TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP TPHCM

**KHOA CÔNG NGHỆ THÔNG TIN**

TP.HCM - Năm 2017

**BÁO CÁO CUỐI KỲ**

**Môn:Phát triển ứng dụng**

**Giáo viên hướng dẫn :**

**Huỳnh Thái Học**

**Đỗ Hà Phương**

**1. Võ Đông Lâm(NT) 15014921**

**2. Nguyễn Công Thịnh 15024631**

IMPORT DATA BY TALEND

**Link GitHub:** <https://github.com/CongThinhNguyen/BaoCaoCuoiKy.git>

**I. Tổng quan:**

**Nhiệm vụ:** Nghiên cứu, tìm hiểu về nhập xuất data và nắm vững chức năng của talend. Thông hiểu được tính ứng dụng thực tế của đồ án và hiện thực áp dụng vào thực tế

**1) Nhiệm vụ của từng thành viên trong nhóm:**

|  |  |  |
| --- | --- | --- |
| Thành viên | Nguyễn Công Thịnh | Võ Đông Lâm |
| Nhiệm vụ của từng thành viên | -Tìm hiểu và cài đặt talent  -Tìm kiếm những tài liệu về talent  -Viết báo cáo sơ lược và gửi về cho nhóm trưởng | **-**Tìm những kiến thức cơ bản về java và cung cấp cho nhóm  -Tìm những tài liệu về Talent cùng bạn Thịnh  -Viết báo cáo từng tuần và gửi cho giáo viên |
| Nhiệm vụ chung | Cài đặt và chạy được những gì mà nhóm đề ra | |

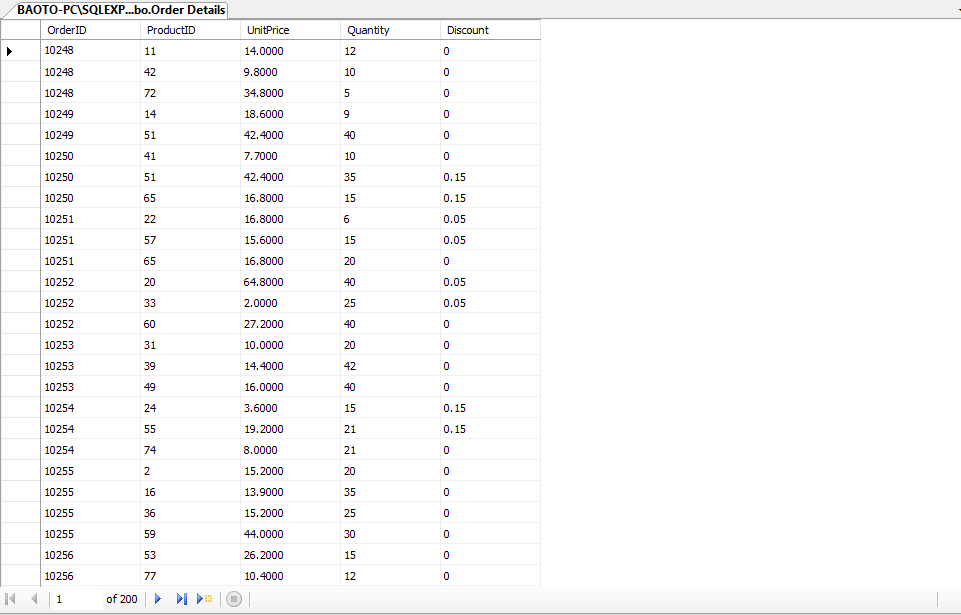
**2) Tiến độ thực hiện:**

**Hoàn thành được các công việc mà giữa kỳ chưa làm được như:**

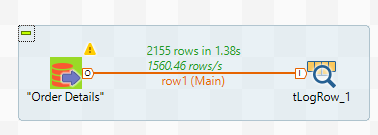
* Đọc dữ liệu từ database File, Thực hiện Truy vấn đữ liệu trong Talend
* Làm việc với XML

**II. Chi Tiết:**

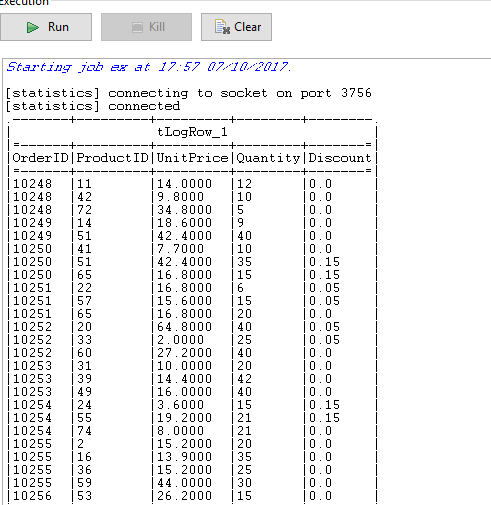
1. **Đọc dữ liệu từ database File (SQL Server)**

Đầu tiên ta phải chuẩn bị 1 file data base có sẵn trong SQL

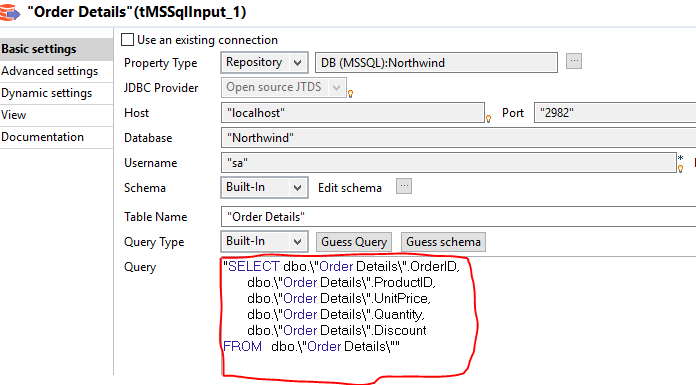
Sau đó vào Talend , Tạo 1 new Job Design . Trong Job ta thêm các thành phần :tMSSqlInput ( thực hiện thêm - kết nối SQL đã trình bày ở trên và cấu hình sau đó thả vào Job Design) và tLogRow . Khi hoàn tất thêm ta tiến hành kết nối 2 thành này với nhau . Cuối cùng bấm Run để chạy .



Và kết quả sẽ được:

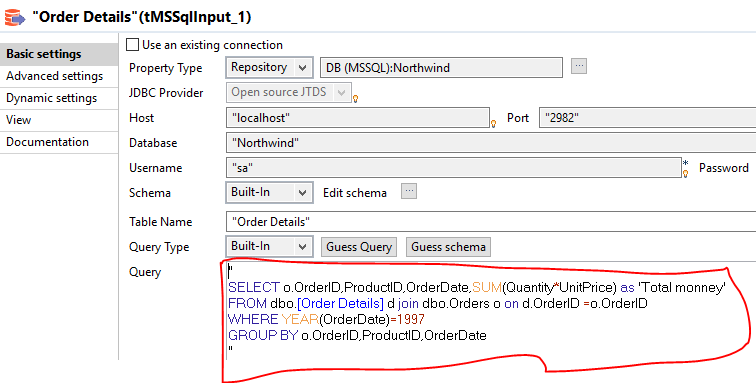


- Ngoài việc hiển thị dữ liệu từ Database File người dùng có thể xuất ra thông tin theo ý muốn với những điều khác nhau hay vì xuất toàn bộ thông tin như trên . Để làm được điều này , ta nhấn double vào thành phần tMSSqlinput . Sau đó hiện lên như sau

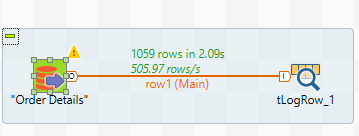


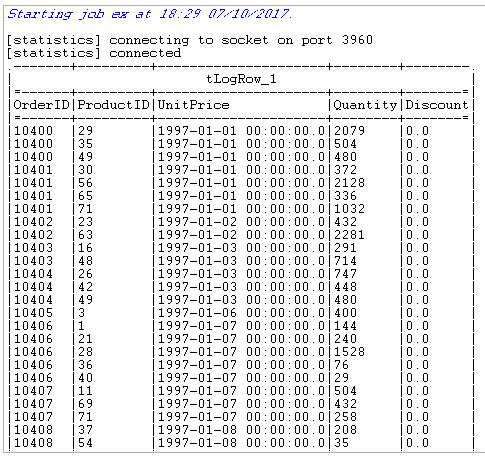
Chỗ khoanh tròn là chỗ cho phép người dùng nhập những câu lệnh để truy vấn dữ liệu

Giả sử , từ bảng dbo.Order Details ta muốn kết nối với bảng khác trong file Database (SQL) ví dụ ở đây là dbo.Orders để xuất ra các mã hóa đơn , các mã sản phẩm , ngày nhập hàng,Tổng tiền hóa đơn với điều kiện rằng hóa đơn đó phải có các sản phẩm có năm nhập hàng là 1997. Thì viết vào câu lệnh vào Query như sau



Sau khi đã hoàn tất . Nhấn Run để xem kết quả

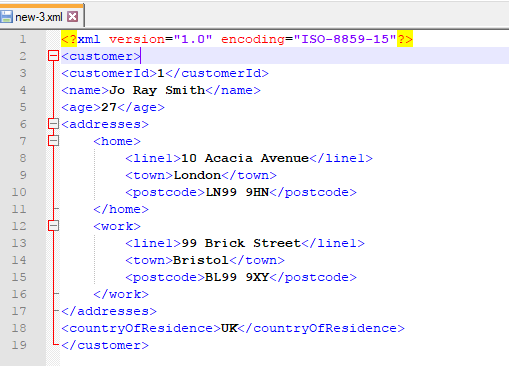


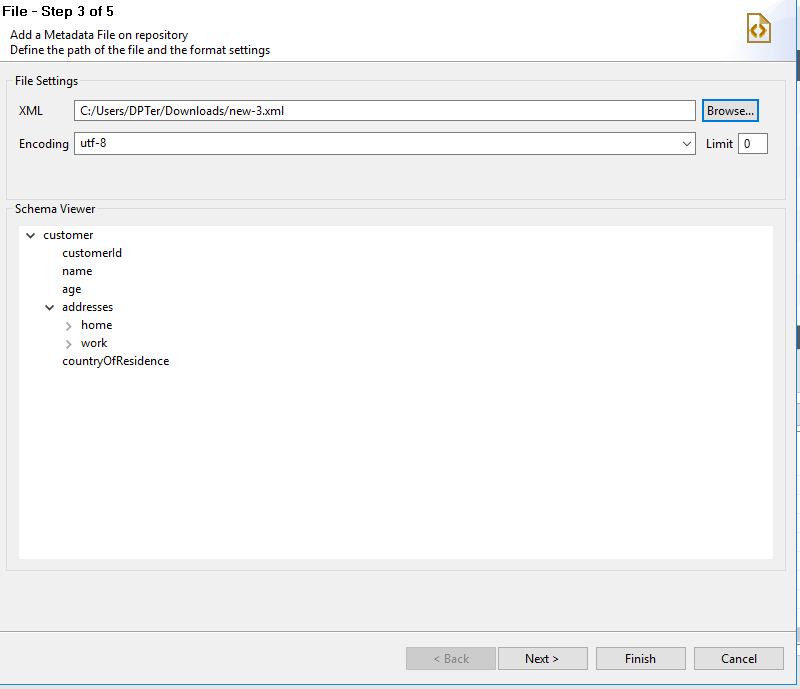


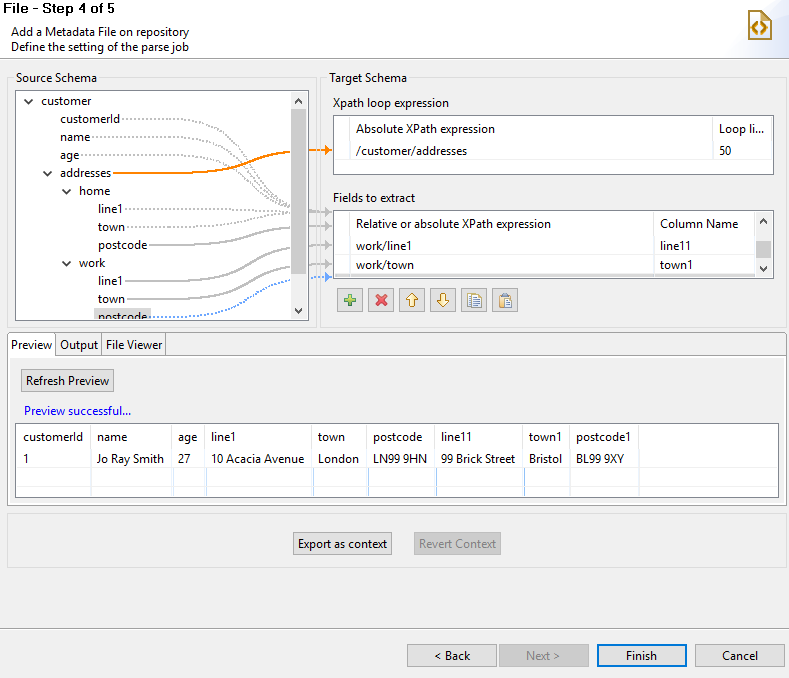
1. **XML**

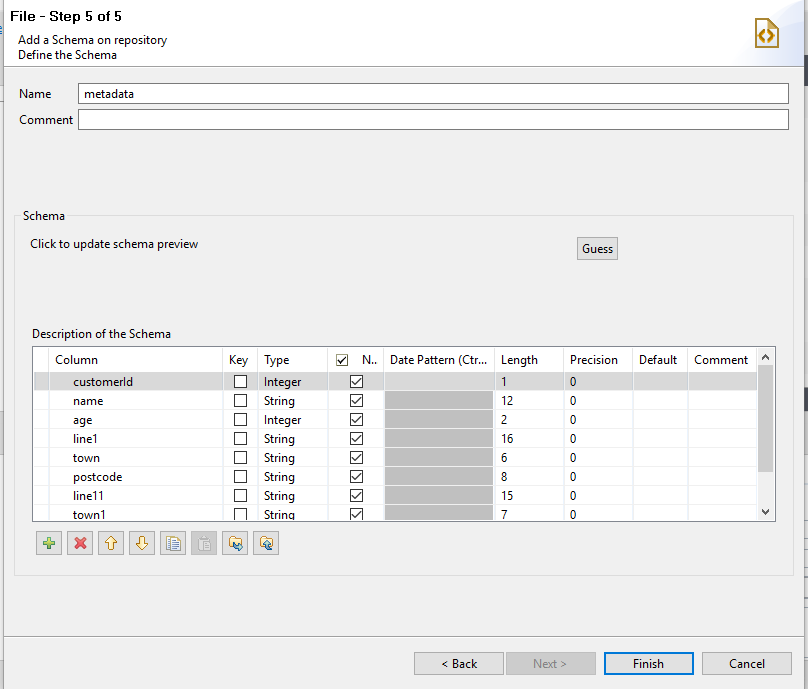
**B1**: Click vào **Metadata** -> **File XML** -> **Create file XML** rồi tạo tên cho file, **Next.**

**B2**: Chọn 1 chế độ để tạo **XML metadata**: **Input** or **Output** , **Next.**

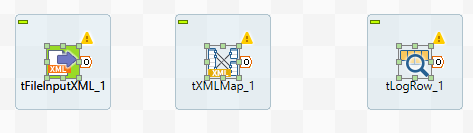
**B3**: Tìm 1 file có đuôi *(.xml)* rồi đưa vào **phần XML** , sau đó **phần Schema Viewer** sẽ hiện lên đoạn code có đuôi *(.xml)* có trên màn hình -> **Next**

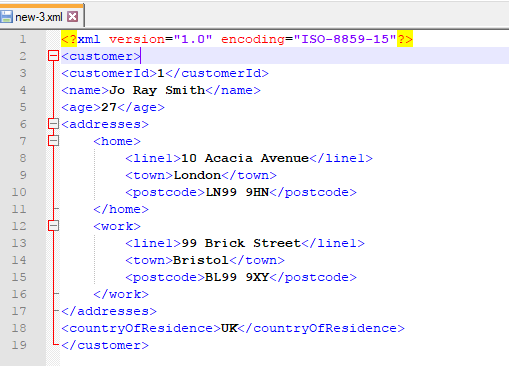


**B4**: Đưa các thông tin chính lên trên, đưa thông tin con xuống dưới -> rồi **click** vào **Preview – Refresh Preview** ,nó sẽ hiện lên bảng mà ta đã đưa vào . **Next**.

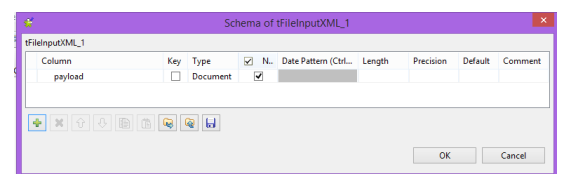
**B5**: Thêm hoặc thay đổi các số liệu( kí tự)trong **phần Description of the schema** -> **Finish**. Đã xong phần tạo **file XML**

Tiếp theo , ta tạo 1 **Job** rồi đưa các tên tài liệu vào **Job** bằng cách “ **click** chuột vào **Job** rồi bấm 1 chữ cái nào đó, nó sẽ hiện lên 1 loạt tên tài liệu mà ta cần tìm”.



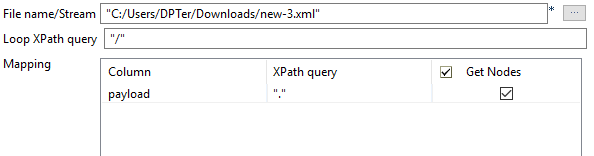
Đây là file có đuôi *(.xml)* [ Notepad++] để đưa vào *tFileInputXML****.***

Tạo mới một Job, kéo thả tFileInputXML vào cửa sổ thiết kế. Thêm trường  
payload có kiểu dữ liệu Document.



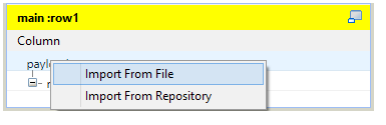
Thay đổi đường dẫn File name/Stream đến file xml vừa tạo và thay đổi LoopXpath query thành “/”.

Thay đổi XPath query thành “.” Và check vào Get Nodes.



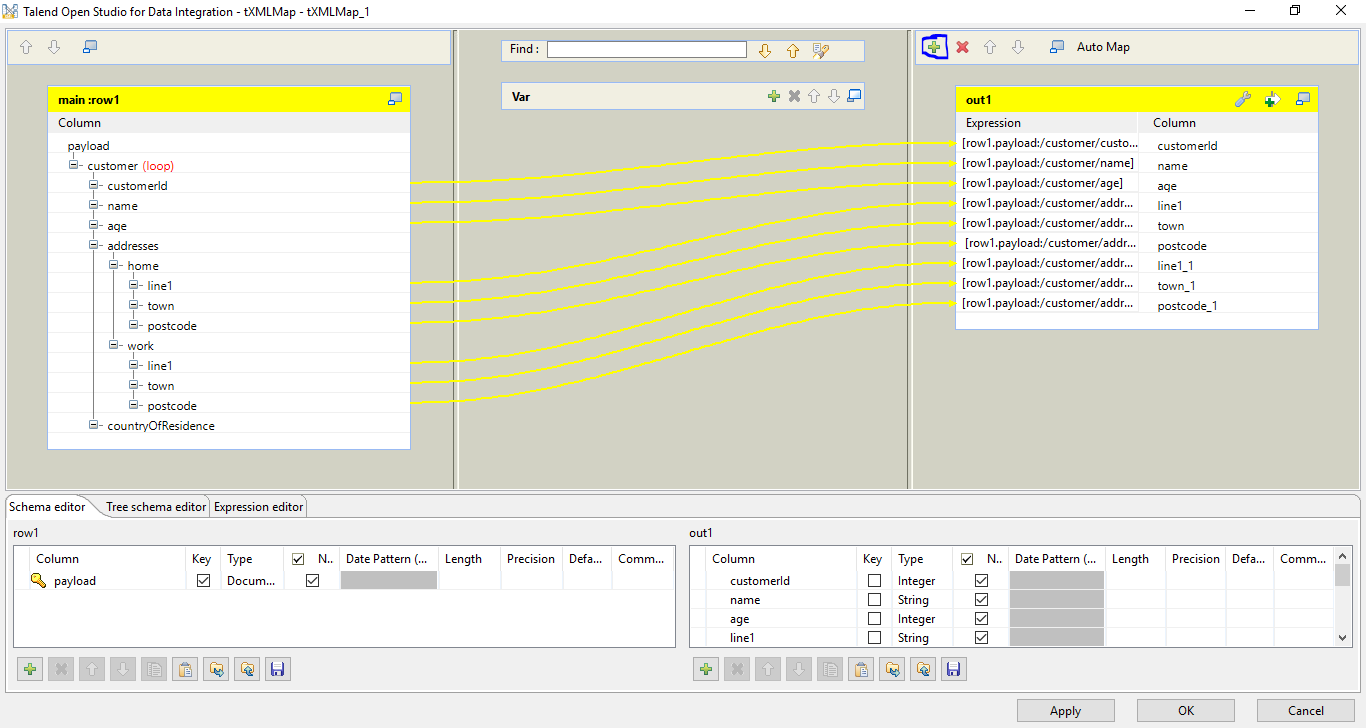
Kéo thả *tXMLMap* vào cửa sổ thiết kế và tạo liên kết từ *tFileInputXML* đến  
*tXMLMap*.

Mở *tXMLMap* và *right-click* và **payload** chọn **Import From File**. Sau đó dẫn  
đường dẫn tới *file xml* vừa tạo, bạn sẽ thấy cấu trúc của *file xml* được hiện ra.



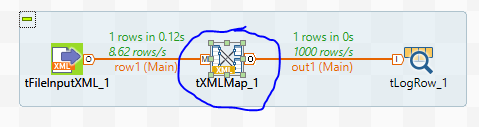
Click vào tXMLMap, sẽ hiện lên 1 bảng:

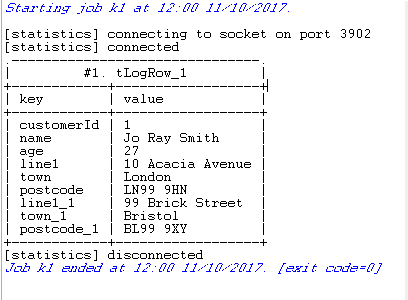
Nhấn  *button* và đặt tên là **customerOut**, sau đó lần lượt kéo thả các  
trường từ *file xml* sang *customerOut*.



Thêm *tLogRow* vào cửa sổ làm việc và liên kết đến *tXMLMap*, sau đó **Run** *Job*bạn sẽ nhìn thấy tất cả dữ liệu *xml* được Talend san phẳng thành một hàng

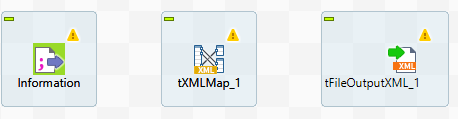
Cuối cùng,**Run**:



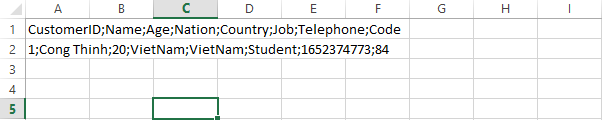


**Sử dụng tXMLMap trong Talend để tạo ra một tài liệu về XML:**

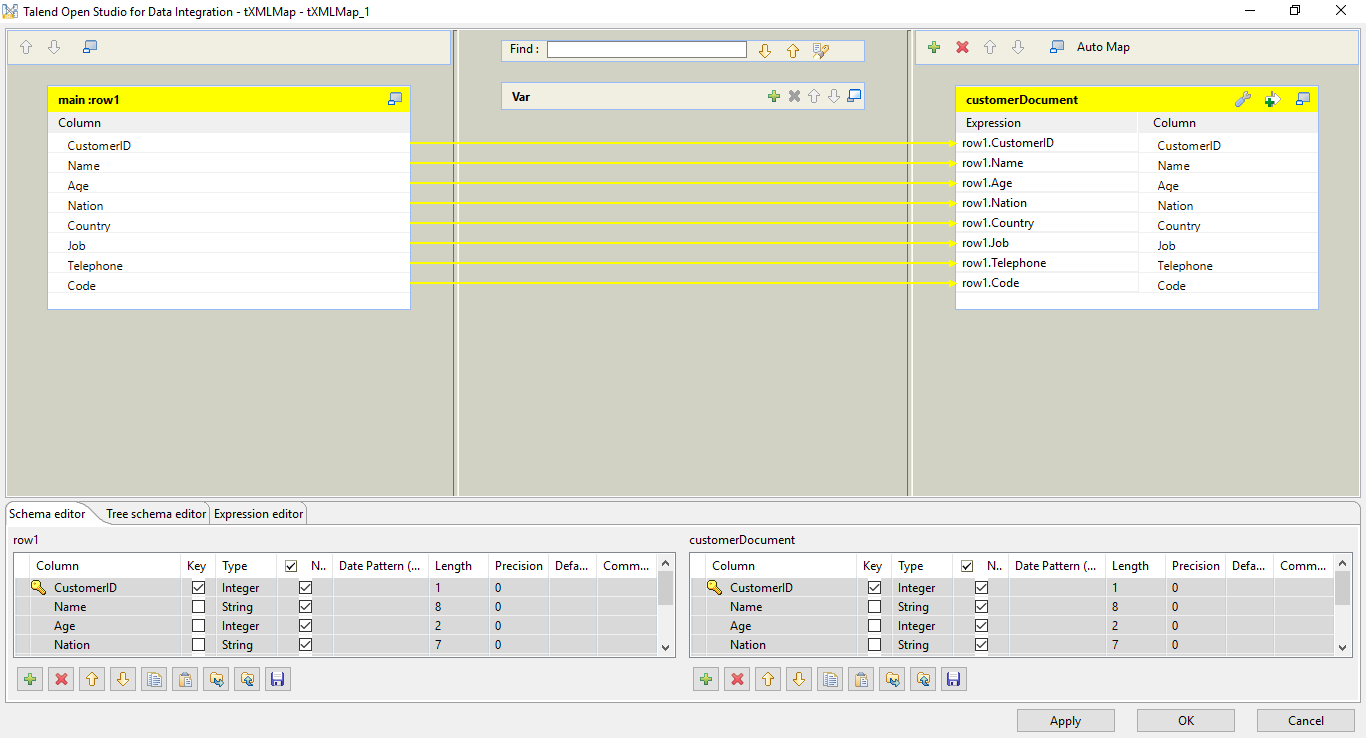
Thêm các thành phân vào Job Design: FileInputDelimited,tXMLMap\_1 và tFileOutputXML



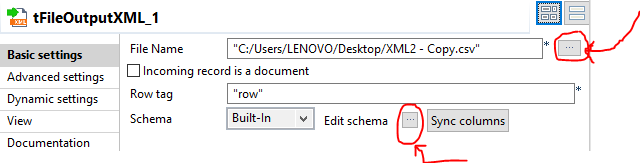
Trong đó tFileInputDelimited đầu vào bằng file XML2.csv có tên Information



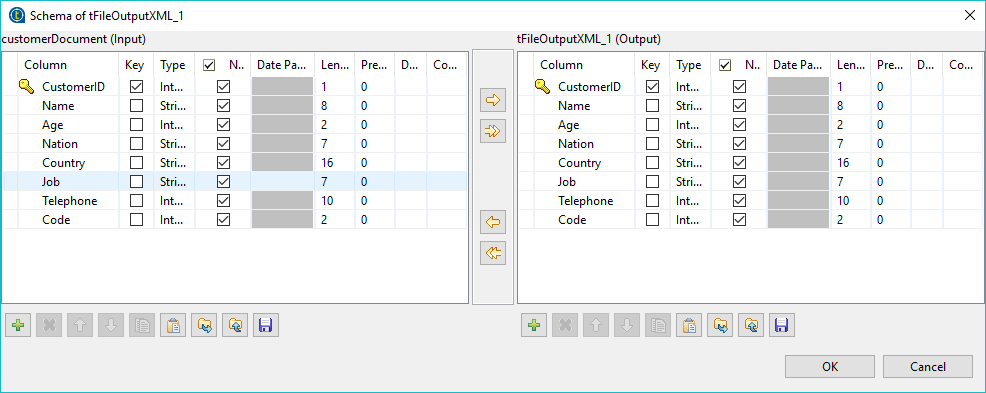
Thực kết nối Job bằng nhấn chuột phải chọn Row 🡪 Main . Sau đó cấu hình tXMLMap\_1 như hình dưới đây



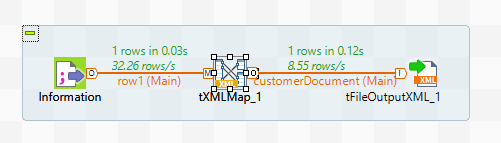
Sau đó double chuột vào tFileOutputXML . Chon Đường dẫn tới 1 file

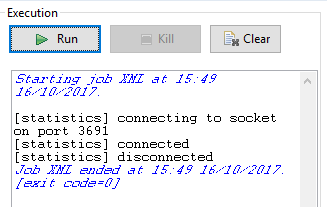


Rồi nhấn nút bên cạnh Edit schema sau đó cấu hình sao cho 2 đầu dữ liệu cùng field.Sau đó nhấn OK

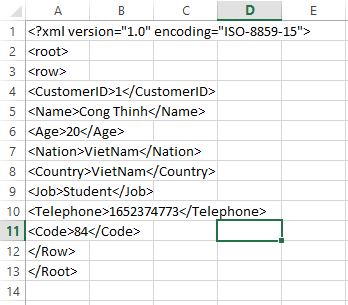


Cuối cùng nhấn Run để chạy Job





Mở file đầu ra để xem kêt quả



**III. Kết Luận:**

1. **Nội dung đã hoàn thành**

* Bắt đầu với một job cơ bản :

+ Tạo 1 job

+ Thêm các phần trong Job

**+** Cấu hình thành phần

**+** Thực hiện Job

* Mapping Data:

+ Đọc dữ liệu trên từ một file và ghi vào kết quả file mới

+ Gom nhóm và trích lọc dữ liệu

* Làm việc với DataBase:

+ Xuất dữ liệu qua Microsoft SQL Server từ 1 file dữ liệu

+ Đọc dữ liệu từ database File

+ Truy vấn đữ liệu trong Talend

* Làm việc với XML:

+ Sử dụng tXMLmap để đọc XML

+ Sử dụng tXMLMap trong Talend để tạo ra một tài liệu về XML

1. **Nội dung chưa hoàn thành**

* Làm việc với Java
* Quản lý File

1. ***Khả năng ứng dụng của Talend trong thực tế***:

*Talend là một bộ công cụ tích hợp dữ liệu mã nguồn mở có khả năng mở rộng quy mô từ các di chuyển tập tin nhỏ cho đến các dự án kho dữ liệu lớn. Nó có các công cụ mạnh mẽ cho ETL, kinh doanh và mô hình hóa công việc, và phân tích tác động, cũng như một khối chuyên nghiệp của trình thuật sĩ để giảm thiểu sự phức tạp*

**Tài liệu kham khảo:**

[1] “*Đồ án tin học ứng dụng tìm hiểu talend*”

(http://123doc.org/document/3363726-do-an-tin-hoc-ung-dung-tim-hieu-talend.htm)

[2] ”*UG\_EN\_DI\_TOSDI*” (Giảng viên cung cấp)