BÁO CÁO ĐỒ ÁN THỰC HÀNH:

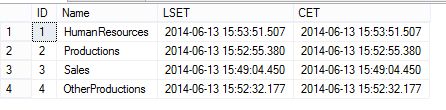
QUẢN LÝ CÔNG TY

BỘ MÔN: BI

GVHD: NGUYỄN THỊ NHƯ ANH

SINH VIÊN THỰC HIỆN :

PHẠM THÀNH CÔNG - MSSV: 1112030

1. **YÊU CẦU BẮT BUỘC**
   1. Thiết Kế Kho Dữ Liệu
      * + 1. Metadata: gồm bảng data\_flow 
          2. Stage: giữ nguyên thiết kế từ Source bao gồm các khoá chính, khoá ngoại, ràng buộc, index...

(lược đồ csdl: file ảnh đính kèm)

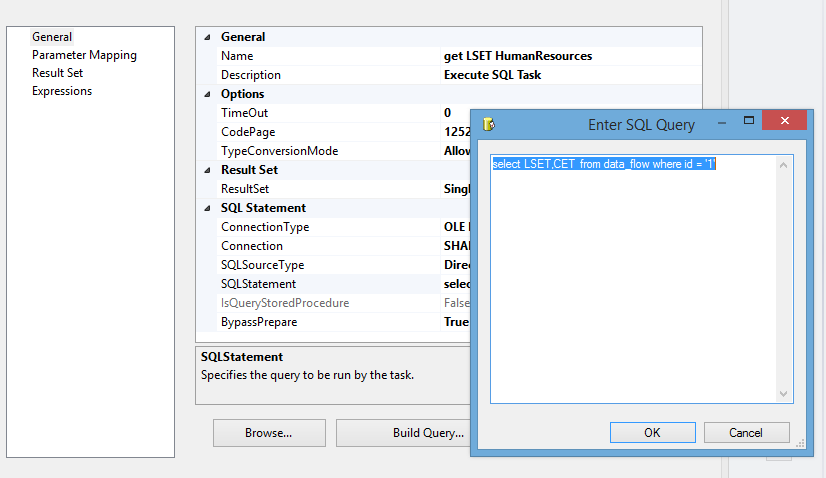
* + - * 1. NDS: thiết kế cơ sở dữ liệu ở đạt dạng chuẩn cao nhất có thể, gồm các công việc tách bảng, thêm khoá chính, khoá ngoại... tránh trùng lắp thông tin.

(lược đồ csdl: file ảnh đính kèm)

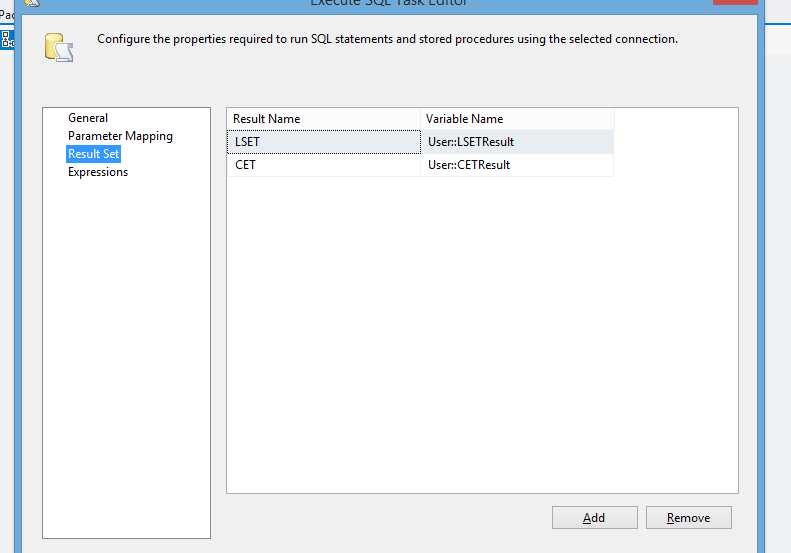
* + - * 1. DDS: thiết kế một DDS duy nhất đảm bảo các nhu cầu phân tích và báo cáo (vì công việc build cube rất tốn thời gian nên việc build càng ít cube thì càng tốt)

(lược đồ csdl: file ảnh đính kèm)

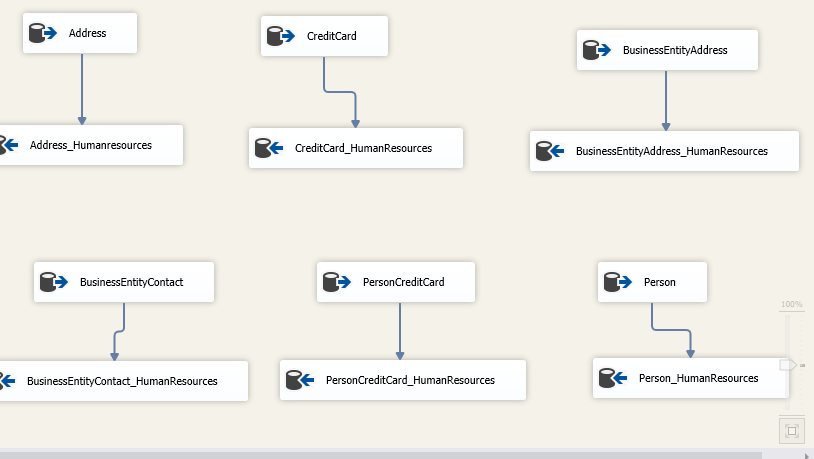
* 1. Thực Hiện ETL Nạp Định Kì Dữ Liệu Vào Kho
     + - 1. Source sang Stage
     1. Cơ sở dữ liệu Humanresources:
        + Cập nhật CET cho bảng data\_flow của Metadata
        + Lấy dữ liệu LSET, CET từ Metadata:
          - Dùng Execute SQL Task để lấy dữ liệu LSET, CET từ metadata dòng có id = ‘1’



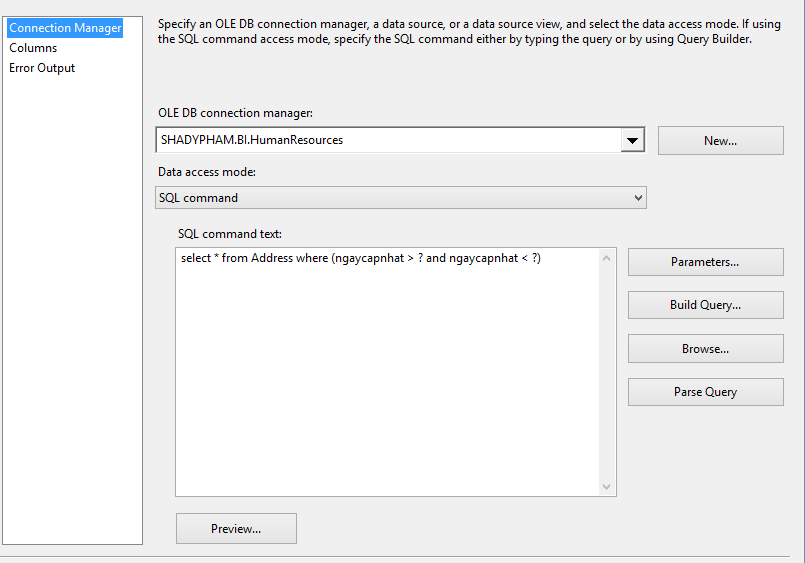
* + - * + Tạo các biến nhận giá trị của LSET, CET



* + - * Đổ dữ liệu từ Source: Humanresources sang Stage

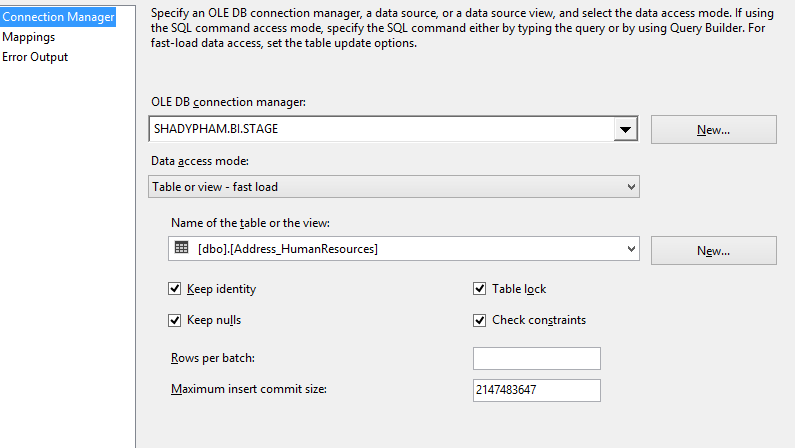


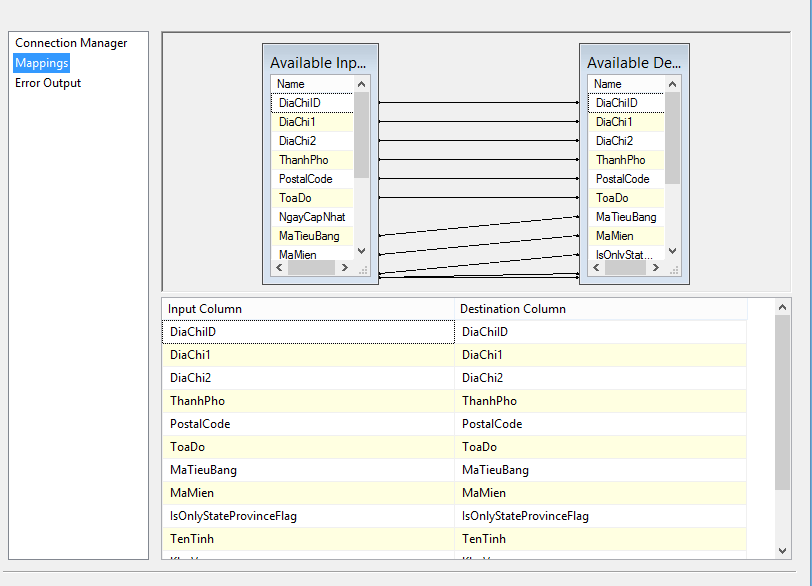
* + - * + Đổ dữ liệu từng bảng từ Source sang bảng Stage tương ứng. Sử dụng OLE DB Source lấy dữ liệu ở Source. Ở đây, chỉ lấy những dòng dữ liệu có ngày cập nhật > LSET và < CET. Lấy vd bảng Address các bảng còn lại tương tự:



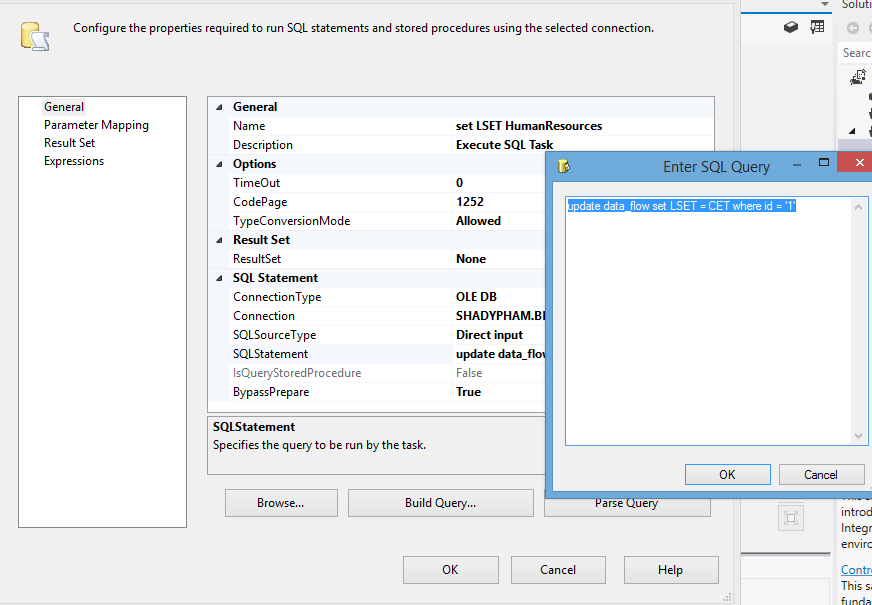
(các dấu ? được thiết lập trong parameters, được lưu LSET và CET ở trên)

* + - * + Ta dùng OLE DB Destination cho việc lấy dữ liệu từ OLE DB Source. Chọn connection đến Stage rồi chọn bảng tương ứng. Sau đó, ta tiến hành mapping từng cột thuộc tính cho khớp:





* + - * Lưu LSET mới vào Metadata:
        + Thực hiện việc cập nhật LSET = CET cho dòng có id = ‘1’ (HumanResource) của Metadata:



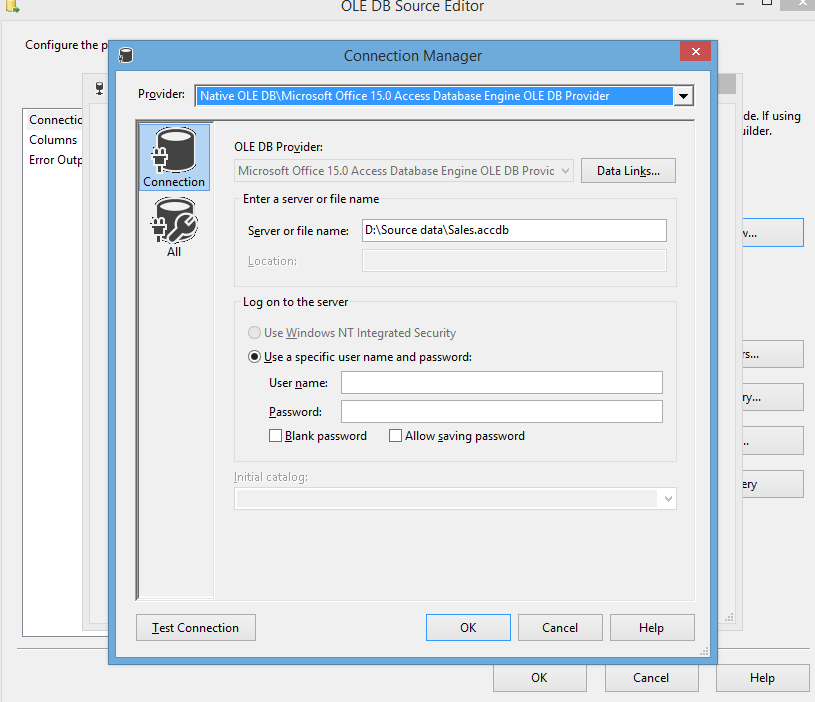
* + 1. Cơ sở dữ liệu Productions :

Vì nơi lưu cơ sở dữ liệu giống nhau (đều là SQL Server) và cấu trúc bảng không có gì khác nên việc đổ dữ liệu từ Source sang Stage của Productions tương tự với HumanResources.

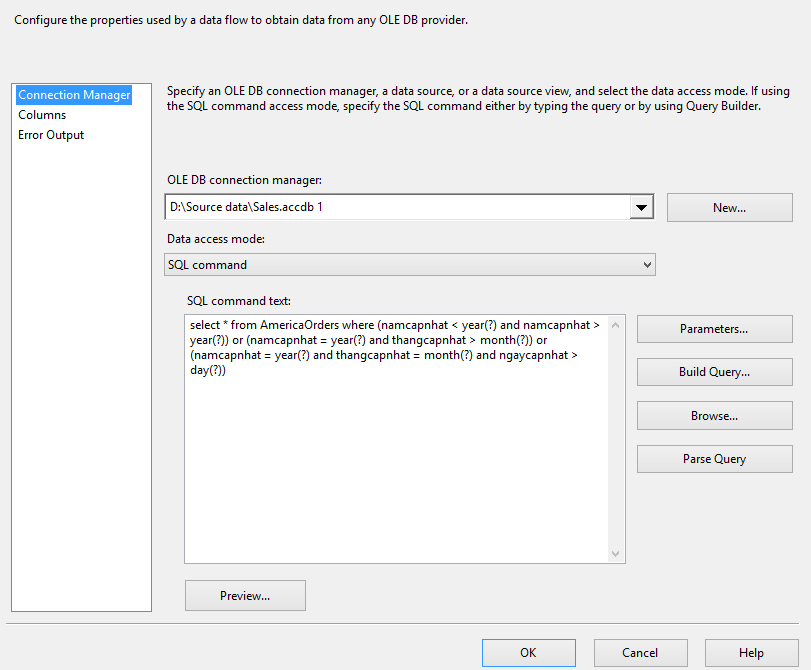
* + - * Lấy dữ liệu LSET, CET từ Metadata:
      * Đổ dữ liệu từ Source: Productions sang Stage
      * Lưu LSET mới vào Metadata:
    1. Cơ sở dữ liệu Sales:

Cơ sở dữ liệu là MS Access nên sẽ khác đôi chút với 2 cơ sở dữ liệu trên

* + - * Lấy dữ liệu LSET, CET từ Metadata:
        + Tương tự HumanResources
      * Đổ dữ liệu từ Source: Sales sang Stage
        + Ở đây, cơ sở dữ liệu được lưu tại MS Access nên việc thiết lập connection khác với SQL Server. Ở Provider: ta chọn MS Office 15.0 Access Database Engine. Phần Server or filename ta điền đường dẫn của file access:



* + - * + Ở bảng AmericaOrders có đôi chút khác biệt với các bảng còn lại, thay vì chỉ có 1 ngaycapnhat thì chia 3 thuộc tính ngaycapnhat, thangcapnhat, namcapnhat:



* + - * + Các bảng còn lại đều có ngaycapnhat nên xử lý tương tự HumanResources bên trên. Cách tạo OLE DB Destination cũng như việc mapping dữ liệu cũng tương tự.
      * Lưu LSET mới vào Metadata:

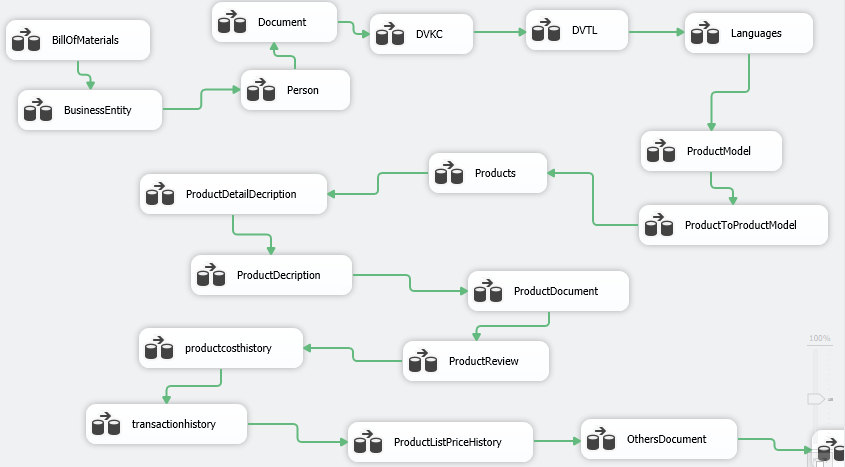
Tương tự HumanResources.

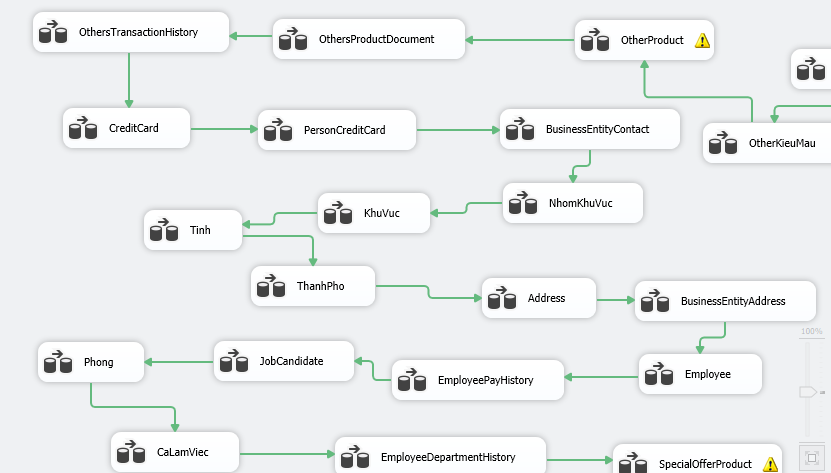
* + - * 1. Stage sang NDS

1. Sơ lược, giải thích Package:

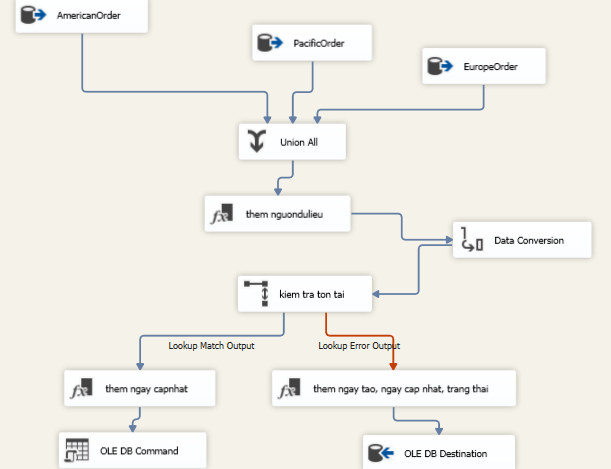
* Chỉ có 1 Package duy nhất đổ các bảng từ stage sang nds.
* Một bảng trong stage sẽ tương ứng với một data flow được tạo
* Thứ tự thực hiện các data flow theo ràng buộc khoá chính, khoá ngoại.
* Package chỉ hoàn tất khi tất cả các bảng từ stage được đổ sang nds theo đúng ràng buộc
* Data flow sau được đổ khi data flow trước đã hoàn thành
* Các tool của SSIS được sử dụng trong phần này gồm:
  + Dataflow Task: tool cần thiết để đổ dữ liệu của một bảng.
  + OLE DB Source: lấy nguồn dữ liệu (ở đây là stage)
  + OLE DB Destination: đổ dữ liệu vào bảng mới (ở đây là nds).
  + OLE DB Command: cập nhật dữ liệu (nếu dữ liệu tồn tại sẵn trong NDS).
  + Devired Column: thêm các cột dữ liệu như ngày cập nhật, nguồn dữ liệu, trạng thái...
  + Lookup: sử dụng để tìm kiểm các mã (kiểm tra tồn tại, tìm kiếm khoá ngoại mới...)
  + Union All: Hợp dữ liệu của nhiều bảng lại với nhau

1. Thứ tự thực hiện và các phép biến đổi sử dụng:
   * Control Flow:
     + Như đã mô tả, thứ tự thực hiện của Control Flow là các bảng treo sẽ được đổ trước, các bảng có tham chiếu khoá ngoại sẽ được đổ sau (tuỳ thuộc vào khoá ngoại)
     + Hình ảnh:





* + Data Flow:
    - Dùng OLE DB Source để kết nối đến bảng tương ứng trong stage.
    - Tiến hành thêm cột nguồn dữ liệu
    - Nếu nhiều bảng gộp lại sẽ phải sử dụng Union All để hợp dữ liệu.
    - Nếu một số thuộc tính trong bảng stage không phù hợp với nds tiến hành Convert.
    - Tiến hành tìm kiếm thuộc tính khoá ngoại mới tồn tại trong nds (nếu có).
    - Kiểm tra tồn tại của dòng dữ liệu đó trong nds : nếu tồn tại tiến hành thêm ngày cập nhật và update lại dòng đó trong nds, nếu không tồn tại sẽ thêm các cột ngày tạo, ngày cập nhật, trạng thái cho dòng dữ liệu đó.
    - Hình ảnh minh hoạ (bảng HinhThucGiaoHang trong NDS):



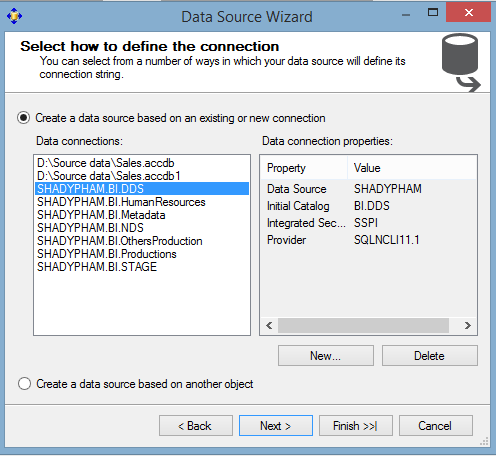
* + - * 1. NDS sang DDS
        2. Giải thích Package:
        + tương tự như phần Stage sang NDS. Tuy nhiên ở đây, thêm một số bảng cần thiết cho việc báo cáo như: ngày, tháng, năm, Sales... và cũng lọc bỏ những bảng không cần thiết cho nhu cầu phân tích.
        + Vì sử dụng DDS để phân tích và build cube cho nên loại bỏ các khoá ngoại có giá trị null của các dòng dữ liệu từ NDS đổ vào.
        1. Thứ tự thực hiện và các phép biến đổi sử dụng:

Tool và cách thức thực hiện tương tự như NDS.

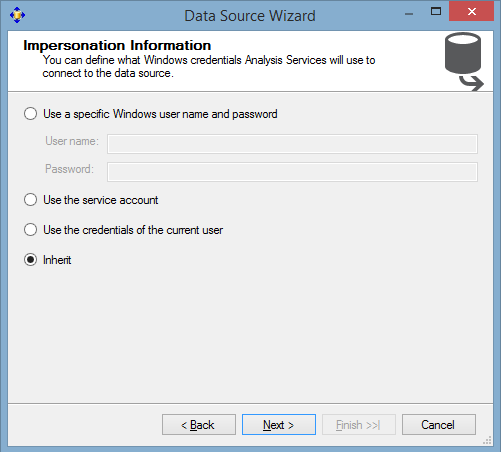
* 1. Tạo cube, thiết kế mô hình đa chiều phục vụ nhu cầu phân tích

Sử dụng mô hình chòm sao để tạo bảng fact và dimension.

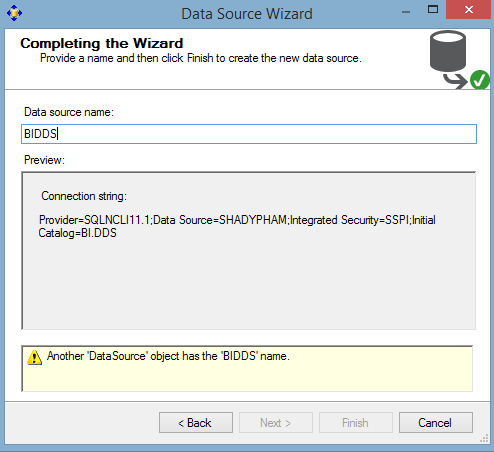
* + - * 1. Tạo data source
* Chọn connection đến DDS đã thiết kế trước đó, bấm next:



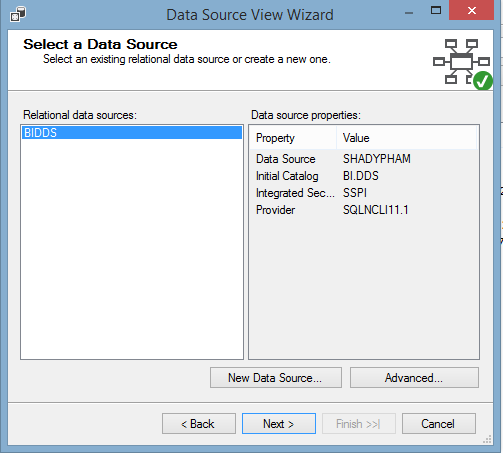
* Chọn Inherit(kế thừa), bước này không quan trọng lắm, chỉ là cách thức để định nghĩa server cho data source, bấm next:



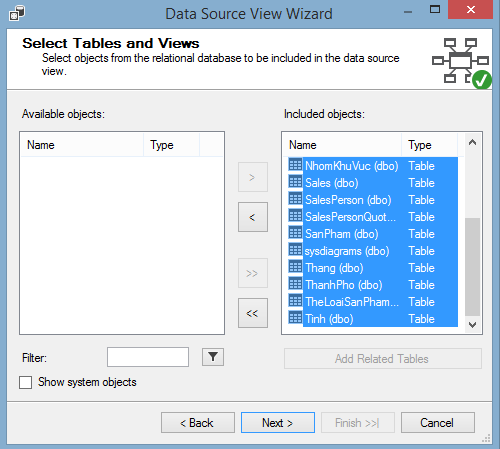
* Đặt tên cho Data Source:



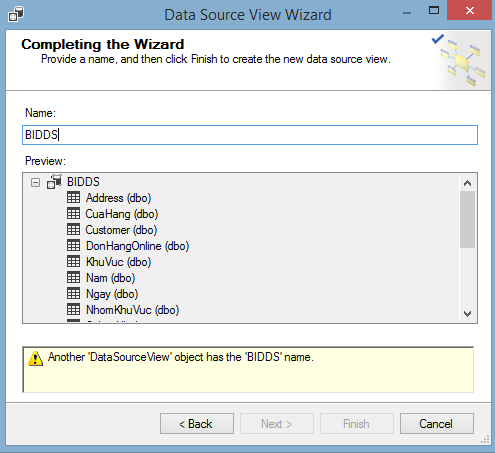
* + - * 1. Tạo data source view
* Chọn data source đã tạo bên trên:



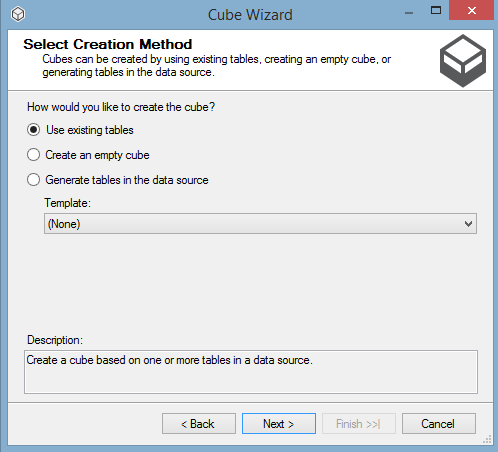
* Chọn các bảng đã tạo trong DDS để tiến hành tạo fact & demension:



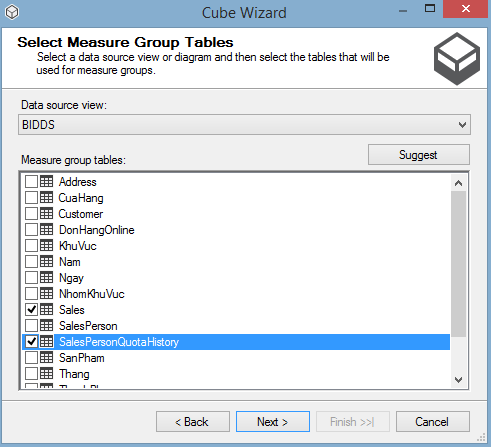
* Đặt tên cho data source view:



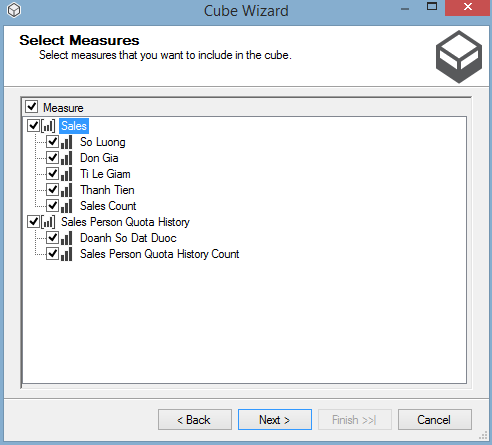
* + - * 1. Tạo cube
      * Chọn các bảng đã tạo trước đó:



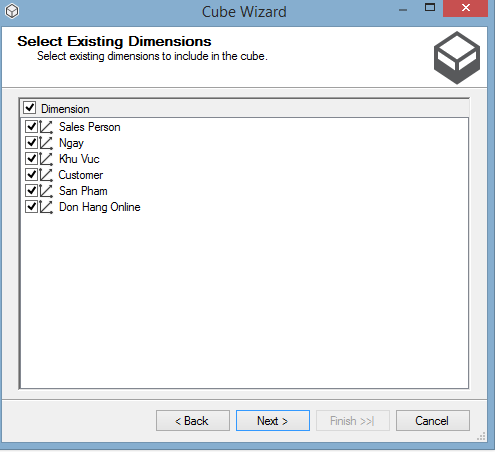
* + - * Chọn bảng Sales và bảng SalesPersonQuotaHistory làm fact: ở đây vì nhu cầu phân tích dựa trên doanh thu cho nên, 2 bảng Sales và SalesPersonQuotaHistory sẽ được đặt làm fact, Sales doanh thu liên quan đến bán hàng sản phẩm, cửa hàng... và SalesPersonQuotaHistory liên quan đến doanh thu, tiền hoa hồng... của nhân viên



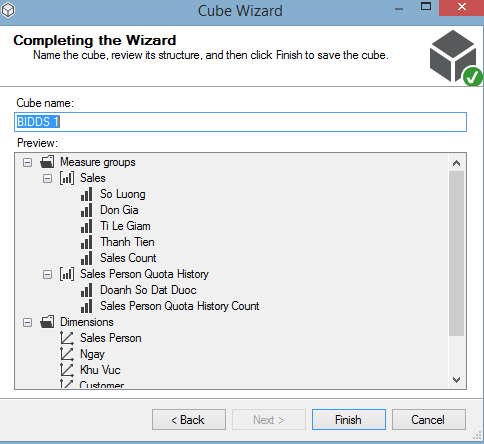
* + - * Xác nhận các độ đo (measures) sử dụng:



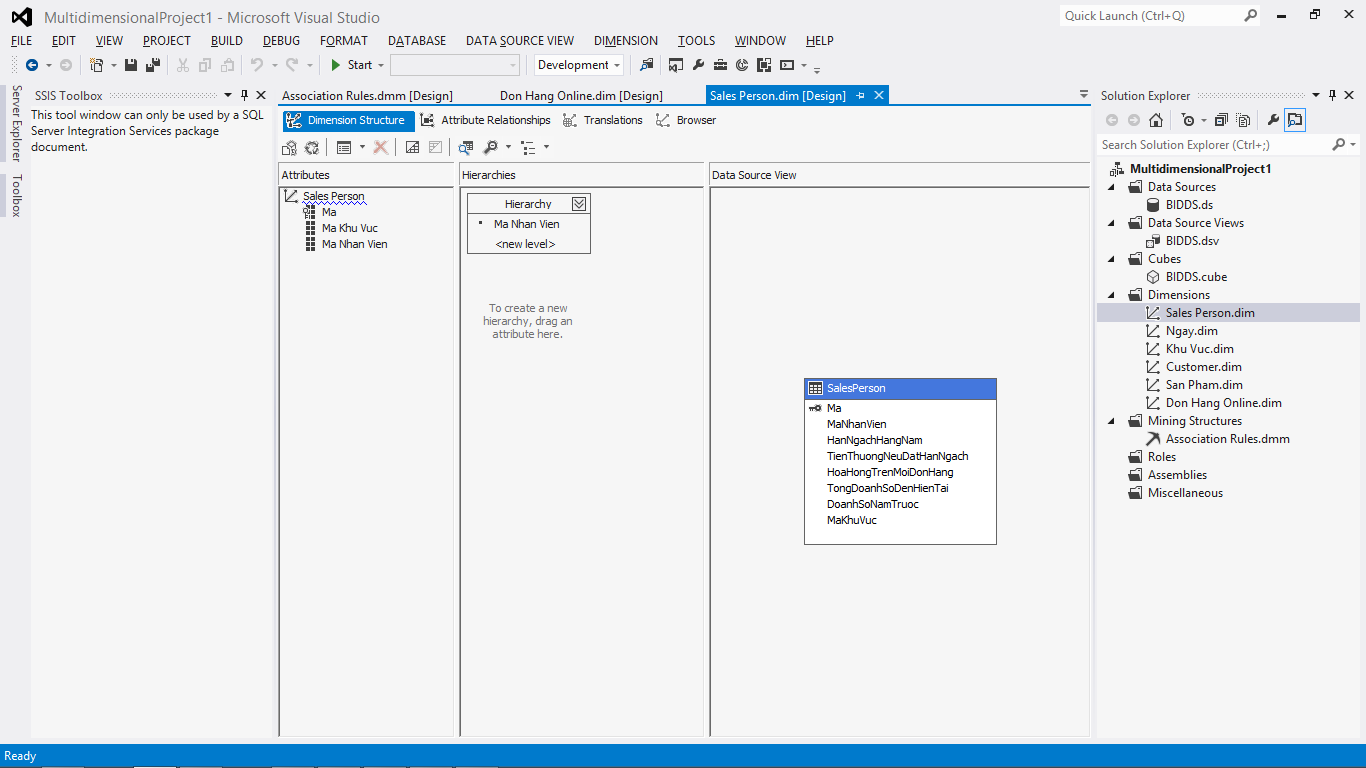
* + - * Xác nhận các dimensions sử dụng:



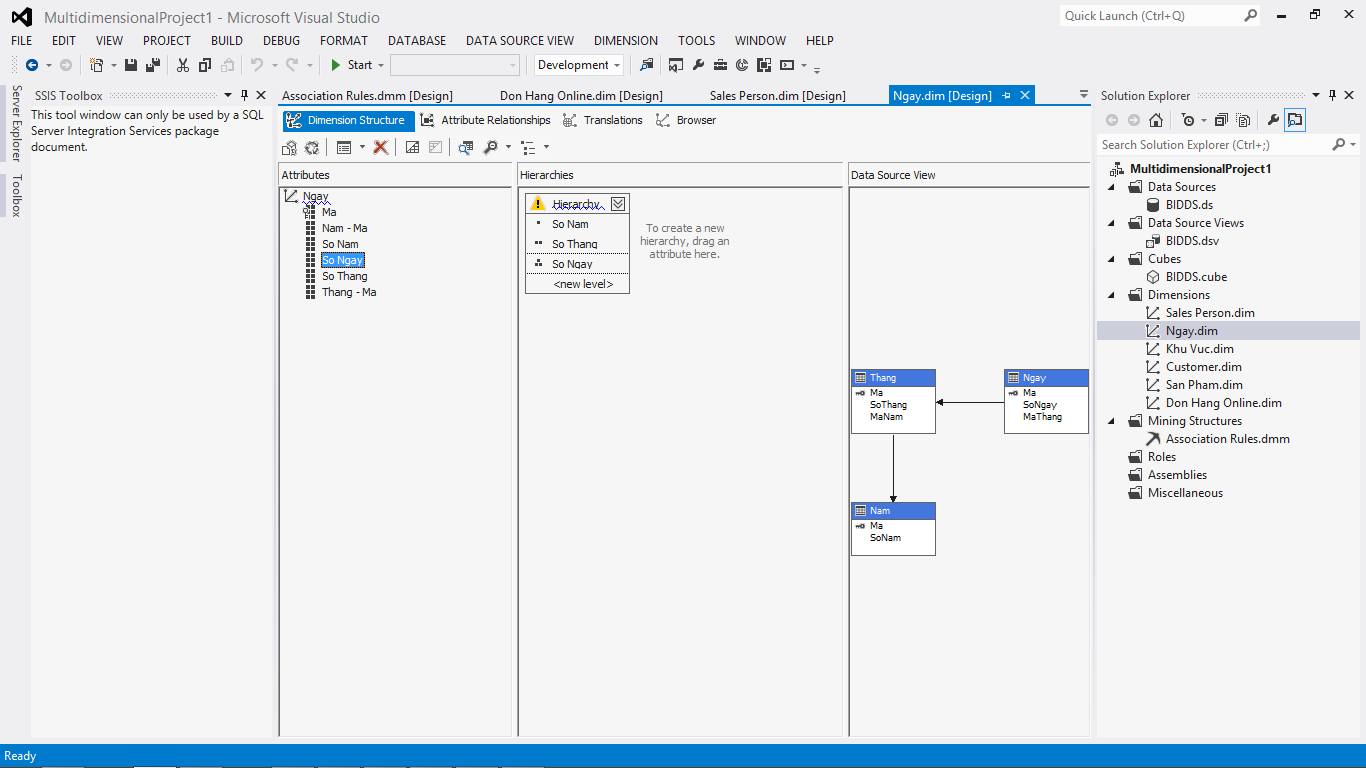
* + - * Đặt tên cho cube:



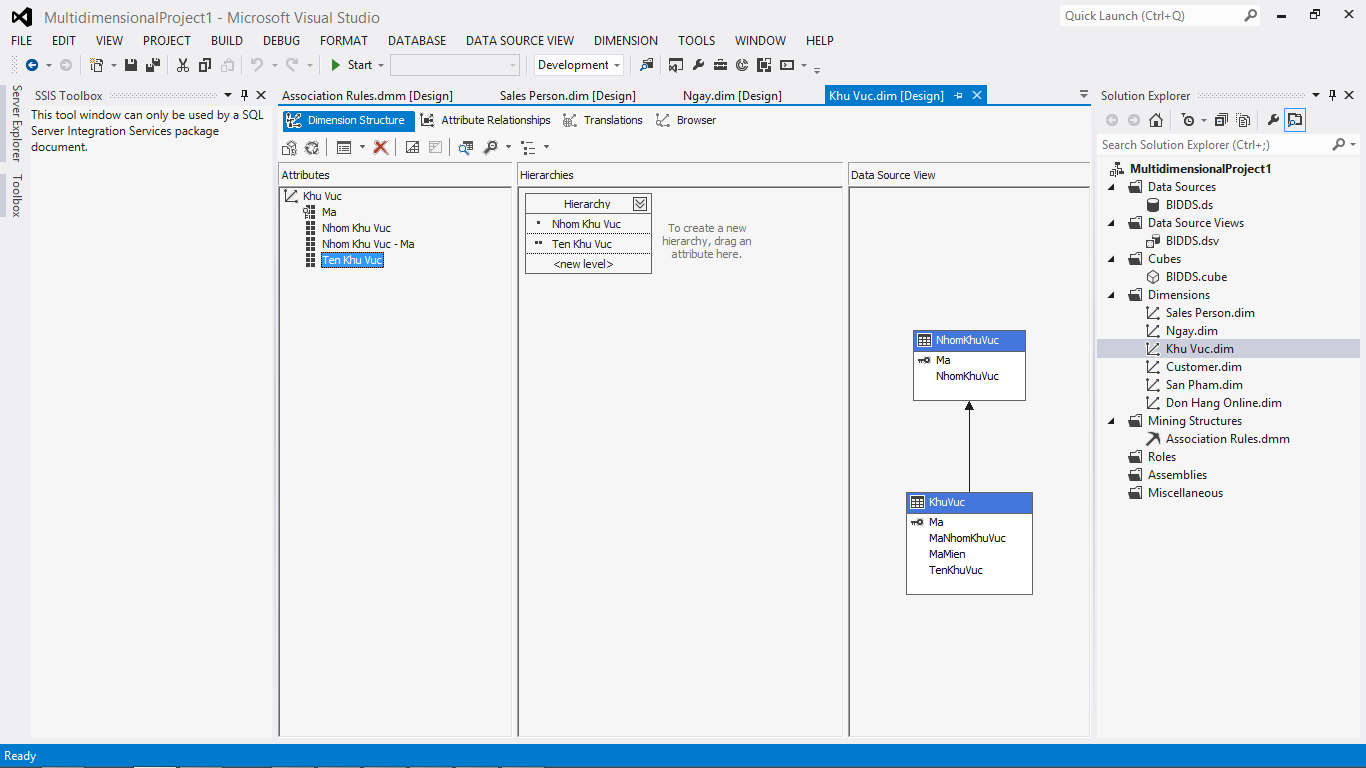
* + - * 1. Tạo chiều cho các bảng Dimension của cube:
      * ở bảng SalesPerson:



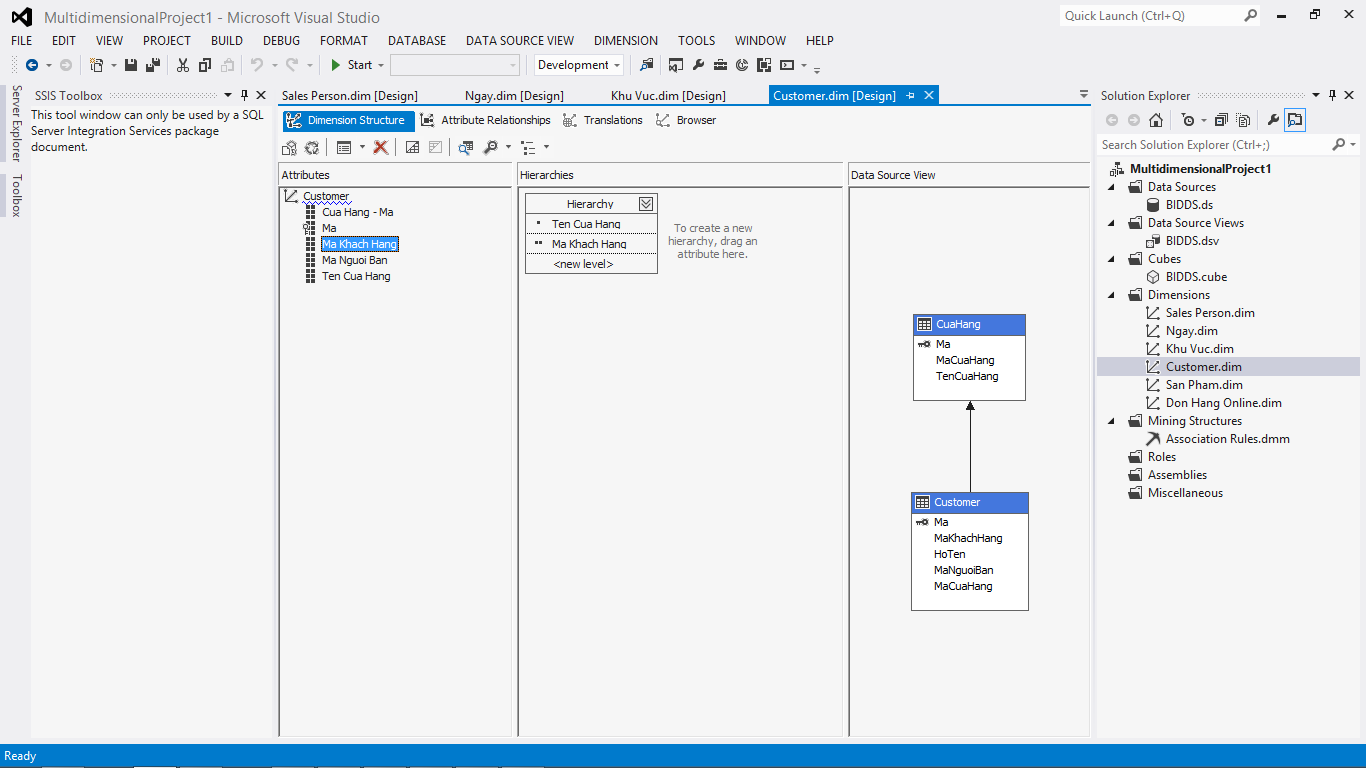
* + - * ở bảng Ngay:



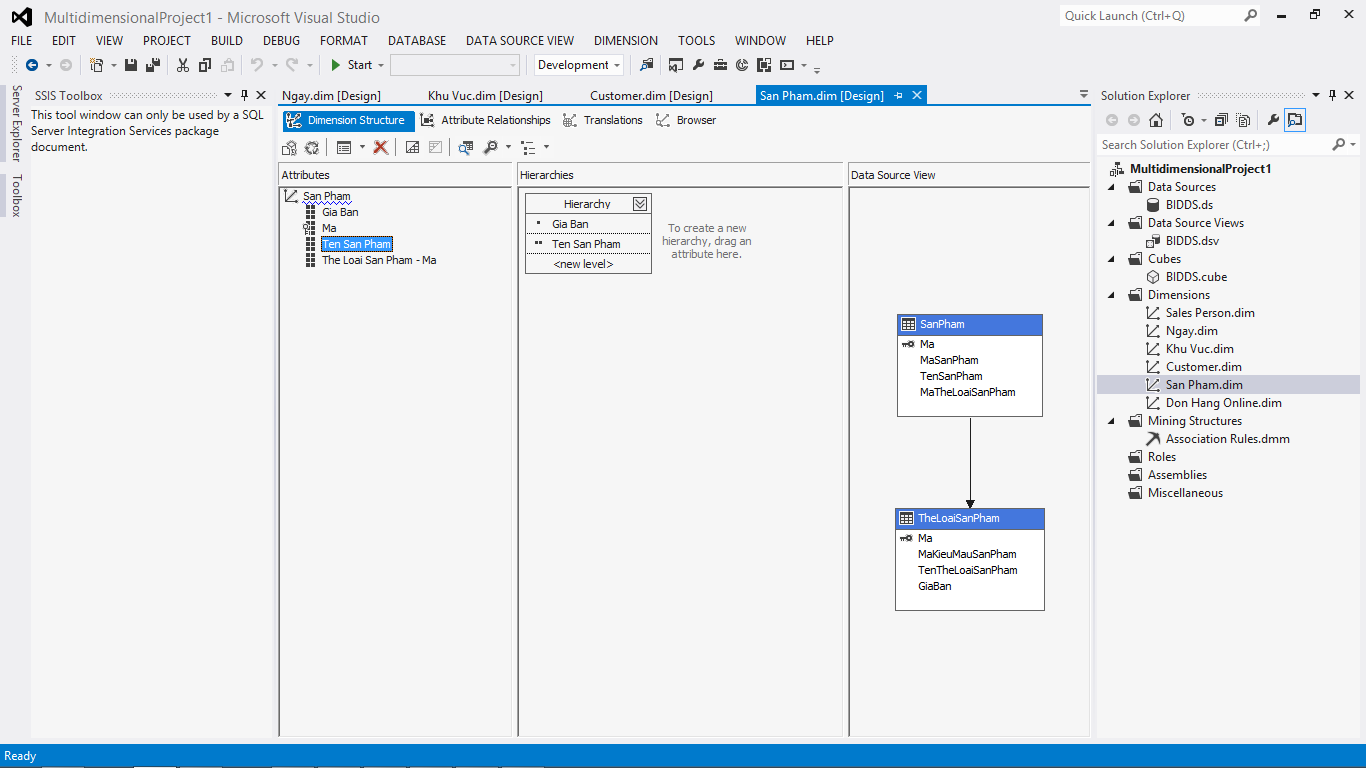
* + - * ở bảng KhuVuc:



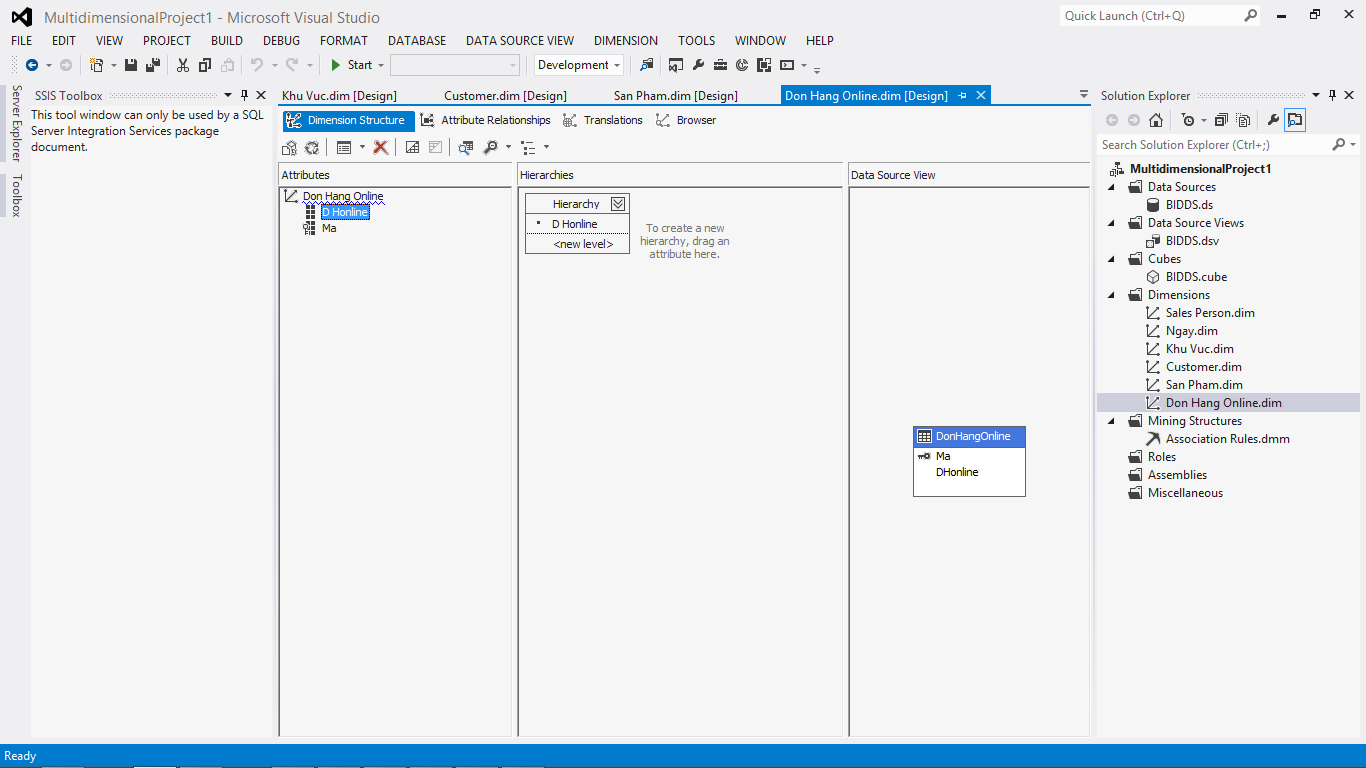
* + - * ở bảng Customer:



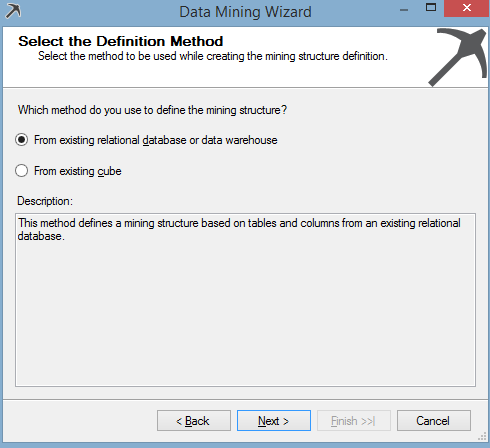
* + - * ở bảng SanPham:



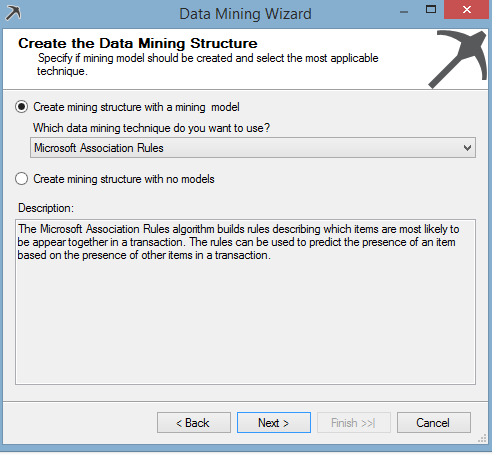
* + - * ở bảng DonHangOnline:



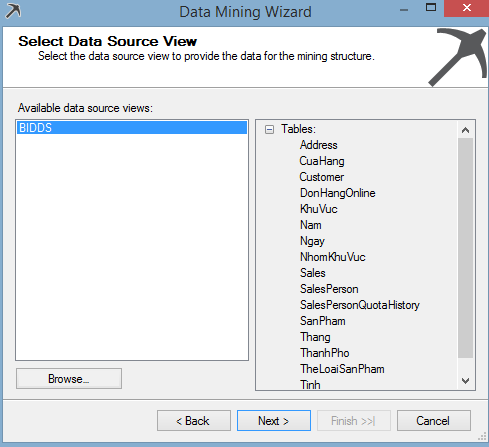
* 1. Thực hiện các phép OLAP truy vấn dữ liệu cube với MDX, SSAS
  2. Khai Thác Dữ Liệu từ Cube phục vụ các nhu cầu truy vấn phân tích trên.
     + - 1. Chọn lựa thuật toán
       - Sử dụng thuật toán Association Rules(luật kết hợp) để dự đoán xu hướng của một khách hàng tại một khu vực, thời điểm cụ thể thường hay mua sản phẩm gì để có thể giới thiệu và quảng cáo.
       - Ta mining trên bảng Sales:
         * Dựa vào các mã customer, mã khu vực, mã ngày, mã người bán ta có thể sinh ra các luật kết hợp với 2 độ đo probability (tần số xuất hiện) và độ đo Importance (tính đúng đắn)
  3. Cài đặt thử nghiệm
     + - Ta tạo mới một mining structure, sử dụng cơ sở dữ liệu đã build bên trên:



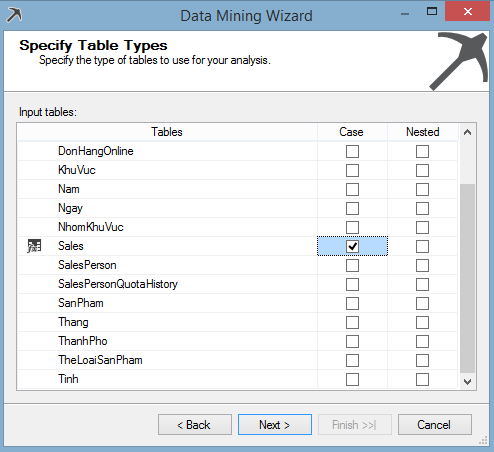
* + - * Chọn lựa thuật toán có sẵn của MS, ở đây ta chọn Association Rules(Phát sinh luật kết hợp):



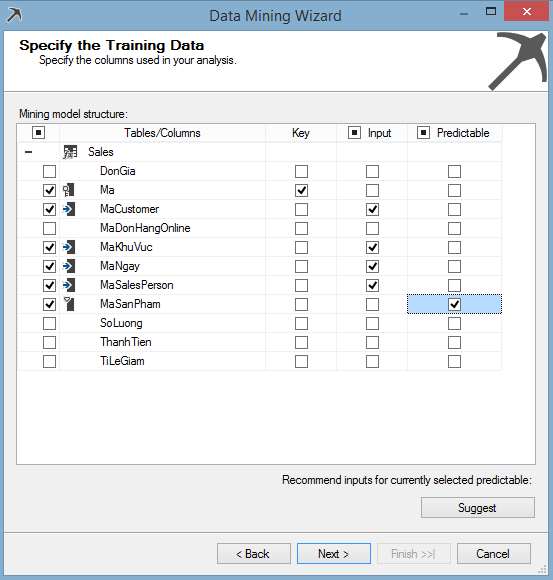
* + - * Chọn data source view:



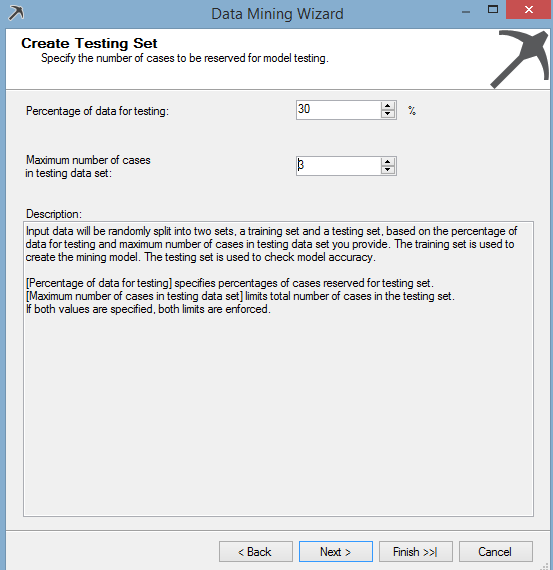
* + - * Chọn lựa bảng để mining, ở đây ta chọn bảng Sales:



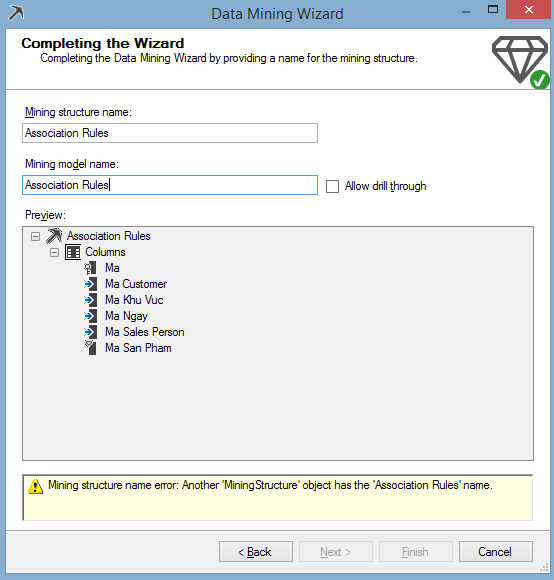
* + - * Chọn key và các giá trị input, predictable(thuộc tính cần sinh luật).



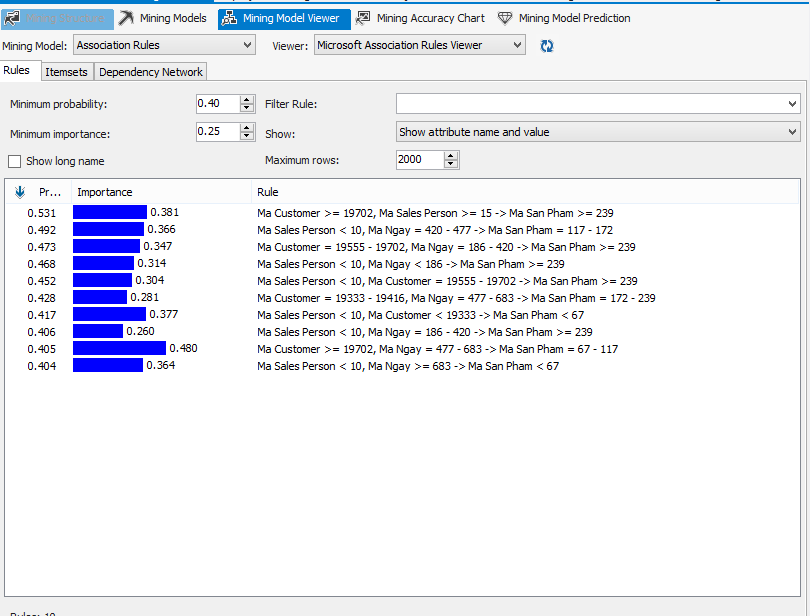
* + - * Lựa chọn độ đo:



* + - * Đặt tên, hoàn tất:



* + - * Sau khi tạo Mining tructure ta tiến hành deploy lại Project.
  1. Kết quả



* + - * ở đây 2 độ đo được sử dụng lần lượt là 0.4 và 0.25
      * như vậy ta thu được các luật kết hợp bên trên.
      * Từ luật kết hợp đó, ta có thể dự đoán được khách hàng tiềm năng hay thời gian tại khu vực đó sản phẩm nào được bán chạy. Vd: đối với những khách hàng có mã lớn hơn 19702 thì cứ vào ngày 477 – 683 sẽ mua sản phẩm có mã từ 67 – 117.

1. **MỨC ĐỘ HOÀN THÀNH**
   1. Đánh giá mức độ hoàn thành:
      * Thiết kế kho dữ liệu
        + Thiết kế Stage hoàn thành tốt, đúng như mục tiêu đề ra. Đánh giá: 100%.
        + Thiết kế NDS tương đối tốt, tránh được các trùng lắp thông tin, đạt dạng chuẩn cao nhất có thể. Đánh giá: 90%.
        + Thiết kế DDS phục vụ được tất cả các nhu cầu phân tích được nêu trong bài. Đánh giá: 90%.
      * Thực hiện ETL nạp dữ liệu định kì vào kho

* Các bảng từ Source đều được đổ chính xác vào Stage, tuy nhiên chỉ mới áp dụng update cho một số bảng, tuỳ các bảng chưa có phần update dữ liệu có thể báo lỗi khi chạy package do trùng lại mã. Đánh giá: 80%.
* Quá trình tách bảng, gộp bảng, liên kết khoá ngoại đều được thực hiện chính xác. Tuy nhiên ở một số bảng có nhiều khoá ngoại, do có thể có nhiều trường hợp null xảy ra cho nên vẫn chưa xử lý hết được các trường hợp đó. Đánh giá: 90%.
* Quá trình NDS sang DDS hoàn thành tốt. Đánh giá: 100%.
  + - Thiết Kế Mô Hình đa chiều phục vụ nhu cầu phân tích
      * Quá trình tạo các data source view, cube, dimension,... không gặp lỗi gì ngoài ý muốn. Đánh giá: 100%.
    - Thực hiện các phép OLAP truy vấn dữ liệu với MDX, SSAS
      * Ở phần này, do không kịp thời gian em chỉ viết các câu truy vấn MDX 3,4,5 trong yêu cầu. Đánh giá: 60%.
    - Khai Thác Dữ Liệu từ Cube phục vụ nhu cầu phân tích
      * Tạo Mining Structure không gặp bất cứ lỗi nào ngoài ý muốn. Có thể các luật sinh ra chưa được cụ thể vì chỉ sử dụng Mã làm đặc trưng(có thể thay thế bằng tên). Đánh giá:90%.

1. **KHÓ KHĂN VÀ HẠN CHẾ**
   1. Chưa rành lắm các tool trong SSIS.
   2. Chưa phân quyền người dùng trên cube.
   3. Chưa áp dụng được chiều thay đổi chậm trong NDS, do phát sinh nhiều lỗi ngoài ý muốn nên quyết định bỏ.
   4. Mất nhiều thời gian ở phần thiết kế DDS và tạo Cube.
2. **NGUỒN THAM KHẢO**

Ngôn ngữ MDX:

<http://msdn.microsoft.com/en-us/library/ms145970.aspx>

tạo Cube với SSAS:

<http://www.codeproject.com/Articles/658912/Create-First-OLAP-Cube-in-SQL-Server-Analysis-Serv>

<http://stackoverflow.com/questions/11722920/ssas-cube-errors-not-deploying>

<http://www.sqlservercentral.com/Forums/Topic1032175-17-1.aspx>