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

**KHOA CÔNG NGHỆ THÔNG TIN**

🙢 🕮 🙠



**BÁO CÁO KHO DỮ LIỆU**

**ĐỀ TÀI: XÂY DỰNG DATA WAREHOUSE TRÊN CSDL CỦA TY BÁN PHẦN CỨNG MÁY TÍNH**

**GV hướng dẫn: Th.S Quách Đình Hoàng**

**Sinh viên thực hiện:**

|  |  |
| --- | --- |
| Ngô Thành Danh | 17133005 |
| Lê Hữu Khuê | 17133034 |
| Phạm Trường Phát | 17133047 |
| Trần Hùng Quân | 17133051 |
| Huỳnh Thảo Tất | 17133058 |

**TP.HCM - Tháng 06/2020**

**MỤC LỤC**

[**MỞ ĐẦU** 1](#_Toc44453191)

[**1.** **Lời nói đầu** 1](#_Toc44453192)

[**2.** **Mục đích nghiên cứu** 1](#_Toc44453193)

[**3.** **Những nội dụng chính** 1](#_Toc44453194)

[**NỘI DUNG** 1](#_Toc44453195)

[**1. Dữ liệu** 1](#_Toc44453196)

[**1.1. Giới thiệu cơ sở dữ liệu** 1](#_Toc44453197)

[**1.2. Lý do cần có một kho dữ liệu** 1](#_Toc44453198)

[**2. Business Process** 2](#_Toc44453199)

[**2.1. Xác Định Business Process** 2](#_Toc44453200)

[2.2. **Thiết kế Star Schema** 4](#_Toc44453201)

[**2.3. Bus Matrix** 4](#_Toc44453202)

[**2.4. Các Dimension Table** 3](#_Toc44453203)

[**2.5. Các bảng Fact** 4](#_Toc44453204)

[**3.** **Quy trình thực hiện và kỹ thuật** 5](#_Toc44453205)

[**3.1.** **Tích hợp dữ liệu** 5](#_Toc44453206)

[**3.2.** **Thiết kế Data Cube** 10](#_Toc44453207)

[**4.** **Tạo báo cáo** 12](#_Toc44453208)

[**KẾT LUẬN** 19](#_Toc44453209)

# **MỞ ĐẦU**

1. **Lời nói đầu**

Trong thời đại công nghệ thông tin phát triển và trở thành một phần không thể thiếu trong cuộc sống , hoạt động kinh doanh. Nhu cầu của con người ngày càng nâng cao, các hoạt động kinh doanh sẽ cạnh tranh một cách khốc liệt. Dữ liệu do hoạt động hằng ngày của con người tạo ra ngày càng lớn. Việc ra một quyết định trong chiến lược kinh doanh là vô cùng quan trọng đòi hỏi phải có sự chính xác cao và nhanh chóng. Chính vì thế doanh nghiệp phải có cách quản lý dữ liệu cụ thể và có tính hiệu quả cao trong hoạt động kinh doanh. Xây dựng kho dữ liệu cho công ty là một biện pháp hiệu quả, kho dữ liệu hỗ trợ việc lưu trữ hiệu quả, phân tích dữ liệu và lập báo cáo hỗ trợ việc ra quyết đinh kinh doanh. Trong báo cào này chúng em thực hiện xây dựng kho dữu liệu trên tập dữ liệu của công ty bán phần cứng máy tính trên hệ quản trị cơ sở dữ liệu SQL Server lập báo cáo cho các Business Process được nhóm chúng em chọn trên tập dữ liệu.

1. **Mục đích nghiên cứu**

Thông qua việc nghiên cứu đề tài này nhóm chúng em muốn xây dựng một kho dữ liệu cho công ty bán phần cứng máy tính giúp lưu trữ dữ liệu cách hiệu quả, tạo ra các báo cáo cho hoạt động kinh doanh. Tìm hiểu thêm kiến thức về kho dữ liệu

1. **Những nội dụng chính**

Đề tài này nhóm chúng em trình bày những nội dung chính sau đây: giới thiệu về cơ sở dữ liệu; mô tả start schema; lý do phải cần có một kho dữ liệu; các Business Process trả lời các câu hỏi nào; quá trình thiết kế kho dữ liệu, tạo truy vấn và báo cáo

**NỘI DUNG**

## **1. Dữ liệu**

### **1.1. Giới thiệu cơ sở dữ liệu**

Cơ sở dữ liệu mẫu có tên OT dựa trên một công ty hư cấu toàn cầu chuyên bán phần cứng máy tính bao gồm lưu trữ, bo mạch chủ, RAM, Card màn hình và CPU.

Công ty duy trì thông tin sản phẩm như tên, chi phí tiêu chuẩn mô tả, giá niêm yết và dòng sản phẩm. Nó cũng theo dõi thông tin hàng tồn kho cho tất cả các sản phẩm bao gồm cả kho nơi sản phẩm có sẵn. Vì công ty hoạt động trên toàn cầu, nên có nhiều kho ở nhiều địa điểm khác nhau trên thế giới.

Công ty ghi lại tất cả thông tin của khách hàng bao gồm tên, địa chỉ và trang web. Mỗi khách hàng có ít nhất một người liên hệ với thông tin chi tiết bao gồm tên, email và điện thoại. Công ty cũng đặt giới hạn tín dụng cho mỗi khách hàng để giới hạn số tiền mà khách hàng có thể nợ.

Bất cứ khi nào khách hàng đưa ra một đơn đặt hàng, một đơn đặt hàng được tạo ra trong cơ sở dữ liệu với trạng thái chờ xử lý. Khi công ty vận chuyển đơn đặt hàng, trạng thái đơn hàng sẽ được chuyển đi. Trong trường hợp khách hàng hủy đơn hàng, trạng thái đơn hàng sẽ bị hủy.

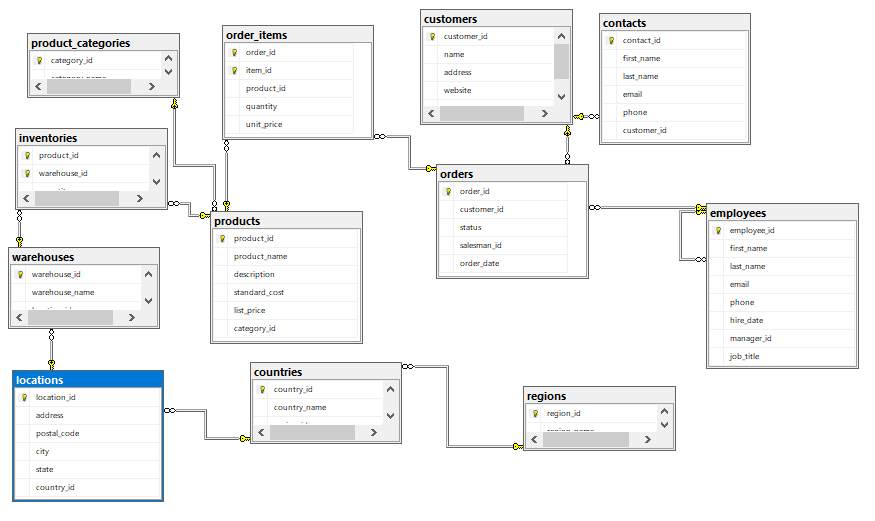
Ngoài thông tin bán hàng, dữ liệu nhân viên được ghi lại với một số thông tin cơ bản như tên, email, điện thoại, chức danh công việc, người quản lý và ngày thuê.

## **1.2. Lý do cần có một kho dữ liệu**

Trong thời đại hiện nay, công nghệ thông tin phát triển dữ liệu tạo do hoạt động hàng ngày của con người tạo ra ngày càng lớn. Nhu cầu lưu trữ dữ liệu hiệu quả là thật sự cần thiết, việc lưu trữ tốt sẽ giúp chúng ta phân tích dữ liệu nhanh và đưa ra các quyết định kinh doanh trong tương lại một cách chính xác. Ngoài ra, kho dữ liệu sẽ giúp quản lý dữ liệu rất tốt (tồn kho, doanh thu, quản lý sản phẩm,..) đối với các doanh nghiệp, công ty và tổ chức.

## **2. Business Process**

### **2.1. Xác Định Business Process**



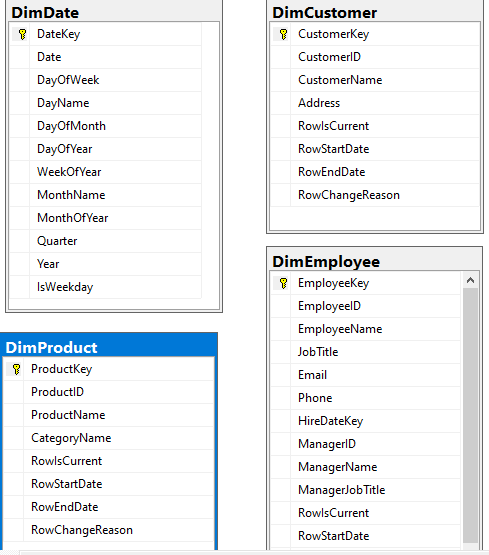
Dựa vào mô tả cở sở dữ liệu cũng như dữ liệu đã cài đặt, nhóm chúng em đã chọn ra ba câu hỏi thú vị để thực hiện kho dữ liệu này. Ba câu hỏi như sau:

* Báo cáo doanh số bán hàng (Sale Report): Người quản lý muốn theo dõi doanh số bán hàng theo khách hàng (customer), nhân viên (employee), sản phẩm (product).
* Phân tích hàng tồn kho (Inventory): Người quản lý muốn kiểm tra xem, tại một thời điểm đó, ở kho đó, tại vị trí đó, thì số lượng sản phẩm (product) còn lại trong kho là bao nhiêu?
* Đánh giá sản phẩm (FeedBack): Có rất nhiều yếu tố để xét xem một sản phẩm là tốt hay không, ở đây, nhóm chúng em chọn trạng thái của của sản phẩm đó trong tập hợp các đơn hàng. Sản phẩm có số lần hủy càng cao thì đánh giá càng thấp và ngược lại.

### **2.2. Các Dimension Table**

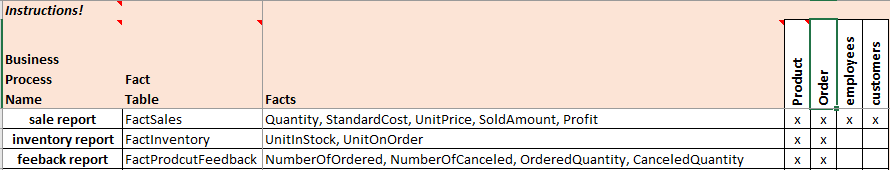
Các bảng Dimension phục vụ trong báo cáo được định nghĩa như sau:

* DimProduct chứa thông tin của sản phẩm gồm ProductID, ProductName, CategoryName
* DimEmployee chứa thông tin của nhân viên gồm EmployeeID, EmployeeName, JobTitle, Email, Phone.
* DimCustomer chứa thông tin khách hàng CustomerID, CustomerName, Address.
* DimDate là bảng chứa các khóa theo ngày, tháng, quý, năm khi thực hiện phân tích dữ liệu sẽ nhanh chóng và thuận tiện.



### **2.3. Bus Matrix**

Để xác định Dimension nào được sử dụng bởi bảng Fact nào, công cụ thường được dùng Bus Matrix. Bus Matrix tương ứng cho các Business Process được chọn bên trên được thể hiện như hình dưới:



### **2.4. Các bảng Fact**

Đối với những Business Process đã được xác định BestSeller, Billing và Inventory. Các bảng Fact phục vụ Business Process được định nghĩa như sau:

FactSales là fact cho business process Sale Report

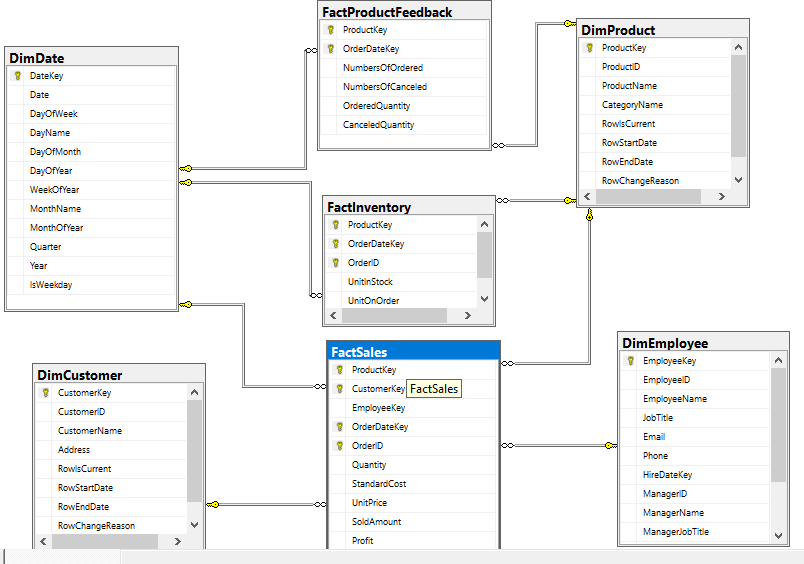
FactInventory là fact cho business process Inventory

FactProductFeedBack là fact cho business process FeedBack

### **2.5**. **Thiết kế Star Schema**

Sau khi chúng ta đã xác định được các Business Process, công việc tiếp theo là xây dựng mô hình kho dữ liệu cho phù hợp, ở đây chọn xây dựng kho bằng mô hình sao Star Schema.

Nhắc lại một chút về kiến thức mô hình Star Schema, mỗi Business process sẽ ứng với một bảng trung tâm, gọi là bảng Fact. Bảng trung tâm sẽ kết nối với những bảng thông tin mở rộng khác để làm rõ thông tin trong nó, những bảng xung quanh nó được gọi là các bảng Dimension. Các bảng Fact có thể chia sẻ các Dimension với nhau, hay có thể nói một Dimension có thể được dùng bởi nhiều bảng Fact.



## **Quy trình thực hiện và kỹ thuật**

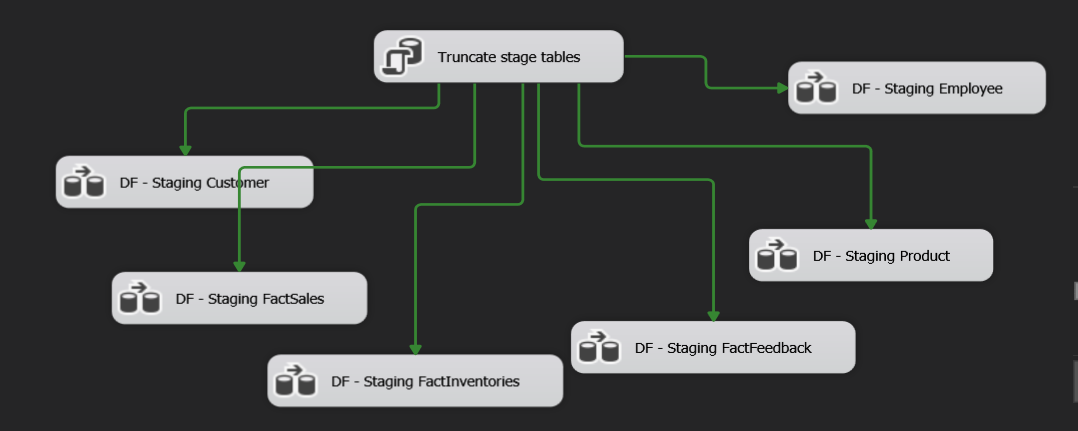
* 1. **Tích hợp dữ liệu**

Để thực hiện việc tích hợp dữ liệu, nhóm chúng em dùng công cụ SSIS của Microsoft nhằm đơn giản hóa các thao tác và hạn chế sao sót. So với sử dụng câu truy vấn SQL để tích hợp và làm sạch dữ liệu thì công cụ SSIS dễ dang sử dụng và cho hiệu qua cao. Trước tích hợp cơ sở dữ liệu vào kho dữ liệu thì chúng ta sẽ qua bước làm sạch dữ liệu (staging).

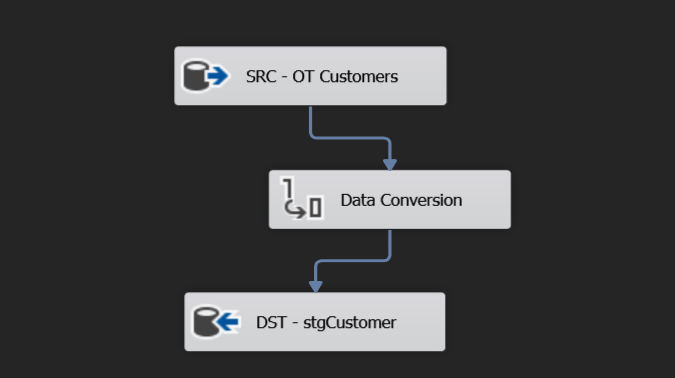
Để thực hiện tích hợp dữ liệu ta sẽ tiến hành các bước sau đây:

Bước 1: Staging cơ sở dữ liệu gốc (OT) sang cơ sở dữ liệu Stage (OT\_Staging):

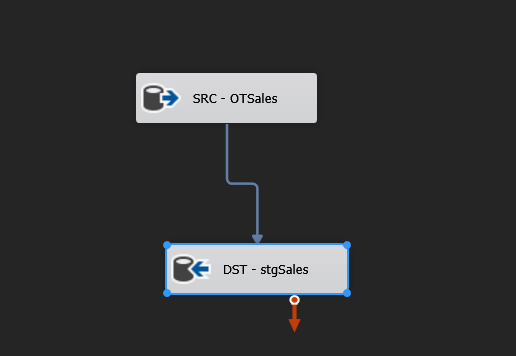
Tạo các Data Flow cho từng Dimension như hình: 3.1.



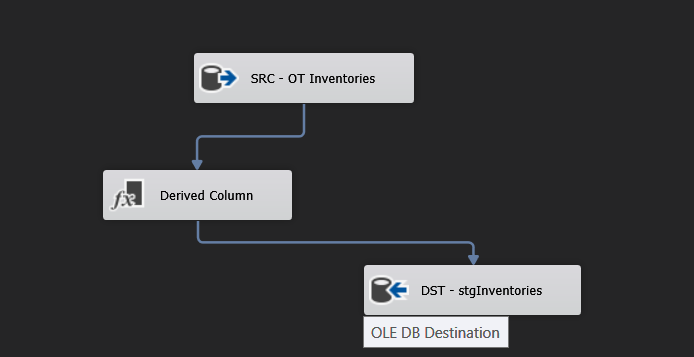
Hình 3.2 Tạo các Data Flow



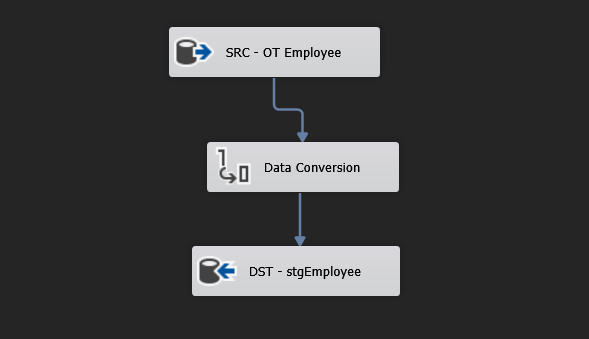
Hình 3.3 Trong DF-Staging Customer



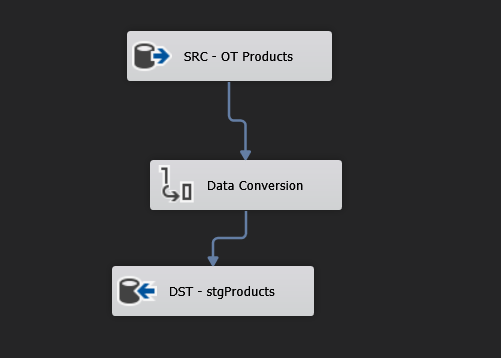
Hình 3.4 Trong DF Staging FactSales



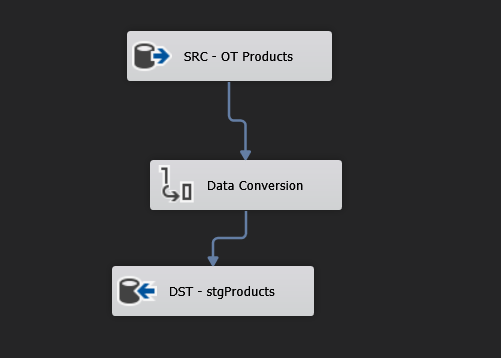
Hình 3.5 Trong DF-Staging FactInventories



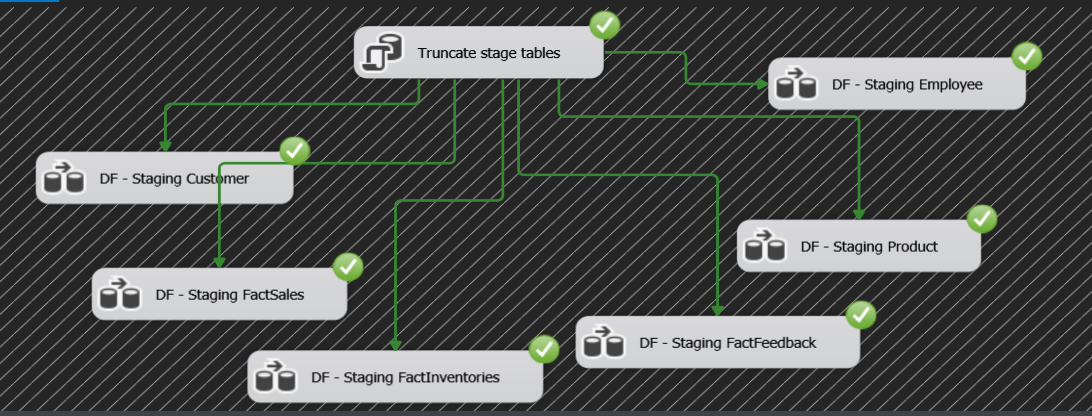
Hình 3.6 Trong DF – Staging Employee



Hình 3.7 Trong DF – Staging Product



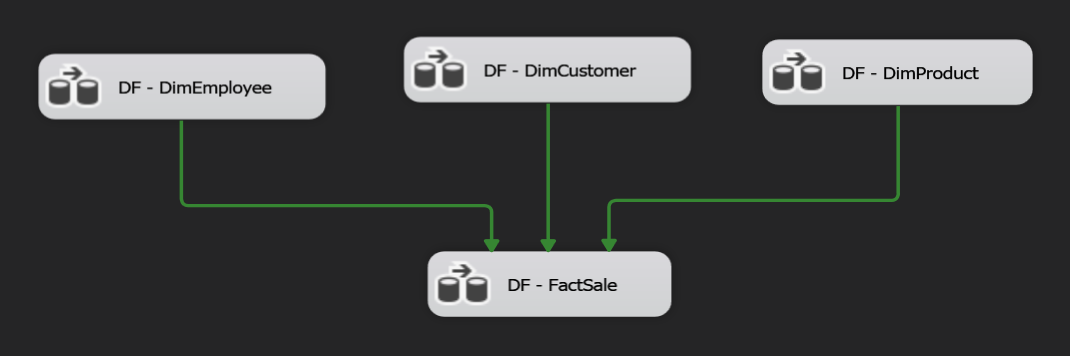
Hình 3.8 Trong DF – Staging Feedback



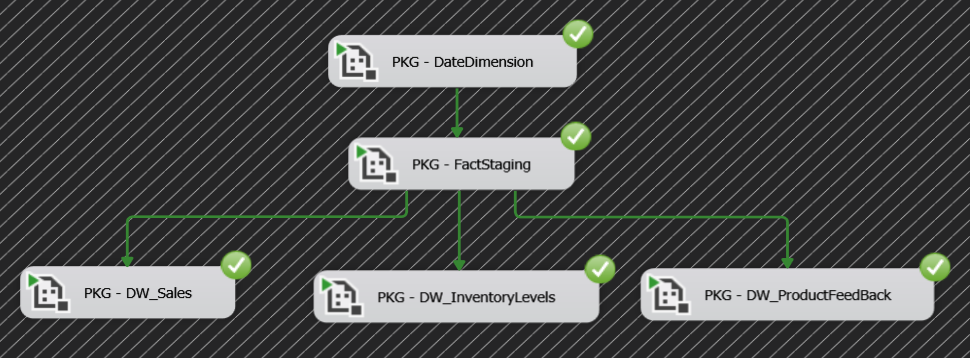
Hình 3.9 Sau khi thực thi package Staging dữ liệu

Bước 2: Load dữ liệu từ cơ sở dữ liệu Stage(OT\_Staging) vào DataWarehouse (OT\_DW)

Tương tư như bước 1 chúng ta tạo các Data Flow ứng với từng Dimension ứng với từng bảng Fact



Hình 3.10 Load dữ liệu từ OT\_Staging vào OT\_DW

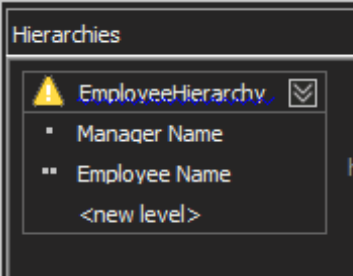
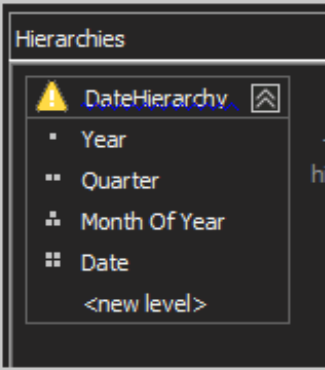
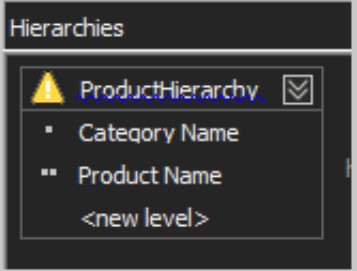


Hình 3.11 Thực thi thành công tất cả các Package

* 1. **Thiết kế Data Cube**

Sau khi dữ liệu được tích hợp vào kho dữ liệu, chúng ta sẽ tiến hành thiết kế Cube phục vục cho quá trình tổng hợp, phân tích dữ liệu, tạo báo cáo.

1. Tạo project: Sử dụng Visual Studio (có extension Microsoft Analysis Service) chọn tạo project Analysis Service Multidimensional and Data Mining Project,
2. Kết nối data source: Sau khi project mới được khởi tạo trong thanh SolutionExplorer chọn Data Sources> New Data Source để kết nối với kho dữ liệu.
3. Tạo data views: Sau khi kết nối với Data source thành công ta chọn DataSource Views>New Data Source View sau đó chọn tất cả table.
4. Tạo cube: Khi có Data Source View ta sẽ sữ dụng các view đó thể tạo cube.  
   Chọn Cubes> New Data Cube, chọn các Fact table và Next.
5. Tạo phân cấp: Khi tạo xong Cube có thể thực hiện trên các Dimension cần  
   thiết, ta sẽ thực, hiện phân cấp trên DimDate. Các Dimension còn tại ta lựa chọn những trường cần hiển thị cho vào phần Attributes. Sau khi hoàn thành các bước trên ta nhấp chuột phải vào project

Mô tả các cây phân cấp

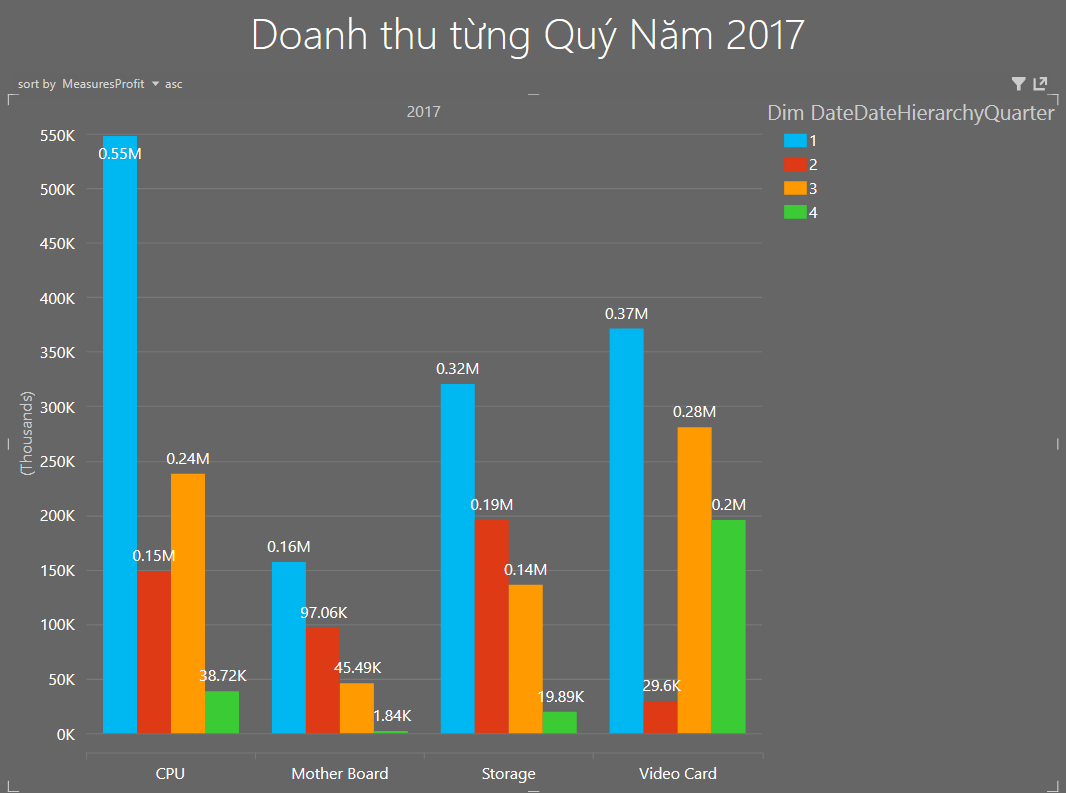
## **Tạo báo cáo**

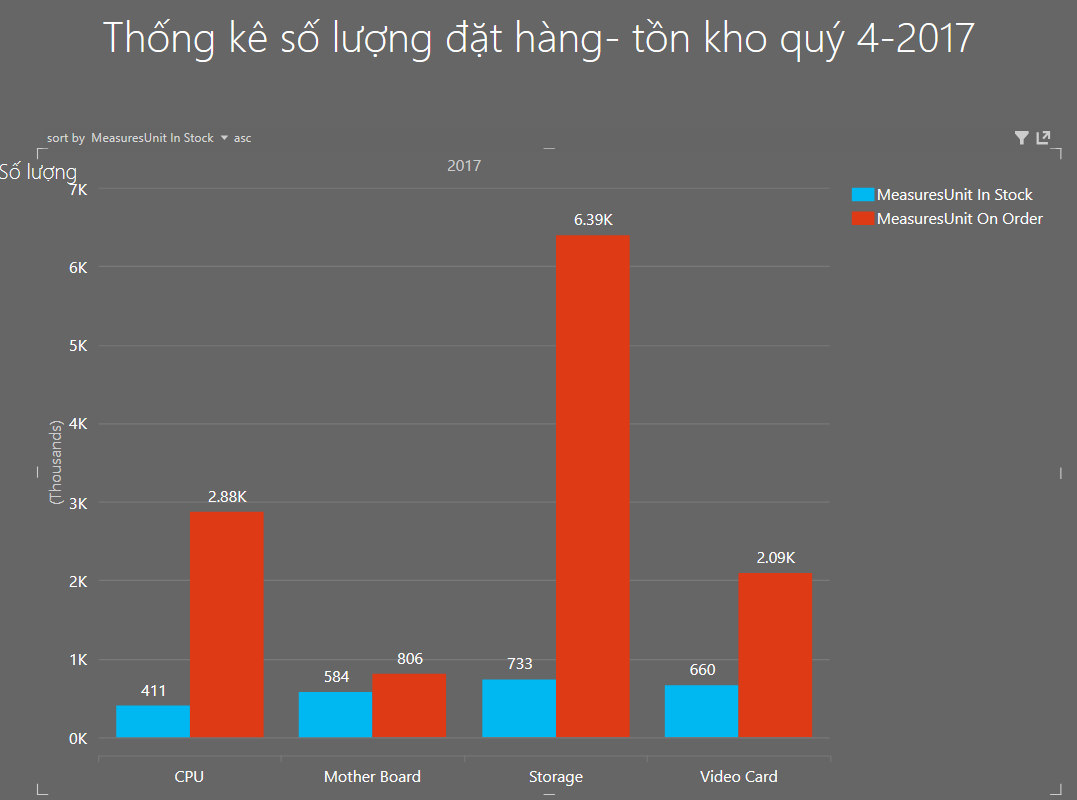
Khi đã có cube ta có thể tổng hợp, phân tích dữ liệu để cho ra các báo cáo hỗ trợ cho việc ra quyết định. Việc xuất báo cáo sẽ được thực hiện bảng Power Pivot của Microsoft Excel. Việc này được thực hiện theo các bước sau:

*Bước 1*: Kết nối với Analysis Service (cube): Power Pivot> Management>Get External Data > From Database > From Analysis Services or PowerPivot.Sau đó chọn và kết nối với cube vừa tạo.  
*Bước 2*: Thiết kế câu truy vấn: Chọn Design và kéo thả các thuộc tính cần thiết vào phần bảng tính toán. Sau khi thiết kế theo ý muốn, chọn OK > Finish để hoàn thành câu truy vấn.

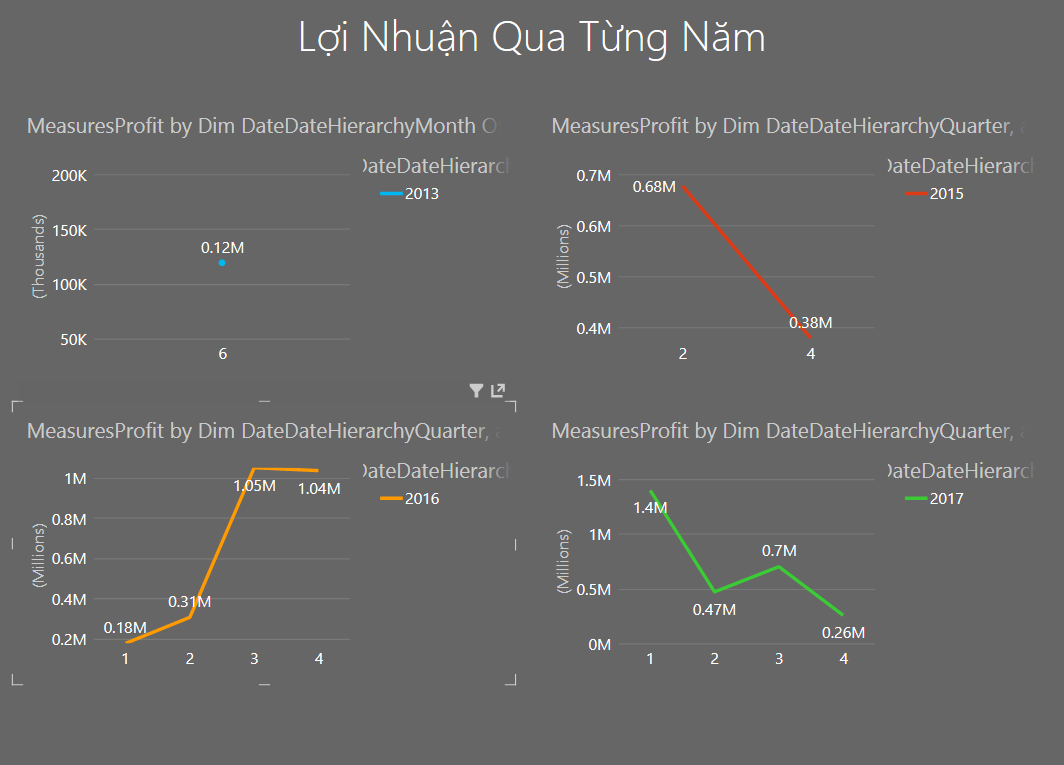
*Bước 3*: Thiết kế báo cáo: Sau khi thiết kế xong câu truy vấn ta có thể thiết kế báo cáo dự trên những dự liệu đã được tổng hợp trong câu truy vấn đó. Có thể tạo những biểu đồ dạng cột hoặc dạng bản đồ, giúp tang phân trực quan cho báo cáo.

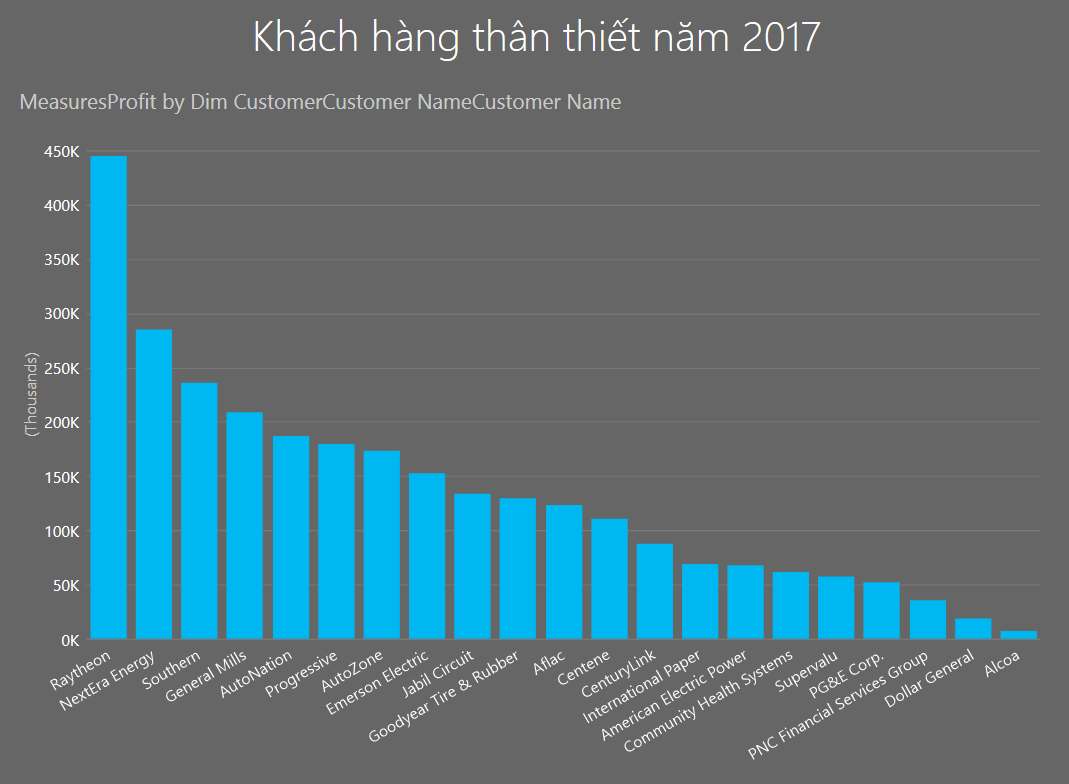
**Dưới đây là một số báo cáo tổng quan:**

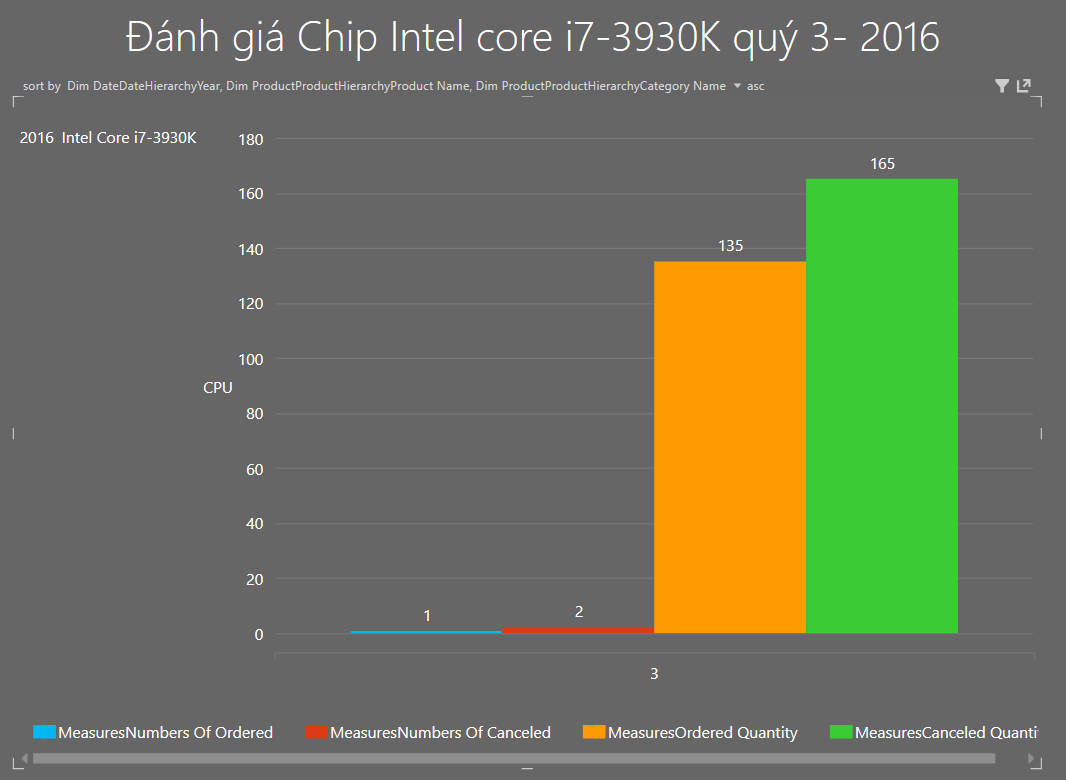


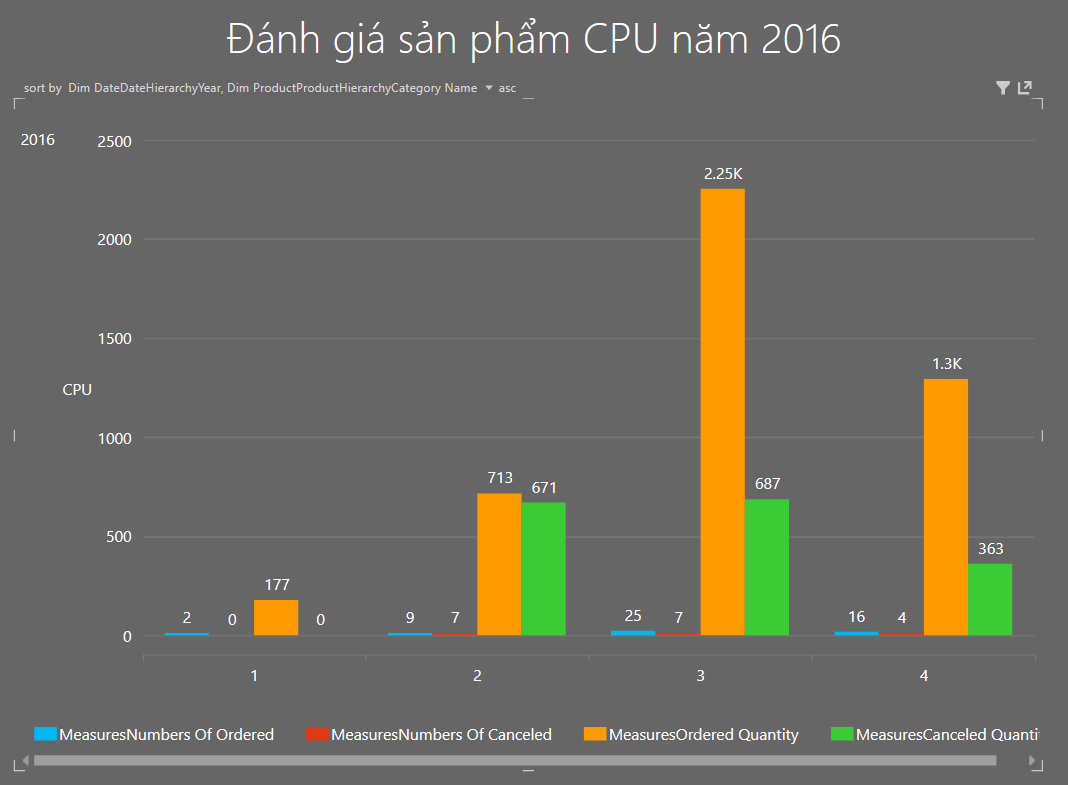












**KẾT LUẬN**

Kho dữ liệu là một công cụ hữu hiệu hỗ trợ cho quá trình ra quyết định của những nhà lãnh đạo liên quan đến các business process của tổ chức. Phần quan trọng nhất trong quá trình xây dựng một kho dữ liệu và việc thiết kế các star schema phù hợp với các business process mà nó phục vụ. Việc tích hợp dữ liệu vào kho dữ liệu cần phải cẩn thận và tốn thời gian để cho các dữ liệu được đồng nhất