**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 Bệnh viện Tai Mũi Họng**

Lớp: **181330A, 181330B, 181330C**

**NHÓM : 10**

THÀNH VIÊN:

1. Trần Minh Tú Trung [51]\_Nhóm trưởng
2. Đào Văn Thắng [41]
3. Đặng Ngọc Sơn [37]

**Ngày nộp: 15/06/2021**

**TP.HCM, ngày 15 Tháng 06 năm 2021**

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

**………………………………………………………………..**

**………………………………………………………………..**

**……………………………………………………………….**

**NHÓM :10**

**ĐIỂM ĐÁNH GIÁ**

1. Trần Minh Tú Trung [51]\_Nhóm trưởng: Thành phần kho dữ liệu Khoa
2. Đào Văn Thắng [41]: Thành phần kho dữ liệu bệnh nhân
3. Đặng Ngọc Sơn [37]: Thành phần kho dữ liệu bác sĩ

**Ngày 15/06/2021**

**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. Trần Minh Tú Trung [51]\_Nhóm trưởng: Thành phần kho dữ liệu Khoa
2. Đào Văn Thắng [41]: Thành phần kho dữ liệu bệnh nhân
3. Đặng Ngọc Sơn [37]: Thành phần kho dữ liệu bác sĩ

**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 Bệnh viện Tai Mũi Họng**

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

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

**Tên sản phẩm đề tài: 51.10.01.TranMinhTuTrung.DAHP.DWHHospital.rar** (Microsoft SQL Server 2014)

**DB Nguồn: 5110TranMinhTuTrungQLBVTMH.bak**

**DB DWH: 5110TranMinhTuTrung HC\_DWH.bak**

**Tên máy chủ DB SQL: .**

**Tên đăng nhập DB SQL: sa**

**Mật khẩu (password) đăng nhập SQL: 123**

**ProjectName BIDS: sttNxHoTenQLKCBBHYT\_DWH (Microsoft Visual Studio 2019) VD: 05N1VXTheQLKCBBHYT\_DWH**

**NHÓM : 10**

THÀNH VIÊN:

1. Trần Minh Tú Trung [51]\_Nhóm trưởng: Thành phần kho dữ liệu Khoa
2. Đào Văn Thắng [41]: Thành phần kho dữ liệu bệnh nhân
3. Đặng Ngọc Sơn [37]: Thành phần kho dữ liệu bác sĩ

{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

Chúng em xin chân thành cảm ơn sự hướng dẫn tận tình của thầy ***Võ Xuân Thể*** đã giúp chúng em hoàn thành bài báo cáo cuối kì bộ môn Kho dữ liệu. Tuy đã cố gắng học tập, tìm hiểu, phân tích thiết kế và xây dựng hệ thống nhưng ắt hẳn thiếu sót là điều không thể tránh khỏi. Chúng em rất mong nhận được sự thông cảm và góp ý chân thành của Thầy. 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](#_Toc74857993)

[DANH MỤC CÁC TỪ VIẾT TẮT 5](#_Toc74857994)

[DANH MỤC CÁC THUẬT NGỮ 6](#_Toc74857995)

[Chương 1. GIỚI THIỆU VỀ ĐỒ ÁN HỌC PHẦN 10](#_Toc74857997)

[1.1. Tổng quan về ĐỒ ÁN HỌC PHẦN 10](#_Toc74857998)

[1.2. Nội dung chuyên môn chính của ĐỒ ÁN HỌC PHẦN 10](#_Toc74857999)

[1.3. Mục tiêu của ĐỒ ÁN HỌC PHẦN 10](#_Toc74858000)

[1.4. Công cụ và nền tảng kỹ thuật thực hiện ĐỒ ÁN HỌC PHẦN 10](#_Toc74858001)

[1.5. Sản phẩm của ĐỒ ÁN HỌC PHẦN 10](#_Toc74858002)

[1.6. Bố cục của báo cáo 10](#_Toc74858003)

[Chương 2. CƠ SỞ LÝ THUYẾT CỦA HỌC PHẦN 11](#_Toc74858004)

[2.1. Tổng quan về kho dữ liệu (DWH) 11](#_Toc74858005)

[2.1.1. Các khái niệm cơ bản về Kho dữ liệu 11](#_Toc74858006)

[2.1.2. Kiến trúc cơ bản của kho dữ liệu: 12](#_Toc74858007)

[2.1.3. Mô hình CSDL đa chiều (Dimensional Modeling) 12](#_Toc74858008)

[2.2. Phân tích và thiết kế Kho dữ liệu 12](#_Toc74858009)

[2.2.1. Phân tích và thiết kế kho dữ liệu theo hướng mô hình hóa nhiều chiều (Dimensional Modeling) 12](#_Toc74858010)

[2.2.2. Các dạng (phương pháp) tích hợp dữ liệu 12](#_Toc74858011)

[2.2.3. Các bước cơ bản trong tiến trình tích hợp dữ liệu 12](#_Toc74858012)

[2.2.4. Các công cụ tích hợp dữ liệu: 13](#_Toc74858013)

[2.2.5. Lập các luồng công việc tích hợp dữ liệu (Data Integration Workflows) 13](#_Toc74858014)

[2.2.6. Hệ thống OLTP (On-Line Transaction Processing = Xử lý giao dịch trực tuyến) 13](#_Toc74858015)

[2.3. Vấn tin khai thác sử dụng Kho dữ liệu 13](#_Toc74858016)

[2.3.1. Vấn tin với các phép toán tổng hợp của SQL trên kho dữ liệu 13](#_Toc74858017)

[2.3.2. Vấn tin với các hàm phân tích của SQL trên kho dữ liệu 13](#_Toc74858018)

[2.3.3. Vấn tin dùng Materialized View để phân tích trên kho dữ liệu 14](#_Toc74858019)

[2.3.4. Hệ thống OLAP 14](#_Toc74858020)

[Chương 3. TỔNG QUAN VỀ KHO DỮ LIỆU TẠI BỆNH VIỆN TAI MŨI HỌNG 14](#_Toc74858021)

[3.1. Ý tưởng hình thành DWH tại Bệnh viện Tai Mũi Họng 14](#_Toc74858022)

[3.2. Giới thiệu chung về Kho dữ liệu 15](#_Toc74858023)

[3.2.1. Mô tả tổng quan về DWH 15](#_Toc74858024)

[3.2.2. Các dữ liệu nguồn hình thành DWH 15](#_Toc74858025)

[3.3. Xác định nhu cần tổ chức và phân tích dữ liệu của toàn bộ DWH 16](#_Toc74858026)

[3.4. Các thành phần phân hệ Kho dữ liệu phân công các cá nhân phụ trách 17](#_Toc74858027)

[Chương 4. PHÂN TÍCH, THIẾT KẾ VÀ KHAI THÁC SỬ DỤNG CÁC PHÂN HỆ KHO DỮ LIỆU TẠI BỆNH VIỆN TAI MŨI HỌNG 17](#_Toc74858028)

[4.1. 41\_ ĐÀO VĂN THẮNG\_ KHO DỮ LIỆU VỀ BỆNH NHÂN 17](#_Toc74858029)

[4.1.1. Mô tả tổng quan về phân hệ DWH – Phân hệ khoa Bệnh viện 17](#_Toc74858030)

[4.1.2. Xác định nguồn dữ liệu tích hợp phân hệ Kho dữ liệu 17](#_Toc74858031)

[4.1.3. Xác định ý tưởng hình thành phân hệ của Kho dữ liệu 18](#_Toc74858032)

[4.1.4. Mô tả hệ thống các DB gốc liên quan phân hệ DWH 18](#_Toc74858033)

[4.1.5. Phân phân tích các DB gốc xác định yêu cầu phân tích DWH của phân hệ 19](#_Toc74858034)

[4.1.6. Thiết kế DB mới tổ chức phân tích Kho dữ liệu cho phân hệ 20](#_Toc74858035)

[4.1.7. Thực hiện các thủ tục tích hợp các DB gốc vào phân hệ DWH 20](#_Toc74858036)

[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](#_Toc74858037)

[4.1.9. Triển khai phân hệ DWH lên BIDS: 28](#_Toc74858038)

[4.1.10. Khai thác sử dụng phân hệ DWH dùng các vân tin SQL 37](#_Toc74858039)

[Chương 5. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN CỦA ĐỒ ÁN HỌC PHẦN 41](#_Toc74858040)

[5.1. Kết luận 41](#_Toc74858041)

[5.1.1. Những kết quả đạt được 41](#_Toc74858042)

[5.1.2. Hạn chế 41](#_Toc74858043)

[5.2. Hướng phát triển 42](#_Toc74858044)

[5.2.1. Hướng khắc phục các hạn chế 42](#_Toc74858045)

[5.2.2. Hướng mở rộng ĐỒ ÁN HỌC PHẦN 42](#_Toc74858046)

[DANH MỤC TÀI LIỆU THAM KHẢO 42](#_Toc74858047)

[CÁC PHỤ LỤC 43](#_Toc74858048)

[ĐÁNH GIÁ KẾT QUẢ HỌC TẬP HỌC PHẦN 44](#_Toc74858049)

# GIỚI THIỆU VỀ ĐỒ ÁN HỌC PHẦN

## Tổng quan về ĐỒ ÁN HỌC PHẦN

Đồ án Học phần môn Kho dữ liệu: **Phát triển mô hình kho dữ liệu Bệnh viện Tai Mũi Họng**

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

Phân tích, thiết kế, chuyển đổi dữ liệu database gốc sang nhà Kho dữ liệu

## Mục tiêu của ĐỒ ÁN HỌC PHẦN

Nắm được cách phân tích thiết kế Data Warehouse theo nhu cầu phân tích dữ liệu

Nắm được kỹ thuật tích hợp dữ liệu từ các nguồn về Data Warehouse

Nắm được kỹ thuật phân tích bằng BIDS, Subtotal,…

## 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)

+ Tableau

## Sản phẩm của ĐỒ ÁN HỌC PHẦN

Kho dữ liệu “HealthCare\_DW”

BIDS Project

Các view phân tích, truy vấn Database gốc, DW

## 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) Kho dữ liệu Bệnh viện Tai Mũi Họng

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)?

Data Warehouse tạm gọi kho dữ liệu là một kỹ thuật thu thập và quản lý dữ liệu từ nhiều nguồn khác nhau để cung cấp những thông tin kinh doanh có ý nghĩa. Data Warehouse là sự pha trộn của các công nghệ và các thành phần hỗ trợ trong việc sử dụng chiến lược dữ liệu.

#### Nguyên lý hình thành DWH

Nguồn dữ liệu của DWH được hình thành từ nhiều loại khác nhau:

* Dữ liệu từ các hệ thống tác nghiệp.
* Hệ thống kế thừa.
* Các nguồn dữ liệu bên ngoài

#### Các đặc điểm cơ bản của DWH

Các đặc điểm cơ bản của DWH

* Tính tích hợp.
* Gắn thời gian.
* Có tính lịch sử.
* Chỉ đọc
* Không biến động
* Dữ liệu tổng hợp và chi tiết

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

DWH về bản chất cũng là một database bình thường, các hệ quản trị cơ sở dữ liệu quản lý và lưu trữ nó như các database thông thường tuy nhiên có hỗ trợ thêm về quản lý dữ liệu lớn và truy vấn. Thực ra nét khác biệt của DWH so với database là ở quan niệm, cách nhìn vấn đề.

### Kiến trúc cơ bản của kho dữ liệu:

#### Tầng đáy: nạp dữ liệu

**Tầng đáy (Data Sources)**: là nơi thu thập, tích hợp dữ liệu từ nhiều nguồn khác nhau sau đó chuẩn hóa về cùng định dạng, làm sạch xử lí dữ liệu để tìm lỗi và sửa và lưu trữ dữ liệu đã tổng hợp.

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

**Tầng giữa (Warehouse)**: cung cấp dịch vụ xử lý phân tích trực tuyến (OLAP – Online Analytical Processing) để thao tác với kho dữ liệu đã được tích hợp. Tầng giữa có thể được cài đặt bằng Relational OLAP, Multidimensional OLAP hay kết hợp cả hai mô hình trên Hybrid OLAP

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

**Tầng trên cùng (Users)**: nơi thực hiện các thao tác truy vấn, báo cáo, phân tích để tìm ra xu hướng, trung bình, tổng hợp…

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

Bản chất đa chiều của các câu hỏi trong nghiệp vụ được phản ánh trong thực tế chẳng hạn như những người quản lí thị trường không được thoả mãn với câu hỏi theo một chiều đơn giản, thay vào đó là những câu hỏi phức tạp. Một cách để quan sát một mô hình dữ liệu nhiều chiều là nhìn nó như một hình khối. Các thành phần chính:

## 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)

### Các bước cơ bản trong tiến trình tích hợp dữ liệu

B1. Trích xuất dữ liệu = thu thập dữ liệu lớn từ nhiều nguồn

B2. Làm sạch = tìm lỗi và sửa sai

B3. So khớp mẫu = Lọc và chuyển đổi dữ liệu = chuyển đổi sang định dạng kho

B4. Lưu trữ dữ liệu theo cấu trúc = sắp xếp, hợp nhất và tổng hợp, ...

### Các công cụ tích hợp dữ liệu:

SQL và phần mềm

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

Tích hợp dữ liệu có 2 hình thức cơ bản đó là:

* Tích hợp để đảm bảo việc trao đổi dữ liệu giữa các hệ thống được thành công (effectiveness), bàn giao hoặc xử lý các công việc tiếp theo theo các luồng công việc định trước.
* Tích hợp các dữ liệu từ các nguồn khác nhau để có cái nhìn tổng thể (gọi là Unified) cho người sử dụng/cấp độ quản lý hoặc chuẩn bị cho các Report cụ thể (dạng Report cho BI).

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

Các chức năng cơ bản của OLTP: cùng với khả năng truy cập và cập nhật các dữ liệu chia sẻ, các hệ thống OLTP còn hỗ trợ các user khả năng truy cập trực tuyến (online), khả năng truy cập tức thời (availability), khả năng phản hồi nhanh chóng (response), và tiết kiệm chi phí đối với từng transaction (low cost).

## Vấn tin khai thác sử dụng Kho dữ liệu

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

Ôn 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 lý 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.

+ Viết truy vấn dùng Materialized View:

* Ôn tập lại khái niệm khung nhìn (views)
* Định nghĩa materialized view và cách thức nó được xử lý
* Các quy tắc khi viết lại truy vấn (query rewrite) dùng materialized view
* Ví dụ áp dụng về viết lại truy vấn

### 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 TẠI BỆNH VIỆN TAI MŨI HỌNG

## Ý tưởng hình thành DWH tại Bệnh viện Tai Mũi Họng

* Bệnh viện Tai Mũi Họng là một trong những bệnh viện hàng đầu của đất nước, đi tiên phong về chất lượng bác sĩ, y tế và có độ uy tín về chất lượng chữa trị. Do đó, lượng bệnh nhân khám chữa bệnh tại bệnh viện rất nhiều, vì thế dữ liệu của bệnh viện rất lớn.
* Để có thể quản lý và có được những thống kê chính xác để có thể duy trì và cải thiện hiệu quả, nhất là trong thời buổi dịch Covid đang ngày càng lớn dần. Bệnh viện cần hình thành Data Warehouse dành cho các Khoa, Bác Sĩ, và hồ sơ bệnh nhân khám bệnh.

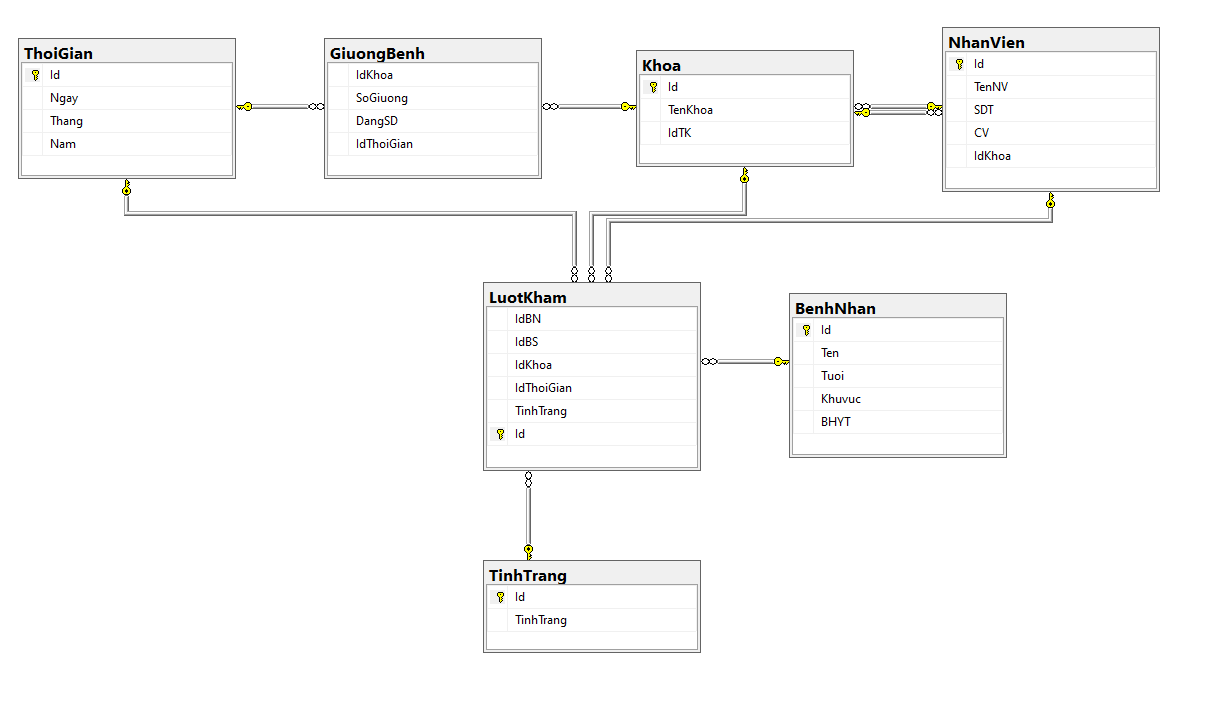
## Giới thiệu chung về Kho dữ liệu

### Mô tả tổng quan về DWH

Một bệnh viện Tai mũi họng chuyên tiếp nhận bệnh nhân hằng có 1 database lưu trữ các thông số từng khoa và lượt khám của từng bệnh nhân.

### Các dữ liệu nguồn hình thành DWH

Diagram của dữ liệu nguồn:



Bao gồm 7 Table:

- GiuongBenh:

+ IdKhoa: ID của khoa sở hữu giường bệnh.

+ SoGiuong: tổng số giường của khoa.

+ DangSD: tổng số giường còn trống của khoa.

+ IdThoiGian: thời điểm.

* Khoa:

+ Id(PK): Id của Khoa.

+ TenKhoa: Tên của khoa.

+ IdTK: Id của trưởng khoa.

* NhanVien:

+ Id(PK): Id của nhân viên.

+ TenNV: tên của nhân viên.

+ SDT: Số điện thoại của nhân viên.

+ CV: chức vụ của nhân viên. (y tá, bác sĩ)

+ IdKhoa: Id Khoa của nhân viên.

* - LuotKham:
* + Id(PK): Id của từng lượt khám.
* + IdBN: Id của bệnh nhân thuộc lượt khám.
* + IdBS: Id của bác sĩ khám.
* + IdKhoa: Id Khoa mà bệnh nhân khám.
* + IdThoiGian: Id thời gian khám.
* + TinhTrang: tình trạng của bệnh nhân.

- BenhNhan:

+ Id(PK): Id của bệnh nhân.

+ Ten: Tên của bệnh nhân.

+ Tuoi: tuổi của bệnh nhân.

+ SDT: Số điện thoại của bệnh nhân.

+ Khuvuc: Khu vực ở của bệnh nhân ( Khu vực 1,2,3 )

+ BHYT: Có hoặc không.

* TinhTrang:

+ Id: Id của tình trạng.

+ TinhTrang: Tình trạng mà bệnh nhân được bác sĩ yêu cầu.

- ThoiGian:

+ Id: DateKey.

+ Ngay

+ Thang

+ Nam

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

- Fact\_BenhNhan:

+ Số lượng khám của một bệnh nhân trong tháng: tìm xem bệnh nhân có lượt khám cao đột biến hay không, để có thể xác định chính xác và tăng hiệu quả chữa bệnh.

+ Số lượt khám của mỗi khu vực và có độ tuổi trung bình là bao nhiêu : thống kê được khu vực có bệnh nhân nào có mắc bệnh nhiều để có phác đồ chữa trị hợp lí.

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

Trần Minh Tú Trung [51]\_Nhóm trưởng: Thành phần kho dữ liệu Khoa

Đào Văn Thắng [41]: Thành phần kho dữ liệu bệnh nhân

Đặng Ngọc Sơn [37]: Thành phần kho dữ liệu bác sĩ

# PHÂN TÍCH, THIẾT KẾ VÀ KHAI THÁC SỬ DỤNG CÁC PHÂN HỆ KHO DỮ LIỆU TẠI BỆNH VIỆN TAI MŨI HỌNG

## ĐÀO VĂN THẮNG[2]\_ KHO DỮ LIỆU VỀ BỆNH NHÂN

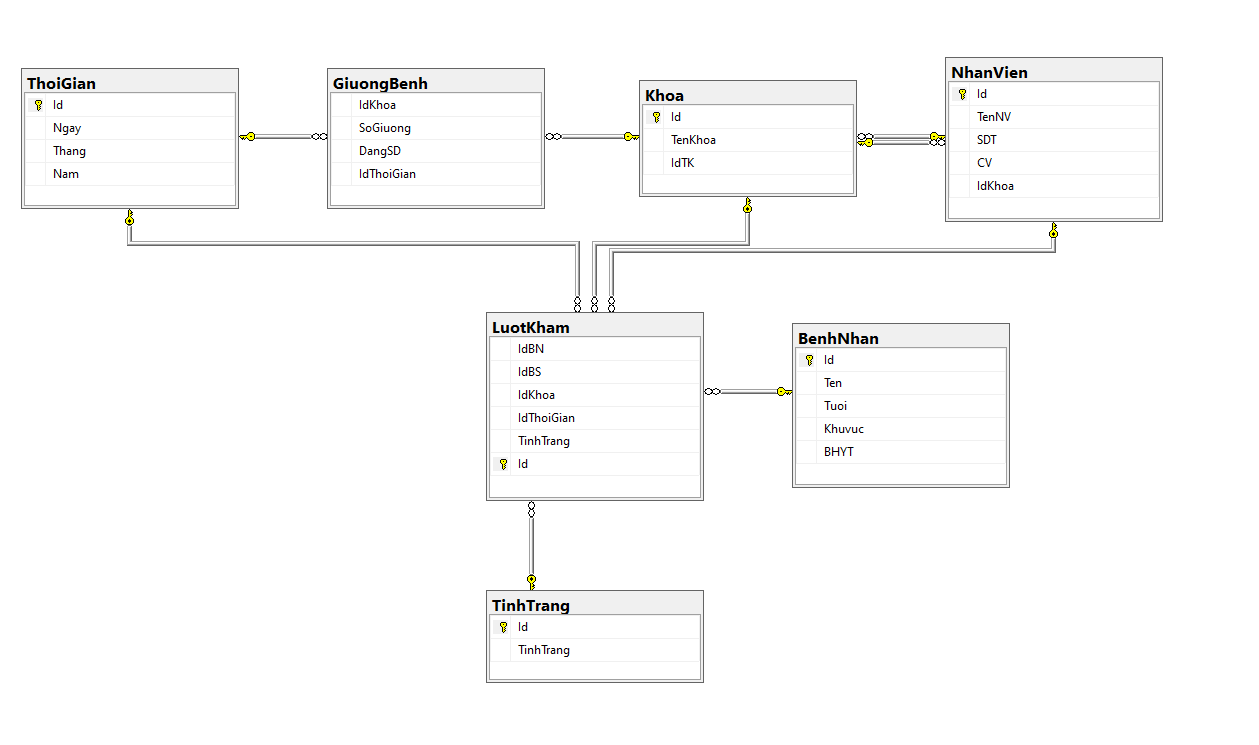
### Mô tả tổng quan về phân hệ DWH – Phân hệ khoa Bệnh viện

* Data Warehouse – Phân hệ Bệnh nhân Bệnh viện bao gồm 1 Fact và 4 Dims. Phân hệ Bệnh Nhân dùng để phân tích, thống kê các số liệu về thông tin bệnh nhân , lượt khám, Độ tuổi trung bình , khu vực của các bệnh nhân đến khám trong bệnh viện Tai Mũi Họng.
* Trong phân hệ, có các phép toán truy suất tổng số lượt khám theo bệnh nhân trong 1 tháng, Tổng số bệnh nhân theo từng khu vực, và độ tuổi trung bình tham gia Khám bệnh.

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

Nguồn dữ liệu phân hệ Kho dữ liệu:

* Diagram:



### Xác định ý tưởng hình thành phân hệ của Kho dữ liệu

+ Tổng số lượt khám theo Bệnh nhân: Theo dõi số lượt khám của các bệnh nhân tại Bệnh viện từ đó đứa ra dự đoàn cũng như những theo dõi đặc biệt dành cho bệnh nhân. Phân loại bệnh nhân dựa trên số lượt khám theo Tháng và theo ngày của Bệnh nhân.

+ Tính số lượt khám theo khu vực : Mỗi bệnh nhân sẽ ở một Khu vực và việc tính số lượt khám theo khu vực giúp cho bệnh viện phán đoán được đối tượng bệnh nhân cần quan tâm và sẽ có hướng hỗ trợ, tổ chức những buổi hội thảo về vấn đề sức khỏe cho từng khu vực.

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

* **BenhNhan:**

+ Id(PK): Id của Bệnh nhân.

+ Ten: Tên của Bệnh nhân.

+ Tuoi: Tuổi của bệnh nhân.

+ Khuvuc: Khu vực của bệnh nhân sinh sống.

+ BHYT : Bệnh nhân đó có bảo hiểm y tế hay là không.

* **- LuotKham:**
* + Id(PK): Id của từng lượt khám.
* + IdBN: Id của bệnh nhân thuộc lượt khám.
* + IdBS: Id của bác sĩ khám.
* + IdKhoa: Id Khoa mà bệnh nhân khám.
* + IdThoiGian: Id thời gian khám.
* + TinhTrang: tình trạng của bệnh nhân.

**- ThoiGian:**

+ Id: DateKey.

+ Ngay

+ Thang

+ Nam.

* **TinhTrang:**

+ Id : Id của Tình trạng

+ TinhTrang: Tên của tình trạng của lượt khám.

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

Xác định các Facts và Dims

**FactBenhNhan:**

+ các IdBN, IdLK, IdTG , IdTT.

+ **TongSoLuotKham =** SUM(**DimLuotKham**.**LuotKham**)

WHERE DimBenhNhan.Id = DimLuotKham.IdBN

**+ DemTheoKhuVuc** = COUNT(**DimBenhNhan.KhuVuc**)

WHERE DimBenhNhan.Id = DimLuotKham.IdBN

+ **DoTuoiTrungBinh** = AVG (DimBenhNhan.Tuoi)

---------[DWH HÌNH SAO: 1st]

DimBenhNhan: Bao gồm id bệnh nhân, tên, Tuổi, Khu vực, BHYT của bệnh nhân.

DimLuotKham: Bao gồm Id lượt khám, Id bệnh nhân, Bác sỹ , Thời gian, Tình trạng

DimTinhTrang: Bao gồm Id tình trạng và tên tình trạng.

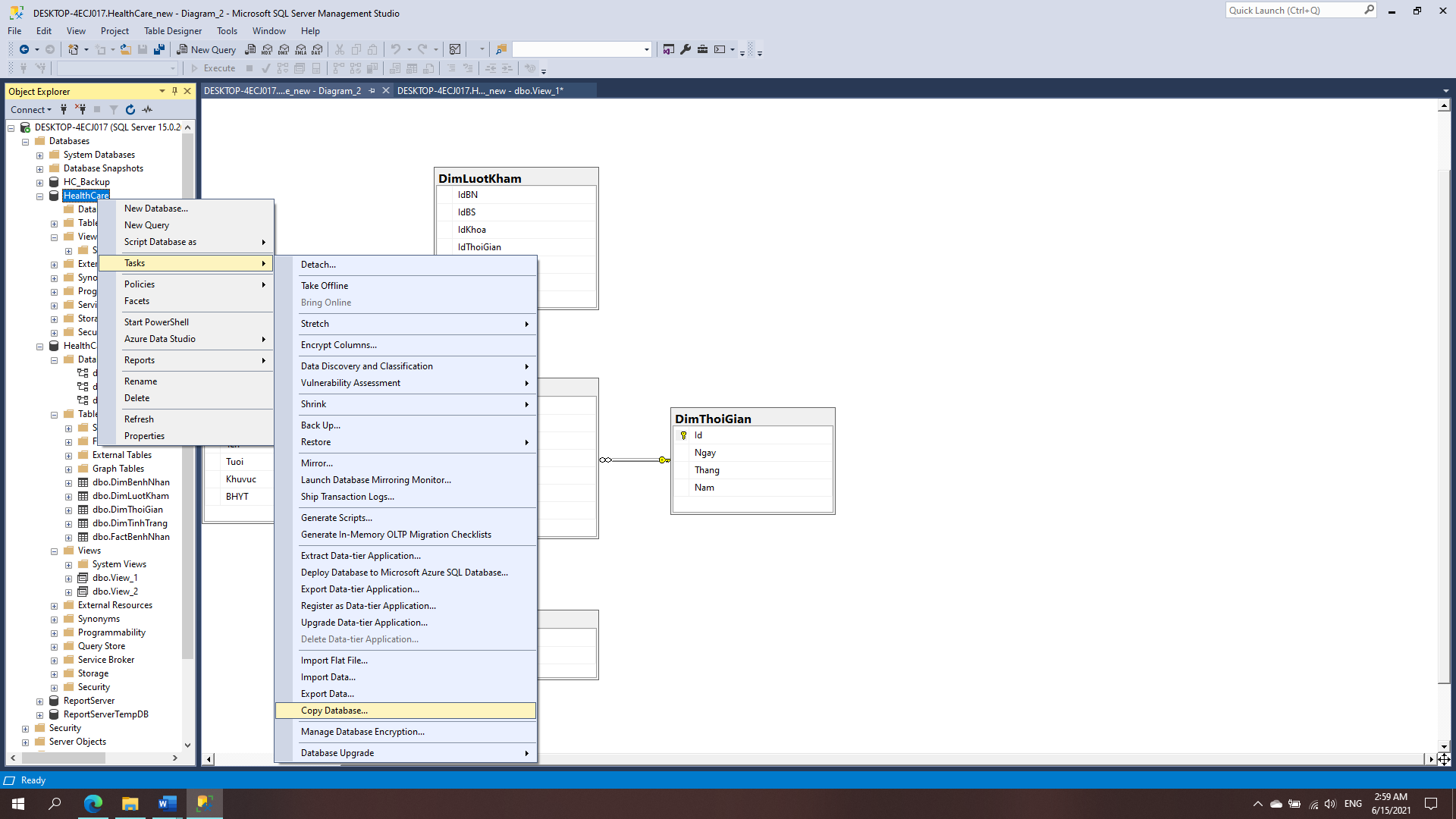
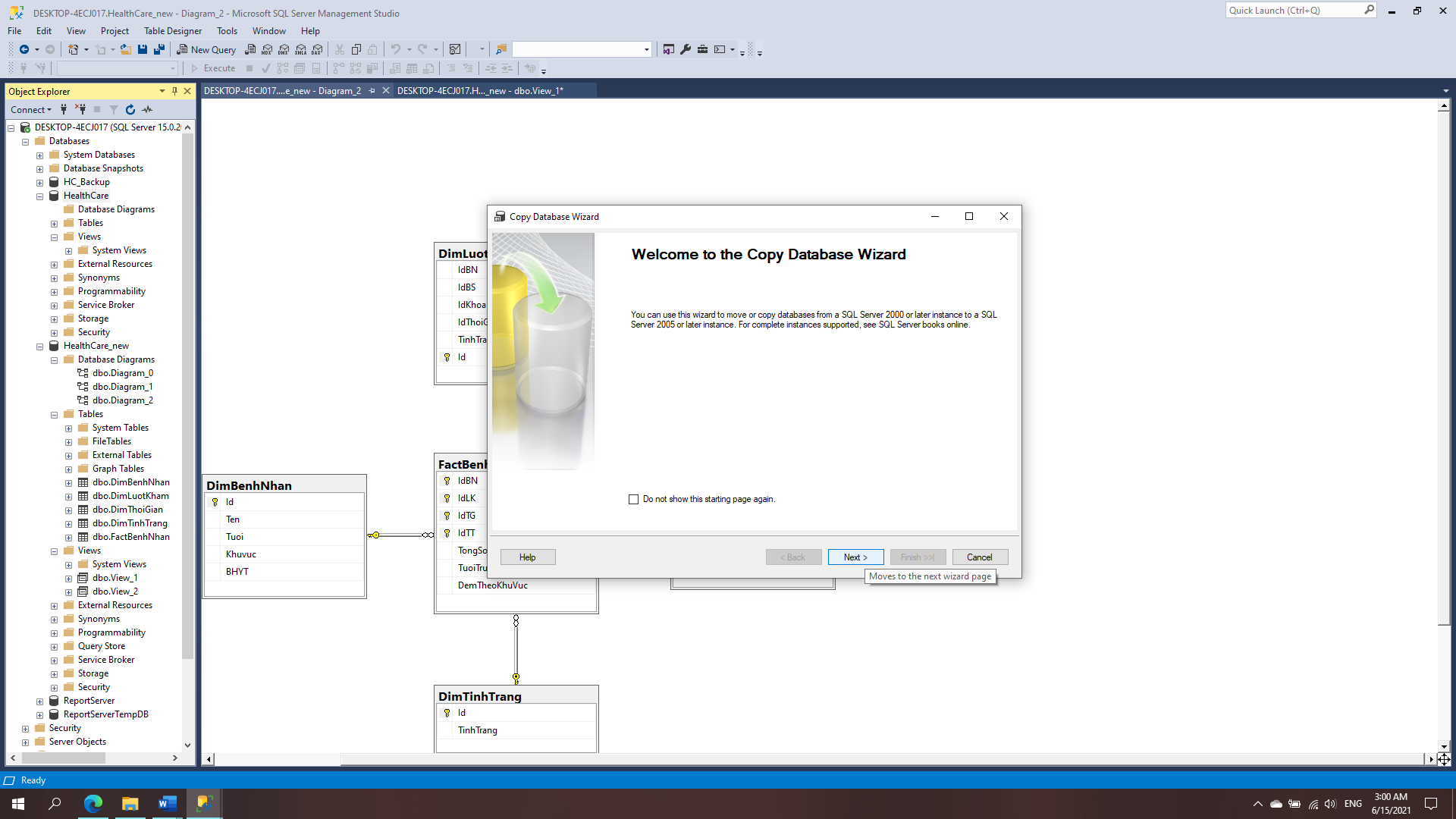
DimThoiGian: Bao gồm Ngày, tháng, năm.

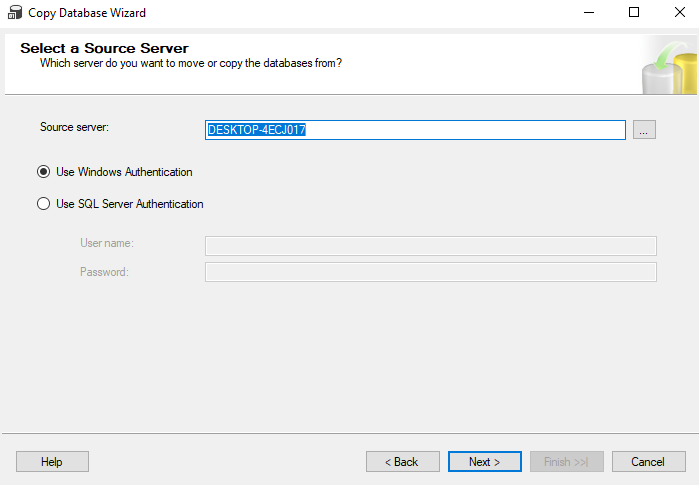
### Thiết kế DB mới tổ chức phân tích Kho dữ liệu cho phân hệ

DataWarehouse phân hệ phân tích **Bệnh nhân:**

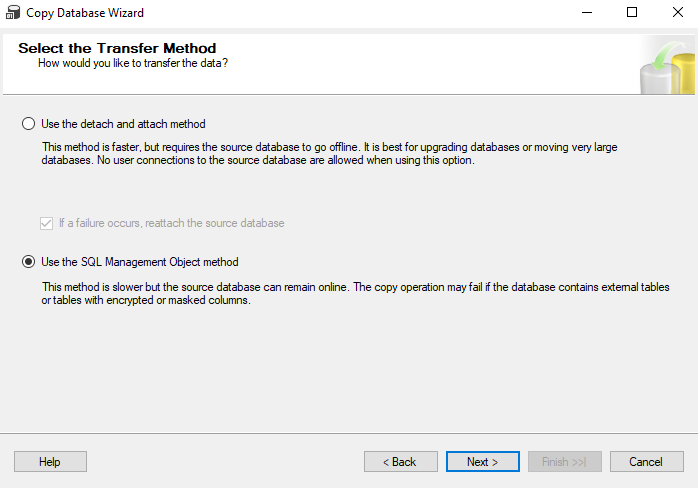
### Thực hiện các thủ tục tích hợp các DB gốc vào phân hệ DWH

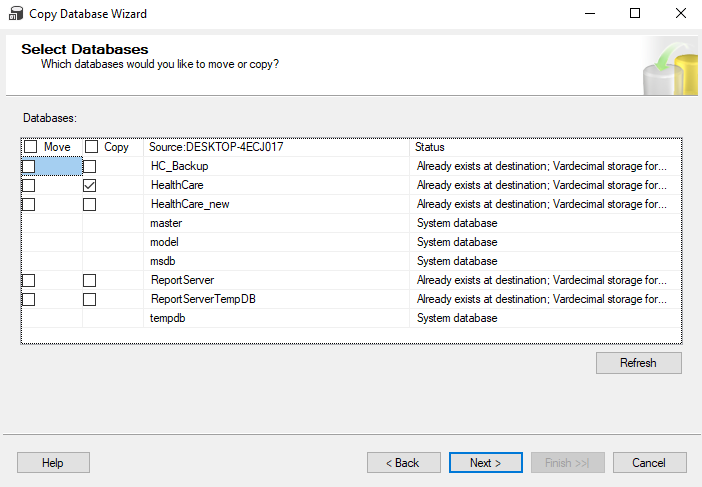
* Thực hiện các thủ tục Copy DataBase cần thiết (để lập bản sao phục vụ phân tích) từ DB gốc sang DB của DHW phân tích.

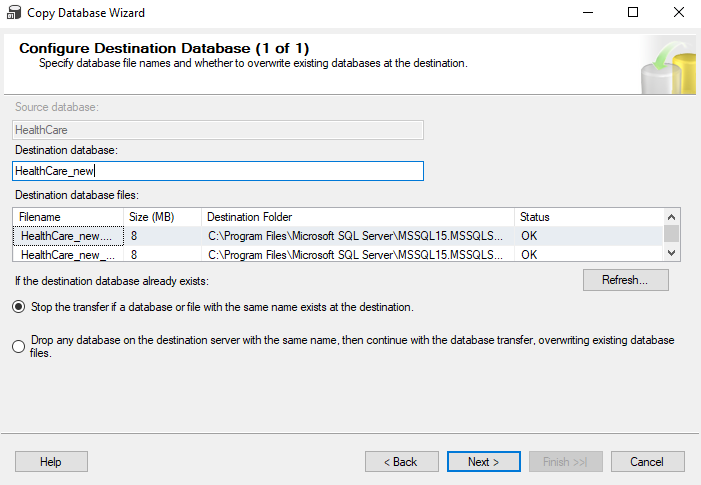


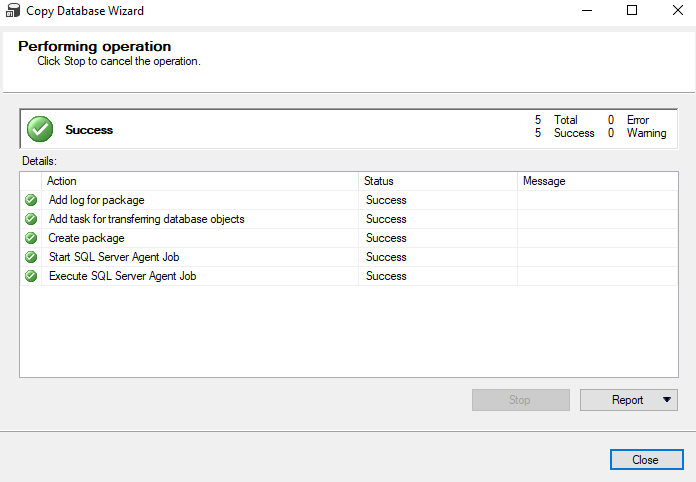


Check in Use the SQL management Object method: có nghĩa là database sau khi copy vẫn chạy một cách độc lập và không bị phụ thuộc vào database gốc và database gốc không bị tình trạng offline









2.Chuyển Database copy về DW

B1: Xóa các khóa ngoại

B2 Xóa các table không cần thiết , giữ lại và đổi tên các table cần dung cho phân hệ.

B3: Tạo bảng Fact và liên kết các khóa chính với các Dim



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

**B1: Tạo View và chọn các thuộc tính cho bảng Fact:**

SELECT bn.Id AS IdBN, tg.Id AS IdTG, CASE WHEN tt.Id IS NULL THEN 1221 ELSE tt.Id END AS IdTT, COUNT(lk.IdBN) AS TongSoLuotKham, AVG(bn.Tuoi) AS DoTuoiKhamTrungBinh, COUNT(DISTINCT bn.Khuvuc) AS DemTheoKhuVuc,

bn.Khuvuc AS [khu vuc]

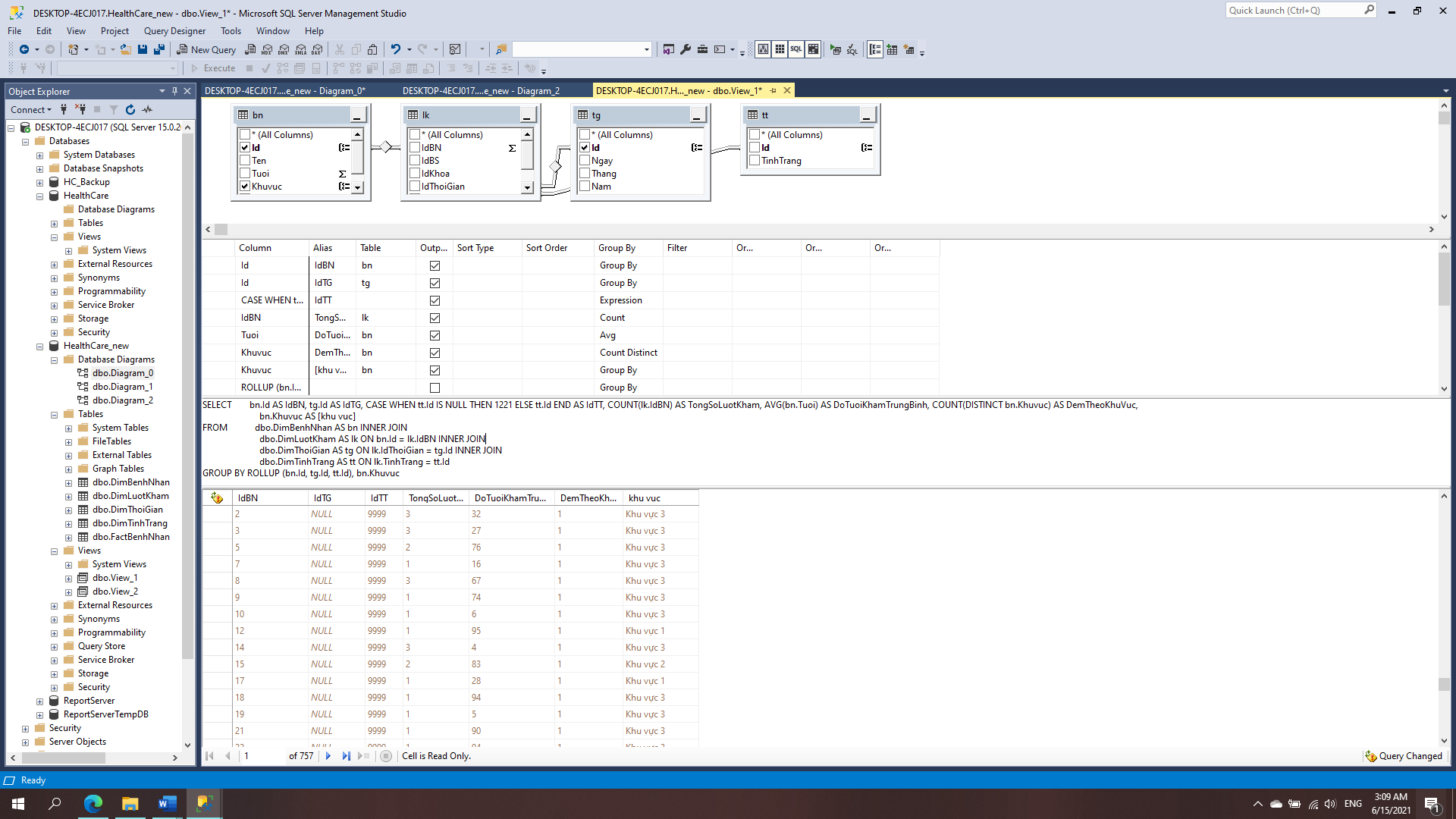
FROM dbo.DimBenhNhan AS bn INNER JOIN

dbo.DimLuotKham AS lk ON bn.Id = lk.IdBN INNER JOIN

dbo.DimThoiGian AS tg ON lk.IdThoiGian = tg.Id INNER JOIN

dbo.DimTinhTrang AS tt ON lk.TinhTrang = tt.Id

GROUP BY ROLLUP (bn.Id, tg.Id, tt.Id), bn.Khuvuc



View\_2:

SELECT lk.IdBN, lk.Id AS IdLK, lk.IdThoiGian AS IdTG, lk.TinhTrang AS IdTT, COUNT(lk.IdBN) AS TongSoLuotKham, AVG(bn.Tuoi) AS TuoiTrungBinh, bn.Khuvuc AS DemTheoKhuVuc

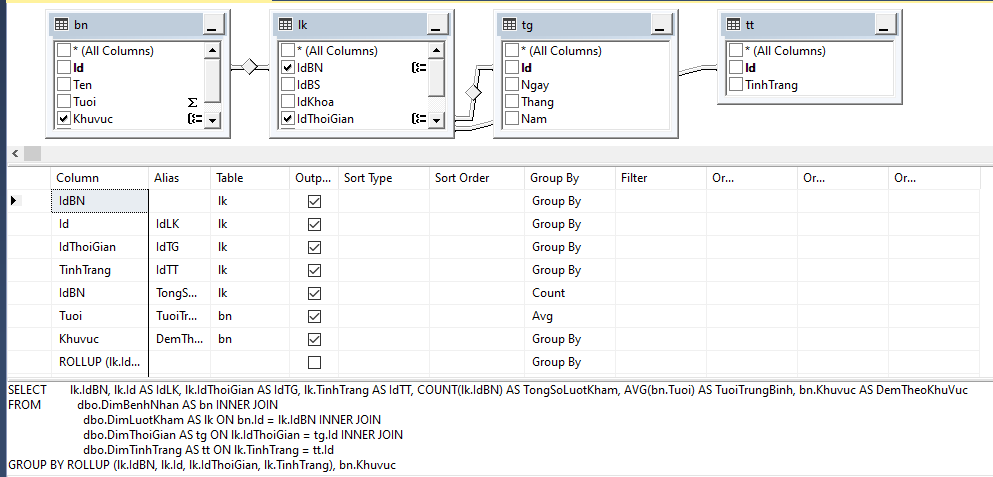
FROM dbo.DimBenhNhan AS bn INNER JOIN

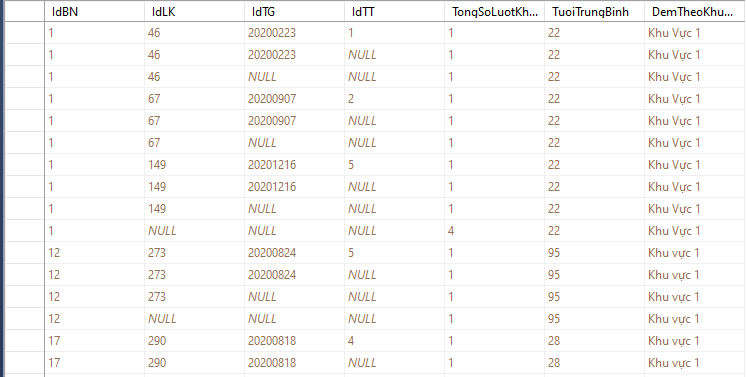
dbo.DimLuotKham AS lk ON bn.Id = lk.IdBN INNER JOIN

dbo.DimThoiGian AS tg ON lk.IdThoiGian = tg.Id INNER JOIN

dbo.DimTinhTrang AS tt ON lk.TinhTrang = tt.Id

GROUP BY ROLLUP (lk.IdBN, lk.Id, lk.IdThoiGian, lk.TinhTrang), bn.Khuvuc



**B2: Execute lệnh tạo View và kiểm tra kết quả:**

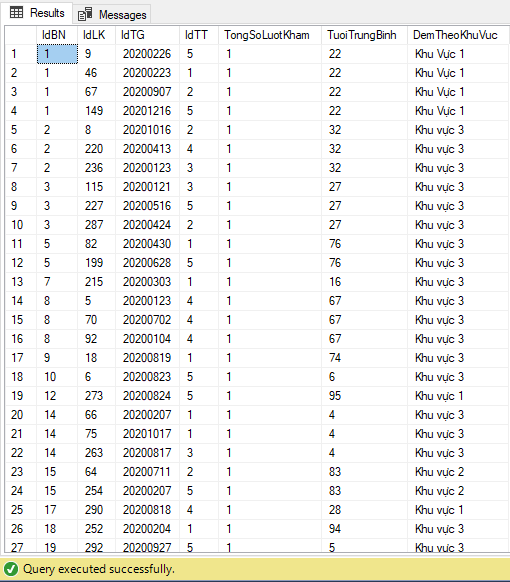
**B3: Insert dữ liệu từ View vào bảng Fact:**

INSERT INTO FactBenhNhan (IdBN,IdLK, IdTG, IdTT , [TongSoLuotKham],[TuoiTrungBinh],[DemTheoKhuVuc])

SELECT IdBN,IdLK,IdTG,IdTT,

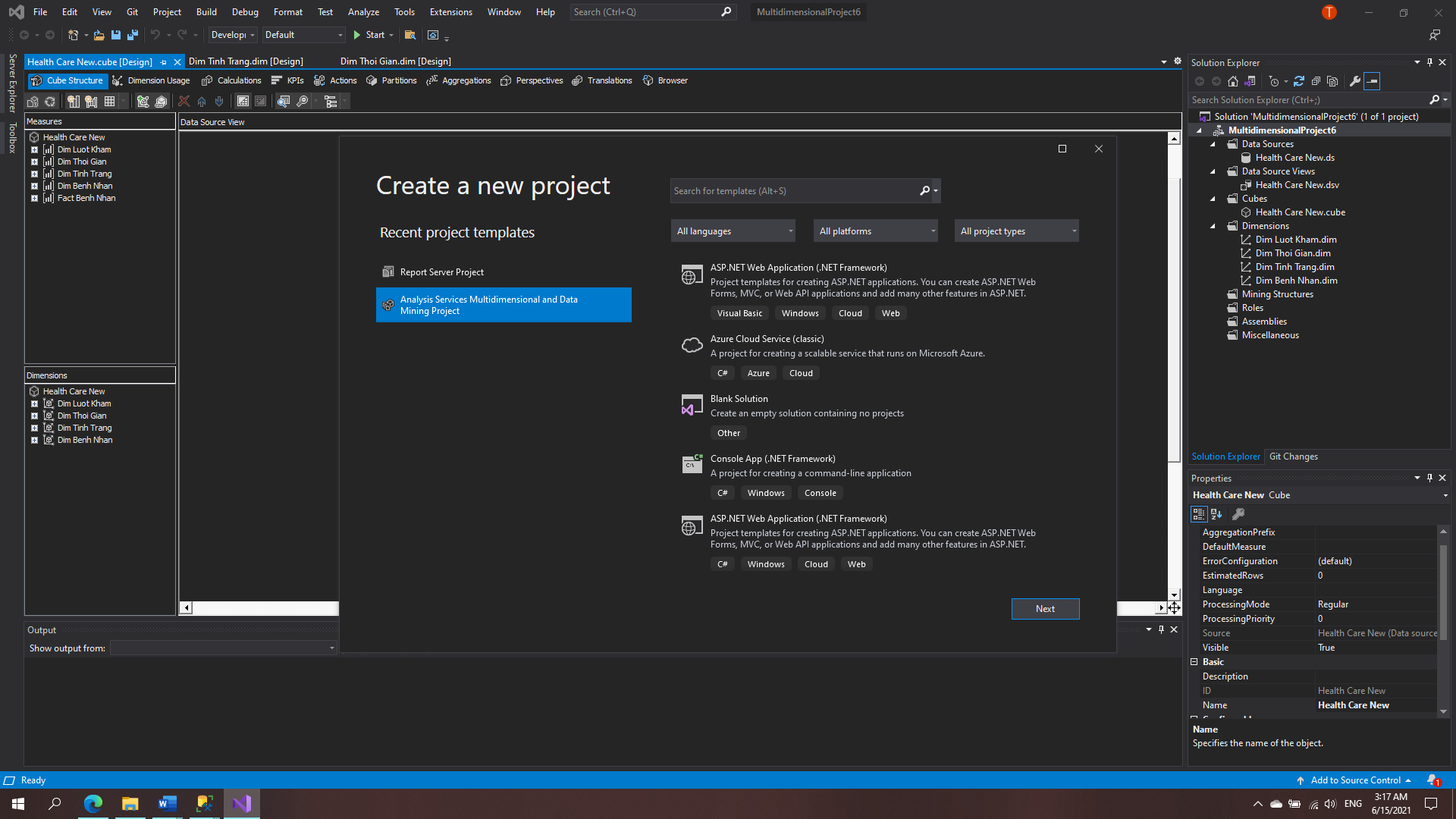
[TongSoLuotKham],[TuoiTrungBinh],[DemTheoKhuVuc]

FROM View\_2 Where IdTT is not NULL

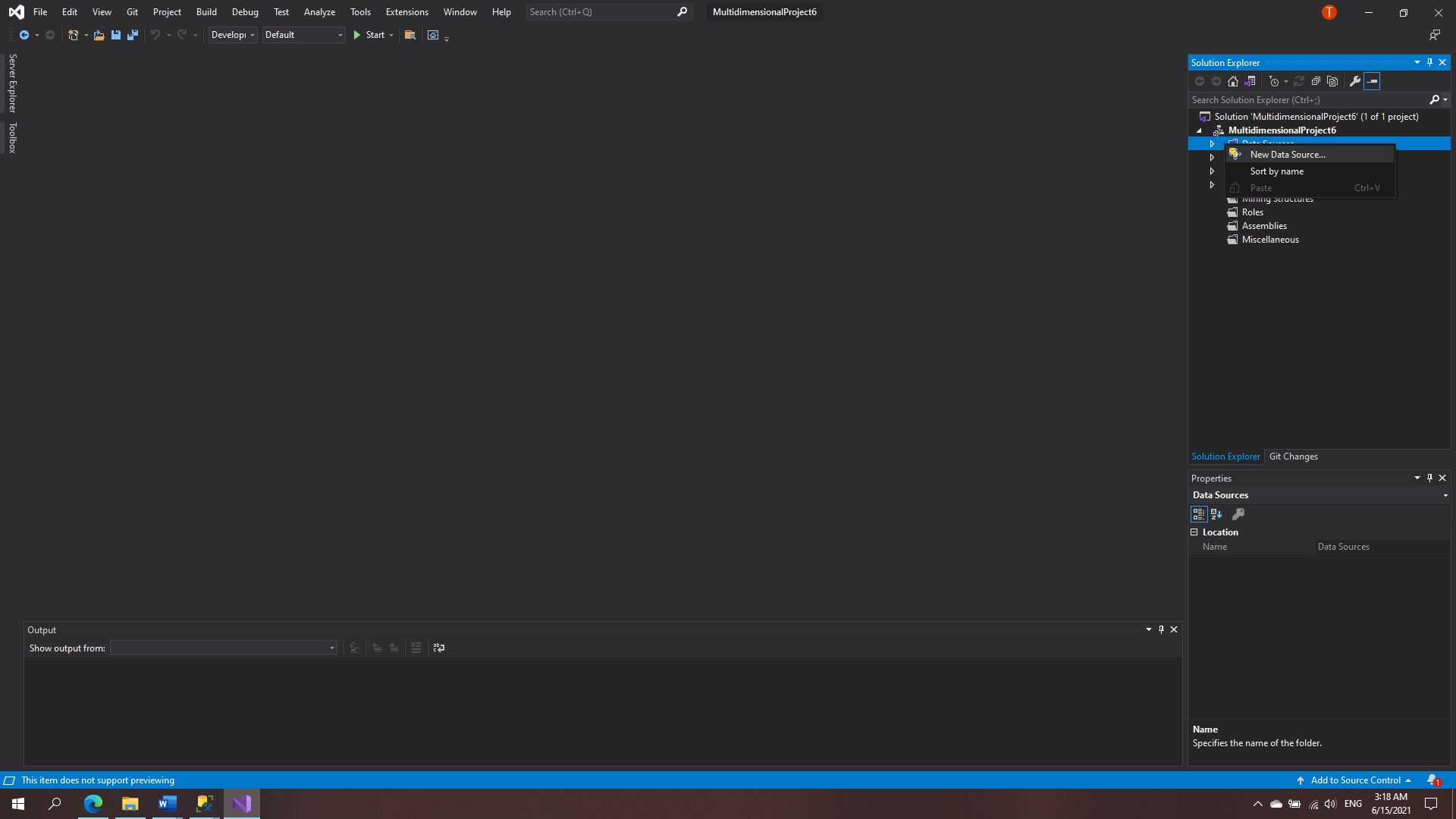


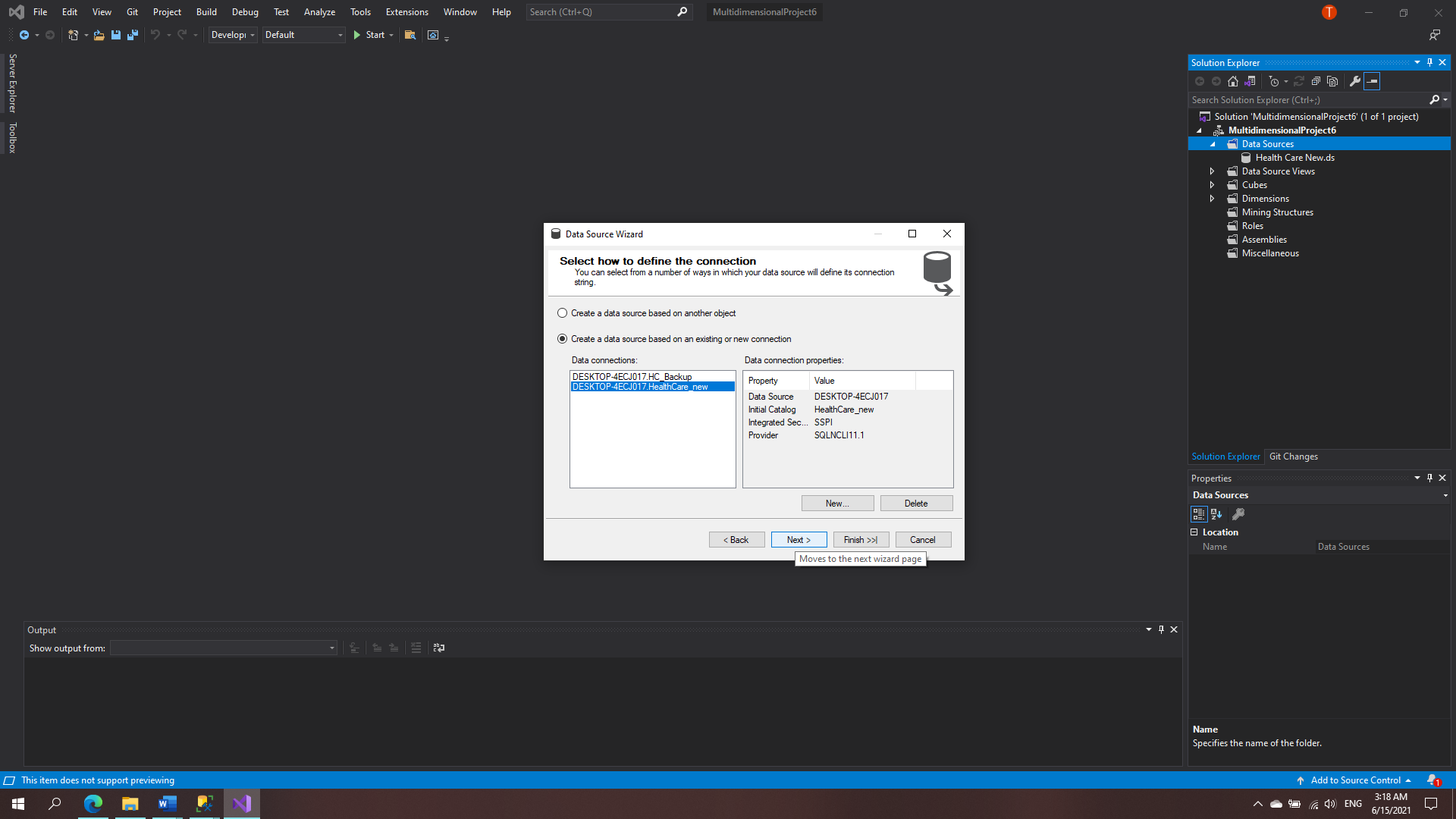
### Triển khai phân hệ DWH lên BIDS:

**B1: Tạo Project BIDS trên Microsoft Visual Studio 2020.**

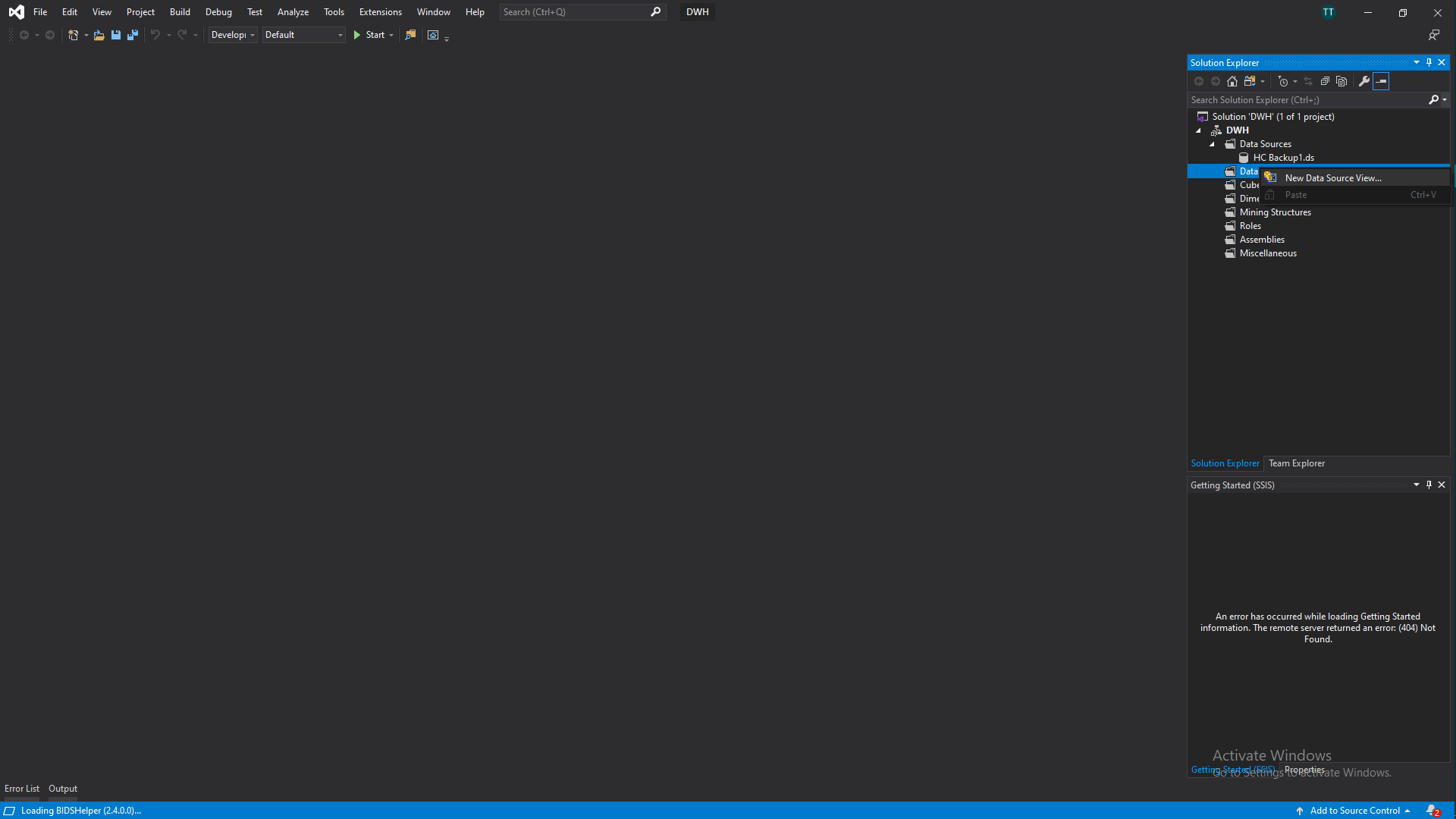


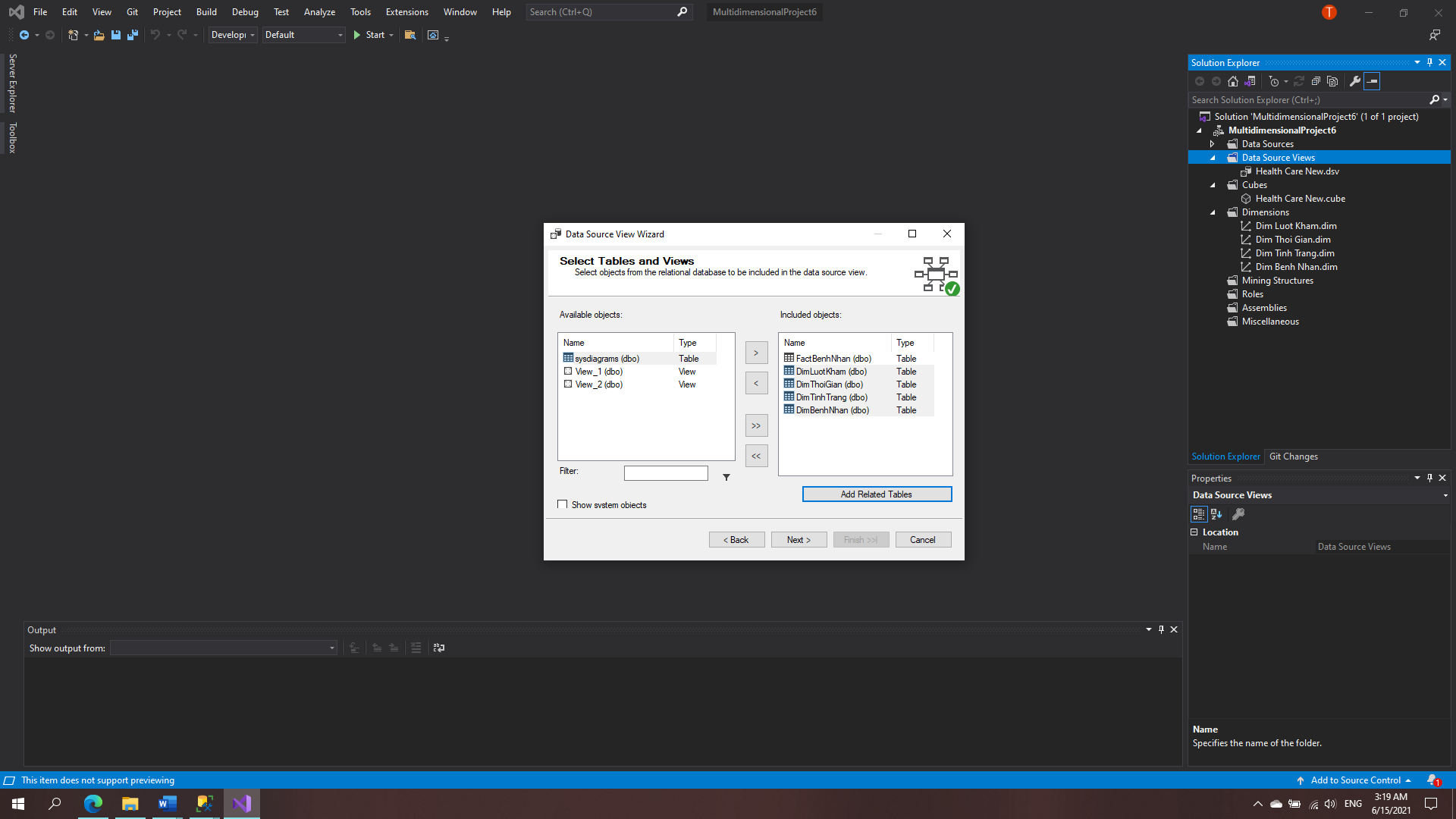
**B2: Tạo DataSource cho Project**





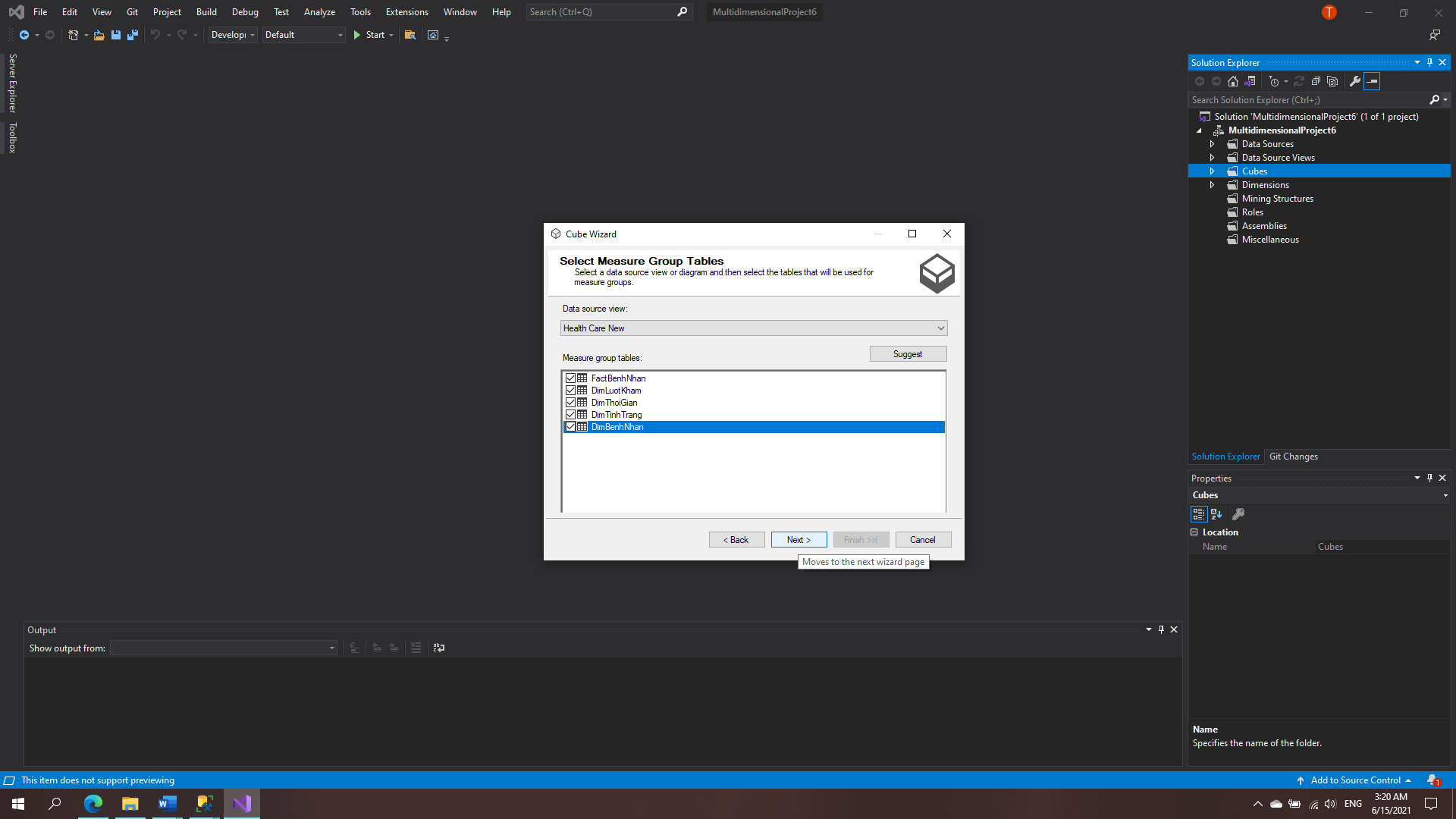
**B3: Tạo DataSourceView**



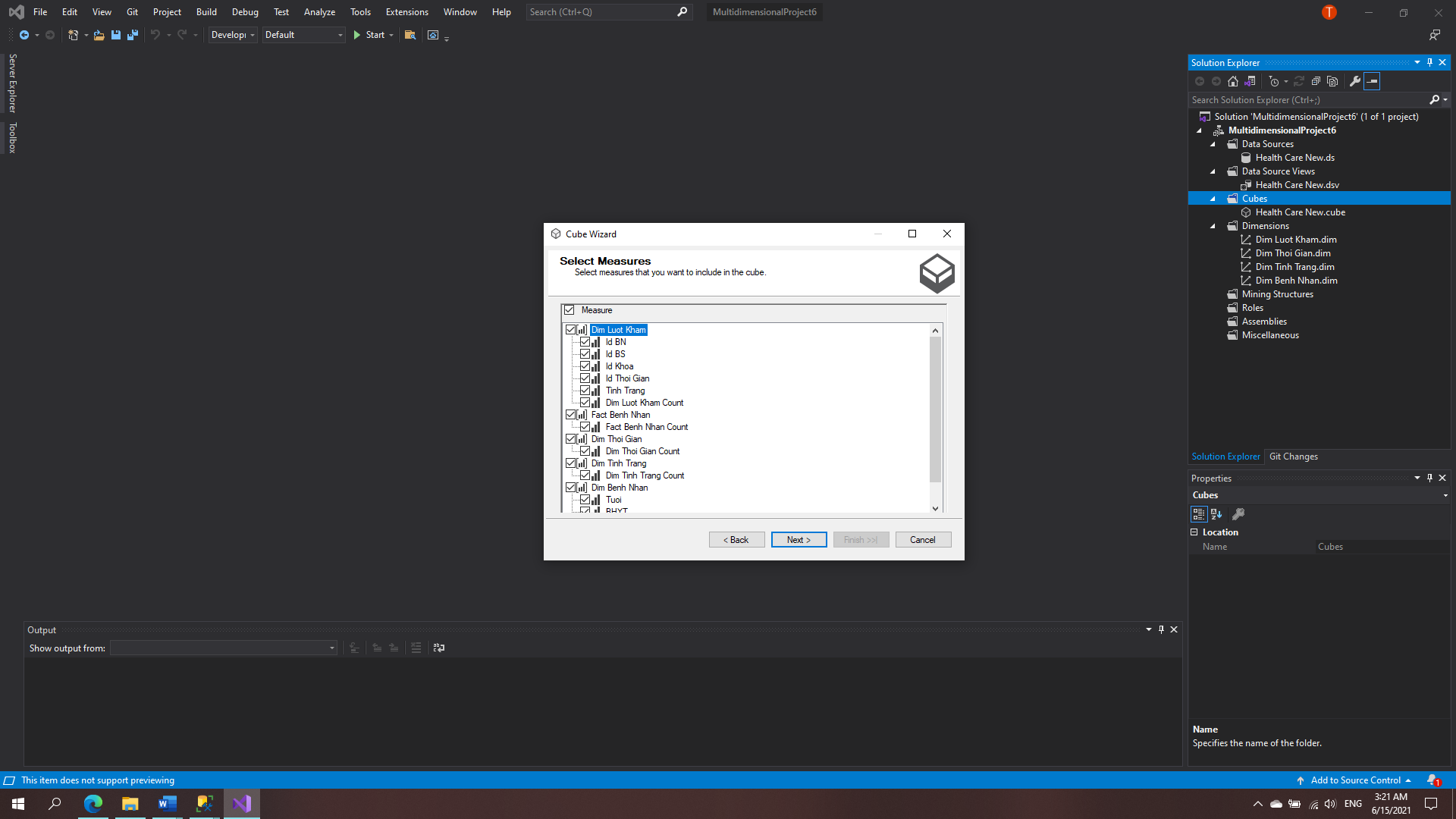


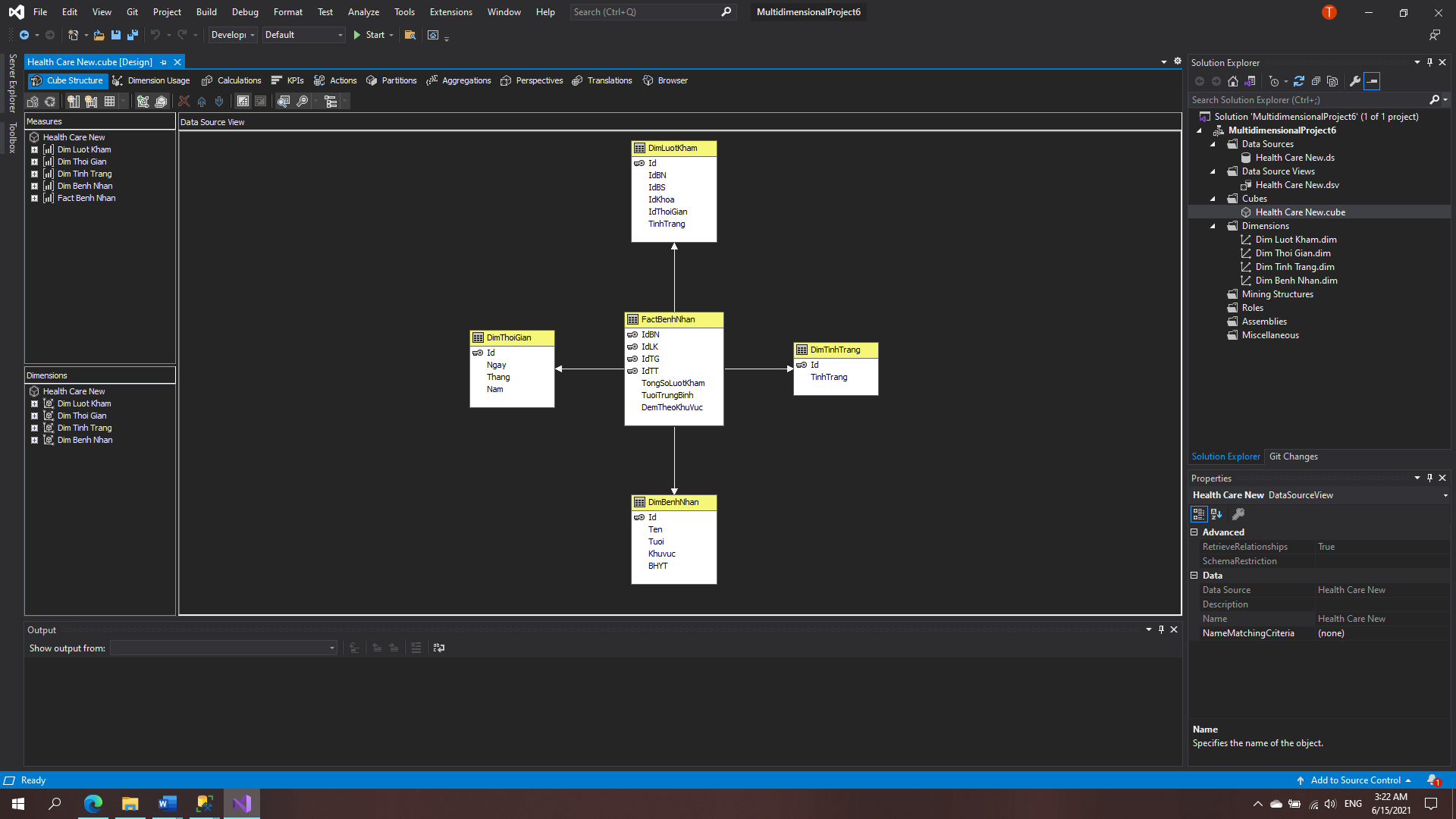
**B4: Tạo Cube:**

**Chọn Measure là bảng Fact:**



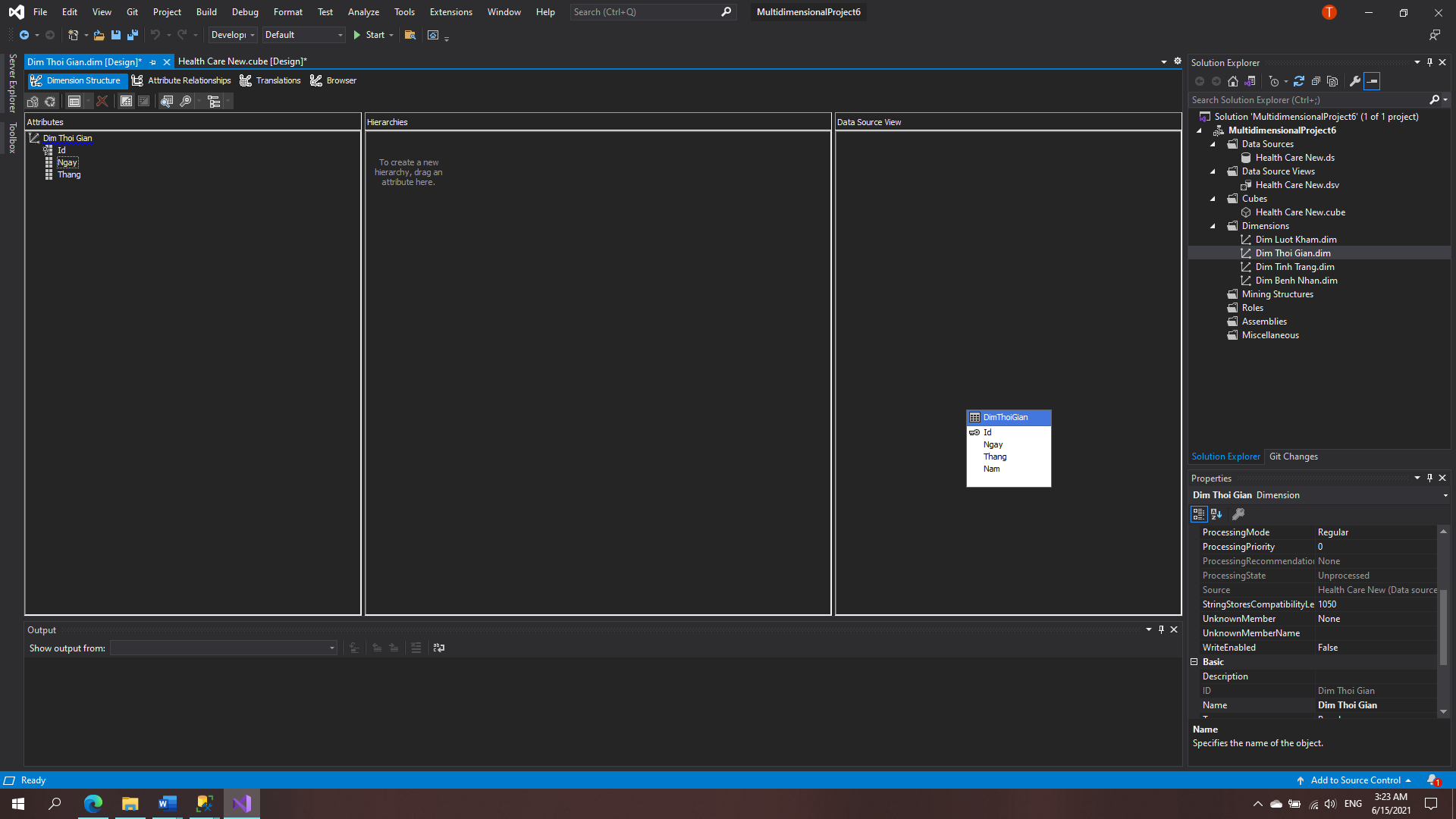
**Chọn Dimension là các bảng Dim**



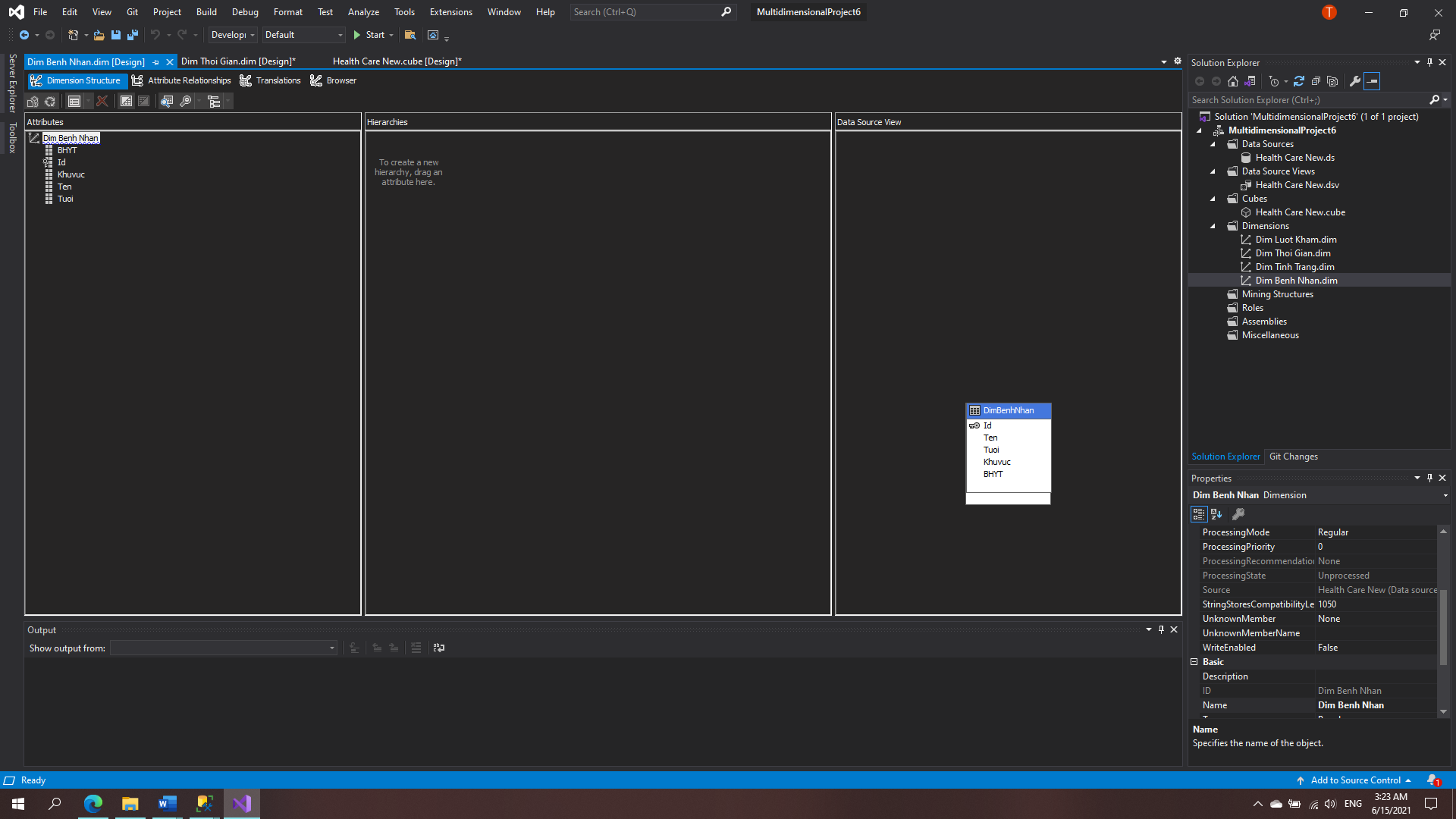


**B5: Xác định các Dimension cần phân tích:**

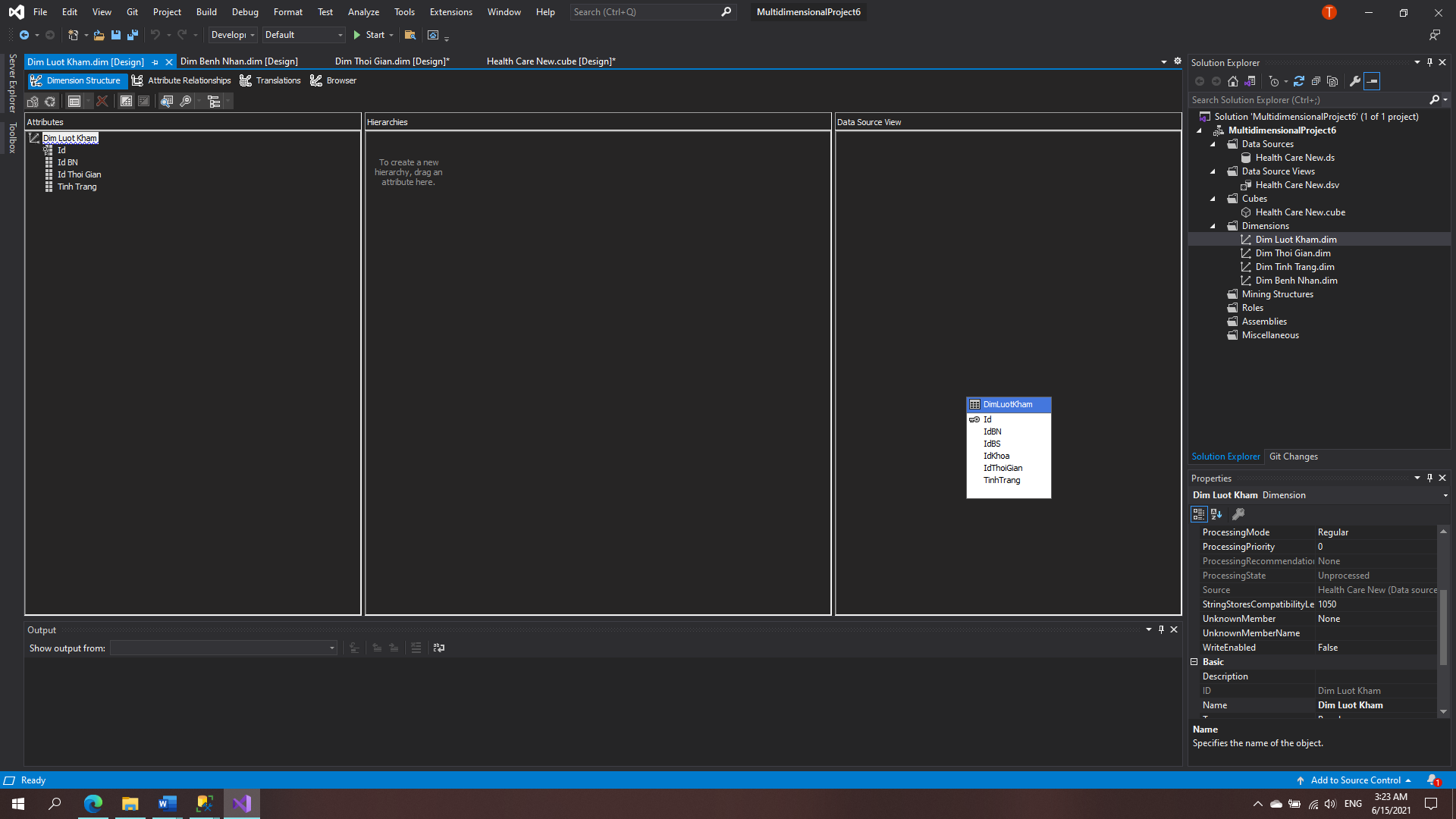
**DimThoiGian:**



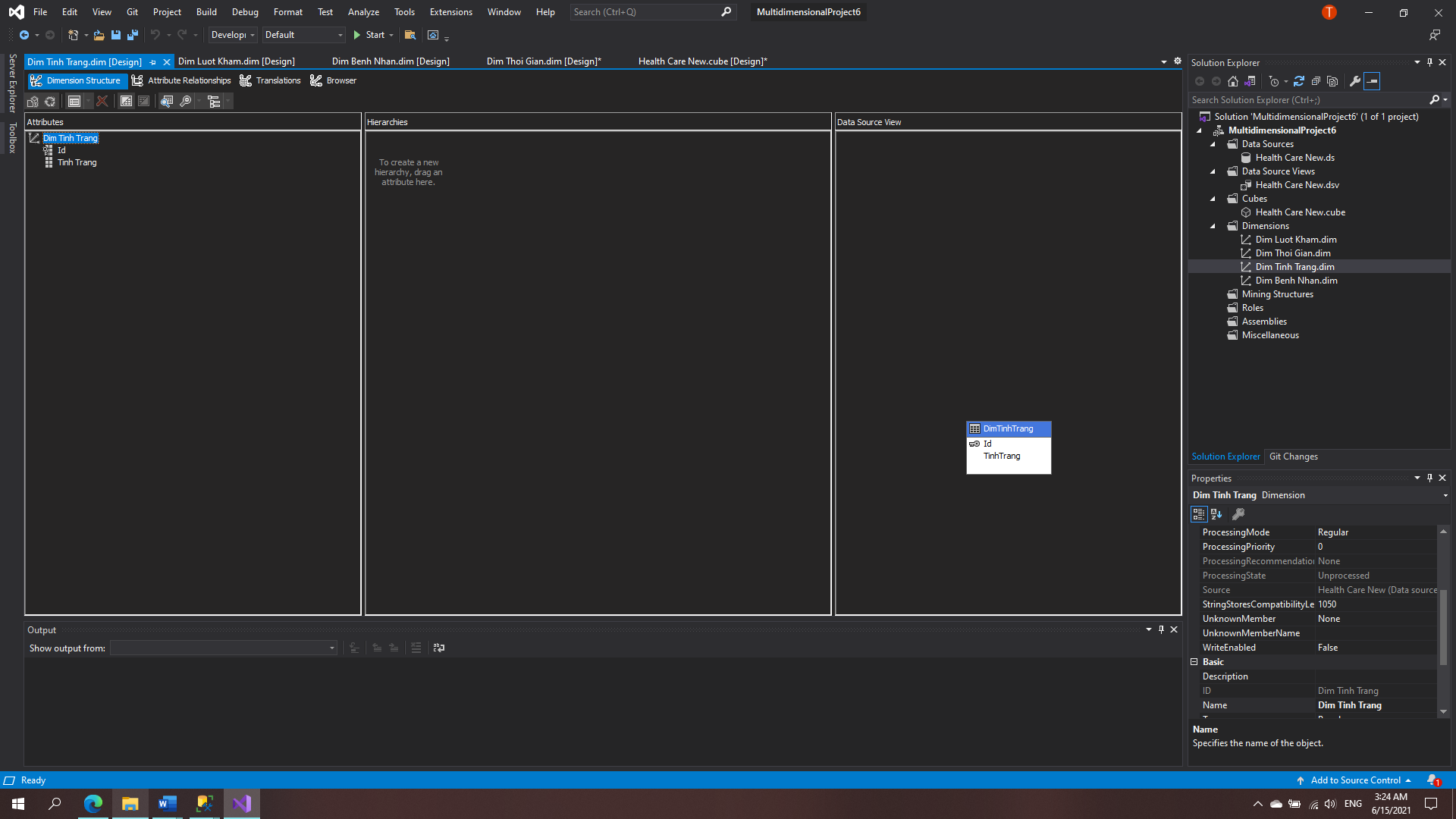
**DimBenhNhan:**



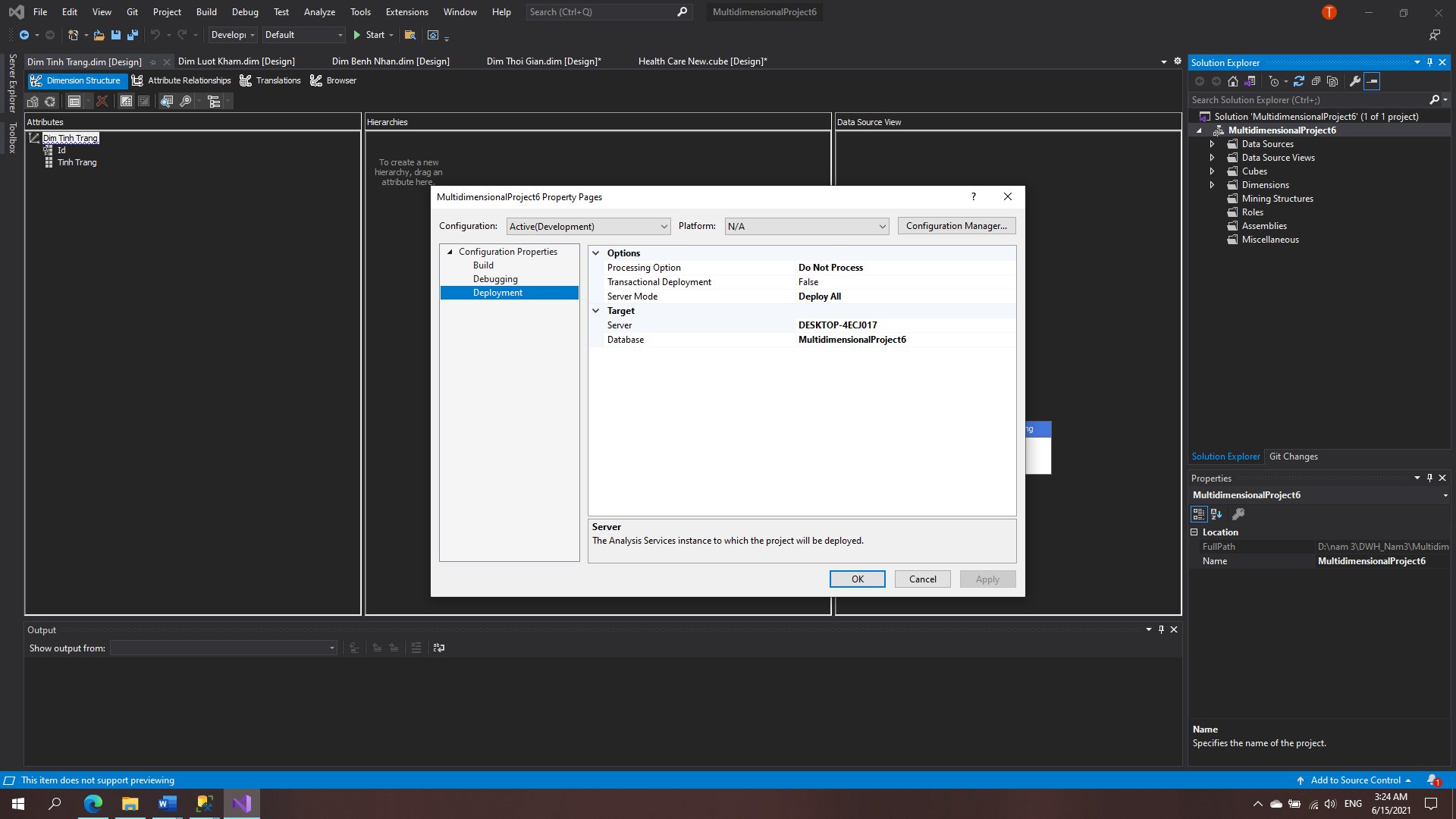
**DimLuotKham:**



**DimTinhTrang:**

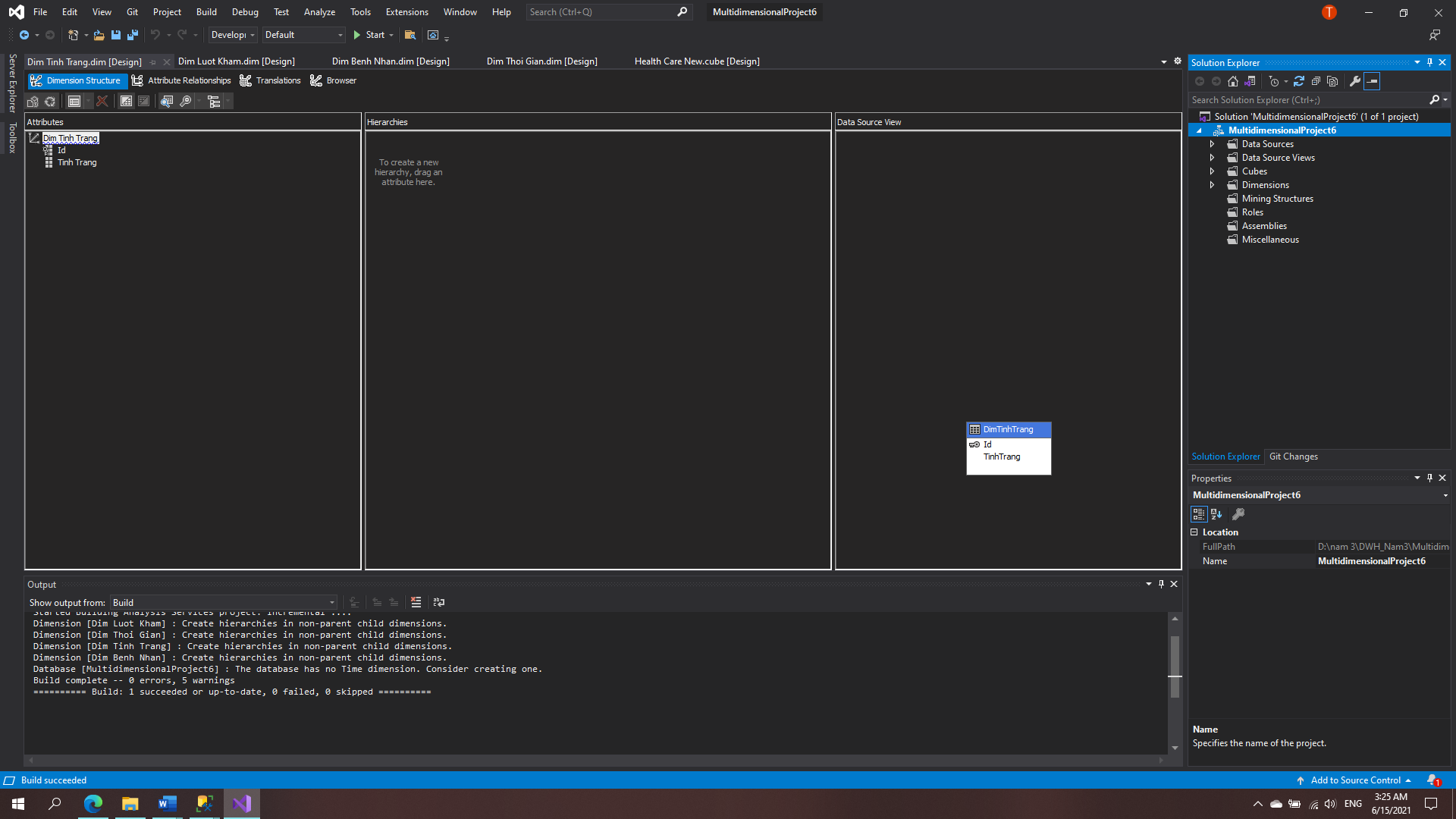


**B6: Thiết lập các Properties cho Project**

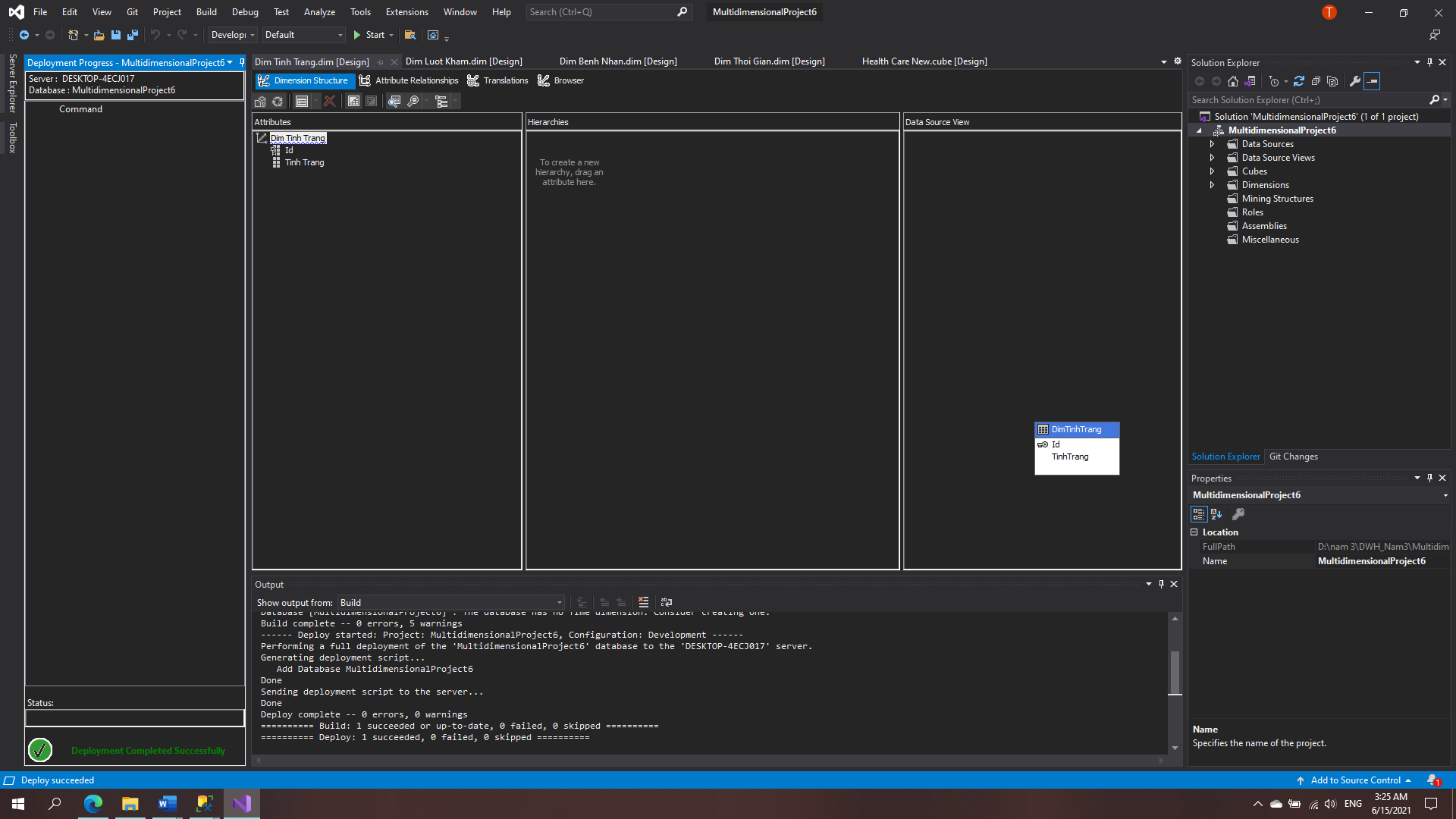


**B7: Process Project:**

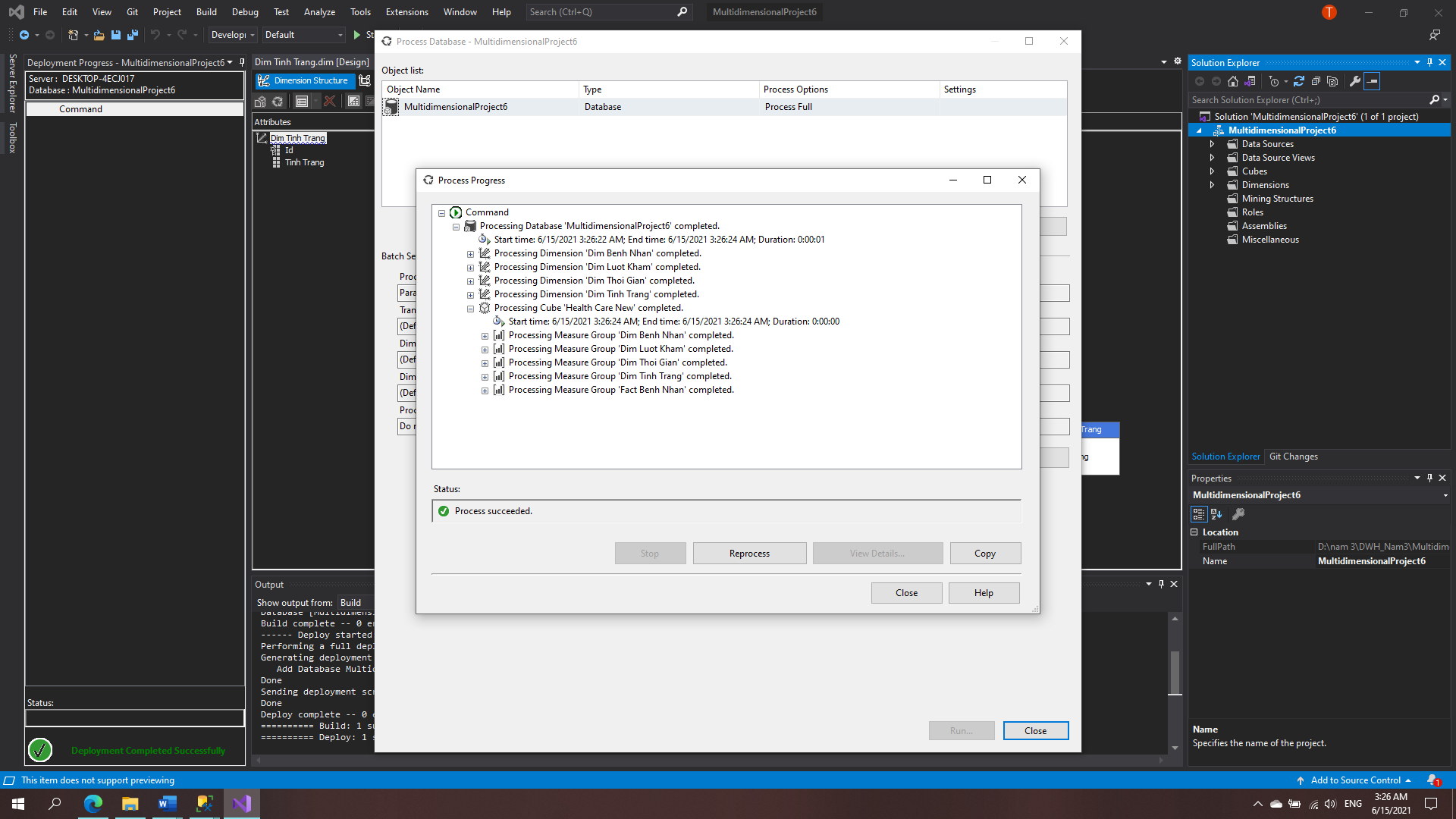
**Build :**



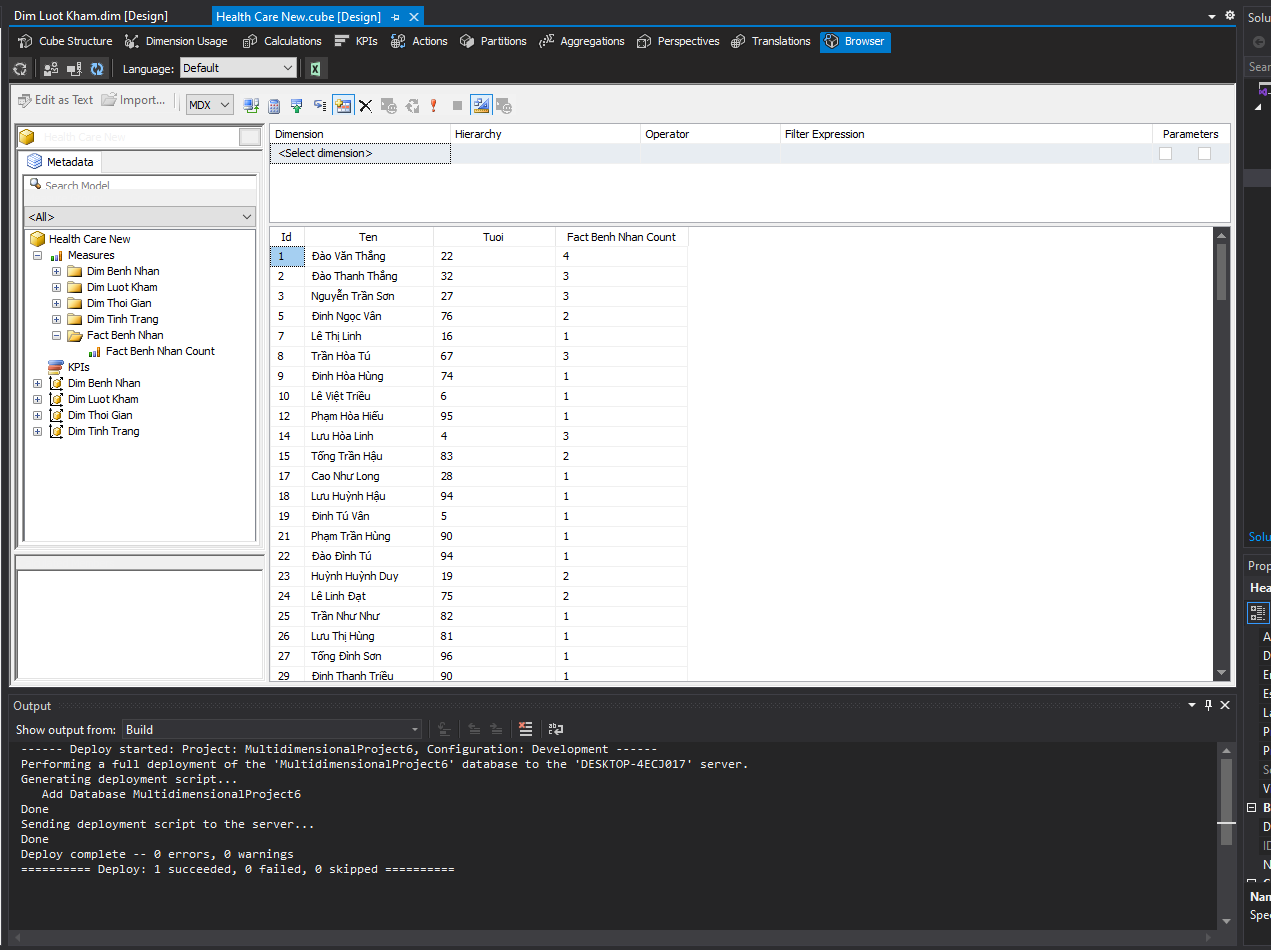
**Deloy:**



**Process:**



**B8: Vào Browser và phân tích dữ liệu**

****

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

**Câu truy vấn 1 : Tính số lần khám mỗi tháng của mỗi bệnh nhân :**

SELECT bn.Id, CASE WHEN tg.Thang IS NULL THEN 'TONGSOLUOTKHAM' ELSE bn.Ten END AS TÊN, bn.Tuoi, CASE WHEN tg.Thang IS NULL THEN 13 ELSE tg.Thang END AS THÁNG, COUNT(lk.Id) AS TongSoLuotKham

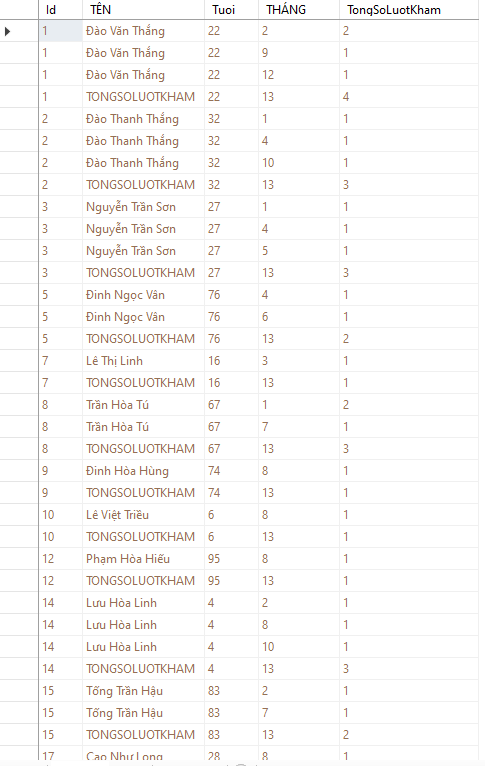
FROM dbo.DimBenhNhan AS bn INNER JOIN

dbo.DimLuotKham AS lk ON bn.Id = lk.IdBN INNER JOIN

dbo.DimThoiGian AS tg ON lk.IdThoiGian = tg.Id INNER JOIN

dbo.DimTinhTrang AS tt ON lk.TinhTrang = tt.Id

GROUP BY bn.Id, bn.Ten, bn.Tuoi, ROLLUP (tg.Thang)



***Câu truy vấn 2:***

**Xem coi số lượt khám ở mỗi tình trạng của 3 khu vực bệnh nhân và độ tuổi trung bình của bệnh nhân:**

SELECT CASE WHEN bn.Khuvuc IS NULL THEN 'Tổng Lượt Khám ' ELSE bn.Khuvuc END AS KhuVuc, tt.TinhTrang, COUNT(bn.Id) AS Expr1, AVG(bn.Tuoi) AS ĐộTuổiTrungBinh

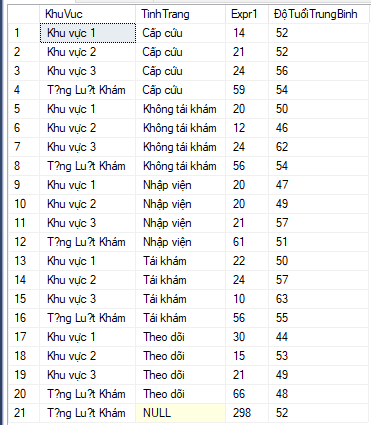
FROM dbo.DimBenhNhan AS bn INNER JOIN

dbo.DimLuotKham AS lk ON bn.Id = lk.IdBN INNER JOIN

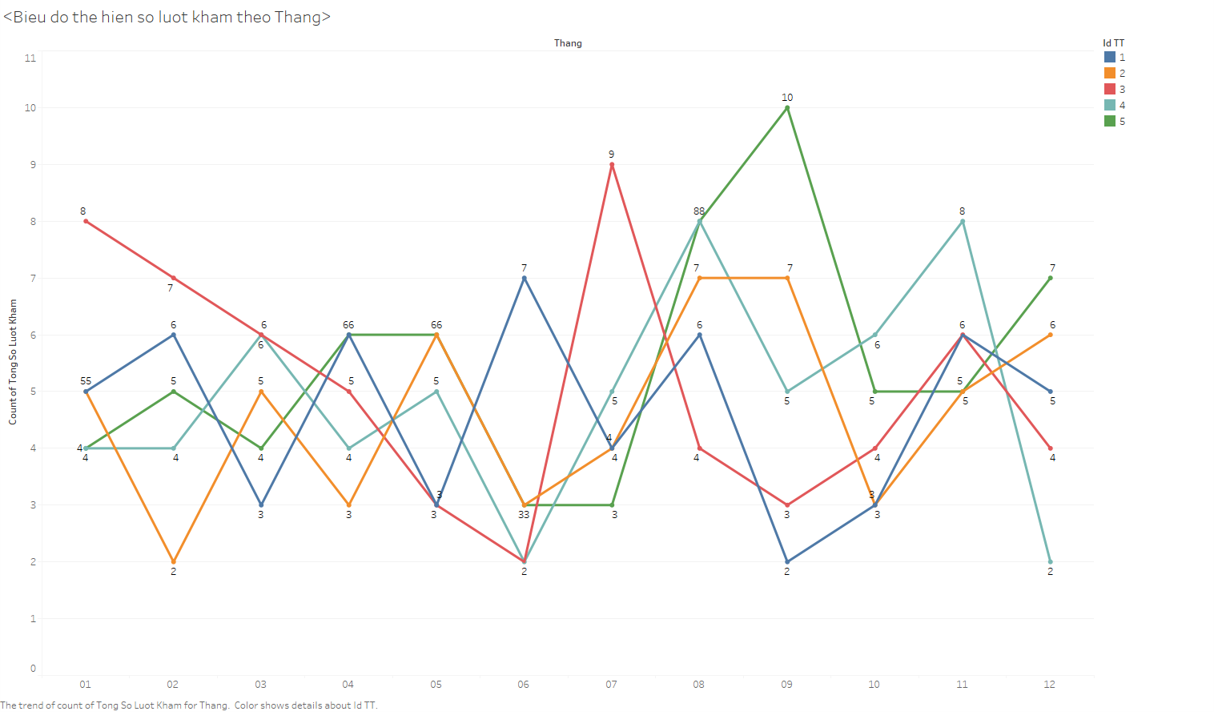
dbo.DimThoiGian AS tg ON lk.IdThoiGian = tg.Id INNER JOIN

dbo.DimTinhTrang AS tt ON lk.TinhTrang = tt.Id

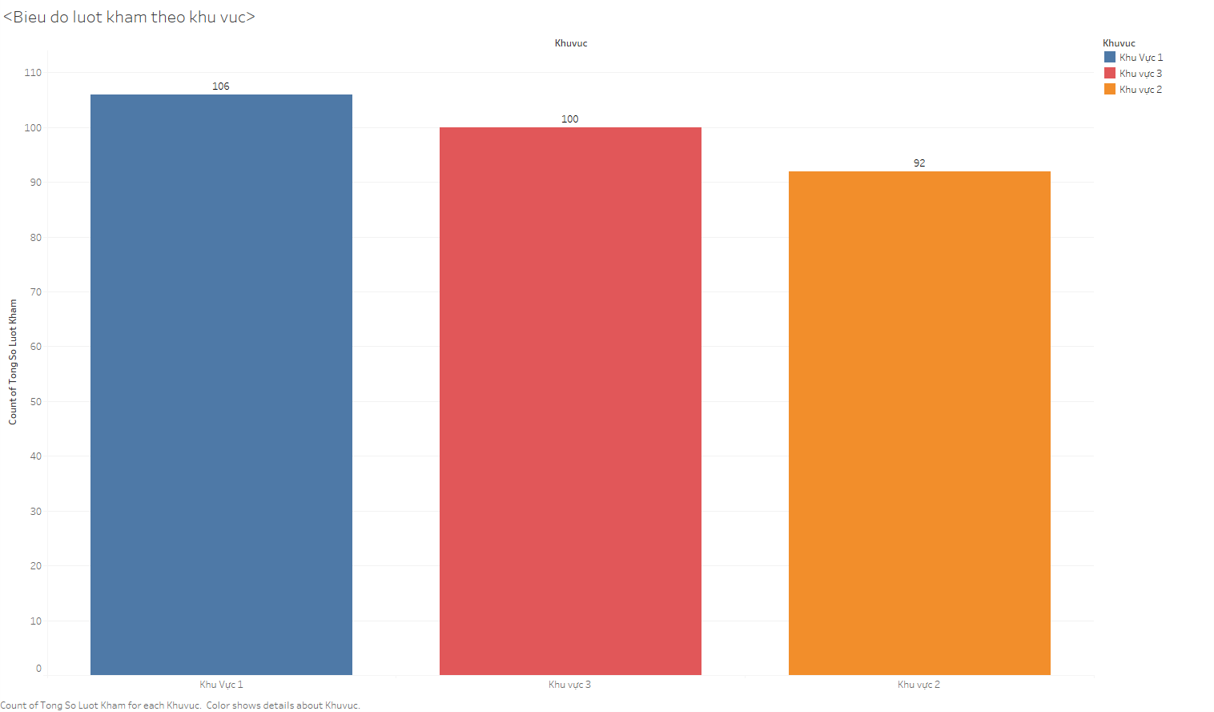
GROUP BY ROLLUP (tt.TinhTrang, bn.Khuvuc)



Biểu đồ thể hiện số lượt khám của 5 tình trạng qua 12 tháng của bệnh viện. Mỗi đường là một tình trạng có thể xem mỗi màu cho mỗi tình trạng, Ví dụ tháng 9 có số ca theo dõi nhiều nhất là 10 lượt , và tháng 2 có số ca không tái khám là nhiều nhất. Cần có sự chuẩn bị và điều chỉnh và chuẩn bị cho từng tình trạng của bệnh viện, dữ liệu quan sát giúp cho những người đứng đầu bệnh viện đưa ra điều chỉnh để có những biển pháp hợp lý cho từng Tháng.



* ***Biểu đồ Cột biểu thị số lượt khám theo từng khu vực 1,2,3:*** được biểu thị bởi 3 màu sắc khác nhau . Giúp cho nhà quản lý có cái nhìn tổng quát về 3 khu vực và mỗi khu vực có số lượng khám khác nhau . Mỗi một côt là một khu vực, thuốc tính của mỗi bệnh nhânkhi đến bệnh viện. Dữ liệu được thu thập từ DWH và xây dựng biếu độ bên dưới., màu xanh Dương là khu vực 1 , màu đỏ là khu vực 2 và màu cam là khu vực 3 .



# 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:

- Biết được khoa nào có lượt khám nhiều nhất và ít nhất

- Biết được khu vực nào có số bệnh nhân nhiều nhất.

- Biết được các loại bệnh thường gặp theo các mùa

- Biết được phân công cơ sở vật chất hợp lý.

- Từ những ý trên, các bệnh viện đưa ra các giải pháp và cải thiện chất lượng phục vụ hơn cho các bệnh nhân, nhất là trong thời gian dịch bệnh bùng phát.

### Hạn chế

* Hạn chế về độ chính xác của Data do Data của bệnh viện là dữ liệu mật, nên các dữ liệu trong bảng BenhNhan và LuotKham là dữ liệu được tạo ngẫu nhiên từ hàm Random trong Python. Tuy là dữ liệu ảo nhưng đã có điều chỉnh hợp lý để tăng độ chính xác dữ liệu cao nhất có thể.
* Chưa thể cập nhật dữ liệu theo thời gian thực.
* Thiết kế Database vẫn chưa được tối ưu hóa, vẫn còn gặp một số bugs lúc đầu, nhưng đã khắc phục và chỉnh sửa.

## Hướng phát triển

### Hướng khắc phục các hạn chế

* Khi đưa dự án vào thực tế thì vấn đề về dữ liệu sẽ được giải quyết vì lúc đó có thể sử dụng dữ liệu trực tiếp tại bệnh viện sử dụng dự án.
* Tiếp thu các cách xây dựng Database trên các kênh để có thể tối ưu hóa và dễ dàng truy xuất.

### Hướng mở rộng ĐỒ ÁN HỌC PHẦN

+ Thay vì sử dụng bộ nhớ ổ cứng một máy tính, ta có thể lưu phân tán dữ liệu khi dữ liệu quá nhiều bằng Framework Hadoop Distributed File System để tăng khả năng chị lỗi cũng như bảo mật dữ liệu.

+ Sử dụng Framework Spark và Framework HiveQL để tạo các Data Pipeline giúp dữ liệu trong DWH được cập nhật một cách tuần tự và nhanh chóng.

# 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 |