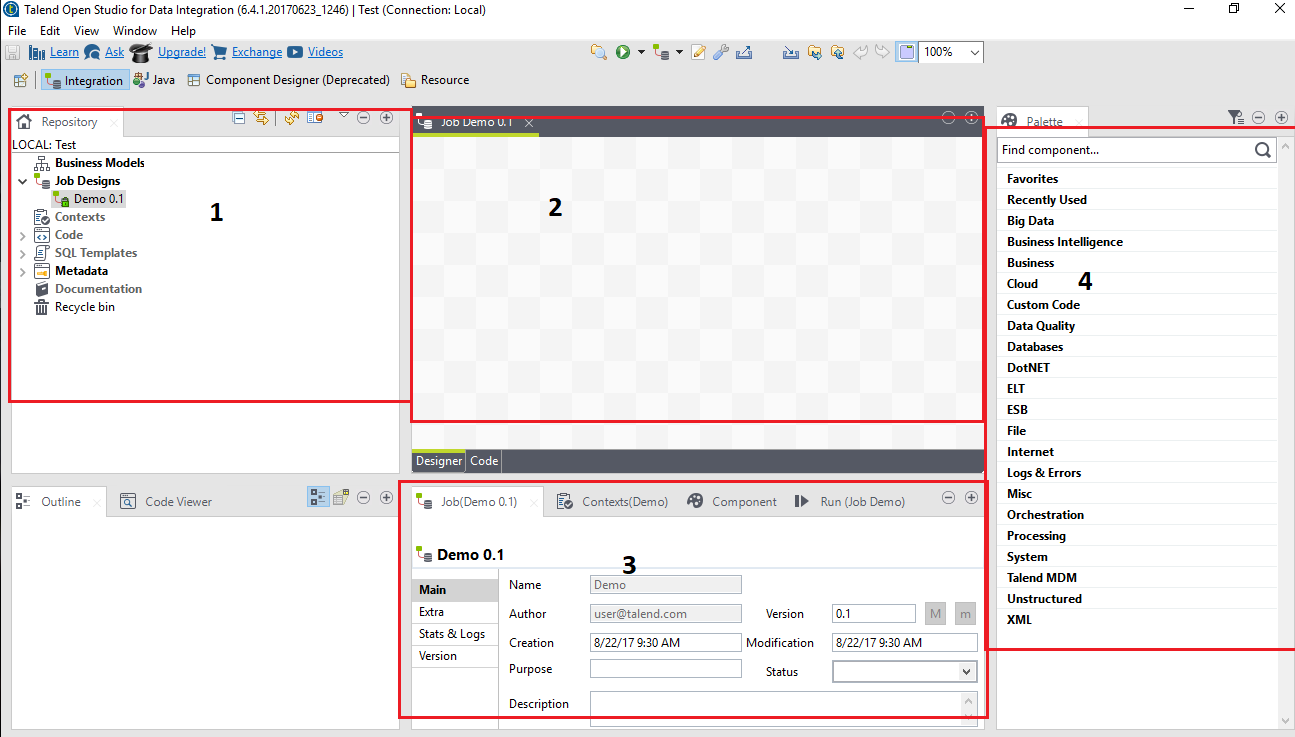
**Tìm hiểu từ Talend**

* Talend là một nhà cung cấp phần mềm mã nguồn mở. cung cấp khả năng tích hợp dữ liệu (*data intergration*), quản lý dữ liệu (*data management*), tích hợp úng dụng doanh nghiệp (*enterprise application intergration)*, các phần mềm và dịch vụ lớn (big data)..
* Talend Open Studio là một tập hợp 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 dữ liệu tích hợp. Cung cấp các nền tảng thônsg nhất làm cho quản lý dữ liệu và tích hợp ứng dụng một cách dễ dàng hơn bằng cách cung cấp một môi trường thống nhất toàn bộ vòng đời của doanh nghiệp qua các ranh giới hoạt đọng của doanh nghiệp.

Sau khi cài đặt được Talen Open Studio for Data Intergration, khởi động Talend chúng ta có thể Create, Open, Delete, Import một Project hoặc xem 1 Demo Project..

Sau khi tạo mới 1 project cửa sổ làm việc xuất hiện 4 menu cho phép ta thao tác dữ liệu:



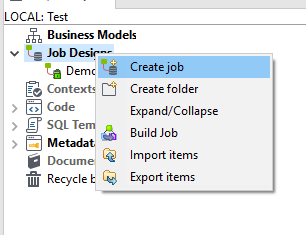
Menu 1: tại đây cho phép người dùng thiết kế a Jobs, hay thao tác trên Database, Hadoop, XML..

Menu 2: thao tác trình lọc, cạp nhật, tải, chỉnh sửa dữ liệu..

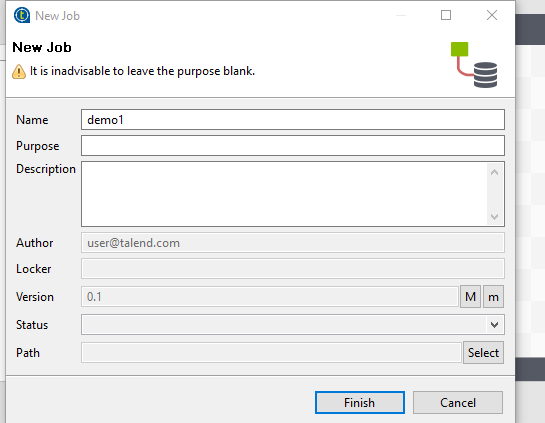
Menu 3: dùng để thiết lập các thông số thoe yêu cầu bài toán được thực hiện, chạy và kiểm lỗi các thao tác đang thức hện

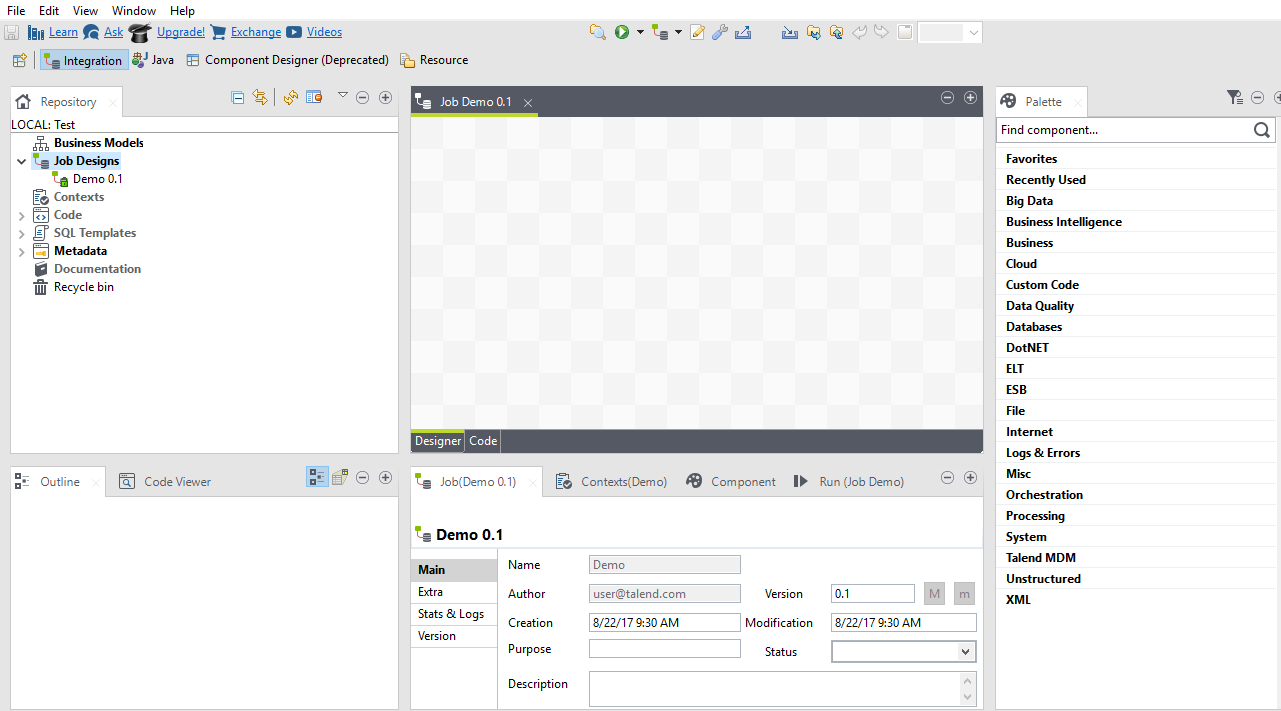
Menu 4: cửa sổ này chứa các thuộc tính để thực hiện các theo tác trên menu 2.

Để làm ra 1 dự án trươc tiên cần Create a jobs



Sau khi click vào Create job hộp thoại new job xuất hiện

3. Một không gian làm việc thiết kế rỗng sẽ mở ra hiển thị tên của Job dưới dạng nhãn tab.



Công việc bạn tạo ra bây giờ được liệt kê dưới nút Job Designs trong khung nhìn Repository tree.

Bạn có thể mở một hoặc nhiều công việc tạo ra bằng cách đơn giản nhấp đúp vào nhãn Job trong khung nhìn Repository tree.

Có một số cách để thêm một thành phần vào không gian làm việc thiết kế. Bạn có thể:

• tìm thành phần của bạn trên Palette bằng cách nhập (các) từ khoá tìm kiếm vào trường tìm kiếm của Palette và thả

Nó vào không gian làm việc thiết kế.

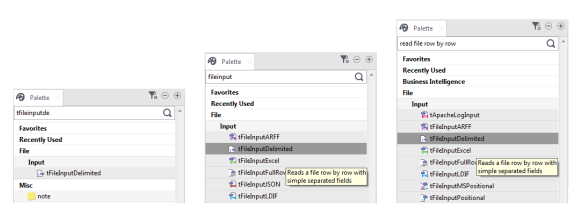
• thêm một thành phần bằng cách gõ trực tiếp (các) từ khoá tìm kiếm của bạn vào không gian làm việc thiết kế.

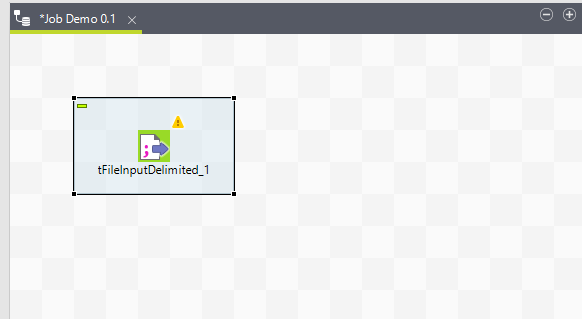
• thêm một thành phần đầu ra bằng cách kéo từ một thành phần đầu vào đã có trên không gian làm việc thiết kế.

• kéo và thả một mục metadata tập trung từ nút Metadata vào không gian làm việc thiết kế, và sau đó chọn thành phần quan tâm từ hộp thoại Thành phần.

Phần này mô tả ba phương pháp đầu tiên. Để biết chi tiết về cách hủy một thành phần từ siêu dữ liệu Node, xem Quản lý siêu dữ liệu.

* khi bạn muốn đọc một tập tin văn bản đầu tiên, bạn sẽ sử dụng thành phần tFileInputDelimited.

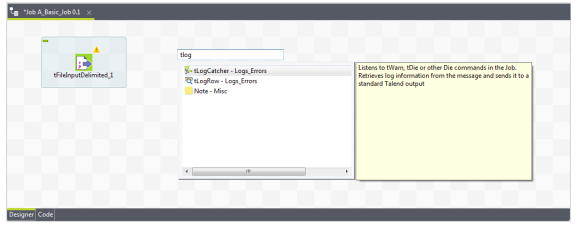




Nhấp vào nơi bạn muốn thêm các thành phần trên không gian làm việc thiết kế, và gõ từ khóa của bạn, có thể Là tên đầy đủ hoặc một phần của hợp phần, hoặc một cụm từ mô tả chức năng của nó nếu bạn không biếtTên. Trong ví dụ của chúng tôi, bắt đầu nhập tlog.

* *Để sử dụng cụm từ mô tả làm từ khoá cho tìm kiếm mờ, đảm bảo tìm kiếm Cũng từ Trợ giúp khi thực hiện Tìm kiếm thành phần hộp kiểm tra được chọn trên khung nhìn Preferences> Palette Settings. Để biết thêm thông tin Xem Tùy chọn Palette (Talend> Palette Settings).*

Một hộp danh sách sẽ xuất hiện bên dưới trường văn bản hiển thị tất cả các thành phần phù hợp theo thứ tự chữ cái.



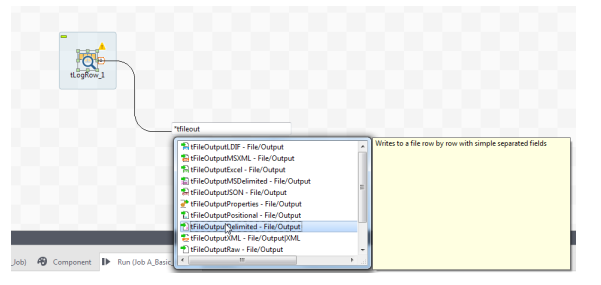
Nhấp đúp vào thành phần mong muốn để thêm nó vào không gian làm việc.

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

1. Nhấp vào thành phần tLogRow để hiển thị biểu tượng o gắn vào nó.

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

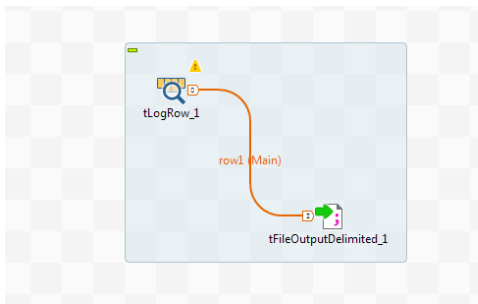
Một trường văn bản và một danh sách thành phần xuất hiện. Danh sách thành phần hiển thị tất cả các thành phần có thể được kết nối Với các thành phần đầu vào



3. 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 của Lãi suất, 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 một Row > Main connection..

*Để sử dụng cụm từ mô tả làm từ khoá cho tìm kiếm mờ, đảm bảo tìm kiếm Cũng từ Trợ giúp khi thực hiện Tìm kiếm thành phần hộp kiểm tra được chọn trên khung nhìn* ***Preferences> Palette Settings****. Để biết thêm thông tin, Xem Tùy chọn* ***Palette (Talend> Palette Settings)***



Bây giờ các thành phần đã được thêm vào không gian làm việc, chúng phải được kết nối với nhau. Các thành phầnKết nối với nhau tạo thành một subjob. Việc làm bao gồm một hoặc nhiều người phụ việc thực hiện các quy trình khác nhau.

Trong ví dụ này, vì các thành phần tLogRow và tFileOutputDelimited đã được kết nối, bạn chỉ cần Để kết nối tFileInputDelimited với thành phần tLogRow.

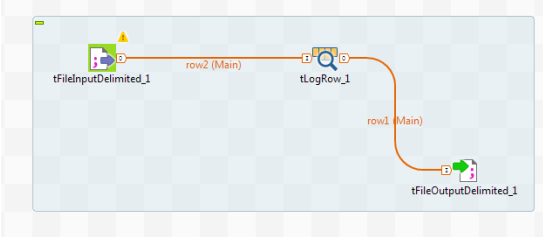
Để kết nối các thành phần với nhau, sử dụng một trong các phương pháp sau:

Nhấp chuột phải và nhấp lại vào

1. Nhấp chuột phải vào thành phần nguồn, tFileInputDelimited trong ví dụ này.

2. Trong trình đơn ngữ cảnh mở ra, chọn loại kết nối bạn muốn sử dụng để liên kết các thành phần, Row> Main trong ví dụ này.

3. Nhấp vào thành phần đích để tạo liên kết, tLogRow trong ví dụ này



Lưu ý rằng một vòng tròn chữ thập màu đen được hiển thị nếu thành phần đích không tương thích với liên kết.



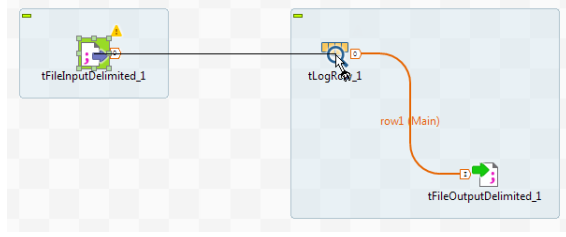
*Theo tính chất và vai trò của các thành phần bạn muốn liên kết với nhau, một số loại liên kết Sẵn sàng. Chỉ có các kết nối được ủy quyền được liệt kê trong trình đơn ngữ cảnh.*

**Kéo và thả**

1. Nhấp vào thành phần đầu vào, tFileInputDelimited trong ví dụ này.

2. Khi biểu tượng O xuất hiện, nhấp chuột vào nó và kéo con trỏ tới phần đích, tLogRow trong ví dụ này.

A Row> Kết nối chính được tạo tự động giữa hai thành phần.



Mặc dù phương pháp này yêu cầu các bước hoạt động ít hơn, nó chỉ hoạt động với các loại kết nối **Row: Main,Lookup, Output, Filter, và Reject,** tùy thuộc vào tính chất và vai trò của các thành phần mà bạn đang kết nối.

Bạn cũng có thể thả các thành phần ở giữa liên kết Hàng. Để biết thêm thông tin, xem Cách thêm một thành phần Giữa hai thành phần kết nối.

Để biết thêm thông tin về cách sử dụng các loại kết nối, xem Sử dụng kết nối.

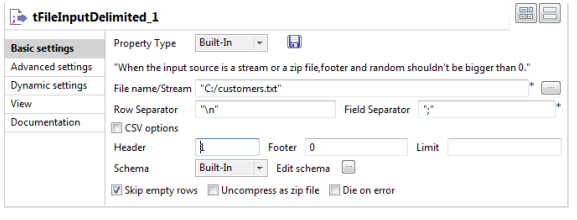
***4.2.4. Cấu hình các thành phần***

Bây giờ các thành phần được liên kết, các thuộc tính của chúng nên được định nghĩa.

Để biết thêm chi tiết về các thuộc tính của các thành phần, xem Cách xác định thuộc tính thành phần.

Cấu hình thành phần tFileInputDelimited

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



2. Nhấp chuột vào nút [...] bên cạnh trường **File Name / Stream**.

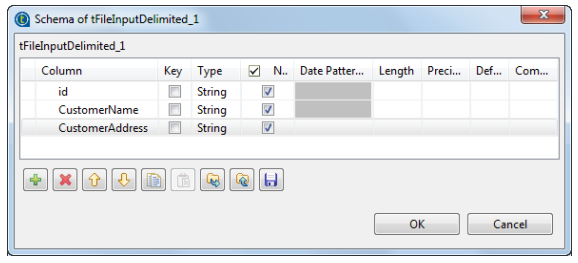
3. Duyệt hệ thống của bạn hoặc nhập đường dẫn đến tệp tin đầu vào, customer.txt trong ví dụ này.

4. Trong trường Header, nhập 1.

5. Nhấp vào nút [...] bên cạnh Chỉnh sửa giản đồ.

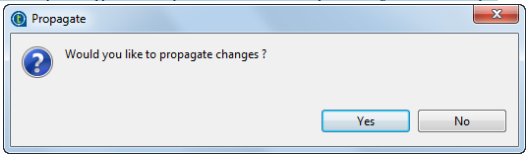
6. Trong Trình chỉnh sửa Giản đồ được mở ra, nhấp ba lần nút [+] để thêm ba cột.

7. Đặt tên cho ba cột id, CustomerName và CustomerAddress tương ứng và nhấn OK để đóng Biên tập viên.



8. Trong cửa sổ pop-up mở ra, nhấp OK chấp nhận việc truyền bá các thay đổi.

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

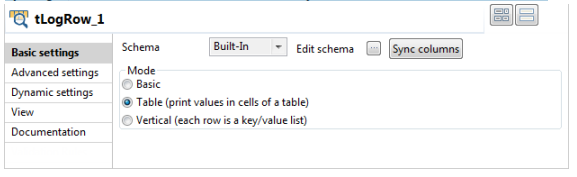


Cấu hình thành phần tLogRow

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

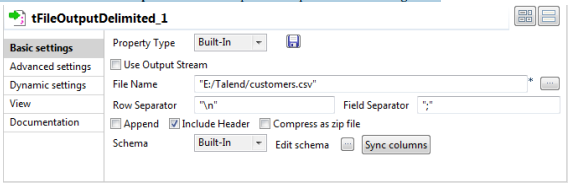
2. Trong vùng Chế độ, chọn Bảng (in giá trị trong ô của bảng).

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



Cấu hình thành phần tFileOutputDelimited

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



2. Nhấp chuột vào nút [...] bên cạnh trường File Name.

3. Duyệt hệ thống của bạn hoặc nhập đường dẫn đến tệp tin đầu ra, customers.csv trong ví dụ này.

4. Chọn hộp kiểm Bao gồm Tiêu đề.

5. Nếu cần, nhấp vào nút Đồng bộ hóa các cột để lấy lược đồ từ thành phần đầu vào.

***4.2.5. Thực hiện công việc***

Bây giờ các thành phần được cấu hình, Job có thể được thực thi.

Để làm như vậy, tiến hành như sau:

1. Nhấn **Ctrl + S** để lưu công việc.

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

Tệp được đọc từng hàng và các trường được giải nén được hiển thị trên bàn điều khiển Chạy và ghi vào tệp được chỉ định Tập tin đầu ra

