Shape

Description automatically generated

**BỘ GIÁO DỤC VÀ ĐÀO TẠO**

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

**KHOA CÔNG NGHỆ THÔNG TIN**

**MÔN KHO DỮ LIỆU**

Logo

Description automatically generated-----🙞🙜🕮🙞🙜-----

**BÁO CÁO ĐỀ TÀI CUỐI KỲ**

**⁃⁃⁃⁃⁃⁃⁃⁃⁃⁃⁃**

**XÂY DỰNG KHO DỮ LIỆU   
QUẢN LÝ KHÁCH SẠN**

**LỚP HỌC PHẦN: DAWH430784\_23\_2\_03CLC**

**HỌC KỲ II – NĂM HỌC 2023-2024**

**Thực hiện: Nhóm 05**

**Giảng viên hướng dẫn : Ths. Nguyễn Văn Thành**

*Thành phố Hồ Chí Minh, tháng 5 năm 2024*

|  |  |
| --- | --- |
| **BỘ GIÁO DỤC VÀ ĐÀO TẠO**  **ĐẠI HỌC SƯ PHẠM KỸ THUẬT TPHCM**  **KHOA CÔNG NGHỆ THÔNG TIN**  **MÔN KHO DỮ LIỆU**  **­­­­­­­­­­­­­­­** | **CỘNG HÒA XÃ HỘI CHỦ NGHĨA**  **VIỆT NAM**  **Độc lập – Tự do – Hạnh phúc** |

-----🙞🙜🕮🙞🙜-----

Logo

Description automatically generated

**BÁO CÁO ĐỀ TÀI CUỐI KỲ**

**⁃⁃⁃⁃⁃⁃⁃⁃⁃⁃⁃**

**XÂY DỰNG KHO DỮ LIỆU   
QUẢN LÝ KHÁCH SẠN**

**LỚP HỌC PHẦN: DAWH430784\_23\_2\_03CLC**

**HỌC KỲ II – NĂM HỌC 2023-2024**

**Thực hiện: Nhóm 05**

**Giảng viên hướng dẫn : Ths. Nguyễn Văn Thành**

*Thành phố Hồ Chí Minh, tháng 5 năm 2024*

|  |  |
| --- | --- |
| **BỘ GIÁO DỤC VÀ ĐÀO TẠO**  **ĐẠI HỌC SƯ PHẠM KỸ THUẬT TPHCM**  **KHOA CÔNG NGHỆ THÔNG TIN**  **MÔN KHO DỮ LIỆU**  **­­­­­­­­­** | **CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM**  **Độc lập – Tự do – Hạnh phúc** |

**DANH SÁCH NHÓM THỰC HIỆN ĐỀ TÀI**

**MÔN KHO DỮ LIỆU**

**HỌC KỲ II NĂM HỌC 2023-2024**

1. **Giảng viên hướng dẫn:** Ths. Nguyễn Văn Thành
2. **Tên đề tài :** Xây dựng kho dữ liệu quản lý khách sạn
3. **Danh sách nhóm viết tiểu luận cuối kỳ:**

|  |  |  |  |
| --- | --- | --- | --- |
| **STT** | **HỌ VÀ TÊN SINH VIÊN** | **Mã số sinh viên** | **Tỉ lệ % tham gia** |
| 01 | Trương Thị Thuỳ Dung | 21110820 | 100% |
| 02 | Trần Văn Bảo Duy | 21110155 | 100% |
| 03 | Bùi Đặng Thuỳ Thương | 21110316 | 100% |
| 04 | Đinh Thị Thúy Quỳnh | 21110284 | 100% |

**Nhận xét của giảng viên:**

**………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………**

*Ngày … tháng 05 năm 2024*

*Giảng viên chấm điểm*

**LỜI CẢM ƠN**

Chúng em xin gửi lời cảm ơn sâu sắc đến Thầy Nguyễn Văn Thành, Thầy đã tận tình giảng dạy trong suốt quá trình tìm hiểu và học tập môn Kho dữ liệu, Trường đại học Sư phạm kỹ thuật Thành phố Hồ Chí Minh. Trong quá trình nghiên cứu và học, Thầy đã nhiệt huyết giảng dạy và những kiến thức mà Thầy đã truyền đạt không chỉ là nền tảng cho quá trình nghiên cứu, thực hiện đồ án mà còn là tư liệu quý báu để chúng em trong quá trình học tập và tìm hiểu sau này cũng như là trên con đường nghề nghiệp phía trước.

Một lần nữa, em xin bày tỏ lòng biết ơn đến Thầy. Xin kính chúc Thầy luôn dồi dào sức khỏe và đạt được nhiều thành công trong công việc.

*Thành phố Hồ Chí Minh, tháng 5 năm 2024*

**KẾ HOẠCH PHÂN CÔNG NHIỆM VỤ THỰC HIỆN ĐỒ ÁN CUỐI KÌ MÔN KHO DỮ LIỆU**

**HỌC KỲ II NĂM HỌC 2023-2024**

1. **Tên đề tài:** Xây dựng kho dữ liệu quản lý khách sạn
2. **Bảng phân công nhiệm vụ:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Trương Thị Thùy Dung** | **Đinh Thị Thúy Quỳnh** | **Bùi Đăng Thùy Thương** | **Trần Văn Bảo Duy** |
| **NHIỆM VỤ** |  |  |  |  |
| **Tìm kiếm tập dữ liệu** | x | x | x | x |
| **Mô hình hóa dữ liệu vào workbook** | x | x | x | x |
| * Dim Customers * Dim Employees |  |  |  | x |
| * Dim Date |  |  | x |  |
| * Fact Sales |  | x |  |  |
| * Fact Bookings | x |  |  |  |
| **Thực hiện ETL đưa dữ liệu vào các dim và fact dùng SSIS** | x |  | x |  |
| * Tạo và đưa dữ liệu vào các dim stages và dim | x |  | x |  |
| * Tạo và đưa dữ liệu vào stgHotelFactBookingvà FactBookings | x |  |  |  |
| * Tạo và đưa dữ liệu vào stgHotelFactSales và FactSales |  |  | x |  |
| * Tạo và đưa dữ liệu vào SSAS, Tạo data source view và cube |  |  |  | x |
| **Tạo cây phân cấp hierarchy cho các bảng dim và các attribute relationships.** |  | x |  | x |
| * Hierarchy cho DimDate |  | x |  |  |
| * Hierarchy cho DimCustomers |  |  |  | x |
| * Hierarchy cho DimEmployees |  |  |  | x |
| **Đặt và thực hiện trả lời các câu hỏi cho Business process: Booking Transaction: Quản lý lượng booking của khách sạn** | x | x | x | x |
| **Thực hiện trả lời các câu hỏi đã đặt ra cho Business process: Booking Transaction: Quản lý lượng booking của khách sạn** | x | x | x | x |
| **Đặt và thực hiện trả lời các câu hỏi cho Business process: Sales Management: Quản lý doanh số của khách sạn** | x | x | x | x |
| **Thực hiện trả lời các câu hỏi đã đặt ra cho Business process: Sales Management: Quản lý doanh số của khách sạn** | x | x | x | x |
| Trả lời câu hỏi dùng SSAS | x |  |  | x |
| Trả lời câu hỏi dùng Power BI Desktop |  | x | x |  |
| Trả lời câu hỏi dùng Excel Pivot | x | x |  |  |

**MỤC LỤC**

-----🙞🙜🕮🙞🙜-----

[**CHƯƠNG 1: TỔNG QUAN VỀ ĐỀ TÀI** 1](#_Toc167024251)

[**1.1.** **Lý do chọn đề tài** 1](#_Toc167024252)

[**1.2.** **Tổng quan về tập dữ liệu** 1](#_Toc167024253)

[**1.2.1.** **Nguồn dữ liệu** 1](#_Toc167024254)

[**1.2.2.** **Mô tả chi tiết tập dữ liệu** 1](#_Toc167024255)

[**CHƯƠNG 2: THIẾT KẾ XÂY DỰNG CƠ SỞ DỮ LIỆU TÁC NGHIỆP (OLAP)** 5](#_Toc167024256)

[**2.1.** **Xác định các Business Process và bảng Fact** 5](#_Toc167024257)

[**2.2.** **Xây dựng các bảng Dimension** 6](#_Toc167024258)

[**2.3.** **Star schema (Lược đồ hình sao)** 7](#_Toc167024259)

[**CHƯƠNG 3: TÍCH HỢP DỮ LIỆU VÀO KHO (SSIS)** 8](#_Toc167024260)

[**3.1.** **Nhập dữ liệu vào bảng Date Dimension** 8](#_Toc167024261)

[**3.1.1.** **Tạo package mới và thiết lập control flow** 8](#_Toc167024262)

[**3.1.2.** **Load dữ liệu từ nguồn vào bảng stgDate** 9](#_Toc167024263)

[**3.1.3.** **Thiết lập the truncate table SQL task** 10](#_Toc167024264)

[**3.1.4.** **Load dữ liệu từ bảng stgDate vào bảng DimDate** 11](#_Toc167024265)

[**3.2.** **Cấu hình cho bảng FactBookings** 13](#_Toc167024266)

[**3.2.1.** **Cấu hình staging cho FactBookings** 13](#_Toc167024267)

[**3.2.2.** **Load dữ liệu vào bảng FactBookings** 16](#_Toc167024268)

[**3.3.** **Cấu hình cho bảng FactSales** 24](#_Toc167024269)

[**3.3.1.** **Cấu hình staging cho FactSales** 24](#_Toc167024270)

[**3.3.2.** **Load dữ liệu vào bảng FactSales** 28](#_Toc167024271)

[**CHƯƠNG 4: PHÂN TÍCH DỮ LIỆU (SSAS)** 35](#_Toc167024272)

[**4.1.** **Xây dựng mô hình** 35](#_Toc167024273)

[**4.1.1.** **Tạo Data Source** 35](#_Toc167024274)

[**4.1.2.** **Tạo Data SourceView** 37](#_Toc167024275)

[**4.2.** **Tạo Data cube** 40](#_Toc167024276)

[**4.3.** **Cấu hình Hierarchy** 42](#_Toc167024277)

[**4.3.1.** **Tạo Hierarchy Dim Date** 42](#_Toc167024278)

[**4.3.2.** **Tạo Hierarchy cho Dim Employees** 43](#_Toc167024279)

[**4.3.3.** **Tạo Hierarchy cho Dim Customers** 44](#_Toc167024280)

[**4.4.** **Thực hiện tạo Cube** 44](#_Toc167024281)

[**4.5.** **Thực hiện phân tích dữ liệu** 45](#_Toc167024282)

[**4.5.1.** **Sử dụng SSAS** 45](#_Toc167024283)

[**4.5.2.** **Sử dụng Pivot Table trong Excel và Power BI** 47](#_Toc167024284)

[**CHƯƠNG 5. KẾT LUẬN** 57](#_Toc167024285)

[**5.1.** **Kết quả đạt được** 57](#_Toc167024286)

[**5.2.** **Những hạn chế** 57](#_Toc167024287)

# 

# **CHƯƠNG 1: TỔNG QUAN VỀ ĐỀ TÀI**

* 1. **Lý do chọn đề tài**

Việc chọn đề tài xây dựng kho dữ liệu cho nhà hàng khách sạn với ứng dụng ETL và các công cụ như SSAS và SSIS mang lại nhiều lợi ích quan trọng. Đầu tiên, ngành nhà hàng và khách sạn thường phải quản lý lượng dữ liệu lớn từ nhiều nguồn khác nhau như đặt phòng, dịch vụ ăn uống và phản hồi của khách hàng.

Việc sử dụng quy trình ETL giúp thu thập, chuyển đổi và nạp dữ liệu vào kho một cách hiệu quả, đảm bảo chất lượng và tính toàn vẹn của dữ liệu. SSIS hỗ trợ xây dựng và thực thi các quy trình ETL, trong khi SSAS cung cấp khả năng phân tích dữ liệu mạnh mẽ, tạo ra các khối dữ liệu phân tích để đưa ra quyết định dựa trên dữ liệu.

Điều này không chỉ tối ưu hóa hoạt động kinh doanh mà còn cải thiện trải nghiệm khách hàng thông qua các chiến dịch tiếp thị tùy chỉnh và dịch vụ cá nhân hóa. Hơn nữa, việc ứng dụng các công cụ công nghệ này giúp nâng cao kỹ năng chuyên môn, đáp ứng nhu cầu ngày càng tăng về phân tích dữ liệu trong ngành Hệ thống thông tin**.**

* 1. **Tổng quan về tập dữ liệu**
     1. **Nguồn dữ liệu**

Nhóm sử dụng Tập dữ liệu Hotel Database được lấy từ trang web Kaggle (kaggle.com), trang web này cung cấp cơ sở dữ liệu được tổng hợp bởi các nguồn dữ liệu lớn và đáng tin cậy.

Đường dẫn tải tập dữ liệu: [Hotel booking analysis Database](https://www.kaggle.com/code/aminizahra/hotel-booking-analysis)

* + 1. **Mô tả chi tiết tập dữ liệu**

Tập dữ liệu gồm thông tin của một khách sạn và việc đặt phòng của khách, dữ liệu gồm 7 bảng

|  |  |  |
| --- | --- | --- |
| **Tên bảng** | **Mô tả** | **Số dòng** |
| BedTypes | Lưu trữ thông tin loại giường | 11 dòng |
| Bookings | Lưu trữ thông tin đặt phòng của khách | 400 dòng |
| Customers | Lưu trữ thông tin của khách | 400 dòng |
| Employees | Lưu trữ thông tin nhân viên | 59 dòng |
| Payments | Lưu trữ thông tin thanh toán của khách | 99 dòng |
| Rooms | Lưu trữ thông tin về phòng | 200 dòng |
| RoomTypes | Lưu trữ thông tin về loại phòng | 1. dòng |

* + - 1. **Bảng BedTypes**

|  |  |
| --- | --- |
| **Thuộc tính** | **Mô tả** |
| BedTypeID | Mã loại giường (Khóa chính) |
| BedType | Tên loại giường |
| Notes | Ghi chú |
| BedCost | Giá của loại giường |

* + - 1. **Bảng Bookings**

|  |  |
| --- | --- |
| **Thuộc tính** | **Mô tả** |
| BookingID | Mã đặt phòng của khách (Khóa chính) |
| ArrvialDate | Ngày đến |
| DepartureDate | Ngày đi |
| RoomNumber | Số phòng |
| CustomerID | Mã khách hàng (Khóa ngoại tham chiếu đến Customers(CustomerID)) |
| ReservationStatus | Trạng thái đặt phòng |
| ReservationStatusDate | Ngày cập nhật trạng thái |
| EmployeeID | Mã nhân viên ( Khóa ngoại tham chiếu đến Employees(EmployeeID) |
| GuestCount | Số lượng khách |

* + - 1. **Bảng Customers**

|  |  |
| --- | --- |
| **Thuộc tính** | **Mô tả** |
| CustomerID | Mã khách hàng (Khóa chính) |
| FirstName | Tên của khách |
| LastName | Họ của khách |
| BirthDate | Ngày sinh của khách |
| Email | Email của khách |
| Phone | Số điện thoại của khách |

* + - 1. **Bảng Employees**

|  |  |
| --- | --- |
| **Thuộc tính** | **Mô tả** |
| EmployeeID | Mã nhân viên (Khóa chính) |
| EmployeeName | Tên nhân viên |
| Title | Vị trí công việc của nhân viên |
| City | Tên thành phố |
| Country | Tên quốc gia |
| HireDate | Ngày tuyển dụng làm việc của nhân viên |

* + - 1. **Bảng Payments**

|  |  |
| --- | --- |
| **Thuộc tính** | **Mô tả** |
| PaymentID | Mã thanh toán (Khóa chính) |
| BookingId | Mã đặt phòng của khách (Khóa ngoại tham chiếu đến Bookings(BookingID) ) |
| PaymentDate | Ngày thanh toán |
| FirstDateOccupied | Ngày đến của khách |
| LastDateOccupied | Ngày rời khách sạn của khách |
| TaxRate | Thuế |
| Duration | Số ngày ở khách sạn |
| Discount | Giảm giá |

* + - 1. **Bảng Rooms**

|  |  |
| --- | --- |
| **Thuộc tính** | **Mô tả** |
| RoomNumber | Số phòng (Khóa chính) |
| RoomTypeID | Mã loại phòng (Khóa ngoại tham chiếu đến RoomTypes(RoomTypeID)) |
| BedTypeID | Mã loại giường (Khóa ngoại tham chiếu đến BedTypes(BedTypeID)) |
| Rate |  |
| Notes | Ghi chú |

* + - 1. **Bảng RoomTypes**

|  |  |
| --- | --- |
| **Thuộc tính** | **Mô tả** |
| RoomTypeID | Mã loại phòng (Khóa chính) |
| RoomType | Tên loại phòng |
| Notes | Ghi chú |
| TypeCost | Giá loại phòng |

# **CHƯƠNG 2: THIẾT KẾ XÂY DỰNG CƠ SỞ DỮ LIỆU TÁC NGHIỆP (OLAP)**

## **Xác định các Business Process và bảng Fact**

Xây dựng Bus Matrix xác định các Business Process, bảng Fact, bảng Dim cần thiết.

A screenshot of a computer

Description automatically generated

Hình 1. Detailed Bus Matrix

**Business Process: Booking Transaction**

A table with text and a list of items

Description automatically generated with medium confidence

Hình 2. Bảng FactBookings

**Business Process : Sales Management**

**A screenshot of a computer

Description automatically generated**

Hình 3. Bảng FactSales

## **Xây dựng các bảng Dimension**

**DimDate**

A screenshot of a computer

Description automatically generated

Hình 4. DimDate

**DimCustomers**

A screenshot of a computer

Description automatically generated

Hình 5. DimCustomers

**DimEmployees**

A screenshot of a computer

Description automatically generated

Hình 6. DimEmployees

* 1. **Star schema (Lược đồ hình sao)**

A screenshot of a computer

Description automatically generated

Hình 7. Lược đồ hình sao

# **CHƯƠNG 3: TÍCH HỢP DỮ LIỆU VÀO KHO (SSIS)**

* 1. **Nhập dữ liệu vào bảng Date Dimension**
     1. **Tạo package mới và thiết lập control flow**

**A screenshot of a computer

Description automatically generated**

* + 1. **Load dữ liệu từ nguồn vào bảng stgDate**

**A screenshot of a computer

Description automatically generated**

* + 1. **Thiết lập the truncate table SQL task**

**A screenshot of a computer

Description automatically generated**

* + 1. **Load dữ liệu từ bảng stgDate vào bảng DimDate**

Bước cuối cùng là định cấu hình giai đoạn cho luồng dữ liệu bảng DimDate. Trong luồng dữ liệu này, sử dụng Type 1 Slowly Changing Dimension để đảm bảo rằng không vô tình đưa cùng một khóa business key nhiều lần vào bảng dimension

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

* 1. **Cấu hình cho bảng FactBookings**
     1. **Cấu hình staging cho FactBookings**
        1. **Tạo package mới và thiết lập control flow**
* **Tạo một Execute SQL Task và 3 Data Flow Task**

**A screenshot of a computer

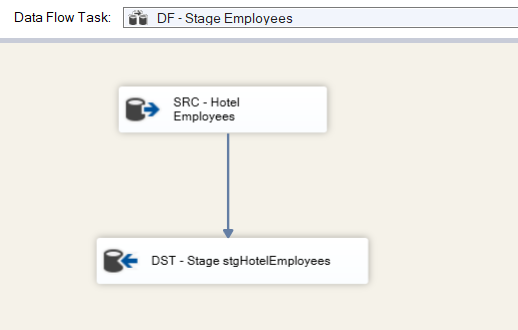
Description automatically generated**

* + - 1. **Load dữ liệu từ nguồn vào bảng stgHotelCustomers**

A screenshot of a computer

Description automatically generated

* + - 1. **Load dữ liệu từ nguồn vào bảng stgHotelEmployees**

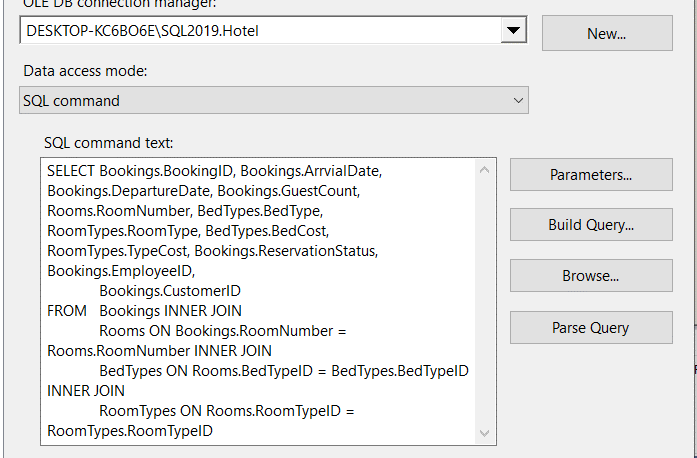
****

* + - 1. **Load dữ liệu từ nguồn vào bảng stgHotelBookings**

A screenshot of a computer

Description automatically generated

* **Dùng SQL command để kết bảng cho stgHotelBookings**

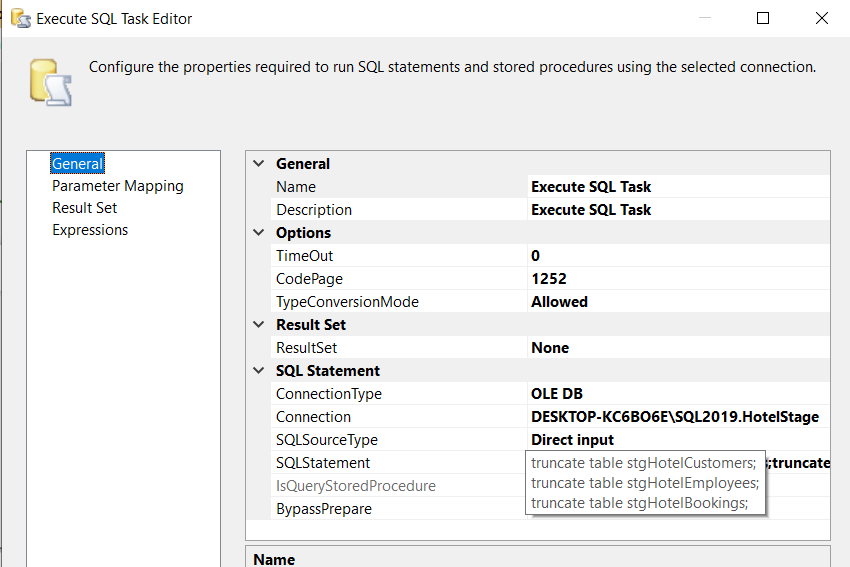


* **Kiểm tra mapping giữa source và destination**

A screenshot of a computer

Description automatically generated

* + - 1. **Thiết lập the truncate table SQL task**

****

* + 1. **Load dữ liệu vào bảng FactBookings**
       1. **Tạo package mới và thiết lập control flow**

**A screenshot of a computer

Description automatically generated**

* + - 1. **Load dữ liệu từ stgCustomers vào DimCustomers**

**A diagram of a process

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

* + - 1. **Load dữ liệu từ stgEmployees vào DimEmployees**

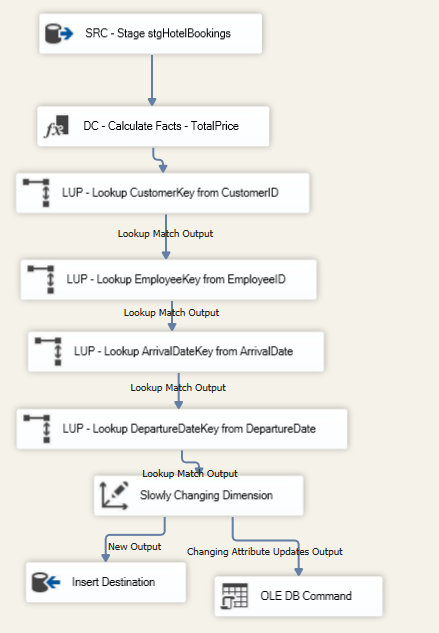
**A screenshot of a diagram

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

* + - 1. **Load dữ liệu từ stgBookings vào bảng FactBookings**

****

* **Tính toán tổng chi phí đặt phòng (TotalPrice)**

**A screenshot of a computer

Description automatically generated**

* **Tìm CustomerKey từ CustomerID**

**A screenshot of a computer

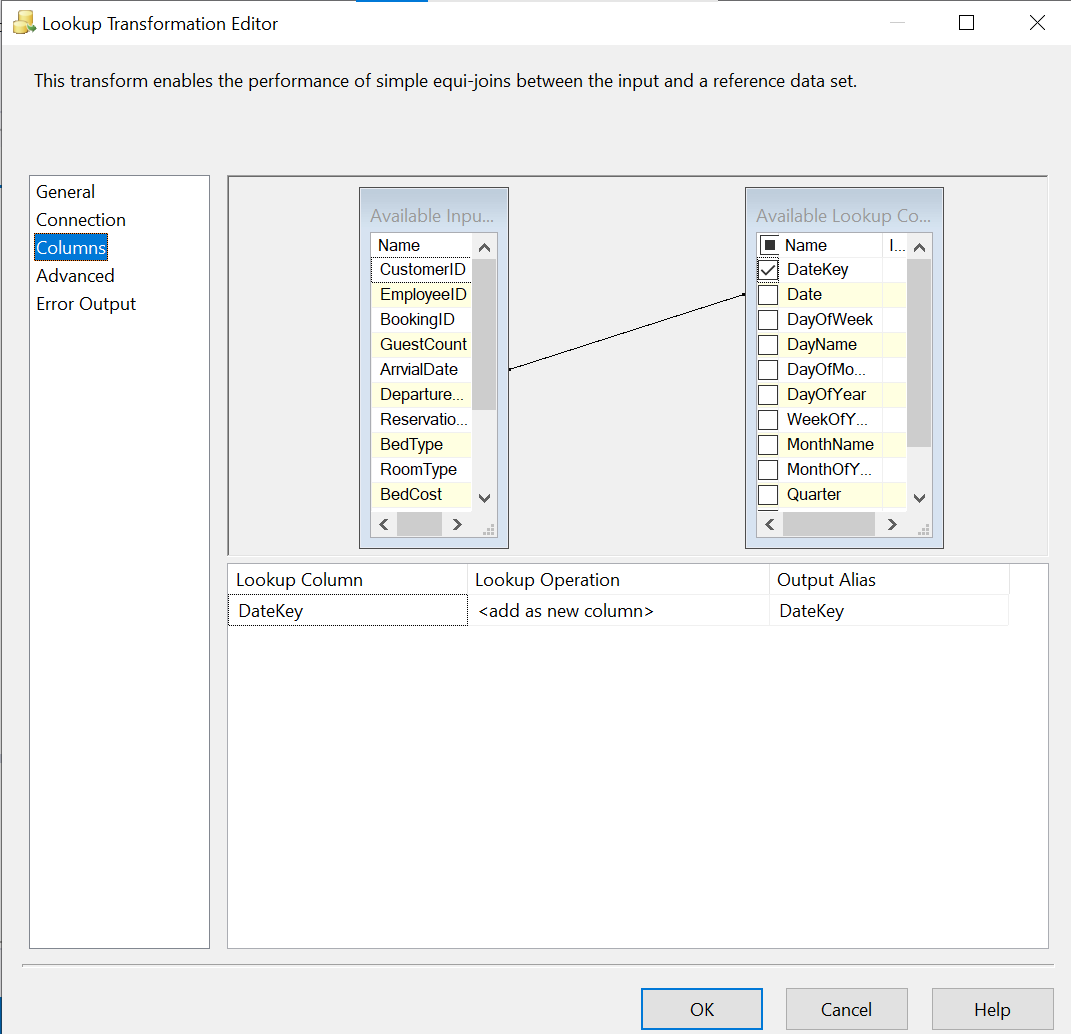
Description automatically generated**

* **Tìm EmployeeKey từ EmployeeID**

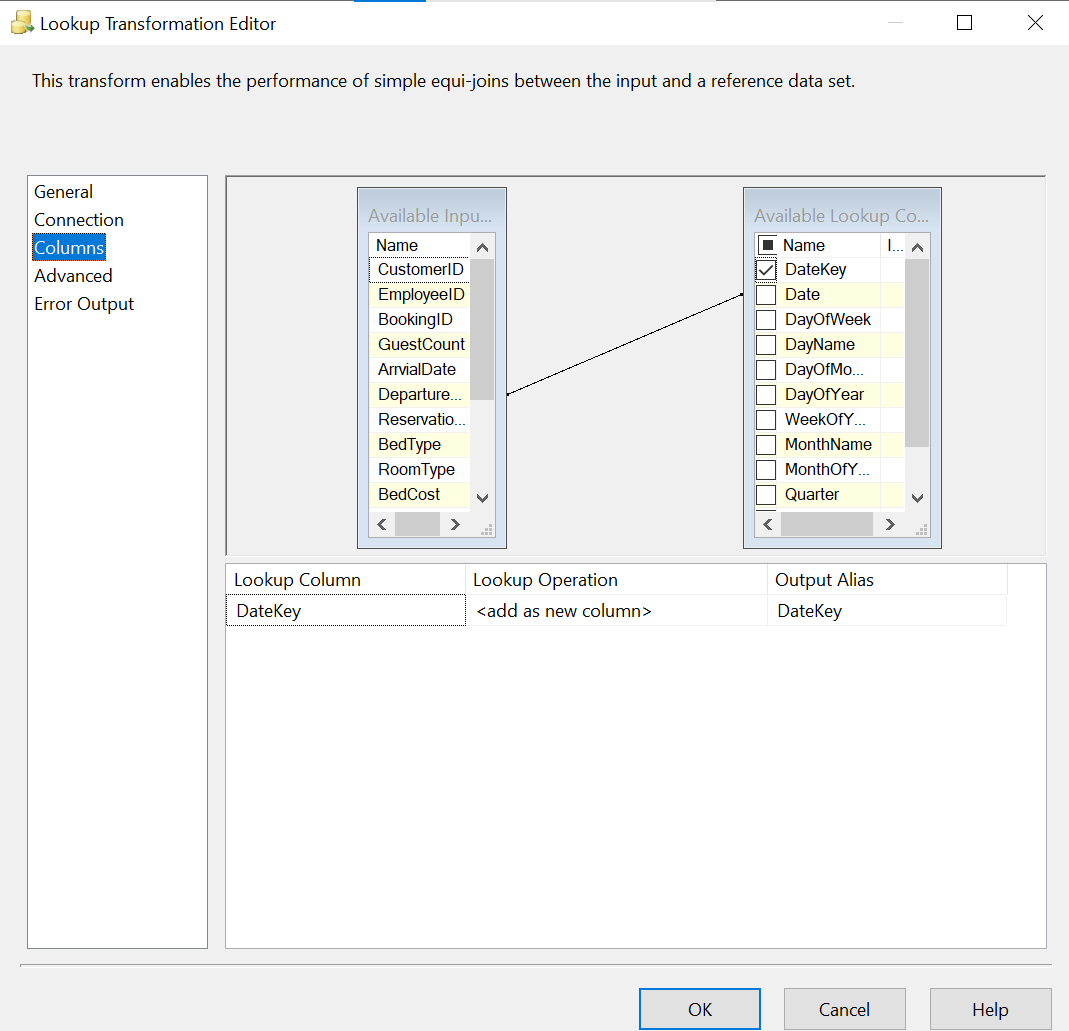
**A screenshot of a computer

Description automatically generated**

* **Tìm ArrivalDateKey từ ArrivalDate**

****

* **Tìm DepartureDateKey từ DepartureDate**

****

* **Slow Changing Dimension**

**A screenshot of a computer

Description automatically generated**

* 1. **Cấu hình cho bảng FactSales**
     1. **Cấu hình staging cho FactSales**
        1. **Tạo package mới và thiết lập control flow**
* **Tạo một Execute SQL Task và 3 Data Flow Task**

**A screenshot of a computer

Description automatically generated**

* + - 1. **Load dữ liệu từ nguồn vào bảng stgHotelCustomers**

A screenshot of a computer

Description automatically generated

* **Kiểm tra mapping giữa source và destination**

A screenshot of a computer

Description automatically generated

* + - 1. **Load dữ liệu từ nguồn vào bảng stgHotelEmployees**

A screenshot of a computer

Description automatically generated

* **Kiểm tra mapping giữa source và destination**

A screenshot of a computer

Description automatically generated

* + - 1. **Load dữ liệu từ nguồn vào bảng stgHotelSales**

**A screenshot of a computer

Description automatically generated**

* **Dùng SQL command để kết bảng cho stgHotelSales**

**A screenshot of a computer

Description automatically generated**

* **Kiểm tra mapping giữa source và destination**

A screenshot of a computer

Description automatically generated

* + - 1. **Thiết lập truncate table SQL task**

**A screenshot of a computer

Description automatically generated**

* + 1. **Load dữ liệu vào bảng FactSales**
       1. **Tạo package mới và thiết lập control flow**

**A screenshot of a computer

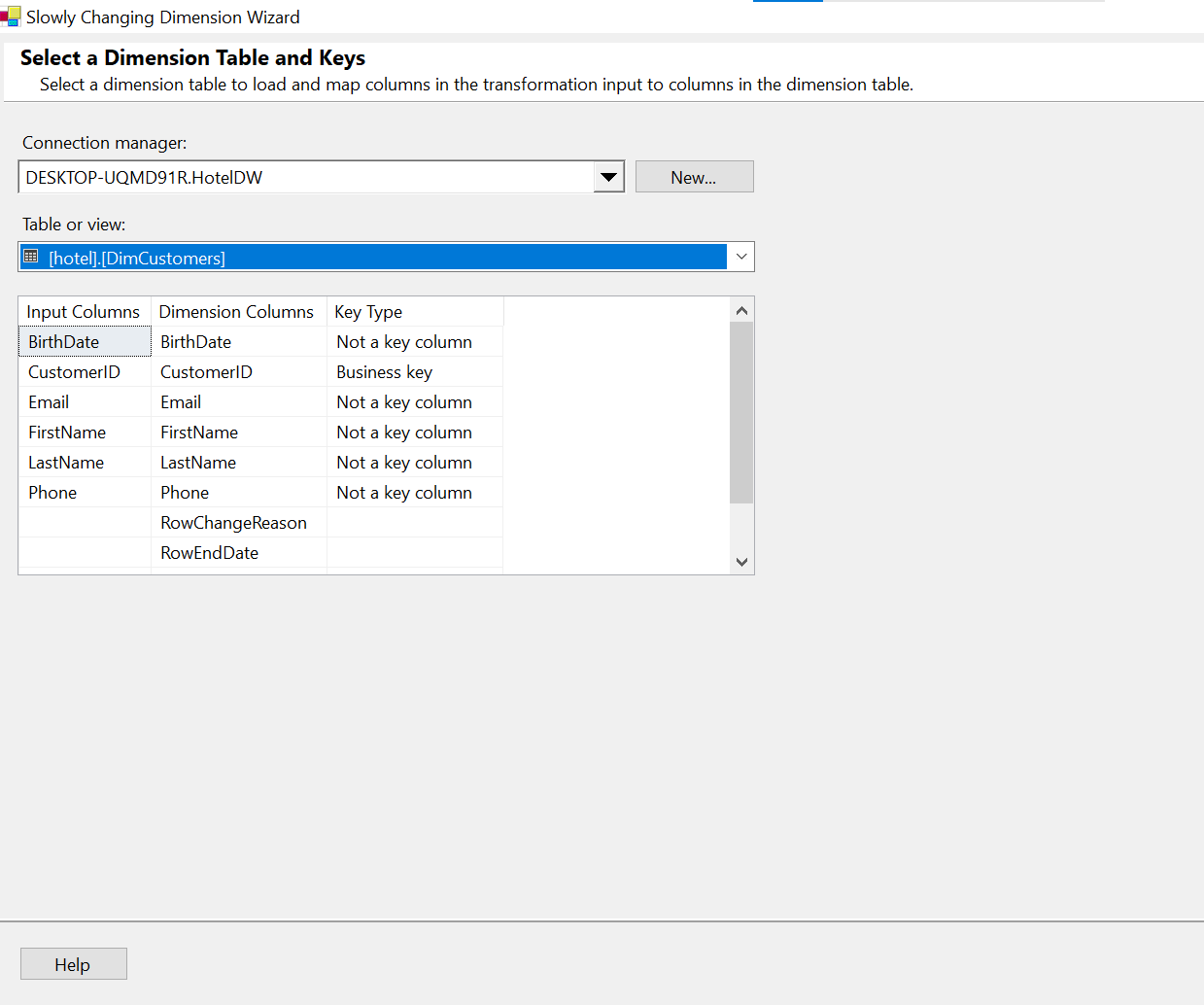
Description automatically generated**

* + - 1. **Load dữ liệu từ stgCustomers vào DimCustomers**

**A diagram of a flowchart

Description automatically generated**

* **DST - Type 2 Slowly Changing Dimension DimCustomers**

****

* + - 1. **Load dữ liệu từ stgEmployees vào DimEmployees**

**A screenshot of a diagram

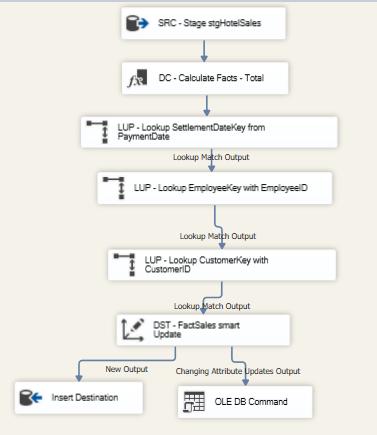
Description automatically generated**

* **DST - Type 2 Slowly Changing Dimension DimEmployees**

**A screenshot of a computer

Description automatically generated**

* + - 1. **Load dữ liệu từ stgSales vào bảng FactSales**

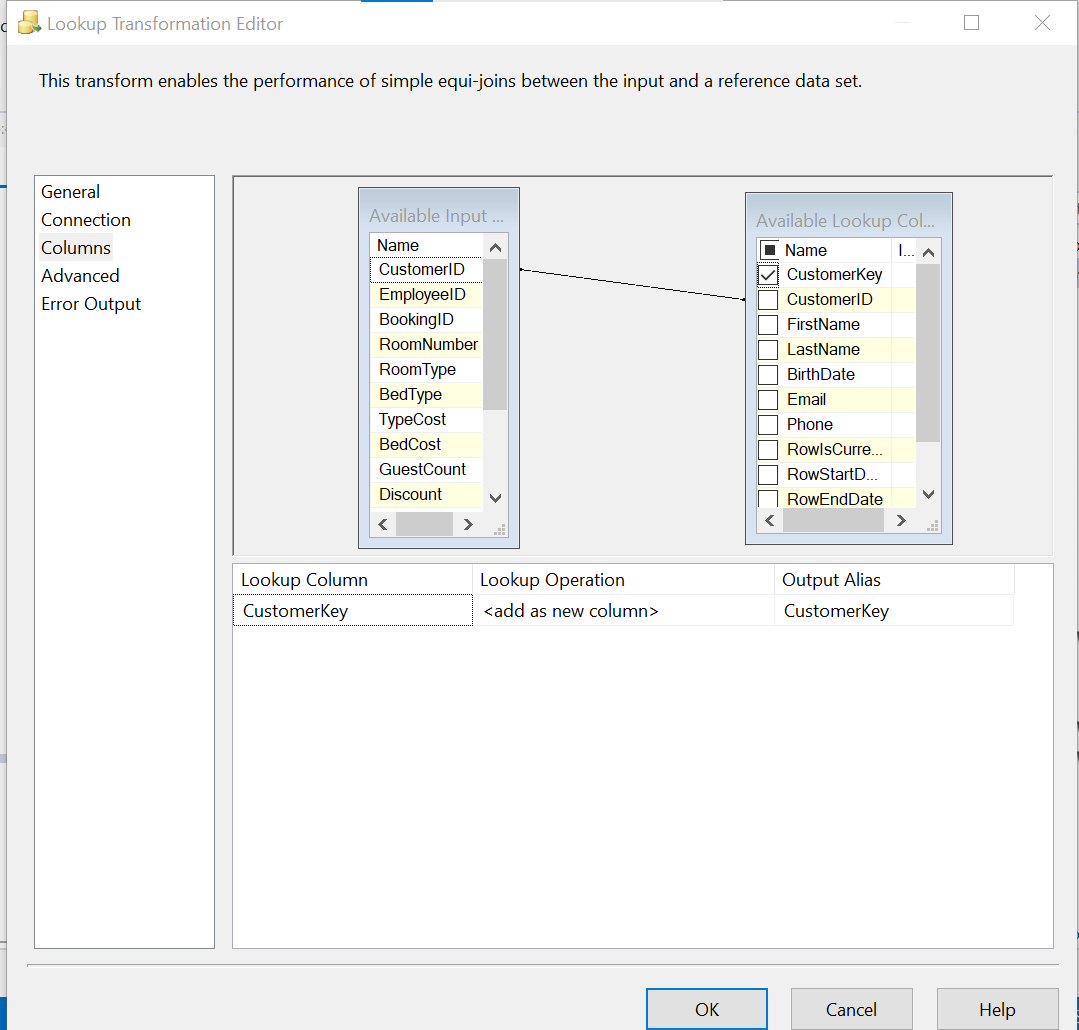
****

* **Tính toán tổng doanh thu của khách sạn (Total)**

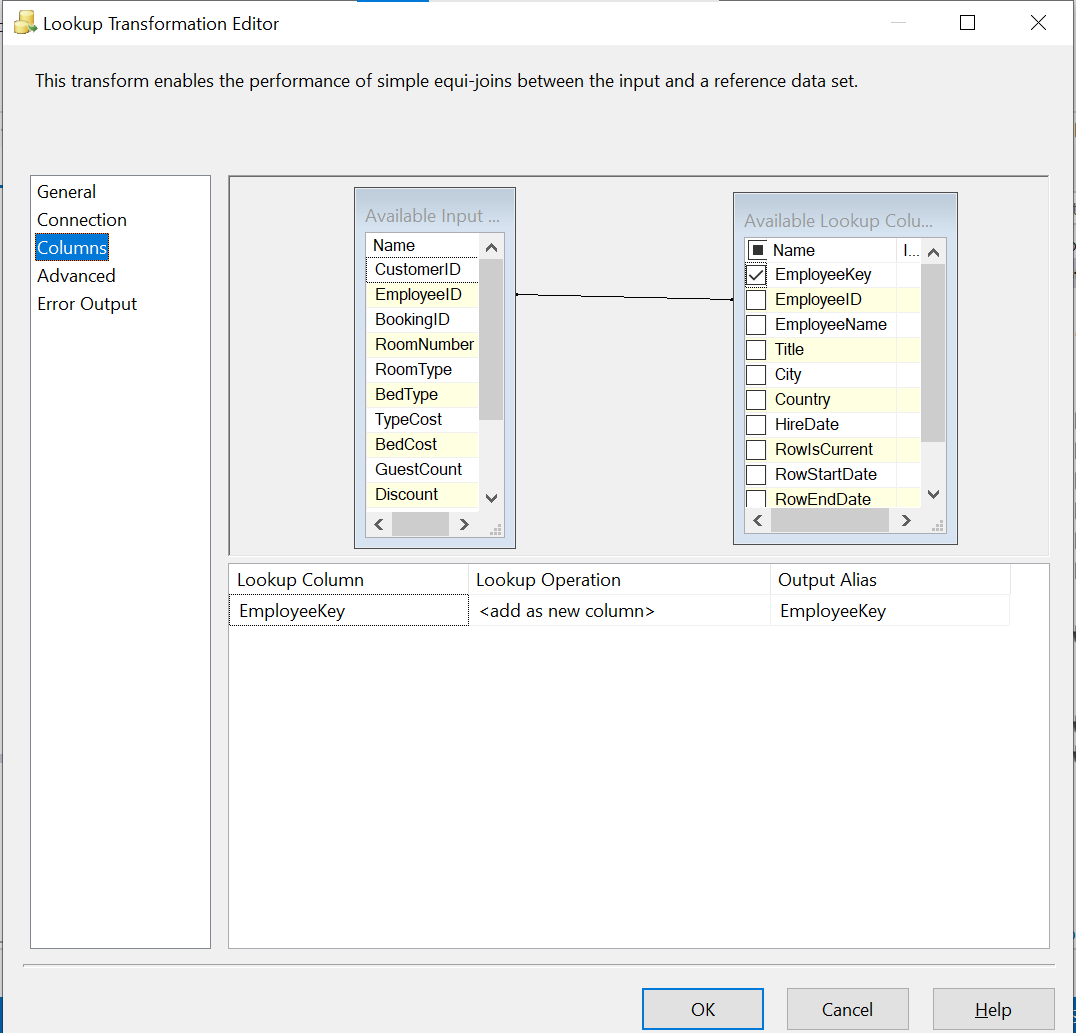
**A screenshot of a computer

Description automatically generated**

* **Tìm CustomerKey từ CustomerID**

****

* **Tìm EmployeeKey từ EmployeeID**

****

* **Tìm SettlementDateKey từ PaymentDate**

**A screenshot of a computer

Description automatically generated**

# **CHƯƠNG 4: PHÂN TÍCH DỮ LIỆU (SSAS)**

* 1. **Xây dựng mô hình**
     1. **Tạo Data Source**

Right Click vào Data Source chọn New Data Source

A screenshot of a computer

Description automatically generated

Một màn hình thông báo hiện lên, chọn Next để thực hiện các bước tiếp theo

A screenshot of a computer

Description automatically generated

Tạo connection với data source

A screenshot of a computer

Description automatically generated

Nếu cần tạo một conection mới thì chọn New để tạo connection với các tùy chọn thích hợp để kết nối vào SQL Server

A screenshot of a computer

Description automatically generated

Tiếp theo khi tạo xong Connection chọn Next. Tại đây tiến hành đặt tên cho Data Source sau đó chọn Finish để kết thúc cấu hình và tạo một Data Source mới

A screenshot of a computer program

Description automatically generated

**Các thông số định nghĩa của Connection**

**Tên của Data source**

* + 1. **Tạo Data SourceView**

Right click Data Source Views chọn New Data Source View

A screenshot of a computer

Description automatically generated

Một màn hình thông báo hiện lên, chọn Next

A screenshot of a computer

Description automatically generated

Tiếp theo chọn một Data Source để tạo View, Data source ở trên vừa tạo sẽ được lựa chọn, nếu chưa có chọn New để thực hiện tiến trình tạo một Data source mới.

Sau đó chọn Next.

A screenshot of a computer

Description automatically generated

Có thể sẽ yêu cầu nhập Password khi tạo kết nối nếu cơ sở dữ liệu của bạn có tạo mật khẩu bảo mật

A screenshot of a computer

Description automatically generated

Tiếp theo chọn các Table cần thiết để tạo View. Trong mục **Included objects**, chọn các bảng Dim và Fact cần thiết

A screenshot of a computer

Description automatically generated

**Vùng hiển thị các bảng đã được chọn để tạo View**

**Các bảng mà đã kết nối được và có thể chọn để tạo view**

Tiếp theo tiến hành đặt tên DataView, sau đó chọn Finish để kết thúc tiến trình

A screenshot of a computer

Description automatically generated

Double-Click vào Data Source View vừa tạo để hiển thị Diagram

A screenshot of a computer

Description automatically generated

* 1. **Tạo Data cube**

Right-click vào Cubes ở khung Solution Explorer. Chon New Cube để tạo Cube mới.

Chọn Data Source View là Hotel DW View . Tích chọn 2 bảng FactSales và FactBooking, nhấn Next

A screenshot of a computer

Description automatically generated

Ở dialog Select Measures, tích chọn những Measure cần thiết

A screenshot of a computer

Description automatically generated

Khởi tạo các Dimension

A screenshot of a computer program

Description automatically generated

Cuối cùng đặt tên Data cube để kết thúc tiến trình

A screenshot of a computer

Description automatically generated

Các bảng Fact, Dimension và quan hệ giữa chúng sẽ được thể hiện trong Diagram

A screenshot of a computer

Description automatically generated

* 1. **Cấu hình Hierarchy**
     1. **Tạo Hierarchy Dim Date**

Thêm thuộc tính vào bảng Dim: Tại table DimDate trong panel Data Source View, kéo thuộc tính cần thiết vào panel Attributes.

Kéo thả các thuộc tính Date, Month và Year từ Attributes vào panel Hierarchies để lần lượt tạo ra các Dtae, Month và Year Hierarchy.

A screenshot of a computer

Description automatically generated

Quan sát thấy trong Year Hierarchy, dữ liệu được sắp xếp theo thứ tự: Năm tháng trong năm số tuần trong tháng giống với Year Hierarchy đã cấu hình.

A person holding a phone

Description automatically generated with medium confidence

* + 1. **Tạo Hierarchy cho Dim Employees**

**A screenshot of a computer

Description automatically generated**

Vào Browser kiểm tra thấy các thông tin đã được sắp xếp theo thứ tự, Tên nhân viên Mã nhân viên Quốc gia Thành phố

**A screenshot of a computer

Description automatically generated**

* + 1. **Tạo Hierarchy cho Dim Customers**

**A screenshot of a computer

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

Vào Browser kiểm tra thấy các thông tin đã được sắp xếp theo thứ tự, ID Customer Name Customer Country City như đã cấu hình trong Hierarchy Customers

* 1. **Thực hiện tạo Cube**

Kết thúc các quá trình cấu hình và nhấn Process để tiến hành tạo Cube với các cấu hình. Và vào Browser để thực hiện các phân tích dữ liệu.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

* 1. **Thực hiện phân tích dữ liệu**
     1. **Sử dụng SSAS**

***Câu hỏi: Cho biết tổng Doanh thu tháng 5, 6, 7 năm 2025 tại khách sạn***

A screenshot of a computer

Description automatically generated

Khi dùng Cube, tất cả doanh thu trong điều kiện đã chọn là tháng 5, 6, 7 và năm 2015 được lọc ra và tổng lại thành 1 giá trị duy nhất, giá tổng sẽ bao gồm giá phòng + loại giường \* phí thuế \* discount để ra total cuối cùng

Từ doanh thu có thể

***Câu hỏi: Cho biết doanh số bánh hàng theo từng khách hàng trong tháng 10 năm 2017***

A screenshot of a computer

Description automatically generated

Các điều kiện được cấu hình từ các bảng Dim, Hierachy:

* Year = 2017
* Month = October
* Gôm nhóm theo tên của từng khách hàng

A screenshot of a computer

Description automatically generated

Từ truy vấn, có thể thấy được kết quả trả về theo cấu trúc đã yêu cầu, bao gồm Customer ID, First Name, Last Name và Total Sales

Từ truy vấn trên chũng ta có thể tương tự và xác định được nhiều vấn đề cần thiết như:

* So sánh doanh số tháng 10 với các tháng trước có thể cho thấy xu hướng tăng hay giảm doanh số. Nếu doanh số của tháng 10 cao hơn các tháng trước, điều này có thể cho thấy hiệu quả của các chiến dịch marketing hoặc sự gia tăng nhu cầu của thị trường
* Đề xuất các chương trình ưu đãi đặc biệt hoặc dịch vụ khách hàng cao cấp cho những khách hàng có doanh số cao,…
  + 1. **Sử dụng Pivot Table trong Excel và Power BI**
       1. ***Câu hỏi: Cho biết xếp hạng các nhân viên nhận booking khách hàng theo quý/ tháng***
* **Sử dụng Pivot Table trong Excel**

Hiển thị theo thời gian quý 3

Ảnh có chứa văn bản, ảnh chụp màn hình, số, Song song

Mô tả được tạo tự động

Ảnh có chứa văn bản, ảnh chụp màn hình, số, màn hình

Mô tả được tạo tự động

Với hình trên thì ta cũng có thể thấy được số liệu nhận booking của các nhân viên theo quý 3 và theo các tháng 7, 8 và 9. Nhìn vào dây có thể thấy rằng doanh thu tháng 9 vào năm 2015 còn khá ít so với tháng 7 và tháng 8.

* **Sử dụng Power BI**

Biểu đồ xếp hạng các nhân viên nhận booking cao nhất theo quý/ tháng

Ảnh có chứa văn bản, ảnh chụp màn hình, biểu đồ, Phông chữ

Mô tả được tạo tự động

Tổng số Khách của 10 Nhân viên dao động từ 8 đến 26.

Ở mức 26, Xavier Neil có Tổng số Khách cao nhất và cao hơn 225.00% so với Yara Flint, người có Tổng số Khách thấp nhất là 8. Xavier Neil chiếm 15.66% tổng số Khách.

* + - 1. ***Câu hỏi: Cho biết doanh thu trên mỗi loại phòng được đặt nhiều nhất theo quý/ tháng***
* **Sử dụng Pivot Table trong Excel**

Ảnh có chứa văn bản, ảnh chụp màn hình, số, màn hình

Mô tả được tạo tự động

Hình trên hiển thị doanh thu của mỗi loại phòng vào tháng 7, 8 và 9. Ta thấy phòng Standard room được khách hàng ưa chuộng và đặt nhiều nhất. Phòng Penthouse được đặt ít nhất với tổng doanh thu là 303.5771. Tổng doanh thu của tất cả loại phòng theo tháng 7, 8 và 9 khá đều nhau. Điều này cho thấy rằng khách sạn có một lượng khách ổn định trong các tháng mùa hè (7, 8, 9). Sự đều đặn trong doanh thu có thể phản ánh hiệu quả của các chiến lược marketing, chất lượng dịch vụ ổn định hoặc không có biến động lớn về nhu cầu du lịch trong giai đoạn này.

* **Sử dụng Power BI**

Xếp hạng loại phòng được đặt nhiều theo tháng.Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ, số

Mô tả được tạo tự động

Ở mức $3,611.7071, phòng Standard có tổng số cao nhất và cao hơn 1,089.72% so với Penthouse, phòng có Tổng số thấp nhất là $303.5771. Loại phòng theo tháng tương đối bằng nhau.

Ta cũng có thể chọn theo quý 4 tháng 12 và chọn loại phòng Bungalow để xem chi tiết về doanh thu theo tháng.

Ảnh có chứa văn bản, ảnh chụp màn hình, phần mềm, số

Mô tả được tạo tự động

* + - 1. ***Câu hỏi: Tỷ lệ trạng thái phòng mà khách hàng đặt tổng chi phí bao nhiêu***

Đã rời đi (Checked-out): Khách hàng đã rời khỏi khách sạn sau khi hoàn thành thời gian lưu trú.

Đã hủy (Cancelled): Đặt phòng đã bị hủy bởi khách hàng hoặc bởi hệ thống do không hoàn thành các yêu cầu cần thiết.

No-show : Trạng thái đặt phòng mà khách hàng đã không xuất hiện tại khách sạn

* **Sử dụng Pivot Table trong Excel**

Ảnh có chứa văn bản, ảnh chụp màn hình, phần mềm, số

Mô tả được tạo tự độngBảng dữ liệu cho thấy chi phí cho ngày trong tuần (Weekday) khách hàng đã Check-out, Canceled và No-show lớn hơn ngày cuối tuần (Weekend). Tổng chi phí khách khàng đã Check-out lớn hơn khách hàng Canceled (hủy) và No-show (không đến).

* **Sử dụng Power BI**

Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ, biểu đồ

Mô tả được tạo tự động

Nhìn hình ta thấy, Check-Out có Tổng Số Tiền cao nhất là $78,313.06, tiếp theo là Canceled với $18,651.52 và No-Show với $612.46. Check-Out chiếm 80.26% Tổng Số Tiền.

Lượng khách thường đến khách sạn nhiều nhất vào ngày nào

* **Sử dụng Pivot Table trong Excel**

Ảnh có chứa văn bản, ảnh chụp màn hình, số, Phông chữ

Mô tả được tạo tự động

Dựa trên thông tin hình ảnh:

**Tổng số khách** trong năm 2015 là **983** người.

Trong đó, số khách vào **ngày thường** là **740** người, chiếm khoảng **75%** tổng số khách.

Số khách vào **cuối tuần** là **243** người, chiếm khoảng **25%** tổng số khách.

**Thứ Sáu** và **Thứ Hai** có lượng khách đến nhiều nhất trong tuần, mỗi ngày là **158** và **153** người tương ứng.

**Thứ Bảy** và **Chủ Nhật** có lượng khách ít hơn, với **124** và **119** người tương ứng.

* **Sử dụng Power BI**

Ảnh có chứa văn bản, biểu đồ, hàng, Sơ đồ

Mô tả được tạo tự động

Vào ngày thứ Sáu, tổng số khách cao nhất trong tuần và cao hơn 37,93% so với ngày thứ Ba, ngày có tổng số khách thấp nhất là 116. Trong suốt 5 ngày, tổng số khách dao động từ 116 đến 160.

# **CHƯƠNG 5. KẾT LUẬN**

* 1. **Kết quả đạt được**

Việc xây dựng 'KHO DỮ LIỆU VỀ KHÁCH SẠN' đã giúp chúng em nâng cao hiểu biết thêm về kiến thức môn học Kho Dữ liệu. Qua quá trình thực hiện, chúng em đã trở nên thành thạo hơn trong việc sử dụng các công cụ hỗ trợ quá trình ETL để đưa dữ liệu vào kho dữ liệu, sư dụng thông thạo trong các công cụ SSIS và SSAS, cũng như các công cụ trực quan hóa và phân thích dữ liệu như Excel Pivot trong Exel và Power BI Desktop.

Dự án này còn giúp chúng em cải thiện kỹ năng phân tích và xử lý thông tin đầu vào, đặt câu hỏi tiền đề, và thảo luận để tìm ra hướng giải quyết và trả lời các câu hỏi.

Những kinh nghiệm thu được từ dự án này là nền tảng vững chắc để chúng em áp dụng vào các dự án thực tế trong tương lai.

* 1. **Những hạn chế**

Trong quá trình thực hiện xây dựng kho dữ liệu, nhóm chúng em đã gặp phải một số khó khăn do hạn chế về mặt khai thác dữ liệu. Tập dữ liệu còn nhiều thông tin chưa được khai thác hết, dẫn đến việc chưa tận dụng được tối đa các dữ liệu sẵn có. Quá trình ETL gặp một số lỗi phải mất nhiều thời gian để giải quyết.

Bên cạnh đó, những bước thực hiện của chúng em chưa đạt đến mức độ nâng cao trong toàn bộ quá trình xây dựng một kho dữ liệu hoàn chỉnh. Điều này ảnh hưởng đến khả năng tạo ra các phân tích và báo cáo chi tiết, chính xác, phục vụ cho việc ra quyết định quản lý và kinh doanh.

Chúng em nhận thấy cần phải tiếp tục học hỏi và nâng cao kiến thức để có thể hoàn thiện hơn trong các dự án tương lai.