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

1. ETL nhà máy Bình Phước
   1. ETL update excel Bán thành phẩm (Project1; ETL\_Excel\_File\_Ban\_Thanh\_Pham.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 các Sheet trong file excel này
* Tạo các bảng tương ứng bên database để chứa dữ liệu đã lấy được

1. Các biến được sử dụng

* ExcelSource: Chứa đường dẫn của file excel nguồn
* ExcelDestination: chứa đường dẫn file excel sau khi copy về

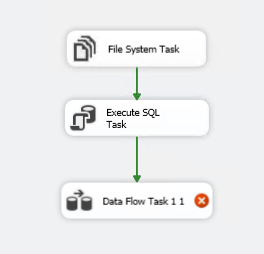
1. Xây dừng mô hình

* Yêu cầu của dữ liệu đầu vào như sau:

+ Đường dẫn và tên của file excel là không đổi

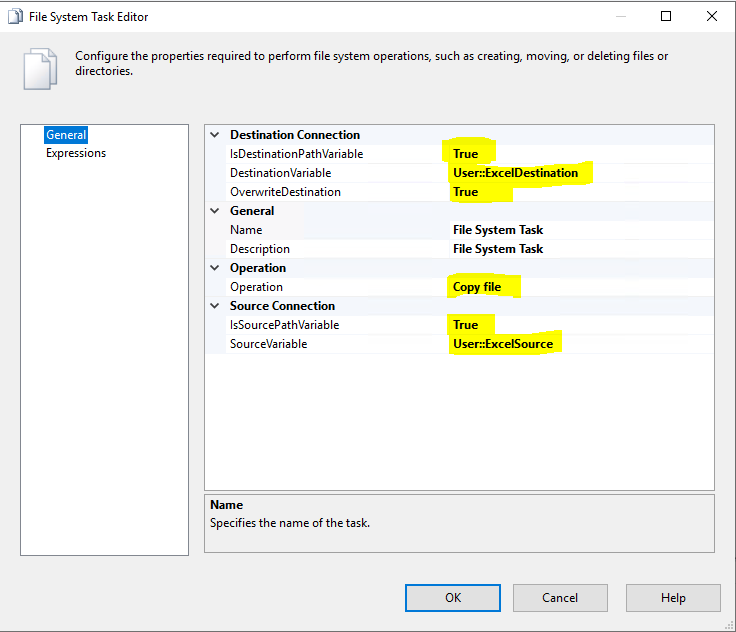
+ Tên các Sheet trong file excel là cố định

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



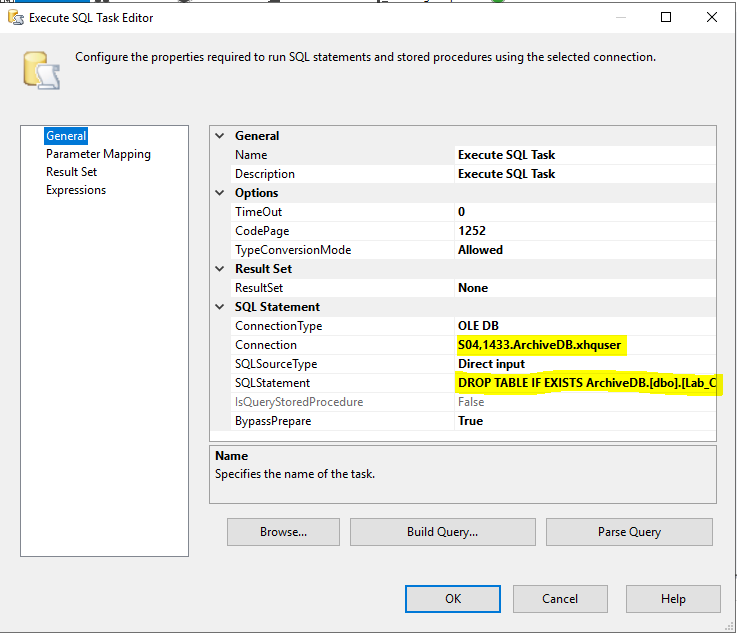
+ Khối 1 (File System task): Copy file từ nguồn về folder trên máy và rename lại thành một tên cố định

Các thông số cần thiết lập:



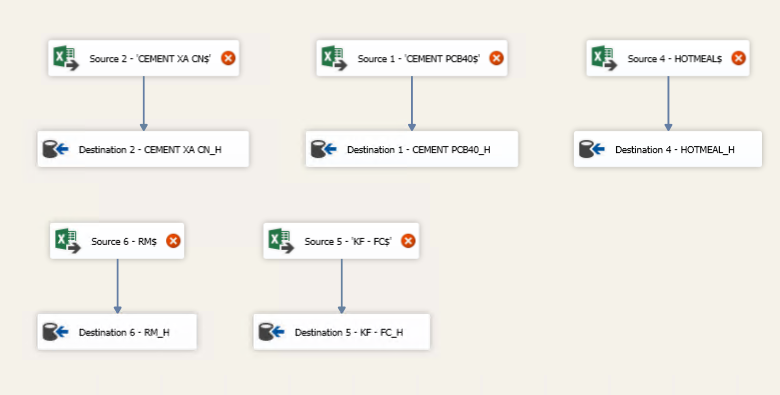
* Chọn Operation là Copy để thực hiện thao tác Copy
* Chọn Source Connection là biến chứa đường dẫn của file excel nguồn (ExcelSource) đã define trước đó
* Chọn Destination Connection là biến chứa đường dẫn của file excel đích, đây cũng chính là file excel chúng ta thao tác trực tiếp để lấy dữ liệu.

+ Khối 2 (Execute SQL Task): Thực hiện thao tác renew lại các bảng chứa giá trị trong SQL server



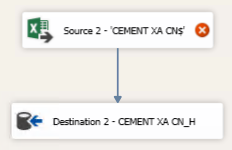
* Phần Connection: tạo connect đến database sử dụng thư viện OLEDB
* Phần SQL Statement: ta sẽ nhập câu lệnh SQL để thực hiện thao tác renew bảng dữ liệu bằng cách Drop bảng chứa dữ liệu và tạo lại bảng đó.

+ Khối 3 (Data Flow Task): Thực hiện đổ dữ liệu từ file excel vào các bảng chứa đã tạo

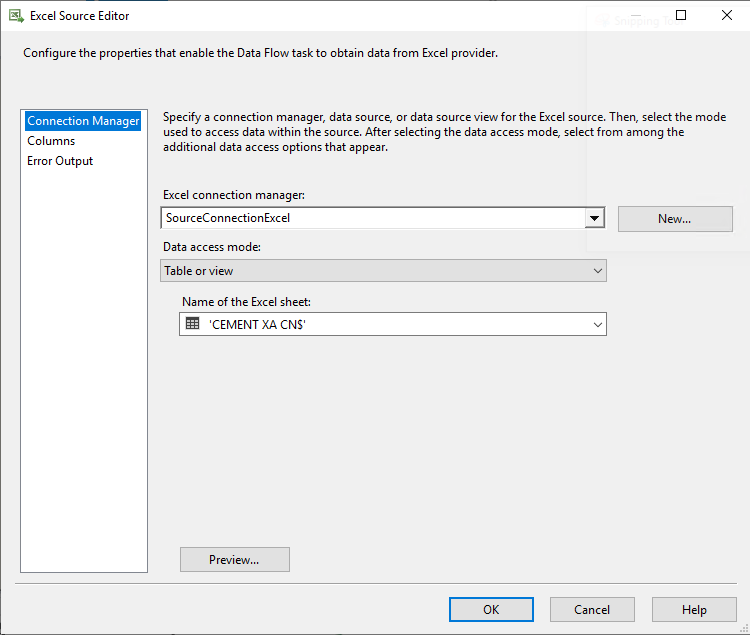


Bên trong ta sẽ tạo các khối để truyền dữ liệu từ excel vào bảng đích.

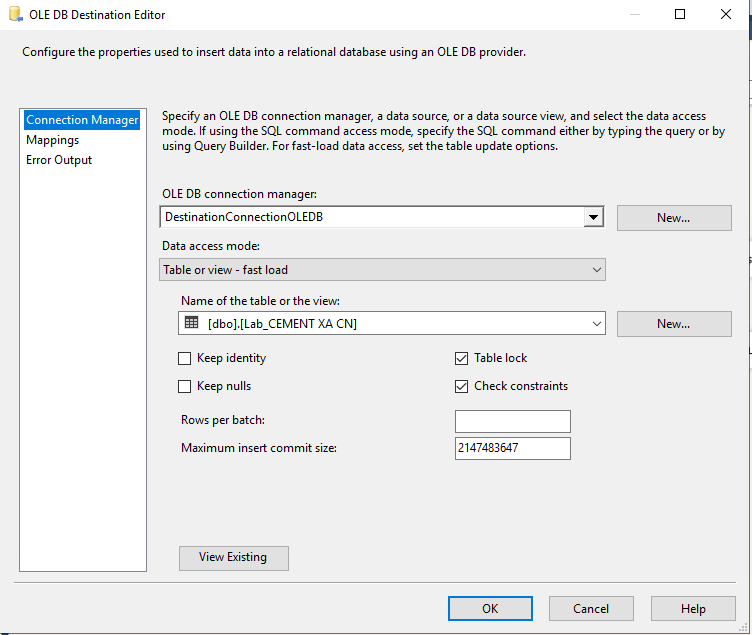
* Đối với mỗi sheet cần lấy dữ liệu, ta sẽ có 2 khối chính



Khối 1: Lấy dữ liệu từ Sheet yêu cầu trong file excel



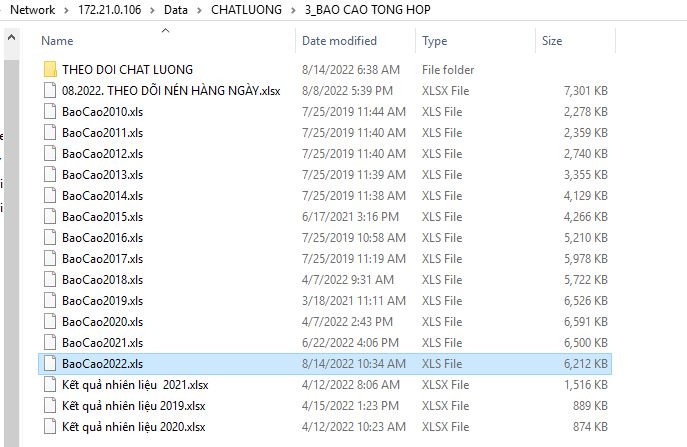
Khối 2: Đẩy dữ liệu đã lấy vào bảng tương ứng



* 1. ETL update excel Báo cáo 2022 (Project1, ETL\_Excel\_File\_Bao\_Cao)

1. Yêu cầu

* Lấy file excel Báo cáo theo đúng năm hiện tại từ đường dẫn \\172.21.0.106\Data\CHATLUONG\3\_BAO CAO TONG HOP

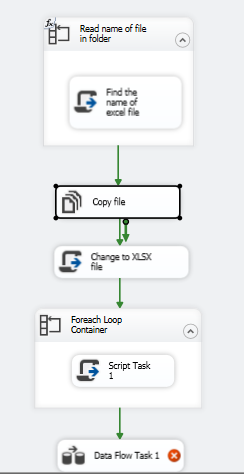


* Đọc dữ liệu từ sheet Monthly\_R và đưa dữ liệu này vào database

1. Các biến được sử dụng

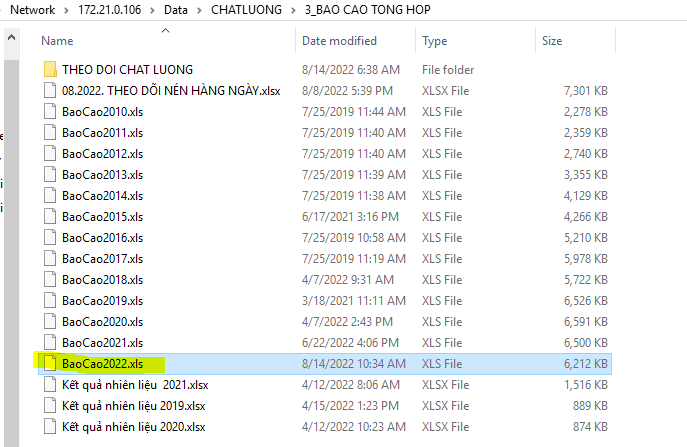
* SheetName: Chứa tên sheet cần đọc dữ liệu
* MainFolder: chứa đường dẫn đến folder tổng chứa dữ liệu nguồn
* FolderDestination: chứa đường dẫn đến folder đích chứa file excel sau khi copy
* FileName: chứa tên của file excel khi copy và rename
* Destination: chứa đường dẫn file excel sau khi copy

1. Xây dừng mô hình



* Theo yêu cầu như trên, ta thấy file excel cần lấy nằm lẫn với rất nhiều file excel, yêu cầu là sẽ lấy đúng file excel có tên chứa năm hiện tại.

VD: năm 2022 ta sẽ lấy file excel



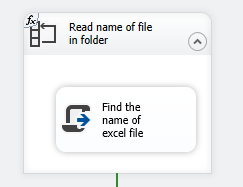
* Theo đó nhiệm vụ đầu tiên sẽ lấy đúng file excel dựa theo tên file theo từng năm:

+ Bước 1: Lấy tên của tất cả các file excel có trong folder cần lấy

+ Bước 2: Cắt chuỗi các tên đó để lấy đuôi là số năm của file excel đó

+ Bước 3: So sánh số năm đó với số năm hiện tại

+ Bước 4: Lấy chính xác file excel yêu cầu



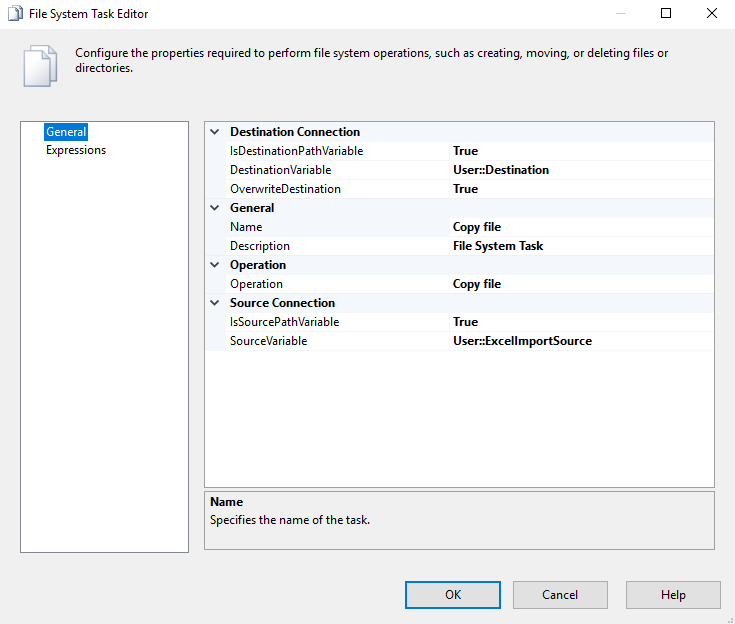
Dùng vòng lặp để lặp qua tất cả các file excel có trong folder nguồn

Khối Script bên trong sẽ có nhiệm vụ cắt tên của file và so sánh với năm hiện tại để quyết định có lấy file excel đó hay không.

Code check tên file



* Sau khi xác định được file excel cần lấy, ta tiến hành copy file về folder cố định và rename lại theo một tên cố định



* Sau khi copy file về máy, ta tiến hành đọc file bằng thư viện OLE DB thì nhận ra file không đọc được và đang ở đuôi là .XLS, ta tiến hành đổi đuôi file thành .XLSX để có thể được được dữ liệu trong file. Để thực hiện điều này ta sẽ tạo một khối script để chuyển đuôi file



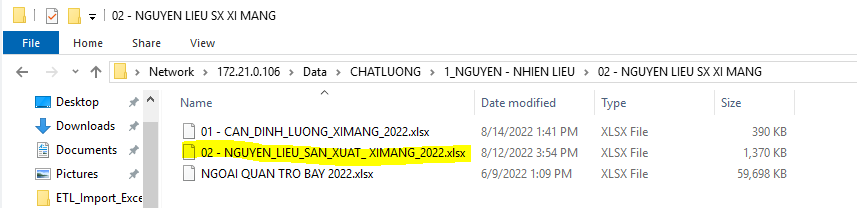
Khối script chuyển đuôi file từ .XLS sang .XLSX



* 1. ETL update độ ẩm xi măng, clinker

1. Yêu cầu

* Lấy file excel dữ liệu xi măng từ đường dẫn \\172.21.0.106\Data\CHATLUONG\1\_NGUYEN - NHIEN LIEU\02 - NGUYEN LIEU SX XI MANG

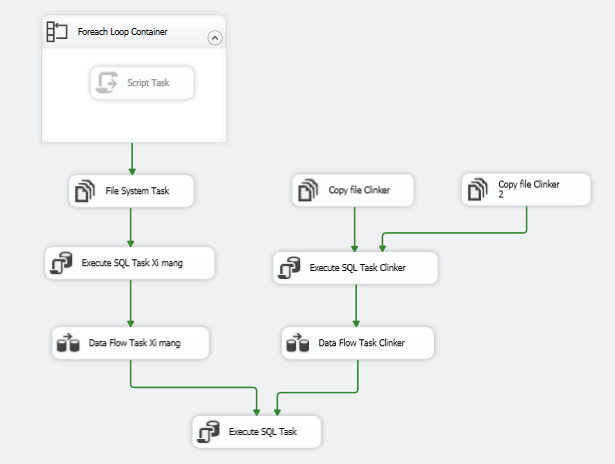


* Đọc dữ liệu trong các sheet của file excel này và đổ vào các bảng tương ứng

1. Các biến được sử dụng

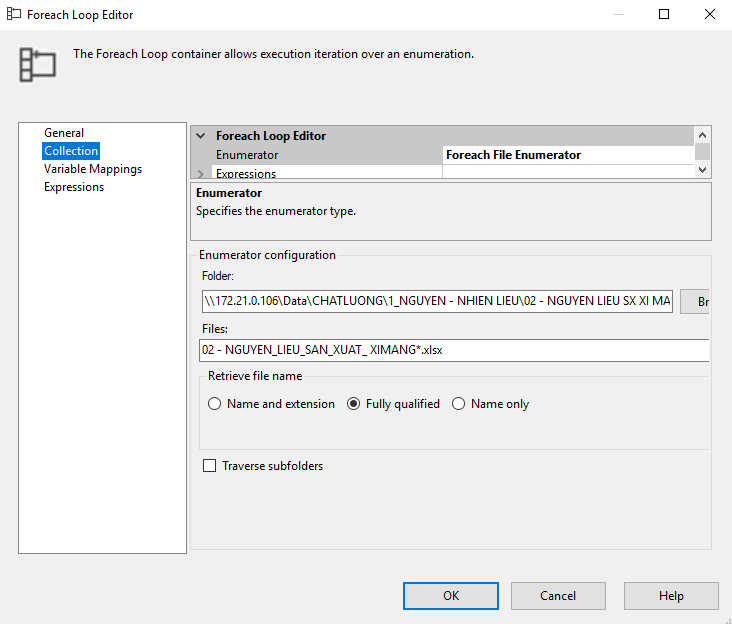
* Source\_Excel\_Clinker: chứa đường dẫn file excel nguồn của Clinker
* Source\_Excel\_Clinker\_2: chứa đường dẫn file excel nguồn thứ 2 của clinker
* File\_excel\_cement\_path: chứa đường dẫn file excel nguồn của xi măng
* Destination\_Excel\_Clinker\_2: chứa đường dẫn file excel đích sau khi copy của clinker
* Destination\_Excel\_Clinker: chứa đường dẫn file excel đích sau khi copy của clinker
* Destination\_Excel\_Cement: chứa đường dẫn file excel đích sau khi copy của xi măng

1. Xây dừng mô hình

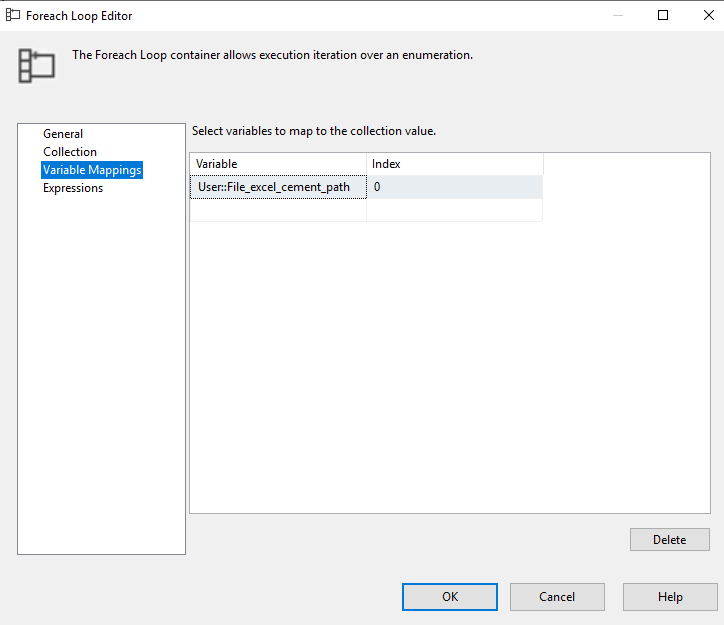


* Mô hình sẽ có 2 nhánh chính:

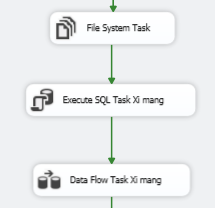
+ Nhánh 1: Copy và update dữ liệu của xưởng xi măng



Tạo vòng lặp để tìm file excel cần tìm

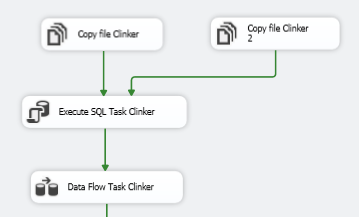


Nếu tìm được sẽ lưu đường dẫn của file đó vào biến

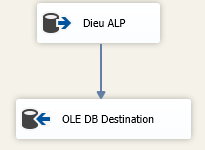


Sau khi đã tìm được file, tiến hành copy file, renew bảng chứa dữ liệu và đổ dữ liệu mới vào

+ Nhánh 2: Copy và update dữ liệu của xưởng Clinker



Copy file excel gốc và đưa về folder cố định đã tạo trong các biến ở trên, sau đó renew bảng chứa dữ liệu và đẩy dữ liệu đọc được vào



Việc đẩy dữ liệu có 2 phần chính: Đọc dữ liệu từ file excel và mapping dữ liệu đó vào bảng đích trong database

* 1. ETL update PXCL (đọc tất cả các sheet trong 1 file excel)

1. Yêu cầu

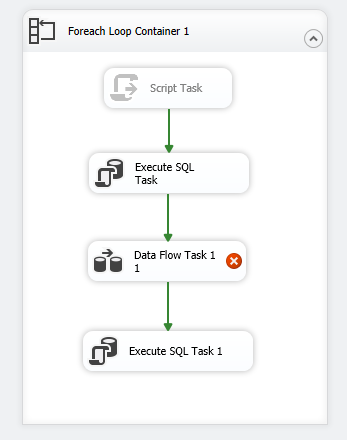
* Đọc file excel có trong folder E:\004\_Databases\Excel PXCL
* Đọc và import từng sheet trong file excel đó vào database, sau đọc lọc dữ liệu tinh và đổ vào một bảng khác

1. Các biến được sử dụng

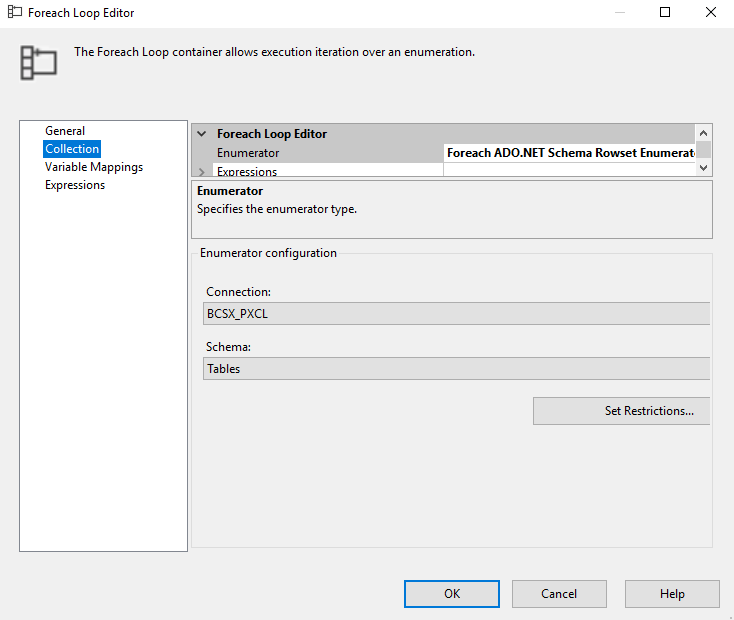
* SQL\_SheetName: chứa câu lệnh query dữ liệu từ các sheet (Biến động)
* SheetName: chứa tên sheet

1. Xây dừng mô hình

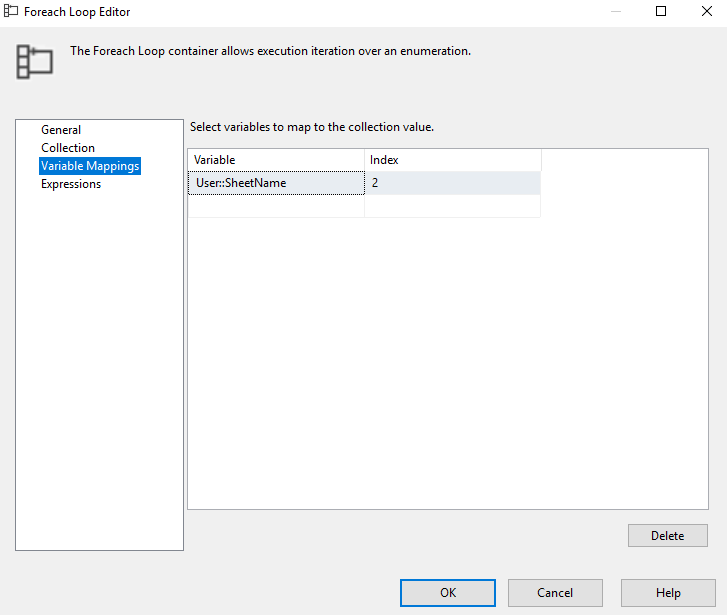
* Theo yêu cầu phải đọc qua toàn bộ các sheet của file excel phải thay đổi giá trị của biến chứa tên sheet để đọc qua toàn bộ sheet



+ Tạo vòng lặp qua các sheet trong file excel

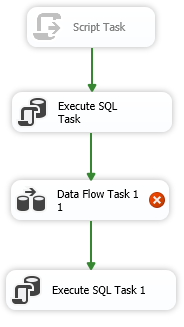


Thông số thiết lập vòng lặp



Với mỗi vòng lặp ta sẽ lưu tên sheet vào biến SheetName đã tạo sẵn.

+ Trong mỗi vòng lặp ta sẽ thực hiện các bước:



B1: In tên của sheet đọc được ra màn hình

B2: Renew bảng dữ liệu trong SQL

B3: Đổ dữ liệu đọc được vào bảng dữ liệu đã renew

B4: Gọi thủ tục xử lý dữ liệu thô thành dữ liệu tinh từ SQL

1. ETL văn phòng chính Hà Tiên (Project2; ETL\_VPC.dtsx)
2. Yêu cầu

* Đọc file excel của từng nhà máy từ folder tổng [\\172.16.0.130\data\BAOCAOSANXUAT\_VICEMHATIEN](file://172.16.0.130/data/BAOCAOSANXUAT_VICEMHATIEN/TNPH/NAM_2022/THANG_07/01072022/TNPH_BCSX_01_07_2022.xlsx)
* Trong mỗi nhà máy sẽ có mỗi folder, vào đúng file excel của từng ngày để lấy dữ liệu của nhà máy

1. Các biến được xử dụng

* ExcelName\_TNPH: Chứa tên của file excel nhà máy Phú Hữu
* ExcelName\_TNLA: Chứa tên của file excel nhà máy Long An
* ExcelName\_TNCR: Chứa tên của file excel nhà máy Cam Ranh
* ExcelName\_NMXMKL: Chứa tên của file excel nhà máy Kiên Lương
* ExcelDestination\_TNPH: Chứa đích đến của file excel nhà máy Phú Hữu sau khi copy
* ExcelDestination\_TNLA: Chứa đích đến của file excel nhà máy Long An sau khi copy
* ExcelDestination\_TNCR: Chứa đích đến của file excel nhà máy Cam Ranh sau khi copy
* ExcelDestination\_NMXMKL: Chứa đích đến của file excel nhà máy Kiên Lương sau khi copy

1. Xây dựng mô hình



* Mỗi nhà máy đều thực hiện qua các bước:

+ Lọc các file excel và lấy đúng file excel đúng tên cần tìm



+ Sau khi đã tìm được file excel cần lấy, ta tiến hành copy file, renew lại bảng chứa dữ liệu, đổ dữ liệu thô từ file excel vào bảng chứa dữ liệu và chạy thủ tục lọc dữ liệu tinh từ dữ liệu vừa đổ.

