Đơn giá |
| Quantity | smallint | Null | Số lượng |
| Discount | real | Null | Giảm giá |

**Employees : Lưu trữ thông tin các nhân viên**

|  |  |  |  |
| --- | --- | --- | --- |
| Tên thuộc tính | Kiểu dữ liệu | Ghi chú | Mô tả |
| EmployeeID | int | Not Null, Primary Key | Mã nhân viên |
| LastName | nvarchar(20) | Not Null | Họ của nhân viên |
| FirstName | nvarchar(10) | Null | Tên của nhân viên |
| Title | nvarchar(30) | Null | Chức danh của nhân viên |
| TitleOfCourtesy | nvarchar(25) | Null | Danh xưng |
| BirthDate | datetime | Null | Ngày sinh của nhân viên |
| HireDate | datetime | Null | Ngày bắt đầu làm việc |
| Address | nvarchar(60) | Null | Địa chỉ của nhân viên |
| City | nvarchar(15) | Null | Thành phố của nhân viên ở |
| Region | nvarchar(15) | Null | Khu vực nhân viên ở |
| PostalCode | nvarchar(10) | Null | Mã bưu chính |
| Country | nvarchar(15) | Null | Quốc gia |
| HomePhone | nvarchar(24) | Null | Số điện thoại |
| Photo | image | Null | Hình ảnh |
| Notes | ntext | Null | Ghi chú về nhân viên |
| ReportsTo | int | Null | Mã số người quản lí trực thuộc của nhân viên. |
| PhotoPath | nvarchar(255) | Not null | Hình ảnh |

## Xác định nhu cần tổ chức và phân tích dữ liệu của toàn bộ DWH

Tổ chức và phân tích dữ liệu của toàn bộ DWH là một việc làm hết sức cần thiết và quan trọng. Nó giúp chúng ta thấy một bức tranh tổng quan của các số liệu thống kê, nó giúp cho người quản lý dễ dàng sử dụng để có thể thấy được những số liệu thống kê mà mình muốn đồng thời dễ dàng cho việc trực quan hóa.DWH giúp việc tổ chức dữ liệu một cách rõ ràng, cho ta những thông tin cần thiết phục vụ cho những kế hoạch sử dụng của doanh nghiệp, cá nhân.

## Các thành phần phân hệ Kho dữ liệu phân công các cá nhân phụ trách

Đề tài chung của Nhóm: **Phát triển mô hình Kho dữ liệu về việc đặt và giao hàng khai thác sử dụng tại các đơn vị kinh doanh có liên quan tới thủ tục giao hàng đặc biệt là kinh doanh trực tuyến**

<50>\_ Phan Thành Trung\_ Phân tích doanh thu và doanh số theo thời gian và đơn hàng.

<23>\_ Huỳnh Thị Hương Ly\_ Phân tích doanh thu, doanh số theo nhân viên.

<25>\_ Lê Đỗ Trà My\_ Phân tích doanh thu, doanh số theo sản phẩm và nhà cung cấp.

<27>\_ Võ Thị Thanh Ngân\_ Phân tích doanh thu, doanh số theo vị trí khu vực khách hàng.

# PHÂN TÍCH, THIẾT KẾ VÀ KHAI THÁC SỬ DỤNG CÁC PHÂN HỆ KHO DỮ LIỆU VỀ VIỆC ĐẶT VÀ GIAO HÀNG KHAI THÁC SỬ DỤNG TẠI CÁC ĐƠN VỊ KINH DOANH CÓ LIÊN QUAN TỚI THỦ TỤC GIAO HÀNG ĐẶC BIỆT LÀ KINH DOANH TRỰC TUYẾN

## **<27>\_ Võ Thị Thanh Ngân\_Phân tích doanh thu, doanh số theo khu vực khách hàng (quốc gia, thành phố)**

### **Mô tả tổng quan về phân hệ DWH phân tích doanh thu, doanh số theo khu vực (quốc gia, thành phố)**

Để xác định nhóm khu vực khách hàng tiềm năng, thuận tiện triển khai kế hoạch kinh doanh phù hợp, là tất yếu trong kinh doanh. Nắm bắt đều đó, ta tiến hành phân tích nhóm khách hàng ở vị trí nào tiềm năng , từ đó, ta xác định được khu vực cần đầu tư, quan tâm khách hàng ở khu vực đó. Ngoài ra, còn tìm ra nhóm khu vực khách hàng kém phát triển, để có biện pháp thích hợp khắc phục những hạn chế.

### **Xác định nguồn dữ liệu tích hợp phân hệ Kho dữ liệu**

Nguồn dữ liệu tích hợp phân hệ Kho dữ liệu là database thông tin về mua bán và giao các sản phẩm của một công ty xuất nhập khẩu xuyên quốc gia từ năm 1996-1998.

### **Xác định ý tưởng hình thành phân hệ của Kho dữ liệu theo hệ thống Kho dl**

Trong vai trò là một công ty kinh doanh, có nhiều giao dịch với các công ty cả trong lẫn ngoài nước, thì việc biết được thống kê về doanh thu và doanh số là điều cực kì quan trọng. Vì chúng là cơ sở để công ty đề ra các chiến lược kinh tế và phát triển hợp lí, cũng như nắm được thị hiếu khách hàng. Theo đó, tôi cho rằng doanh thu và doanh số sẽ chịu ảnh hưởng từ các yếu tố khu vực khách hàng, sản phẩm, phân loại sản phẩm và nhân viên công ty. Do vậy, tôi muốn thực hiện các phân tích doanh thu, doanh số theo vị trí khu vực của khách hàng, cụ thể ở đây là quốc gia và thành phố.

### **IMG_256Mô tả hệ thống các DB gốc liên quan phân hệ DWH**

Ở đây ta quan tâm đến bảng:

+ Customers: quan trọng nhất là thuộc tính **[Country]** và **[City]**, đối tượng chính ta đang quan tâm.

+ Oders: là bảng trung gian để liên kết bảng Order\_Details với Customer.

+ Order\_Details: để lấy các thuộc tính [UnitPrice]-Giá trên 1 đơn vị, [Quantity]-Số lượng, [Discount]-giảm giá. Để tính:

--Doanh thu = (Giá bán \* Số lượng) - Giảm giá

--Doanh thu = Số lượng

### **Phân phân tích các DB gốc xác định yêu cầu phân tích DWH của phân hệ**

Ở đây ta quan đến doanh thu, doanh số theo khu vực, nên chỉ cần 2 bảng từ DB gốc là Customers và Order để làm 2 dim.

Xác định các Facts và Dims:

**Doanh thu, doanh số theo vị trí khu vực của khách hàng**

**Fact\_Location:**

+ Primary key: Coutry, City, id\_order, id\_customer

+ **DoanhThu (Revenue) =** SUM(unitprice\*quantity) - Discount WHERE order\_details

**+ DoanhSo (Quantity)** = SUM(quantity) WHERE order\_details

**Dim\_customers: Xác định danh tính khách hàng thuộc khu vực nào**

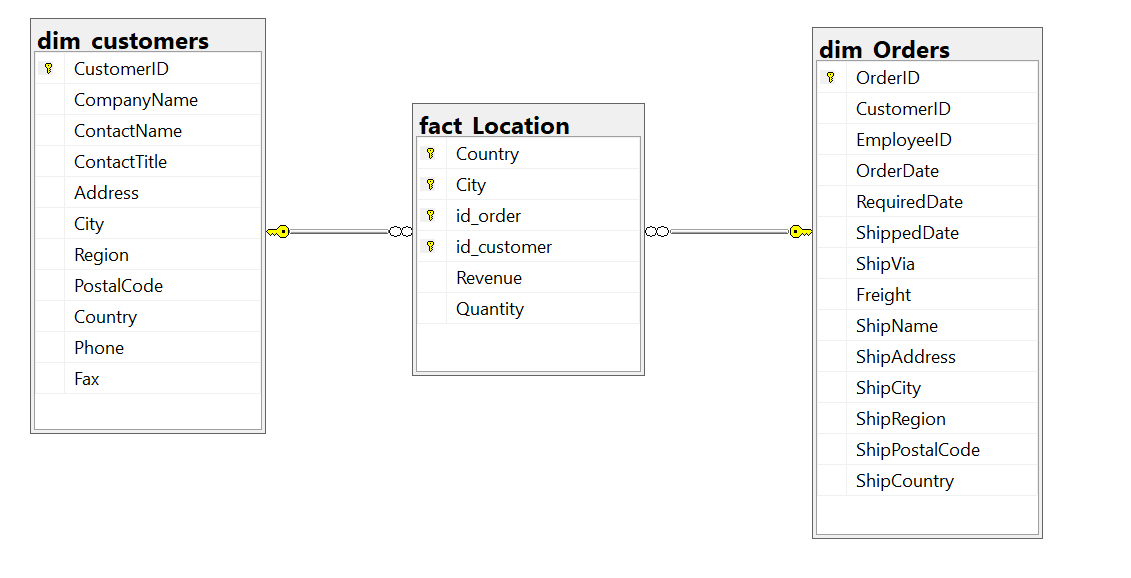
**[CustomerID]- PK** , [CompanyName], [ContactName] , [ContactTitle], [Address],[City], [Region], [PostalCode], [Country], [Phone], [Fax]

**Dim\_orders: Xác định thông tin đơn đặt hàng**

**[OrderID]-PK** ,[CustomerID],[EmployeeID] ,[OrderDate] ,[RequiredDate],

[ShippedDate] ,[ShipVia] ,[Freight] ,[ShipName] ,[ShipAddress],[ShipCity] , [ShipRegion], [ShipPostalCode], [ShipCountry]

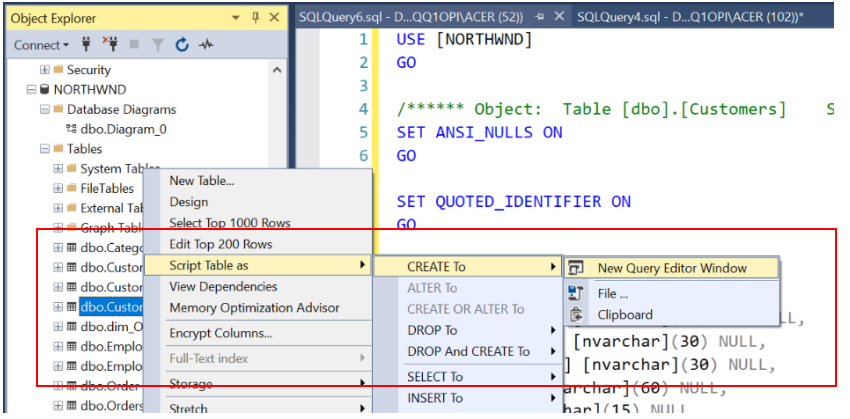
### **Thiết kế DB mới tổ chức phân tích Kho dữ liệu cho phân hệ**

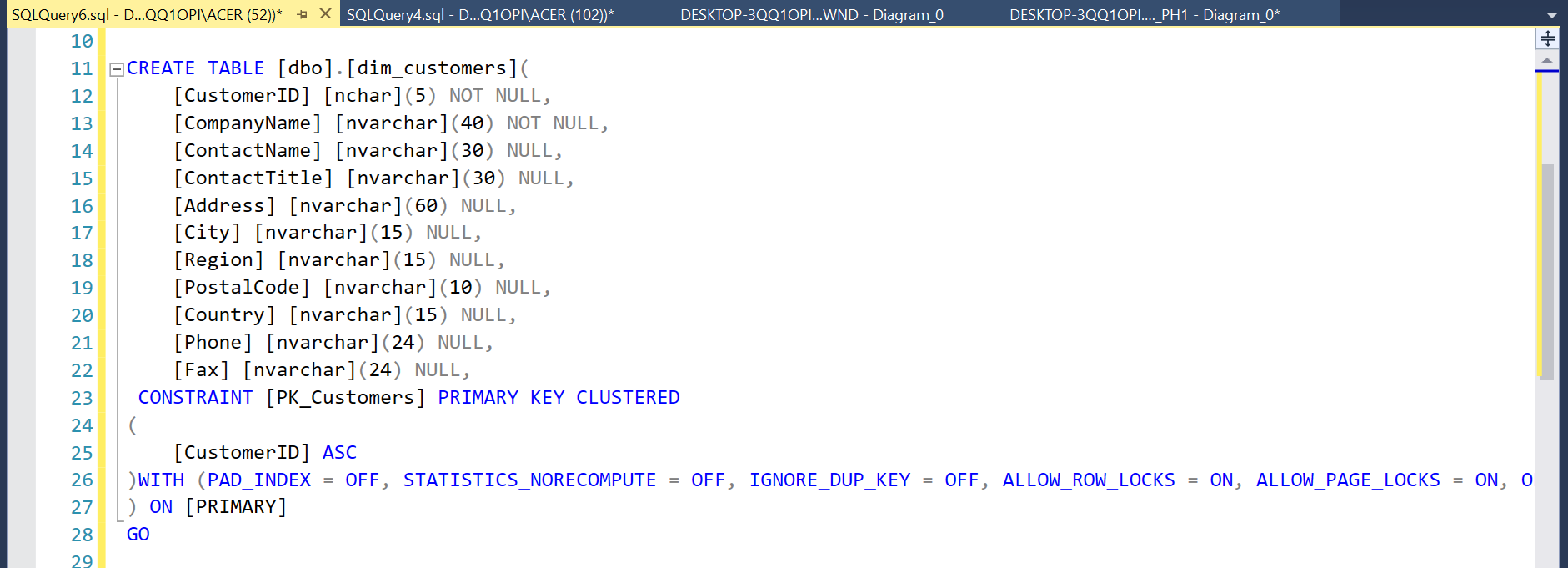


### **Thực hiện các thủ tục tích hợp các DB gốc vào phân hệ DWH**

- Do ta chỉ lấy 2 bảng từ DB gốc, nên sử dụng cách dưới dây để lấy dữ liệu, thay vì copy toàn bộ DB, để tránh mất thời gian và thừa bảng.

**-** Tạo bảng dim\_customers bằng cách lấy script từ DB bảng gốc, tương tự với dim\_order:

chuột phải vào bảng ở DB gốc và làm theo hướng dẫn hình dưới đây:

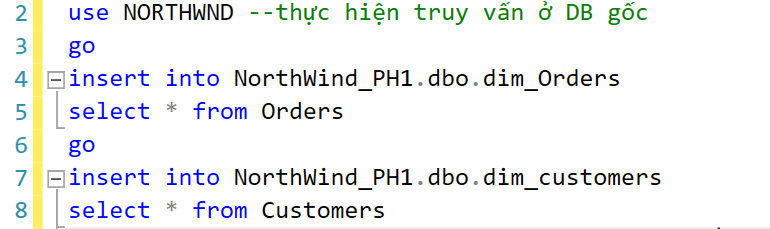
- Xóa các dòng thừa đi và đổi tên bảng,ta được lệnh truy vấn tạo bảng:

- Nạp dữ liệu vào dim, với cú pháp truy vấn đơn giản:

Insert into <tên\_DB\_đích.tên bảng> (tên cột muốn lấy)

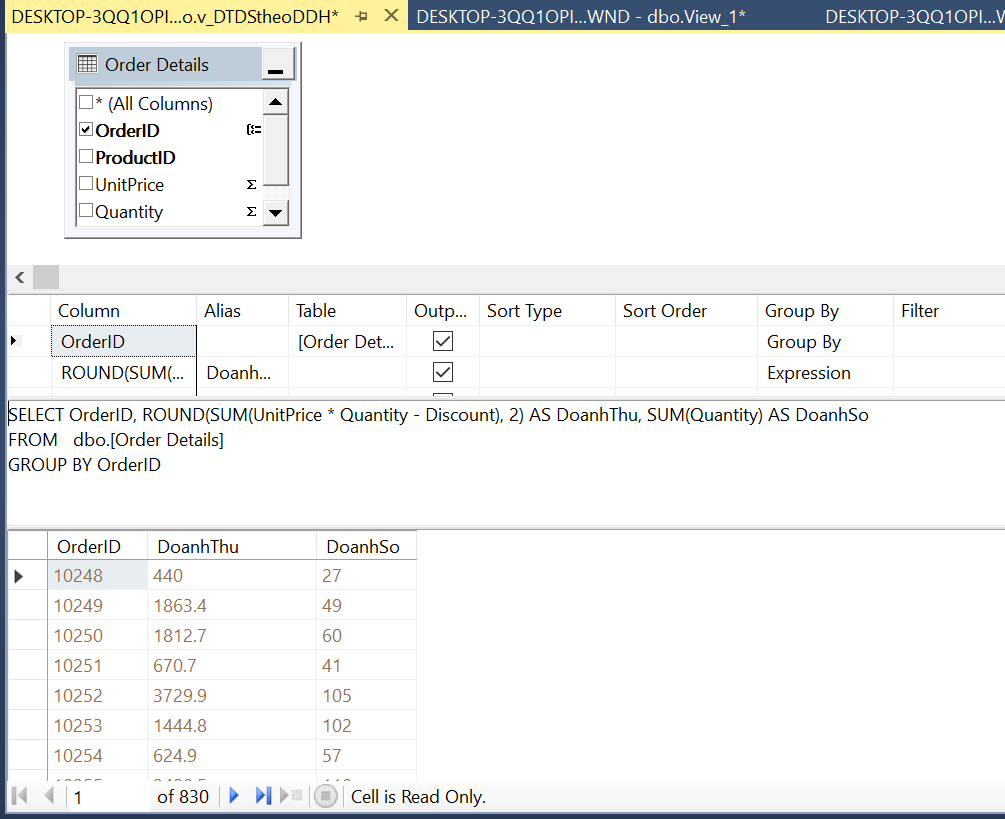
Select <tên cột muốn insert> from <tên bảng nguồn>

- Ở trong trường hợp này là:

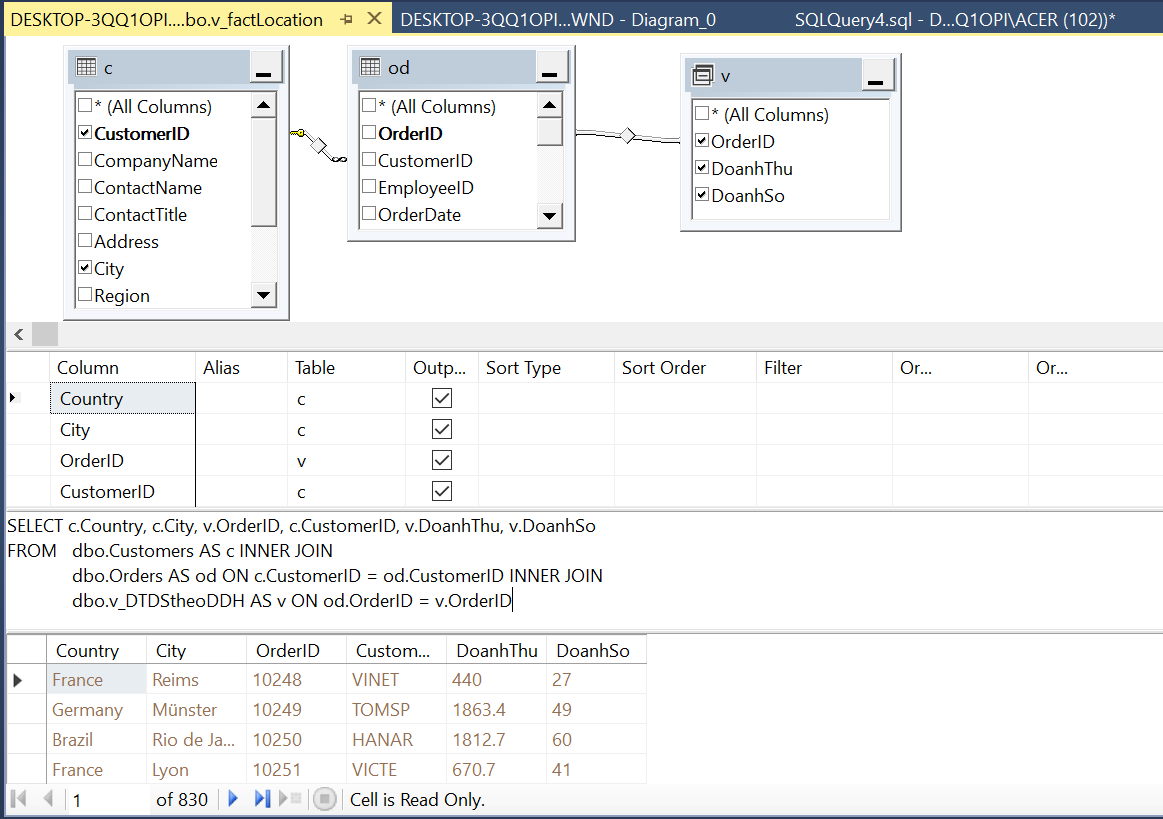


### **Thiết lập các Views tính toán cần thiết để nạp dữ liệu từ DB gốc vào các Factors của phân hệ DWH**

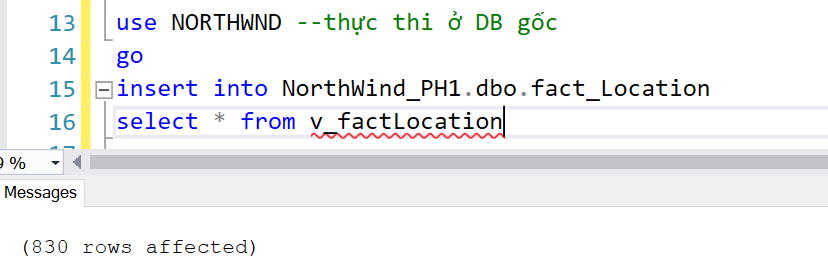
**-** Tạo view v\_DTDStheoDDH: Tính doanh thu, doanh số theo Đơn đặt hàng (orderID)



- Tạo view v\_factLocation để nạp dữ liệu vào fact:

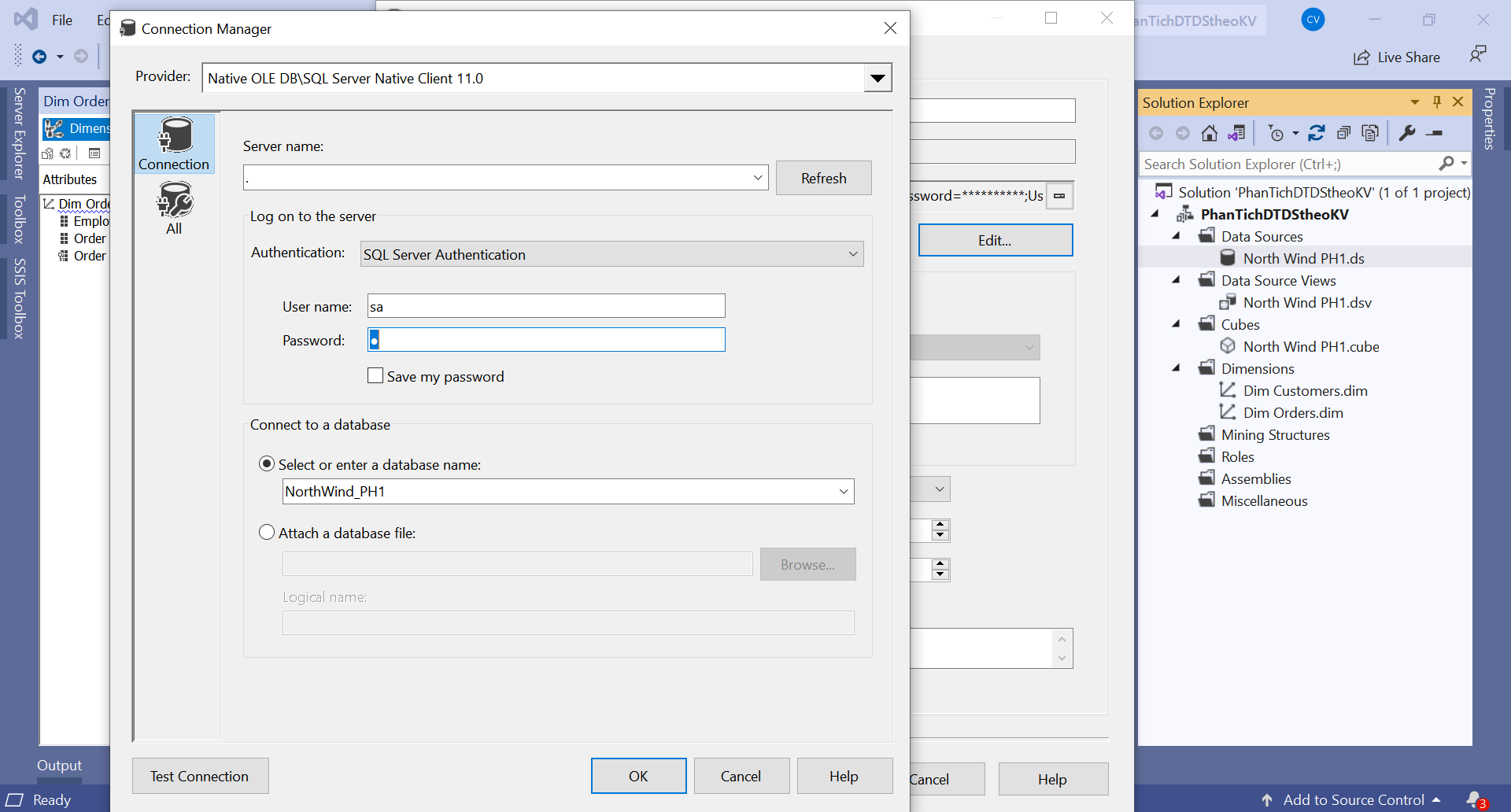


- Tương tự ở dim, ta lấy dữ liệu từ view ở trên insert vào bảng fact:

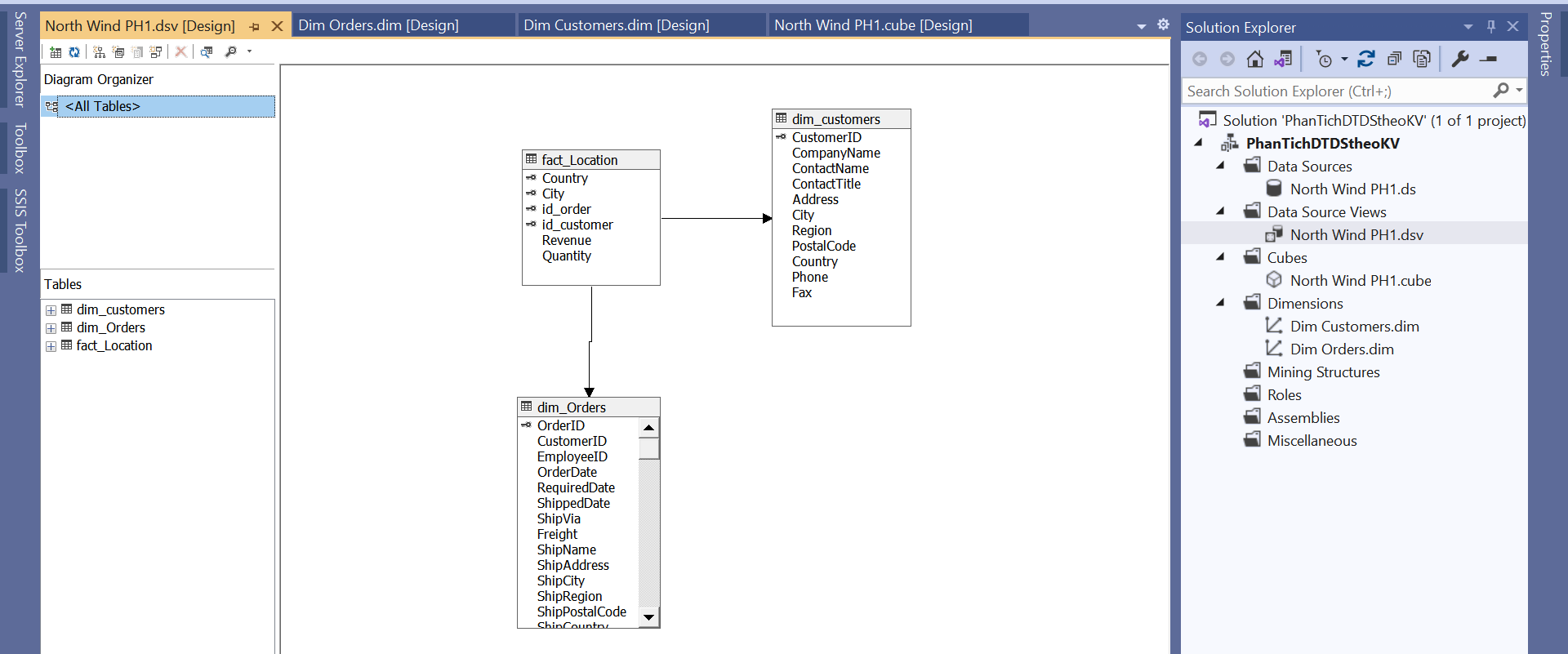


### **Triển khai phân hệ DWH lên BIDS**

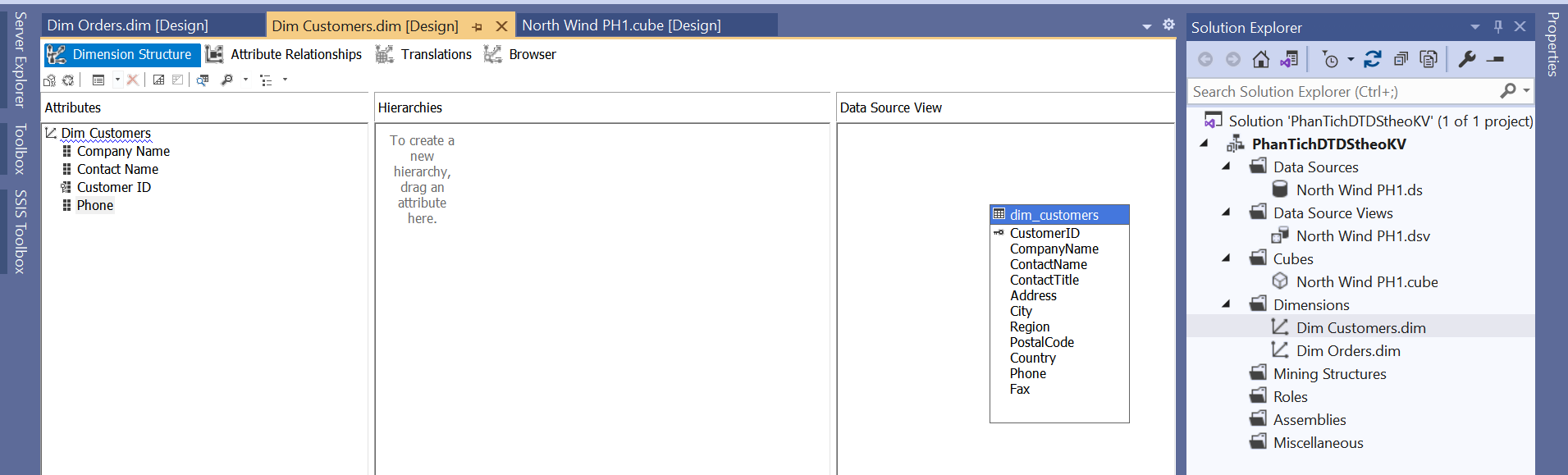
**-** Tạo DataSource:

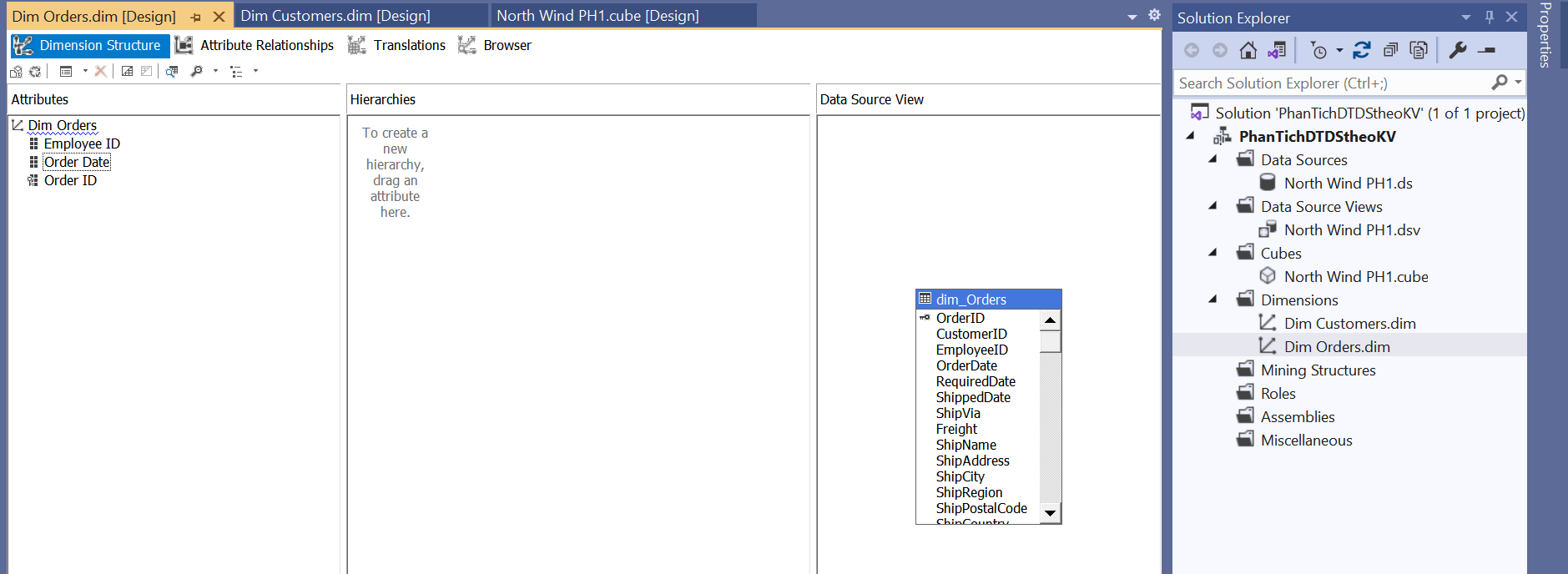


- Tạo view cho DataSource:

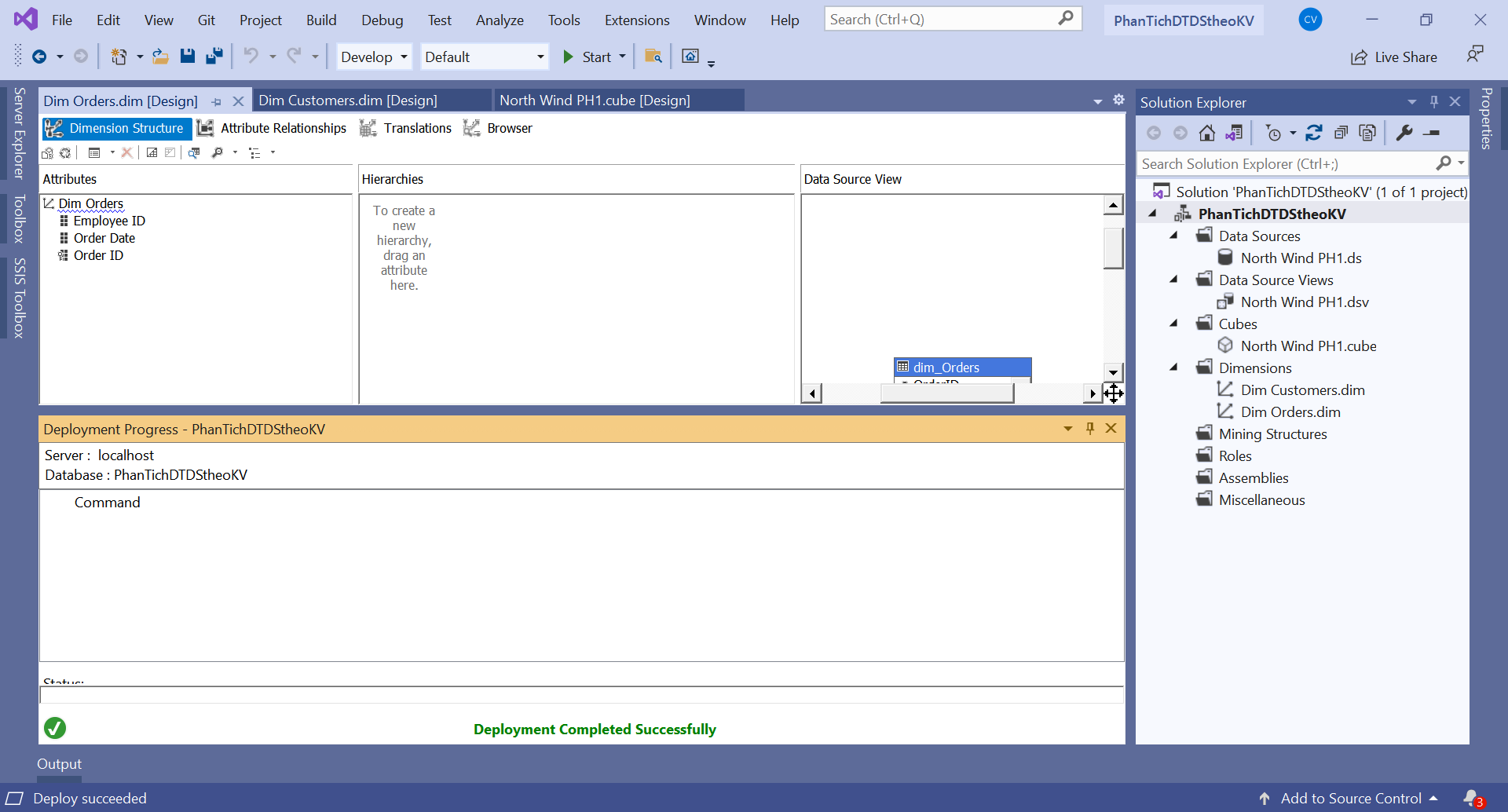


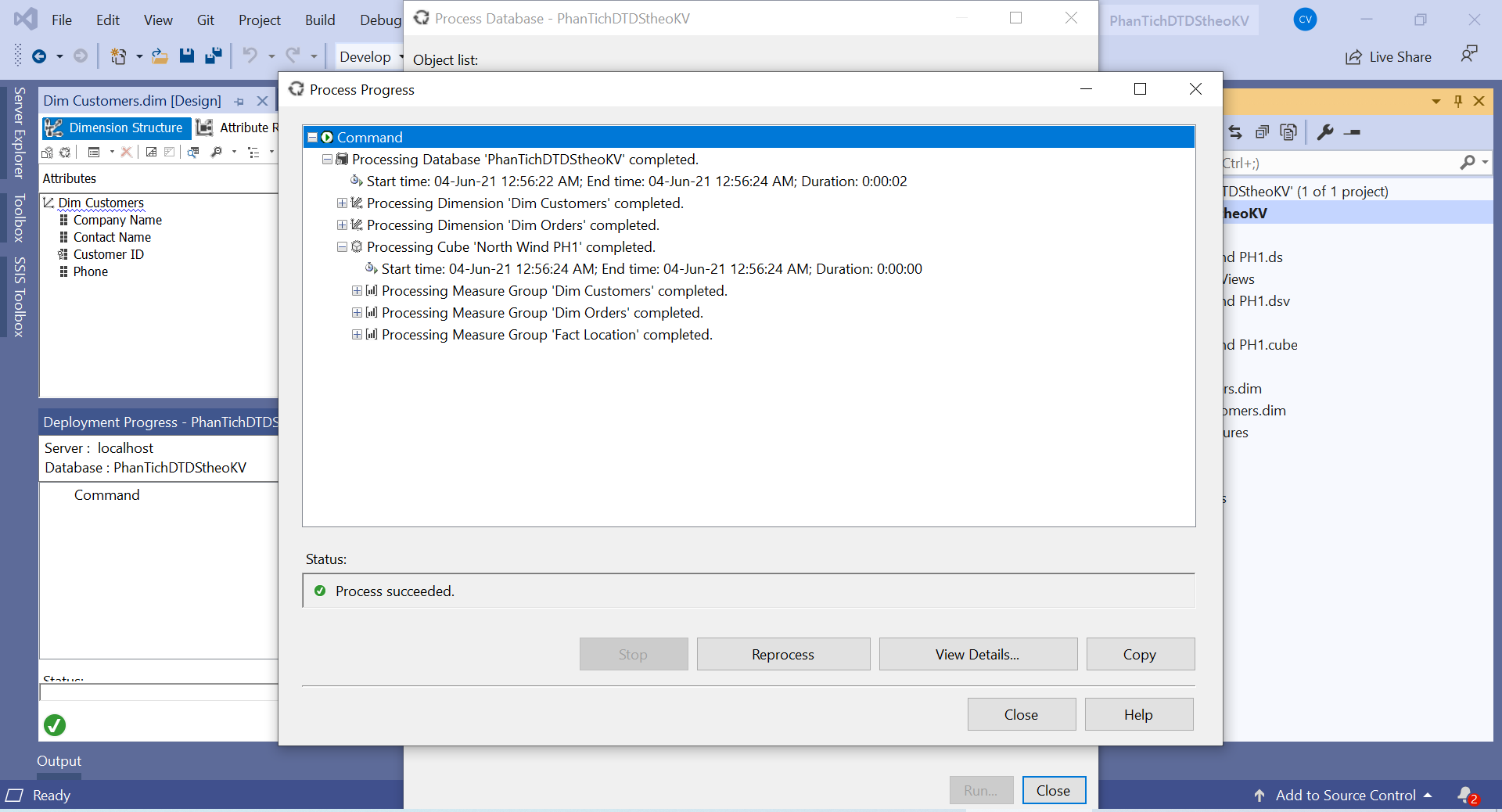
**-** Chọn các thuộc tính ở các dim để phân tích:





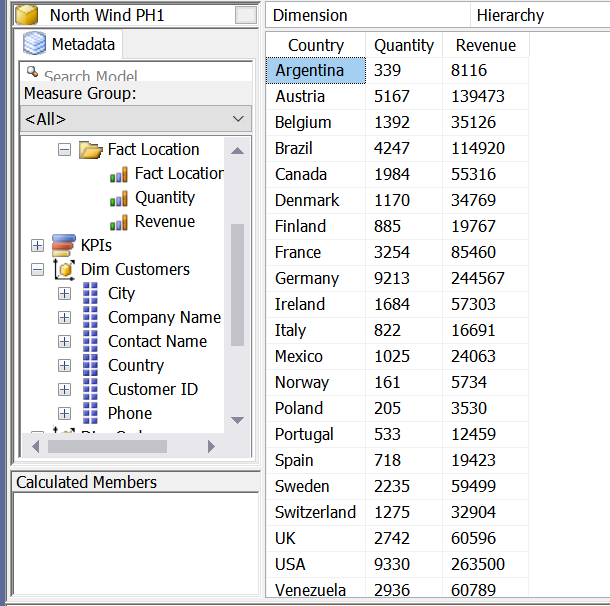
**-** Tiến hành Deloy và Process kho dữ liệu:



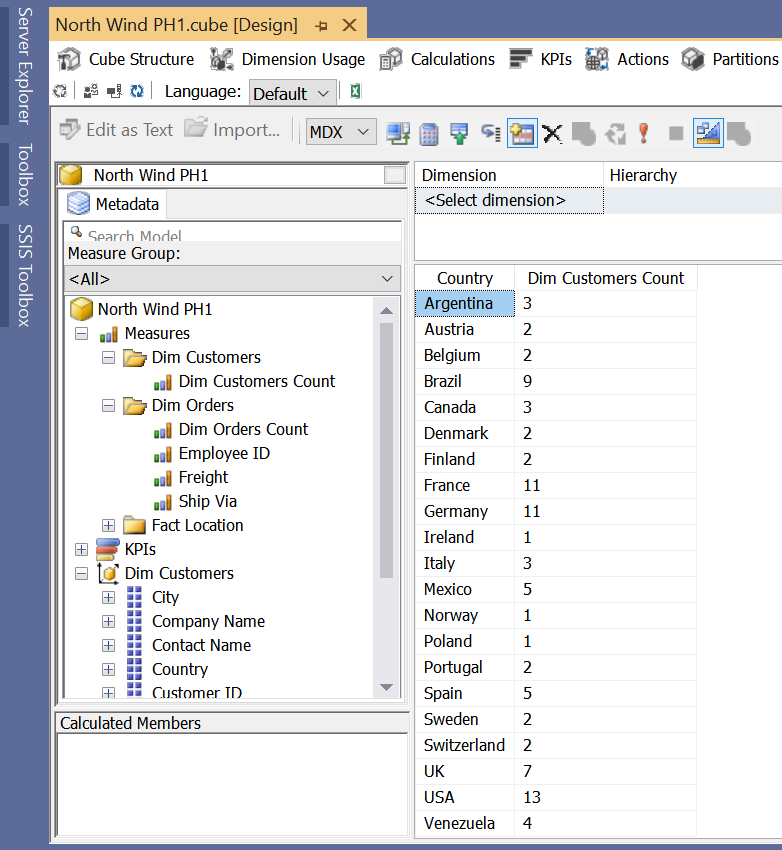


#### \*\*\*Phân tích số liệu trên Brower BIDS:

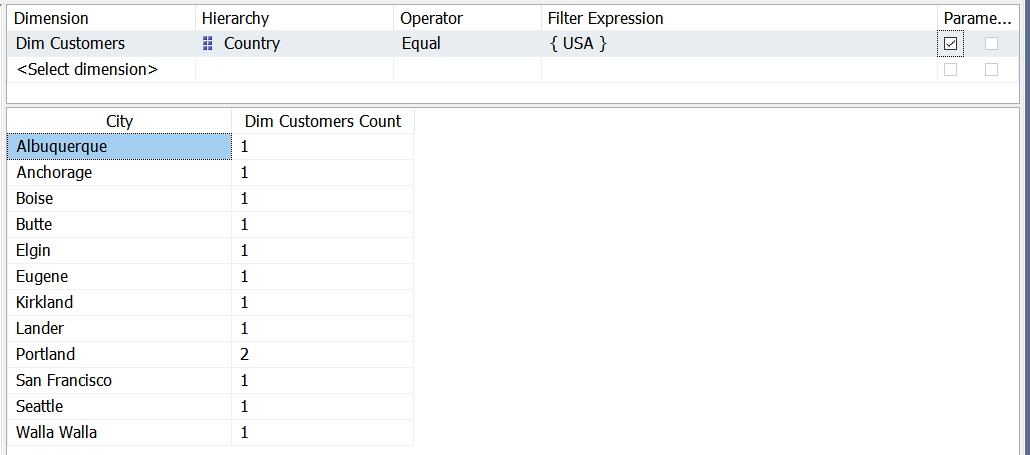
- Doanh thu, doanh số ở các khu vực:



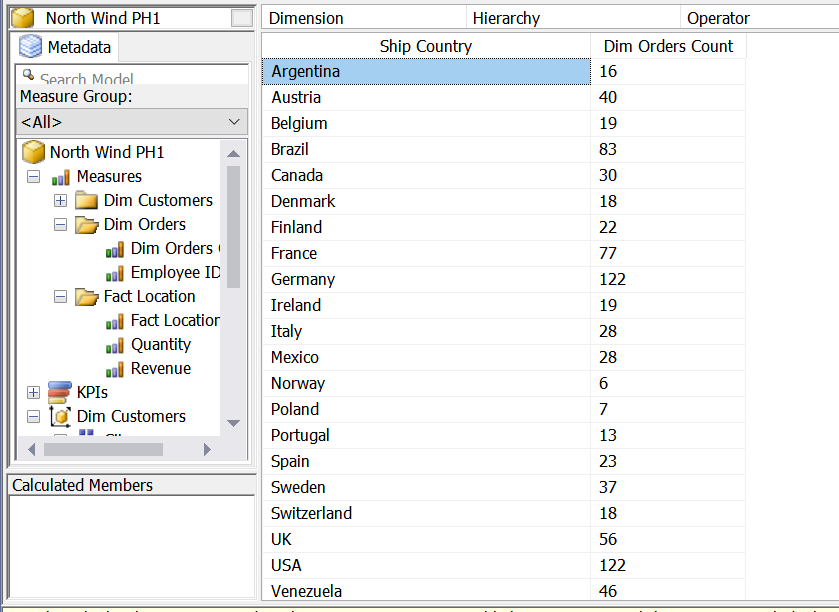
- Số lượng khách hàng ở các quốc gia:



- Số lượng khách hàng theo thành phố ở quốc gia “USA”-chiếm số khách hàng đông đảo ở công ty. Nhận xét: khách hàng phân bố khá đồng đều, riêng thành phố Poland nhỉn hơn một tí:

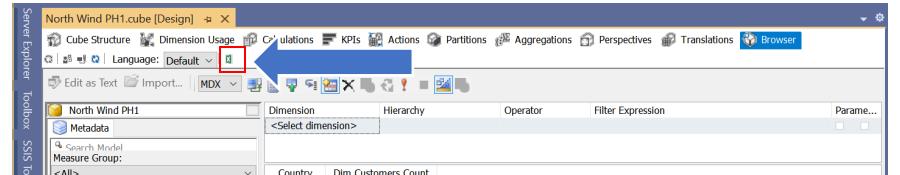


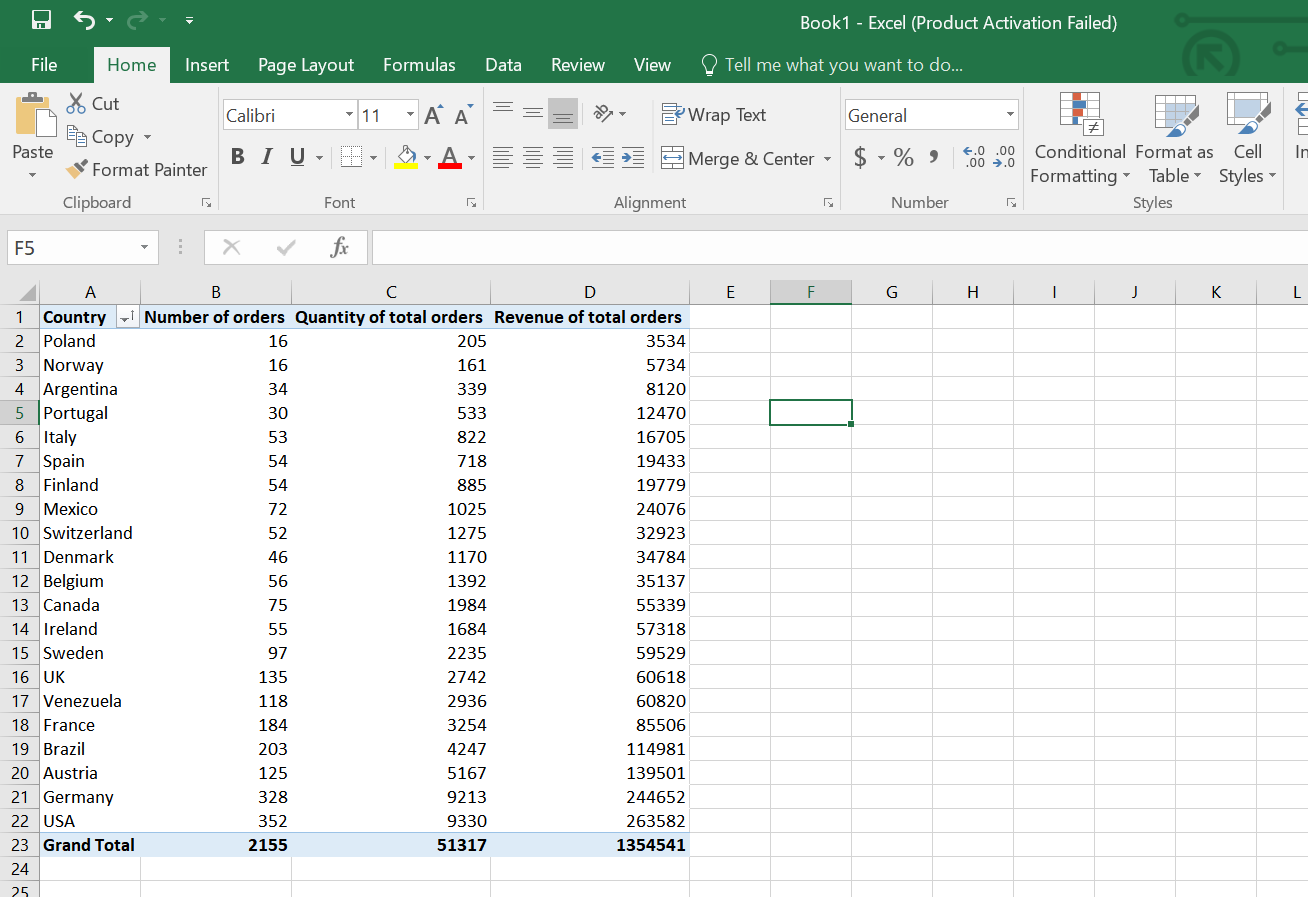
- Số lượng đơn đặt hàng theo quốc gia:



#### \*\*\*Phân tích trong Excel:

- Click vào biểu tượng này, chuyển hướng sang trang tính excel:





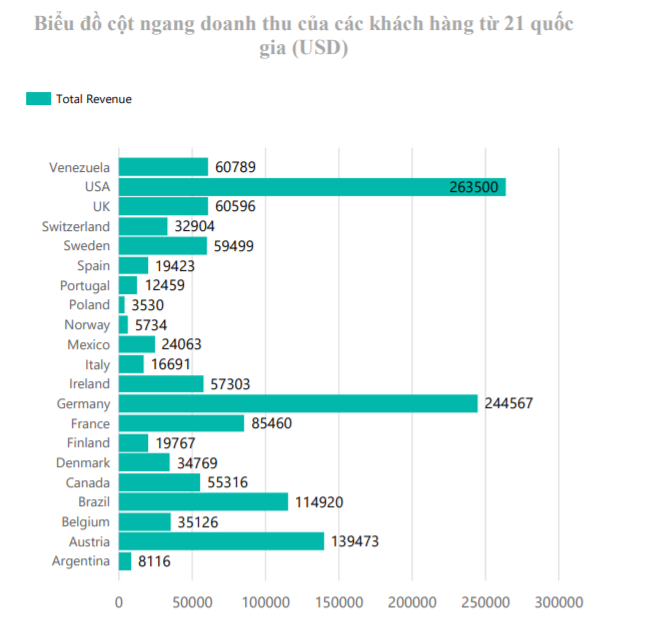
- 4 cột lần lượt là: tên quốc gia, số lượng đơn đặt hàng, lợi nhuận tổng đơn đặt hàng mang lại, tổng số lượng sản phẩm các đơn đặt hàng.

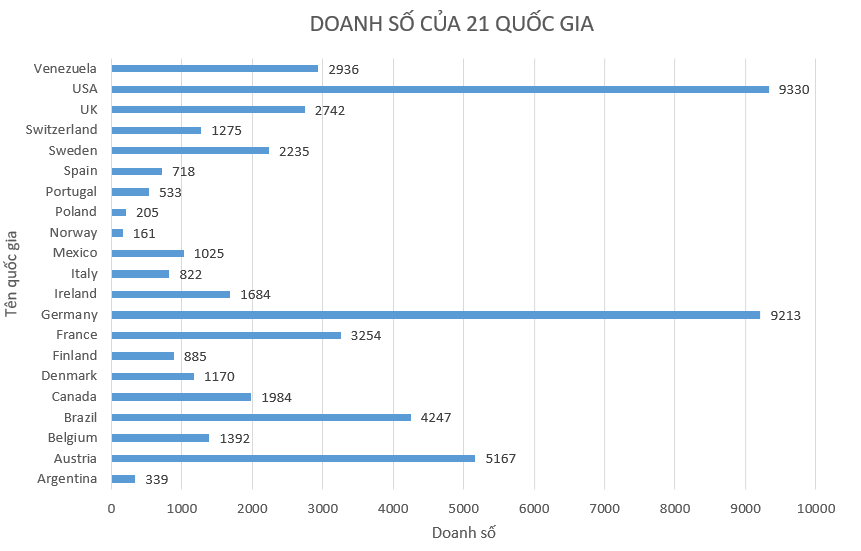
- Trong đó, Poland đặt hàng ít nhất cũng như mang lại nhỏ nhất. Ngược lại, USA thì lớn nhất. Điều này, dễ dàng giải thích do sự chênh lệnh phát triển kinh tế, với USA là một cường quốc lớn.

#### \*\*\*Phân tích bằng biểu đồ:

- Nhìn chung, “USA” và “Germerny”, vượt khá xa các nước còn lại.

- “Poland” và “Norway”, chiếm khoảng khá nhỏ.

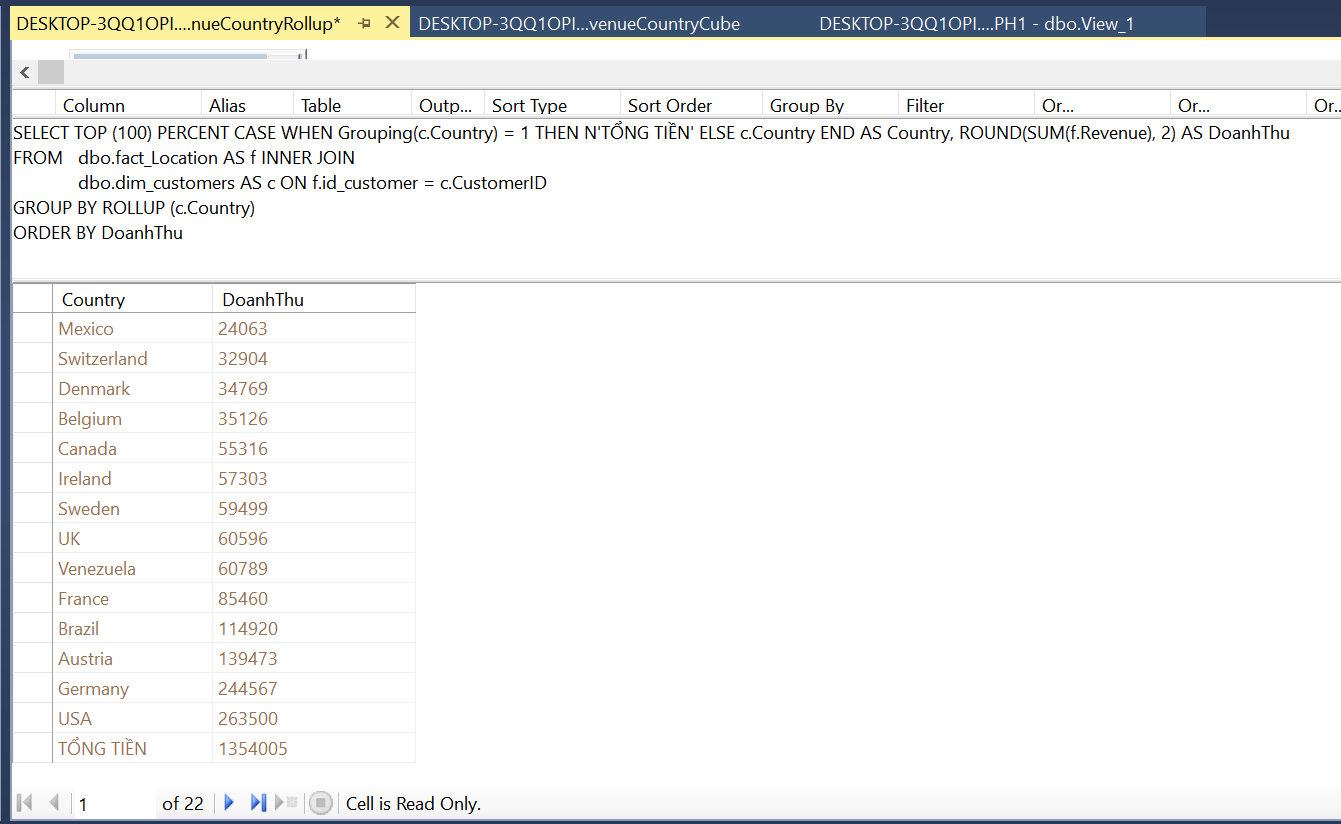




### **Khai thác sử dụng phân hệ DWH dùng các vấn tin SQL**

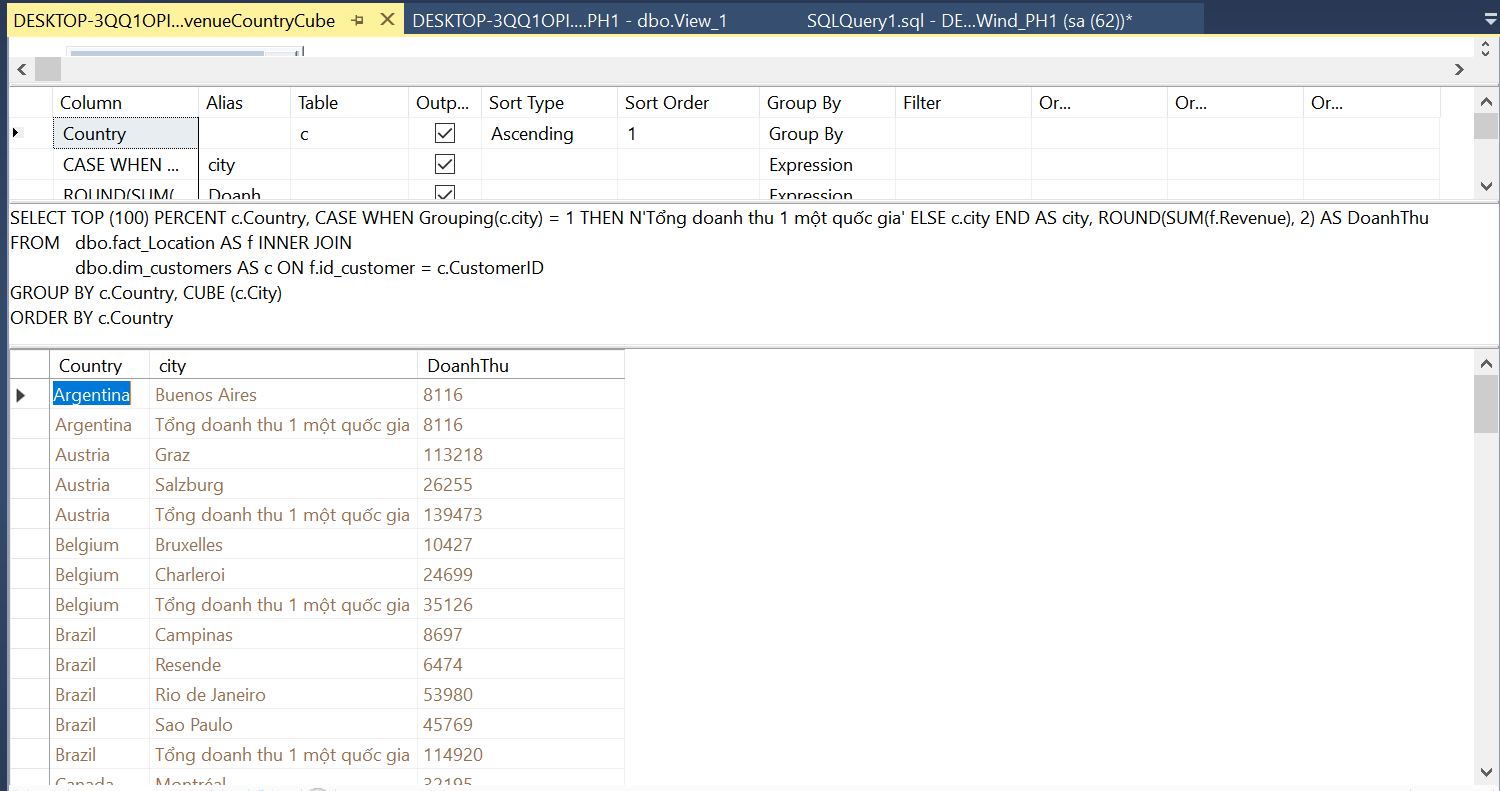
#### \*\*\*Sử dụng “ROLLUP”

- Lấy tổng doanh thu từng quốc gia và ra tổng doanh thu cuối cùng của cả công ty:



#### \*\*\*Sử dụng CUBE

**-** Lấy doanh thu theo 1 nhóm các thành phố ở mỗi quốc gia:

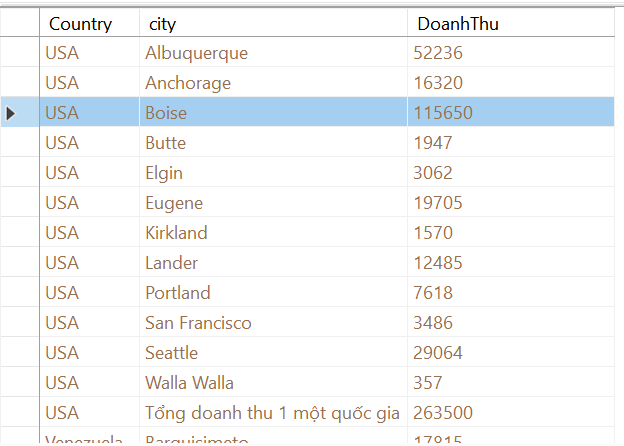


- Kết quả như nhau nếu dùng “group by Grouping Sets((c.Country, c.City),c.Country)”

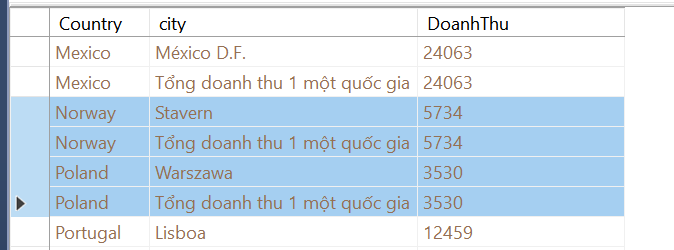
=> Nếu muốn thống kê doanh số, chỉ cần thay biến “Revenue” thành “Quantity”.

=> Nhận xét chung:

- “USA” chiếm phần nhiều doanh thu cũng như doanh số nhiều nhất, tập trung chủ yếu ở thành phố “Boise”. Trong tương lai, công ty nên có những ưu đãi với khu vực khách hàng tiềm năng này.

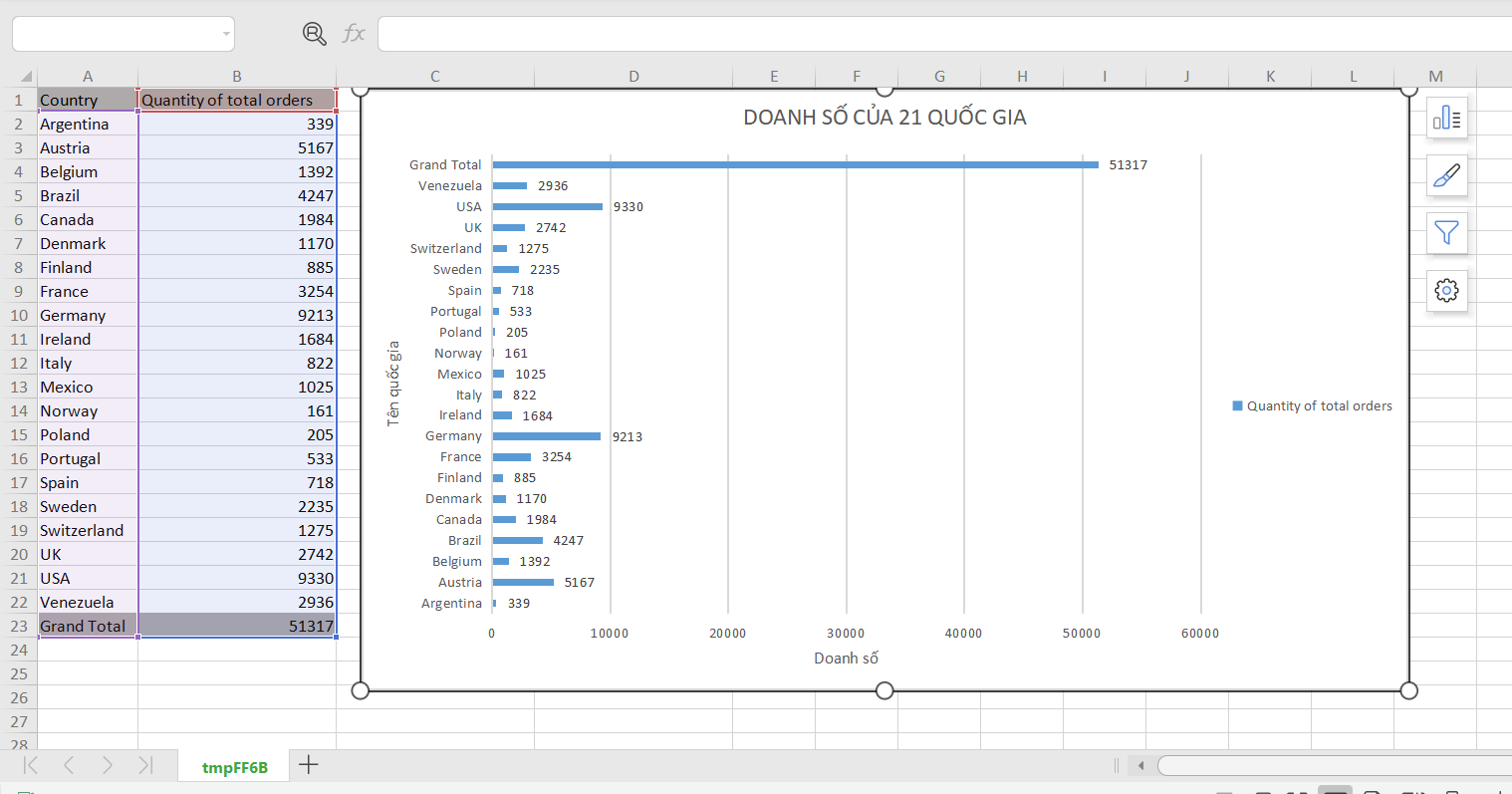


- Bên cạnh đó, thì ở một số quốc gia, doanh thu còn hạn chế.

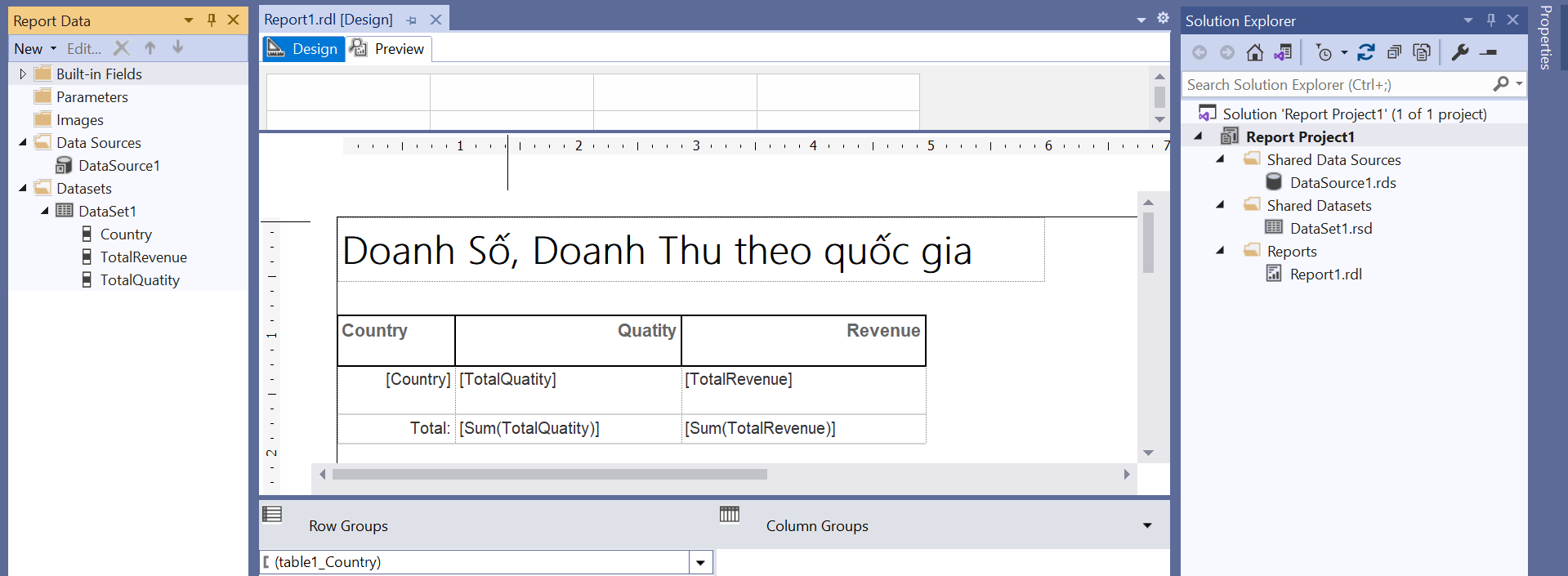


### Trực quan hóa dữ liệu DWH:

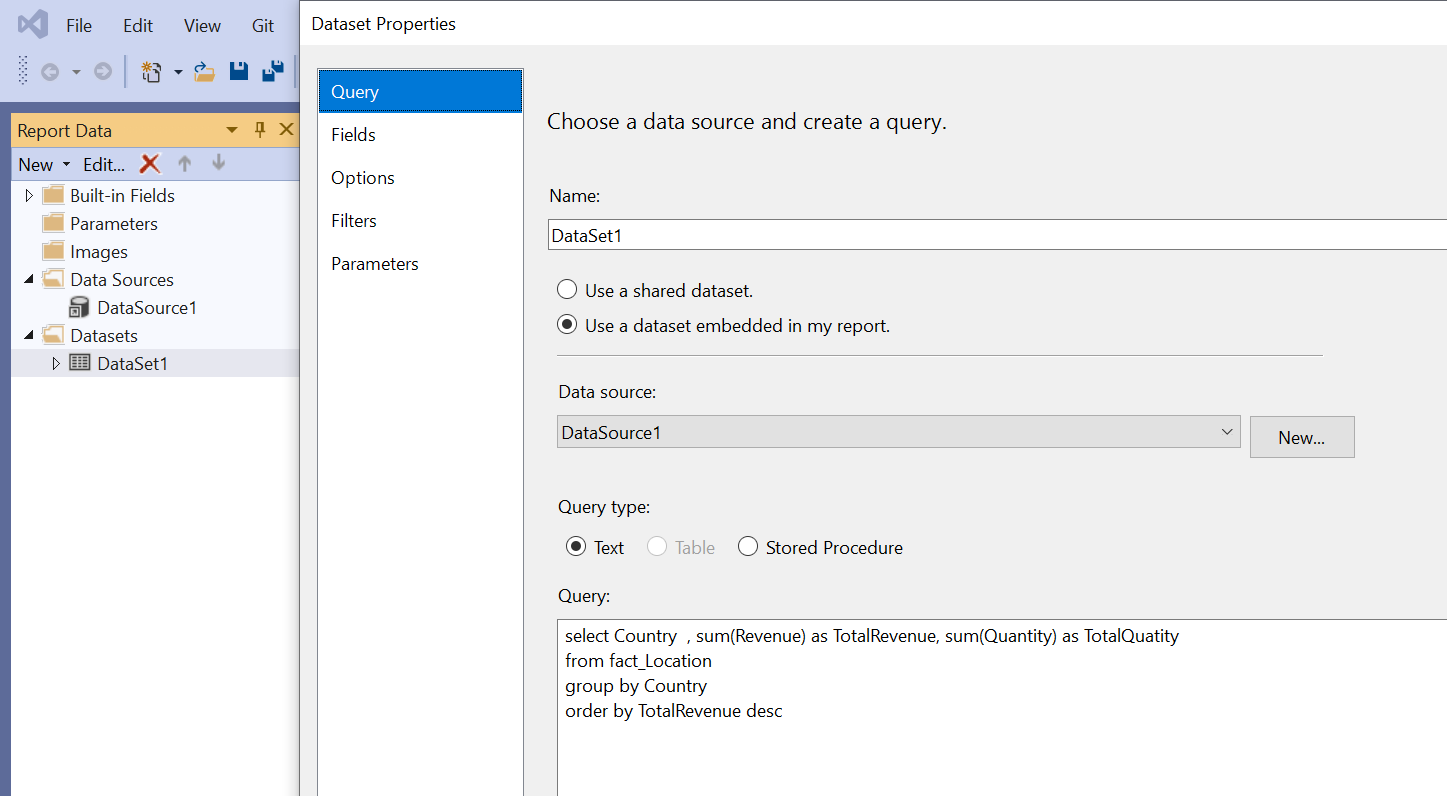
#### \*\*\*Sử dụng Microsoft Excel:



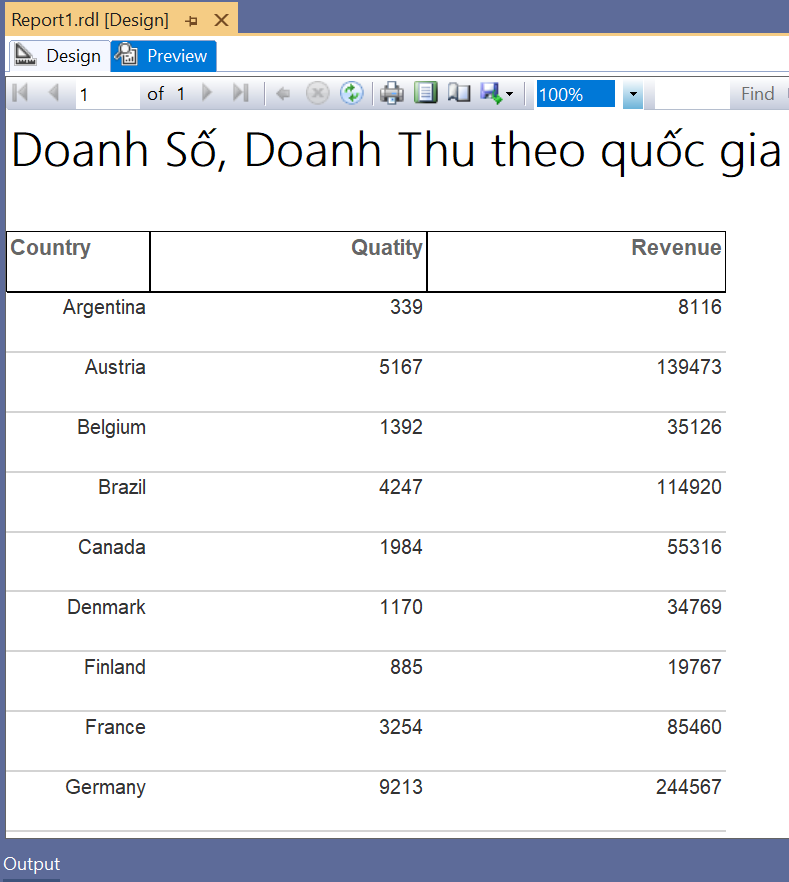
#### \*\*\*Sử dụng Microsoft Reporting Services Project:

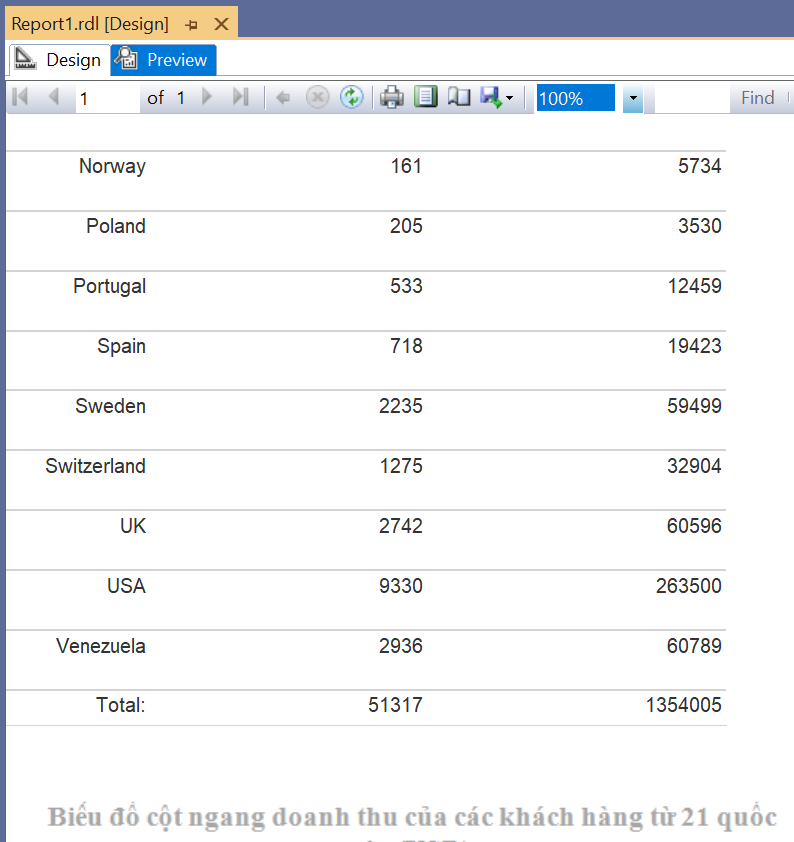


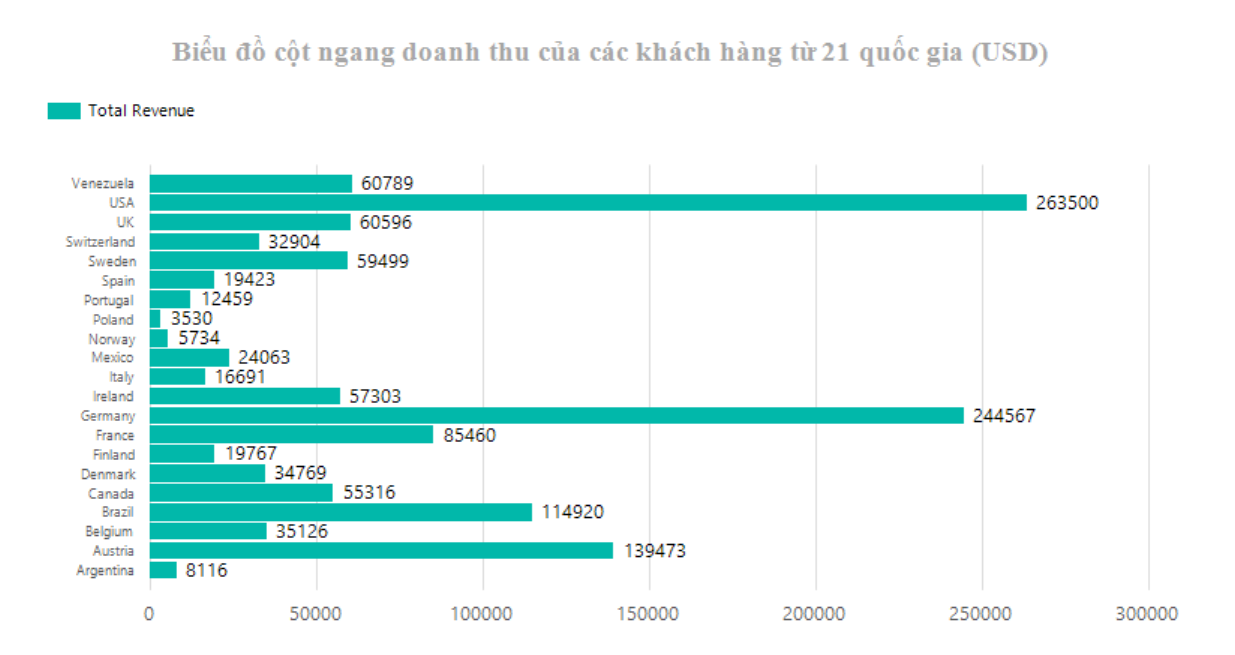
- Sau khi thiết lập DataSource, tiếp theo tùy chọn vùng DataSet để trực quan dữ liệu bằng câu lệnh SQL:



- Cuối cùng, xuất ra được báo cáo và biểu đồ tương ứng:

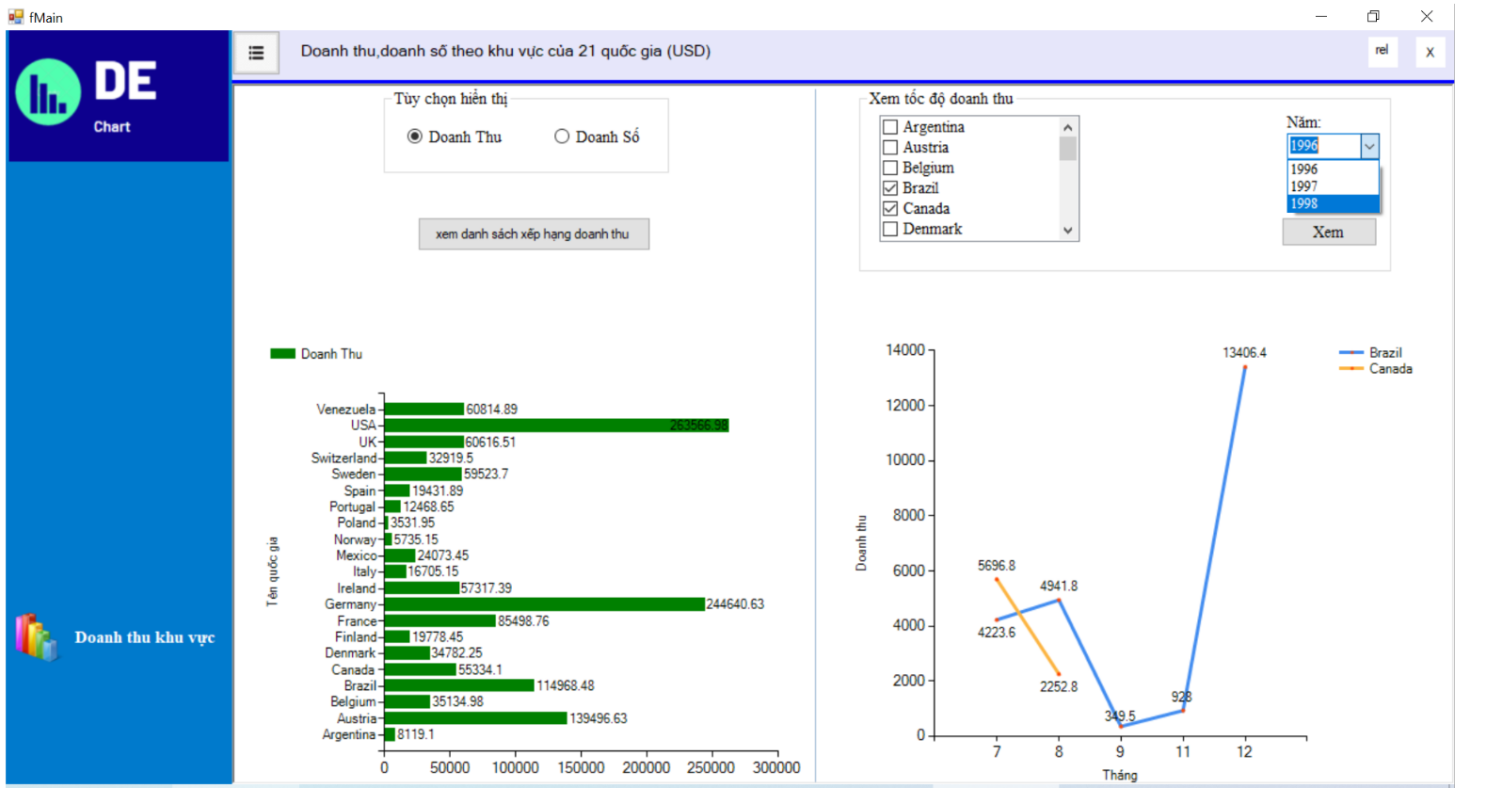






#### \*\*\*Sử dụng Windows Form App C#:

- Ở đây có thể tùy chọn hiểu thị doanh thu với thao tác dễ dàng với những quốc gia và năm mình muốn :



# KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN CỦA ĐỒ ÁN HỌC PHẦN

## Kết luận

### Những kết quả đạt được

- Thiết kế được mô hình kho dữ liệu cho công ty xuất nhập khẩu xuyên quốc gia.

- Sử dụng công cụ SSIS để tích hợp và chuyển đổi dữ liệu từ cơ sở dữ liệu nguồn vào kho dữ liệu đã thiết kế.

- Tạo view để tích hợp dữ liệu từ cơ sở dữ liệu nguồn vào bảng Fact trong DWH

- Dùng các câu lệnh vấn tin để tổng hợp doanh số và doanh thu theo thời gian.

- Dùng công cụ BIDS để khai thác sử dụng nhà kho nhằm đưa ra những thống kê về doanh thu, doanh số,… của công ty theo thời gian và theo đơn hàng.

### Hạn chế

- Chưa tận dụng được hết các chức năng của BIDS

- Có lỗi kĩ thuật khi thực hiện tích hợp dữ liệu từ cơ sở dữ liệu nguồn vào DWH nhưng đã khắc phục được.

- Dữ liệu nguồn chưa lớn như mong muốn ban đầu.

## Hướng phát triển

### Hướng khắc phục các hạn chế

- Tìm hiểu thêm và các chức năng cũng như luyện tập nhiều hơn để có thể sử dụng hết các tính năng nổi bật và cần thiết của BIDS.

- Nắm rõ lí thuyết bài học, thực hành thêm để có thể nắm và hiểu rõ được cách hoạt động của DWH.

- Tiếp tục thu thập và mở rộng dữ liệu nguồn.

### Hướng mở rộng ĐỒ ÁN HỌC PHẦN

- Sử dụng các công cụ để trực quan hóa dữ liệu nhằm thuận tiện cho người xem.

- Mở rộng nghiên cứu và tích hợp thêm các thuộc tính mới về khách hàng , đối tác giao hàng để có những chiến lược và chiến thuật kinh doanh phù hơn cho doanh nghiệp.

# DANH MỤC TÀI LIỆU THAM KHẢO

1. Võ Xuân Thể (2021), *Bài giảng học phần Kho dữ liệu*, TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT TP.HCM (lưu hành nội bộ).

# ĐÁNH GIÁ KẾT QUẢ HỌC TẬP HỌC PHẦN

ĐỒ ÁN HP THAY CHO BÀI THI KẾT THÚC HỌC PHẦN

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Hình thức KT** | **Nội dung** | **Chuẩn đầu ra đánh giá** | **Trình độ năng lực** | **Phương pháp đánh giá** | **Công cụ đánh giá** | **Tỉ lệ (%)** |
| **BÀI TẬP LẬP TRÌNH** | | | | | |  |
| BL#1 | Bài tập lập trình theo từng Buổi thực hành = tính bình quân  Kiểm tra viết | G1.1 | 3 | Bài tập  Kiểm tra | Bài tập  Kiểm tra | 50 |
| G2.1 | 3 |
| G3.1  G3.2 | 4  3 |
| G4.1 | 5 |
| **ĐỒ ÁN HỌC PHẦN (Project)** | | | | | |  |
| ĐA#1 | Phát triển một hệ thống KHO DỮ LIỆU đơn giản:  + Đề tài Nhóm có phân công phân hệ cụ thể cho từng thành viên thuộc nhóm  + Xây dựng và thực nghiệm Hệ thống Kho dữ liệu theo nhóm và phân hệ cá nhân  => báo cáo vào ngày thi. | G1.1 | 3 | Báo cáo, Demo và thuyết trình | Rubric | 50 |
| G2.1 | 3 |
| G3.1 | 4 |
| G3.2  G4.1 | 3  5 |
| G2.1 | 4 |
| G3.1 | 4 |
| G3.2  G4.1 | 3  5 |