**I. Giới thiệu**

**1.1. Mục tiêu**

Thiết kế và xây dựng một kho dữ liệu với mục đích hỗ trợ ra quyết định cho hệ thống xử lý đơn đặt hàng của khách hàng cho một công ty A.

**1.2. Phạm vi**

Phạm vi kho dữ liệu chỉ nằm trong các chức năng chính như quản lý cửa hàng, quản lý khách hàng, quản lý đơn hàng. Các chức năng khác như quản lý nhân viên không nằm trong phạm vi này.

**II. Phân tích yêu cầu**

**2.1. Yêu cầu nghiệp vụ**

Kho dữ liệu được xây dựng để cung cấp các xử lý trực tuyến cho các yêu cầu chọn lựa của khách hàng.

* Tìm tất cả các cửa hàng cùng với thành phố, bang, số điện thoại, mô tả, kích cỡ, trọng lượng và đơn giá của tất cả các mặt hàng được bán ở kho đó.
* Tìm tất cả các đơn đặt hàng với tên khách hàng và ngày đặt hàng được thực hiện bởi khách hàng đó
* Tìm tất cả các cửa hàng cùng với tên thành phố và số điện thoại mà có bán các mặt hàng được đặt bởi một khách hàng nào đó
* Tìm địa chỉ văn phòng đại diện với tên thành phố, bang của tất cả các cửa hàng lưu kho một mặt hàng nào đó với số lượng trên mức cụ thể.
* Với mỗi một đơn đặt hàng của khách, liệt kê các mặt hàng được đặt cùng với mô tả, mã cửa hàng, tên thành phố và các cửa hàng có bán mặt hàng đó.
* Tìm thành phố và bang mà một khách hàng nào đó sinh sống
* Tìm mức độ tồn kho của một mặt hàng cụ thể tại tất cả các cửa hàng ở một thành phố cụ thể nào đó
* Tìm các mặt hàng, số lượng đặt, khách hàng, cửa hàng và thành phố của một đơn đặt hàng.
* Tìm các khách hàng du lịch, khách hàng đặt theo đường bưu điện và khách hàng thuộc cả hai loại
* Tổng số lượng mỗi mặt hàng được lưu trữ tại mỗi thành phố. Sau mỗi lần một mặt hàng được phân phối (được bán), doanh nghiệp này cần biết tổng số lượng mỗi mặt hàng còn lại tại tất cả các cửa hàng trong một thành phố.

**2.2. Đặc tả chức năng**

Đầu vào của kho dữ liệu là cơ sở dữ liệu nguồn đã được tích hợp.

Sơ đồ cơ sở dữ liệu nguồn:

Đầu ra của kho dữ liệu chính là các khối dữ liệu để phục vụ cho việc phân tích trực tuyến OLAP.

**III. Thiết kế kho dữ liệu**

**3.1. Khái niệm kho dữ liệu**

Kho dữ liệu được định nghĩa theo nhiều cách khác nhau. Một trong những định nghĩa phổ biến nhất là định nghĩa của W.H.Inmon được phát biểu như sau: “Một kho dữ liệu là một bộ dữ liệu hướng chủ đề, tích hợp, biến động theo thời gian và không mất đi được sử dụng để hỗ trợ quá trình ra quyết định quản lý.”

**3.2. Các bước thiết kế kho dữ liệu**

***Bước 1: Chuyển đổi các cơ sở dữ liệu quan hệ sang các mô hình thực thể liên kết mở rộng.***

Cơ sở dữ liệu Văn phòng đại diện bao gồm những quan hệ với các lược đồ như sau:

Khách hàng (Mã KH, Tên KH, Mã Thành phố, Ngày đặt hàng đầu tiên)

Khách hàng du lịch (\*Mã KH, Hướng dẫn viên du lịch, Thời gian)

Khách hàng bưu điện (\* Mã KH, Địa chỉ bưu điện, Thời gian)

Các bước chuyển đổi:

Bước 1.1: Xác định các quan hệ, khóa và các trường

PR1: Khách hàng

Giải thích: Do quan hệ Khách hàng có khóa chính không thuộc khóa của quan hệ khác.

PR2: Khách hàng du lịch và khách hàng bưu điện

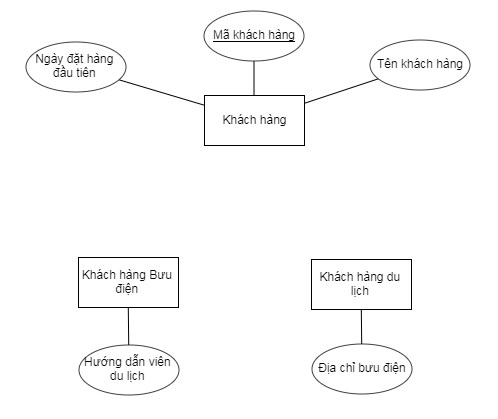
Giải thích: Quan hệ Khách hàng du lịch và Khách hàng bưu điện có khóa chính là khóa của quan hệ Khách hàng.

Bước 1.2: Ánh xạ các quan hệ và chuyển đổi

PR1 được ánh xạ sang là một thực thể Khách hàng.

PR2 được ánh xạ sang thực thể kiểu con. Do một khách hàng chỉ có thể là một trong hai loại khách hàng du lịch hoặc khách hàng bưu điện tại một thành phố. Nên mối quan hệ giữa các thực thể Khách hàng, Khách hàng du lịch và Khách hàng bưu điện là khái quát hóa không giao nhau của dữ liệu.

Mô hình EER cho CSDL Văn phòng đại diện:



d

Cơ sở dữ liệu Bán hàng gồm những quan hệ với lược đồ như sau:

Văn phòng đại diện (Mã Thành phố, Tên Thành phố, Địa chỉ VP, Bang, Thời gian)

Cửa hàng (Mã cửa hàng, \* Mã Thành phố, Số điện thoại, Thời gian)

Mặt hàng (Mã MH, Mô tả, Kích cỡ, Trọng lượng, Giá, Thời gian)

Mặt hàng\_được lưu trữ (\* Mã cửa hàng, \* Mã mặt hàng, Số lượng trong kho, Thời gian)

Đơn đặt hàng (Mã đơn, Ngày đặt hàng, Mã Khách hàng)

Mặt hàng được đặt (\* Mã đơn, \* Mã mặt hàng, Số lượng đặt, Giá đặt, Thời gian).

Các bước chuyển đổi:

Bước 1.1: Xác định các quan hệ, khóa và trường

PR1: Văn phòng đại diện, Cửa hàng, Mặt Hàng, Đơn đặt hàng

Giải thích: Do các quan hệ Văn phòng đại diện, Cửa hàng, Mặt Hàng, Đơn đặt hàng có khóa chính không chứa khóa của các quan hệ khác.

SR1: Mặt hàng được lưu trữ và Mặt hàng được đặt

Giải thích: Do quan hệ Mặt hàng được lưu trữ và Mặt hàng được đặt có khóa chính là khóa chính của quan hệ Cửa hàng, Mặt Hàng, Đơn đặt hàng.

FKA: Mã Thành phố

Giải thích: Thuộc tính mã thành phố là khóa chính của quan hệ Văn phòng đại diện và đồng thời là khóa ngoại của quan hệ Cửa hàng.

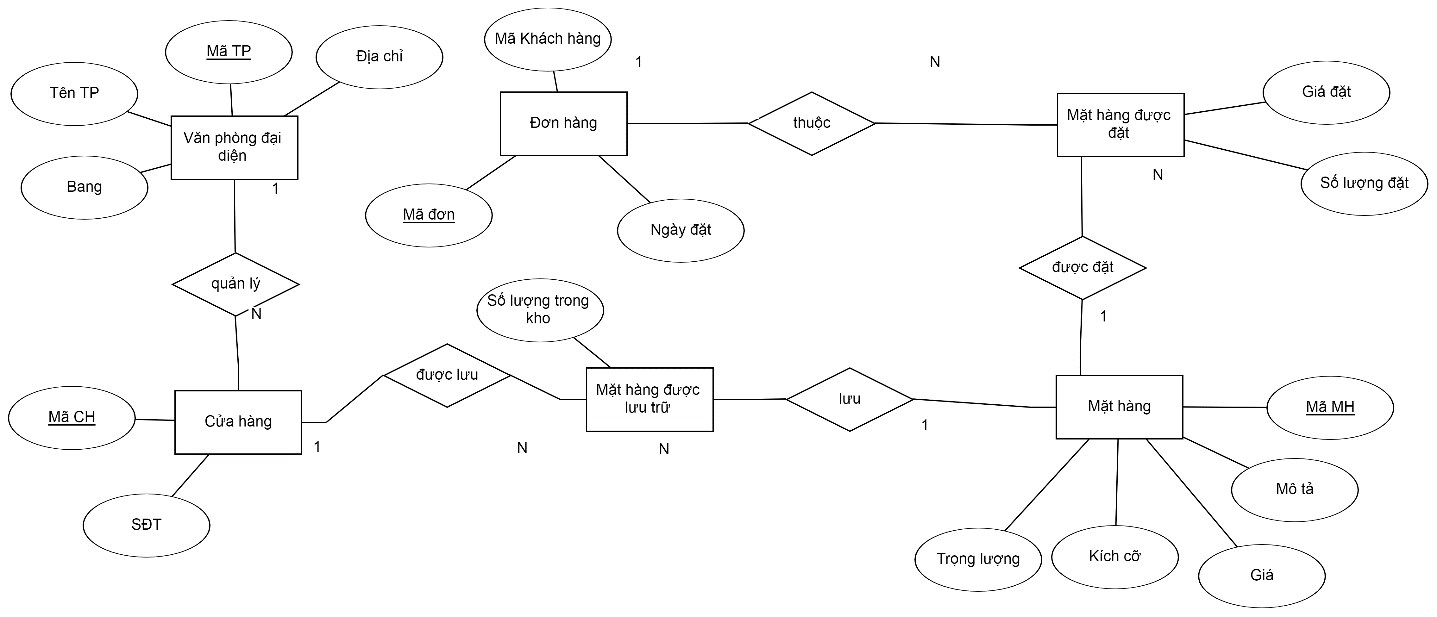
Bước 1.2: Ánh xạ và chuyển đổi các quan hệ

PR1 đại diện cho các thực thể. Do đó các PR1 là Văn phòng đại diện, Cửa hàng, Mặt Hàng, Đơn đặt hàng được ánh xạ sang các thực thể tương ứng là Văn phòng đại diện, Cửa hàng, Mặt Hàng, Đơn đặt hàng.

SR1 đại diện cho các liên kết. Các SR1 là Mặt hàng được lưu trữ và Mặt hàng được đặt được ánh xạ thành quan hệ hai ngôi Mặt hàng được lưu trữ và Mặt hàng được đặt có khóa chính là các khóa chính của quan hệ Cửa hàng và Mặt Hàng; Đơn hàng và Mặt hàng.

FKA: thuộc tính khóa ngoại. Do đó ánh xạ FKA sang một mối quan hệ giữa Văn phòng đại diện và Cửa hàng.

Mô hình EER cho CSDL Bán hàng:



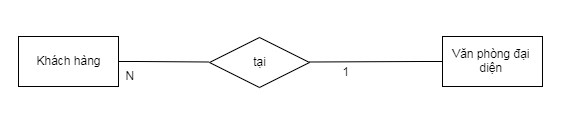
***Bước 2: Tích hợp các mô hình thực thể liên kết ở bước 1 thành một mô hình thực thể liên kết thống nhất.***

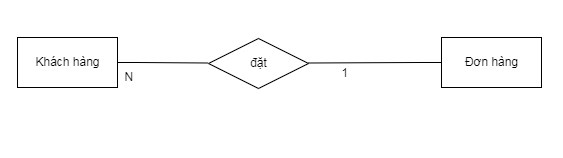
Bước 2.1: Giải quyết xung đột:

Ta thấy hai lược đồ trên không có xung đột về đồng âm, đồng nghĩa, xung đột kiểu dữ liệu, xung đột khóa, lực lượng và không có xung đột về thực thể yếu, thực thể kiểu con.

Bước 2.2: Trộn các thực thể dùng quan hệ hai ngôi

Ta thấy Thực thể Khách hàng có thuộc tính Mã Thành Phố không phải khóa chính. Thực thể Văn phòng đại diện có thuộc tính Mã Thành Phố là khóa chính. Suy ra thực thể Văn phòng đại diện và thực thể Khách hàng có quan hệ như sau:

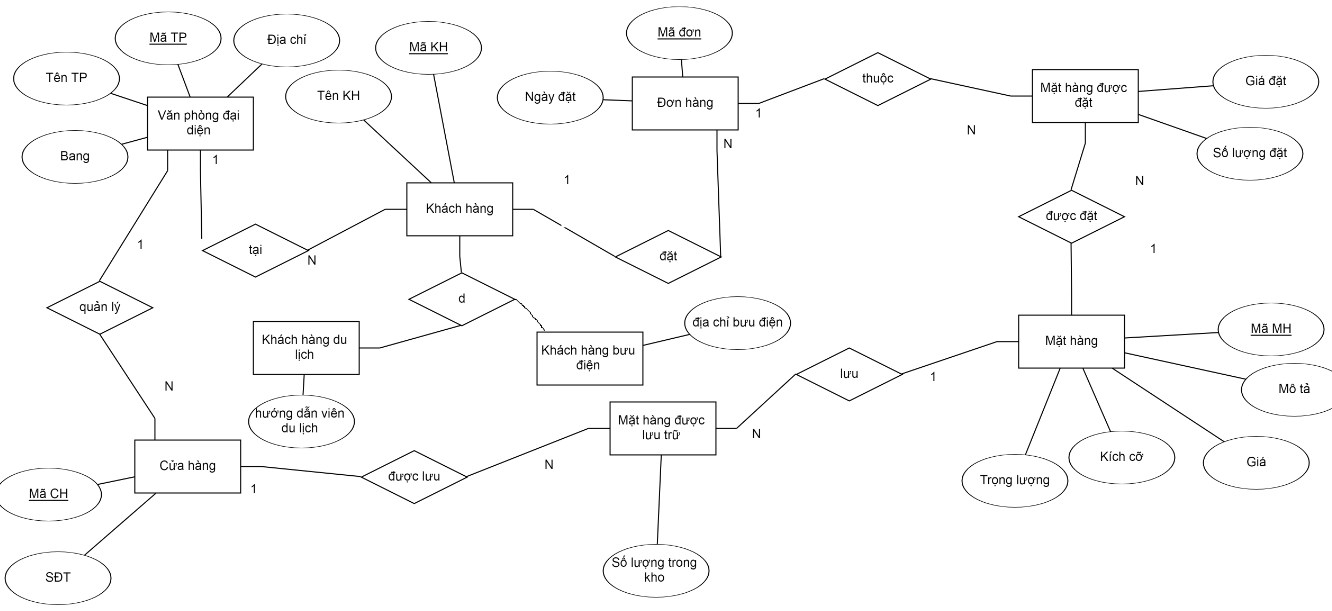


Tương tự, thực thể Khách hàng có thuộc tính Mã khách hàng là khóa chính, thực thể Đơn hàng có thuộc tính Mã khách hàng nhưng không phải là khóa chính. Do đó ta thể hiện quan hệ hai thực thể này như sau

Bước 2.3: Trộn các quan hệ

Ta thấy các EER đã thỏa mãn được quan hệ.

Tích hợp hai mô hình EER trên ta được mô hình EER thống nhất như sau:



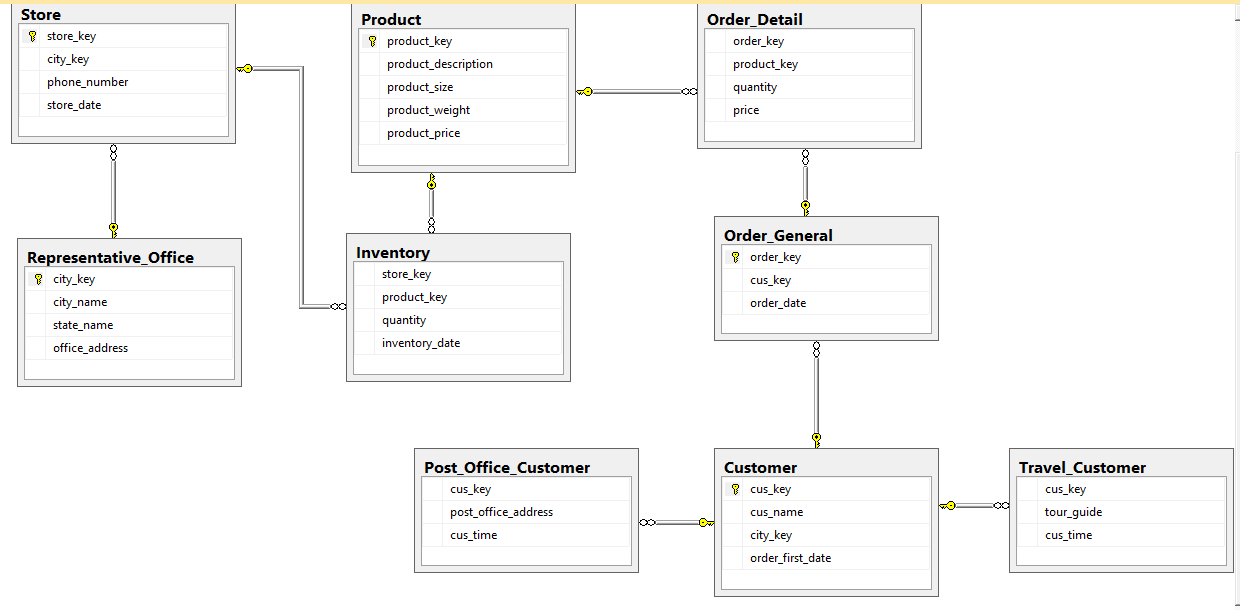
***Bước 3: Chuyển đổi mô hình thực thể liên kết mở rộng thành cơ sở dữ liệu tích hợp.***

Từ mô hình thực thể liên kết ở bước 2, ta ánh xạ các thực thể và quan hệ sang các bảng trong một cơ sở dữ liệu tích hợp.

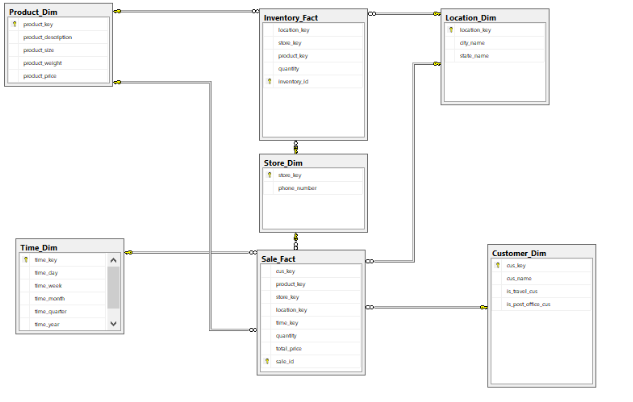
Cụ thể như sau:

* Thực thể Văn phòng đại diện thành bảng Representative Office
* Thực thể Cửa hàng thành bảng Store
* Thực thể Khách hàng thành bảng Customer
* Thực thể Đơn hàng thành bảng Order
* Thực thể Mặt hàng thành bảng Product
* Thực thể Mặt hàng được lưu trữ thành bảng Inventory
* Thực thể Mặt hàng được đặt thành bảng Order\_Detail
* Thực thể Khách hàng du lịch và Khách hàng bưu điện được lưu trữ thành hai bảng tương ứng là Travel\_Customer và Post\_Office\_Customer.

Ta có lược đồ cơ sở dữ liệu sau khi tích hợp như sau:



***Bước 4: Thiết kế mô hình hình sao cho kho dữ liệu***



***Bước 5: Ánh xạ cơ sở dữ liệu tích hợp vào kho dữ liệu***

1. Xây dựng các bảng Dimension
   * Dimension Store

Khóa chính của dim\_store là store\_key. Khóa này được ánh xạ từ thuộc tính store\_id của bảng Store từ cơ sở dữ liệu nguồn sang.

Các thuộc tính phone\_number được ánh xạ từ trường dữ liệu phone\_number của bảng Store trong cơ sở dữ liệu nguồn.

CREATE TABLE STORE\_DIM (

store\_key int,

phone\_num int

PRIMARY KEY (store\_key)

);

INSERT INTO STORE\_DIM

(store\_key, phone\_num)

SELECT Store.store\_key, Store.phone\_num

FROM Store

* + Dimension Product

Khóa chính của dim\_product là prod\_key. Khóa này được ánh xạ từ thuộc tính product\_key trong bảng Product cơ sở dữ liệu nguồn

Các thuộc tính product\_descreption, product\_price, product\_weight được ánh xạ từ trường dữ liệu product\_descreption, product\_price, product\_weight của bảng Product trong cơ sở dữ liệu nguồn.

Cú pháp tạo bảng dim\_product:

CREATE TABLE PRODUCT\_DIM (

product\_key int,

product\_size int,

product\_descreption varchar,

product\_weight float,

product\_price float,

PRIMARY KEY (product\_key)

);

INSERT INTO PRODUCT\_DIM

(product\_key, product\_size, product\_descreption, product\_weight, product\_price)

SELECT product\_key, product\_size, product\_descreption, product\_weight, product\_price

FROM Product

* + Dimension Order

Khóa chính của dim\_order là order\_key. Khóa này được ánh xạ từ thuộc tính order\_key trong bảng Order\_general từ cơ sở dữ liệu nguồn

Các thuộc tính order\_date được ánh xạ từ trường dữ liệu order\_date của bảng Order\_general trong cơ sở dữ liệu nguồn.

Cú pháp tạo bảng dim\_order:

CREATE TABLE DIM\_ORDER (

order\_key int,

order\_date date,

PRIMARY KEY (order\_key)

);

INSERT INTO DIM\_ORDER

(order\_key, order\_date)

SELECT order\_key, order\_date

FROM Order\_general

* + Dimension Customer

Khóa chính của dim\_customer là cus\_key. Khóa này ánh xạ từ thuộc tính khóa là cus\_key của bảng Customer trong cơ sở dữ liệu nguồn.

Các thuộc tính cus-name, được ánh xạ từ trường dữ liệu cus\_name của bảng Customer trong cơ sở dữ liệu nguồn.

* + Dimension Time

Khóa chính của dim\_time là time\_key. Khóa này được các luồng xử lý dữ liệu nội tại của kho dữ liệu sinh ra.

Các thuộc tính mô tả: day, week, month, quarter và year.

Cú pháp tạo bảng DIM\_TIME

CREATE TABLE DIM\_TIME (

Time\_key int,

Day int,

Week int,

Month int,

Year int,

PRIMARY KEY (time\_key)

);

1. Xây dựng bảng Fact
   * Sale\_Fact

Bảng Sale\_Fact có khóa chính là sale\_id.

Các thuộc tính time\_key, store\_key, order\_key, product\_key, customer\_key được ánh xạ từ các khóa chính time\_key, store\_key, order\_key, product\_key, customer\_key của các chiều tương ứng.

Các độ đo là quantity: tổng số lượng hàng bán được và totalprice: tổng số tiền bán được.

Cú pháp tạo bảng Fact

CREATE TABLE SALE\_FACT (

Sale\_id int

cus\_key int,

product\_key int

store\_key int

location\_key int

time\_key int

order\_key int

quantity int

total\_price float

PRIMARY KEY (sale\_id)

);

INSERT\_INTO SALE\_FACT (

cus\_key, product\_key, store\_key, location\_key, time\_key, order\_key,

quantity, total\_price )

SELECT cus\_key, product\_key, store\_key, location\_key, order\_key

FROM SALE\_FACT

INNER JOIN Dim\_Customer ON Dim\_Customer.cus\_key = Sale\_Fact.cus\_key

INNER JOIN Dim\_Product ON Sale\_Fact.product\_key= Dim\_Product. product\_key

INNER JOIN Dim\_Store ON Sale\_Fact.store\_key= Dim\_Store. store\_key

INNER JOIN Dim\_Location ON Sale\_Fact.location\_key= Dim\_Location. location\_key

INNER JOIN Dim\_Order ON Sale\_Fact.order\_key= Dim\_Order. order\_key

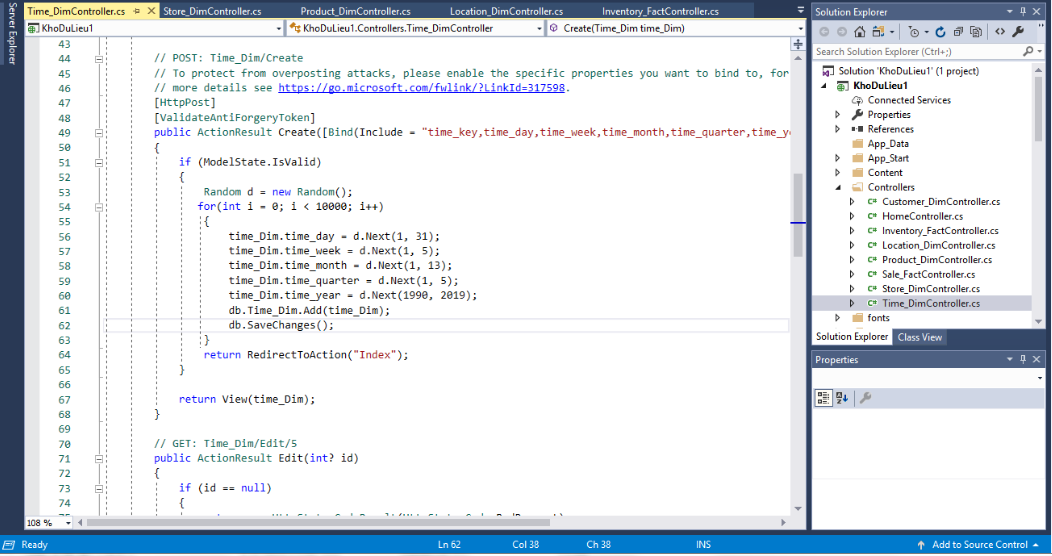
* + Inventory\_Fact

Bảng Fact này lưu giữ tổng số lượng hàng còn lại trong kho. Khóa chính của bảng này được cấu thành từ hai khóa của hai chiều dim\_prod và dim\_store. Cụ thể là prod\_key và store\_key.

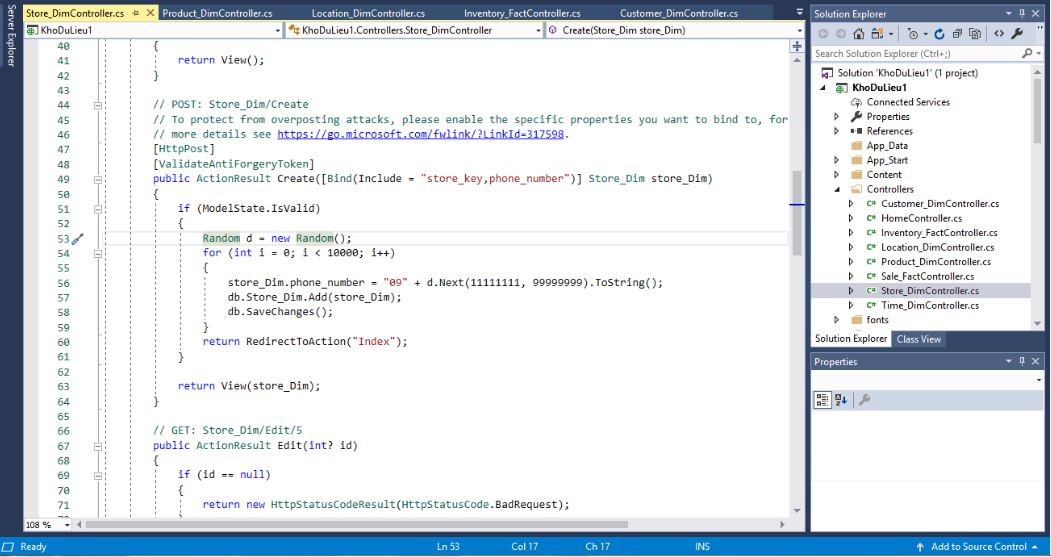
Tạo bảng Inventory\_Fact tương tự như Sale\_Fact.

***Bước 6: Sinh dữ liệu cho kho dữ liệu***

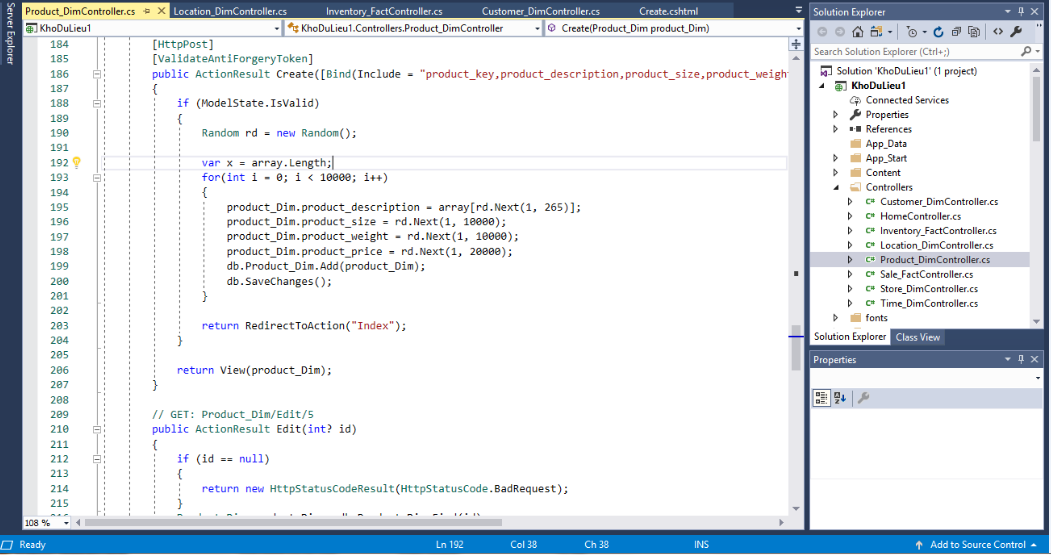
* + Sinh dữ liệu cho chiều Dim\_Time



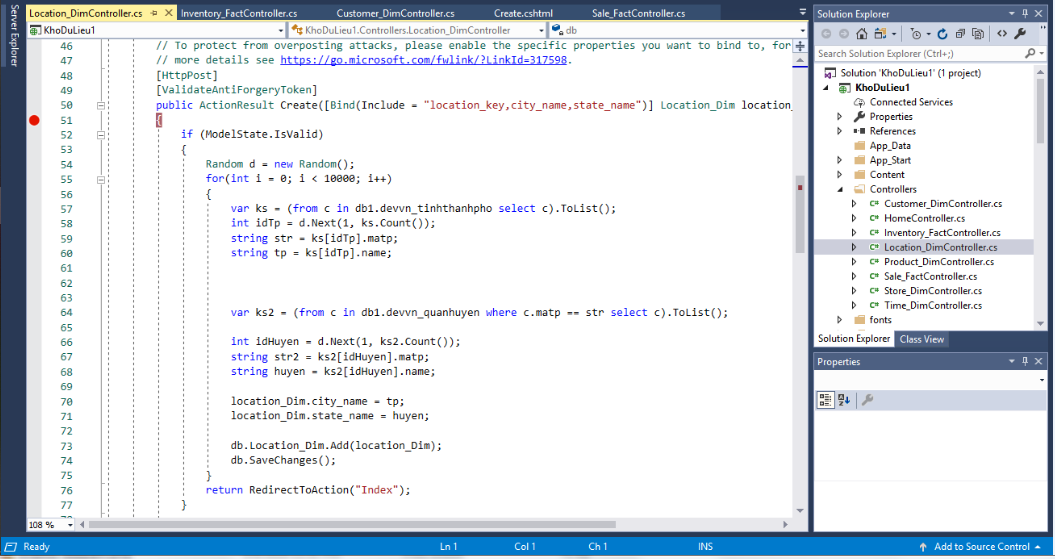
* + Sinh dữ liệu cho chiều Dim\_Store



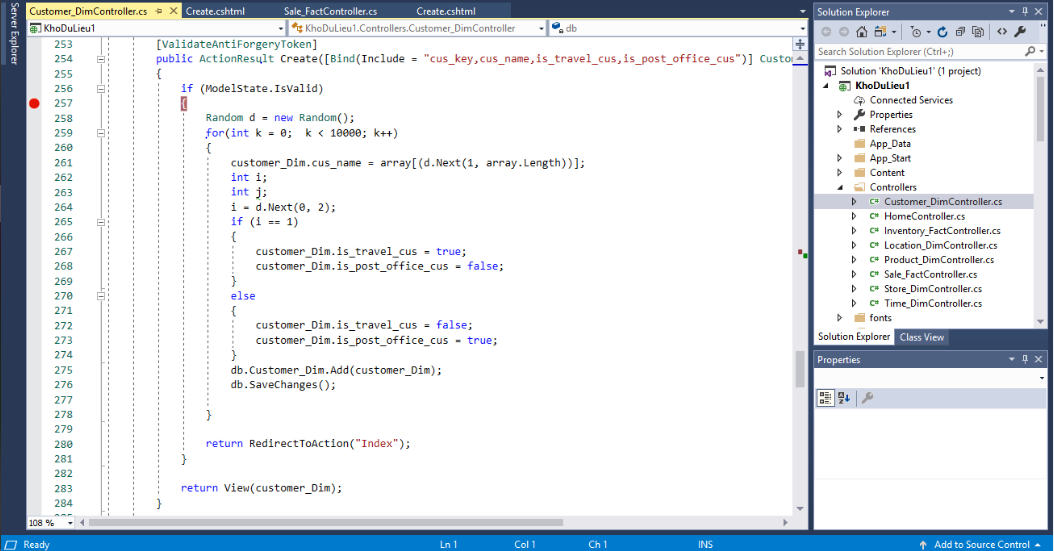
* + Sinh dữ liệu cho chiều Dim\_Product



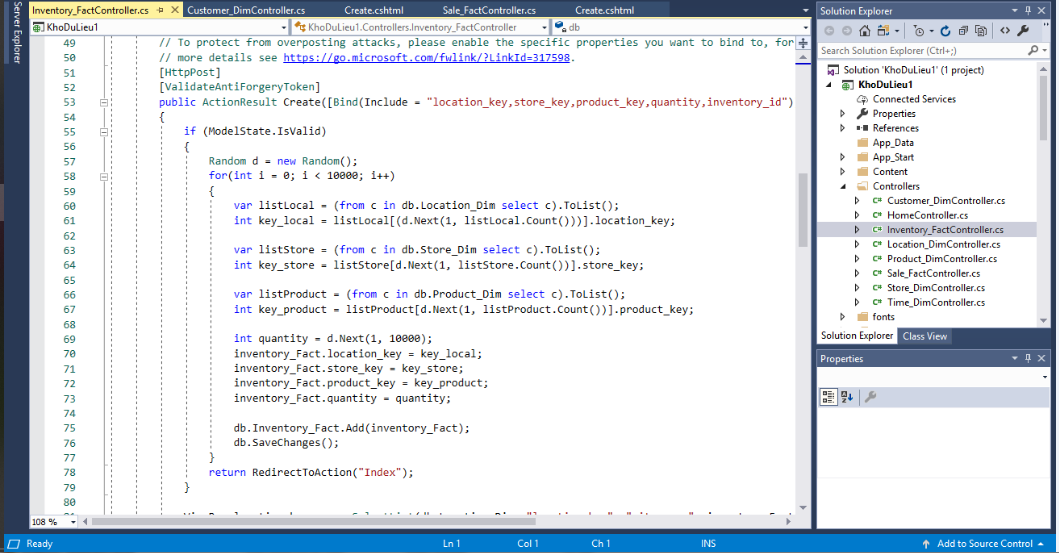
* + Sinh dữ liệu cho chiều Dim\_Location



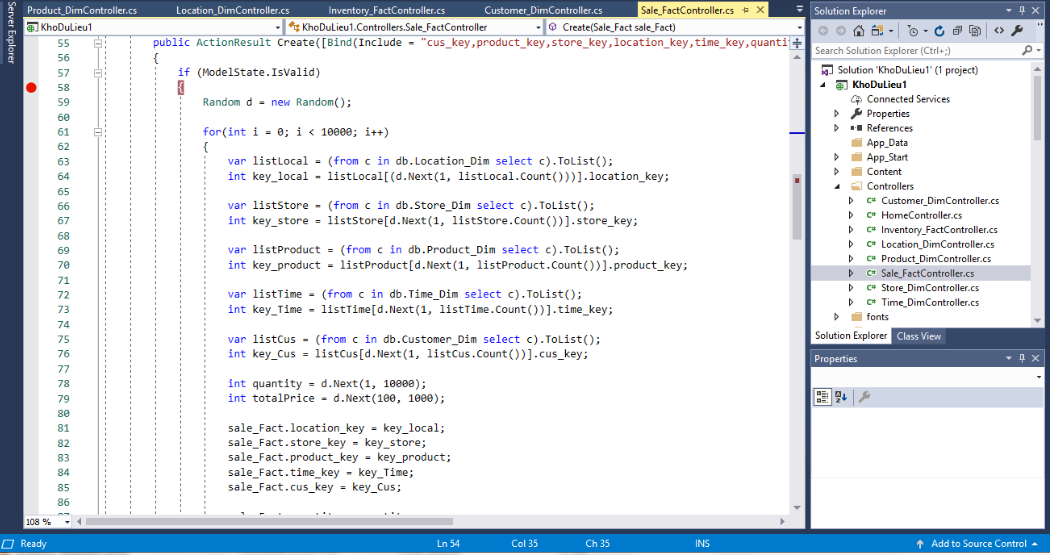
* + Sinh dữ liệu cho chiều Dim\_Customer



* + Sinh dữ liệu cho bảng Inventory\_Fact



* + Sinh dữ liệu cho bảng Sale\_Fact



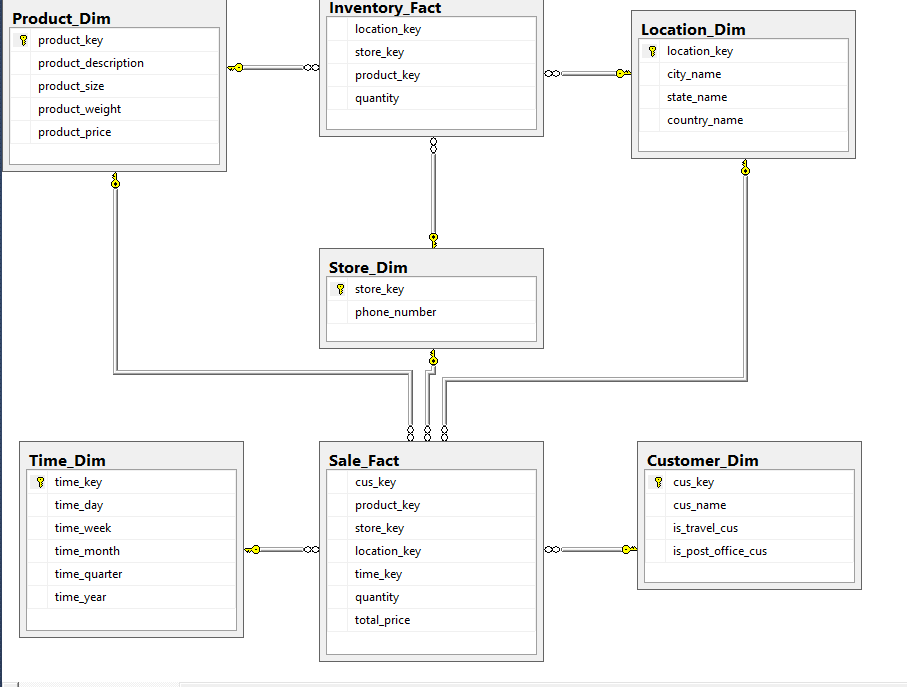
***Bước 7: Tạo phân cấp cho các chiều và hình thành các khối dữ liệu***

* + Chiều dữ liệu Product có một phân cấp là Category -> Product. Ý nghĩa của sự phân cấp này nói lên rẳng mỗi loại sản phẩm sẽ có nhiều sản phẩm khác nhau., vì thế khi biết số lượng hàng bán được của mỗi sản phẩm thì sẽ tổng hợp được số hàng bán được của mỗi loại sản phẩm.
  + Chiều Location có một phân cấp là City -> Store. Chiều phân cấp này thể hiện một thành phố sẽ có một hoặc nhiều cửa hàng. Do đó nếu biết số lượng hàng bán được, số lượng hàng còn lại ở mỗi cửa hàng sẽ biết được số lượng hàng bán được, số lượng hàng còn lại tại mỗi thành phố.
  + Chiều thời gian có phân cấp là: year -> quarter -> month-> week-> day. Chiều phân cấp này cho phép nếu biết được số tiền bán được của mỗi sản phẩm theo ngày sẽ tính được số tiền bán được theo tuần, theo tháng, theo quý và theo năm.

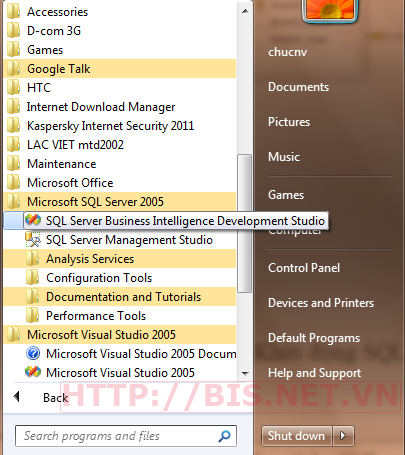
1. **Cài đặt các khối dữ liệu**

Các bước thực hiện:

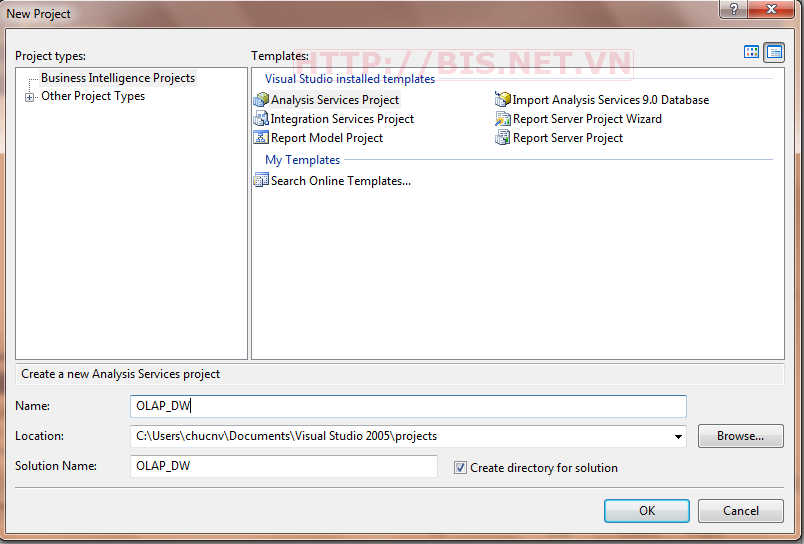
Khởi động SQL Server Management Studio và tạo CSDL có tên **DW** như sau và nhập vào các bảng một số records để phân tích.



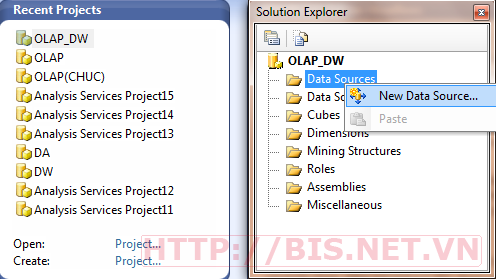
Khởi động SQL Server Business Intelligence Development Studio



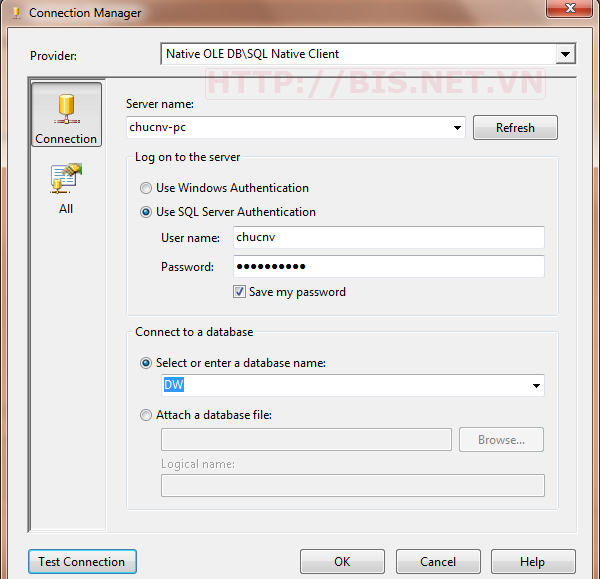
Tạo một Analysis Services Project mới có tên “**OLAP\_DW**”



Trong của sổ Solution Explorer của Project **OLAP\_DW**, bấm phím phải chuột vào Data Source để tạo một bộ kết nối đến dữ liệu dùng cho phân tích.

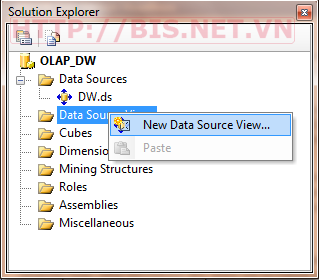


Xác định các tham số kết nối đến kho dữ liệu có tên “DW” đã tạo ra trong SQL Server Management Studio.

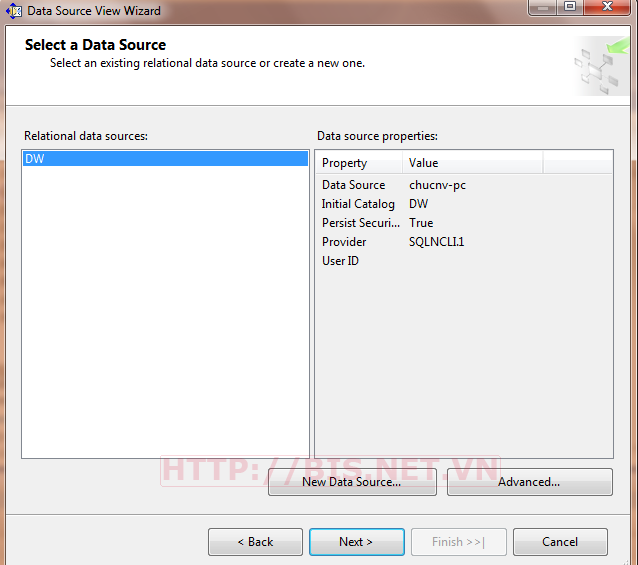


Đặt tên cho Data Source vàm bấm Finish để hoàn thành việc kết nối đến cơ sở dữ liệu.

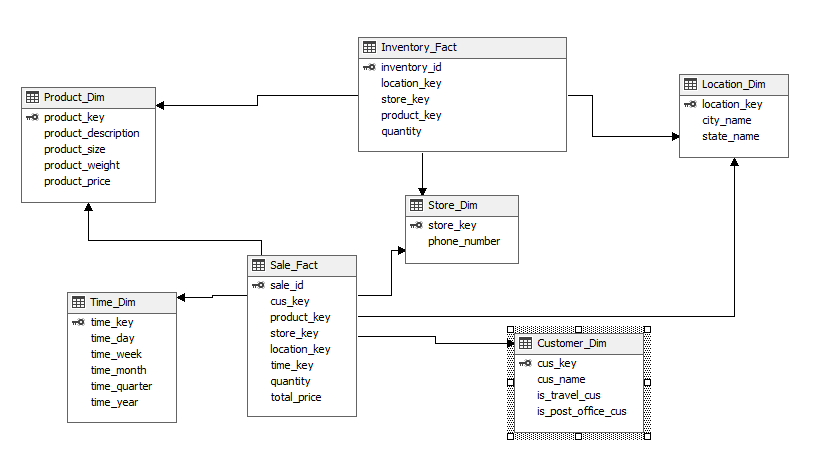
 Tạo Data Source View để lấy các bảng dữ liệu cần thiêt cần cho phân tích. Bấm phím phải chuột vào Data Source View trong của sổ Solution Explorer chọn New Data Source View



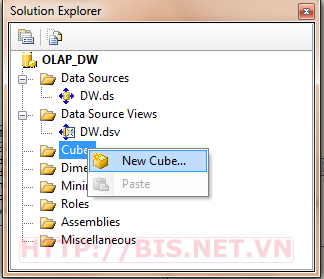
Xác định nguồn dữ liệu (Data Source) cần lấy là DW mới vừa tạo ra ở bước trước



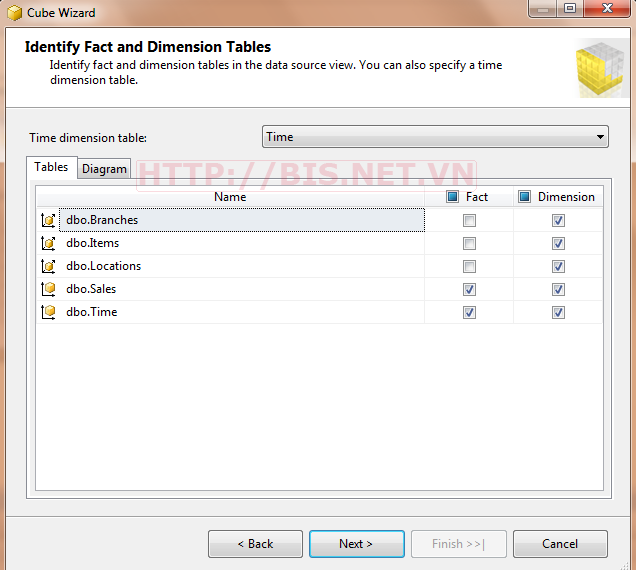
Sau khi hoàn thành, các bảng Fact và Dimension như sau:



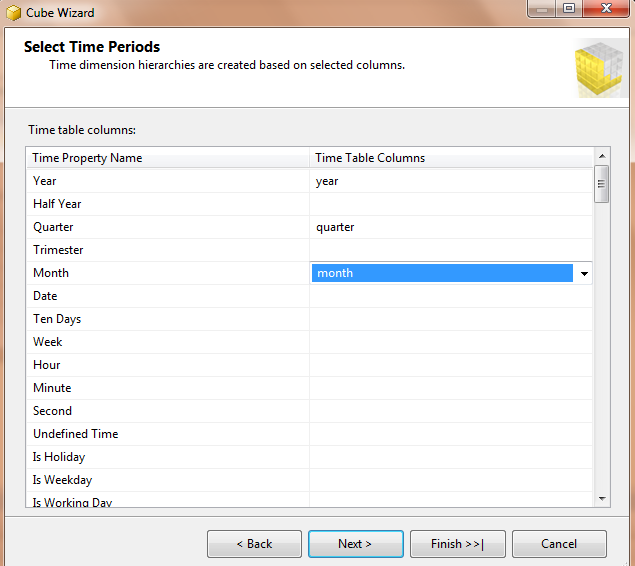
Sau khi tao Data Source và Data Source View ta tạo dữ liệu khối cho phân tích bằng cách bấm chuột phải lên Cube trong Solution Explorer và chọn New Cube



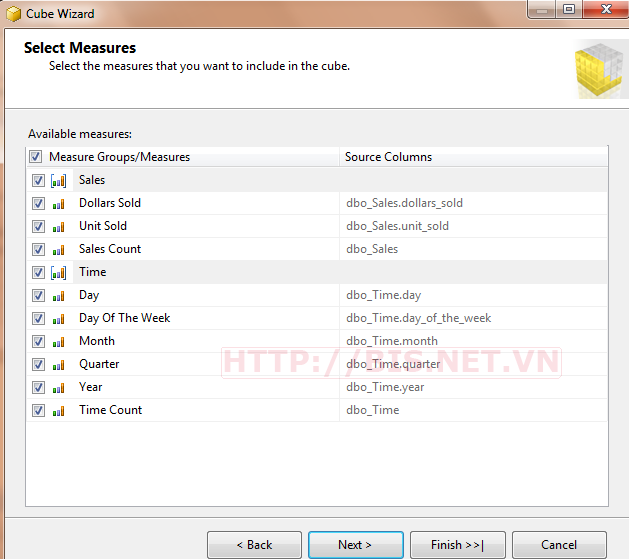
Chọn Next và chọn nguồn dữ liệu cho Khối (DW), hệ thống sẽ tự động dò tìm fact và Dimension Tables



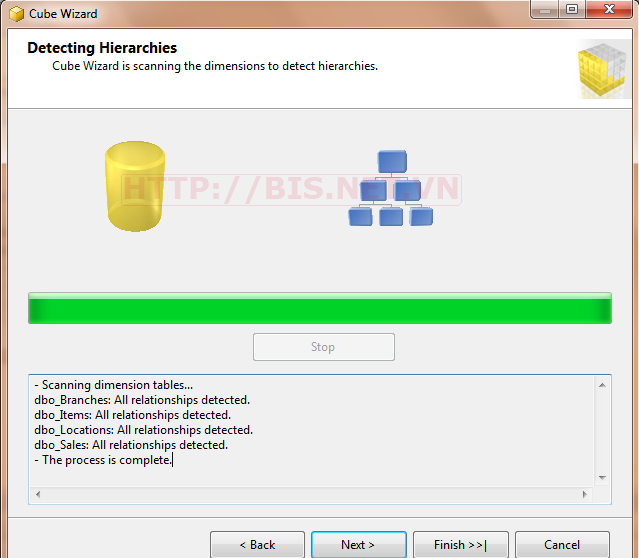
Bấm Next để thiết lập chiều thời gian. Chú ý, thời gian là một chiều rất quan trọng trong kho dữ liệu nói chúng và phân tích OLAP. Vì vậy nếu bạn không xác định chiều thời gian thì hệ thống sẽ tự động tạo ra một chiều thời gian để quản lý.



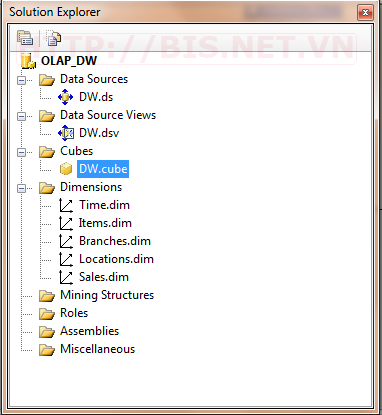
Bấm Next để xác định các độ đo (Measure) cho phân tích.



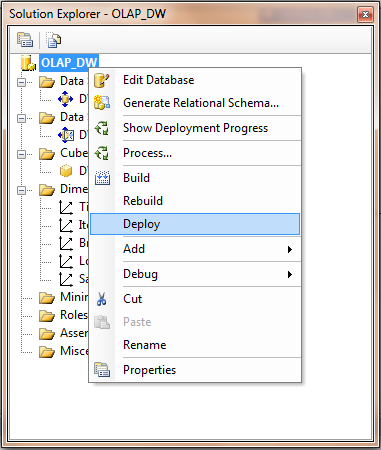
Bấm Next, hệ thống sẽ tự động phát hiện các cấu trúc phân cấp (Hierarchies) trong các Dimesion Tables



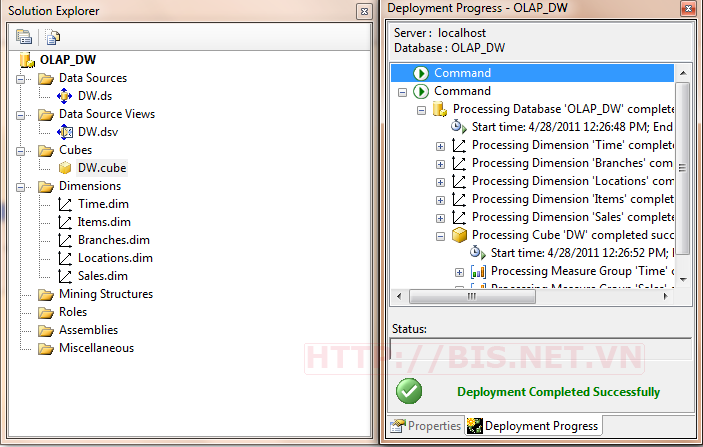
Đặt tên khối (DW) và bấm finish để sinh ra khối. Khối dữ liệu với các chiều được sinh ra



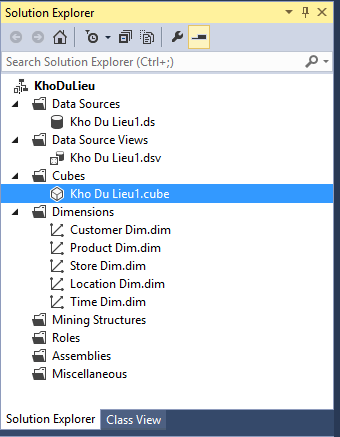
Sau khi tạo ra khối dữ liệu cho phân tích, để thực thi OLAP ta bấm phím phải chuột vào tên project trong Solution Explorer và chọn **Deploy**



Project được thực thi thành công như sau



Sau khi thực thi xong project, để thực hiện các phân tích OLAP phục vụ cho công tác quản lý, bấm phím phải chuột vào Cube trong Solution Explorer chọn **Browse** để xuất hiện mà hình phân tích:



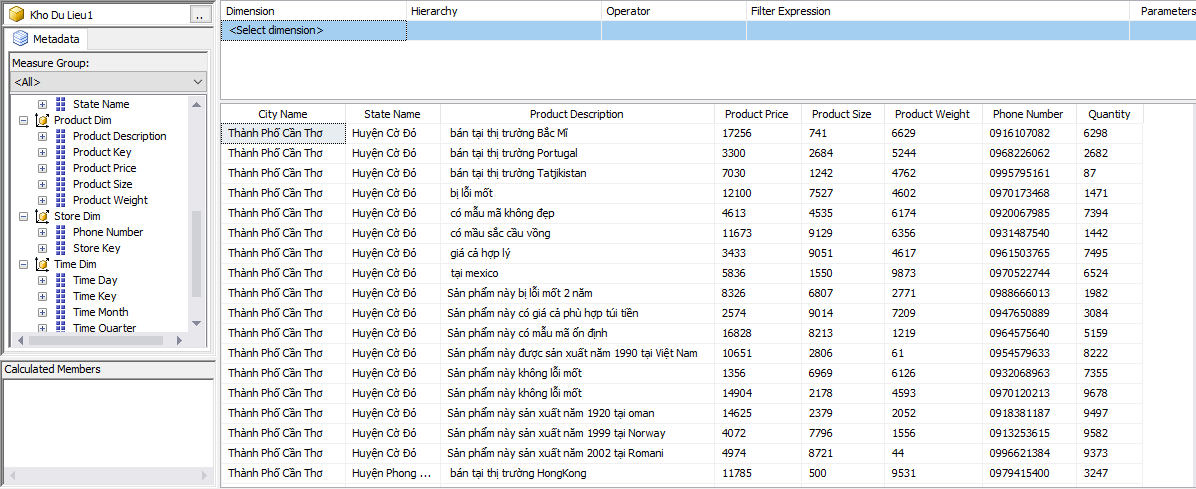
1. **Các phân tích OLAP**

Màn hình phân tích OLAP như sau:

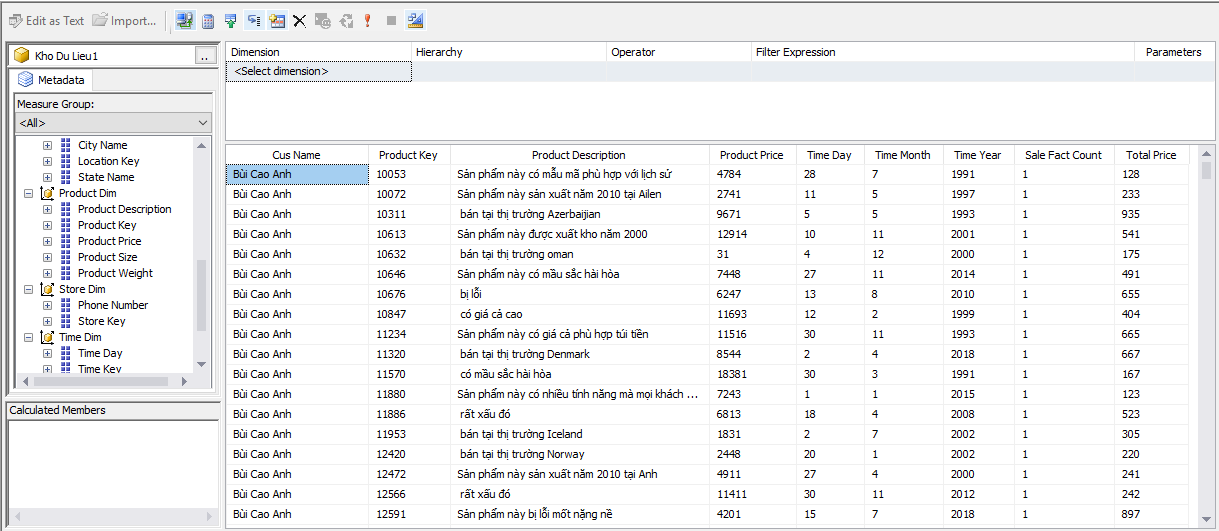
Panel bên trái chứa các Measure và các Dimensions đã định nghĩa khi xây dựng khối.

Panel bên phải chia làm 2 cửa sổ, cửa sổ phía trên dùng để xác định các điều kiện để phân tích, cửa sổ phía dưới chứa kết quả các measure khi ta kéo thả (drag and drop) các measure từ panel bên trái qua. Tùy theo mục đích phân tích mà chúng ta xác lập các biểu thức phân tích cho phù hợp.

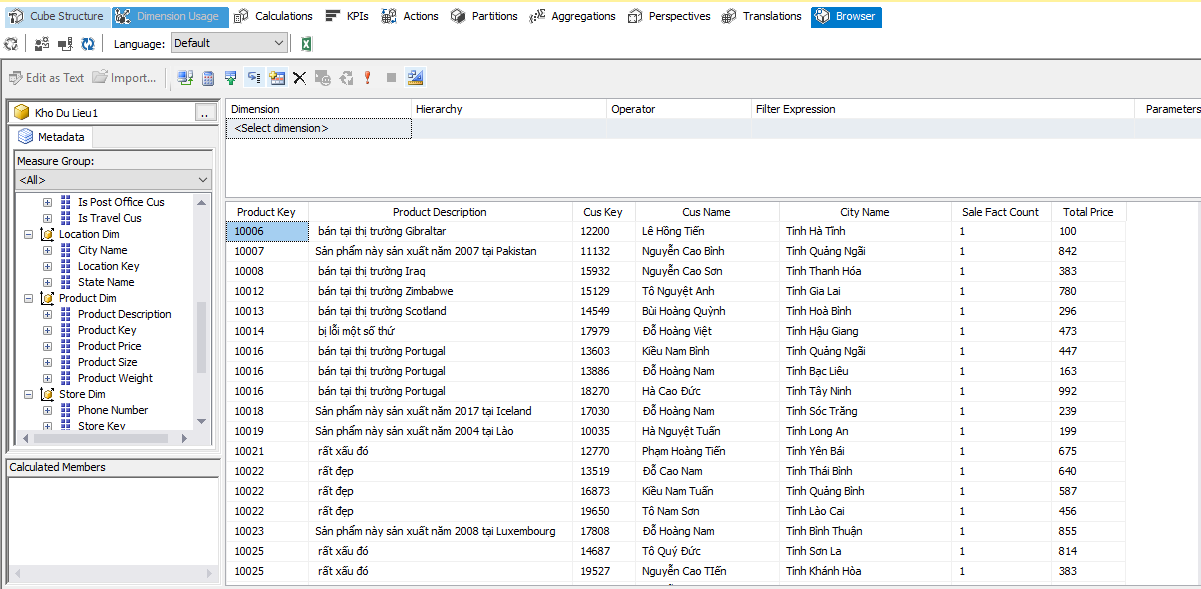
Tìm tất cả các cửa hàng cùng với thành phố, bang, số điện thoại, mô tả, kích cỡ, trọng lượng và đơn giá của tất cả các mặt hàng được bán ở kho đó



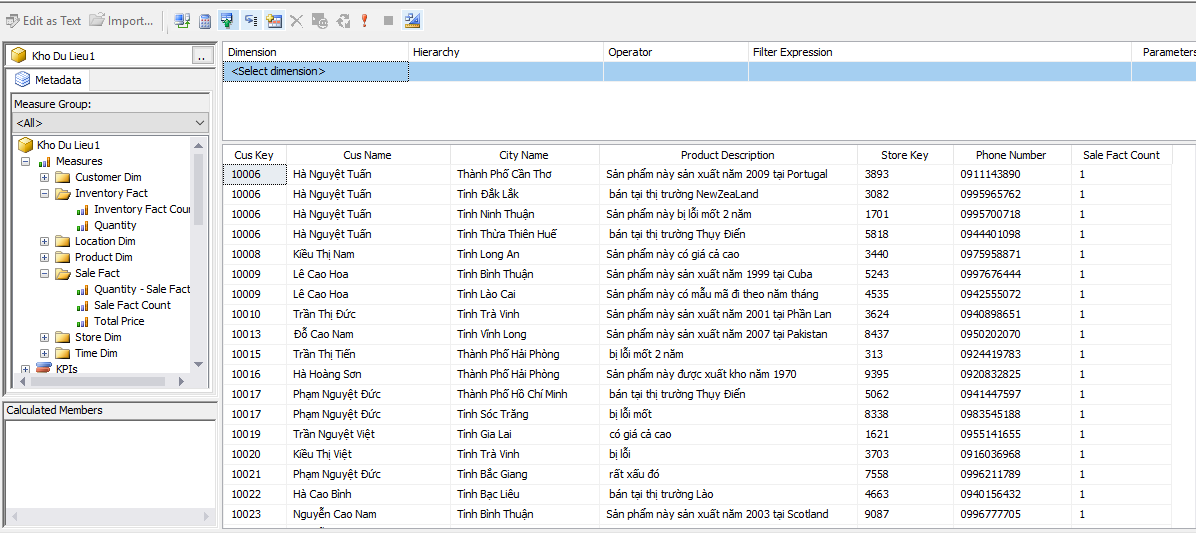
Tìm tất cả các đơn đặt hàng với tên khách hàng và ngày đặt hàng được thực hiện bởi khách hàng đó

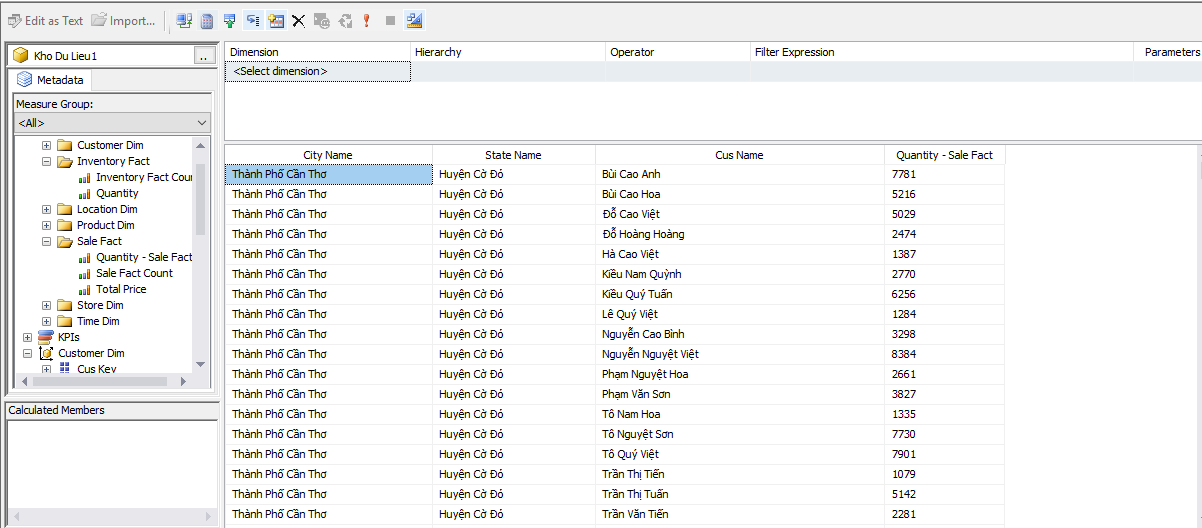


Tìm tất cả các cửa hàng cùng với tên thành phố và số điện thoại mà có bán các mặt hàng được đặt bởi một khách hàng nào đó



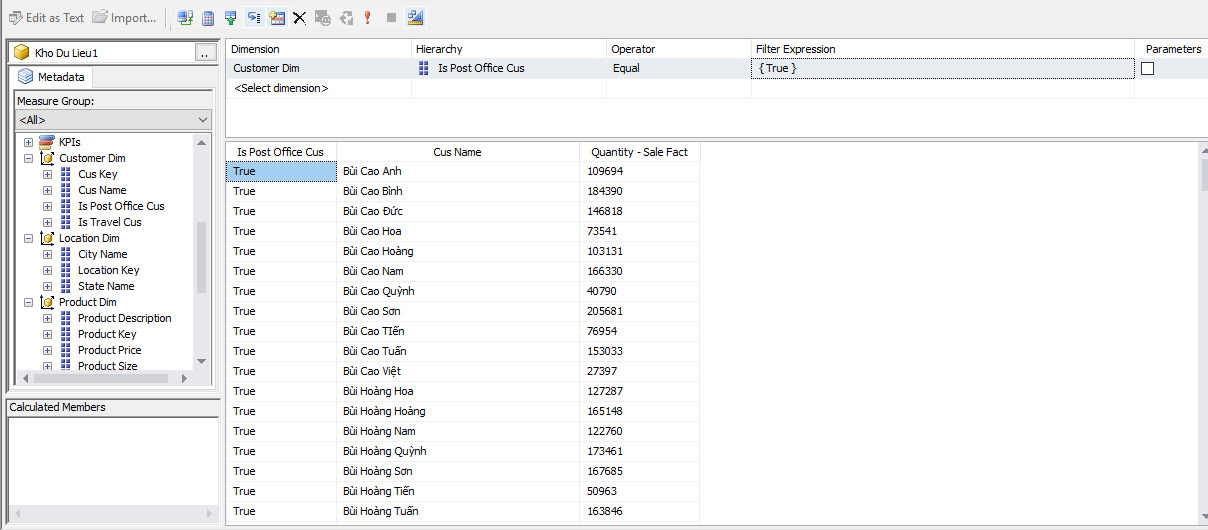
Với mỗi một đơn đặt hàng của khách, liệt kê các mặt hàng được đặt cùng với mô tả, mã cửa hàng, tên thành phố và các cửa hàng có bán mặt hàng đó.



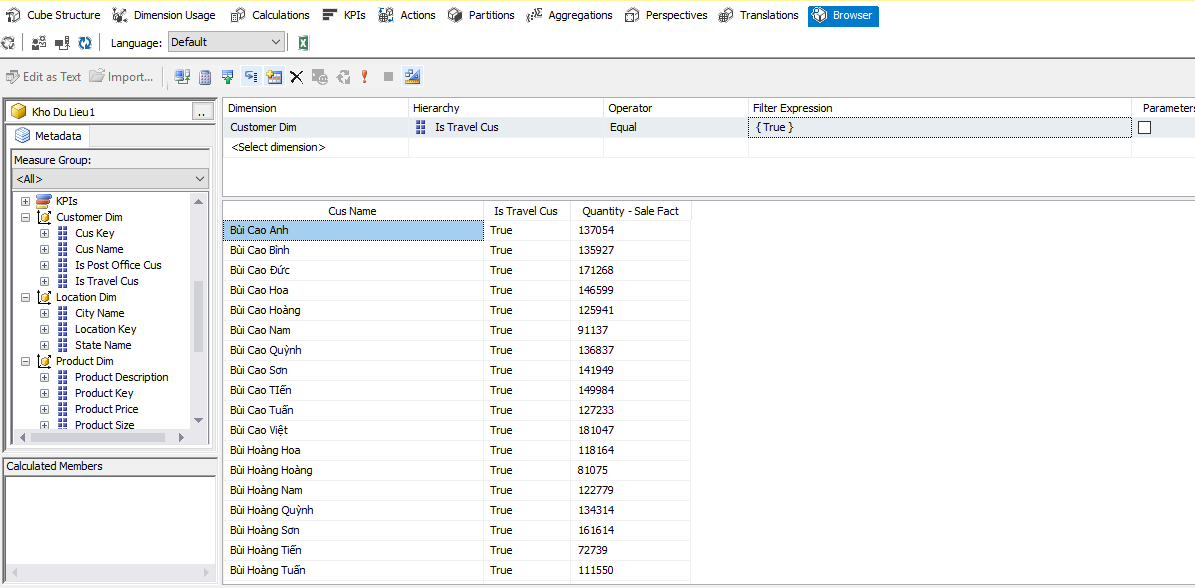


Tìm các khách hàng du lịch, khách hàng đặt theo đường bưu điện và khách hàng thuộc cả hai loại

* Khách hàng bưu điện



* Khách hàng du lịch



1. **Kết luận**

Báo cáo đã trình bày được các bước để thiết kế một kho dữ liệu cho một bài toán đơn giản. Từ đó đã xây dựng và tạo các báo cáo phân tích trực tuyến OLAP trên server.

Tuy nhiên, báo cáo chưa đạt được mục đích cuối cùng là tạo trang Web phân tích trực tuyến phục vụ cho người dùng.