TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP TPHCM

**KHOA CÔNG NGHỆ THÔNG TIN**

TPHCM- Năm 2017

**BÁO CÁO**

**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**

IMPORT DATA BY TALEND

Page

1. Tổng quan về báo cáo đồ án:
2. Nhiệm vụ phân chia 3
3. Tiến độ thực hiện
4. Nội dung : 4
5. Giới thiệu Talend
6. Giới thiệu chung về Talend Open Studio
7. Lịch sử ra đời của Talend 5
8. Các sản phẩm của Talend 8

-Dòng sản phẩm quản lý dữ liệu

-Dòng sản phẩm tích hợp ứng dụng

2.Bắt đầu một dự án Talend 10

a) Cài đặt Talend

b) Giới thiệu về Project trong Talend 12

- Tạo một Project 13

- Import project 15

- Mở project 19

- Xóa project 20

- Export project 21

3.Thiết kế Business Model (Mô hình kinh doanh) 22

a)Business Model là gì

b)Mở hoặc tạo 1 Business Model 23

c)Giao diện và các thành phần trong Business Model 24

d)Chỉnh sửa 29

- Đổi tên

- Sao chép

- Di chuyển

- Xóa

4.Thế nào là Job Design ?

5.Bắt đầu với 1 Job đơn giản

a)Tạo 1 Job 30

b)Thêm các thành phần trong Job 32

- tFlieInputDelomited

- tFile OutputDelimited

- tLogRow

c)Cấu hình thành phần 35 - tFileInputDelimited

- tFileOutputDelimited -tLogRow

d)Thực hiện Job 37

e)Demo

- Đọc và hiển thị từ 1 file - Từ 1 file xuất thông tin qua 1 file khác 44

6.Mapping data 46

a) Giới thiệu Mapping Data

b)Các thao tác(demo) trên Mapping Data

- Đọc dữ liệu 1 từ 1 file và ghi kết quả vào file mới

- Gom nhóm và trích lọc dữ liệu

7.Database 57

a)Demo xuất thông tin qua Database từ file dữ liệu

b) Đọc dữ liệu Database 62

c)Thực hiện truy vấn dữ liệu 65

d)Tạo bảng trong Talend 68

8.Làm việc với Java 71

a)Viết chương trình Java đơn giản trong Talend

b)Tiếp cận The context & biến Gloal Map trong Java

9.Làm việc với XML 72

a)Đọc dữ liệu XML

b)Từ 1 file tạo ra 1 file XML 80

10.Quản lý File 82

a)Sử dụng Regular Expresstion

b)Thêm một bản ghi mới vào file 85

c)Di chuyển,sao chép và xóa file trong folders 89

11.Code Review và xuất Souce Code 90

12.Tổng kết

III.Tài liệu tham khảo :

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

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

1. **Tổng quan đồ án:**

**1)Nhiệm vụ thành viên:**

|  |  |
| --- | --- |
| Thành viên | Nhiệm vụ |
| Tô Nguyễn Bảo | -Hoàn thiện báo cáo cho các tuần .  -Chỉnh sửa và hoàn thiện đồ án.  - Demo project , Job Design,Maping Data,SQL server vào thứ 3 thực hành  - Tạo Github và quản lý đồ án trên GitHub |
| Trịnh Hưng Khang | -Tìm hiểu Talend  - Thiết kế Business Model  -Tìm hiểu cách thêm và cấu hình trong Job Design |
| Nguyễn Duy Phong | -Làm quen với dự án trong Talend  - Tạo một Job Design  - Tìm hiểu về mapping data |

* Mức độ hoàn thành trong các tuần theo từng thành viên

|  |  |  |
| --- | --- | --- |
| Thành viên | Mức độ hoàn thành | Ghi chú |
| Tô Nguyễn Bảo | 100% |  |
| Trịnh Hưng Khang | 70% | Không thực hiện nhiệm vụ trong tuần 4 và có số bài làm chưa đầy đủ và sơ sài |
| Nguyễn Duy Phong | 90% | Còn thiếu chưa được đầy đủ |

***2)Tiến độ thực hiện đồ án***

-Tiến độ hoàn thành : 50%

-Những nội dung đã được:

+Làm quen với Talend Open Studio for Data Integration

+Tạo , thêm , xóa,mở,import,export một Project trong Talend

+Thiết kế một Business Model

+Bắt đầu làm việc với một Job đơn giản

\* Tạo 1 Job đồng thời thêm và cấu hình các thành phần vào Job

\* Kết nối các thành phần trong Job với nhau

\* Chạy Job và demo

+Mapping Data và làm việc với database

-Những trắc trở trong việc thực hiện đồ án:

+ Các thành viên không có chủ động trong việc thực hiện đồ án . Giao cái gì thì làm cái đó không giao thì không làm . Có khi có giao mà không làm không thực hiện mà không có lí do chính đáng.Chưa biết tìm kiếm tài liệu . Có tài liệu mà không biết sử dụng tài liệu triệt để dẫn đến làm việc sai mục đích đề ra ban đầu

+ Gặp trở ngại trong việc chạy 1 job . Khi đưa một tệp csv vào trong Talend và hiển thị nó , Hoặc xử lí xuất thông tin từ CSV này sang 1 CSV .

+ Kết nối với SQL server là gặp trắc trở nhất cũng như xuất thông tin và hiển thị đọc nội dung

1. **Nội dung báo cáo đồ án giữa kì :**

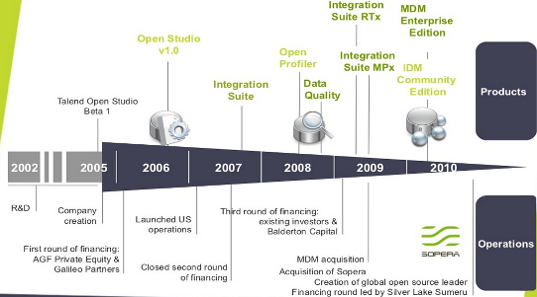
**1.*Giới thiệu Talend.***

**a)** *Giới thiệu chung về Talend:*

* Talend là một nhà cung cấp phần mềm mã nguồn mở. Talend cung cấp khả năng tích hợp dữ liệu (*data integration*), quản lý dữ liệu (*data management*), tích hợp ứng dụng doanh nghiệp (*enterprise application integration*), các phần mềm và các dịch vụ dữ liệu lớn (*big data*).
* Talend Open Studio là một tập hợp mạnh mẽ và linh hoạt của các sản phẩm mã nguồn mở để phát triển, thử nghiệm, triển khai, quản trị các dự án quản lý dữ liệu và các dự án ứng dụng tích hợp. Talend cung cấp các nền tảng thống nhất làm cho quản lý dữ liệu và tích hợp ứng dụng dễ dàng hơn bằng cách cung cấp một môi trường thống nhất quản lý toàn bộ vòng đời của doanh nghiệp qua các ranh giới hoạt động của doanh nghiệp.
* Talend - nhà lãnh đạo toàn cầu trong hội nhập mã nguồn mở, cung cấp giải pháp tích hợp mà thực sự mở rộng cho bất kỳ loại thách thức tích hợp, bất kỳ khối lượng dữ liệu và bất kỳ phạm vi dự án dù đơn giản hay phức tạp. Chỉ có số liệu đánh giá cao khả năng mở rộng của Talend, ứng dụng và quy trình kinh doanh nền tảng tích hợp khả năng mở rộng của Talend. Talend kết hợp các dự án và các công nghệ để tăng tốc đáng kể thời gian đưa ra giá trị cho các hoạt động kinh doanh.
* Sẵn sàng cho môi trường dữ liệu lớn, kiến trúc linh hoạt của Talend dễ dàng thích nghi với các nền tảng công nghệ thông tin trong tương lai. Sản phẩm của Talend giảm đáng kể rào cản áp dụng cho các doanh nghiệp muốn đóng gói các giải pháp mạnh mẽ để thách thức hoạt động như tích hợp dữ liệu **(Data Integration**), chất lượng dữ liệu (**Data Quality)**, quản lý dữ liệu tổng thể (**Master Data Management - MDM**), xe buýt phục vụ doanh nghiệp (**Enterprise Service Bus - ESB**), quản lý qui trình kinh doanh (**Business Process Management - BPM**) và kho dữ liệu lớn (**Big Data**). Talend thúc đẩy và mở rộng công nghệ hàng đầu *Apache*. Mã nguồn mở Talend cho ESB và các giải pháp nguồn mở SOA giúp các tổ chức xây dựng linh hoạt kiến trúc doanh nghiệp hiệu suất cao đó là tích hợp và cho phép các dịch vụ ứng dụng phân tán.
* Không giống như các nhà cung cấp truyền thống cung cấp các giải pháp đóng cửa và rời rạc, Talend cung cấp một nền tảng mở và linh hoạt, được hỗ trợ bởi một mô hình thuê bao dựa trên giá trị dự đoán và khả năng mở rộng .

**b.** *Lịch sử ra đời của Talend:*

* Talend được thành lập vào năm 2005 bởi Bertrand Diard và Fabrice Bonan và được phát triển không ngừng đến ngày nay.



**Hình 1: Lịch sử hình thành và phát triển của Talend.**

* Talend là nhà cung cấp mã nguồn mở thương mại đầu tiên của các phần mềm tích hợp dữ liệu. Các nhà cung cấp khác có thể gia nhập thị trường này bao gồm **Apatar**, **Jitterbit**, **Pentaho**, và **SnapLogic**. Các nhà cung cấp tích hợp dữ liệu mã nguồn mở không bao gồm phần mềm **AG**, **Ab Initio**, **SAS Institute**, **[Pervasive](http://en.wikipedia.org/wiki/Pervasive_Software" \o "Pervasive Software)****[Software](http://en.wikipedia.org/wiki/Pervasive_Software" \o "Pervasive Software)**, **IBM**, **Informatica** và **SAP**.



**Hình 2: Một số đối tác của Talend.**

* Talend có trụ sở chính đặt tại thành phố Redwood, California, Talend có văn phòng tại Bắc Mỹ, châu Âu và châu Á và một mạng lưới quốc tế của các đối tác kỹ thuật và dịch vụ. Talend có hơn 400 nhân viên tại 14 văn phòng trên 7 quốc gia.
* Sản phẩm đầu tiên của công ty, Talend Open Studio f[or Data Integration](http://en.wikipedia.org/wiki/Talend_Open_Studio_for_Data_Integration), đã được đưa ra trong tháng 10 năm 2006, dưới tên trước đây của nó: Talend Open Studio. Vào tháng Giêng năm 2008, nó đã được tải về hơn 1 triệu lần. Vào tháng Giêng năm 2012, sản phẩm đạt 20 triệu tải và công ty đã có hơn 3500 khách hàng trên toàn thế giới. Ngày nay, Talend có hơn 4000 khách hàng. Khách hàng bao gồm eBay, Virgin Mobile, Sony Online Entertainment, Deutsche Post và Allianz……



**Hình 3: Những khách hàng đã lựa chọn sử dụng Talend.**

* Ngoài ra, Talend còn có những đối tác là các nhà sản xuất thiết bị (những nhà chế tạo, sửa chữa, đóng gói và bán ra thiết bị phần cứng) và các đối tác công nghệ.



**Hình 4: Các đối tác là các nhà sản xuất thiết bị.**

****

**Hình 5: Các đối tác là các nhà công nghệ của Talend.**

*c) Các sản phẩm của Talend.*

Talend hiện có 3 dòng sản phẩm:

*-Dòng sản phẩm quản lý dữ liệu.*

* Talend Open Studio for Big Data: kết hợp công nghệ dữ liệu lớn vào một môi trường mã nguồn mở thống nhất đơn giản hóa việc tải dữ liệu, khai thác, chuyển đổi, xử lý dữ liệu lớn và đa dạng.
* Talend Enterprise Big Data: một giải pháp tích hợp dữ liệu lớn mở rộng của Talend Open Studio cho dữ liệu lớn với khả năng làm việc nhóm và quản lý.
* Talend Platform for Big Data: mạnh mẽ, linh hoạt tích hợp dữ liệu lớn là giải pháp chất lượng dữ liệu mà đơn giản hoá việc tải, khai thác, xử lý dữ liệu lớn và đa dạng để bạn có thể đưa ra nhiều quyết định kịp thời hơn.
* Talend Open Studio for Data Integration: một ứng dụng mã nguồn mở cho công việc thiết kế tích hợp dữ liệu với một môi trường phát triển đồ họa, được thiết kế để kết hợp, chuyển đổi và cập nhật dữ liệu tại các địa điểm khác nhau trên toàn doanh nghiệp. Nó là một sản phẩm tích hợp dữ liệu của Talend cung cấp các tích hợp mạnh mẽ, linh hoạt nên các doanh nghiệp có thể ngừng lo lắng về cơ sở dữ liệu và các ứng dụng đang nói chuyện với nhau và thay vì tối đa hóa giá trị của việc sử dụng dữ liệu.
* Talend Enterprise Data Integration: là một sản phẩm mở rộng của Talend Open Studio for Data Integration với việc bổ sung phân phối, hợp tác phát triển các tính năng cũng như khả năng nắm bắt sự thay đổi dữ liệu. Trong khi các sản phẩm Talend Open Studio được thiết kế chủ yếu cho một nhà phát triển duy nhất làm việc trên một máy tính duy nhất, Enterprise cung cấp một kiến trúc phân tán cho phép nhiều nhà phát triển để chia sẽ tài nguyên và kết hợp các nỗ lực vào một dự án duy nhất.
* Talend Platform for Data Management: Giải pháp quản lý dữ liệu của Talend biến nguồn khác nhau, không thống nhất và bản sao của dữ liệu vào nguồn đáng tin cậy của thông tin hợp nhất, vì vậy doanh nghiệp của bạn có thể đáp ứng tốt hơn và tự tin hơn trong các quyết định hàng ngày.
* Talend Platform for Data Services: Tạo thành một dữ liệu thống nhất toàn diện, ứng dụng và giải pháp tích hợp dịch vụ giảm tác động của việc thay đổi cấu trúc dữ liệu trong khi thực hiện việc quản lý dữ liệu trên các lĩnh vực dễ dàng hơn.
* Talend Open Studio for MDM: MDM với một cách tiếp cận linh hoạt và mã nguồn mở để làm chủ các dự án dữ liệu. Nó cung cấp một nền tảng các chức năng quản lý dữ liệu tổng thể về tích hợp, chất lượng và làm chủ bất kỳ dữ liệu nào.
* Talend Platform for Master Data Management: Giải pháp quản lý dữ liệu tổng thể của Talend của biến các nguồn khác nhau, thông tin không phù hợp trên toàn doanh nghiệp của bạn thành một duy nhất, đáng tin cậy “phiên bản của sự thật”, gia tăng sự tự tin trong quyết định thực hiện của bạn.
* Talend Open Studio for Data Quality: một công cụ hồ sơ dữ liệu mã nguồn mở xem xét các nội dung, cấu trúc và chất lượng của các dữ liệu có cấu trúc phức tạp. Với giải pháp này, các tổ chức hiểu và cải thiện tính đầy đủ, chính xác và toàn vẹn của dữ liệu, vì vậy bạn phải tự tin hơn trong các thông tin bạn nhận được và quyết định thực hiện.

*-Dòng sản phẩm tích hợp ứng dụng.*

* Talend ESB Standard Edition: là một mã nguồn mở Apache dựa trên nguồn ESB.
* Talend Open Studio for ESB: là một xe buýt phục vụ doanh nghiệp và một lớp kết nối dựa trên tiêu chuẩn được sử dụng để tích hợp thêm hệ thống phân phối trên chức năng, doanh nghiệp và ranh giới địa lý. Nó được hổ trợ bởi các dự án tích hợp mã nguồn mở Apache CXF, Apache Camel và Apache ActiveMQ.
* Talend Enterprise ESB: là sản phẩm mở rộng của Talend Open Studio for ESB với việc triển khai và quản lý các chức năng tiên tiến hơn.
* Talend Platform for Enterprise Integration: quản lý thống nhất quy trình kinh doanh, tích hợp ứng dụng và dữ liệu để các doanh nghiệp tăng năng suất kinh doanh, cung cấp các dự án nhanh hơn, và chi phí vận hành thấp hơn.

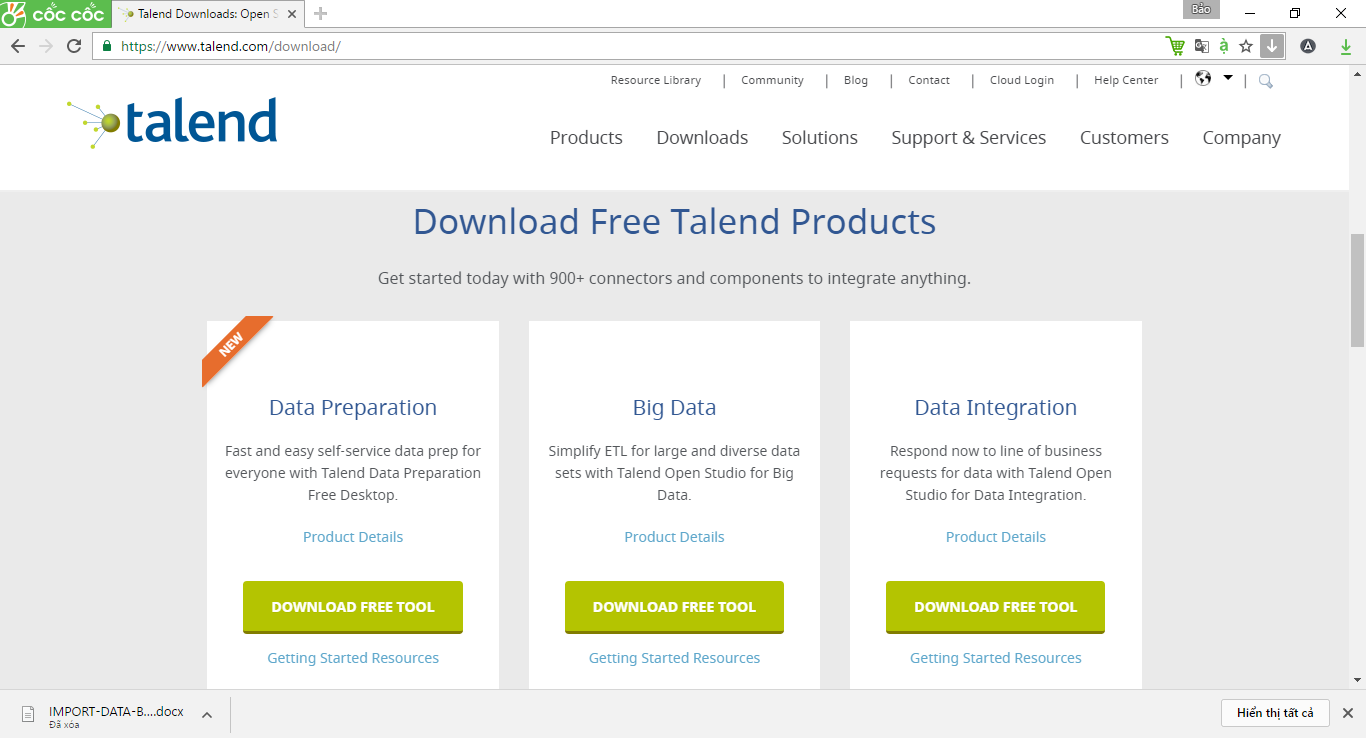
*-Dòng sản phẩm tích hợp ứng dụng dựa trên nền điện toán đám mây.*

* Talend Platform for Hybrid Cloud: Talend Cloud là một tập hợp toàn diện, thống nhất các ứng dụng tích hợp và các công cụ quản lý dữ liệu cho phép kết nối nhanh chóng, điều chỉnh và quản lý, ứng dụng dựa trên đám mây và SAAS trong một doanh nghiệp trả về kết quả nhanh hơn.

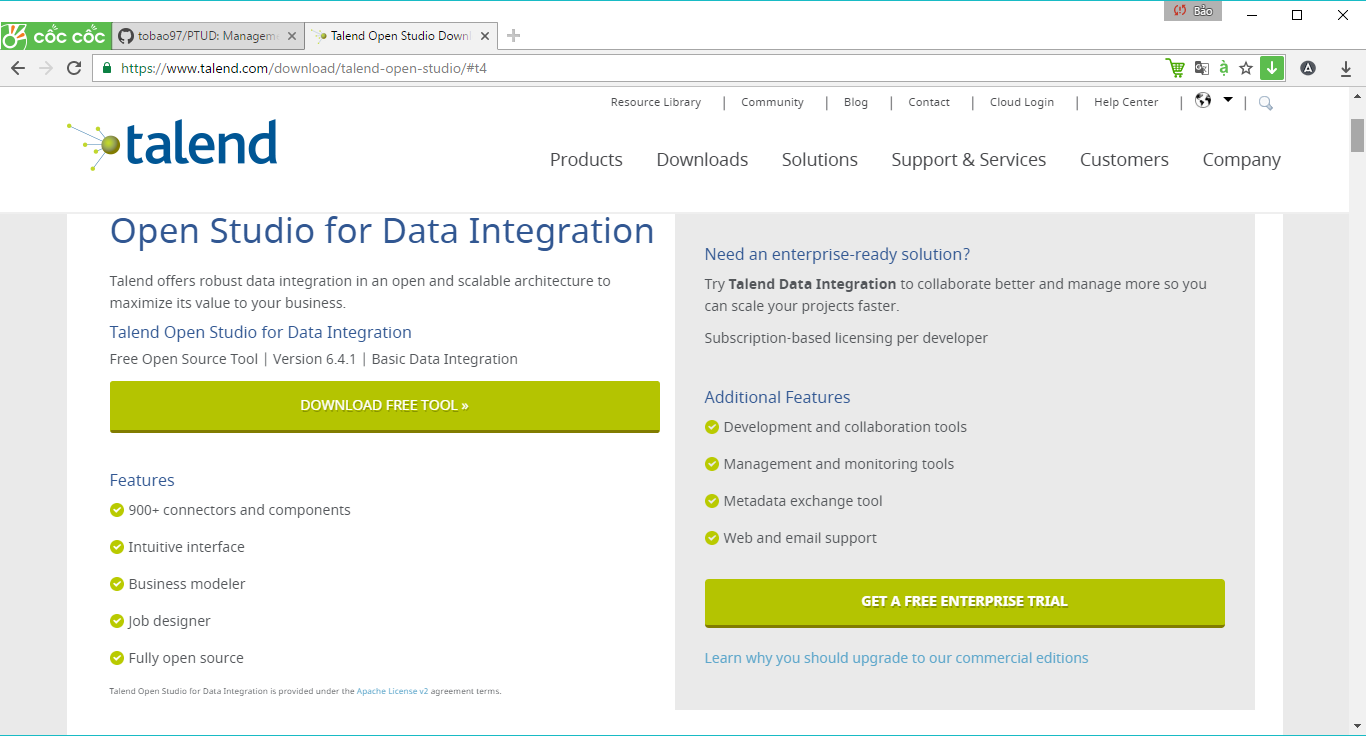
**2. Bắt đầu với một dự án Talend.**

**a.** *Cài đặt Talend.*

* Tải phần mềm Talend là một công việc đơn giản, Talend được tải về từ website [www.talend.com](http://www.talend.com/).

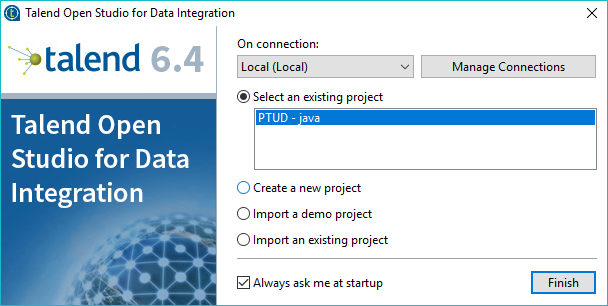


* Hơn thế nữa, tại trang web của Talend bạn có thể tải về các phiên bản phù hợp cho việc phát triển ứng dụng của bạn, nhưng cho dù bạn đang sử dụng Talend Open Studio for Big Data hoặc Talend Open Studio for Data Integration… thì tất cả các phát triển của bạn sẽ diễn ra trong môi trường Talend Studio.

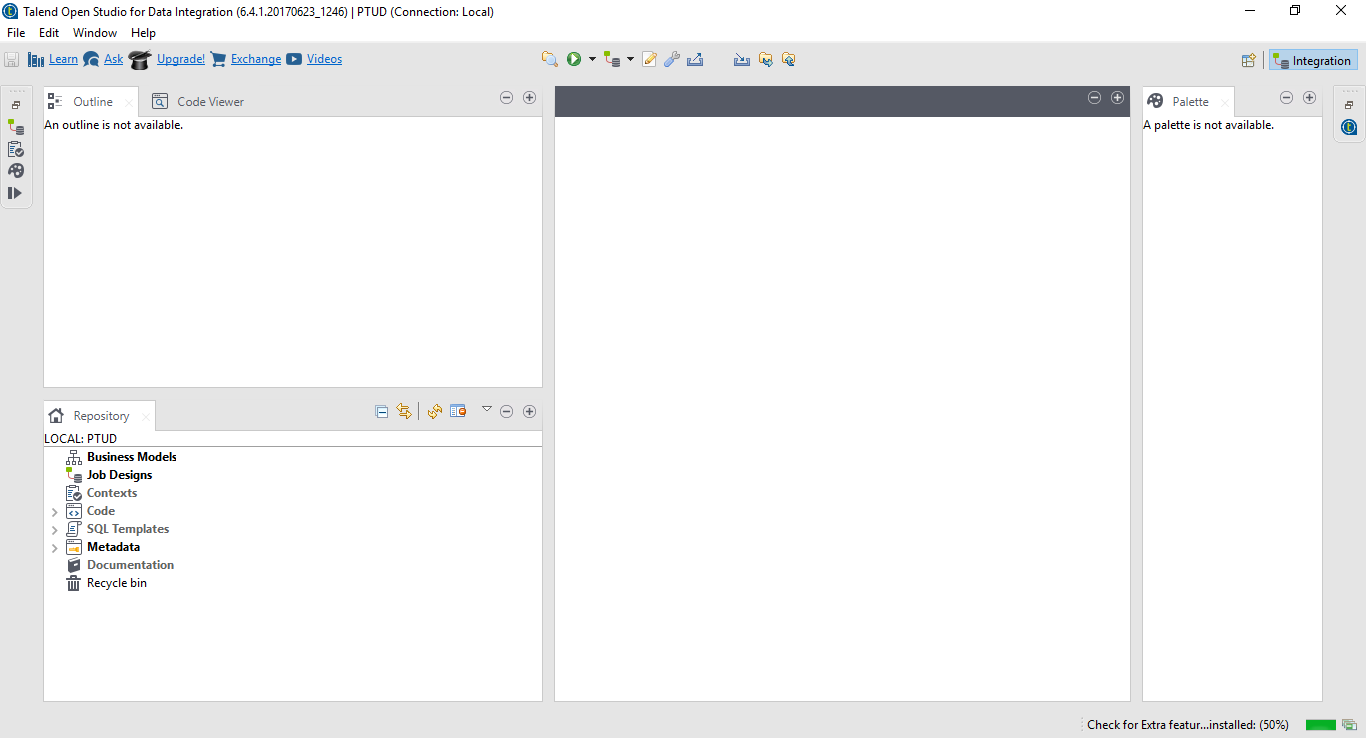


* Yêu cầu trước khi cài đặt : Phải cài đặt JDK(Java SE Develoment Kit) trước khi cài đặt Talend





* Giao diện của Talend



*b.Giới thiệu về Project trong Talend*

Trong Talend Studio,cấu trúc vật lý cao nhất để lưu trữ tất cả các loại công việc tích hợp dữ liệu khác nhau, Jobs, metadata,

thói quen, vv là "dự án".

Từ cửa sổ đăng nhập của Studio, bạn có thể:

+Tạo dự án

Khi bạn khởi chạy Studio lần đầu tiên, không có dự án mặc định được liệt kê. Bạn cần tạo một dự án

mà sẽ giữ tất cả tích hợp dữ liệu Jobs và các mô hình kinh doanh mà bạn thiết kế trong trường hợp hiện tại của Studio.

Bạn có thể tạo nhiều dự án như bạn cần để lưu trữ dữ liệu của bạn trong các trường hợp khác nhau của Studio của bạn.

Khi tạo một dự án mới, một thư mục cây được tạo tự động trong thư mục workspace trên kho của bạn

Máy chủ. Điều này sẽ tương ứng với khung nhìn Repository được hiển thị trên cửa sổ chính của Studio.

Để biết thêm thông tin, hãy xem Cách tạo dự án.

• Nhập dự án Demo để khám phá các tính năng của Talend Studio dựa trên các mẫu khác nhau sẵn sàng sử dụng

Việc làm. Khi bạn nhập dự án Demo, nó sẽ tự động được cài đặt trong thư mục không gian làm việc của hiện tại

phiên họp của Studio.

Để biết thêm thông tin, hãy xem Làm thế nào để nhập một dự án trình diễn.

• Nhập các dự án bạn đã tạo ra với phiên bản Talend Studio trước đây vào Studio Talend hiện tại của bạn

thư mục không gian làm việc.

Để biết thêm thông tin, hãy xem Cách nhập dự án.

• mở một dự án mà bạn đã tạo hoặc nhập vào Studio.

Để biết thêm thông tin, hãy xem Cách mở dự án.

• xóa các dự án địa phương mà bạn đã tạo ra hoặc nhập khẩu và bạn không cần nữa.

Để biết thêm thông tin, hãy xem Cách xóa dự án.

Một khi bạn khởi chạy Talend Studio, bạn có thể xuất các tài nguyên của một hoặc nhiều dự án được tạo ra trong hiện tại

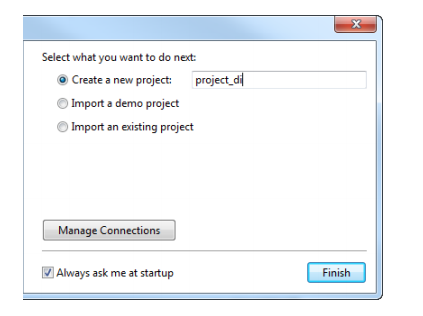
ví dụ của Studio. Để biết thêm thông tin, xem Cách xuất khẩu một dự án.

**\*Làm thế nào để tạo một dự án**

Để tạo một dự án tại thời điểm khởi động ban đầu của Studio, hãy thực hiện theo các bước sau:

B1: Khởi chạy Talend Studio.

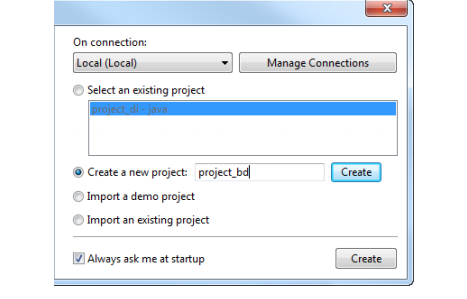
B2: Trên cửa sổ đăng nhập, chọn tùy chọn **Create a new project** và nhập tên dự án vào trường.



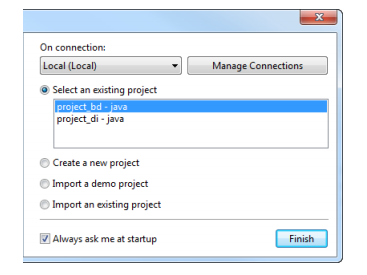
B3.:Nhấn Finish để tạo dự án và mở nó trong Studio.

Để tạo một dự án mới sau khi khởi động ban đầu của Studio, hãy thực hiện theo các bước sau:

B3.1: Trên cửa sổ đăng nhập, chọn tùy chọn **Create a new project** và nhập tên dự án vào trường.



B3.2: Nhấp vào **Create** để tạo dự án. Dự án mới được tạo ra được hiển thị trong danh sách các dự án hiện có



B3.3: Chọn dự án trong danh sách và nhấn Finish để mở dự án trong Studio.

Sau đó, nếu bạn muốn chuyển đổi giữa các dự án, trên thanh trình đơn Studio, sử dụng kết hợp File> Switch Project hoặc Workspace.

**\*Làm thế nào để Import một demo dự án**

Bạn có thể nhập một hoặc nhiều dự án demo bao gồm nhiều mẫu sẵn sàng để sử dụng Jobs vào Talend của bạn

Studio để giúp bạn hiểu được các chức năng của các thành phần Talend khác nhau.

Để nhập một dự án demo, tiến hành như sau:

B1: Khi phát hành Talend Studio của bạn, chọn **Import a demo project** vào cửa sổ đăng nhập Studio

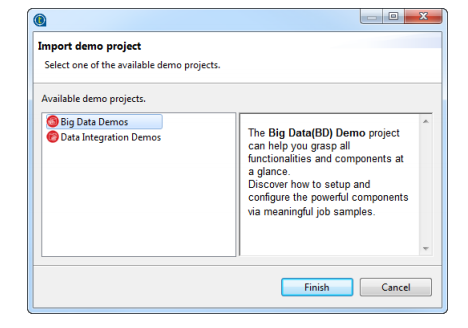
và nhấp vào Select, hoặc nhấp vào liên kết Demos trên cửa sổ chào mừng, để mở hộp thoại [Import demo project]

cái hộp.

Sau khi khởi chạy Studio, nhấp vào nút  trên thanh công cụ hoặc chọn **Help** > **Welcome**  từ trình đơn Studio

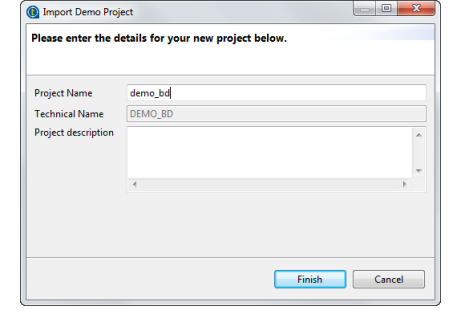
bar để mở cửa sổ chào mừng và sau đó nhấp vào liên kết Demos, để mở hộp thoại [Import demo project]

B2: Trong hộp thoại [Import Demo Project], chọn dự án demo bạn muốn import và xem mô tả trên bảng điều khiển bên phải. Các dự án demo có sẵn trong hộp thoại có thể khác nhau tùy thuộc vào sản phẩm bạn đang sử dụng.



B3:Nhấn Finish để đóng hộp thoại

B4: Trong hộp thoại mới mở ra, nhập tên dự án mới và thông tin mô tả nếu cần.



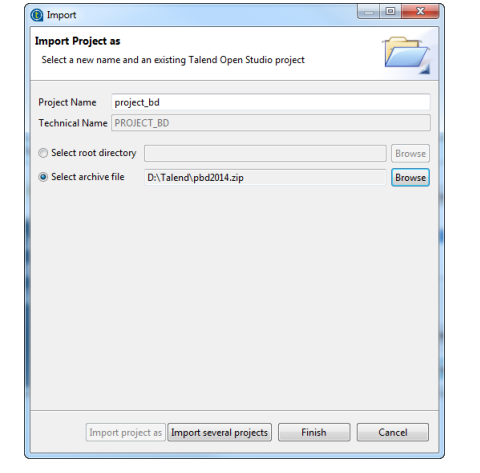
B5: Nhấn Finish để tạo dự án. Tất cả các mẫu của dự án demo được nhập khẩu vào dự án vừa mới tạo, và tên của dự án mới dự án được hiển thị trong danh sách Dự án trên màn hình đăng nhập. Cách nhập dự án 10 Talend Open Studio để Tích hợp Dữ liệu Hướng dẫn sử dụng

B6: Để mở dự án giới thiệu nhập khẩu trong Talend Studio, trở lại cửa sổ đăng nhập, chọn nó từ Dự án danh sách và sau đó nhấp vào Kết thúc. Các mẫu Job trong dự án demo mở sẽ được tự động nhập vào thư mục workspace của bạn và được tạo sẵn trong khung nhìn Repository tree dưới thư mục Job Designs.

**\*Cách import một dự án**

Trong Talend Studio, bạn có thể nhập một hoặc nhiều dự án mà bạn đã tạo ra với phiên bản trước của Studio. Để nhập một dự án đơn lẻ, hãy thực hiện theo các bước sau:

B1; Từ cửa sổ đăng nhập Studio, chọn Import an existing project rồi nhấp vào select để mở phần [Import]



B2: Nhấp vào **Import project as** và nhập tên cho dự án mới của bạn trong trường **Project Name**.

B3: Nhấp **Select root directory** hoặc **Select archive file** tùy thuộc vào nguồn mà bạn muốn nhập từ.

B4.:Nhấn **Browse...** để chọn thư mục không gian làm việc / tập tin lưu trữ của thư mục dự án cụ thể. Theo mặc định,vùng làm việc được lựa chọn là một trong những phiên bản hiện tại. Duyệt đến để đạt đến không gian làm việc phát hành trước

thư mục hoặc tệp lưu trữ có chứa các dự án nhập khẩu.

B5: Nhấn **Finish** để xác nhận hoạt động và trở lại cửa sổ đăng nhập.

***\*Để import một số dự án đồng thời, hãy thực hiện theo các bước sau:***

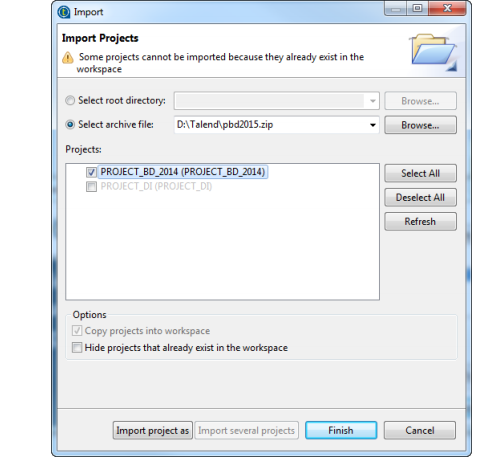
B1.Từ cửa sổ đăng nhập Studio, chọn **Import an existing project** rồi nhấp vào **Select** để mở phần **[Import]** wizard.

B2. Nhấp vào **Import several projects**.

B3. Nhấp **Select root directory** or **Select archive file** tùy thuộc vào nguồn mà bạn muốn nhập từ.

B4. Nhấn **Browse...** để chọn thư mục không gian làm việc / tập tin lưu trữ của thư mục dự án cụ thể. Theo mặc định, vùng làm việc được lựa chọn là một trong những phiên bản hiện tại. Duyệt đến để đạt đến không gian làm việc phát hành trước

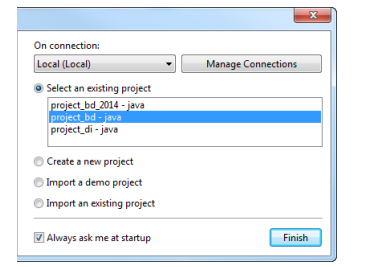
thư mục hoặc tệp lưu trữ có chứa các dự án nhập khẩu.



B5. Chọn **Copy projects into workspace** để tạo bản sao của dự án nhập khẩu thay vì di chuyển Nó. Tùy chọn này chỉ khả dụng khi bạn nhập một số dự án từ thư mục gốc. Nếu bạn muốn xóa các thư mục dự án ban đầu khỏi thư mục không gian làm việc Talend Studio bạn nhập từ, xóa hộp kiểm này. Nhưng chúng tôi khuyên bạn nên giữ nó cho mục đích sao lưu. Làm thế nào để mở một dự án 12 Talend Open Studio để Tích hợp Dữ liệu Hướng dẫn sử dụng

B6. Chọn **Hide projects that already exist in the workspace** trong không gian làm việc để ẩn các dự án đã có từ Danh sách **Project**. Tùy chọn này chỉ khả dụng khi bạn nhập một vài dự án.

B7. Từ danh sách Các dự án, chọn các dự án cần nhập và nhấp vào finish để xác nhận hoạt động. Khi Import dự án thành công, tên của các dự án Import được hiển thị trong danh sách Dự án của việc đăng nhập cửa sổ



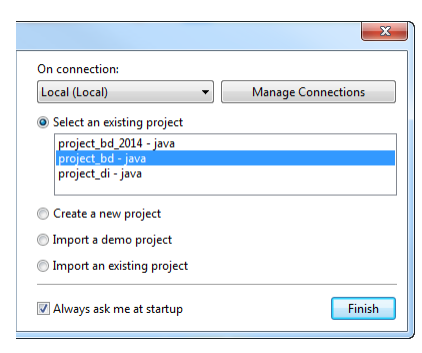


\***Làm thế nào để mở một dự án.**

Khi bạn khởi chạy Talend Studio lần đầu tiên, không có tên dự án nào được hiển thị trong danh sách Dự án. Đầu tiên bạn cần tạo một dự án hoặc nhập một dự án trình diễn để đưa vào Danh sách dự án với các tên dự án tương ứng mà bạn có thể mở trong Studio.

Để mở một dự án trong Studio Talend:

Trên màn hình đăng nhập Studio, chọn dự án quan tâm từ danh sách dự án và nhấp vào Finish.



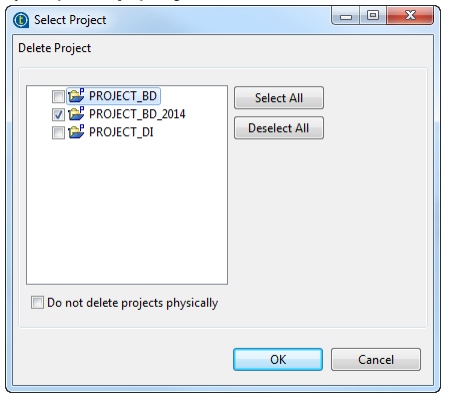
Thanh tiến trình xuất hiện. Chờ cho đến khi công việc hoàn thành và cửa sổ chính của Talend Studio sẽ mở ra.



Khi bạn mở dự án được nhập từ phiên bản trước của Studio, một cửa sổ thông tin xuất hiện để liệt kê một đoạn ngắn mô tả các nhiệm vụ di chuyển thành công.

**\*Làm thế nào để xóa một dự án**

B1.Trên màn hình đăng nhập, nhấp **Manage Connections**, sau đó trên hộp thoại mở ra nhấp **Delete Existing** **Project(s)** để mở hộp thoại **[Select Project]**.



B2.Chọn (các) hộp kiểm của (các) dự án mà bạn muốn xóa.

B3. Nhấn OK để xác nhận việc xóa.

Danh sách dự án trên cửa sổ đăng nhập được làm mới cho phù hợp.



Hãy cẩn thận, hành động này là không thể đảo ngược. Khi bạn nhấp vào OK, không có cách nào để hồi phục dự án đã xóa.

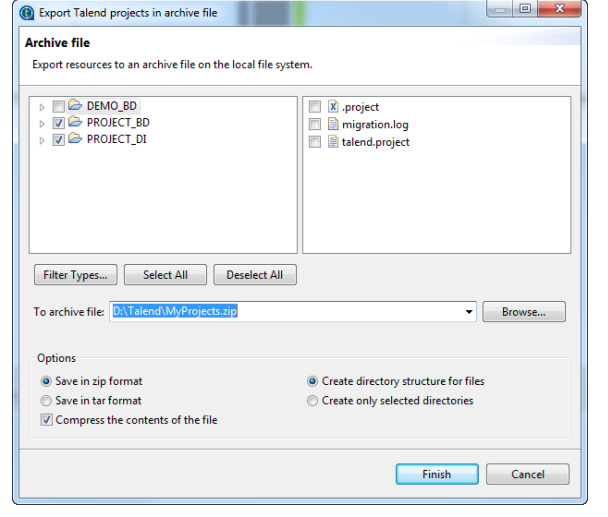


Nếu bạn chọn **Do not delete projects physically**, bạn chỉ có thể xóa các dự án đã chọn khỏi danh sách dự án và vẫn có nó / họ trong thư mục không gian làm việc của Talend Studio. Vì vậy, bạn có thể phục hồi các xóa (các) dự án bất kỳ lúc nào bằng cách sử dụng **Import existing project(s) as local** làm tùy chọn cục bộ trong danh sách Dự án từ cửa sổ đăng nhập.

**\*Làm thế nào để Export một dự án**

Talend Studio cho phép bạn xuất các dự án được tạo ra hoặc nhập khẩu trong trường hợp hiện tại của Talend Studio.

1.Trên thanh công cụ của cửa sổ chính của Studio, nhấp  để mở các dự **[Export Talend projects in archive file]** trong hộp thoại.



2.Chọn hộp kiểm của dự án bạn muốn xuất. Bạn chỉ có thể lựa chọn các phần của dự án liên kết **Filter Types...**., nếu cần (cho người dùng cấp cao).

3. Trong trường **To archive file**, nhập tên của hoặc duyệt đến tệp lưu trữ mà bạn muốn xuất khẩu các dự án được lựa chọn

4. Trong vùng Option, chọn định dạng nén và kiểu cấu trúc mà bạn muốn.

5. Nhấn vào Finish để xác nhận các thay đổi.

Tệp lưu trữ chứa các dự án đã xuất được tạo ở nơi đã xác định.

**3.Thiết kế** Business Model **(mô hình kinh doanh)**

Quan điểm hội nhập của Talend Studio cung cấp công cụ tốt nhất để chính thức hóa mô tả kinh doanh trong việc xây dựng khối và các mối quan hệ của họ. Talend Studio cho phép thiết kế hệ thống, kết nối, quy trình và yêu cầu sử dụng ký hiệu quy trình làm việc chuẩn thông qua một thư viện đồ họa trực quan về hình dạng và liên kết.

Chương này nhằm vào các nhà quản lý kinh doanh, các nhà hoạch định chính sách hoặc các nhà phát triển những người muốn mô hình quản lý lưu lượng của họ cần ở cấp độ vĩ mô.

*a)Business Model**là gì:*

Các mô hình kinh doanh của Talend cho phép các bên tham gia dự án tích hợp dữ liệu biểu diễn bằng đồ hoạ các nhu cầu của họ bất kể yêu cầu thực hiện kỹ thuật. Mô hình Kinh doanh giúp nhân viên CNTT hiểu những nhu cầu này đã thể hiện và chuyển chúng thành quy trình kỹ thuật (Việc làm). Chúng thường bao gồm cả hệ thống và quy trình đã hoạt động trong doanh nghiệp, cũng như những quy trình cần thiết trong tương lai.

Thiết kế mô hình kinh doanh là một phần của các thực tiễn tốt nhất của doanh nghiệp mà các tổ chức nên áp dụng từ rất sớm giai đoạn của một dự án tích hợp dữ liệu để đảm bảo sự thành công của nó. Bởi vì các mô hình kinh doanh thường giúp phát hiện và giải quyết nhanh chóng các nút thắt cổ chai và điểm yếu, chúng giúp hạn chế sự bội chi ngân sách và / hoặc giảm đầu tư trả trước. Sau đó, trong và sau khi thực hiện dự án, mô hình kinh doanh có thể được xem xét và chỉnh sửa để phản ánh bất kỳ thay đổi được yêu cầu.

Một mô hình kinh doanh là một cái nhìn phi kỹ thuật về nhu cầu công việc kinh doanh.

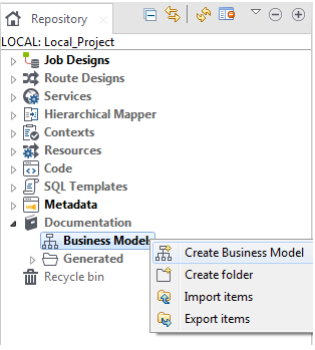
Nói chung, một mô hình kinh doanh điển hình sẽ bao gồm các hệ thống chiến lược hoặc quy trình đã có và chạy trong công ty cũng như nhu cầu mới. Bạn có thể biểu tượng cho các hệ thống, quy trình và nhu cầu sử dụng nhiều hình dạng và tạo ra các kết nối giữa chúng. Có khả năng, tất cả chúng có thể dễ dàng mô tả bằng cách sử dụng thuộc tính kho và định dạng công cụ.

Trong không gian làm việc thiết kế của Quan điểm tích hợp của Talend Studio, bạn có thể sử dụng nhiều công cụ để:

• Rút ra nhu cầu kinh doanh của bạn,

• Tạo và gán nhiều kho cho các đối tượng mô hình của bạn,

• Xác định các thuộc tính mô hình kinh doanh của các đối tượng mô hình của bạn.

*b) Opening or creating a Business Model*

Mở Talend Studio theo thủ tục được nêu chi tiết trong Hướng dẫn Bắt đầu.

Trong khung Repository tree của Integration perspective, kích chuột phải vào Documentation> Business Models Nodes

Chọn Expand/Collapse để hiển thị tất cả các Business Models hiện tại (nếu có)

**\* How to open a Business Model**

-Nhấp đúp vào tên của the Business Model

-The Business Model đã chọn mở ra trên không gian làm việc thiết kế

**\*. How to create a Business Model**

**- Nhấn chuột phải vào nút Business Models và chọn Create Business Model.**The creation wizard hướng dẫn bạn qua các bước để tạo ra một Business Model mới

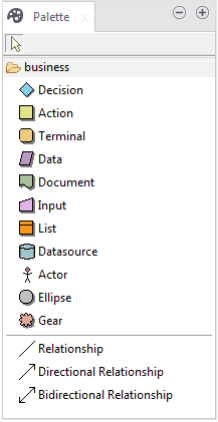
- Nhập các thuộc tính Business Model : **Name,** **Purpose,** **Description,** **Author,** **Locker,** **Version,** **Status,** **Path**

**-** The **Modeler mở ra trên không gian làm việc thiết kế rỗng. Bạn có thể tạo ra nhiều mô hình như bạn muốn và mở tất cả**



*c) Modeling a Business Model*

Nếu bạn có nhiều tab được mở trên không gian làm việc thiết kế của bạn, hãy nhấp vào tab có liên quan để hiển thị thích hợp thông tin mô hình.

Trong giao diện Business Model , bạn có thể xem thông tin liên quan đến mô hình hoạt động. Sử dụng Palette để thả các hình dạng có liên quan trên không gian làm việc thiết kế và kết nối chúng cùng với các chi nhánh và sắp xếp hoặc cải tiến mô hình hình ảnh bằng cách phóng to hoặc thu nhỏ.

Palette này cung cấp các đại diện đồ họa cho các đối tượng tương tác trong Business Model. Các đối tượng có thể có các loại khác nhau, từ hệ thống chiến lược đến tài liệu đầu ra hoặc bước quyết định. Mỗi người có một vai trò cụ thể trong mô hình kinh doanh của bạn theo mô tả, định nghĩa và nhiệm vụ bạn đưa ra cho nó.

Tất cả các đối tượng được thể hiện trong Palette dưới dạng hình dạng, và có thể được bao gồm trong mô hình.

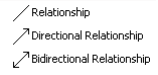
Lưu ý rằng bạn phải nhấp vào thư mục business để hiển thị thư viện hình dạng trên Palette.

**\*Shapes**

Chọn hình dạng tương ứng với đối tượng liên quan mà bạn muốn đưa vào Business Model của bạn. Nhấn đúp chuột nó hoặc nhấp vào hình dạng trong Palette và thả nó vào khu vực mô hình hóa.

Ngoài ra, để truy cập nhanh vào thư viện hình dạng, hãy giữ con trỏ ở khu vực mô hình cho một vài giây để hiển thị thanh công cụ truy cập nhanh:

**\* Connecting shapes**

Khi thiết kế, bạn muốn thực hiện mối quan hệ giữa một hình dạng nguồn và một hình dạng đích

Có hai cách để kết nối hình dạng trong không gian làm việc thiết kế của bạn:

\*Cách làm

B1. Đơn giản chỉ cần di chuyển con trỏ chuột qua một hình dạng mà bạn đã bỏ trên không gian làm việc thiết kế của bạn, theo thứ tự

để hiển thị các mũi tên kết nối đôi.

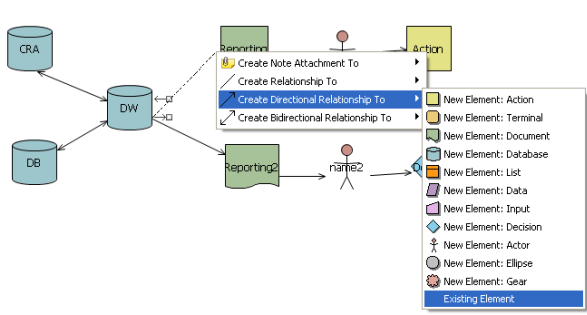
B2. Chọn mũi tên liên quan để thực hiện kết nối hướng chính xác nếu cần thiết. Kết nối hình dạng Studio Mở Talend để Tích hợp Dữ liệu Hướng dẫn sử dụng 23

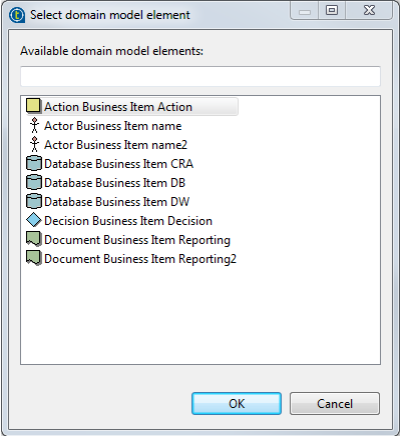
B3. Kéo một liên kết tới một vùng trống của không gian làm việc thiết kế và phát hành để hiển thị kết nối bật lên

thực đơn.

B4. Chọn kết nối phù hợp từ danh sách. Bạn có thể chọn giữa Create Relationship To, Create Directional Relationship To hoặc Create Bidirectional Relationship To.

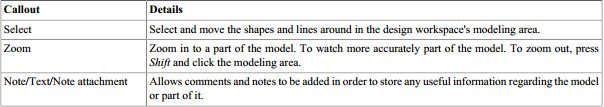
B5. Sau đó, chọn phần tử thích hợp để kết nối, trong số các mục được liệt kê.



Bạn có thể tạo kết nối với một phần tử hiện có của mô hình. Chọn Existing Element trong trình đơn bật lên và chọn phần tử hiện có mà bạn muốn kết nối trong hộp danh mục hiển thị.

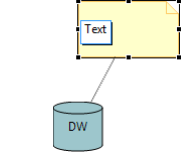
**\* How to comment and arrange a model**

**Các công cụ của Palette cho phép bạn tùy chỉnh mô hình của bạn:**



**\*How to add a note or free text**

Để thêm một ghi chú, chọn biểu tượng Ghi chú trong Palette, được gắn ở bên phải của không gian làm việc thiết kế.

Ngoài ra, nhấp chuột phải vào mô hình hoặc hình dạng mà bạn muốn liên kết ghi chú đó, và chọn Thêm Ghi chú. Hoặc chọn

Lưu ý công cụ trong thanh công cụ truy cập nhanh.

Một lưu ý dính sẽ hiển thị trên khu vực mô hình hóa. Nếu ghi chú được liên kết với một hình dạng đặc biệt, một dòng sẽ được tự động

vẽ theo hình dạng.

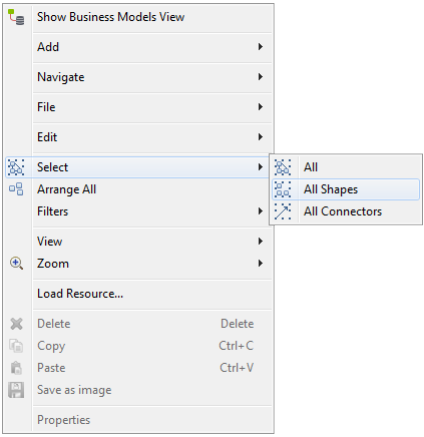
Gõ vào văn bản trong hộp đầu vào hoặc, nếu thứ hai không hiển thị, gõ trực tiếp vào lưu ý dính.

Nếu bạn muốn liên kết ghi chú và hình dạng cụ thể của mô hình của mình, hãy nhấp vào mũi tên xuống bên cạnh công cụ Ghi chú trên

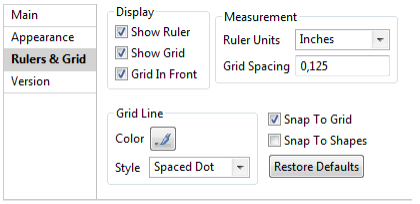
Palette và chọn Note attachment. Kéo mũi tên màu đen về phía một vùng trống của không gian làm việc thiết kế, và giải phóng. Trình đơn bật lên cung cấp cho bạn để đính kèm một Chú ý mới cho hình dạng được lựa chọn. Bạn cũng có thể chọn tính năng Thêm văn bản để nhập văn bản miễn phí trực tiếp vào khu vực mô hình hóa. Bạn có thể truy cập tính năng trong trình đơn thả xuống Ghi chú của Palette hoặc thông qua một phím tắt nằm bên cạnh tính năng Thêm Ghi chú trên

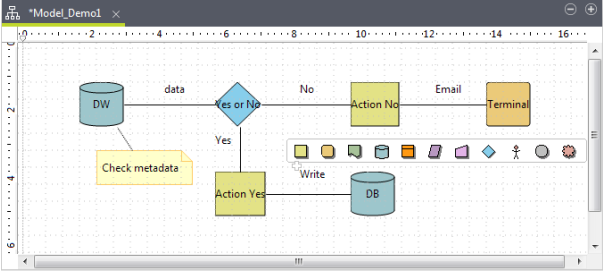
thanh công cụ truy cập nhanh

**\* How to arrange the model view**



**\* Rulers and Grid tab**





*d)Editing a Business Model*

**\* Đổi tên một Business Model**

B1. Nhấp chuột phải vào nhãn mô hình kinh doanh liên quan trên khung nhìn **Repository** tree sau đó chọn **Edit properties** hiển thị thông tin thuộc tính **Main** tương ứng trong hộp thoại [Chỉnh thuộc tính].

B2. Chỉnh sửa tên mô hình trong trường **Name**, sau đó nhấp vào **Finish** để đóng hộp thoại. Nhãn mô hình thay đổi tự động trong khung nhìn Repository tree và sẽ được phản ánh trên tab model của không gian làm việc thiết kế,lần tiếp theo bạn mở mô hình kinh doanh.Nếu Mô hình Kinh doanh mở, thông tin trong hộp thoại [Chỉnh sửa thuộc tính] sẽ là chỉ đọc, vì vậy bạn sẽ không có thể chỉnh sửa chúng.

**\*Copy and paste a Business Model**

B1. Trong **Repository > Business model**, nhấp chuột phải vào tên Mô hình Kinh doanh để Copy và chọn Sao chép trong phần trình đơn bật lên hoặc nhấn **Ctrl + C.**

B2. Nhấp chuột phải vào nơi bạn muốn dán Model Kinh doanh của bạn, và chọn Paste.

**\*Di chuyển Business Model**

B1. Để di chuyển mô hình kinh doanh từ vị trí này sang vị trí khác trong thư mục dự án mô hình kinh doanh của bạn, hãy chọn một Business  
Model trong **Repository > Business Models**

B2. Sau đó chỉ cần thả nó vào vị trí mới.

**\*Xóa một Business Model**

B1. Nhấp chuột phải vào tên của mô hình cần xóa và chọn **Delete** trong popup menu.

B2. Ngoài ra, chỉ cần chọn mô hình kinh doanh có liên quan, sau đó bỏ nó vào Thùng rác(**Recycle bin)** của **Repository** tree view.

e) Lưu một Business Model

Nhấn chọn **File > Save** hoặc nhấn tổ hợp phim **Ctrl+S**

Để lưu một mô hình kinh doanh và tăng phiên bản của nó cùng một lúc:

1. Chọn **File>Save as...**. một hộp thoại Save As xuất hiện
2. Bên cạnh trường Version, nhấp vào nút M để tăng phiên bản chính và nút m để tăng phiên bản phụ.

3. Nhấp vào Finish để xác nhận tính sửa đổi

**4.Thế nào là Job Design ?**

Là lớp runnable của Business Model . Nó là một thiết kế đồ họa, của một hoặc nhiều thành phần kết nối với nhau, cho phép bạn thiết lập và chạy quy trình quản lý luồng dữ liệu. Một công việc thiết kế chuyển đổi nhu cầu kinh doanh vào mã code , thói quen và các chương trình, nói cách khác nó kỹ thuật thực hiện luồng dữ liệu của bạn.

Công việc bạn thiết kế có thể giải quyết tất cả các nguồn và mục tiêu khác nhau mà bạn cần cho quá trình tích hợp dữ liệu và bất kỳ quá trình liên quan khác.Khi bạn thiết kế một Job trong Talend Studio, bạn có thể:

• Đưa ra các hành động tích hợp dữ liệu bằng cách sử dụng một thư viện các thành phần kỹ thuật.

• Thay đổi cài đặt mặc định của các thành phần hoặc tạo các thành phần mới hoặc gia đình của các thành phần phù hợp với nhu cầu chính xác.

• Thiết lập kết nối và mối quan hệ giữa các thành phần để xác định trình tự và bản chất của hành động.

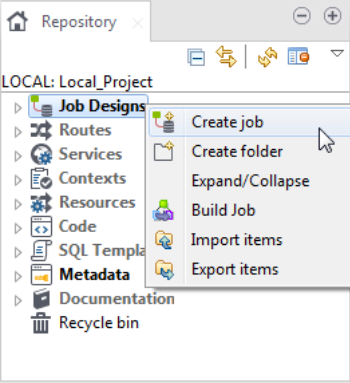
• Truy cập mã vào bất kỳ lúc nào để chỉnh sửa hoặc ghi lại các thành phần trong Job được thiết kế.

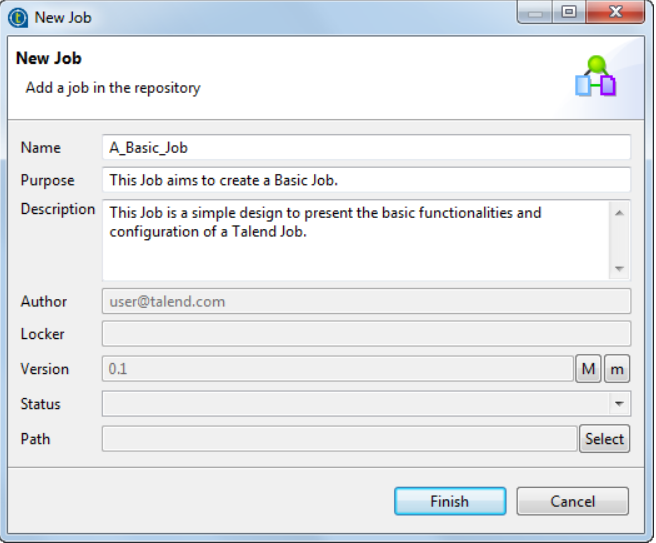
• Tạo và thêm các mục vào kho lưu trữ để sử dụng lại và chia sẻ mục đích (trong các dự án khác hoặc công việc hoặc với các người dùng).

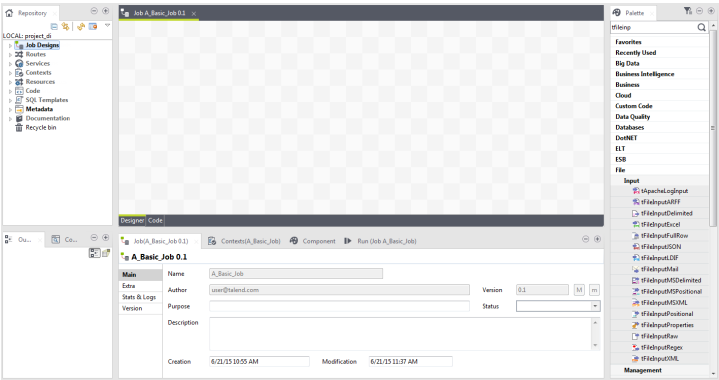
**5.Bắt đầu với 1 Job cơ bản:**

*a.Tạo 1 Job:*

Sau khi xây dựng một dự án. Chúng ta bắt đầu xây dựng một **Job** - đơn vị thực thi  
bất cứ những gì mà bạn đang xây dựng trong Talend đó là **Job**. Về bản chất,  
**Talend Job** là một *Class Java* đơn. Job ghi nhớ các giá trị điều này giúp cho việc  
hiểu công việc của bạn như thế nào, phạm vi thông tin có sẵn cho bạn cũng như  
phát triển thông tin đó.  
Job bao gồm một hoặc nhiều **Job** (*SubJobs*). Một *SubJob* là một thành phần hoặc  
một số thành phần được kết hợp bởi dòng dữ liệu (*dataflow*). Một *Job* sẽ có ít nhất  
một *SubJob*.  
**Tạo mới một Talend Job**: Để tạo ra một **Talend Job** mới, *right-click* **Job  
Designs** trong khung *Talend Repository Browser* và lựa chọn **Create Job** từ  
*Context Menu*.



Một hộp thoại xuất hiện :

Nhập thông tin các Field : Name ( tên job) , Purpose(mục đích của job) , Description(Mô tả về job) . Nhấn **Finish** để kết thúc công việc tạo Job

*b.Thêm các phần trong Job*

Khi Job được tạo ra, các thành phần phải được thêm vào không gian làm việc thiết kế: tFileInputDelimited , tLogRow và tFileOutputDelimited

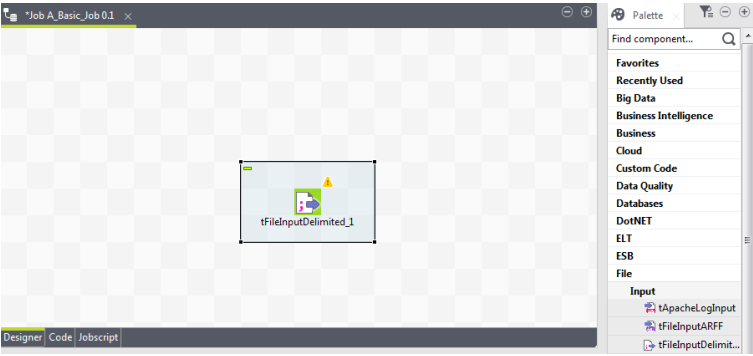
\* Dropping the first component from the Palette

B1: Nhập (các) từ khoá tìm kiếm vào trường tìm kiếm của Palette và nhấn Enter để xác thực tìm kiếm của bạn.

Từ khóa (s) có thể là một phần hoặc toàn bộ tên của thành phần, hoặc một cụm từ mô tả chức năng của nó nếu bạn không biết tên của nó, ví dụ, tfileinputde, fileinput, hoặc read file row by row

B2:Chọn thành phần mà bạn muốn sử dụng và nhấp vào không gian làm việc thiết kế mà bạn muốn thả thành phần.

Lưu ý rằng bạn cũng có thể thả một ghi chú vào Job giống như cách bạn thả các thành phần. Mỗi thành phần mới được thêm vào được hiển thị trong một hộp màu xanh để cho thấy rằng nó như là một cá nhân Subjob



,,c c

**\* Adding the second component by typing on the design workspace**

Thành phần thứ hai của Job của chúng ta sẽ được thêm vào bằng cách gõ tên của nó trực tiếp vào workspace, thay vì thả nó từ Palette hoặc từ nút Metadata.

Để thêm một thành phần trực tiếp vào không gian làm việc, hãy thực hiện như sau:

B1: Nhấp vào nơi bạn muốn thêm các thành phần trên không gian làm việc thiết kế, và gõ từ khóa của bạn, có thể

là tên đầy đủ hoặc một phần của hợp phần, hoặc một cụm từ mô tả chức năng của nó nếu bạn không biết Tên. Trong ví dụ của chúng tôi, bắt đầu nhập tlog.

B2: Nhấp đúp vào thành phần mong muốn để thêm nó vào không gian làm việc, tLogRow trong ví dụ của chúng ta

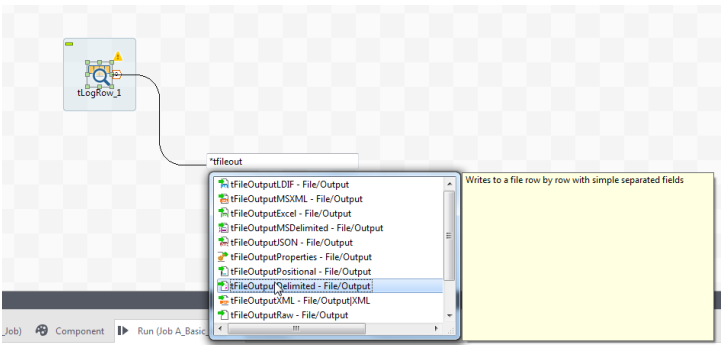
\*Adding an output component by dragging from an inputone

Bây giờ bạn sẽ thêm thành phần thứ ba, một tFileOutputDelimited, để ghi dữ liệu được đọc từ tệp nguồn vàomột tập tin văn bản khác. Chúng ta sẽ thêm thành phần này bằng cách kéo từ thành phần tLogRow, nó đóng vai trò đầu vào thành phần mới sẽ được thêm vào.

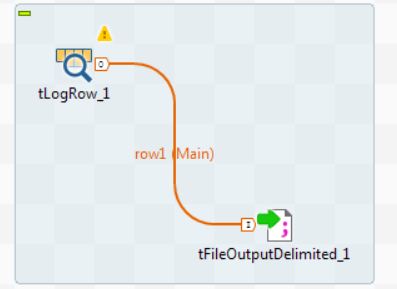
B1: Nhấp vào thành phần **tLogRow** để hiển thị biểu tượng **o** được gắn vào nó.

2. Kéo và thả biểu tượng **o** nơi bạn muốn thêm một thành phần mới.

Một trường văn bản và một danh sách thành phần xuất hiện. Danh sách thành phần hiển thị tất cả các thành phần có thể được kết nối với các thành phần đầu vào.



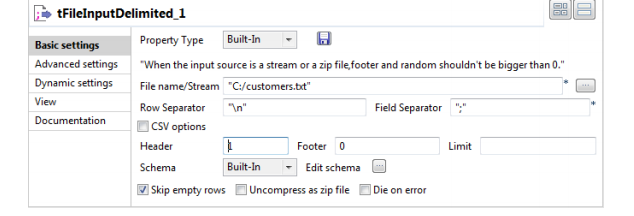
B3. Để thu hẹp tìm kiếm, nhập trường văn bản tên của thành phần mà bạn muốn thêm hoặc một phần của nó, hoặc một cụm từ mô tả chức năng của thành phần nếu bạn không biết tên của nó, và sau đó nhấp đúp vào thành phần tFileOutputDelimited trong ví dụ này, trên danh sách thành phần để thêm nó vào không gian làm việc thiết kế.Thành phần mới được tự động kết nối với thành phần đầu vào tLogRow, sử dụng a Row > Main connection



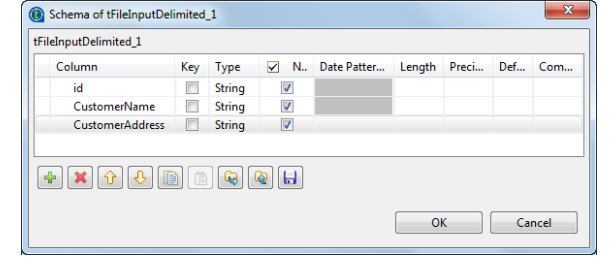
*c.Cấu hình thành phần*

\*Cấu hình thành phần **tFileInputDelimited**

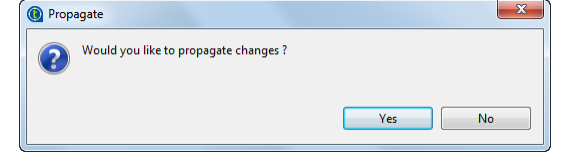
1. Nhấp đúp vào thành phần **tFileInputDelimited** để mở khung nhìn **Basic settings.**



1. Nhấp chuột vào nút [...] bên cạnh trường **File Name / Stream**.
2. Duyệt hệ thống của bạn hoặc nhập đường dẫn đến tập tin đầu vào, có đuôi là .txt .
3. Trong **Header** nhập 1.
4. Nhấp chuột vào nút […]bên cạnh **Edit schema.**
5. Trong **Schema Editor** được mở ra, nhấp ba lần nút [**+**] để thêm ba cột
6. Đặt tên cho ba cột id, CustomerName và CustomerAddress tương ứng và nhấn **OK** để đóng



Điều này cho phép bạn sao chép lược đồ bạn đã tạo cho thành phần tiếp theo, **tLogRow** trong ví dụ này

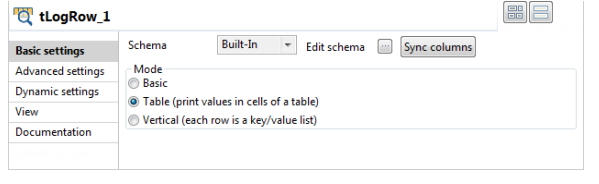


Cấu hình thành phần **tLogRow**

1. Nhấp đúp vào thành phần **tLogRow** để mở khung nhìn **Basic settings.**

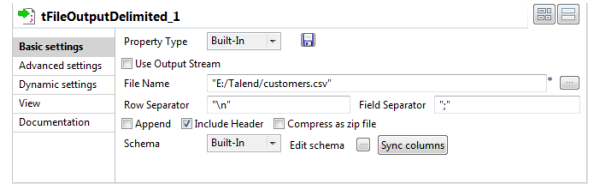
2. Trong vùng **Mode**, chọn **Table (print values in cells of a table).**

Bằng cách làm như vậy, nội dung của tệp tin co đuôi .txt sẽ được in trong một bảng và do đó dễ đọc hơn.



\*Cấu hình thành phần **tFileOutputDelimited**

1. Nhấp đúp vào thành phần t**FileOutputDelimited** để mở khung nhìn **Basic settings.**



1. Nhấp vào nút [...] bên cạnh **File Name**

3. Duyệt hệ thống của bạn hoặc nhập đường dẫn đến tệp tin đầu ra, file csv.

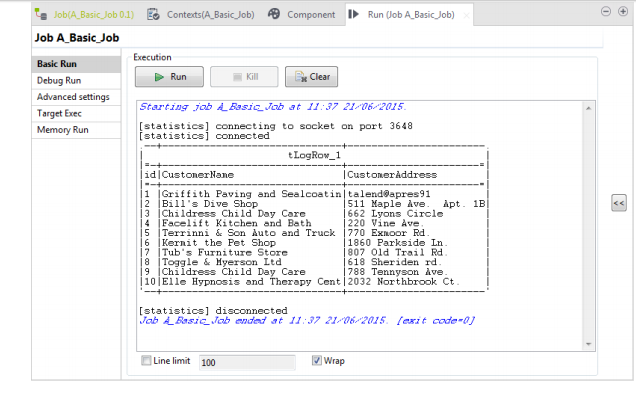
4. Chọn **Include Header**

5. Nếu cần, nhấp vào **Sync columns** để lấy lược đồ từ thành phần đầu vào.

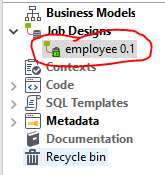
*d.Thực hiện Job:*

1. Nhấn **Ctrl + S** để lưu job

2. Vào tab **Run**, và nhấp vào Run để thực hiện công việc.



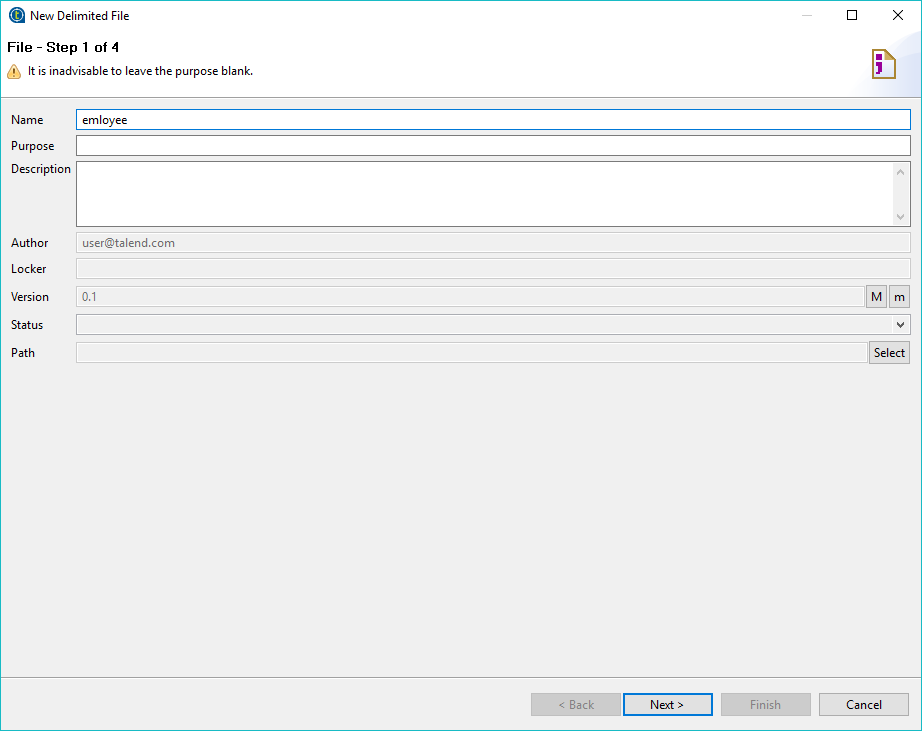
*e)Demo :*

 \*Đọc và hiển thị dữ liệu từ 1 file

Đầu tiên ta tạo một Job có tên **employee** và thêm các thành phần **tFileInputDelimited** vào trong job :

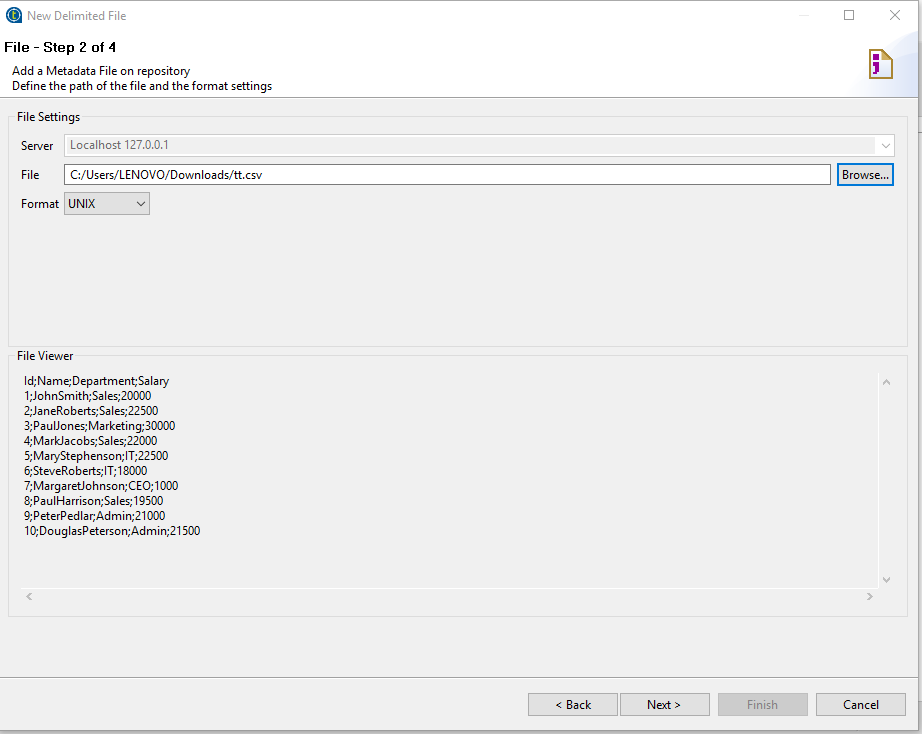


Mô tả được tạo với mức tin cậy rất caoB1:Bấm vào dấu > ở Metadata nhấn chuột phải File delimited chọn Create File delimited .

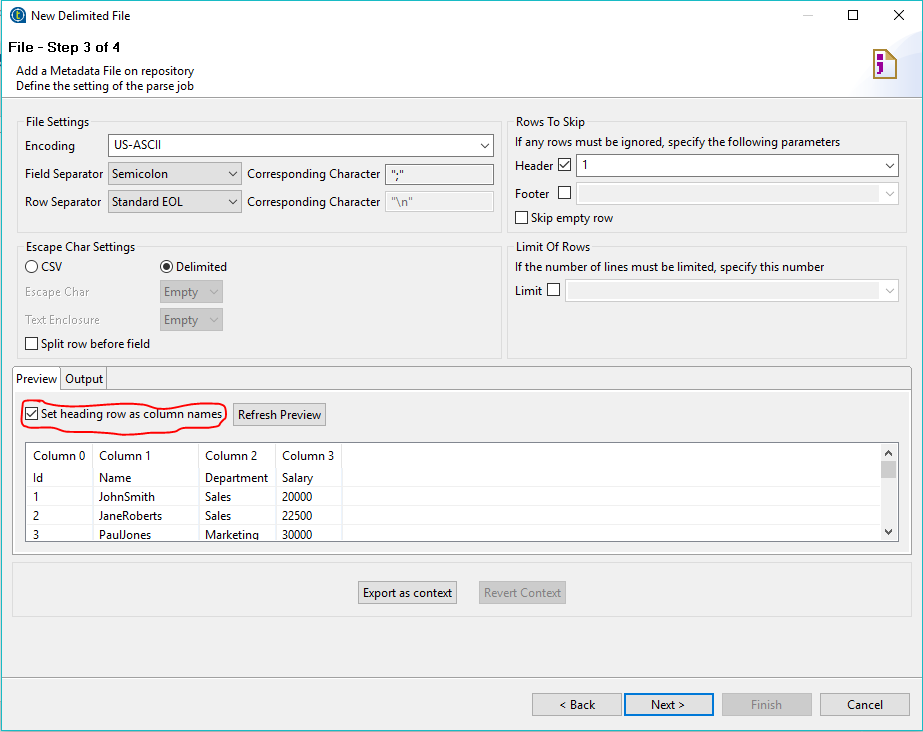


B2:Nhập vào các trường Name(tên),Purpose(mục đích),Description(mô tả) . Sau khi nhập bấm **Next>** để tiếp tục

B3:Chuyển qua hộp thoại tiếp theo . Nhấn **Browse..** để chọn file . Sau khi chọn xong , nó sẽ File Viewer cho xem lại dữ liệu của file đó . Nhấn Next> để tiếp tục



B4:Đánh dấu Set heading row as column names và ta thấy preview dữ liệu hiển thị

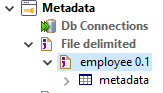


B5:Thực hiện định nghĩa Schema cho employee



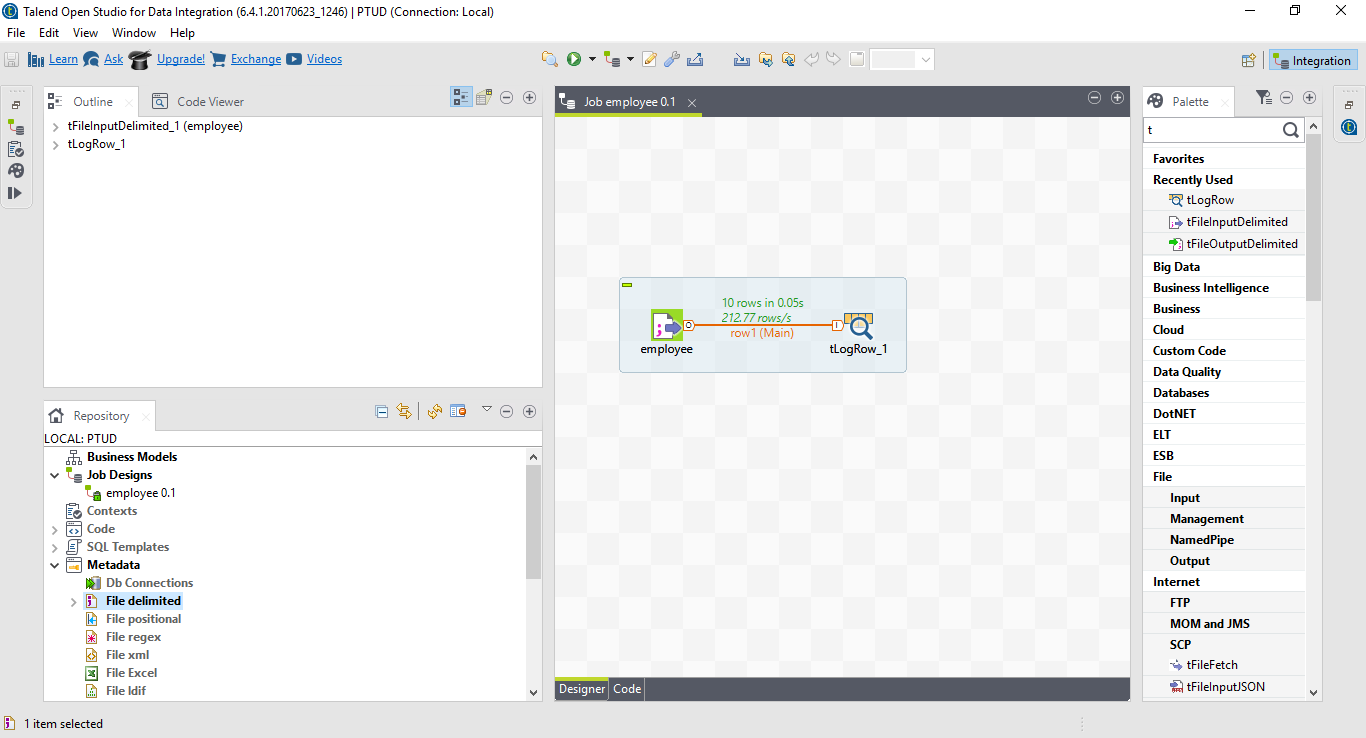
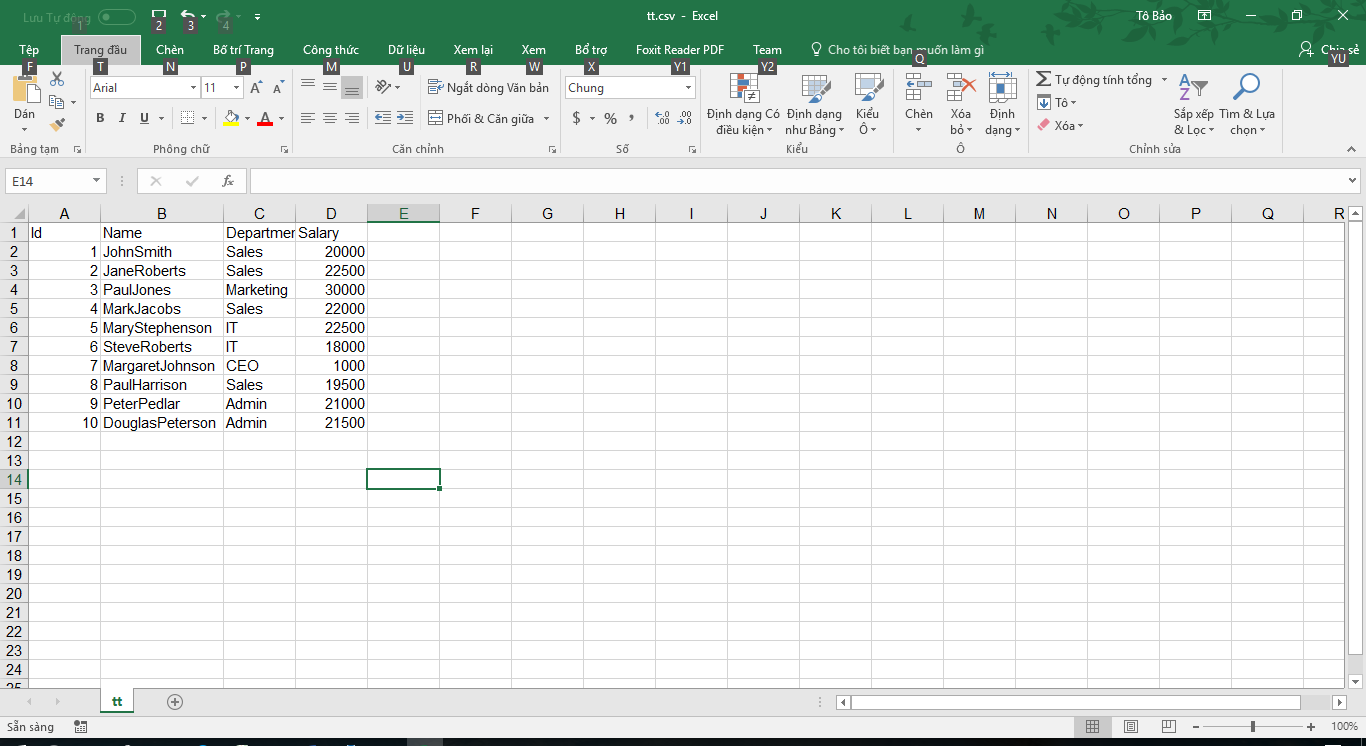
B6:Nhấn Finish để kết thúc

🡺Đây là cách thứ 2 để tạo ra thành phần tFileInputDelimited thay vì thiết kế bằng cách lựa chọn chúng từ Components Palette. Trong cửa sổ Palette tìm đến thành phần tFileInputDelimited trong phần thêm thành phần ở trong mục b

B7:Sau khi Finish, nhấn dấu bên trái Meta data , nhấn và kéo thả chuột employee 0.1 vào vùng thiết kế Job Design

Sau khi thêm tFileInputDelimited tiếp tục thêm thành phần tLogRow

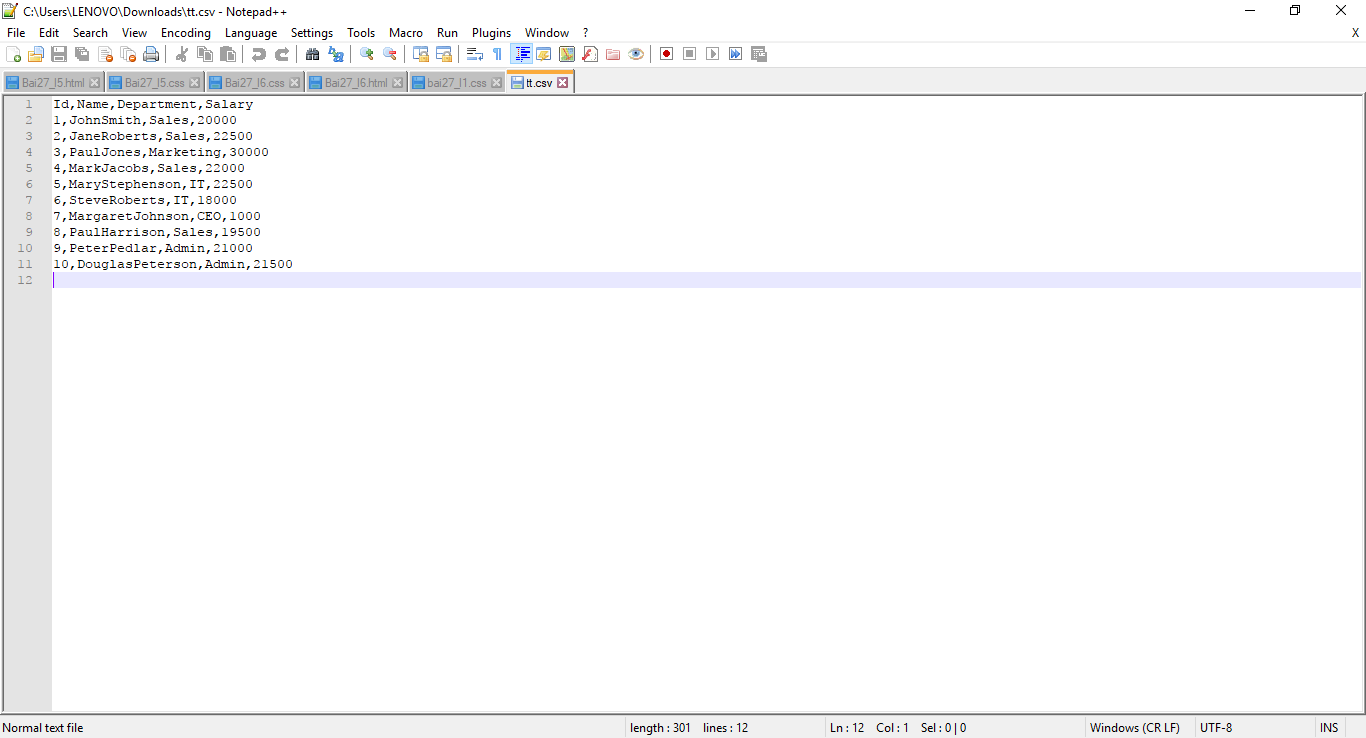
Khi đã có 2 thành phần tFileInputDelimited và tLogRow tiến hành kết nối 2 phần lại với nhau . Nhấn chuột phải **tFileInputDelimited** rồi Row 🡪 Main

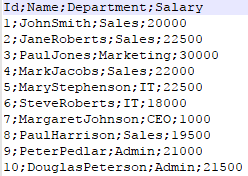
\*Lưu ý : Khi Sử dụng csv có sẵn và mở file csv ta được như sau :

Nó có các thuộc tính gồm Id,Name,Department,Salary

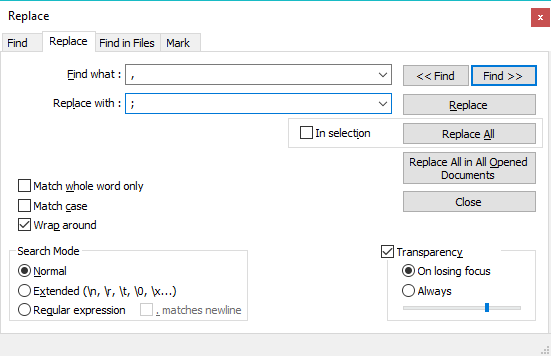
Để đưa và hiển thị toàn bộ thông tin trong Talend ta thực hiện các bước sau đây :

B1:Nhấn chuột phải vào file tt.css chọn Edit with Notepad++ hoặc Open with 🡪 Notepad



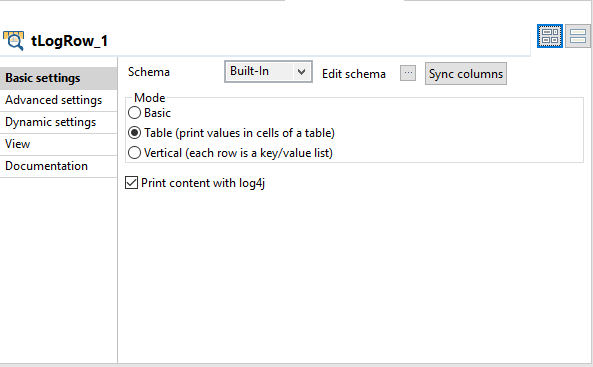
B2:Bôi đen một dấu phẩy nhấn Ctrl+H sau đó điền vào Replace with dấu ; rồi chọn Replace All

Lúc này tất cả dấu cả , trở thành dấu ;

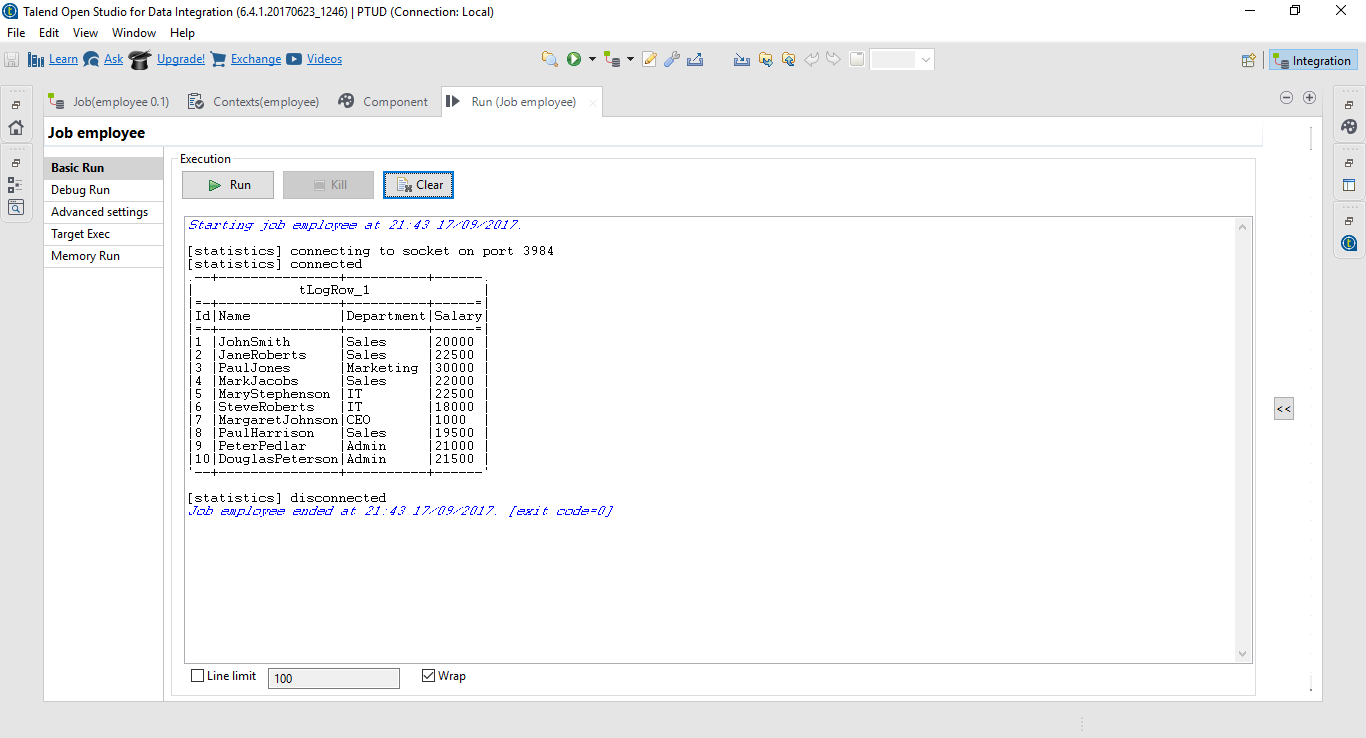


B3: vào File 🡪 Save để lưu lại

Sau đó double chuột vào tLogRow Mode chọn Table:

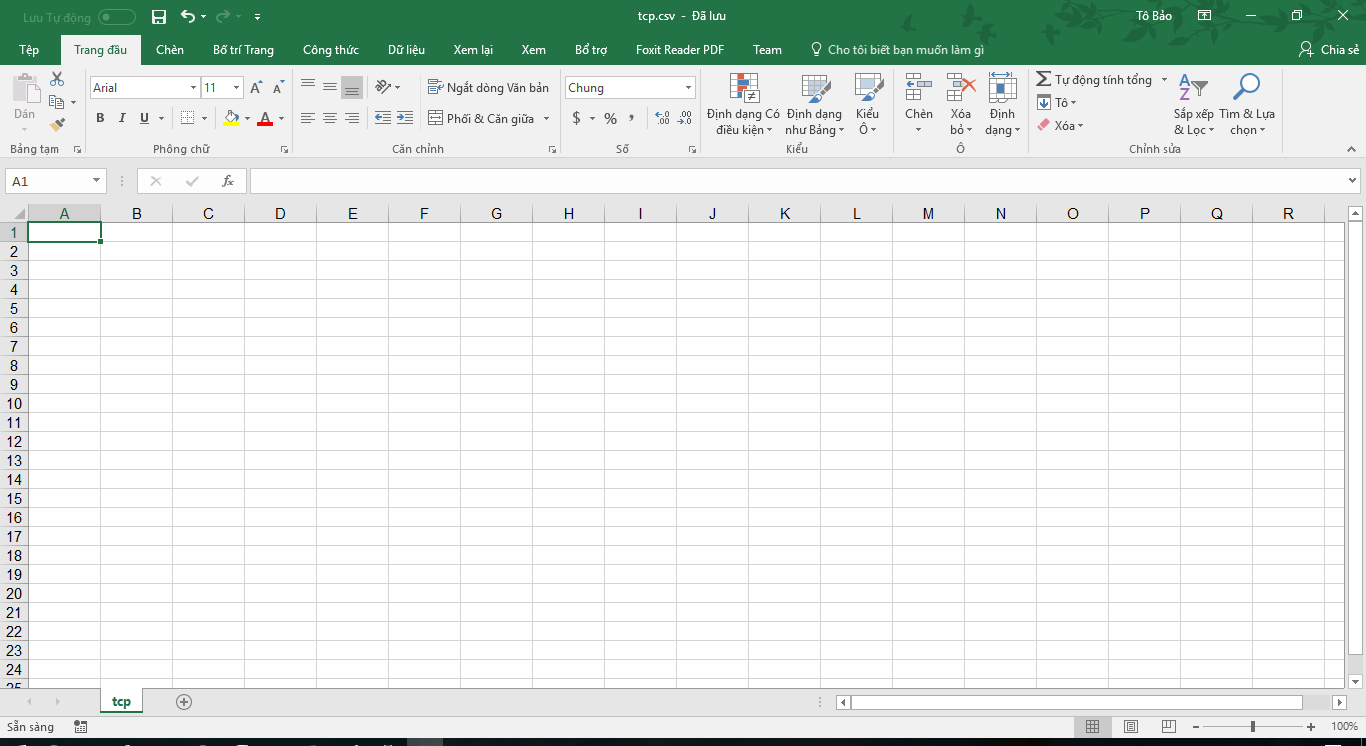


Cuối cùng nhấn Run  để Chạy Job trên thanh 

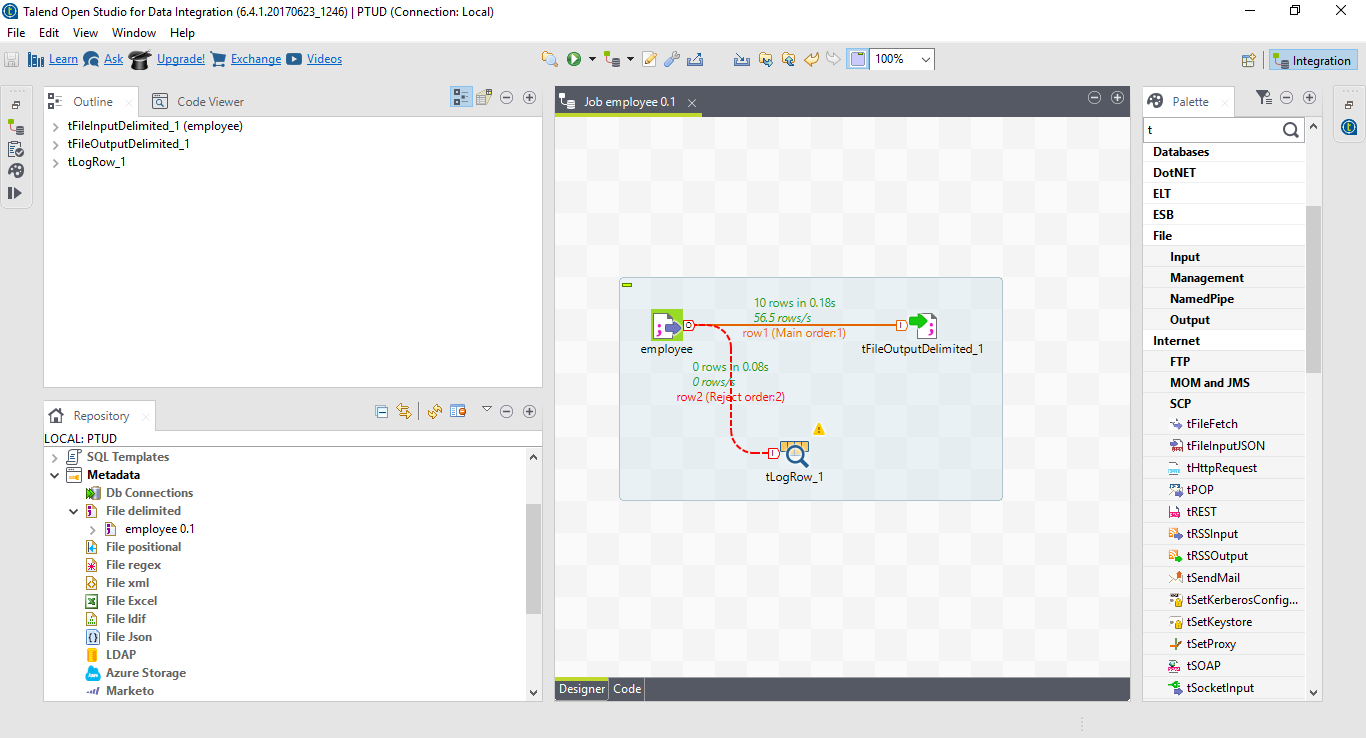


\*Xử lí sao chép thông tin từ 1 csv sang 1 csv khác

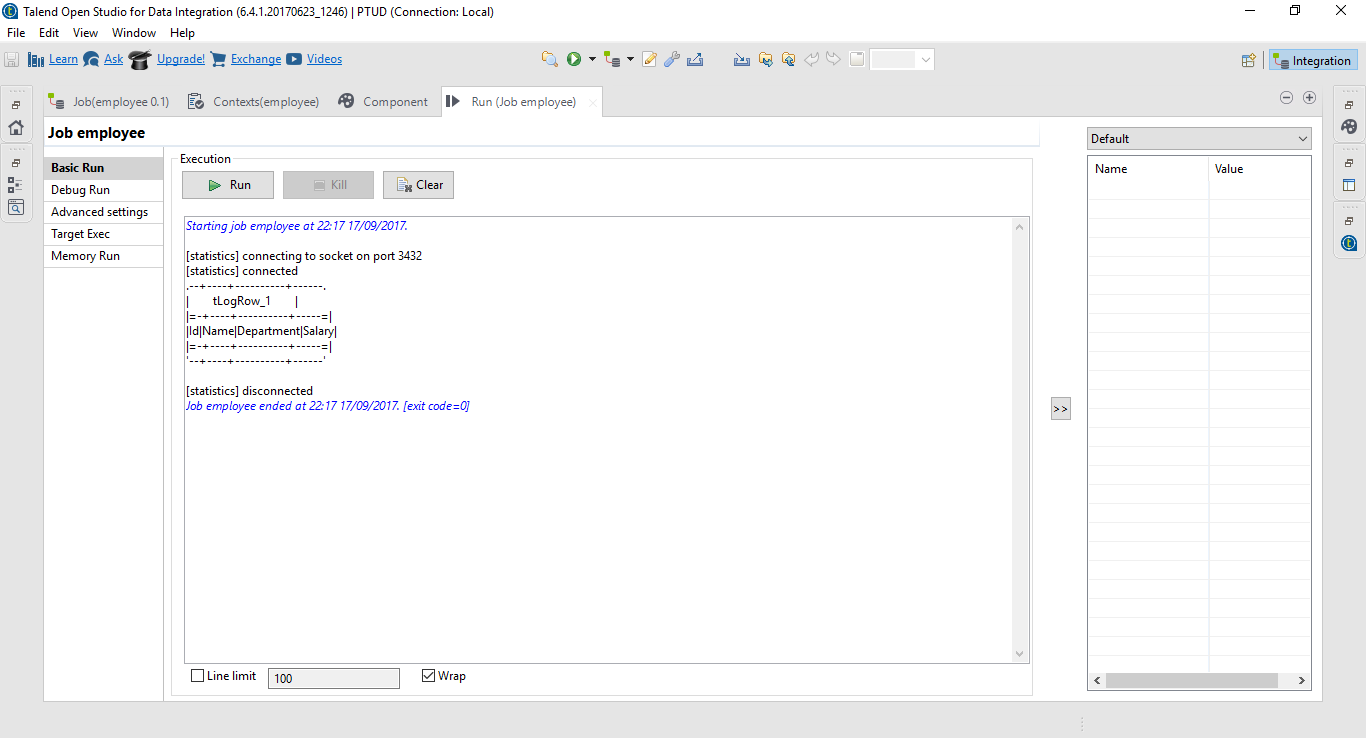
Đầu tiên phải chuẩn bị 1 file csv rỗng không có dữ liệu hoặc 1 file có dữ liệu nhưng bắt buộc dữ liệu đó phải có các thuộc tính và kiểu dữ liệu giống với tFileInputDelimited.Ở đây ta chuẩn bị 1 file csv rỗng

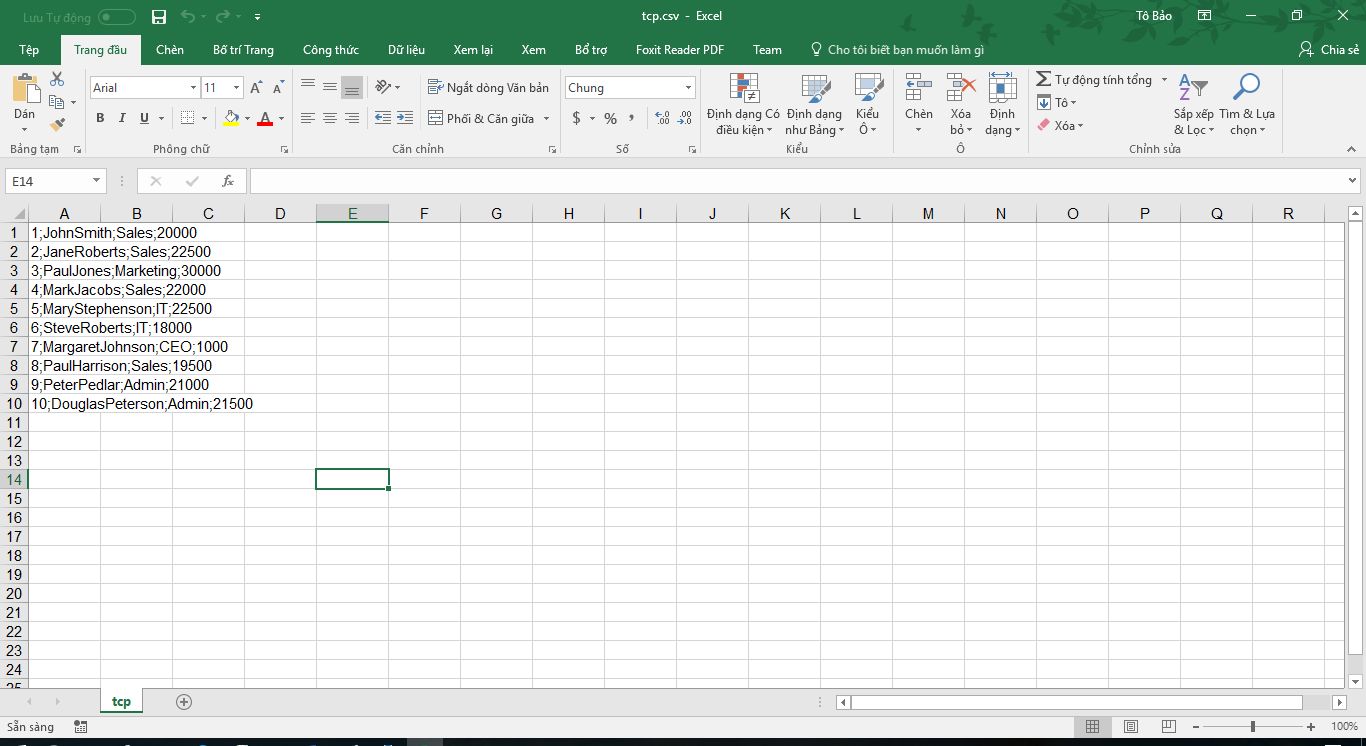


Thực hiện cấu hình tFileOutputDelimited với file csv rỗng và kết nối các các thành phần tFileInputDelimited và tLogRow



Sau đó chạy Job ta được như sau



Xong khi chạy xong , mở lại file csv rỗng đó Ta thấy toàn bộ thông tin trong csv hoàn toàn giống với csv trong tFileInputDelimited

**6.Mapping Data:**

*a)Giới thiệu về Mapping Data*

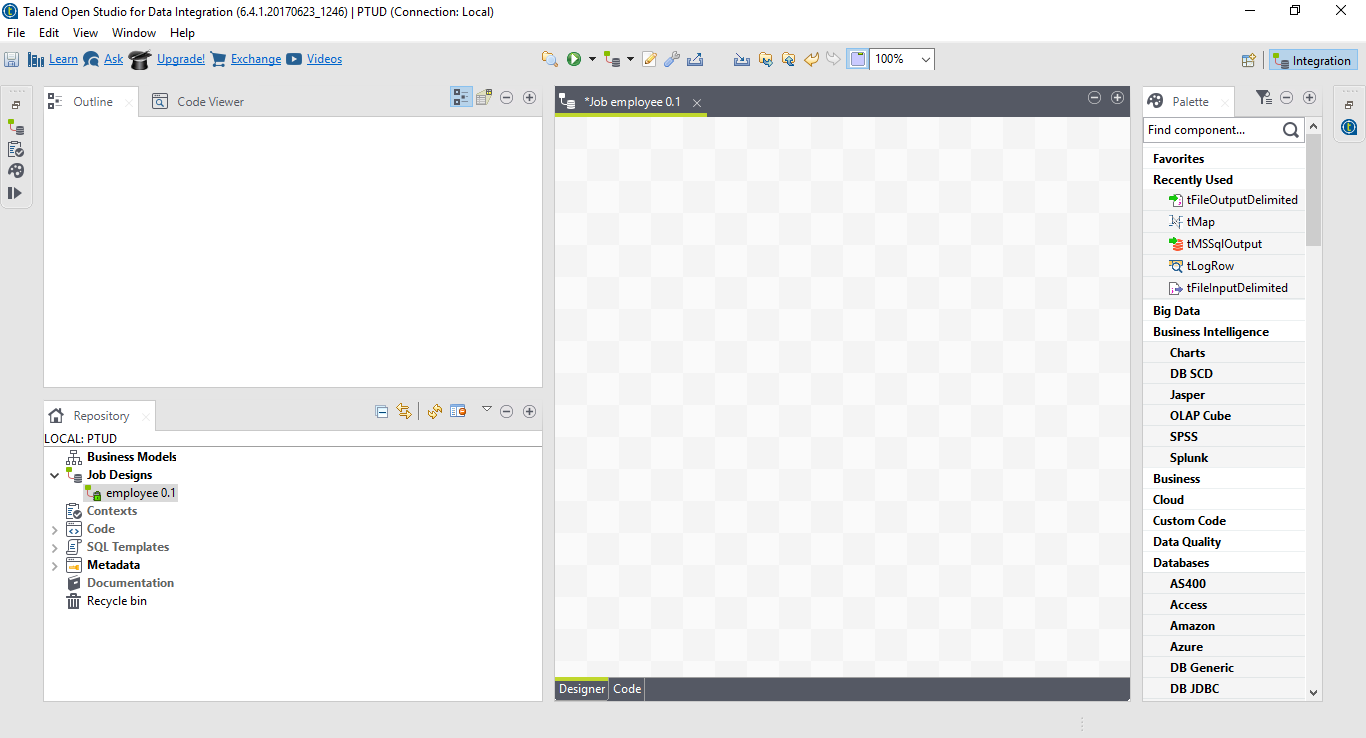
Mappy Data chủ yếu là thao tác với các *tMap component* mà thường là xử lý  
chính là thành phần trung tâm của bất kỳ việc chuyển đổi nào của *Talend Job*.  
Các *tMap component* có khả năng chuyển đổi rộng lớn và do đó trở thành công  
cụ tích hợp dữ liệu cho nhiều nhà phát triển lựa chọn. Trong đó *tMap  
component* có khả năng:

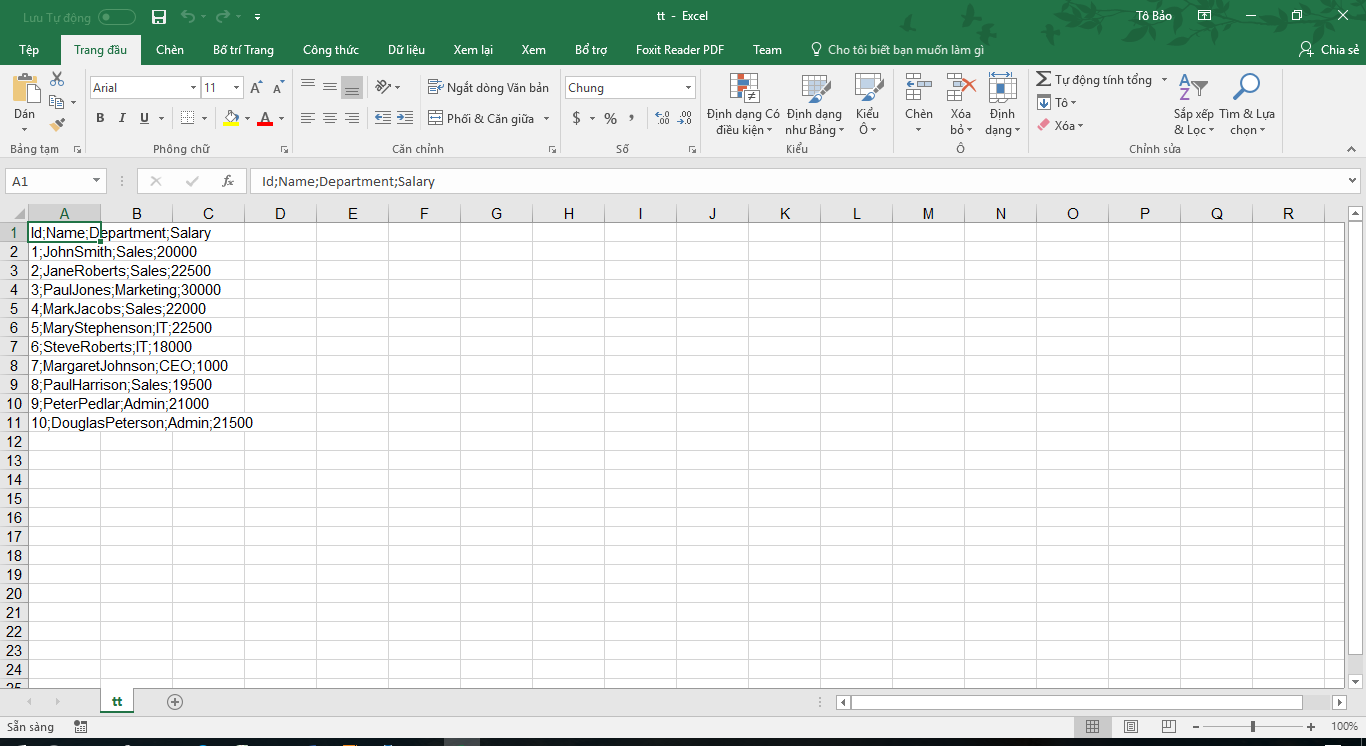
- Thêm và xoá các cột.  
 - Áp dụng quy tắc chuyển đổi một hoặc nhiều cột.  
 -Lọc dữ liệu đầu vào và đầu ra.  
 -Kết hợp dữ liệu từ nhiều nguồn vào một hoặc nhiều kết quả đầu ra.  
 - Phân chia nguồn dữ liệu thành nhiều kết quả đầu ra.

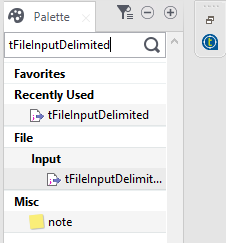
*b)Các thao tác trên Mappting Data:*

**\*Đọc dữ liệu trên từ một file và ghi vào kết quả file mới :**

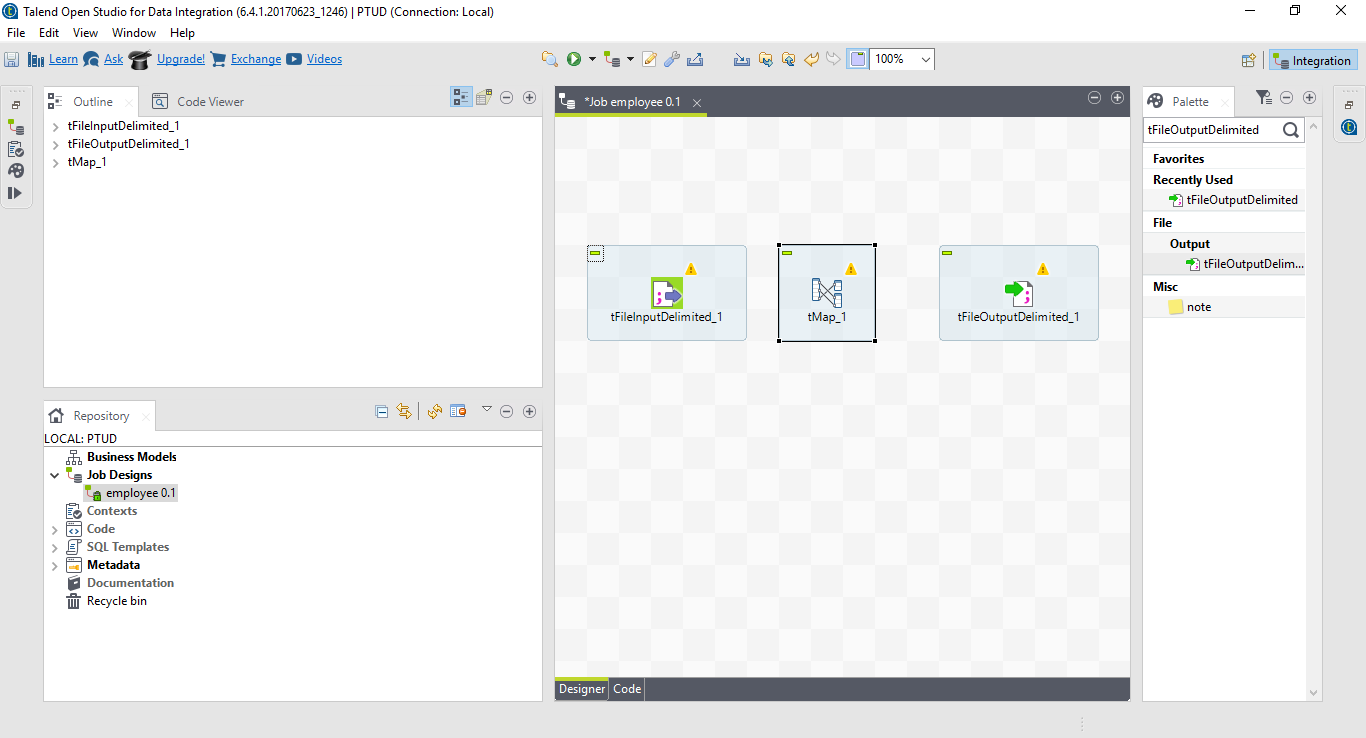
B1:Tạo 1 job mới hoặc tái sử dụng lại job cũ và xóa các thành phần trong Job.Ở đây sẽ sử dụng lại Job Design có tên employee trong phần demo



B2:Sử dụng lại file tt.csv

B3: Thêm các thành phần vào cửa sổ thiết kế bằng cách lựa chọn chúng từ Components Palette. Trong cửa sổ Palette tìm đến 3 thành phần ***tFileInputDelimited, tMap và tFileOutputDelimited***

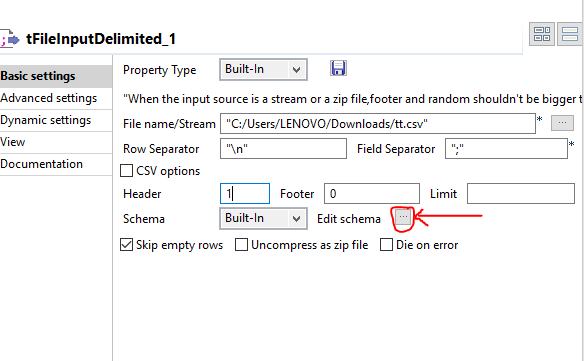
Lần lượt kéo thả chúng vào cửa sổ thiết kế sau khi tìm được các phần

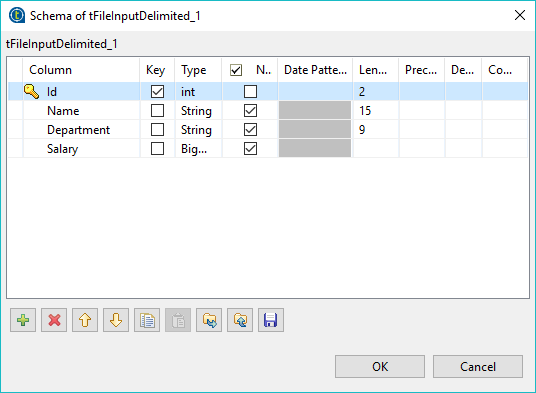


B4:Cấu hình thành phần tFileInputDelimited bằng cách double chuột vào các thành phần đó chọn đường dẫn



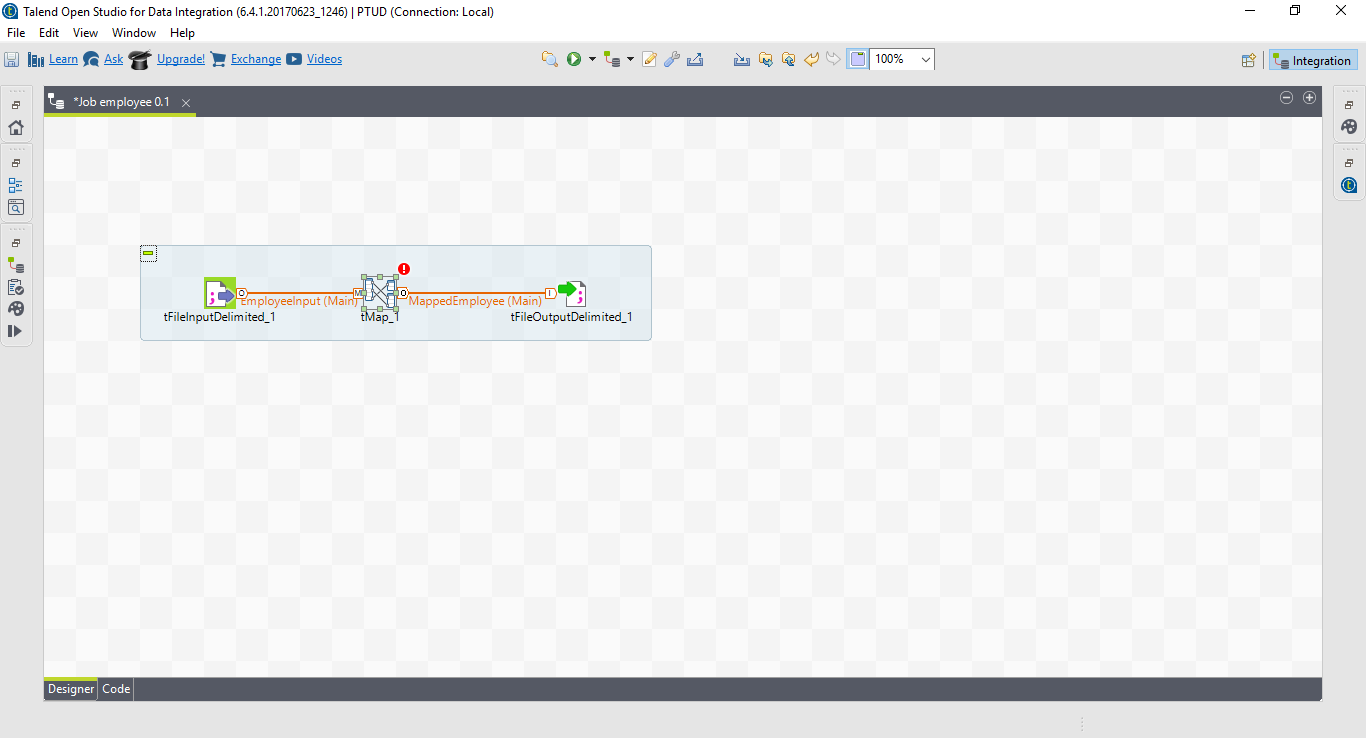
Sau đó Định nghĩa **Schema** cho **EmployeeInputFile:** chọn bên phải **Edit  
Schema.** Talend mở ra cho bạn hộp thoại mới.



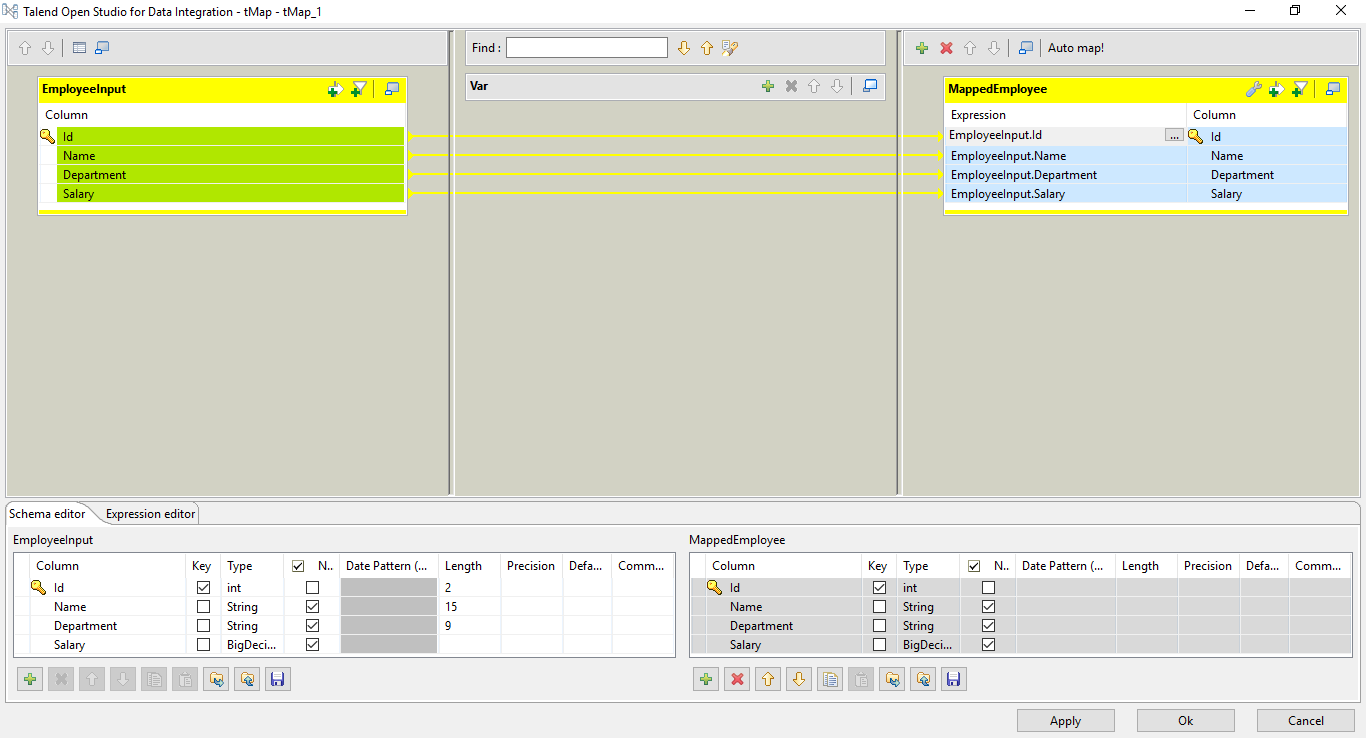


Nhấn OK để hoàn thành

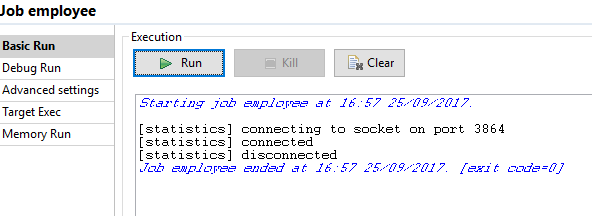
B5:Tiến hành kết nối tFileInputDelimited với tMap bằng cách nhấn chuột phải và chọn Row🡪Main . Sau đó kết nối với tFileOutputDelimited bằng cách nhấn chuột phải ở tMap chọn Row🡪\*New Output\*(Main).

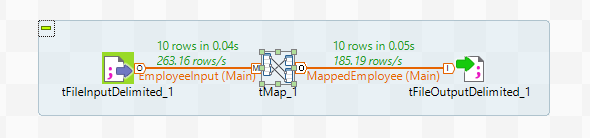
Khi bạn chọn lựa chọn **New Output** Talend yêu cầu bạn phải đặt tên cho *Output*vừa tạo, bạn đặt tên là **MappedEmployee** và sửa nhãn **row1(Main)** thành  
**EmployeeInput** .

B6:Cấu hình tMap: để cấu hình bạn lựa chọn vào chính nó và chọn thẻ *Component properties*, sau đó chọn bên phải **Map Editor h**oặc *double-click* vào trong cửa sổ thiết kế của bạn. Ngay lập tức một hộp thoại xuất hiện cho phép bạn cấu hình tMap.  
Theo hướng dẫn này, bạn sẽ tạo cấu hình cho cấu trúc của *file* đầu ra giống như  
cấu trúc của *file* đầu vào.Sau khi cấu hình xong nhấn **Apply** hiện lên Thông báo Propagate:Would you like to propagate changes? Chọn **Yes** để lưu thay đổi và nhấn **OK** để kết thúc

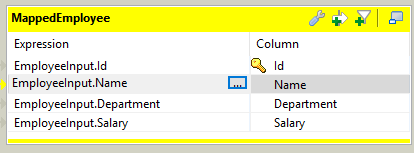
 B7:Cấu hình tFileOutputDelimited : Double chuột vào thành phần và chọn đường dẫn đến file kết quả cần truy xuất dữ liệu từ file input

B8: Run Job :Nhấn nút run để chạy chương trình

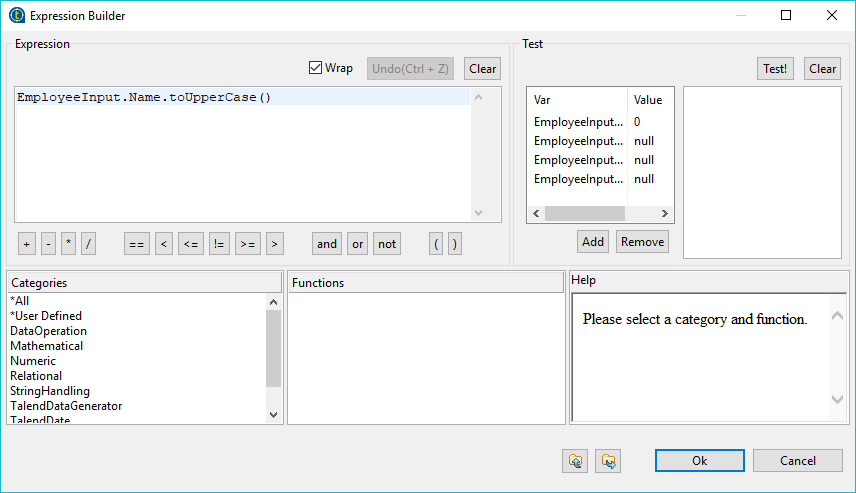


Vùng thiết kế Job Design sau khi chạy chương trình

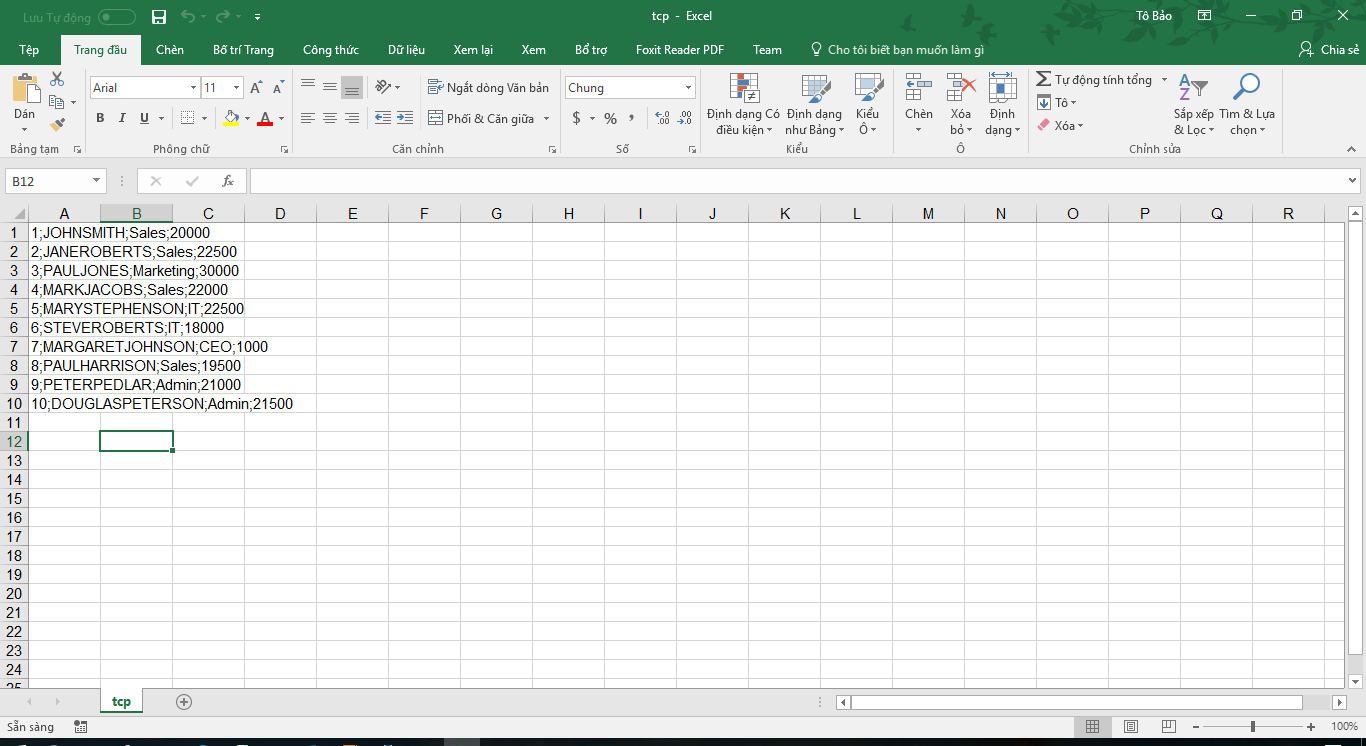
B9:Mở rộng – Thêm điều kiện dữ liệu cho file output: Lựa chọn lại  
**tMap** và cấu hình lại như sau: *click* bên phải **EmployeeInput.Name**trong khung **MappedEmployee.**



Sau khi *click* một hộp thoại mới xuất hiện, đây là cửa sổ để chúng ta thiết lập  
các điều kiện . .Ví dụ ở đây ta chuyển tất các dữ liệu trong Name thành chữ hoa bằng cách Thay đổi biểu thức ***Name***như sau ***EmployeeInput.Name.toUpperCase()****.* Khi bạn nhập, bạn sẽ nhận thấy rằng Talend bật lên một cửa sổ hiển thị các phương pháp có sẵn cho các hiện tượng *Class.*Nhấn **Ok** để hoàn thành

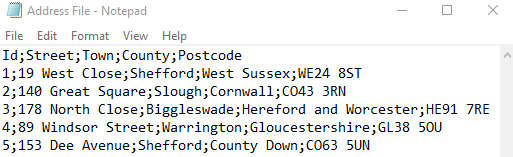


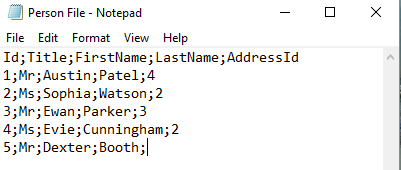
Sau khi cấu hình xong , chạy lại chương trình và sau đó mở file kết quả ta thấy tất cả tên đã thành chữ hoa



\*Gom nhóm và trích lọc dữ liệu : Ngoài chức năng chuyển đổi dữ liệu đầu ra thì *tMap* còn có thể gom nhóm và trích lọc dữ liệu

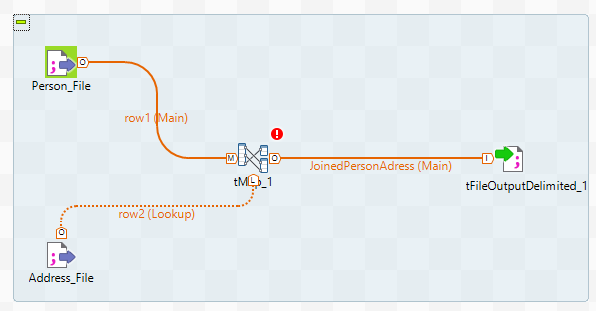
B1: Tạo 2 *file* dữ liệu **Address File.txt** và **Person File.txt** có cấu trúc như  
sau.



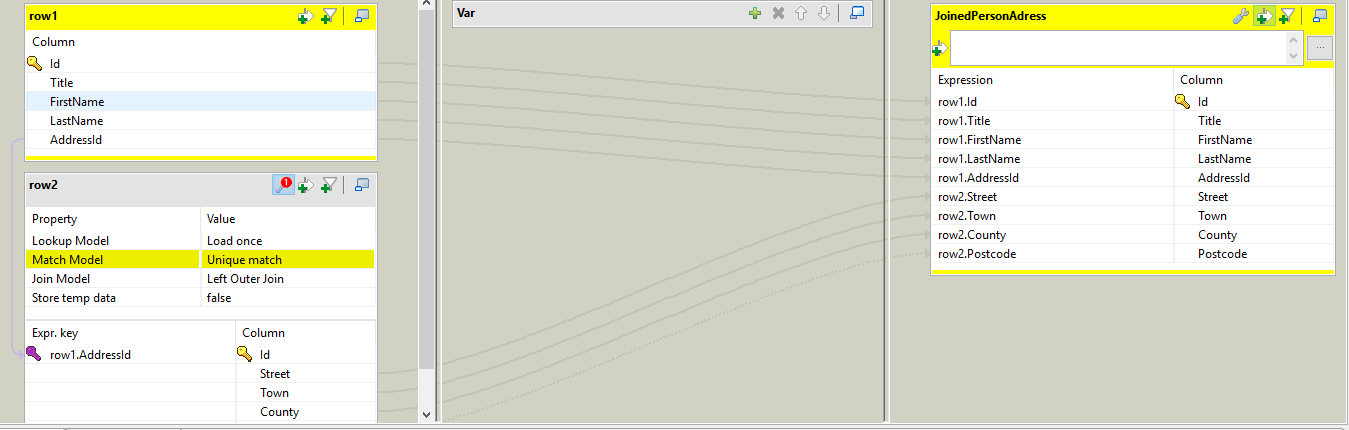


B2: Thiết kế **Job** và cấu hình các *components* như sau:  
 **Person File (tFileInputDelimited\_1):** chứa đường dẫn tới *Person File.txt*.  
**Address File (tFileInputDelimited\_2):** chứa đường dẫn tới *Address File.txt*.  
**tFileOutputDelimited\_1:** chứa đường dẫn tới file kết quả

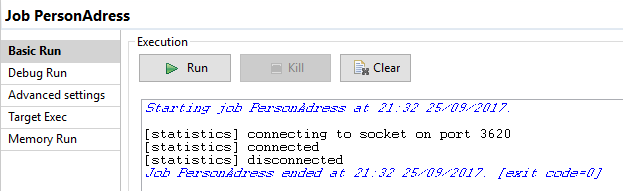
**tMap\_1**



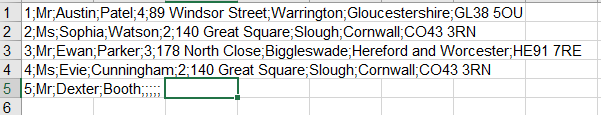
B3: Tiếp theo, chúng ta thực hiện cấu hình cho **MapPersonAddress.**Như có thể thấy hình chụp dưới đây, 2 files đã được thêm vào *Job* và đã được **join** theo **AddressId.** Bạn có thể thực hiện điều này bằng cách kéo thả cột là khóa ngoại của bảng này vào khóa chính của bảng kia hoặc nhập bằng tay.



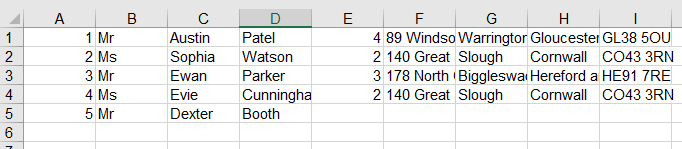
B4: Sau khi cấu hình xong thì **Run** xem thử kết quả

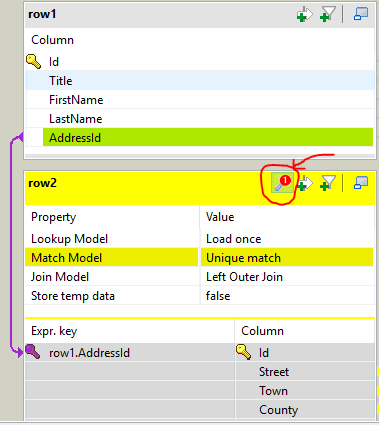


Mở file output ta được như sau :



Chú ý : Trong csv muốn hiển thị thành từng dòng từng cột thì nhấn chuột phải vào file chọn Edit with Nopad++ sau đó bôi đen dấu ; và nhấn Ctrl+H điền vào ô Replace with dấu , rồi nhấn Relace All . Xong nhấn Close để hộp thoại . Lưu lại và mở lại File ta được như sau



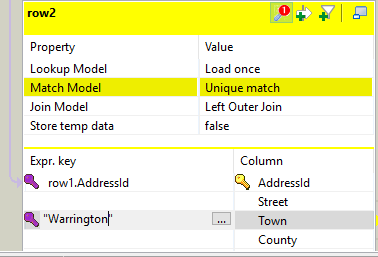
B5:Thiết lập dữ liệu đầu vào (hoặc thiết lập các điều kiện thiết lập)

Bằng cách lựa chọn *button (tMap settings)* đầu tiên. Mỗi tùy chọn đầu vào đều có các thiết lập đặc trưng, thông thường sẽ được giử thiết lập mặc định. Nếu khi bạn thay đổi các thiết lập đó thì bạn sẽ nhận thấy *button* được khoanh tròn màu đỏ trong hình bên :

**Lookup Model**: Thông thường bạn lựa chọn tải dữ liệu dò tìm từ **Input** chỉ một lần, giá trị mặc định của **Lookup Model** là **Load Once**, nếu bạn có trường hợp cần tải lại dữ liệu từ *Input* cho mỗi *row* khóa chính thì bạn có lựa chọn **Reload at each row or Reload at each row (cache),** tùy thuộc vào kích thước của đầu vào của bạn, một trong hai tùy chọn đó sẽ ảnh hưởng nghiêm trọng đến thông lượng *Job* của bạn.

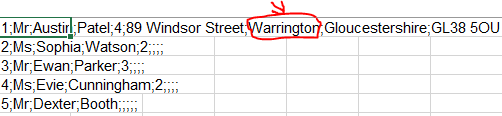
**Match Model**: Mặc định **Match Model** là **Unique match**, nếu *row* dò tìm chính của bạn phù hợp với nhiều dòng trong *file Input* thì chỉ có dòng phù hợp cuối cùng là xuất hiện trong *file Output.* Tùy chọn thứ hai là **First match, chỉ có dòng dò tìm phù hợp đầu tiên xuất hiện trong *file Output.* Tùy chọn cuối cùng là All matches tất cả các dòng phù hợp sẽ xuất hiện trong *file output*.**

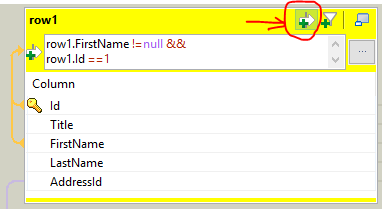
**Join Model**: Mặc định **Join** **Model** là ***Left Outer Join***. Nếu không có *row* dò tìm phù hợp xuất hiện trong file *Input* thì *row* là *primary* của *Input* vẫn được giữ lại trong *file* *Output*. Nếu bạn muốn thực hiện một **Antijoin** (dò tìm có trong A mà không có trong B) thì lựa chọn *option* và loại trừ các dòng ra, bởi *output* ra một giá trị quan trọng từ cái nhìn vào *Input* của bạn và sau đó không bao gồm hàng có giá trị này là *null*. Tùy chọn thứ hai có sẵn là **Inner Join,** trong trường hợp này thì chỉ có những *row* phù hợp sẽ được *output* ra.

**Store temp data**: Nếu bạn đang xử lý dữ liệu lớn, bạn có thể tìm thấy nó hữu ích khi thiết lập tùy chọn này *true*, để tiết kiệm bộ nhớ. Khi thiết lập là *true*, bạn cũng sẽ cần phải thiết lập **Temp data directory path**, việc này bạn có thể được tìm thấy trên *tab* **Advanced settings** *component* của *tMap*. Lưu ý rằng điều này có thể có tác động tiêu cực đối với thời gian tổng thể *Job* của bạn.

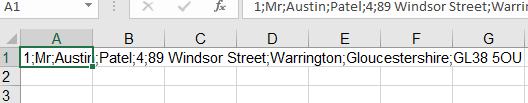
B4: Thiết lập điều kiện trích lọc như sau.

**Run** lại *Job* của bạn và xem kết quả. Với điều kiện trích lọc như trên thì tất cả các dòng trong *Person File* sẽ xuất hiện trong f*ile Output* tuy nhiên chỉ có **Address.Town==”Warrington”:**



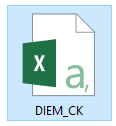
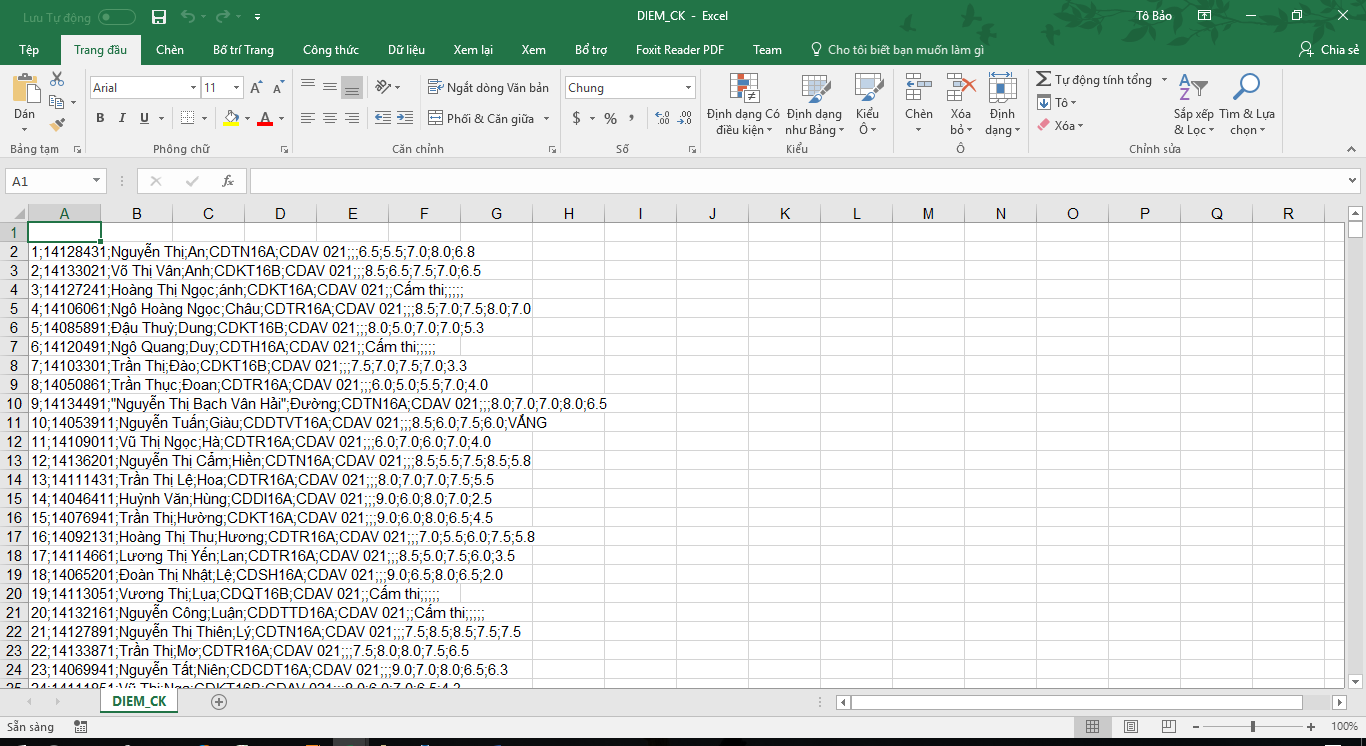
B5: Nhấn **button được đánh dấu tròn bên hình** trong hộp thoại **tMap Expression Editor, đây là nơi bạn thiết lập các điều kiện trích lọc phức tạp nhưng Talend lại không có biểu hiện gì khác biệt khi bạn tắt cửa sổ Expression Filter, chính vì lý do này tôi cũng tránh sử dụng các bộ lọc ẩn.**

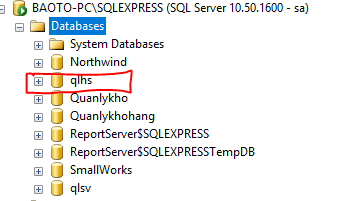
**B6:Run Job ta nhận được kết quả là hiển thị tất cả các First name có Id là 1**



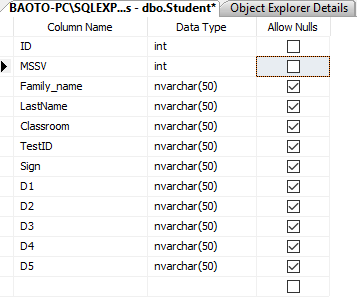
**7)Làm việc với DataBase** :

*a)Demo Xuất dữ liệu qua Microsoft SQL Server từ 1 file dữ liệu*

 Ở đây ta có một file csv Diem\_CK.csv

Mở - kết nối SQL Server tạo 1 database có tên qlsv(hay tên gì đó tùy)

Sau đó tạo một Table có tên Student.Cấu trúc các trường phải giống như bên csv



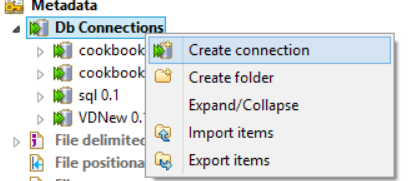
Vào Talend tạo 1 Job mới và thêm 3 thành phần :

-**tFileInputDelimited** có tên là Diem chứa file dữ liệu Diem\_CK.csv

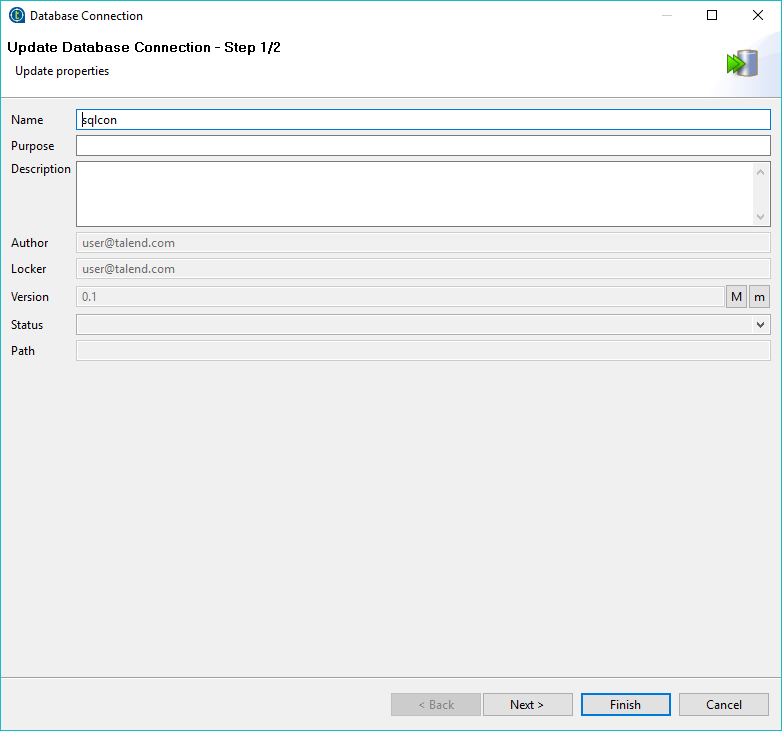
**-tMap\_1**

- **tMSSqlserverOutput** có tên sqlcon

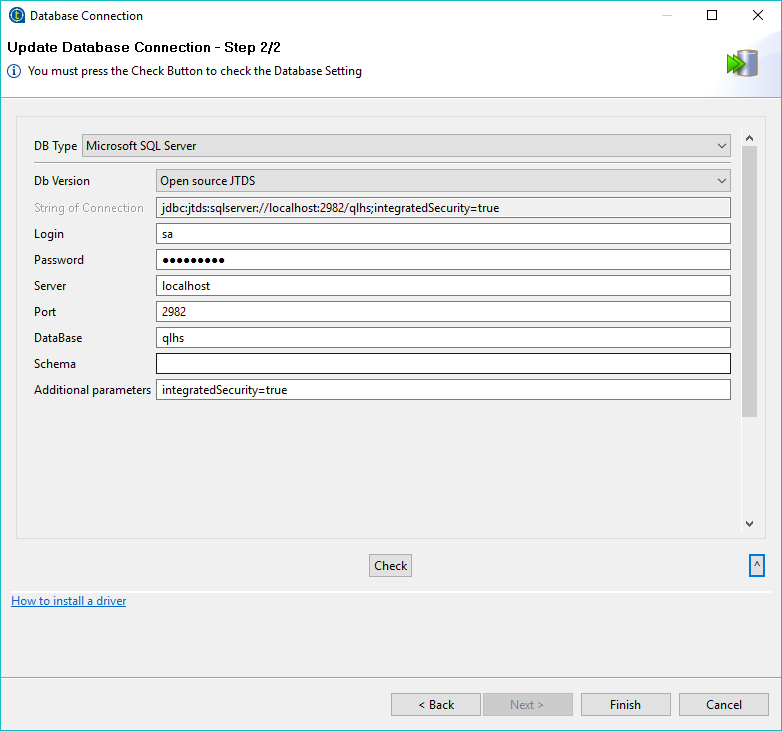
Việc thêm **tFileInputDelimited và tMap1** đã được hướng dẫn ở những phần trước . Đến phần này sẽ trình bày cách thêm thành phần **tMSSqlserverOutput**

**B1:** Từ **Metadata**, *Right-click* vào ***Db Connections***và chọn **Create connection**

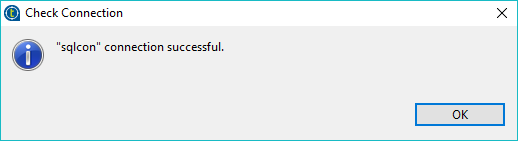
B2;Một cửa sổ mới xuất hiện . Nhập Name (tên ) là sqlcon ( muốn đặt gì cũng được) . Purose(Mục đích) và Description(mô tả) có thể bỏ qua . Sau đó nhấn next



B3:Một cửa sổ tạo mới kết nối đến cơ sở dữ liệu được mở ra. Đặt tên cho kết nối  
và nhấn *Next*.  
Lựa chọn **DB Type** là *Microsoft SQL Server*:  
***Login***: Tài khoản được phân quyền từ *Microsoft SQL Server* (có thể bỏ qua).  
***Password***: Mật khẩu của tài khoản SQL (có thể bỏ qua).  
***Server***: Tên kết nối *server* của *SQL* hoặc có thể nhập địa chỉ IP kết nối đến  
server.  
***Port***: Talend sẽ tự động sinh ra *port* cho kết nối của bạn.  
***DataBase***: Tên *database* bạn cần thao tác từ Talend đến SQL.  
***Schema***: Tên này thường là **dbo** (Tên này đứng trước tên *table* trong *sql*, ví dụ:  
***dbo. Categories***).  
***Additional parameters:*** ***integratedSecurity=true***.(quan trọng phải có)



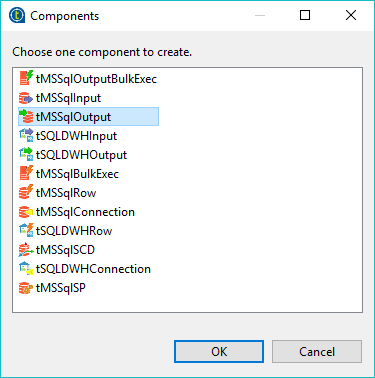
B4:Nhấn check để kiểm tra đã kết nối với SQL chưa . Nếu đã kết nối sẽ đưa ra thông báo sau đây

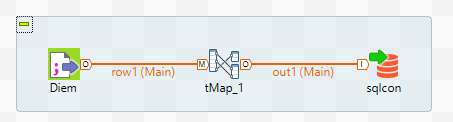


\*Lưu ý : Ban đầu có thể xảy ra lỗi vì ko kết nối được . Lí do là bạn chưa tải về JTDS hoặc những lỗi nhập từ Post hay Server

B5:Nhấn OK và Finish để hoàn thành việc thêm MSSqlserverOutput. Vậy ta đã hoàn thành việc thêm các phần vào Job Design . Sau đó kéo thả vào Job Design sẽ

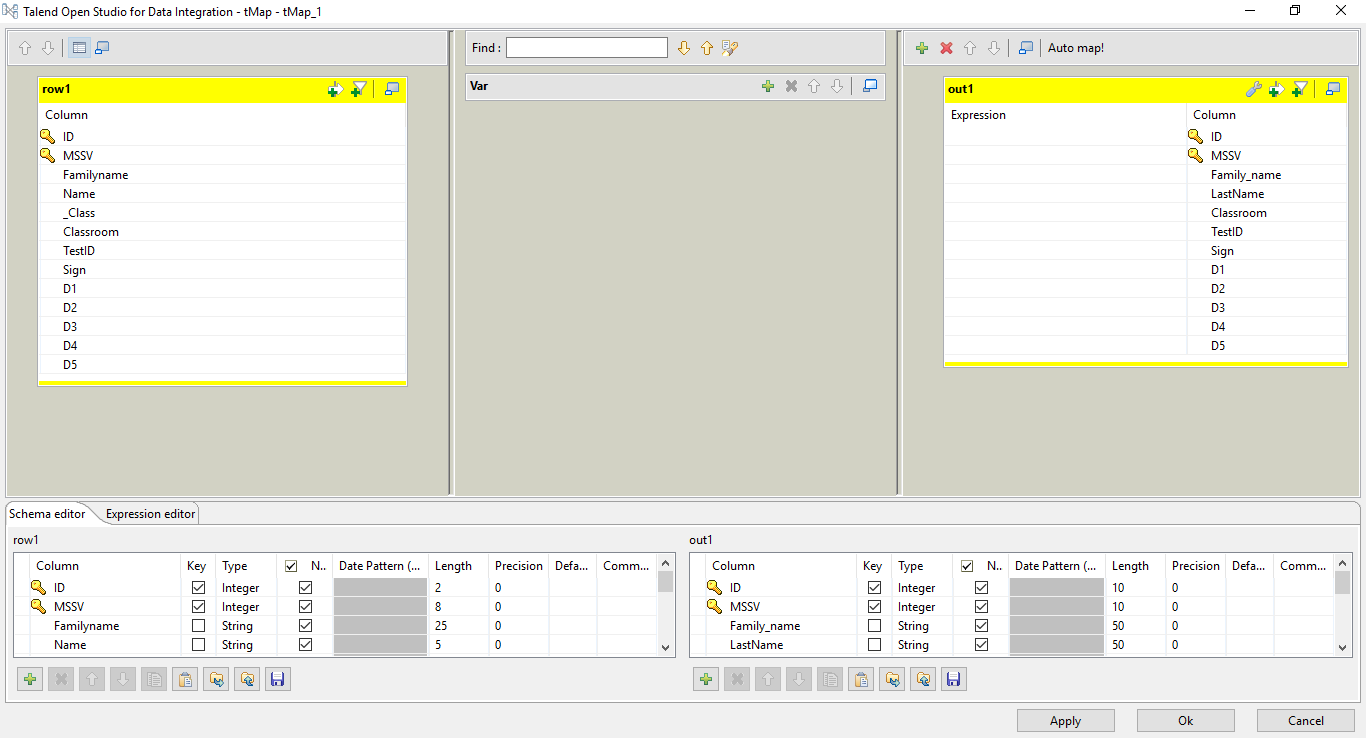
Hiện lên cửa sổ ở dưới và chọn tMSSqlOutput .Vậy việc thêm thành phần đã xong





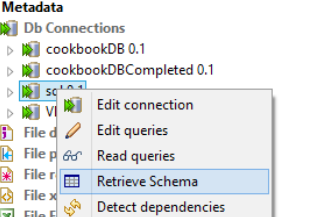
Sau đó kết nối các thành phần với nhau như hình trên

Bước tiếp theo là cấu hình các thành phần . Đầu tiên cấu hình tMap\_1 như sau

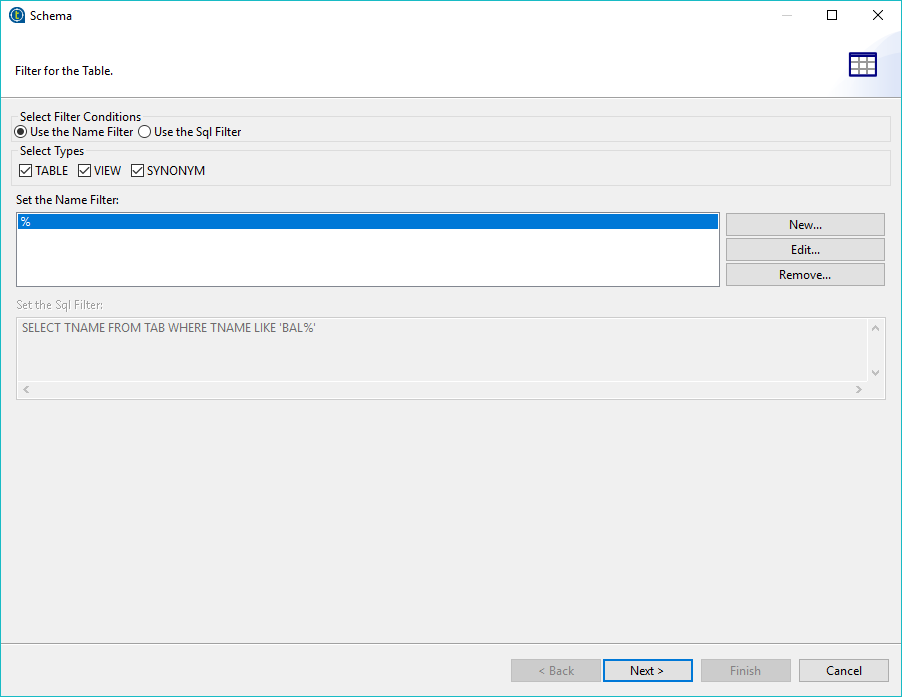


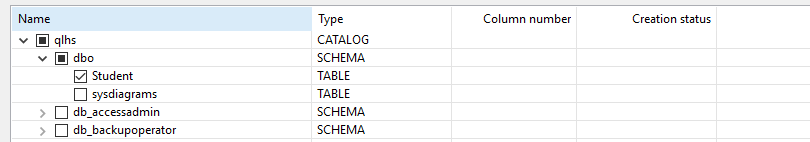
Sau đó cấu hình sqlcon

B1: Từ *connection* đã tạo, *Right-click* và lựa chọn **Retrieve Schema**



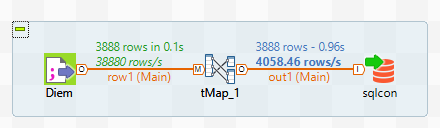
B2:Một cửa sổ xuất hiện và bỏ qua cửa sổ này Bấm Next> để qua cửa sổ 2

B3:Tại của sổ 2 . Nhấn chuột vào ‘>’ sổ ra danh sách các bảng trong SQL . Và chọn bảng Student tạo ban đầu

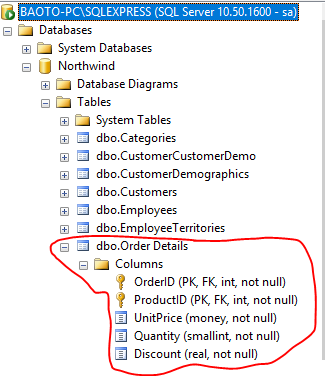


B4:Nhấn Finish để hoàn thành cấu hình

B5:Sau đó Nhấn button Run để chạy

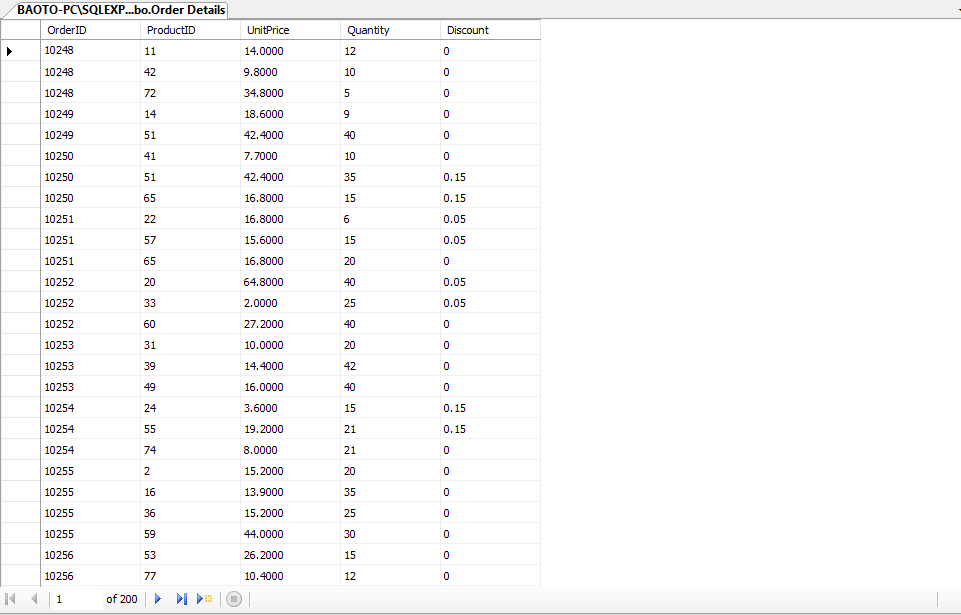


Mở lại SQL Server vào Database mở dbo.Student để xem kết quả

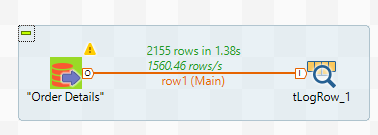
***b*)*Đọc dữ liệu từ database File (SQL Server)***

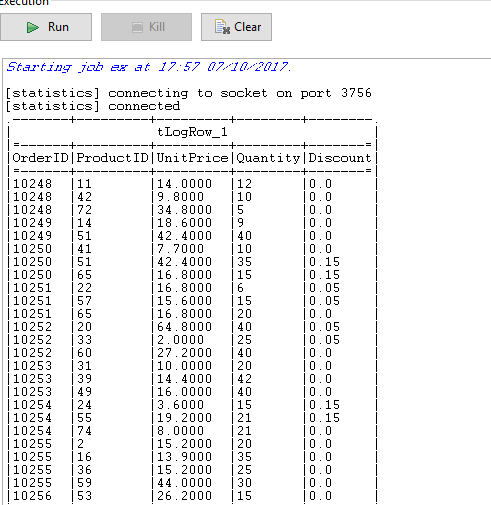
**-**Đầu tiên ta phải chuẩn bị 1 file data base có sẵn trong SQL . Ở đây ta có Northwind Datatabase . Với các bảng ta chọn dbo.Order Details

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 .

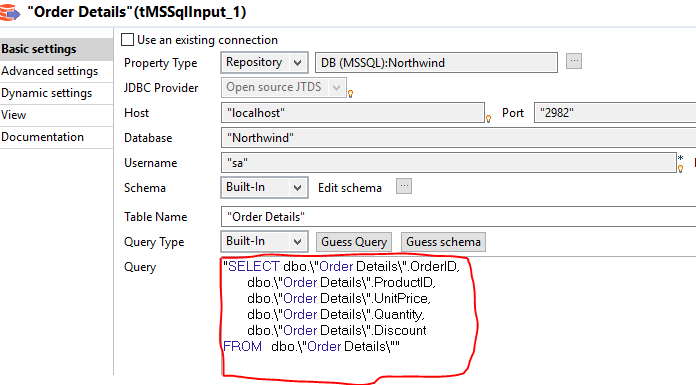




Sau khi chạy xong ta thấy toàn bộ dữ liệu trong Database được hiển thị

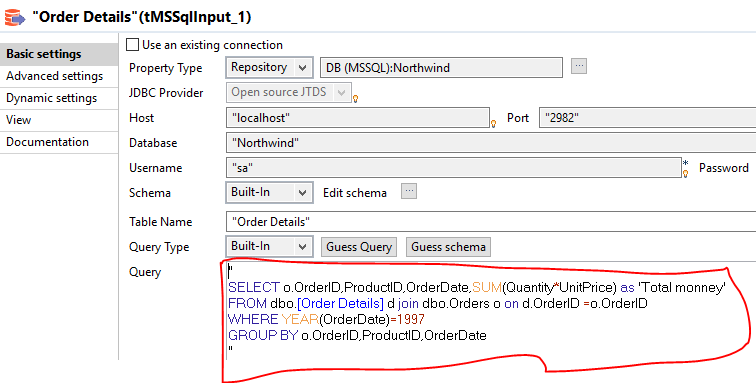
*c)Thực hiện Truy vấn đữ liệu trong Talend*

- 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

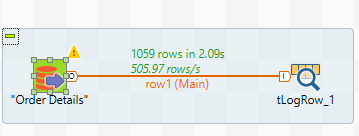


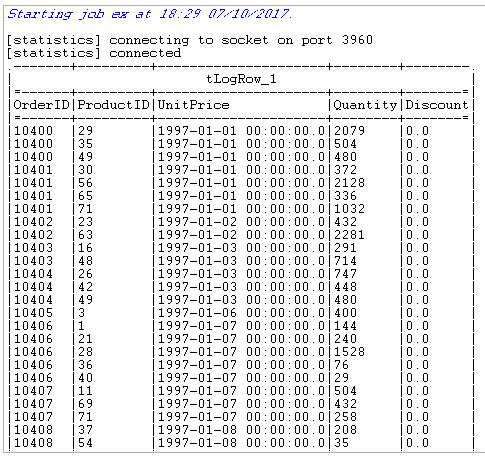
Chỗ khoanh tròn là chỗ cho phép người dùng nhập những câu lệnh để truy vấn dữ liệu

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

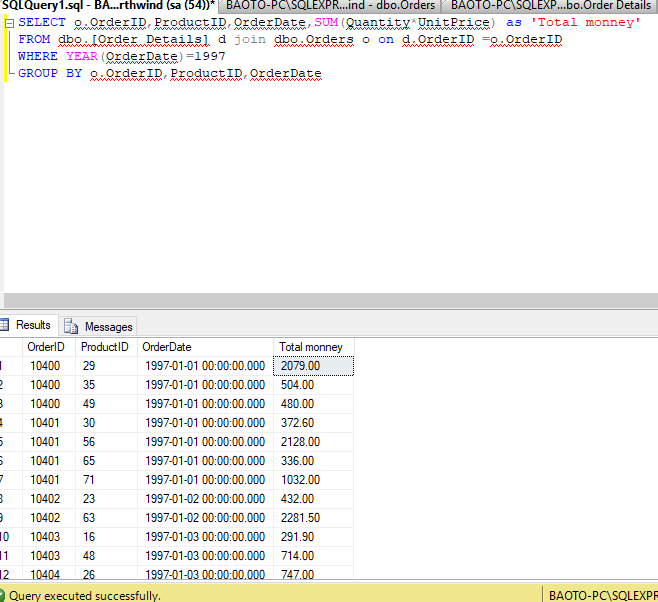


Sau khi đã hoàn tất . Nhấn Run để xem kết quả

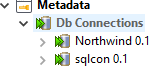


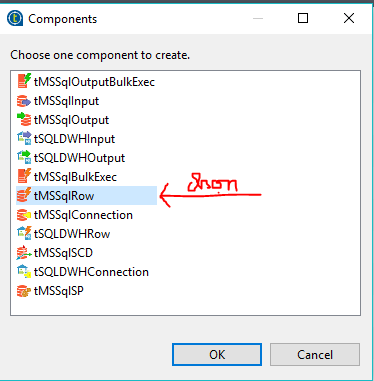


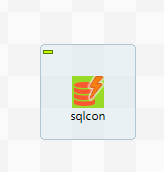
Đây là kết quả sau khi thay đổi câu truy vấn . Nó không còn xuất toàn bộ thông tin nữa mà chỉ xuất theo truy vấn của người dùng. Nếu chưa chắc chắn về kết quả thì viết lại câu đó vào SQL Server ta sẽ ra kết quả tương tự

\*

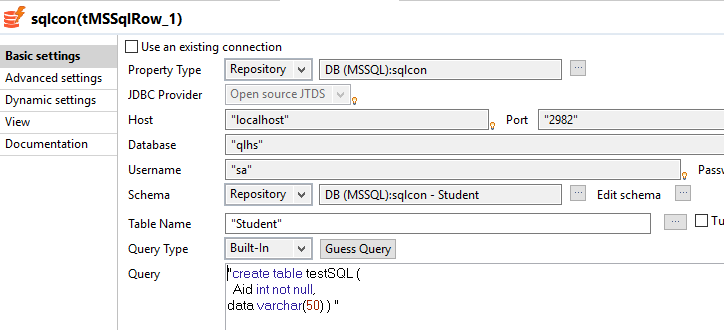
d)Thực hiện tạo bảng database từ Talend :

 - 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

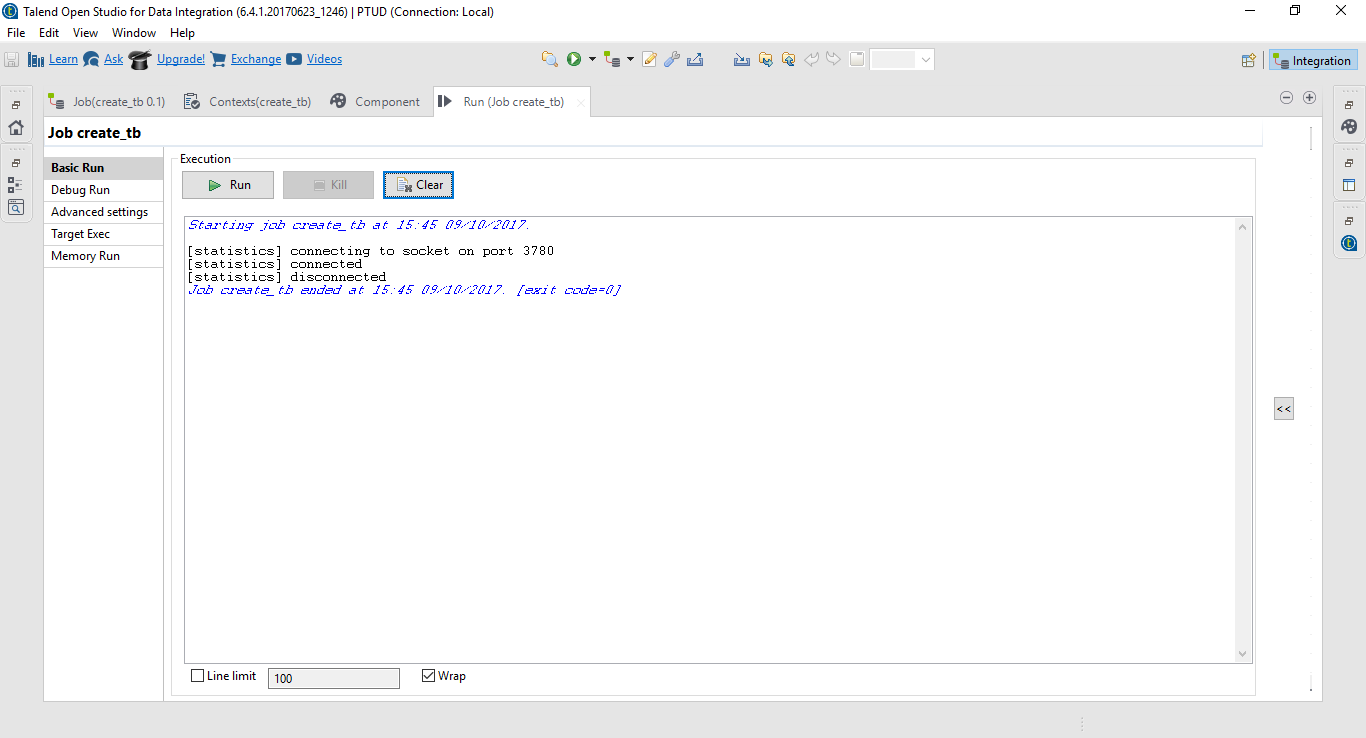


Nhấn double vào tMSSqlRow . Sau đó gõ lại câu lệnh sau vào chỗ Query

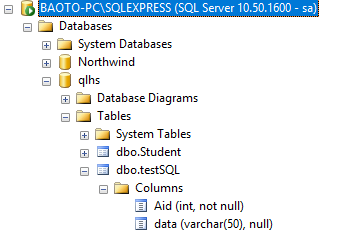


Có nghĩa rằng ta tạo 1 bảng có 2 trường dữ liệu là Aid và data

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

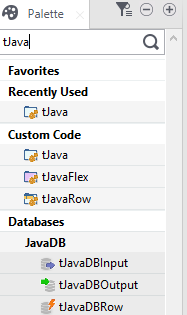


Vậy là ta đã tạo thành công bảng của database từ Talend . Để chứng thực ta vào SQL mở qlhs database ta thấy dbo.testSQL đã được thêm vào

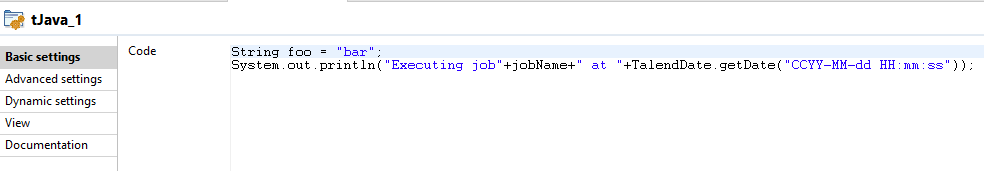


**8.Làm việc với Java :**

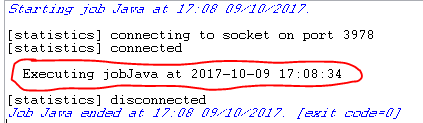
*a)Viết chương trình java cơ bản :*

 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"));

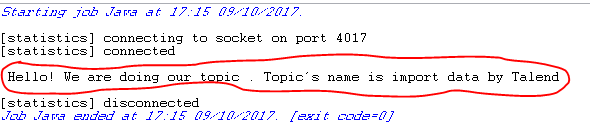
Sau đó vào quay lại màn hình chính của Job Design . Ta nhấn run để xem kết quả



Vậy là Câu lệnh đã được thực hiện

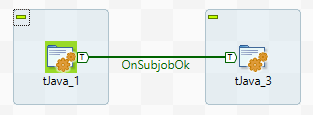
Vậy sẽ thử một ví dụ khác nhé ! Ở đây ta muốn in ra một câu trên Talend như sau “Hello ! We are doing our topic . Topic’s name is import data by Talend” thì cũng làm tương tự double chuột tJava rồi viết code vào như hình dưới đây

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"));***

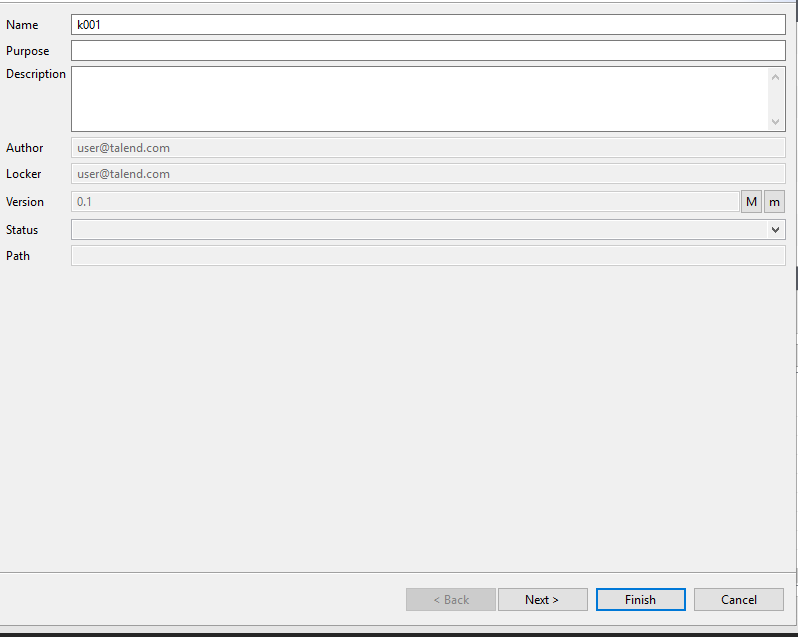
Cuối cùng nhấn Run Job để chạy sẽ nhìn thấy các biến được cài đặt trong lần đầu tiên sẽ được in chính xác trong lần thứ hai.  
Các biến *context* và *globalMap* được lưu trữ như trên toàn cầu có sẵn trong *Java  
hashmaps*. Điều này cho phép các giá trị được tham chiếu trong bất kỳ các  
*components*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

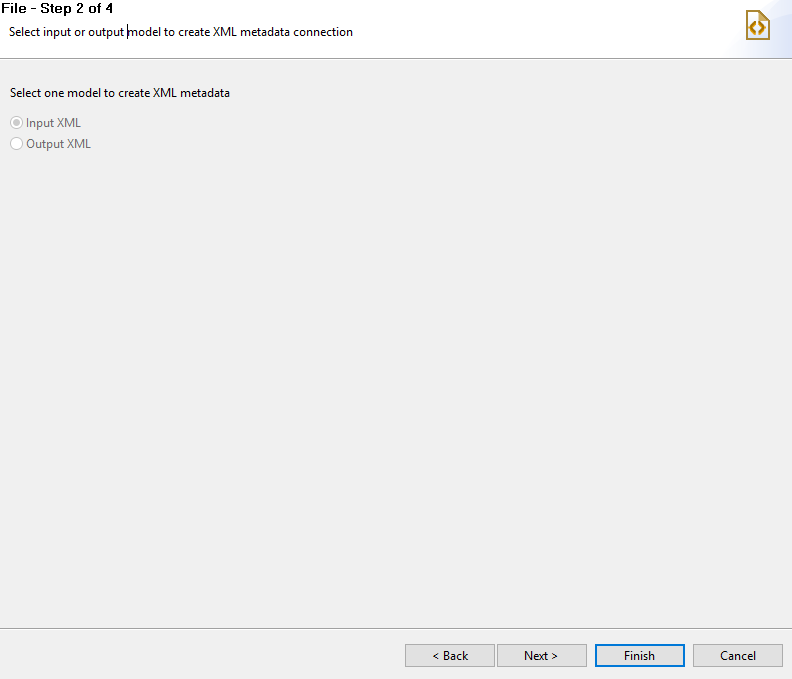
**9.Làm việc với XML:**

a)Sử dụng tXMLmap để đọc XML:

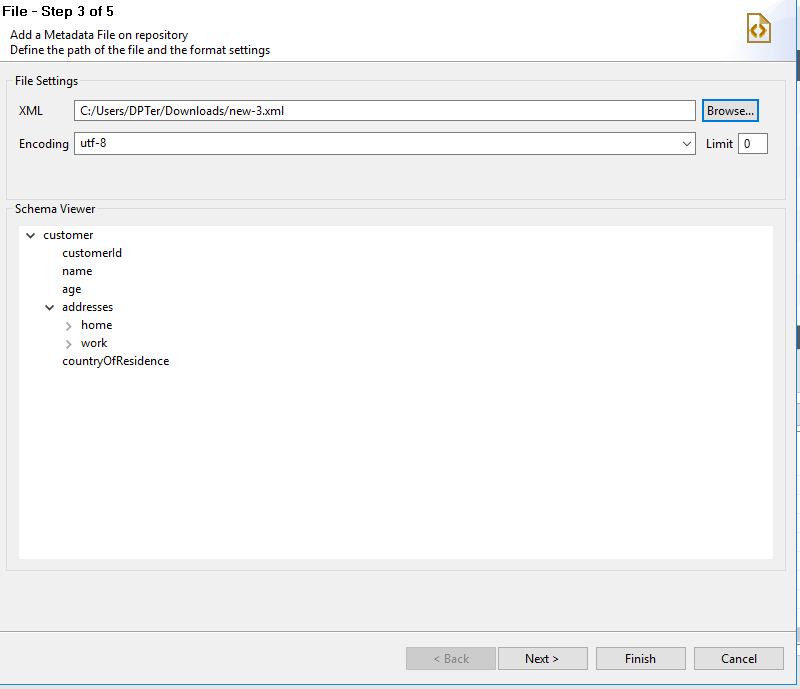
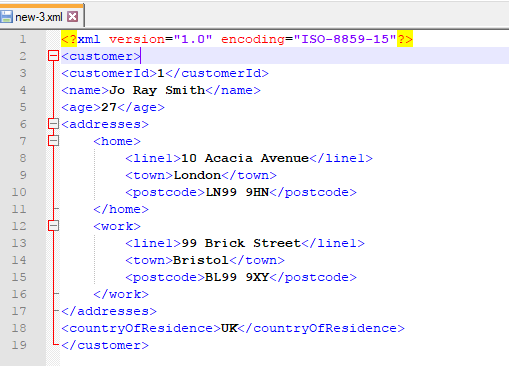
Để vào được XML trong talend , ta phải tạo file xml:

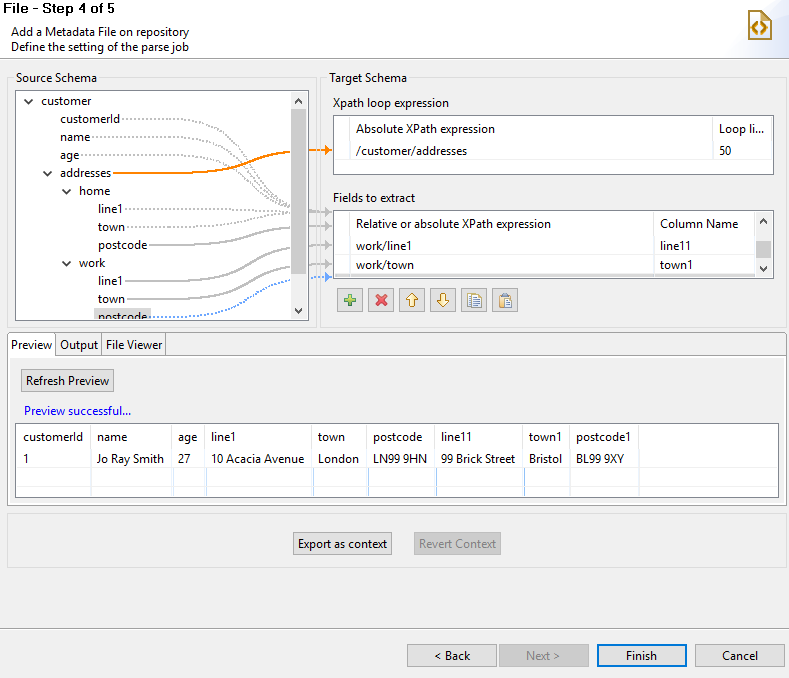
**B1**: Click vào **Metadata** -> **File XML** -> **Create file XML**(click chuột trái), rồi tạo tên cho file, **Next.**

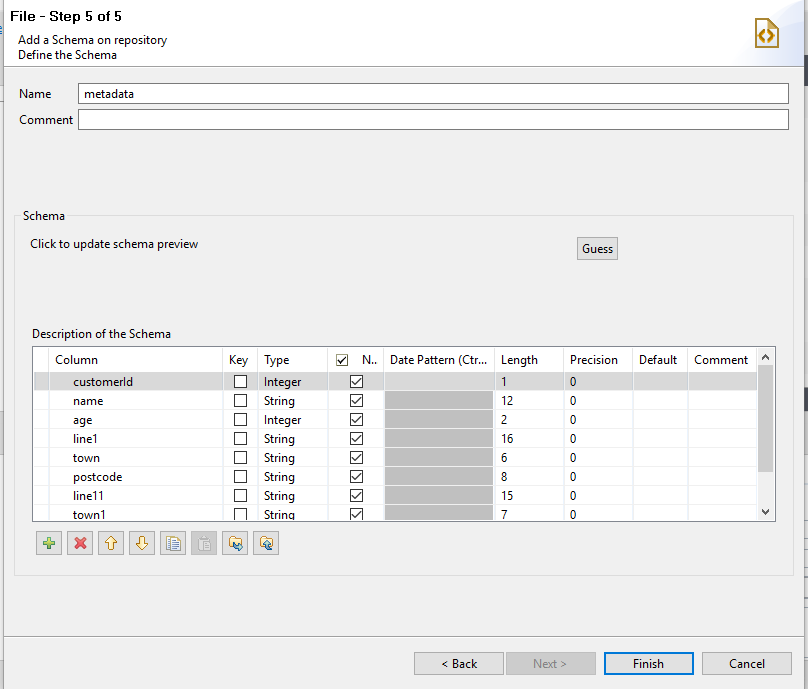
**B2**: Chọn 1 chế độ để tạo **XML metadata**: **Input** or **Output** , **Next.**



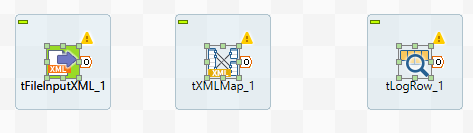
**B3**: Tìm 1 file có đuôi *(.xml)* rồi đưa vào **phần XML** , sau đó **phần Schema Viewer** sẽ hiện lên đoạn code có đuôi *(.xml)* có trên màn hình -> **Next**

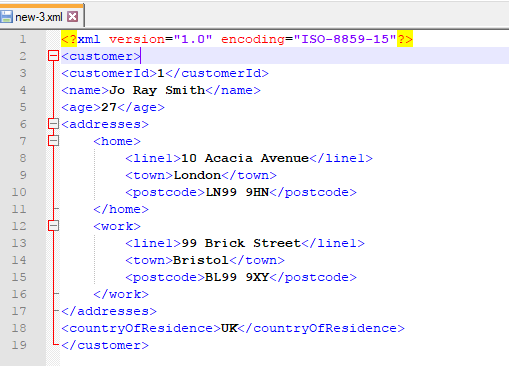


**B4**: Đưa các thông tin chính lên trên, đưa thông tin con xuống dưới -> rồi **click** vào **Preview – Refresh Preview** ,nó sẽ hiện lên bảng mà ta đã đưa vào . **Next**.

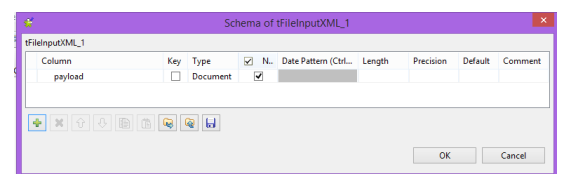
**B5**: Thêm hoặc thay đổi các **số liệu( kí tự)** trong **phần Description of the schema** -> **Finish**. Đã xong phần tạo **file XML**

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”.



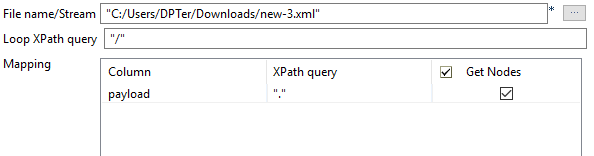
Đây là file có đuôi *(.xml)* [ Notepad++] để đưa vào *tFileInputXML****.***

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



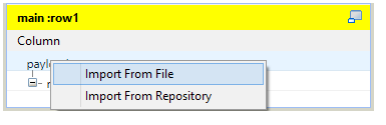
Thay đổi đường dẫn File name/Stream đến file xml vừa tạo và thay đổi LoopXpath query thành “/”.

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



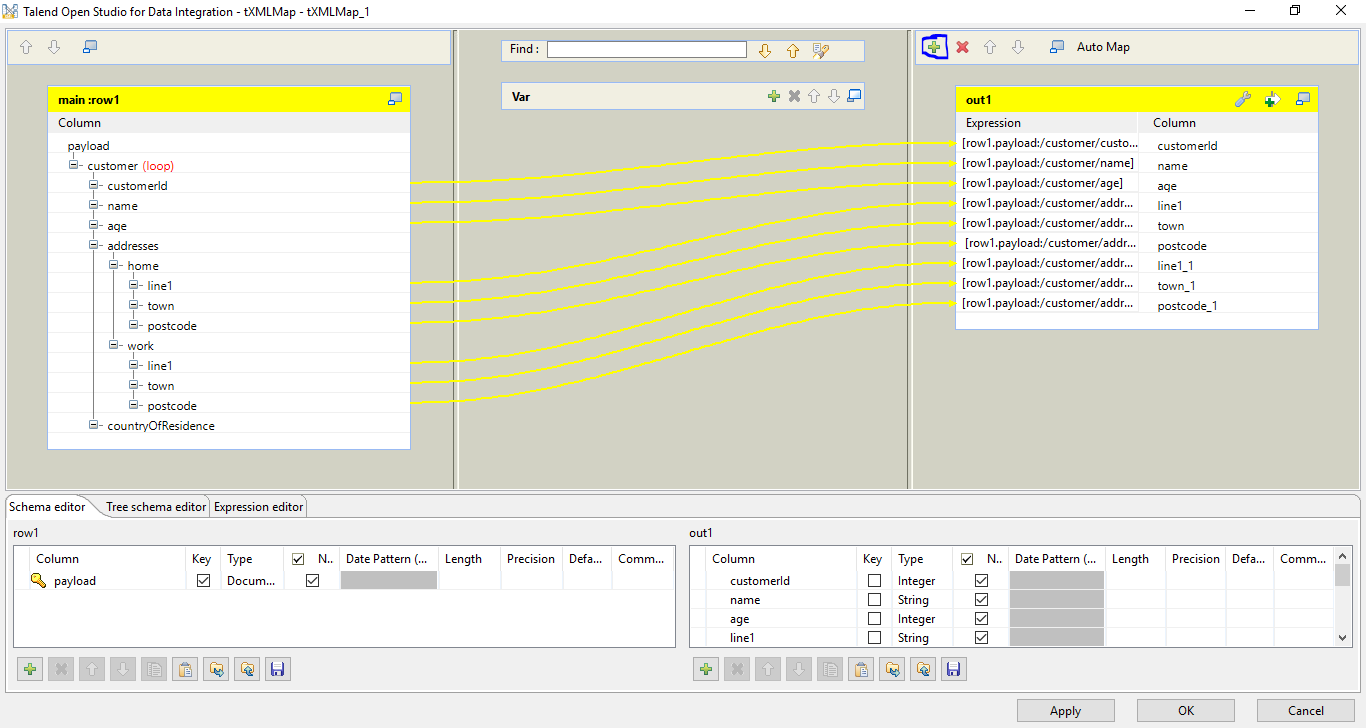
Kéo thả *tXMLMap* vào cửa sổ thiết kế và tạo liên kết từ *tFileInputXML* đến  
*tXMLMap*.

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.

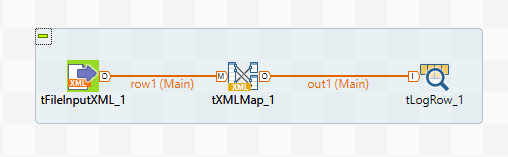


Click vào tXMLMap, nó sẽ hiện lên 1 bảng:

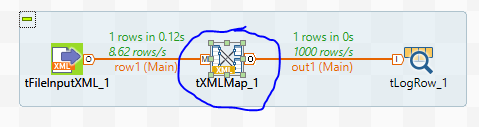
Nhấn 🞧 *button* và đặt tên là **customerOut**, sau đó lần lượt kéo thả các  
trường từ *file xml* sang *customerOut*.

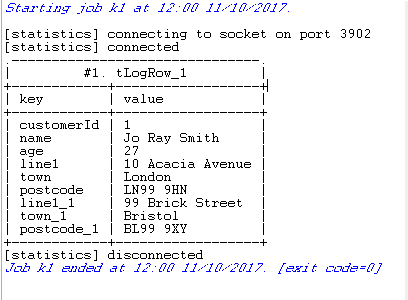


Thêm *tLogRow* vào cửa sổ làm việc và liên kết đến *tXMLMap*, sau đó **Run** *Job*bạn sẽ nhìn thấy tất cả dữ liệu *xml* được Talend san phẳng thành một hàng.



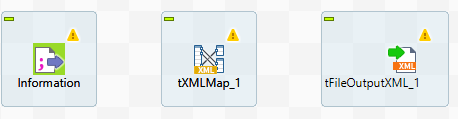
Cuối cùng,**Run**:



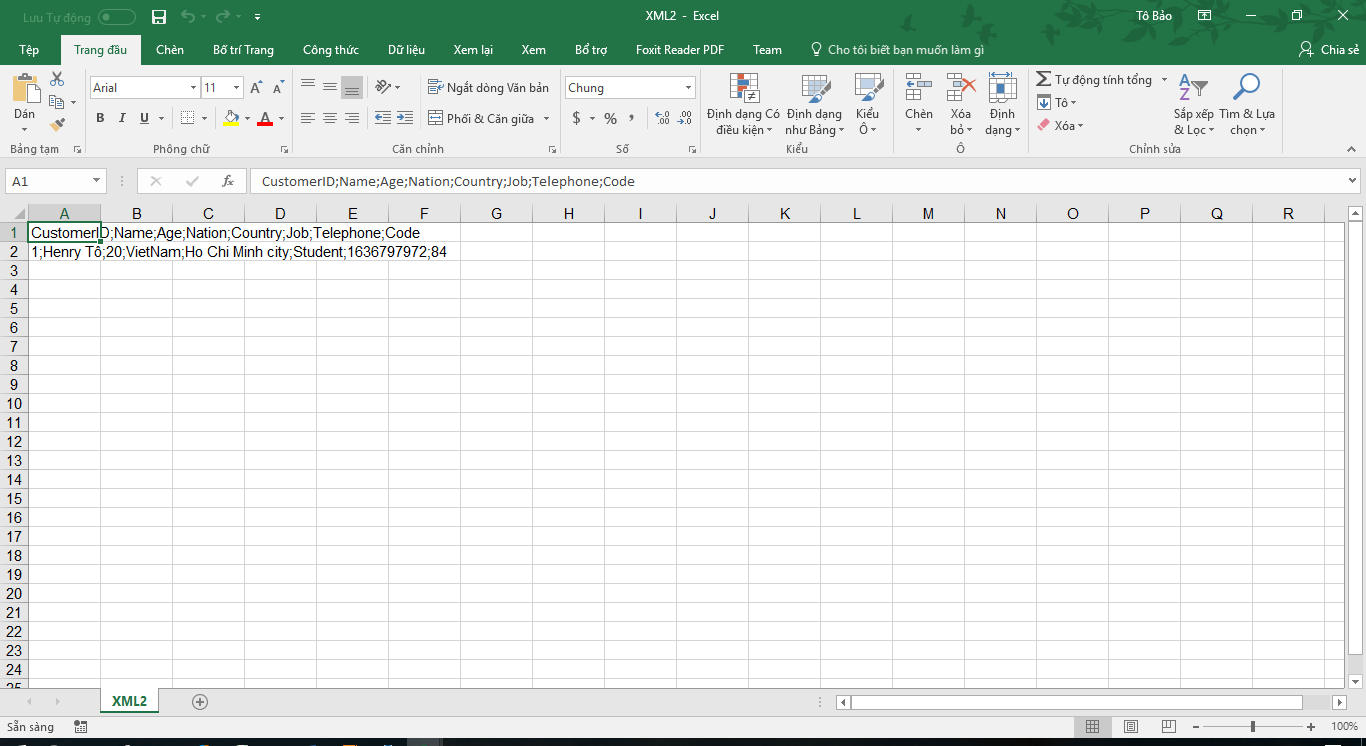


b)Sử dụng tXMLMap trong Talend để tạo ra một tài liệu về XML:

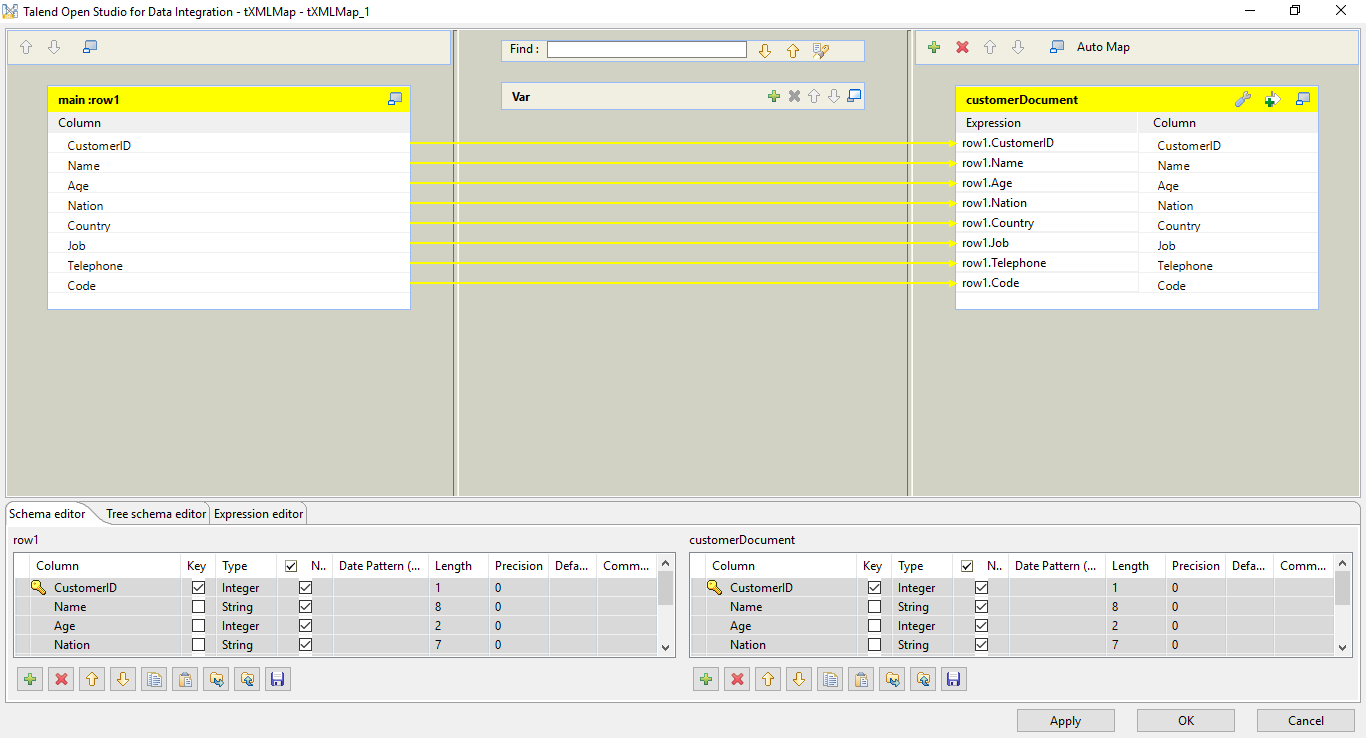
Đầu tiên thêm các thành phân vào Job Design :tFileInputDelimited,tXMLMap\_1 và tFileOutputXML



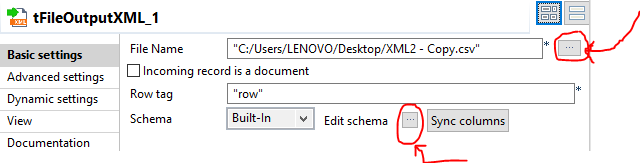
Trong đó tFileInputDelimited đầu vào bằng file XML2.csv có tên Information



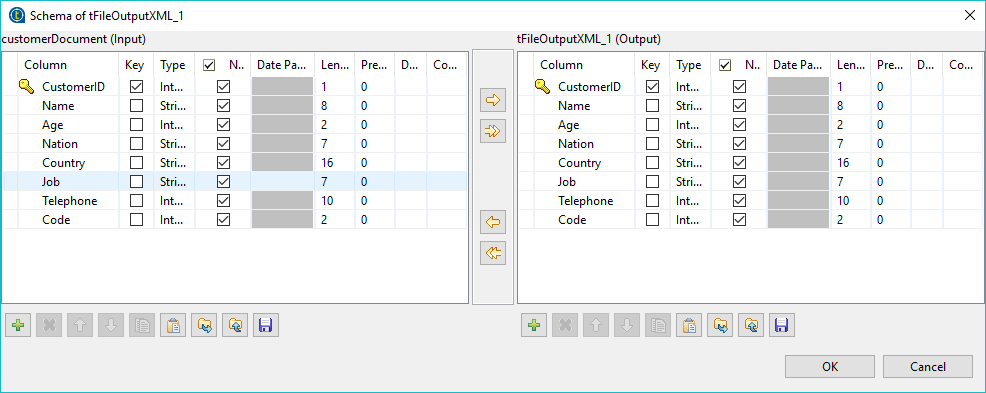
Thực kết nối Job bằng nhấn chuột phải chọn Row 🡪 Main . Sau đó cấu hình tXMLMap\_1 như hình dưới đây



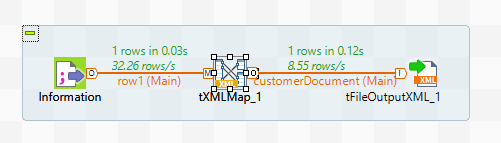
Sau đó double chuột vào tFileOutputXML . Chon Đường dẫn tới 1 file

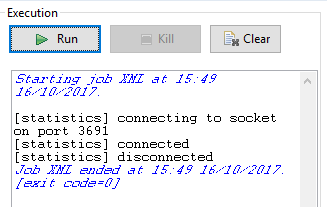


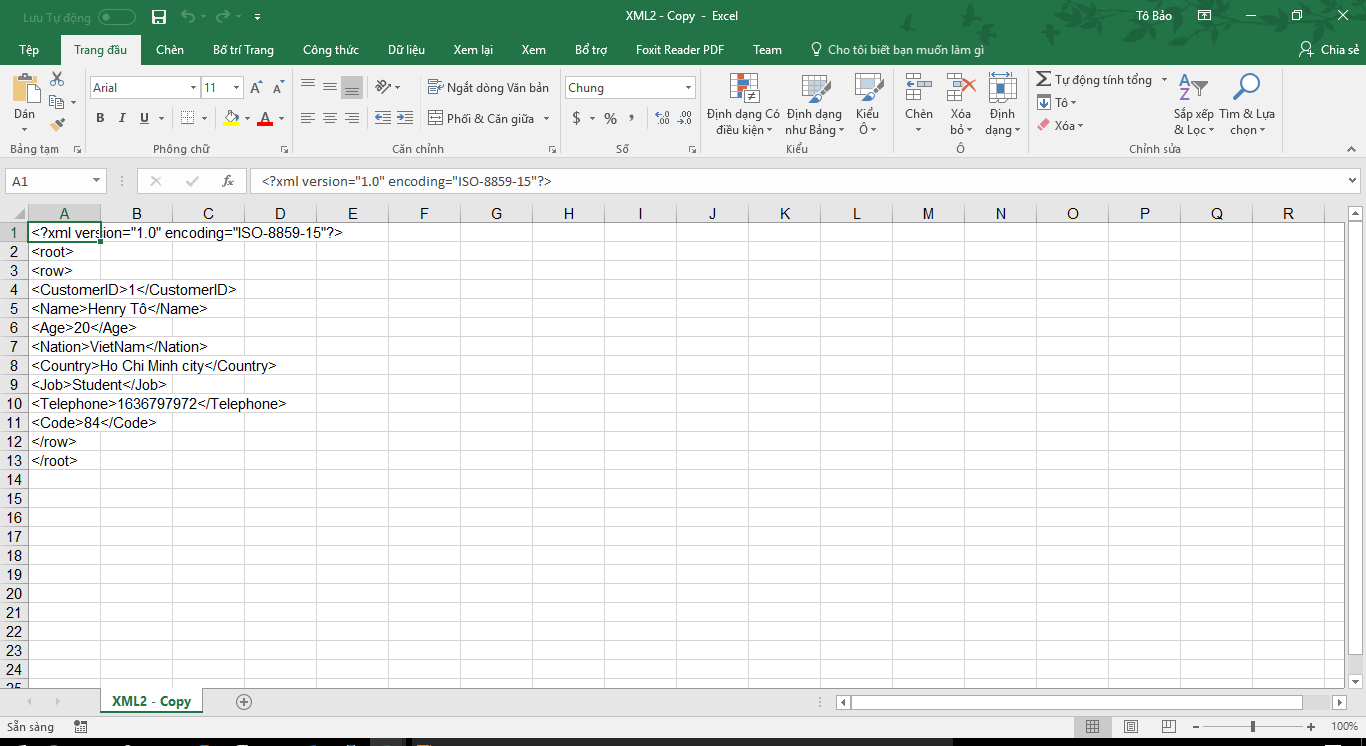
Rồi nhấn nút bên cạnh Edit schema sau đó cấu hình sao cho 2 đầu dữ liệu cùng field.Sau đó nhấn OK



Cuối cùng nhấn Run để chạy Job





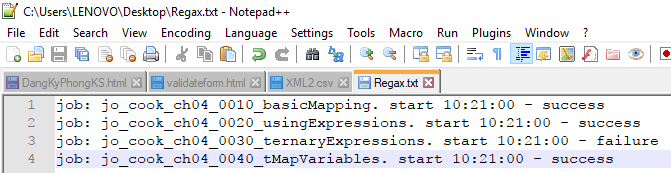
Mở file đầu ra để xem kêt quả 

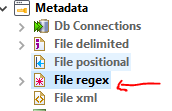
**10)Quản lý file**

a)Sử dụng Regular expression:

**- Regular expression** (*regex*) là một phương pháp mạnh mẽ cho mô hình kết hợp và  
thay thế trong nhiều ngôn ngữ lập trình. Một sử dụng thú vị cho các Regular  
expression là khi giao dịch với các định dạng đầu vào bất thường mà khó có thể  
mô tả bằng cách sử dụng định dạng giới hạn bình thường hoặc tập tin chiều rộng  
không cố định. Công thức này cho thấy cách *regex* có thể được sử dụng để xác  
định một tập hợp các cột đầu vào từ một dòng đầu vào có cấu trúc.

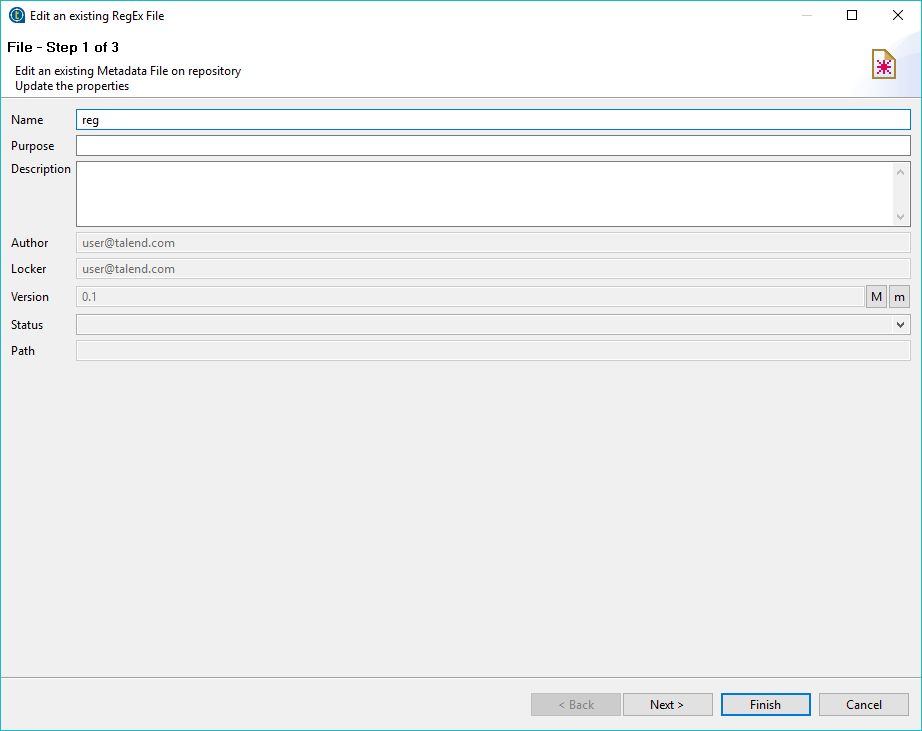
Đầu tiên có 1 file txt có cấu trúc như sau :

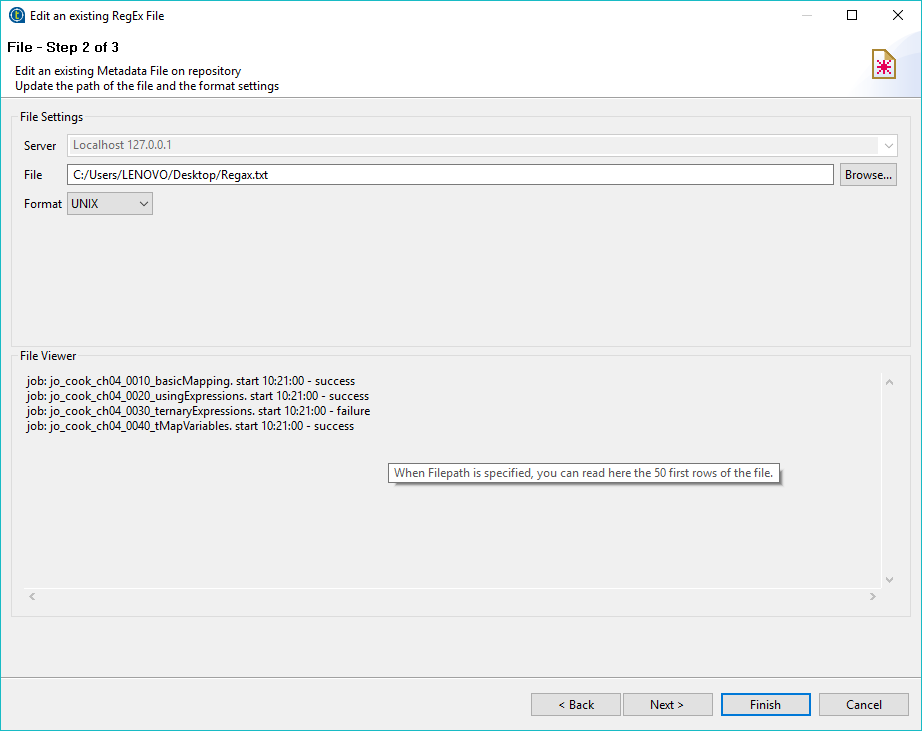


Tạo một Job mới thêm các thành phần : trong đó **tFileInputRegex**chứa dường  
dẫn đến file *txt đó* và **tLogRow**

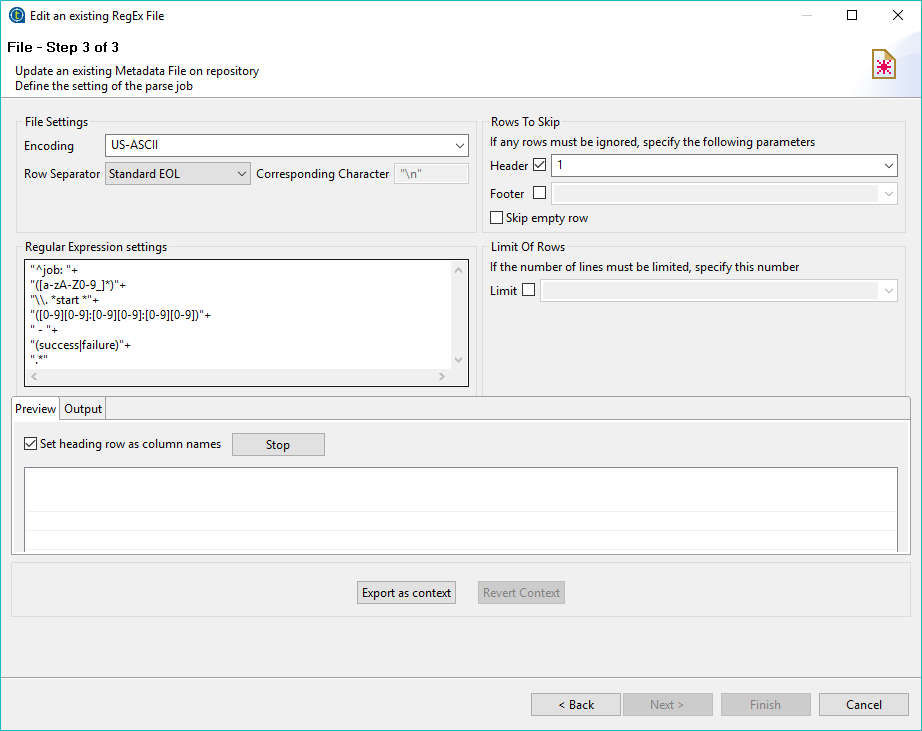
Để thêm thành phần tFileInputRegex- Trong metadata ta nhấn chuột phải File Regex và chọn Create file regex

Rồi nhập tên vào field Name rồi nhấn Next>

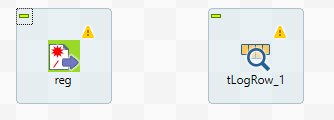


Sau đó chọn đường dẫn tới file txt bằng cách nhấn chuột vào Browse...

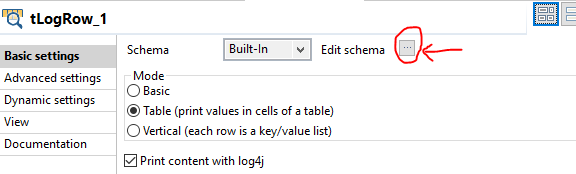
Rồi nhấn Next > để qua bước tiếp theo . Nhập đoạn code vào Regular Express setting



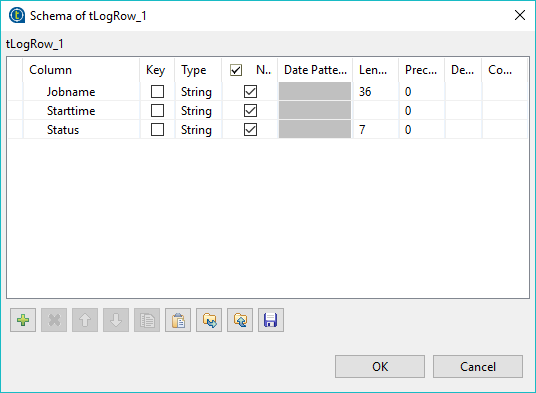
Nhấn Finish để hoàn tất . Sau đó kéo vào Job Design và thêm thành phần tLogRow



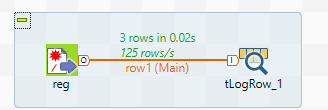
Ta double chuột tLogRow để cấu hình

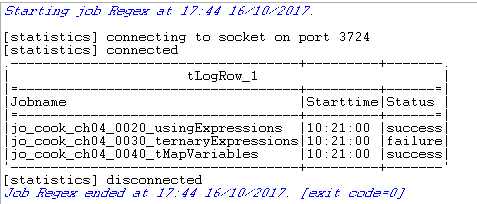


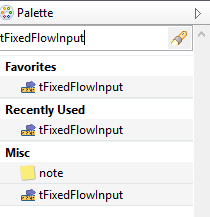
Check vào Table trong Mode và nhấn vào bên cạnh Edit schema

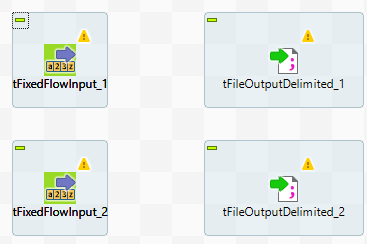


Thêm và Nhập tên các field.Sau đó nhấn OK và chạy chương trình



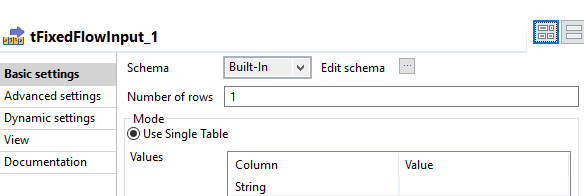
   
b)Thêm một bản ghi mới vào File

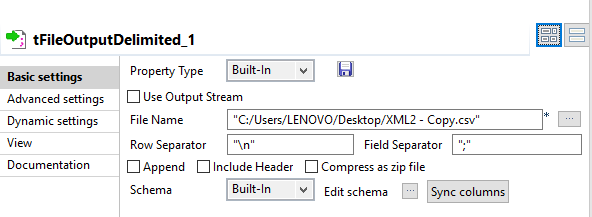
Tạo một Job mới đồng thời thêm các thành phần tFixedFlowInput và tFileOutputDelimited ( bằng cách gõ vào Palette sau đó kéo vào Job Design)như sau



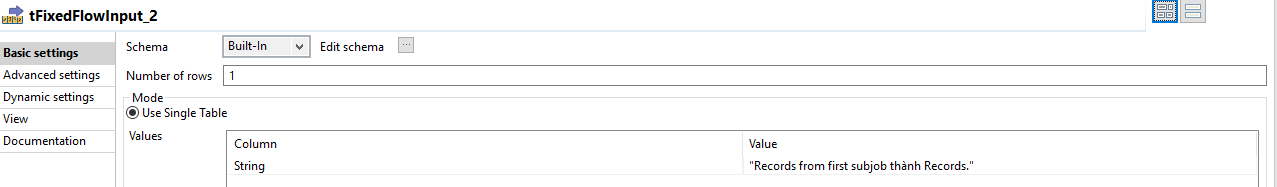
Sau đó kết nối với các thành phần với nhau và cấu hình từng thành phần

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

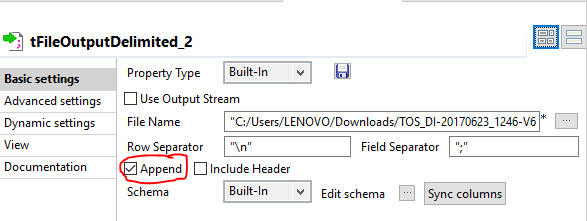


Chọn đường dẫn tới file cần ghi dữ liệu mới

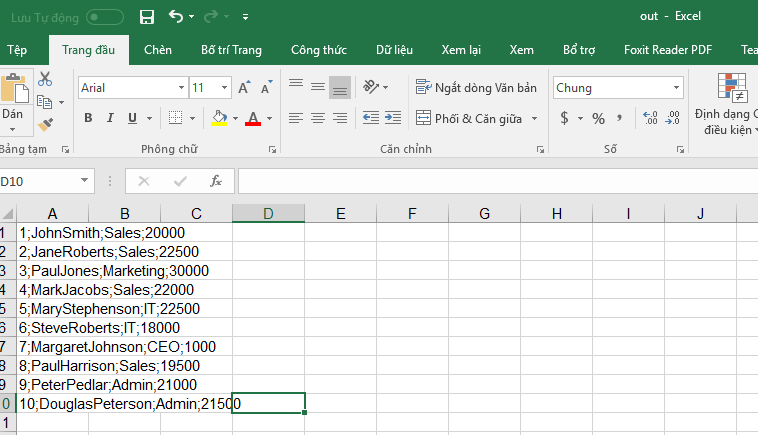
Tương tự tFixedFlowInput\_1 ,ta thêm một Column tên là String có giá trị là một chuỗi "Records from first subjob thành Records."



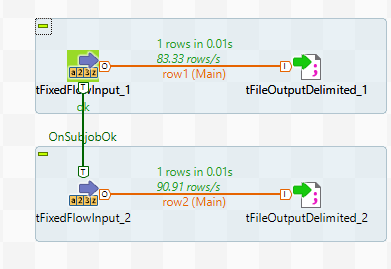
Lưu ý ở tFileOutputDelimitedw đánh check vào Append . Sau đó chọn đường dẫn đến file

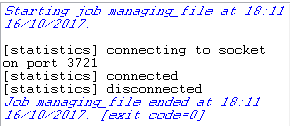


File đầu ra khi chưa run Job

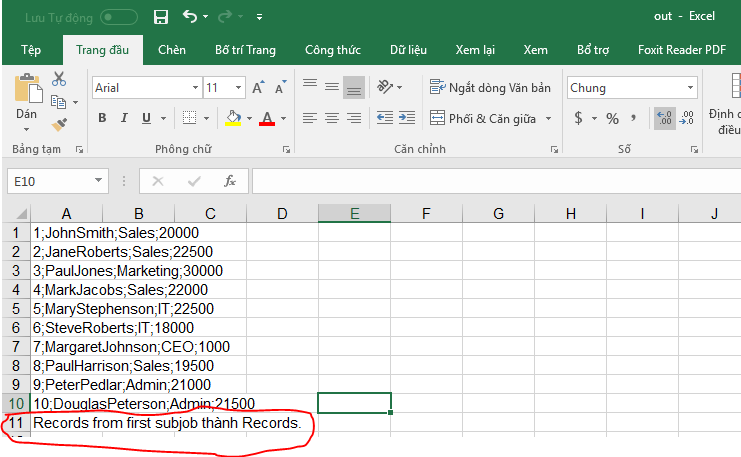


Cuối cùng đã hoàn tất cấu hình nhấn Run Job



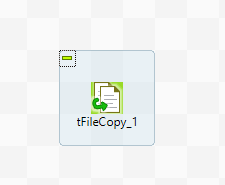


Vậy là sau khi chạy Job , mở lại các file đầu ra ta thấy một bản ghi mới đã được thêm vào . Dưới đây là file của tFileOutputDelimited\_2

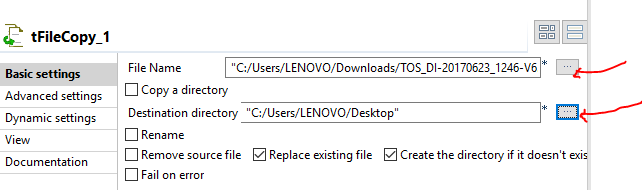


Còn đối với file của tFileOutputDelimited\_1 do nhận giá trị NULL nên ko có bản ghi mới

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

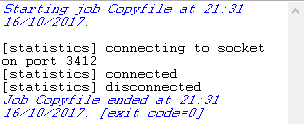
 \*Sao chếp 1 file vào thư mục khác:

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



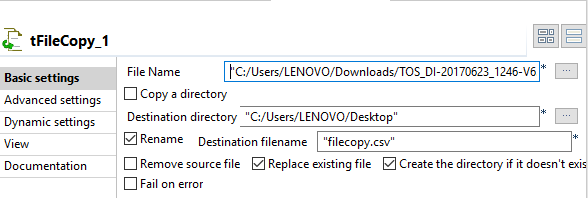
Thiết lập giá trị **File name**: chứa đường dẫn đến *file* cần copy.  
Thiết lập giá trị **Destination directory**: là nơi lưu tập tin được copy ra.

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

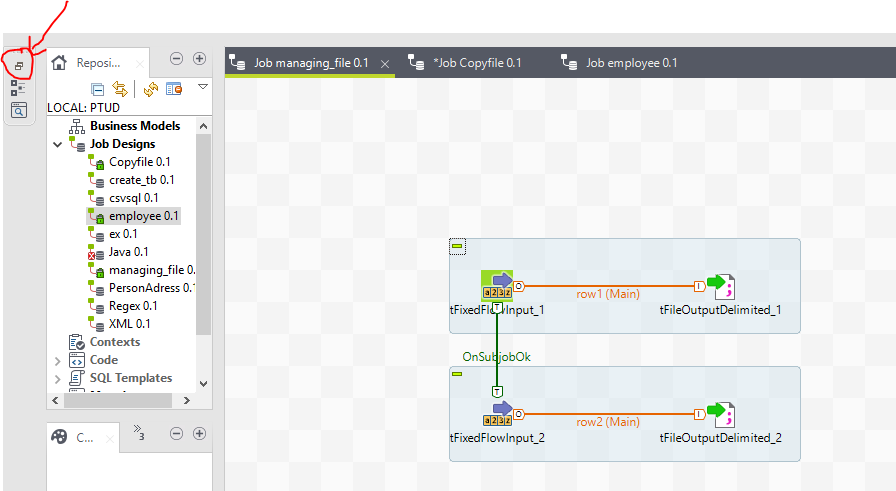
Cũng làm như trên . Nhưng khi double , ta check vào Rename và nhập tên mới trong Destination filename . Sau đó Run Job



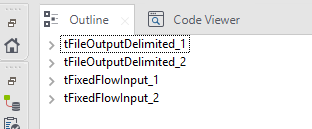
Nếu bạn muốn sau khi copy và đổi tên , bạn muốn file ở đường dẫn cũ . Nhấn check vào Remove source file . Cách này tương đương với di chuyển file hoặc xóa file

**11.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 mỗi Job Design thực hiện .Để làm việc này ta mở một Job và nhấn vào Restore như hình như dưới



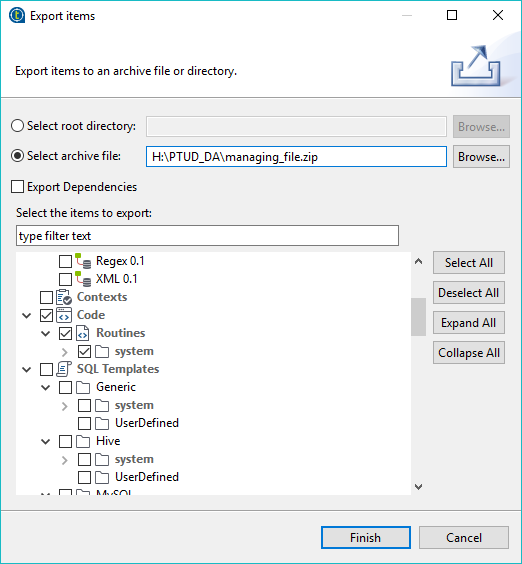
Khi nhấn vào xuất hiện một hộp thoại có 2 tag OutLine ( chứa các thành trong Job Design) và Code Viewer ( chứa code ) . Ta nhấn vào Code Viewer để xem code



Như vậy trong Talend ta 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

12)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. 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. Từ Talend 4 nâng cấp thanh Bar với việc bổ sung các công cụ tổng thể quản lý dữ liệu và hỗ trợ cho Apache Hadoop Distributed File System, đặt nền móng cho quy mô lớn, khả năng phân tích "dữ liệu lớn". Mở khóa tất cả các tính năng mà mức độ doanh nghiệp đòi hỏi một giấy phép đăng ký kinh doanh, nhưng mã nguồn mở miễn phí Talend Open Studio, Talend Open Profiler, và Talend MDM hiện tại vượt ra ngoài định hướng khung cơ bản.