**Phần I: Các ETL trong dự án Hà Tiên**

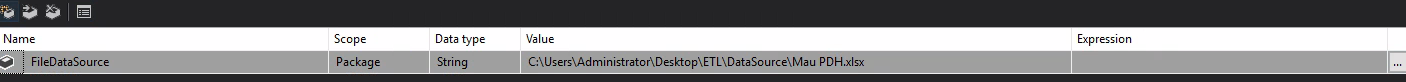
1. ETL nhà máy Bình Phước
   1. ETL update excel Nhật Trình Công Đoạn Xi Măng (C:\Users\Administrator\Desktop\ETL\ETL\ETL\Package.dtsx)
2. Nhiệm vụ

* Lấy file excel từ nguồn [\\172.21.0.106\Data\CHATLUONG\2\_BAN THANH PHAM\BAN THANH PHAM.xlsx](file://172.21.0.106/Data/CHATLUONG/2_BAN%20THANH%20PHAM/BAN%20THANH%20PHAM.xlsx)
* Đọc dữ liệu từ form báo cáo mẫu ở Mau\_PDH trong file excel này
* Tạo các bảng tương ứng bên database để chứa dữ liệu đã lấy được

1. Các biến được sử dụng

* ExcelDSource: chứa đường dẫn file excel sau khi copy về

Tạo biến chứ ExcelSource:

 Name: FileDataSource

Scope:Package

DataType: String

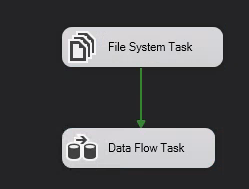
Value: C:\Users\Administrator\Desktop\ETL\DataSource\Mau PDH.xlsx

1. Xây dừng mô hình

* Yêu cầu của dữ liệu đầu vào như sau:

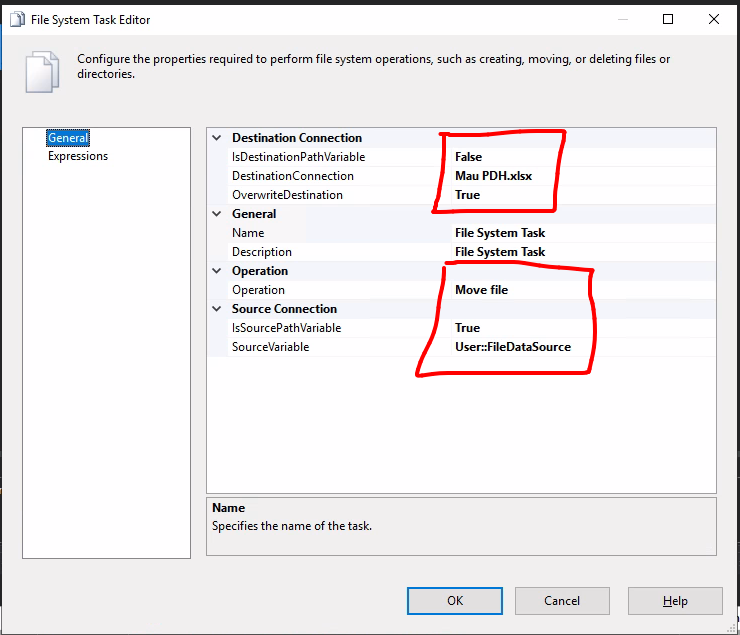
+ Đường dẫn và tên của file excel là không đổi

+ Tên các Sheet trong file excel là cố định

* Dựa vào các yêu cầu trên ta xây dựng ETL thành 3 khối:

+ Khối 1 (File System task): Move file từ nguồn về folder trên máy và ghi đè lên file sẵn có

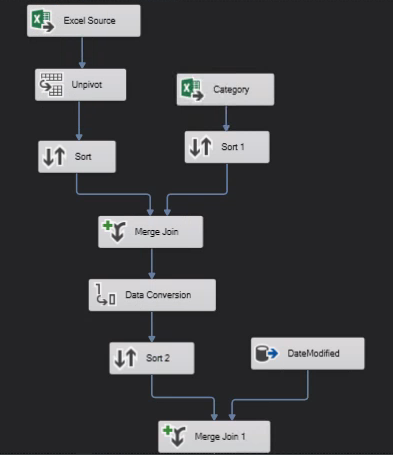
Các thông số cần thiết lập:

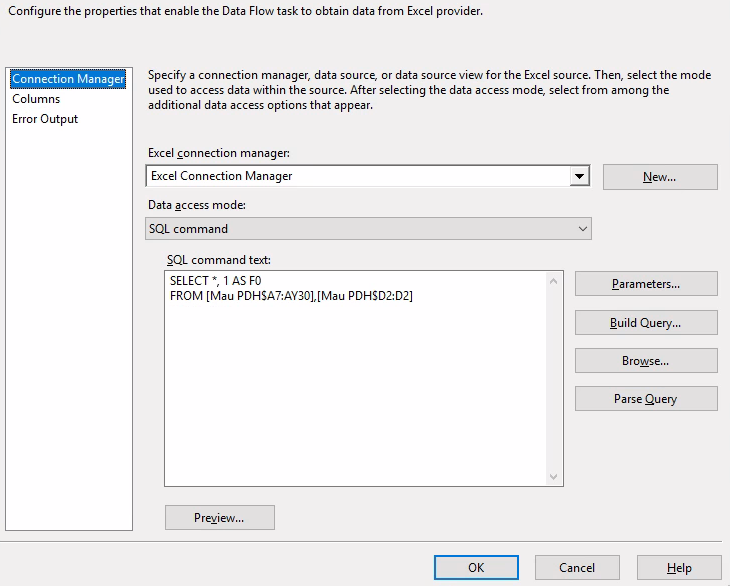
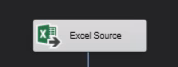


* Chọn Operation là Move File để thực hiện thao tác Move (Ghi đè lên file trước)
* Chọn Source Connection là biến chứa đường dẫn của file excel nguồn (ExcelSource) đã define trước đó
* Chọn Destination Connection là Đường dẫn đến file cố định theo một form có sẵn để tránh xảy ra sai sót khi khởi chạy ETL

+ Khối 2 (Data Flow Task): Thực hiện thao tác lấy dữ liệu báo cáo được chỉnh sửa, thực hiện xử lý dữ liệu để so sánh sự tồn tại để thực hiện insert hay update dữ liệu. Bên trong khối 2 chứ nhiều khối nhỏ thực hiện xử lý dữ liệu.

Chia thành 2 giai đoạn thực hiện;

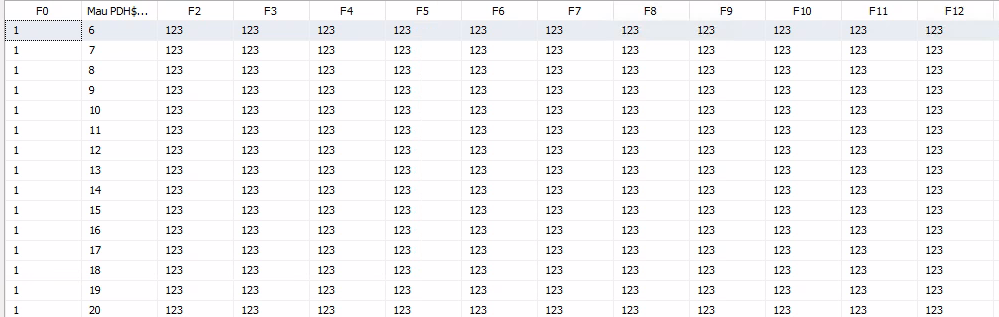
* Giai đoạn 1: Xử lý dữ liệu để phù hợp lưu trữ ở Database

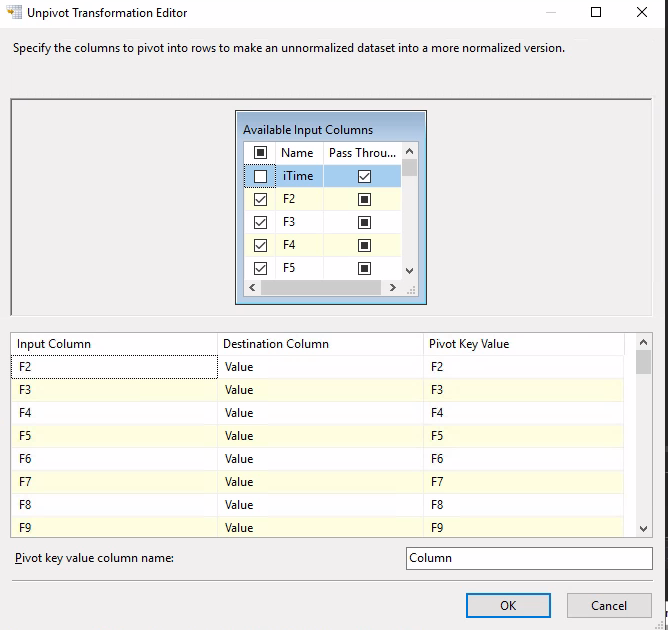
Khối 1: Excel Source hoặc Category lấy dữ liệu từ file Mau PDH.xlsx ở sheet Mau PDH

Lấy dữ liệu bảng tương ứng với form báo cáo và thời gian của báo cáo đó. Chỉnh sửa Tên cột để dễ xử lý nhận biết. **F + số** là giá trị theo từng TagName mỗi giờ trong ngày( ngoại lệ: F0 là cột tạo giá trị 1 để lấy UserNameModiFied), iTime giờ trong ngày, dtDate ngày của báo cáo.







Khối 2: Unpivot là lất bảng thực hiện chuyển đổi dữ liệu sang dạng quy định

Các tùy chỉnh cần lưu ý:

Input column: giá trị đầu vào

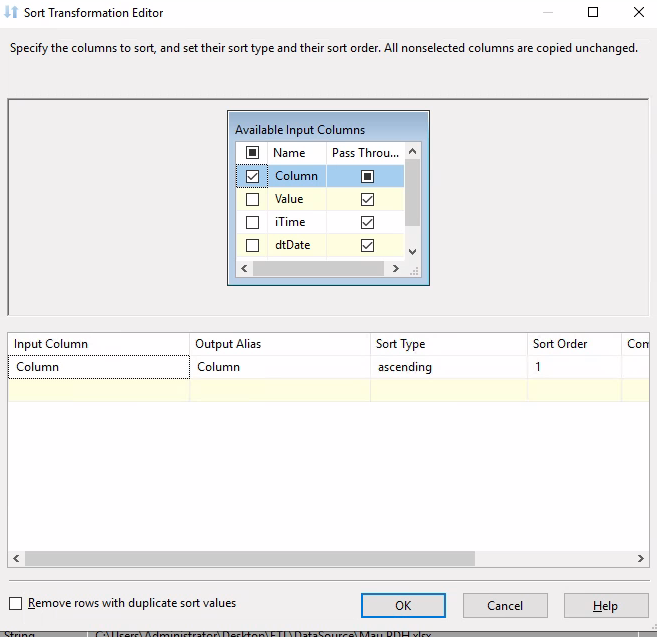
Destination column: giá trị đầu ra có thể tuyển biến tên cho cột mới

Pivot Key Value mặc định lấy tên cột trước khi unpivot.

 Để unpivot click chọn ô

Khối 3: gồm các khối Sort, Sort 1, Sort 2 các khối thực hiện công việc sắp xếp tăng hoặc giảm theo 1 cột được chỉ định:

Tùy chỉnh cần lưu ý:

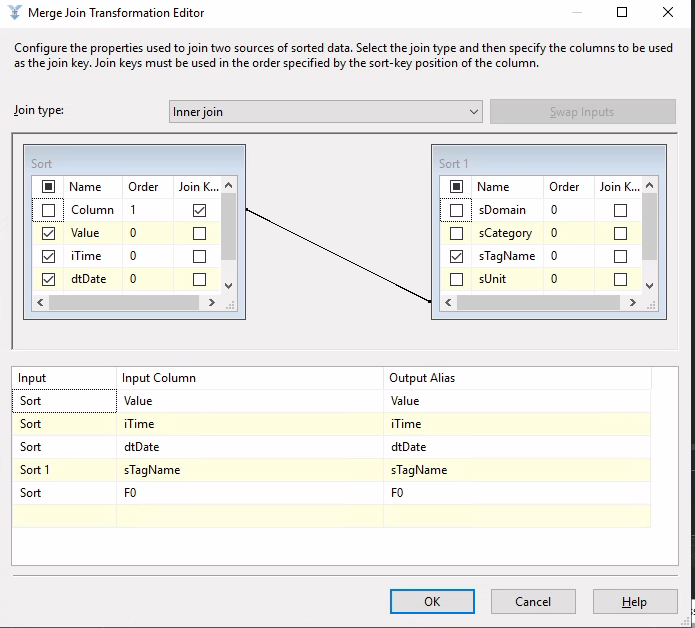


CLICK chọn cột muốn các cột còn lại sắp xếp theo và chọn ascending tăng dần, descending giảm dần.

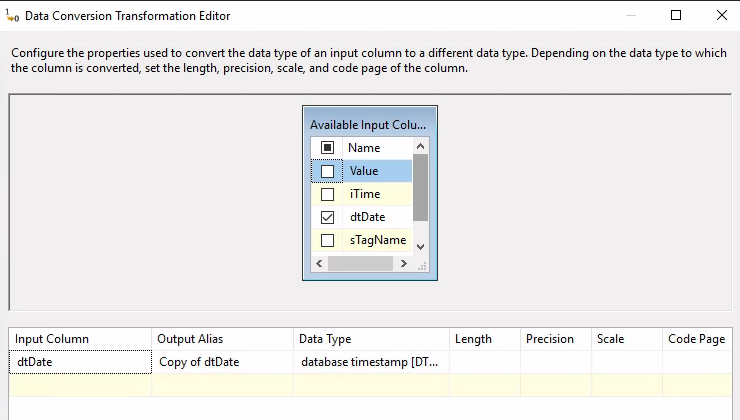
Khối 4: Merge Join, Merge Join 1 thực hiện việc ghép 2 bảng dữ liệu lại dựa trên 1 trường dữ liệu chung.

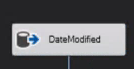
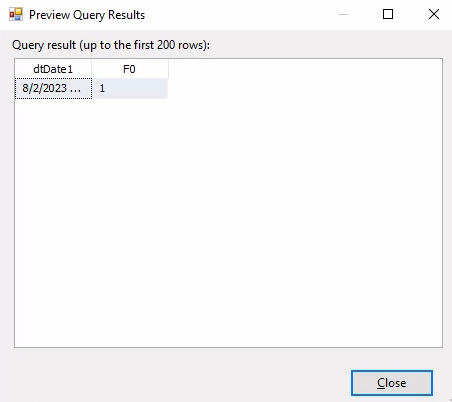
Một số lưu ý:

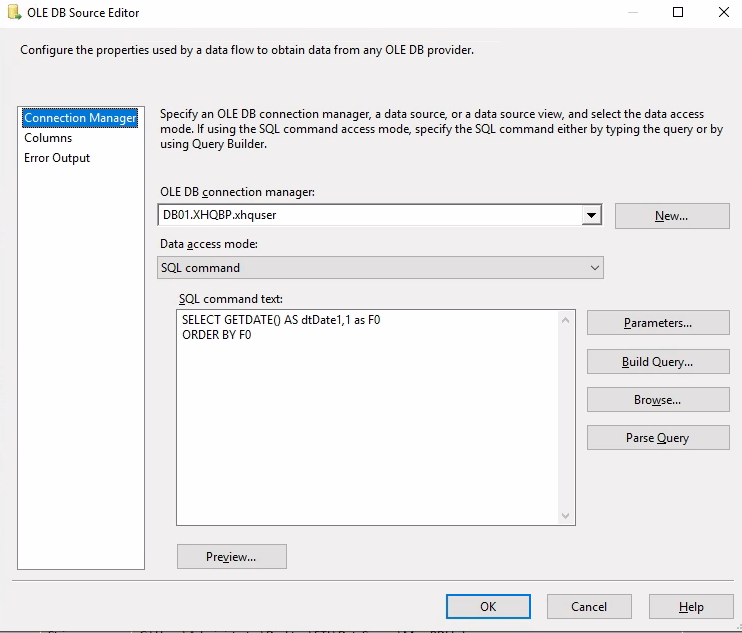
* 2 bảng dữ liệu khi merge join cần được sắp xếp trước
* 2 bảng dữ liệu cần có một trường dữ liệu chung để mapping
* Có thể click chọn các trường dữ liệu muốn lấy ta ở 2 bảng và có thêm đặt tên cột ở Output Alias



Khối 5: Chỉnh sửa, thay đổi kiểu dữ liệu để có thể thực hiện việc kiểm tra so sánh cụ thể ở đây là dtDate

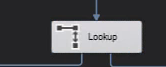


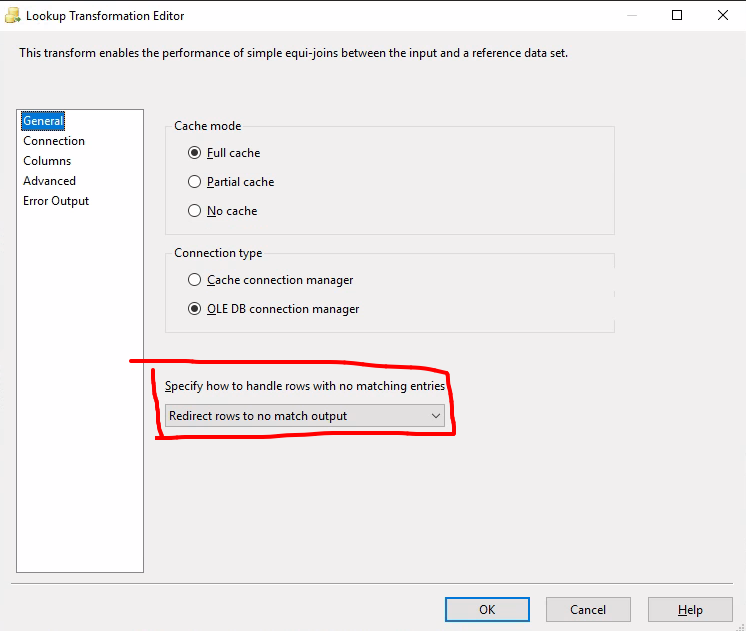
Khối 6: DateModified lấy thời gian thực hiện việc khởi chạy ETL để kiểm soát sự thay đổi.

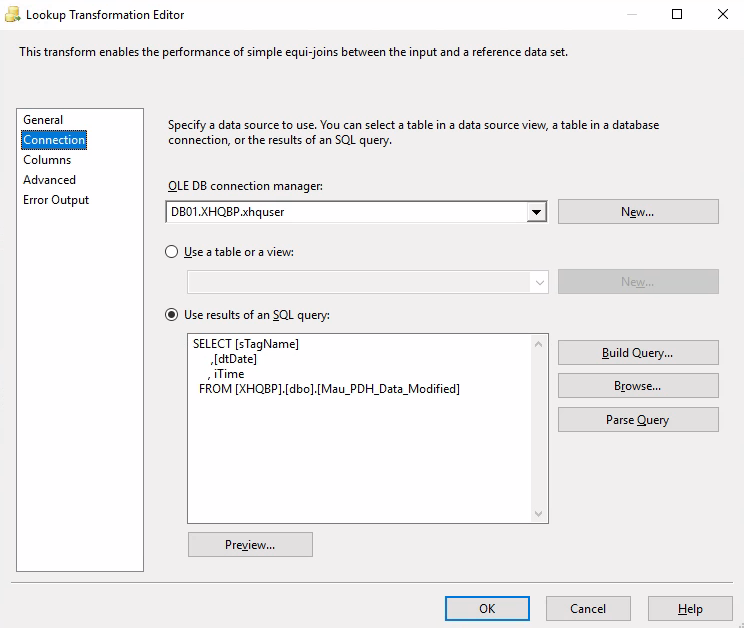


Giai đoạn 2: Kiểm tra sự tồn tại của dữ liệu trong data base để thực hiện Insert hay Update

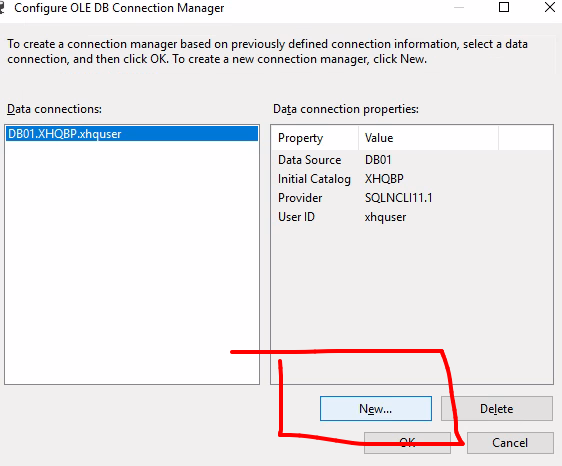
Khối 1: Lookup kiểm tra xuống database

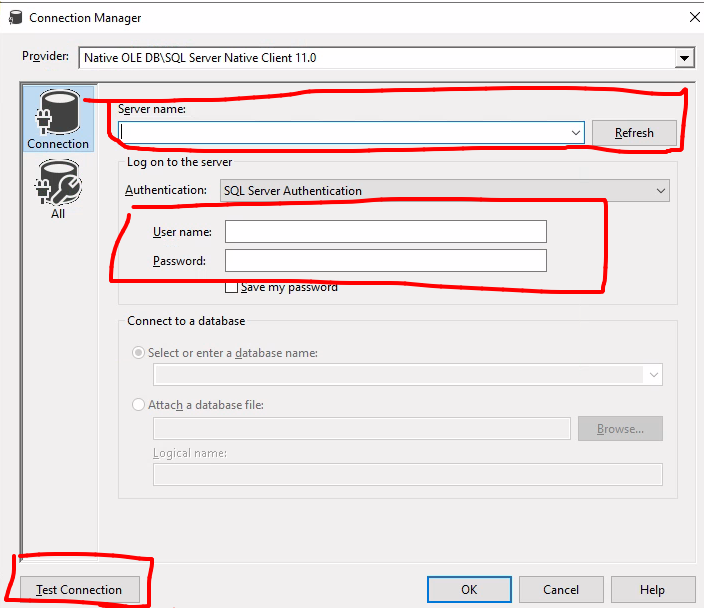
Một số tùy chỉnh cần chú ý:



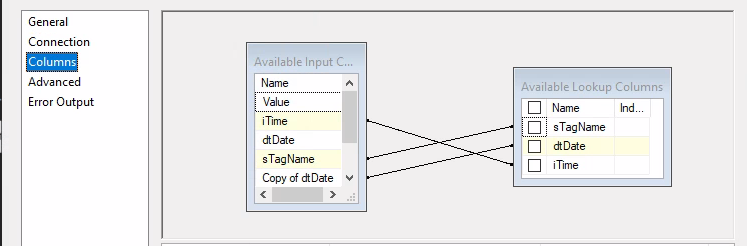
Connetion thực hiện tạo connection tới database chưa bảng cần kiểm tra để thực hiện insert hoặc update.

Click chọn new

Tiếp tục click chọn new

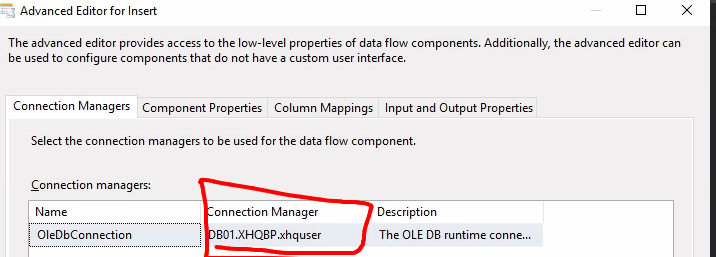


Những thông số cần chú ý:

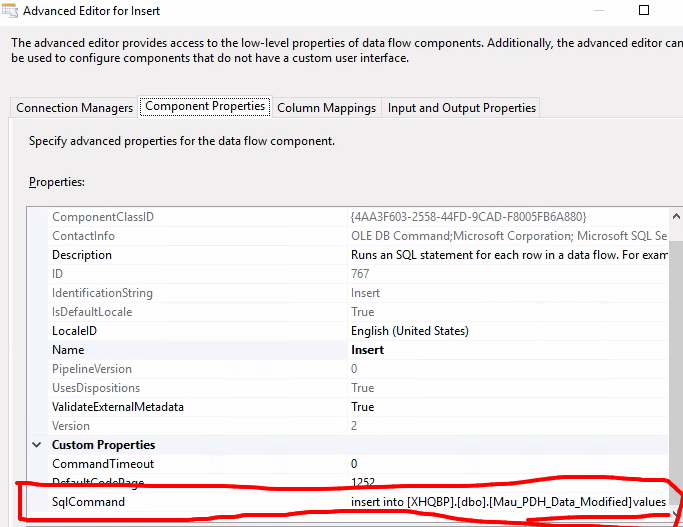
* Server name: server cần kết nối
* Click chọn SQL Server Authentuccation: để cung cấp User name và Password
* Kiểm tra kết nối với Test Connection

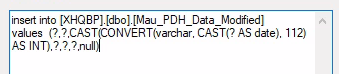
Columns thực hiện việc mapping từ bảng dữ liệu xử lý với table của database theo các trường mapping tương ứng có thể kéo thả đến click chuột phải chọn Edit mapping.

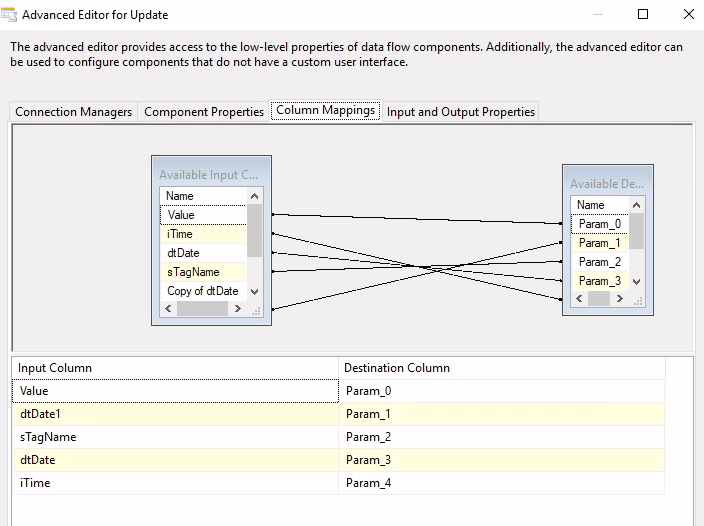
Khối 2: Insert kiểm tra nếu chưa có báo cáo thay đổi của ngày đó thì thêm dữ liệu vào database



Chọn kết nối tới database đã tạo



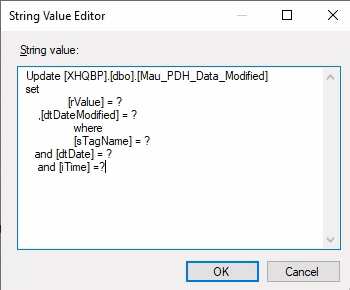
Sqlcomand viết lệnh insert có điều kiện bằng việc kiểm tra sTagName, iDate, iTime như sau;



Tùy chỉnh column Mapping đúng với các tham số truyền vào.

Khối 3: Update kiểm tra nếu có dũ liệu báo cáo thay đổi của ngày đó thì sẽ thực hiện cập nhật dữ liệu

Các phần Connection và column mapping tương tự với Insert.

Sqlcommand viết lệnh Update có điều kiện kiểm tra sTagName, iDate, iTime như sau:

Tổng kết

HẾT