**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 online**

Lớp: **181330A, 181330B**

**NHÓM : 4**

THÀNH VIÊN: 1.Nguyễn Văn A [stt]\_NT; 2. Nguyễn Thị B [stt] . . . .

. . . .

{bản mềm: nhóm trưởng để nguyên ds sau; bản mềm các thành viên: chỉ để tên +… của mình[đúng số tt/nhóm], Xóa các thành viên khác}

**Ngày nộp: …../06/2021**

**TP.HCM, ngày 15 Tháng 06 năm 2021**

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

**………………………………………………………………..**

**………………………………………………………………..**

**……………………………………………………………….**

**NHÓM : 4**

**ĐIỂM ĐÁNH GIÁ**

{bản mềm: nhóm trưởng để nguyên ds sau; bản mềm các thành viên: chỉ để tên +… của mình[đúng số tt/nhóm], Xóa các thành viên khác: NẾU DS DÀI = Trang riêng: 1 lần :BỎ DS BÌA & BÌA LÓT}

1. Nguyễn Văn A [stt]: **:………….(…………………)………..**
2. Nguyễn Thị B [stt]: **:………….(…………………)………….**
3. . . . .

**Ngày ……../06/2021**

**Giảng viên Ký tên**

THÀNH VIÊN:

{bản mềm: nhóm trưởng để nguyên ds sau; bản mềm các thành viên: chỉ để tên +… của mình[đúng số tt/nhóm], Xóa các thành viên khác}

1. Nguyễn Văn A [stt]\_Nhóm trưởng: Thành phần kho dữ liệu nhân sự - tiền lương
2. Nguyễn Thị B [stt]: Thành phần kho dữ liệu bán hàng (Dims, Facts)
3. . . . .

**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 online**

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

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

**Tên sản phẩm đề tài: 50.N4.01.PhanThanhTrung.DAHP.DWH.rar** (Microsoft SQL Server 2019)

**DB Nguồn: 50N4PhanThanhTrungSaleDADWH**

**DB DWH: 50N4PhanThanhTrungSaleDWH**

**Tên máy chủ DB SQL: .**

**Tên đăng nhập DB SQL: sa**

**Mật khẩu (password) đăng nhập SQL: thanhtrung1256**

**ProjectName BIDS: 50N4PhanThanhTrungQLKCBBHYT\_DWH**

**NHÓM : 4**

THÀNH VIÊN: 1.Nguyễn Văn A [stt]\_NT; 2. Nguyễn Thị B [stt] . . . .

. . . .

{bản mềm: nhóm trưởng để nguyên ds sau; bản mềm các thành viên: chỉ để tên +… của mình[đúng số tt/nhóm], Xóa các thành viên khác}

**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](#_Toc73666776)

[DANH MỤC CÁC TỪ VIẾT TẮT 5](#_Toc73666777)

[DANH MỤC CÁC THUẬT NGỮ 6](#_Toc73666778)

[MỤC LỤC 7](#_Toc73666779)

[Chương 1. GIỚI THIỆU VỀ ĐỒ ÁN HỌC PHẦN 10](#_Toc73666780)

[1.1. Tổng quan về ĐỒ ÁN HỌC PHẦN 10](#_Toc73666781)

[1.2. Nội dung chuyên môn chính của ĐỒ ÁN HỌC PHẦN 10](#_Toc73666782)

[1.3. Mục tiêu của ĐỒ ÁN HỌC PHẦN 10](#_Toc73666783)

[1.4. Công cụ và nền tảng kỹ thuật thực hiện ĐỒ ÁN HỌC PHẦN 10](#_Toc73666784)

[1.5. Sản phẩm của ĐỒ ÁN HỌC PHẦN 10](#_Toc73666785)

[1.6. Bố cục của báo cáo 10](#_Toc73666786)

[Chương 2. CƠ SỞ LÝ THUYẾT CỦA HỌC PHẦN 11](#_Toc73666787)

[2.1. Tổng quan về kho dữ liệu (DWH) 11](#_Toc73666788)

[2.1.1. Các khái niệm cơ bản về Kho dữ liệu 11](#_Toc73666789)

[2.1.2. Kiến trúc cơ bản của kho dữ liệu: 11](#_Toc73666790)

[2.1.3. Mô hình CSDL đa chiều (Dimensional Modeling) 11](#_Toc73666791)

[2.2. Phân tích và thiết kế Kho dữ liệu 11](#_Toc73666792)

[2.2.1. Phân tích và thiết kế kho dữ liệu theo hướng mô hình hóa nhiều chiều (Dimensional Modeling) 11](#_Toc73666793)

[2.2.2. Các dạng (phương pháp) tích hợp dữ liệu 11](#_Toc73666794)

[2.2.3. Các bước cơ bản trong tiến trình tích hợp dữ liệu 11](#_Toc73666795)

[2.2.4. Các công cụ tích hợp dữ liệu: 11](#_Toc73666796)

[2.2.5. Lập các luồng công việc tích hợp dữ liệu (Data Integration Workflows) 11](#_Toc73666797)

[2.2.6. Hệ thống OLTP (On-Line Transaction Processing = Xử lý giao dịch trực tuyến) 11](#_Toc73666798)

[2.3. Vấn tin khai thác sử dụng Kho dữ liệu 11](#_Toc73666799)

[2.3.1. Vấn tin với các phép toán tổng hợp của SQL trên kho dữ liệu 11](#_Toc73666800)

[2.3.2. Vấn tin với các hàm phân tích của SQL trên kho dữ liệu 11](#_Toc73666801)

[2.3.3. Vấn tin dùng Materialized View để phân tích trên kho dữ liệu 11](#_Toc73666802)

[2.3.4. Hệ thống OLAP 11](#_Toc73666803)

[Chương 3. TỔNG QUAN VỀ KHO DỮ LIỆU …<theo tên đề tài>…… TẠI ……<theo tên đề tài>………… 12](#_Toc73666804)

[3.1. Ý tưởng hình thành DWH …<theo tên đề tài>…… tại ……<theo tên đề tài>………… 12](#_Toc73666805)

[3.2. Giới thiệu chung về Kho dữ liệu 12](#_Toc73666806)

[3.2.1. Mô tả tổng quan về DWH 12](#_Toc73666807)

[3.2.2. Các dữ liệu nguồn hình thành DWH 12](#_Toc73666808)

[3.3. Xác định nhu cần tổ chức và phân tích dữ liệu của toàn bộ DWH 12](#_Toc73666809)

[3.4. Các thành phần phân hệ Kho dữ liệu phân công các cá nhân phụ trách 12](#_Toc73666810)

[Chương 4. PHÂN TÍCH, THIẾT KẾ VÀ KHAI THÁC SỬ DỤNG CÁC PHÂN HỆ KHO DỮ LIỆU …<theo tên đề tài>…… TẠI ……<theo tên đề tài>………… 13](#_Toc73666811)

[4.1. <STT>\_ Ho Tên 1\_ <thành phần kho dữ liệu đươc phân công> 13](#_Toc73666812)

[4.1.1. Mô tả tổng quan về phân hệ DWH…<tên phân hệ DWH>… 13](#_Toc73666813)

[4.1.2. Xác định nguồn dữ liệu tích hợp phân hệ Kho dữ liệu 13](#_Toc73666814)

[4.1.3. Xác định ý tưởng hình thành phân hệ của Kho dữ liệu 13](#_Toc73666815)

[4.1.4. Mô tả hệ thống các DB gốc liên quan phân hệ DWH 13](#_Toc73666816)

[4.1.5. Phân phân tích các DB gốc xác định yêu cầu phân tích DWH của phân hệ 13](#_Toc73666817)

[4.1.6. Thiết kế DB mới tổ chức phân tích Kho dữ liệu cho phân hệ 14](#_Toc73666818)

[4.1.7. Thực hiện các thủ tục tích hợp các DB gốc vào phân hệ DWH 15](#_Toc73666819)

[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 15](#_Toc73666820)

[4.1.9. Triển khai phân hệ DWH lên BIDS 15](#_Toc73666821)

[4.1.10. Khai thác sử dụng phân hệ DWH dùng các vân tin SQL 15](#_Toc73666822)

[4.2. <STT>\_ Ho Tên 2\_ <thành phần kho dữ liệu đươc phân công> 15](#_Toc73666823)

[Chương 5. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN CỦA ĐỒ ÁN HỌC PHẦN 16](#_Toc73666824)

[5.1. Kết luận 16](#_Toc73666825)

[5.1.1. Những kết quả đạt được 16](#_Toc73666826)

[5.1.2. Hạn chế 16](#_Toc73666827)

[5.2. Hướng phát triển 16](#_Toc73666828)

[5.2.1. Hướng khắc phục các hạn chế 16](#_Toc73666829)

[5.2.2. Hướng mở rộng ĐỒ ÁN HỌC PHẦN 16](#_Toc73666830)

[DANH MỤC TÀI LIỆU THAM KHẢO 17](#_Toc73666831)

[CÁC PHỤ LỤC 18](#_Toc73666832)

[ĐÁNH GIÁ KẾT QUẢ HỌC TẬP HỌC PHẦN 19](#_Toc73666833)

# 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 nhứng 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:

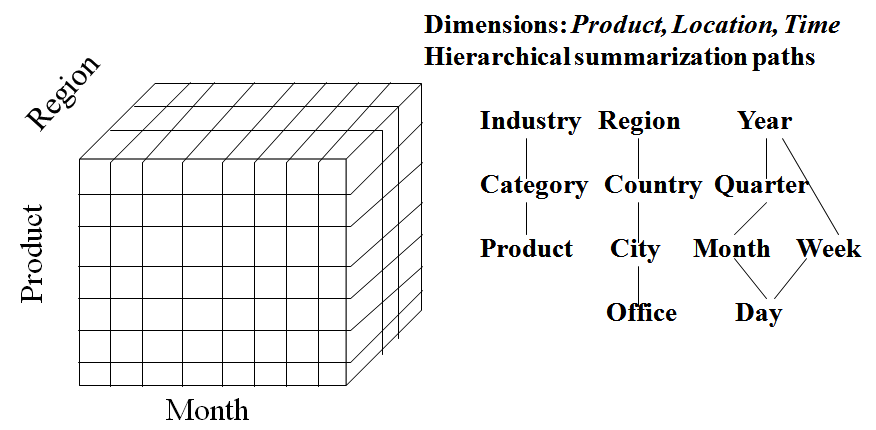
#### Tầng đáy: nạp dữ liệu

#### 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

## **4**

**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

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

### Mô tả tổng quan về phân hệ DWH Phân tích doanh thu và doanh số theo nhân viên…

Phân hệ bao gồm 3 dim : dim date , dim employee , dim product và 1 bảng fact

### 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

Trong kinh doanh, doanh thu luôn là yếu tố được quan tâm hàng đầu vì nó trực tiếp tạo ra lợi nhuận. Theo đó mức doanh thu, doanh số của từng cá nhân nhân viên luôn được người chủ doanh nghiêp chú ý cân nhắc vì nó cấu thành chung vào doanh thu toàn doanh nghiệp. Vì vậy, việc phân tích mức doanh thu, doanh số trên từng nhân viên là một việc quan trọng để có thể đánh giá và đề xuất mức hợp lí cho mỗi cá nhân

### Mô tả hệ thống các DB gốc liên quan phân hệ DWH

* Bảng Employees(nhân viên): bảng nhân viên bao gồm thông tin nhân viên như họ tên , nơi ở cũng như id người quản lý của nhân viên đó
* Bảng Oders : bảng lưu trữ thông tin các đơn đặt hàng gồm các thông tin về id người mua , id người nhân viên nhận hàng
* Bảng product : chi tiết sản phẩm gồm thông tin sản phẩm và tên sản phẩm,..
* Bảng chi tiết đơn hàng (order\_items): chi tiết đơn hàng gồm thông tin sản phẩm và số lượng sản phẩm lẫn tỉ lệ giảm giá nhằm để biết doanh thu theo sản phẩm.

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

* Dim\_employee: Dim về nhân viên được lấy từ bảng nhân viên(employee) ở DB gốc.
* Đim\_product: Đim về thông tin sản phẩm được lấy từ bảng product
* Dim\_Date: dim về thời gian giao dịch mua bán

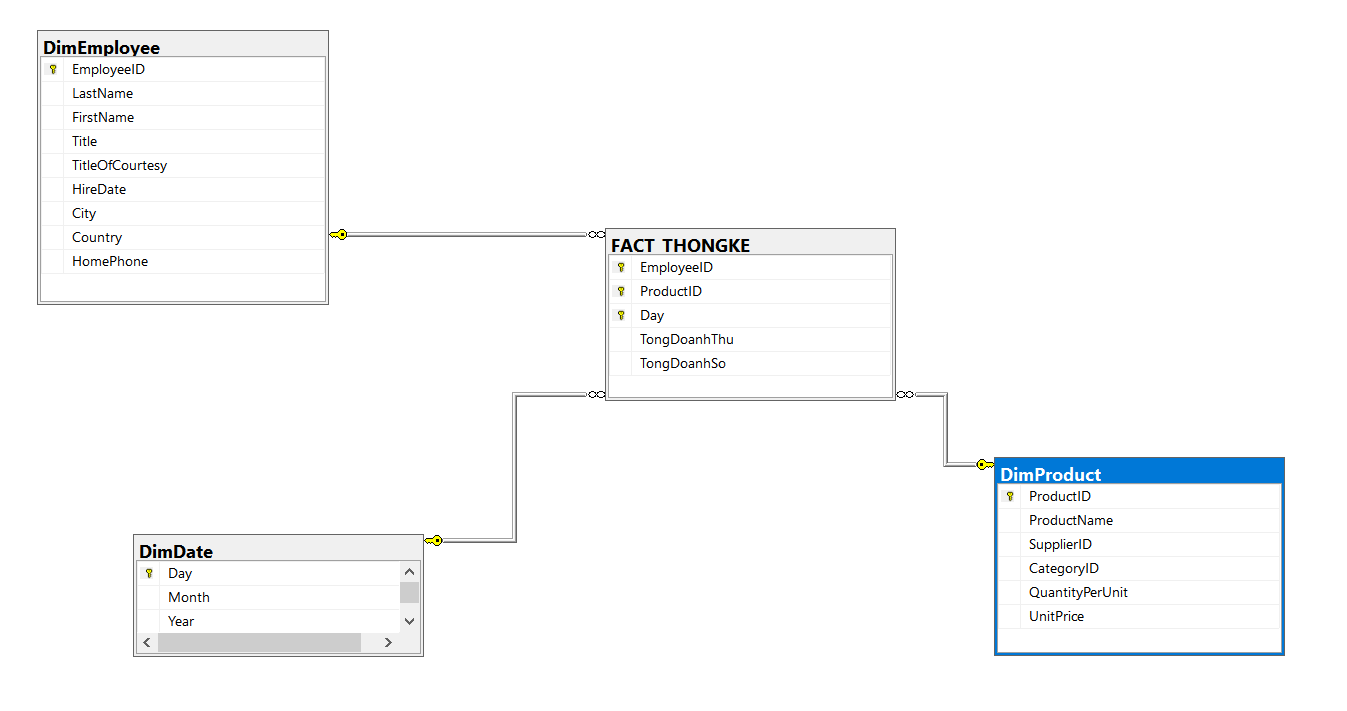
**Factor thống kê**

+ EmployeeID,ProductID,day

+ **DoanhThu =**  SUM(order.quantity\*oder.UnitPrice\*(1-oder.Discount)).

**+ DoanhSo** = SUM(order.quatity)

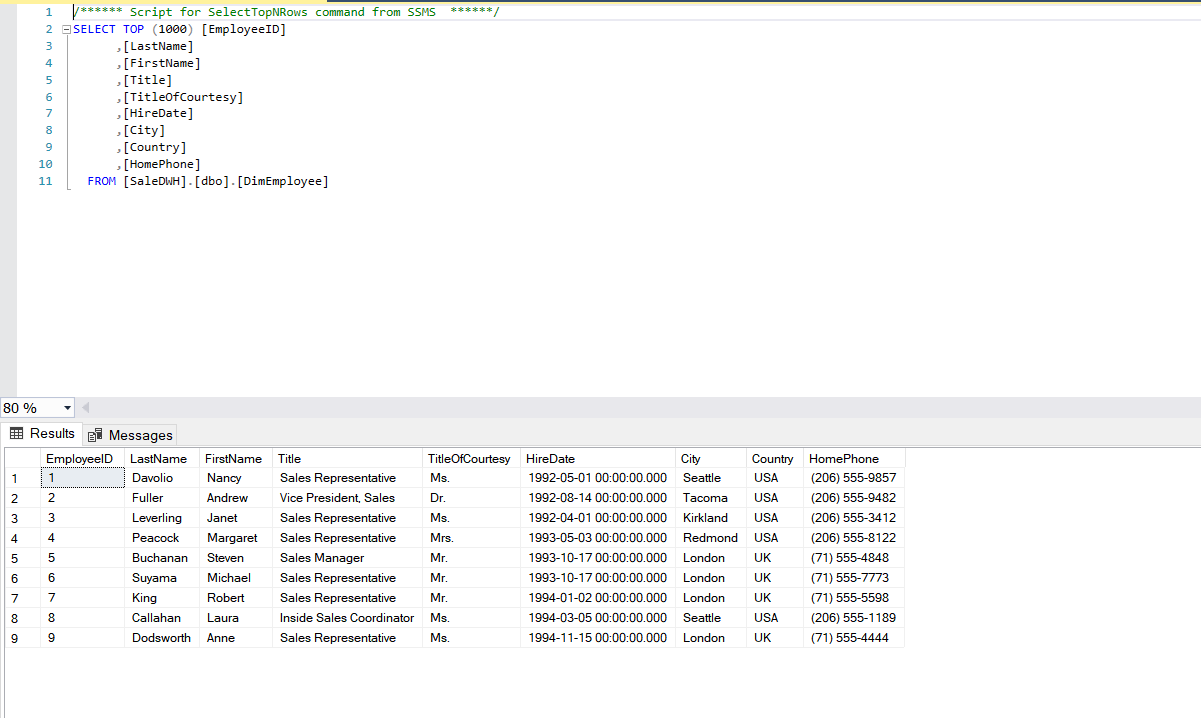
### Thiết kế DB mới tổ chức phân tích Kho dữ liệu cho phân hệ



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

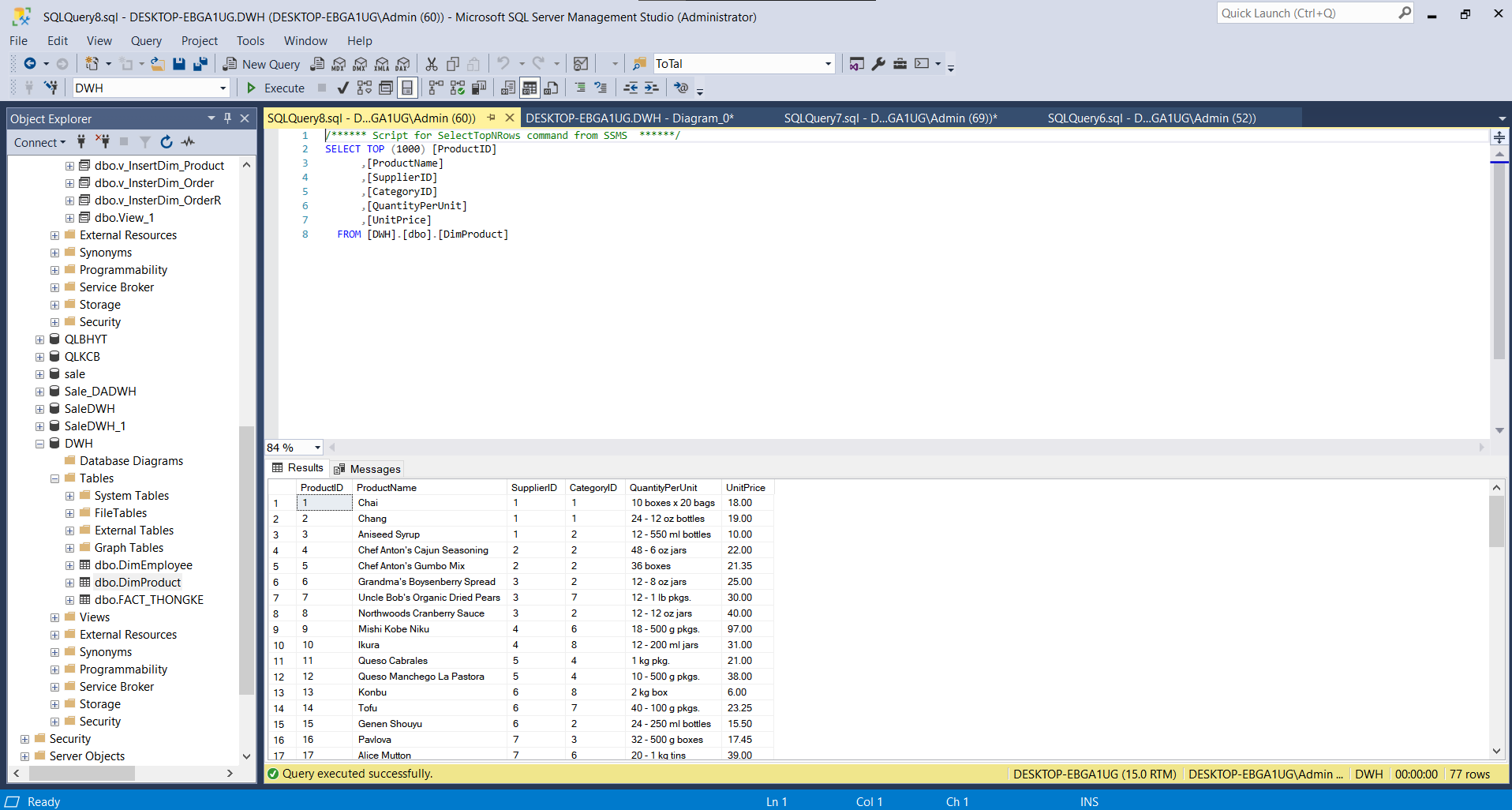
* Đưa dữ liệu vào dim\_employee

|  |
| --- |
| create view v\_InsertDim\_Emp as (  select [EmployeeID], [LastName], [FirstName], [Title], [TitleOfCourtesy], [HireDate], [City], [Country], [HomePhone]  from Employees  )  go  insert into SaleDWH.dbo.DimEmployee  select \* from v\_InsterDim\_Emp |



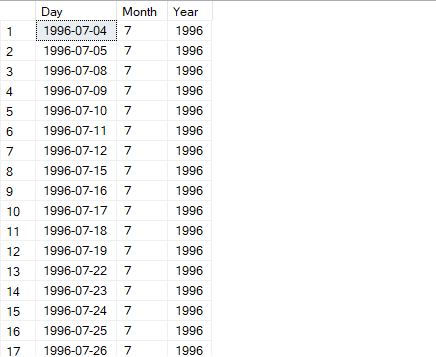
* Đưa dữ liệu và dim\_product

|  |
| --- |
| create view v\_InsertDim\_Product as  (select a.ProductID,a.ProductName,a.SupplierID,a.CategoryID,a.QuantityPerUnit,a.UnitPrice from Products as a)  go  insert into SaleDWH2.dbo.DimProduct  select \* from v\_InsertDim\_Product |



* Đưa dữ liệu vào dimDate

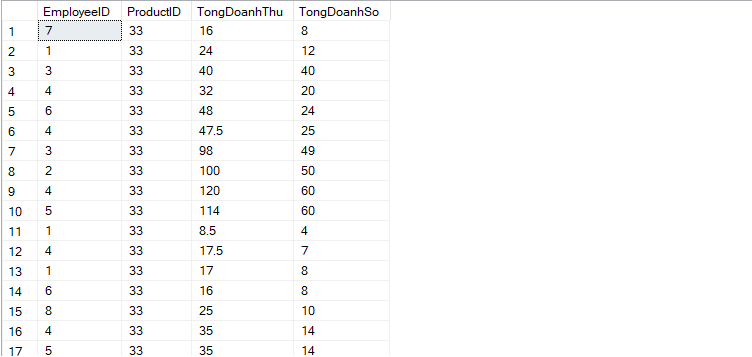
|  |
| --- |
| insert into SaleDWH2.dbo.Date  select \* from v\_InsertDim\_Date |



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

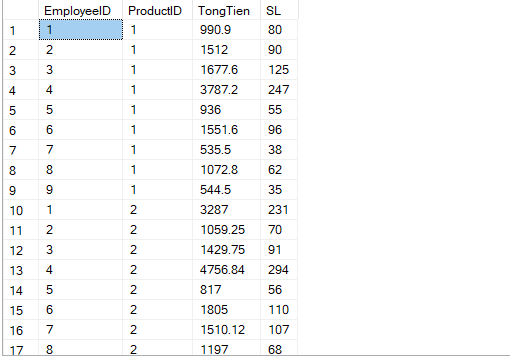
* Tạo view tính tổng doanh thu,doanh số theo theo nhân viên với từng sản phẩm của từng đơn hàng

|  |
| --- |
| SELECT c.EmployeeID, b.ProductID, (b.UnitPrice \* b.Quantity) \* (1 - b.Discount) AS TongDoanhThu, SUM(b.Quantity) AS TongDoanhSo  FROM dbo.Products AS a INNER JOIN  dbo.[Order Details] AS b ON a.ProductID = b.ProductID INNER JOIN  dbo.Orders AS d ON b.OrderID = d.OrderID INNER JOIN  dbo.Employees AS c ON d.EmployeeID = c.EmployeeID  GROUP BY b.UnitPrice, b.Quantity, b.Discount, b.ProductID, c.EmployeeID |

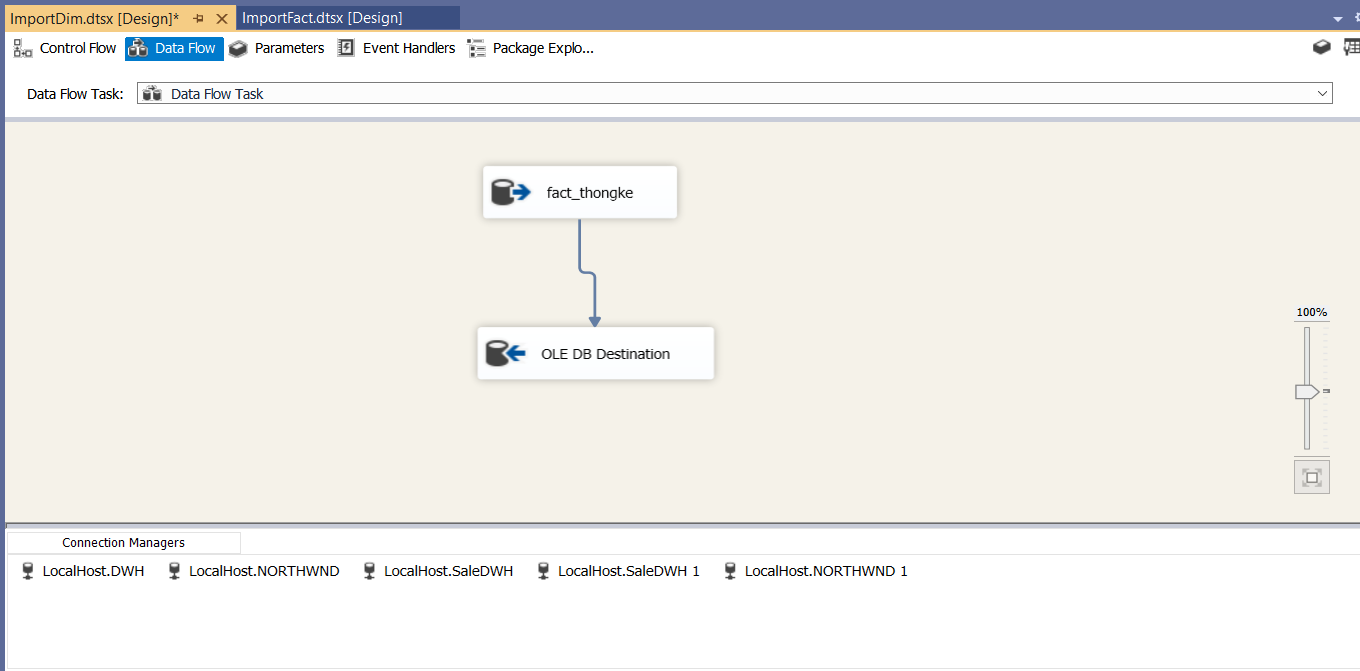


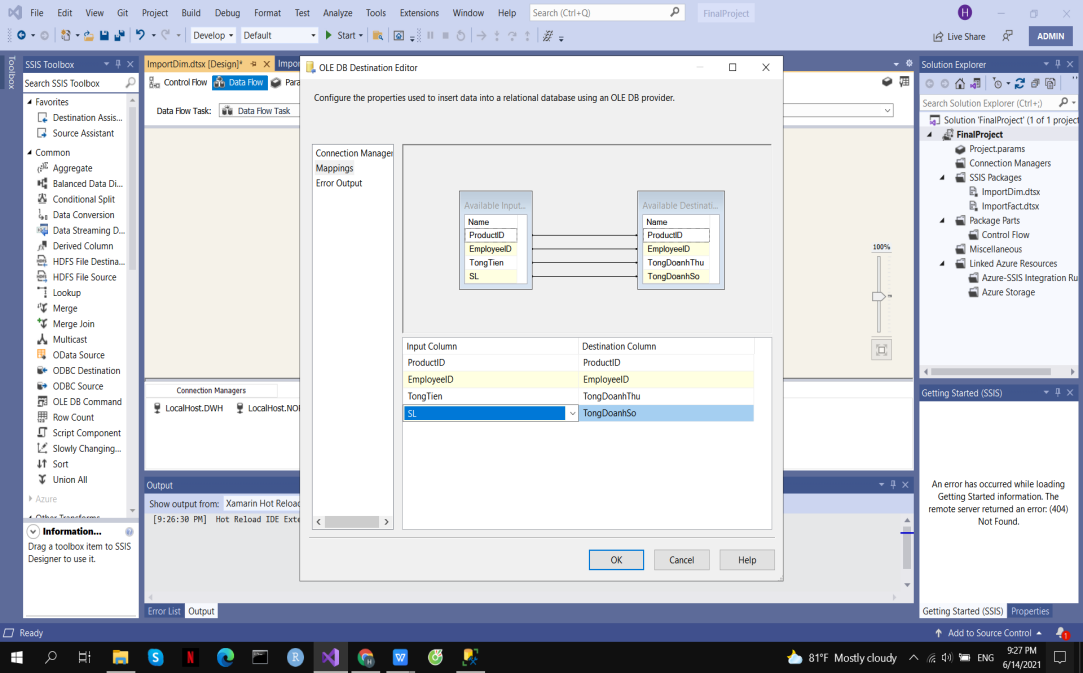
* Tạo view tính doanh thu , doanh số theo nhân viên theo sản phẩm

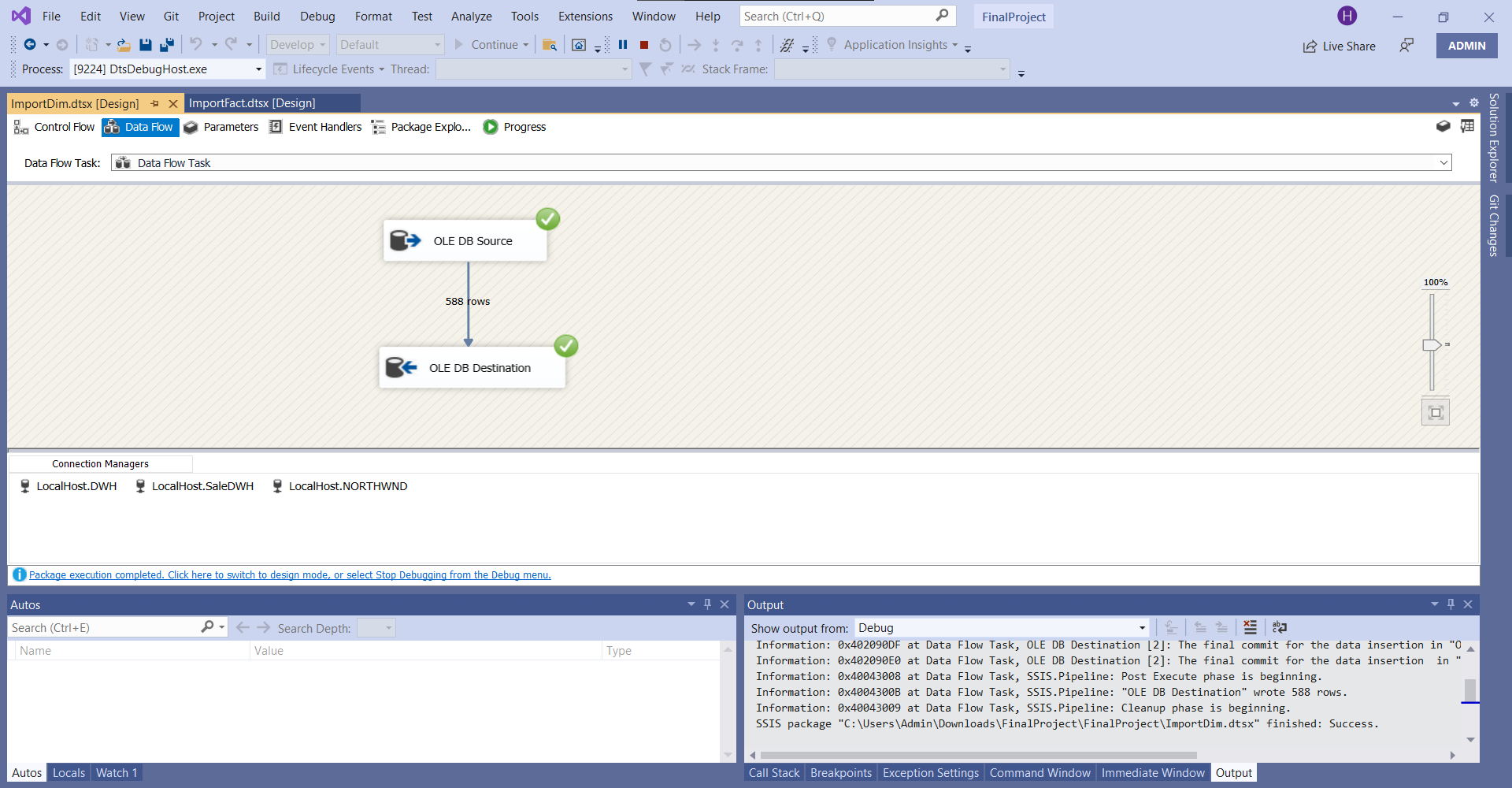
|  |
| --- |
| SELECT EmployeeID, ProductID, ROUND(SUM(TongDoanhThu), 2) AS TongTien, SUM(TongDoanhSo) AS SL  FROM dbo.tinhtong AS k  GROUP BY EmployeeID, ProductID  GROUP BY b.UnitPrice, b.Quantity, b.Discount, b.ProductID, c.EmployeeID |



* Nạp dữ liệu vào bảng fact

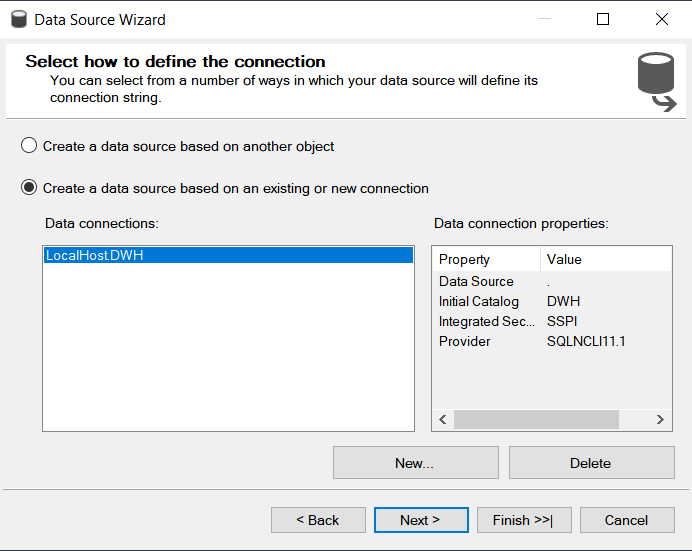




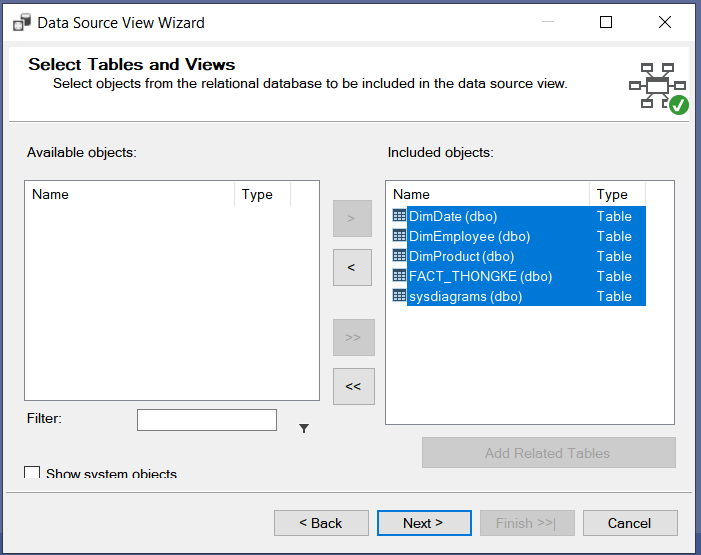


### Triển khai phân hệ DWH lên BIDS

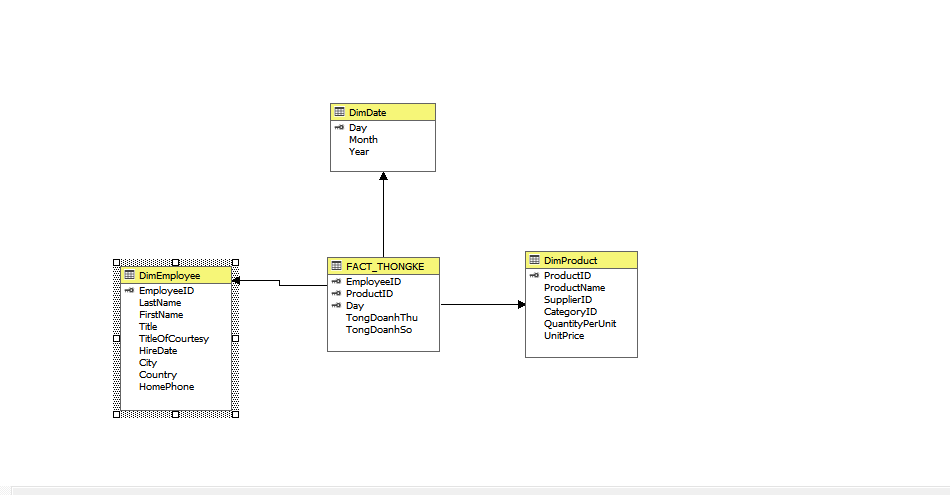
* Data source



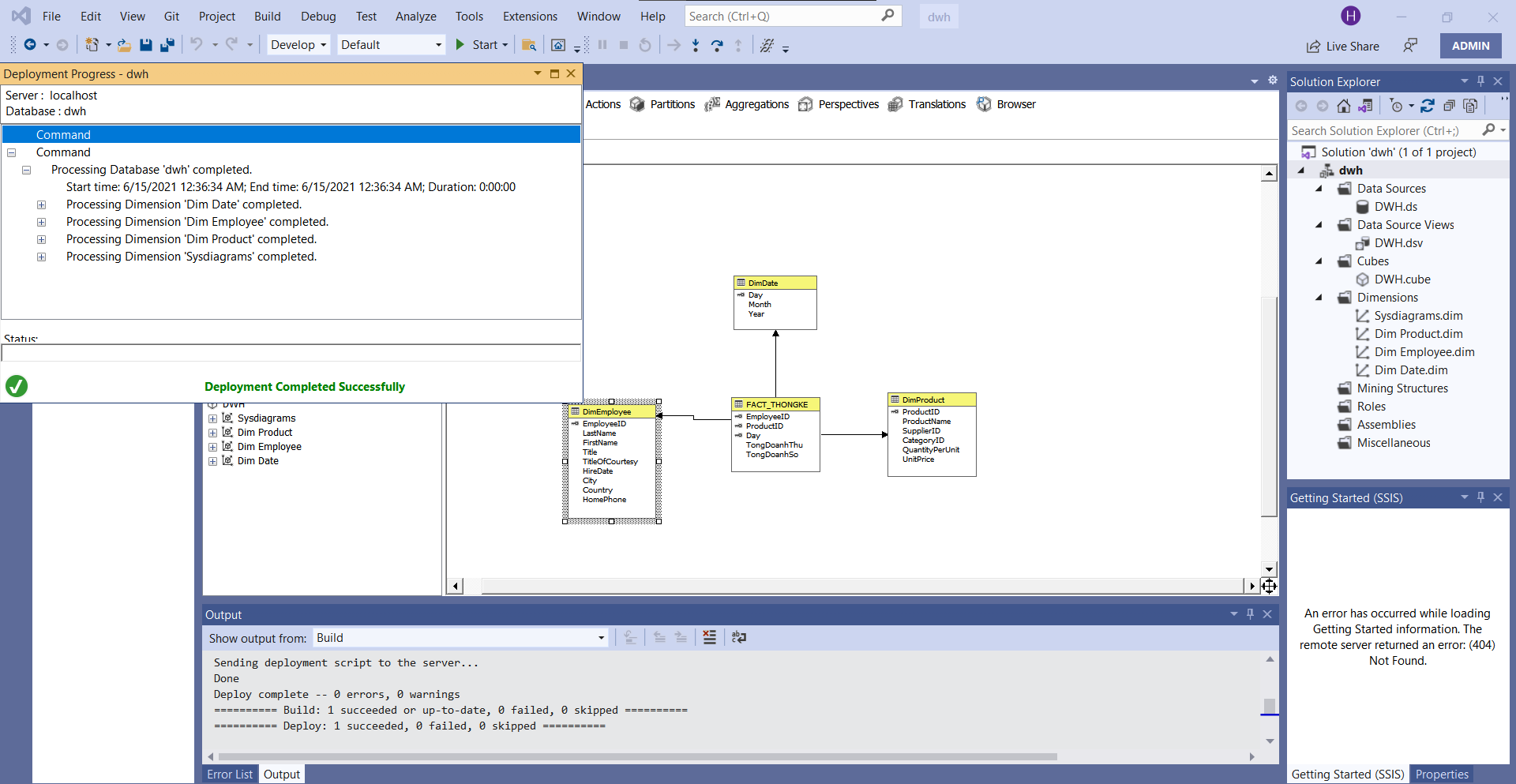
* Data source view



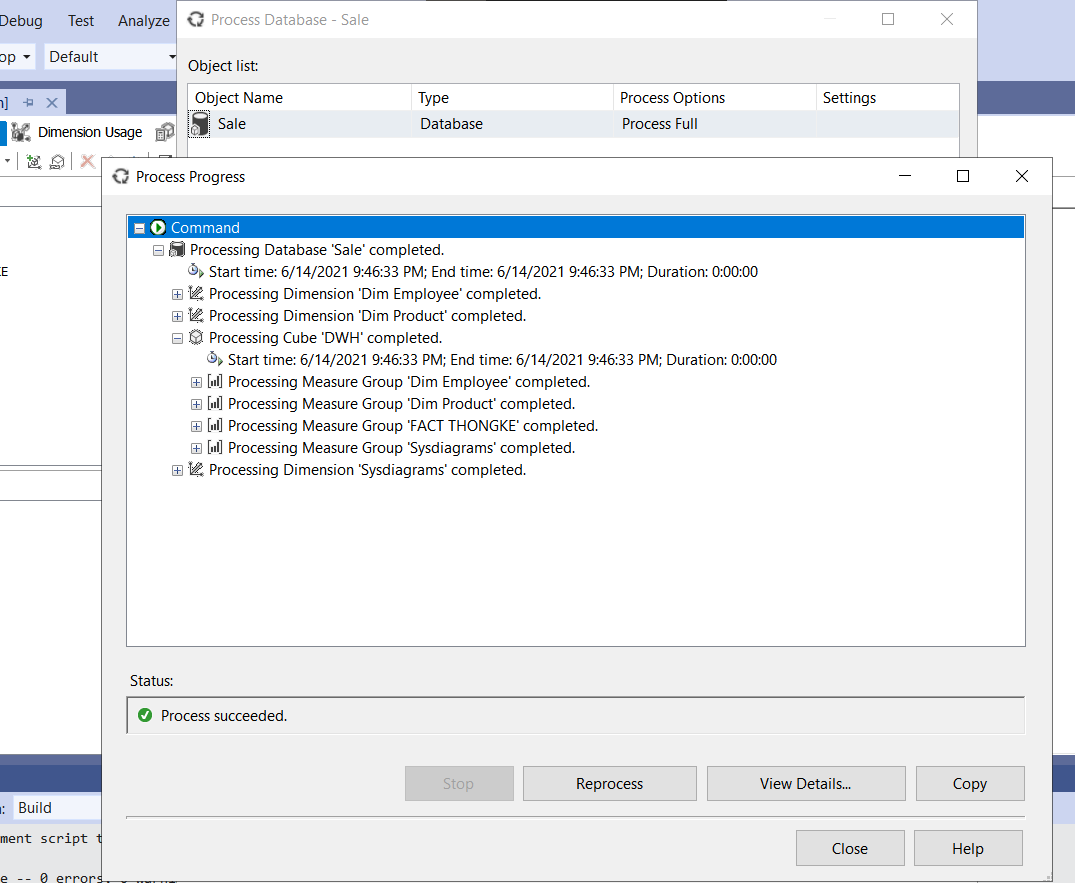
* Cube



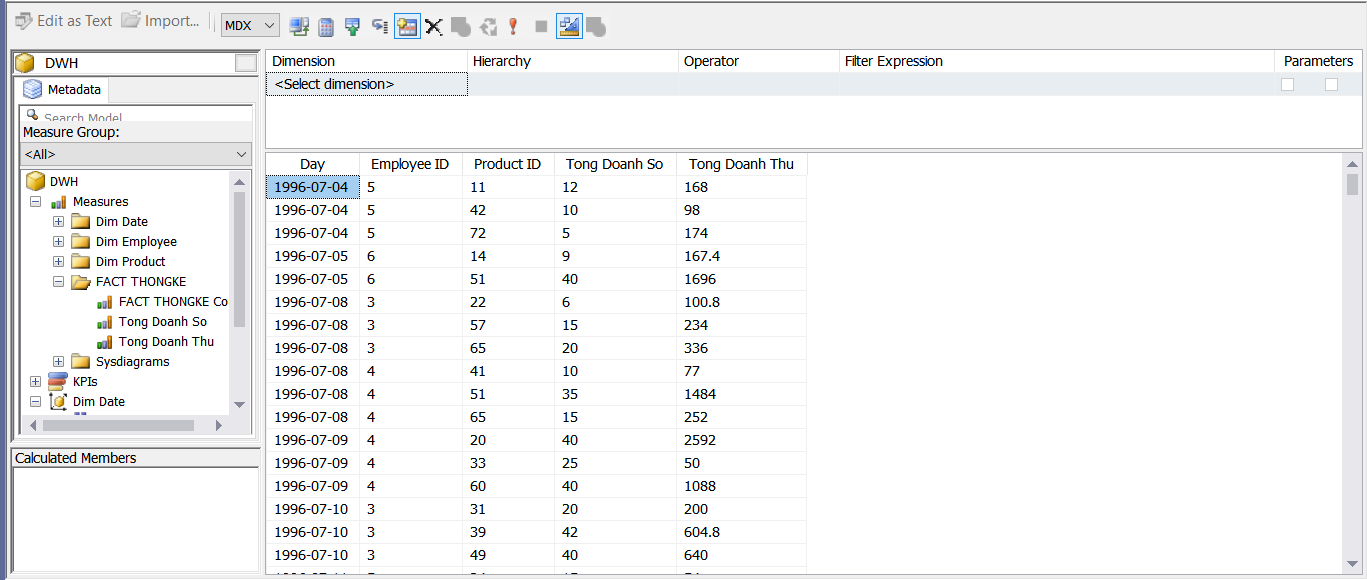
* Deloy



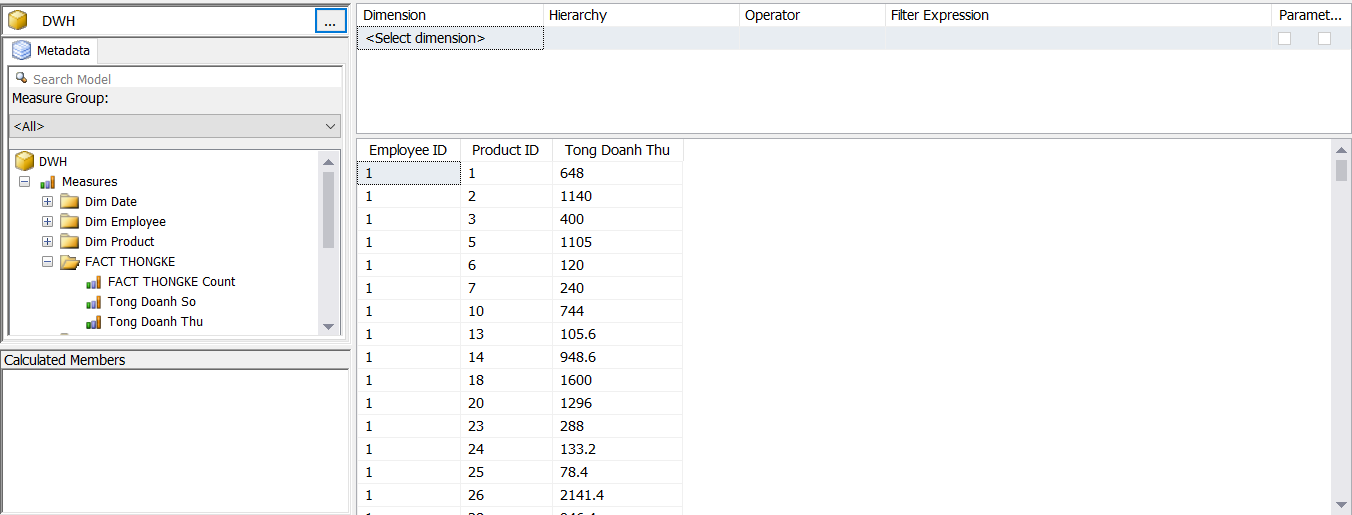
* Process



* Kết quả phân tích:
* Doanh thu ,doanh số của nhân viên theo thời gian cụ thể



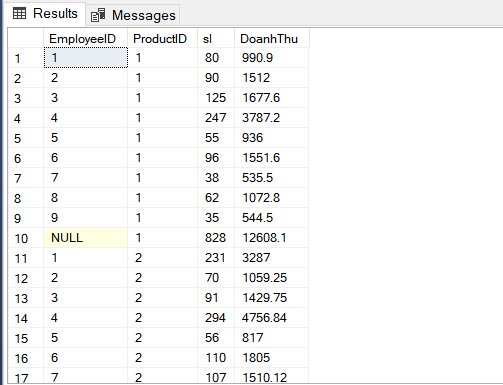
* Doanh thu, doanh số của nhân viên theo id sản phẩm



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

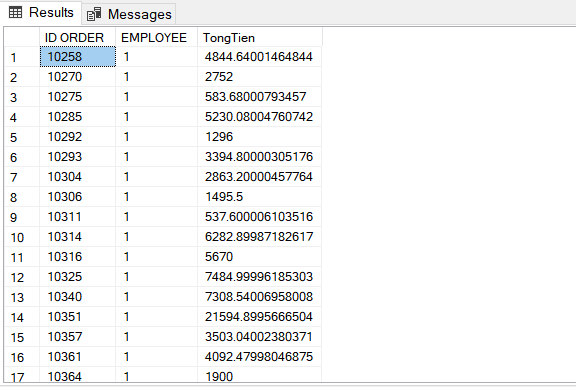
* Tính tổng doanh số , doanh thu của từng sản phẩm

|  |
| --- |
| select k.EmployeeID , k.ProductID , round(sum(k.TongDoanhSo),2) as sl ,  round(sum(k.TongDoanhThu),2) as DoanhThu  from tinhtong as k  group by ROLLUP(k.EmployeeID), rollup(k.ProductID) |



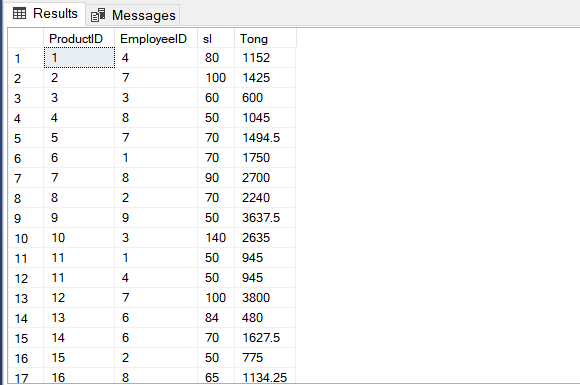
* Tính tổng số tiền đơn hàng theo mã nhân viên

|  |
| --- |
| SELECT CASE WHEN Grouping(k.OrderID) = 1 THEN N'TONG TIEN CUA DON HANG THEO NV' ELSE k.OrderID END AS [ID ORDER], u.EmployeeID AS [EMPLOYEE],    SUM(t.Subtotal) AS TongTien  FROM dbo.[Order Details] AS k INNER JOIN    dbo.[Order Subtotals] AS t ON k.OrderID = t.OrderID  inner join dbo.Employees as u on u.EmployeeID=t.EmployeeID  GROUP BY ROLLUP(k.OrderID),u.EmployeeID |



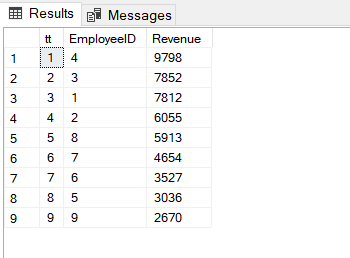
* Nhân viên có doanh thu cao nhất phân theo mã sản phẩm

|  |
| --- |
| select a.ProductID, b.EmployeeID, a.sl,a.Tong  from (  select k.ProductID , max(k.TongDoanhSo) as sl , max(k.TongDoanhThu) as tong  from tinhtong as k  GROUP by k.ProductID) as a, tinhtong as b  where a.ProductID=b.ProductID and a.tong=b.TongDoanhThu  order by a.ProductID ,b.EmployeeID |

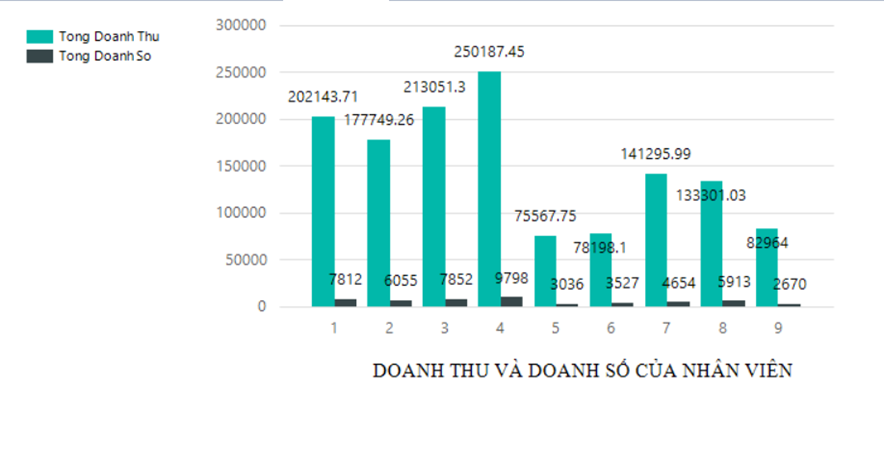


* Xếp hạng doanh thu của nhân viên

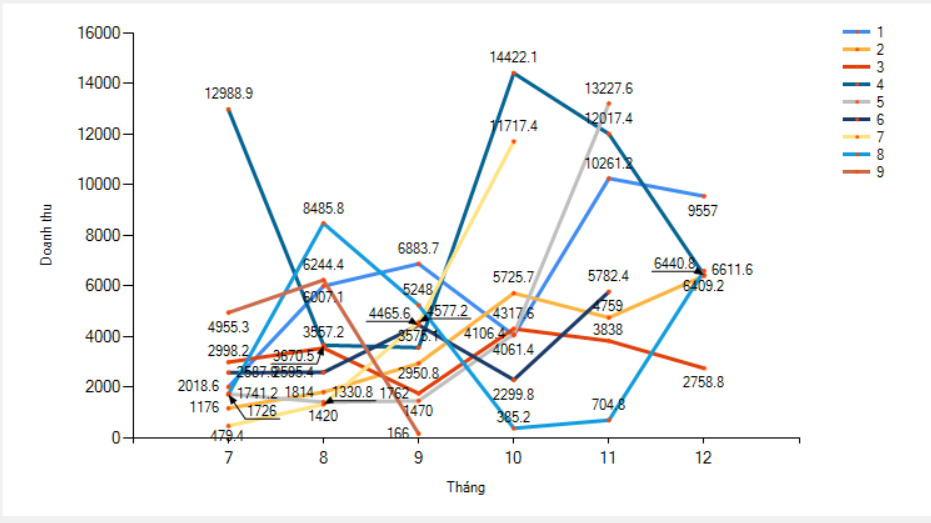
|  |
| --- |
| SELECT rank() OVER (ORDER BY sum(f.TotalQuantity) desc ) AS tt, cus.EmployeeID, SUM(f.TotalQuantity) AS Revenue  FROM dbo.FACT\_Statistic AS f INNER JOIN  dbo.DimEmployee AS cus ON f.EmployeeID = cus.EmployeeID  GROUP BY cus.EmployeeID |



* + 1. Trực quang dữ liệu



* Biểu đồ hiển thị “doanh thu và doanh số của nhân viên” ta thấy được nhân viên có id 4 bán hàng vượt trôị hơn hẳn với doanh thu là 250187.45$ và doanh số bán được là 9798 mặt hàng



* Biểu đồ hiển thị doanh thu của nhân viên theo thời gian (tháng/năm) cho thấy trong năm 1996 nhân viên có id là 6 bán hàng cao hơn hẳng nhưng không ổn định

## <STT>\_ Ho Tên 2\_ <thành phần kho dữ liệu đươc phân công>

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

## Kết luận

### Những kết quả đạt được

ĐỒ ÁN HỌC PHẦN đã giải quyết được vấn đề gì. Gạch đầu dòng nói những ý đã làm được.

### Hạn chế

Nêu những vấn đề còn tồn tại.

Nêu nhưng nội dung THEO yêu cầu của ĐỒ ÁN HỌC PHẦN nhưng bản thân chưa thực hiện được, hoặc đã thực hiện, nhưng chưa tốt và chưa hài lòng.

Cố gắng tự khen mình, không nên chê, không “vạch áo cho người khác xem lưng” (nói khéo để bỏ qua vấn đề còn vướng mắc): nhưng KHÔNG khen ẩu

## Hướng phát triển

### Hướng khắc phục các hạn chế

Nêu hướng khắc phục tất cả những tồn tại, hạn chế đã nêu trên (trong phần 5.1.2. của kết luận).

+ Có bao nhiêu hạn chế là bấy nhiêu hướng khắc phục; không bỏ qua hạn chế nào

+ Hướng khắc phục chỉ là giải pháp đề xuất, không nhất thiết đúng và cũng nhất thiết là thực hiện được (chỉ là giải pháp = phương pháp giải quyết vấn đề do cá nhân đề xuất); nhưng KHÔNG đề xuất những giải pháp mâu thuẫn, vô lý, quá phi thực tế, . . .

### Hướng mở rộng ĐỒ ÁN HỌC PHẦN

+ Hướng mở rộng để nghiên cứu thêm cho ĐỒ ÁN HỌC PHẦN.

+ (chú ý: khác với 5.1.2 hạn chế) : Nêu những nội dung có thể thực hiện thêm NGOÀI những yều cầu của ĐỒ ÁN HỌC PHẦN nhằm tăng giá trị của ĐỒ ÁN HỌC PHẦN về lý luận hoặc/và thực tiễn.

# 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).

# CÁC PHỤ LỤC

(nếu có)

# 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 |