TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP TPHCM

**KHOA CÔNG NGHỆ THÔNG TIN**

TPHCM- Năm 2017

**BÁO CÁO CUỐI KÌ**

**Môn:Phát triển ứng dụng**

**MỤC LỤC**

**Giáo viên hướng dẫn :**

**Huỳnh Thái Học**

**Đỗ Hà Phương**

(Link GitHub: <https://github.com/tobao97/PTUD.git>)

**1. Tô Nguyễn Bảo(NT) 15027521**

**2. Trịnh Hưng Khang 15035581**

**3.Nguyễn Duy Phong 15030081**

**DHCNTT11A**

IMPORT DATA BY TALEND

**trang**

1. **Tổng quan báo cáo : 2**
2. *Bảng phân chia nhiệm vụ*
3. *Tình hình chung và những khó khăn khi thực hiện topic*
4. **Nội dung báo cáo :**
5. Làm việc với Database **3**
6. Đọc dữ liệu từ SQL server
7. Thực hiện truy vấn dữ liệu trong Talend **4**
8. Thực hiện tạo một bảng trong Talend **7**
9. Làm việc với Java
10. Viết chương trình Java cơ bản **9**
11. Tiếp cận the context và biến Global trong Java **10**
12. Làm việc với XML
13. Sử dụng tXMLmap để đọc XML **11**
14. Sử dụng tXMLmap trong Talend để tạo ra 1 fle XML **15**
15. Quản lý file
16. Thêm bản ghi mới vào file **17**
17. Di chuyển,sao chép , đổi tên và xóa thư mục **19**
18. Code Review và xuất Source Code **20**
19. Tổng kết **21**

TÀI LIỆU THAM KHẢO :

1. <https://info.talend.com/rs/talend/images/UG_EN_DI_TOSDI.pdf>

2. <http://123doc.org/document/3363726-do-an-tin-hoc-ung-dung-tim-hieu-talend.html>

3. <https://www.youtube.com/playlist?list=PL2PXs_iGAtLuawQmgJVeqwXieG-STXchn>

4. <https://www.youtube.com/watch?v=pXFazmPcqX8&t=63s>

5. <https://www.talendbyexample.com/talend-xml-reference.html>

6. <https://www.talendforge.org>

1. **Tổng quan báo cáo :**
2. ***Bảng phân chia nhiệm vụ :***

|  |  |  |
| --- | --- | --- |
| Tuần | Nhiệm vụ được giao | Đánh giá và mức độ hoàn thành |
| 6 | \*Tô Nguyễn Bảo : Tiếp tục làm việc với database . Demo project , làm và nộp báo cáo tuần 6 | 100% |
| 7 | \*Nguyễn Duy Phong:Làm việc với Java . Demo project , làm và nộp báo cáo tuần 7 | 100% |
| 8 | Trịnh Hưng Khang:Làm việc với XML . . Demo project , làm và nộp báo cáo tuần 8 | 100%  Tiến trình thực hiện của nhóm hơi chậm so với dự kiến |
| 9 | \*Tô Nguyễn Bảo : Quản lý file  \*Nguyễn Duy Phong : Tìm hiểu Code review trong Talend | 100% |
| 10 | \*Trịnh Hưng Khang : Xuất Soure Code  \*Các thành viên còn lại tổng hợp báo cáo, demo project ,thuyết trình trong buổi kiểm tra thực hành | 100% |
| Cuối kì | Tất cả thành viên cùng nhau thảo luận , chọn lọc – thực hiện báo cáo cuối kì | 100% |

1. ***Tình hình chung và những khó khăn trong quá trình hiện topic :***

-Các thành viên thực hiện một cách nghiêm túc , biết cách làm việc và hoàn thành tốt nhiệm vụ được giao .

- Sự chủ động của các thành viên trong quá trình thực hiện topic còn bị hạn chế và bị phụ thuộc . Chưa có sự cải thiện rõ rệt so với ban đầu . Nhưng có cố gắng .

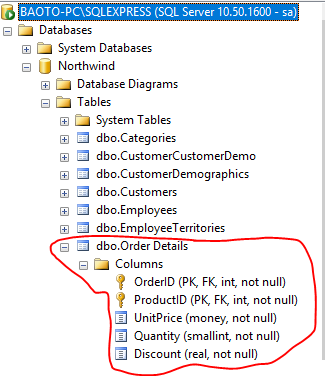
- Mỗi tuần , việc demo project đều gặp nhiều khó khăn trong quá trình thực hiện – phát sinh nhiều lỗi . Nên phải có sự giúp đỡ của các thành viên khác trong nhóm để hỗ trợ một phần , nhằm giải quyết – khắc phục những lỗi và giải quyết vấn đề một cách nhanh nhất .

- Talend còn khá mới mẻ . Đồng thời một trở ngại rất lớn , nhìn chung trình độ ngoại ngữ của nhóm còn khá yếu . Cho nên gặp trắc trở trong quá trình tìm tài liệu tham khảo và thực hiện nó dù đã có tài liệu .

-Vì điều kiện không cho phép nên Laptop của các tất cả thành viên không đủ mạnh . Do Talend khá nặng khi chạy nên thường bị treo , đứng , đơ máy hay loading rất lâu . Phải tốn rất nhiều thời gian để chờ mới thực hiện được . .

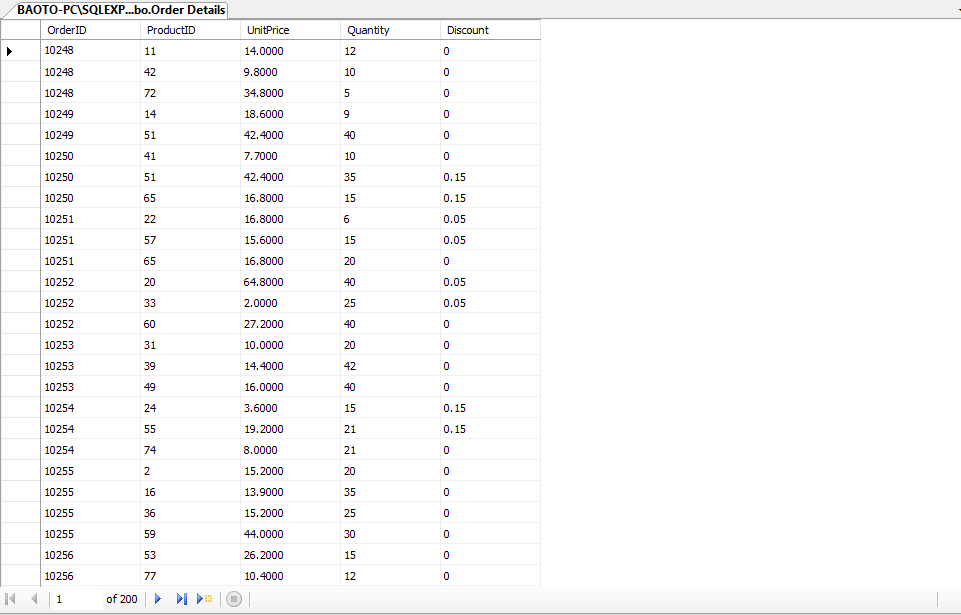
- Việc viết báo cáo cũng là một trong khó khăn của nhóm .Đa số nhóm demo bằng hình ảnh và chú thích chi tiết cho các bước thực hiện. Nhưng báo cáo bị giới hạn số trang khá ít nên nhóm phải cân đo , cân nhấc và lược bỏ rất nhiều

1. **Nội dung báo cáo :**
2. ***Làm việc với database:***

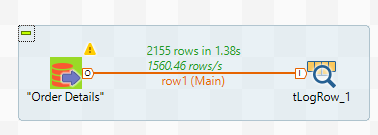
***a*)*Đọc dữ liệu từ database File (SQL Server)***

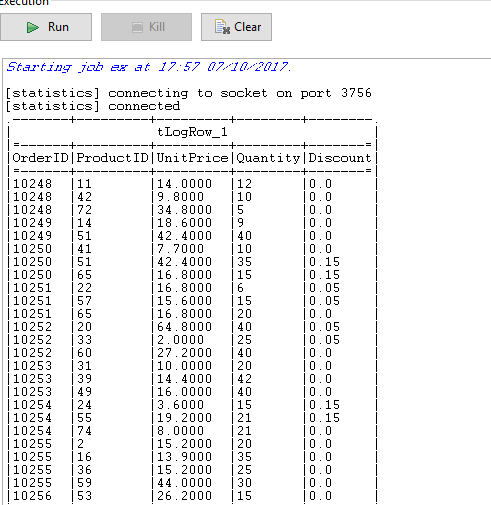
**-**Đầu tiên ta phải chuẩn bị 1 file data base có sẵn trong SQL . Ở đây ta có Northwind Datatabase(Được dùng trong môn hệ CSDL) . Với các bảng ta chọn dbo.Order Details

Mở bảng Order Detail (chi tiết hóa đơn ) ta thấy dữ liệu sau đây . Thông tin gồm có OrdersID(mã hóa đơn),ProductID(mã sản phẩm),UniPrice(Đơn giá),Quantity(Số lượng),Discount(Giảm giá)



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 .

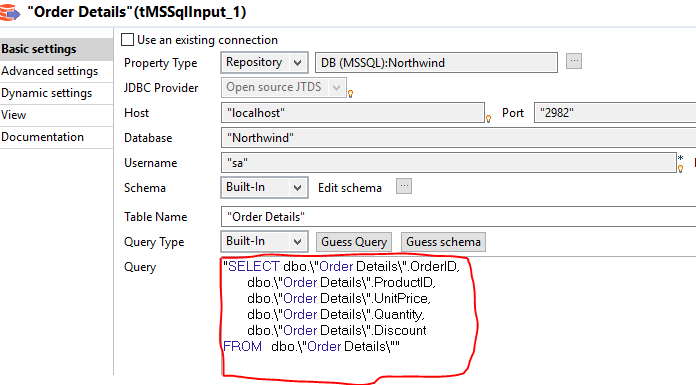




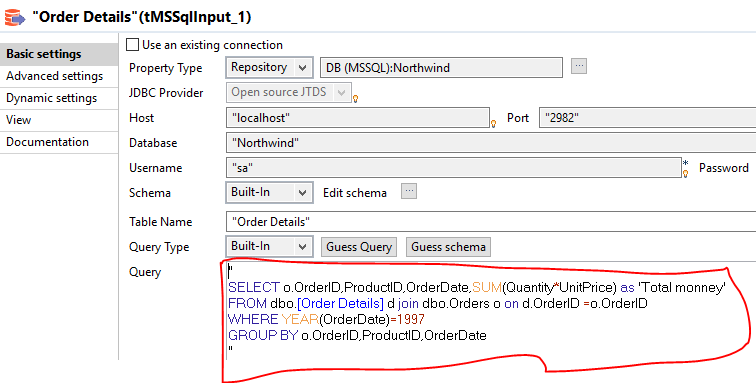
Sau khi chạy xong ta thấy toàn bộ dữ liệu trong Database được hiển thị

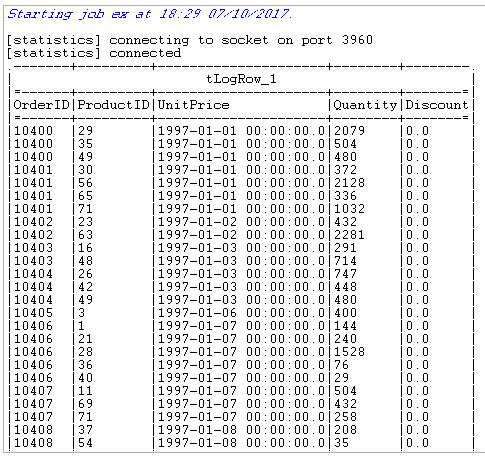
***b)Thực hiện Truy vấn đữ liệu trong Talend***

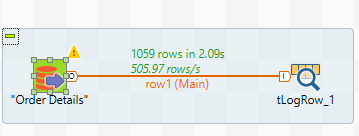
- Ngoài việc hiển thị dữ liệu từ Database File người dùng có thể xuất ra thông tin theo ý muốn với những điều khác nhau hay vì xuất toàn bộ thông tin như trên . Để làm được điều này , ta nhấn double vào thành phần tMSSqlinput . Sau đó hiện lên như sau



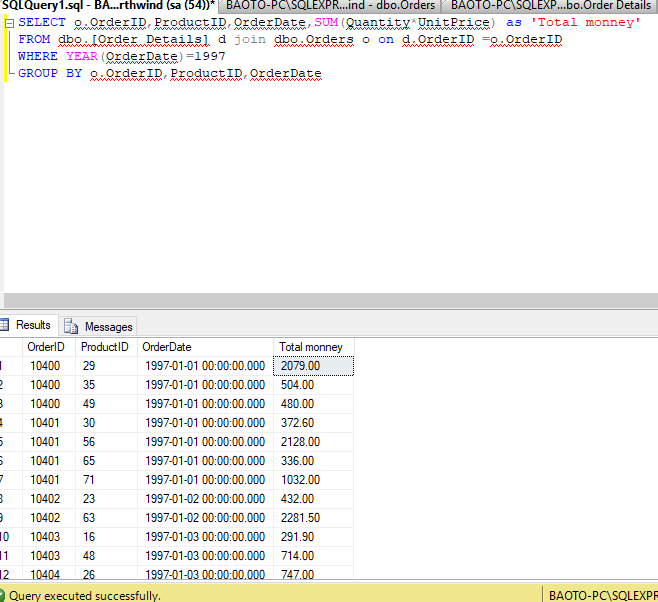
Chỗ khoanh tròn là chỗ cho phép người dùng nhập những câu lệnh để truy vấn dữ liệu

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

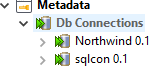
Sau khi đã hoàn tất . Nhấn Run để xem kết quả

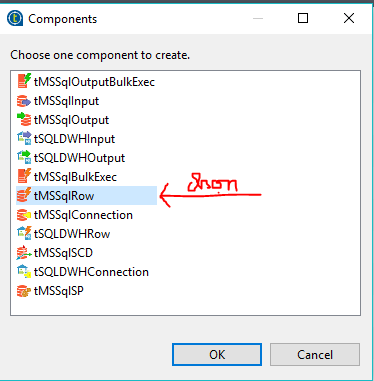


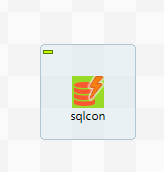
Đây là kết quả sau khi thay đổi câu truy vấn . Nó không còn xuất toàn bộ thông tin nữa mà chỉ xuất theo truy vấn của người dùng. Nếu chưa chắc chắn về kết quả thì viết lại câu đó vào SQL Server ta sẽ ra kết quả tương tự

\*

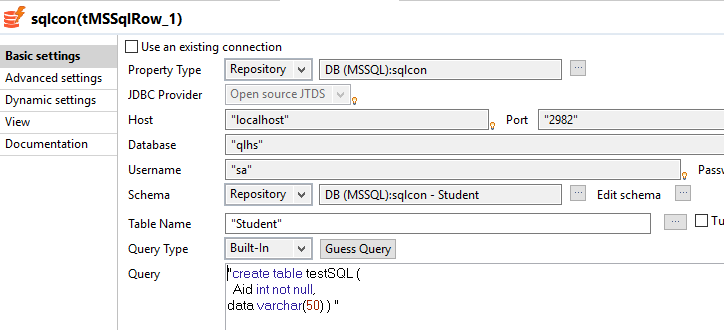
***c)Thực hiện tạo bảng database từ Talend :***

 - Tạo một Job mới và tiếp tục tạo DB conection – kết nối SQL.Ở đây chúng ta sẽ dùng lại sqlcon khi thao tác xuất liệu qua SQL từ 1 file dữ liệu . Trong DB conection , chon và kéo DB conecton vừa tại vào Job Design . Sau đó hiện lên hộp thoại . Nhấn chọn tMSSqlRow

Nhấn Ok để hoàn tất . Giờ trên trên màn hình Job Desin

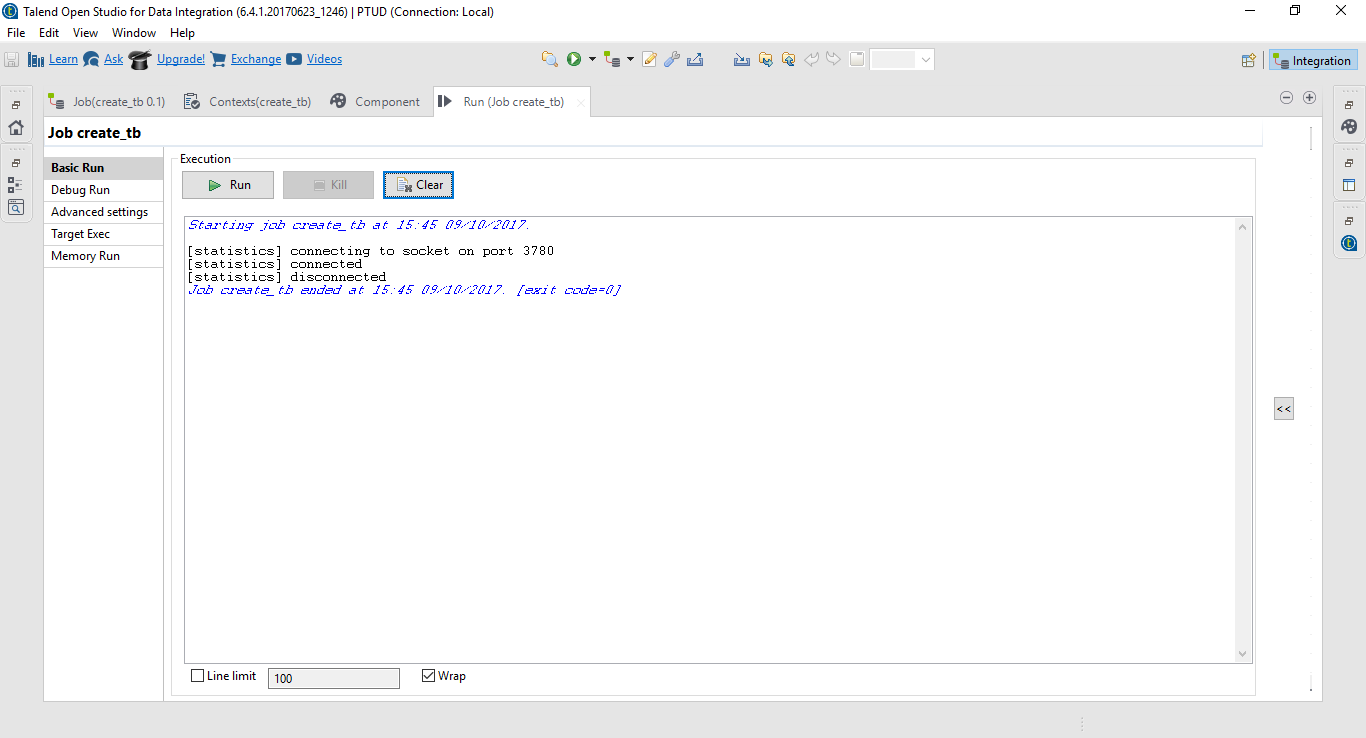


Nhấn double vào tMSSqlRow . Sau đó gõ lại câu lệnh sau vào chỗ Query

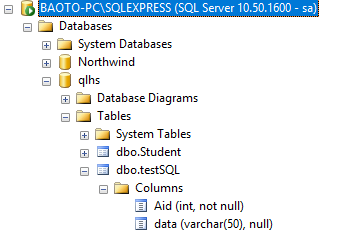


Có nghĩa rằng ta tạo 1 bảng có 2 trường dữ liệu là Aid và data

Quay lại màn hình chính nhấn Run để chạy . Sau chạy hiện lên như sau

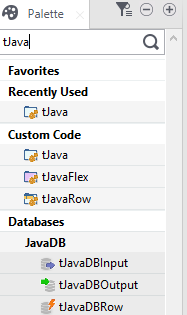


Vậy là ta đã tạo thành công bảng của database từ Talend . Để chứng thực ta vào SQL mở qlhs database ta thấy dbo.testSQL đã được thêm vào

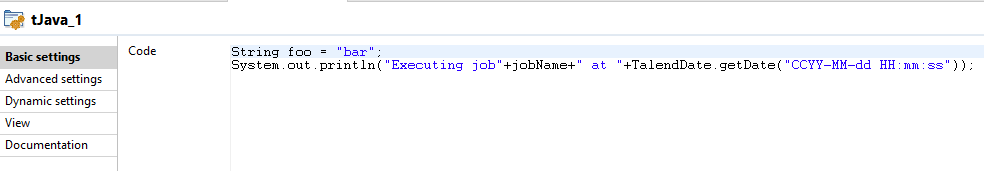


***2.Làm việc với Java :***

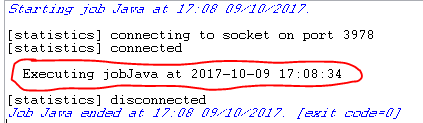
***a)Viết chương trình java cơ bản :***

Đầu tiên gõ tJava vào Palete rồi nhấn vào biểu tượng kính lúp tìm kiếm . Sau đó rê chuột tới tJava kéo vào Job Design . Ta được như hình bên

Nhấn double vào tJava\_1 . Trong code ta gõ vào code Java cơ bản như sau :

String foo = "bar";System.out.println("Executing job"+jobName+" at "+TalendDate.getDate("CCYY-MM-dd HH:mm:ss"));

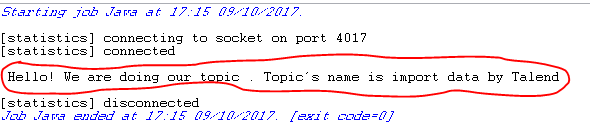
Sau đó vào quay lại màn hình chính của Job Design . Ta nhấn run để xem kết quả



Vậy là Câu lệnh đã được thực hiện

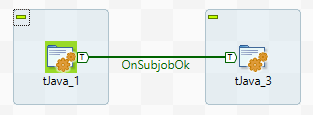
Vậy sẽ thử một ví dụ khác nhé ! Ở đây ta muốn in ra một câu trên Talend như sau “Hello ! We are doing our topic . Topic’s name is import data by Talend” thì cũng làm tương tự double chuột tJava rồi viết code vào như hình dưới đây

Sau đó chạy Job để xem kết quả



***b)Tiếp cận the context và biến globalMap trong Java***

Ta thêm một tJava vào Job Design . Sau đó chọn và nhấn chuột phải vào tJava\_1 rê chuột vào Trigger -> on SubJob Ok. Ta được như sau

Sau đó double chuột vào một trong 2 . Sau đó chọn Context(Java) sẽ nhìn thấy một biến tên là testValue .Quay lại Component gõ vào từng đoạn code tương ứng :

\*tJava\_1:

***System.out.println("tJava\_1");  
context.testValue ="testValue is now initialized";***

***globalMap.put("gmTestValue", "gmTestValue is now initialized");***

***\****tJava\_3:

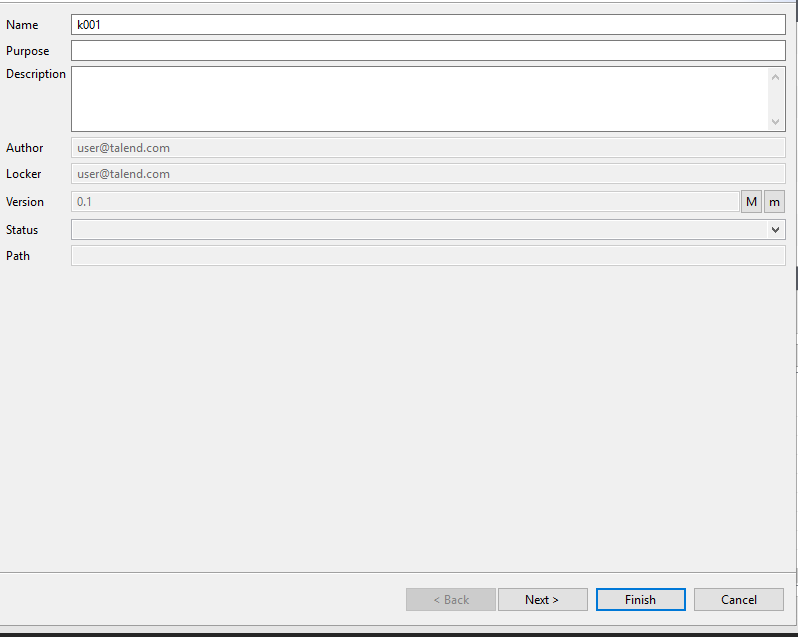
***System.out.println("tJava\_2");  
System.out.println("context.testValue is: "+context.testValue);  
System.out.println("gmTestValue is: "+(String)  
globalMap.get("gmTestValue"));***

Cuối cùng nhấn Run Job để chạy sẽ nhìn thấy các biến được cài đặt trong lần đầu tiên sẽ được in chính xác trong lần thứ hai.Các biến *context* và *globalMap* được lưu trữ như trên toàn cầu có sẵn trong *Java  
hashmaps*. Điều này cho phép các giá trị được tham chiếu trong bất kỳ các  
*components* khác, chẳng hạn như *tMap*, *tFixedFlowInput*, và *tFileInputDelimite*.

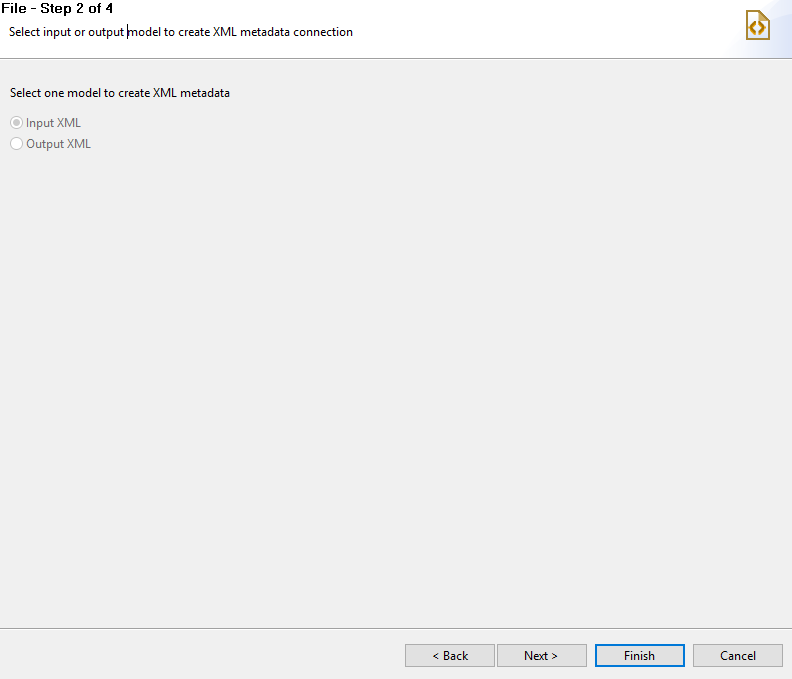
***3.Làm việc với XML:***

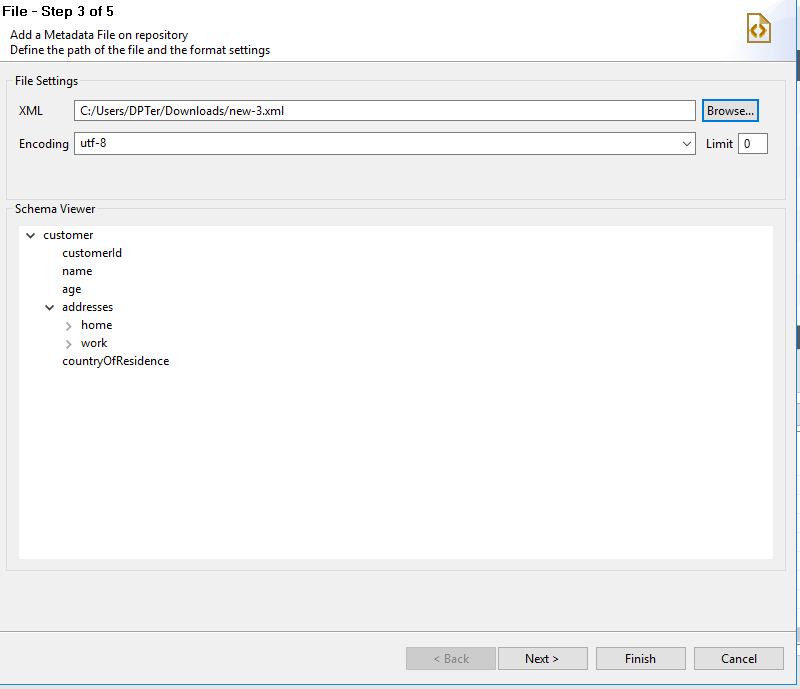
***a)Sử dụng tXMLmap để đọc XML:***

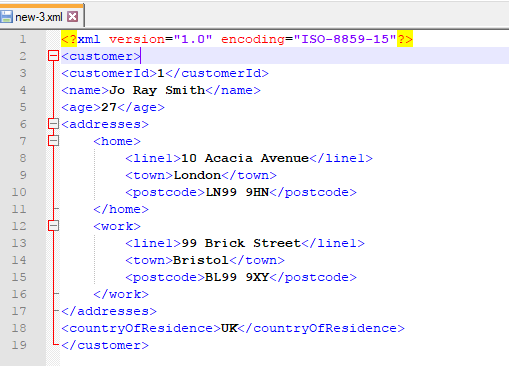
Để vào được XML trong talend , ta phải tạo file xml:

**B1**: Click vào **Metadata** -> **File XML** -> **Create file XML**(click chuột trái), rồi tạo tên cho file, **Next.**

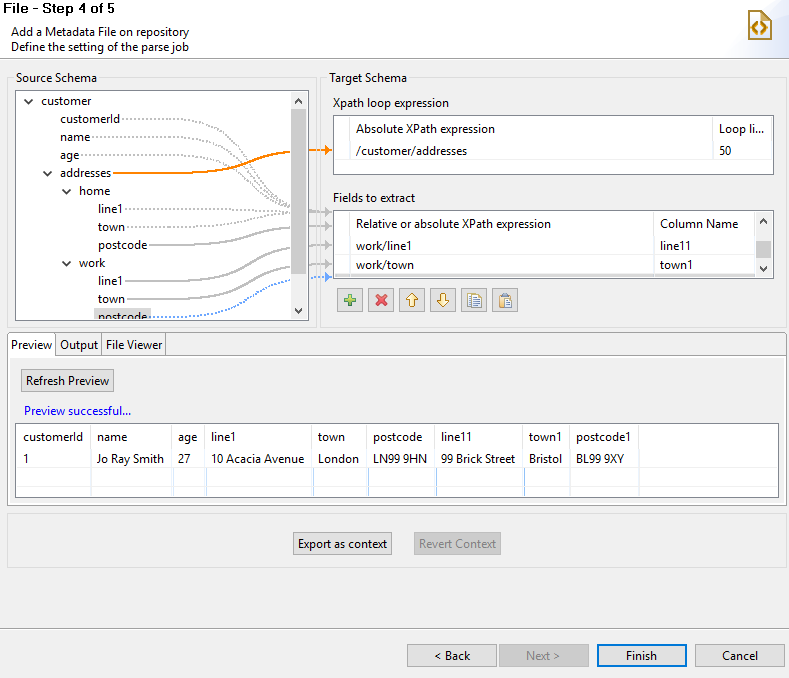
**B2**: Chọn 1 chế độ để tạo **XML metadata**: **Input** or **Output** , **Next.**

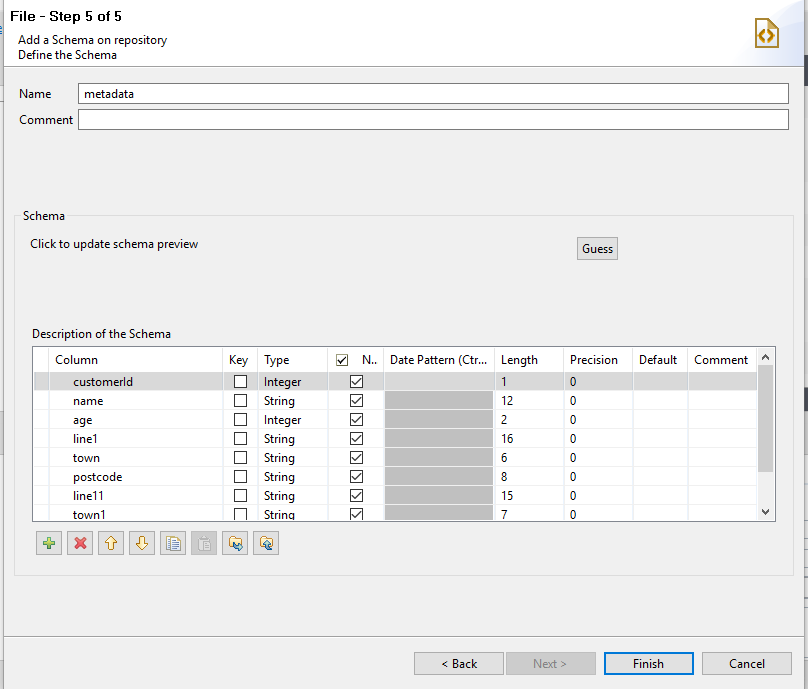


**B3**: Tìm 1 file có đuôi *(.xml)* rồi đưa vào **phần XML** , sau đó **phần Schema Viewer** sẽ hiện lên đoạn code có đuôi *(.xml)* có trên màn hình -> **Next**

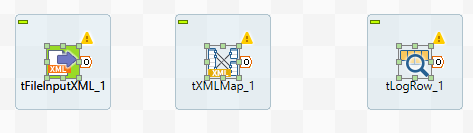
Đây là file .xml

**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**.

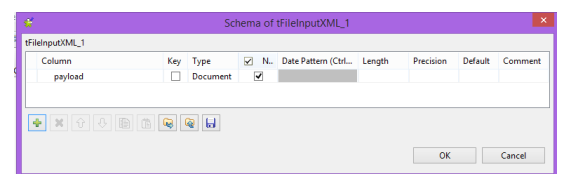


**B5**: Thêm hoặc thay đổi các **số liệu( kí tự)** trong **phần Description of the schema** -> **Finish**. Đã xong phần tạo **file XML**

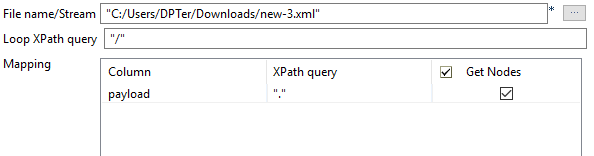
Tiếp theo , ta tạo 1 **Job** rồi đưa các tên tài liệu vào **Job** bằng cách “ **click** chuột vào **Job** rồi bấm 1 chữ cái nào đó, nó sẽ hiện lên 1 loạt tên tài liệu mà ta cần tìm”.



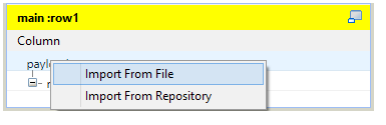
Tạo mới một Job, kéo thả tFileInputXML vào cửa sổ thiết kế. Thêm trường  
payload có kiểu dữ liệu Document.



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.

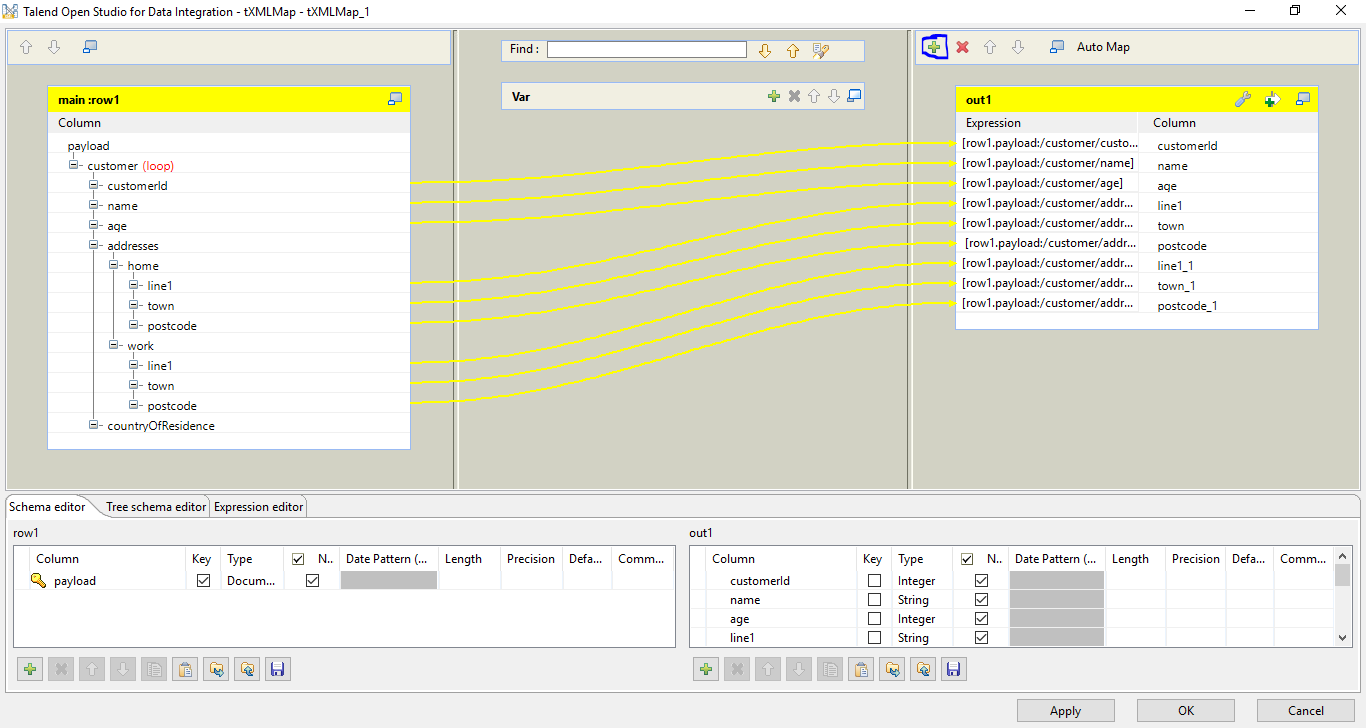


Kéo thả *tXMLMap* vào cửa sổ thiết kế và tạo liên kết từ *tFileInputXML* đến  
*tXMLMap*.Mở *tXMLMap* và *right-click* và **payload** chọn **Import From File**. Sau đó dẫnđường dẫn tới *file xml* vừa tạo, bạn sẽ thấy cấu trúc của *file xml* được hiện ra.

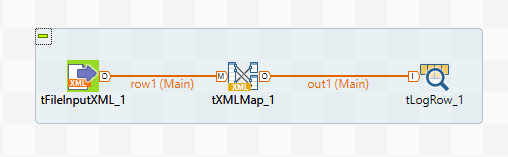


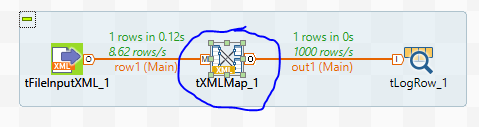
Click vào tXMLMap, nó sẽ hiện lên 1 bảng:

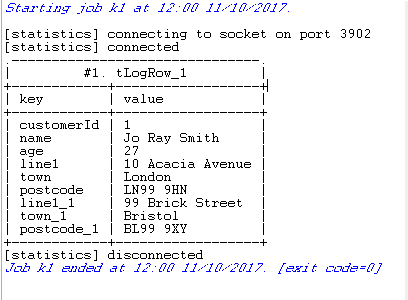
Nhấn 🞧 *button* và đặt tên là **customerOut**, sau đó lần lượt kéo thả các  
trường từ *file xml* sang *customerOut*.



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.

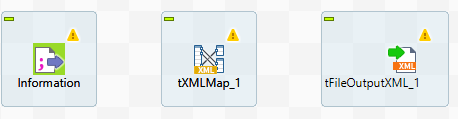


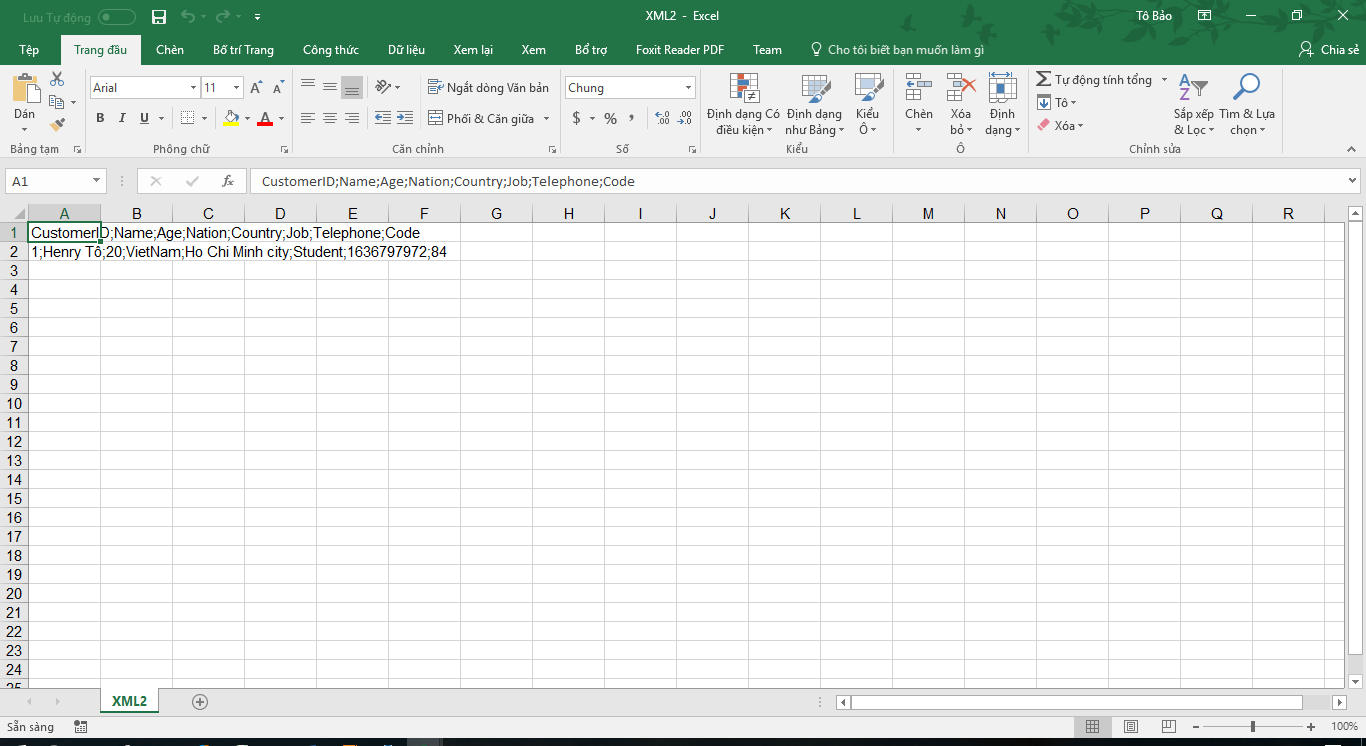
Cuối cùng,**Run**:

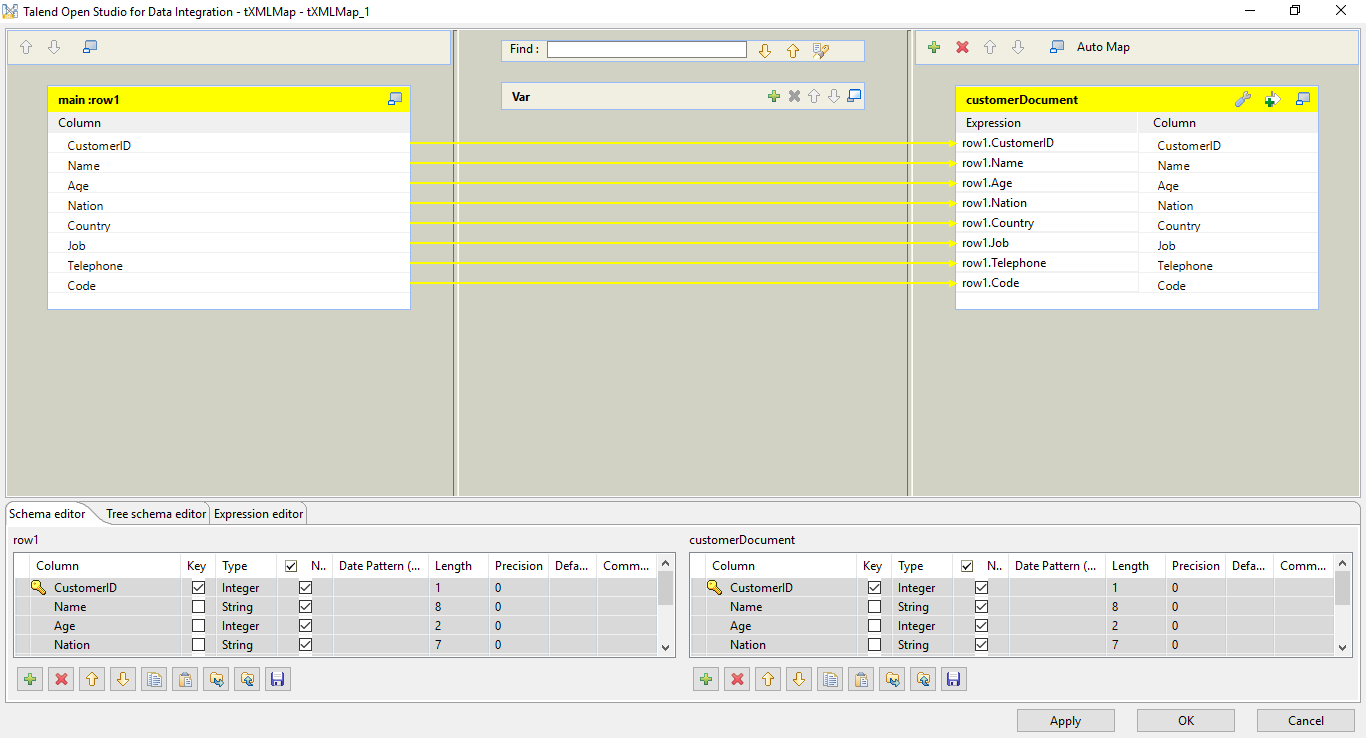


***b)Sử dụng tXMLMap trong Talend để tạo ra một tài liệu về XML:***

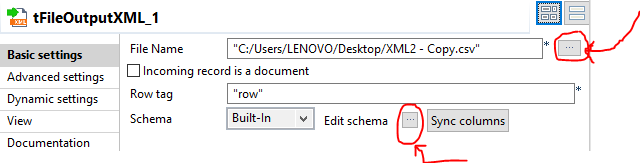
Đầu tiên thêm các thành phân vào Job Design :tFileInputDelimited,tXMLMap\_1 và tFileOutputXML



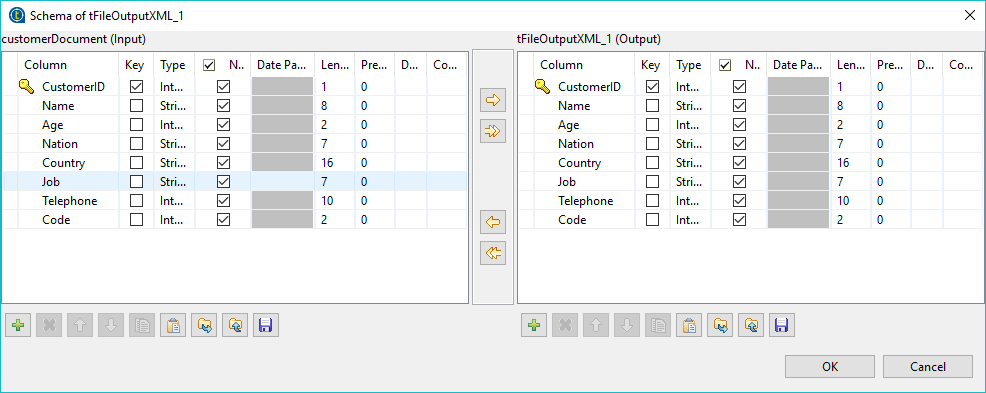
 Trong đó tFileInputDelimited đầu vào bằng file XML2.csv có tên Information

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

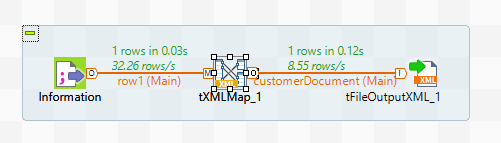
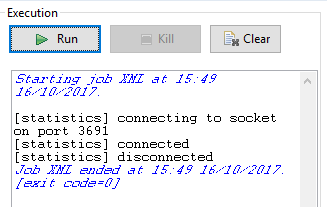
Sau đó double chuột vào tFileOutputXML . Chon Đường dẫn tới 1 file

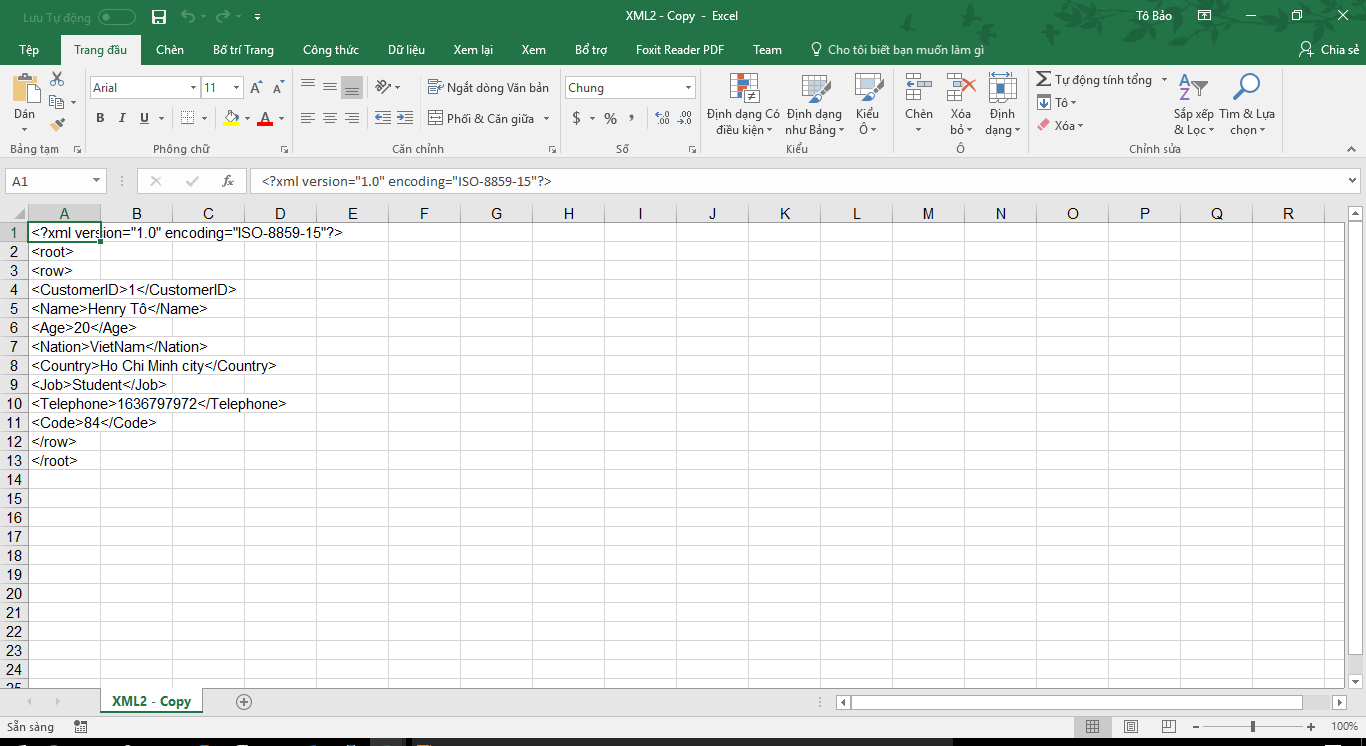


Rồi nhấn nút bên cạnh Edit schema sau đó cấu hình sao cho 2 đầu dữ liệu cùng field.Sau đó nhấn OK

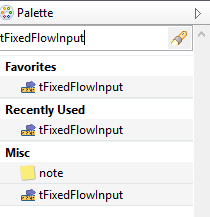


Cuối cùng nhấn Run để chạy Job

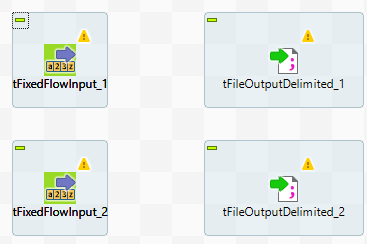


Mở file đầu ra để xem kêt quả

***4.Quản lý file***

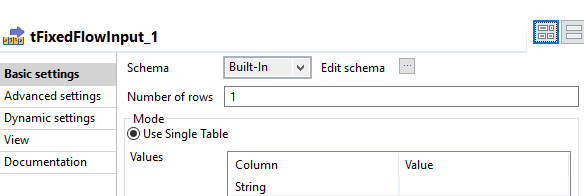
*** a)Thêm một bản ghi mới vào File***

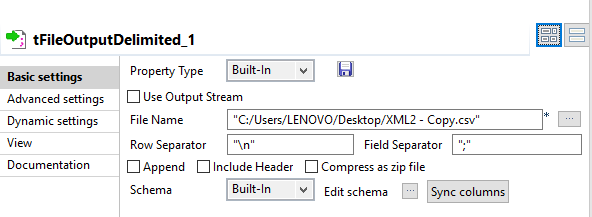
Tạo một Job mới đồng thời thêm các thành phần tFixedFlowInput và tFileOutputDelimited ( bằng cách gõ vào Palette sau đó kéo vào Job Design)như sau



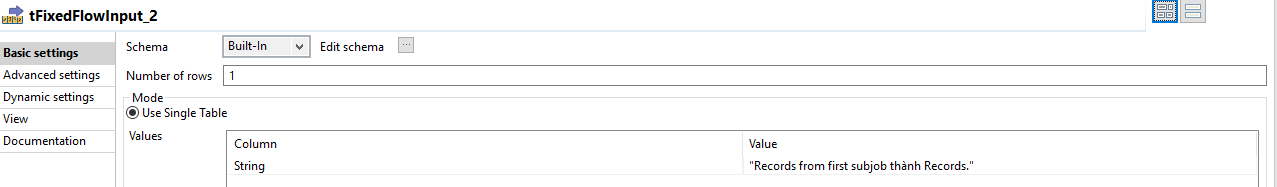
Sau đó kết nối với các thành phần với nhau và cấu hình từng thành phần

Ta thêm các field bằng cách nhấn nút bên Edit schema sau đó thêm các giá trị value cho từng field . Ở đây ta thêm một Column tên là String có giá trị là NULL.Lưu ý phải check vào Use Single Table

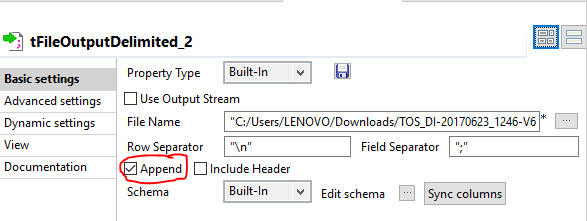


Chọn đường dẫn tới file cần ghi dữ liệu mới­

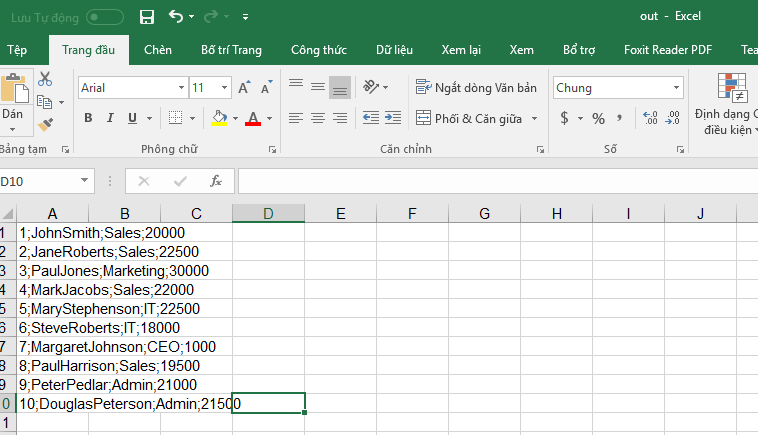
Tương tự tFixedFlowInput\_2 ,ta thêm một Column tên là String có giá trị là một chuỗi "Records from first subjob thành Records."



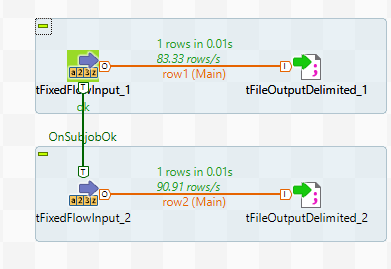
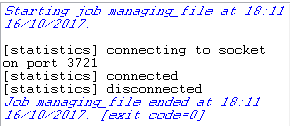
Lưu ý ở tFileOutputDelimited đánh check vào Append . Sau đó chọn đường dẫn đến file



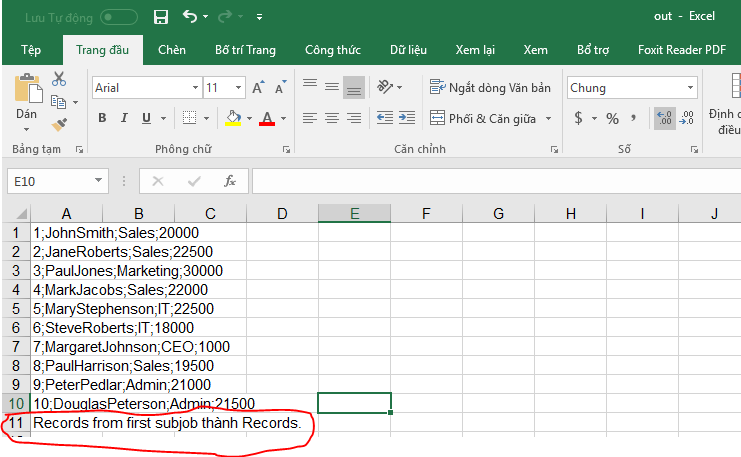
File đầu ra khi chưa run Job



Cuối cùng đã hoàn tất cấu hình nhấn Run Job

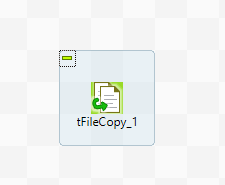


Vậy là sau khi chạy Job , mở lại các file đầu ra ta thấy một bản ghi mới đã được thêm vào . Dưới đây là file của tFileOutputDelimited\_2

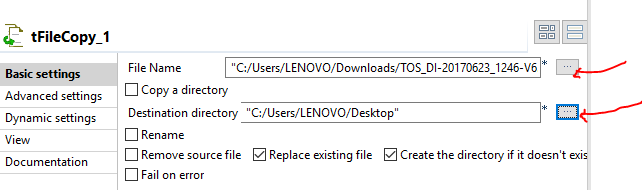


Còn đối với file của tFileOutputDelimited\_1 do nhận giá trị NULL nên ko có bản ghi mới

***b)Di chuyển ,Sao chép,đổi tên và xóa file – folders***

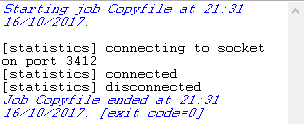
 \*Sao chếp 1 file vào thư mục khác:

Tạo 1 Job mới , và thêm thành phần tFileCopy . Double chuột vào tFileCopy



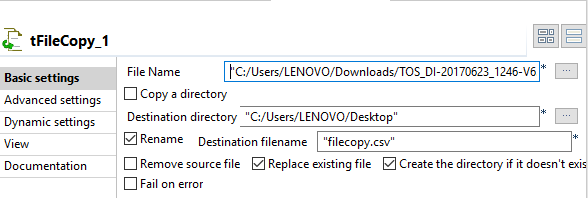
Thiết lập giá trị **File name**: chứa đường dẫn đến *file* cần copy.  
Thiết lập giá trị **Destination directory**: là nơi lưu tập tin được copy ra.

Sau đó nhấn Run Job . Sau đó mở đường dẫn ở **Destination directory** để sẽ thấy file đã đươc copy



\*Sao chép 1 file với một tên khác khi qua đường dẫn mới

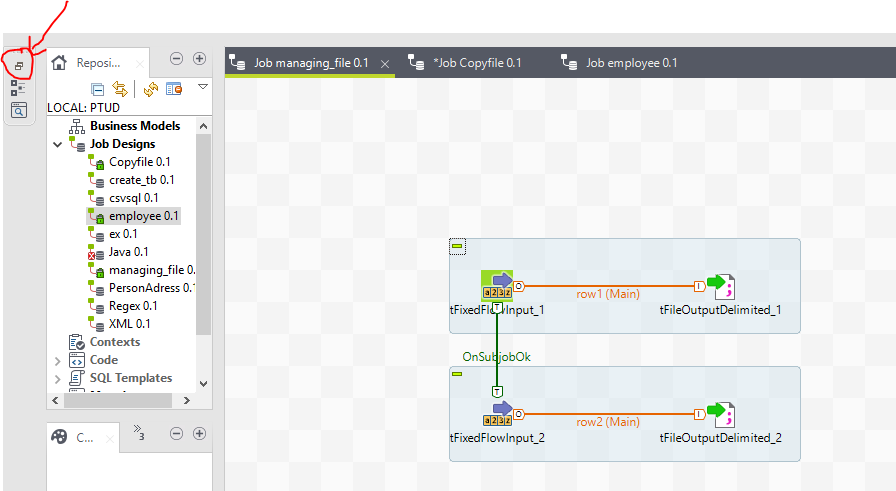
Cũng làm như trên . Nhưng khi double , ta check vào Rename và nhập tên mới trong Destination filename . Sau đó Run Job



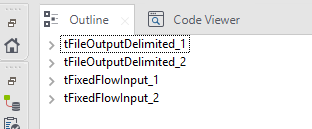
Nếu bạn muốn sau khi copy và đổi tên , bạn muốn file ở đường dẫn cũ . Nhấn check vào Remove source file . Cách này tương đương với di chuyển file hoặc xóa file

**5.Code Review và xuất souce code**

Đa số ta làm việc trên Talend là dùng tools công cụ để thực hiện . Nhưng ta có thể xem và thực hiện code cho mỗi Job Design .Để làm việc này ta mở một Job và nhấn vào Restore như hình như dưới



Khi nhấn vào xuất hiện một hộp thoại có 2 tag OutLine ( chứa các thành trong Job Design) và Code Viewer ( chứa code ) . Ta nhấn vào Code Viewer để xem code

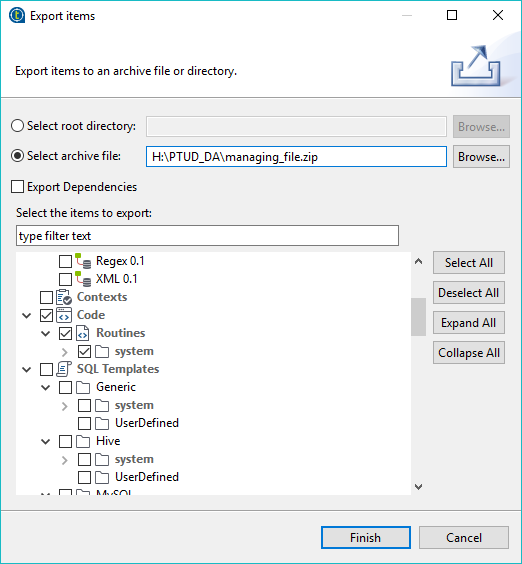


Như vậy trong Talend ta còn có thể xuất soure code ra một một tệp nén bằng các như sau . Chọn 1 một Job cần xuất Souce code sau đó chọn Export items

Ảnh có chứa ảnh chụp màn hình, máy tính

Mô tả được tạo với mức tin cậy rất cao

Một hộp thoại xuất hiện . Chọn Select archive file – nhấn Browse… để chọn đường lưu source code . Và check vào Code như hình dưới . Cuối cùng nhấn Finish



6.Tổng kết :

[**Talend**](http://www.talend.com/products-data-integration/talend-products.php)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.Qua tìm hiểu ta tổng kết các chức năng cơ bản : Thiết kế Business Model (Mô hình kinh doanh), Job Design, Mapping data,làm việc với Database ,Làm việc với Java , XML, Quản lý File ,….và nhiều thao tác phức tạp khác .

(Toàn bộ báo cáo : <https://github.com/tobao97/PTUD/blob/master/T%E1%BB%95ng%20h%E1%BB%A3p%20b%C3%A1o%20c%C3%A1o(full).docx> )