**BÁO CÁO TUẦN 4**

**Môn:Phát triển ứng dụng**

**Topic 2 : Import data by Talend**

**\*\*Group:**

**1.Nguyễn Duy Phong 15030081**

**2.Tô Nguyễn Bảo(NT) 15027521**

**3.Trịnh Hưng Khang 15035581**

**I.Kế hoạch đặt ra và trạng thái hoàn thành**

|  |  |  |
| --- | --- | --- |
| Nhiệm vụ | Thành viên đảm nhận | Hoàn thanh mục tiêu |
| -Tìm hiểu thế nào Job Design ?  -Tìm hiểu và nêu các bước  +Tạo 1 Job và thêm các thành phần vào trong Job  +Loại bỏ thành phần đầu tiên từ Palette  + Thêm thành phần thứ hai trong Job bằng cách gõ vào thiết kế không gian làm việc  +Thêm một thành phần đầu ra bằng cách kéo từ một đầu vào  một | **Tô Nguyễn Bảo** | Hoàn thành |
| \*Nêu các bước tìm hiểu:  - Kết nối các thành phần trong Job với nhau  - Cấu hình các thành phần trong Job  +Cấu hình thành phần tFileInputDelimited  + Cấu hình thành phần tLogRow  + Cấu hình tFileOutputDelimited component  - Cho thực hiện Executing the Job | **Trịnh Hưng Khang** | Không thực hiện làm trễ báo cáo . Buộc 2 thành viên còn lại phải làm thay |
| -Tiến hành các thao ở trên. Chụp hình các thao tác thực hiện  - Đồng thời demo các bước làm trên cho các thành viên nhóm ở buổi thực hành  -Đọc được file CSV và lưu được CSV vào các file khác | **Nguyễn Duy Phong** | Đang tiến hành tới tuần sau |

**II.Nội dung thực hiện**

1. **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.

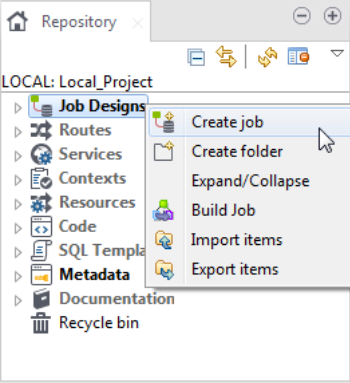
• Thiết lập kết nối và mối quan hệ giữa các thành phần để xác định trình tự và bản chất của hành động.

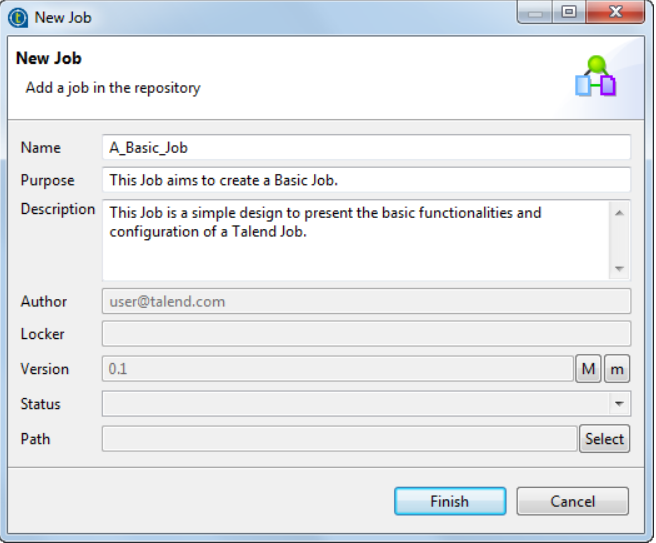
• Truy cập mã vào bất kỳ lúc nào để chỉnh sửa hoặc ghi lại các thành phần trong Job được thiết kế.

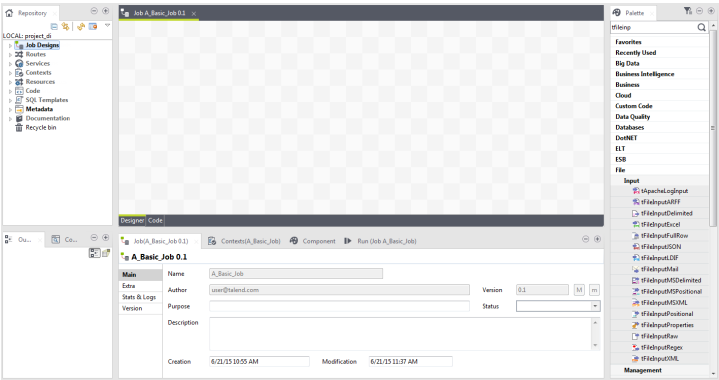
• Tạo và thêm các mục vào kho lưu trữ để sử dụng lại và chia sẻ mục đích (trong các dự án khác hoặc công việc hoặc với các người dùng).

1. **Bắt đầu với 1 Job cơ bản:**
2. ***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

1. ***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

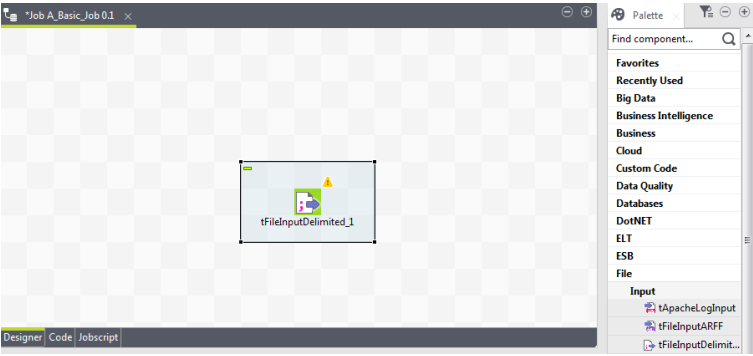
\* Dropping the first component from the Palette

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



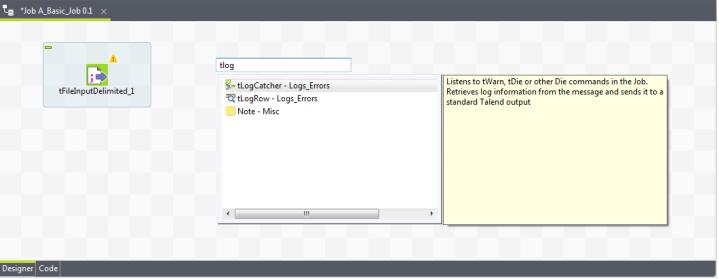
,,c c

**\* 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

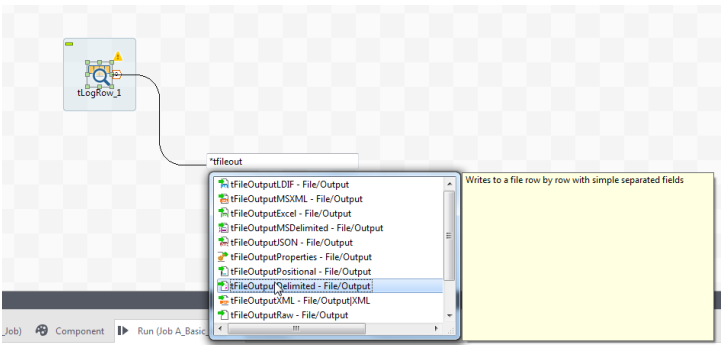
\*Adding an output component by dragging from an inputone

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.

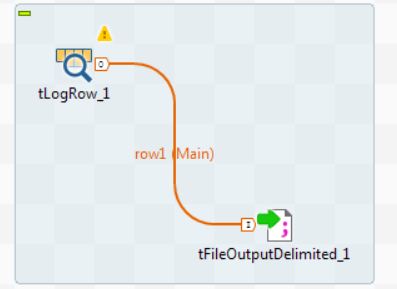
B1: Nhấp vào thành phần **tLogRow** để hiển thị biểu tượng **o** được gắn vào nó.

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.



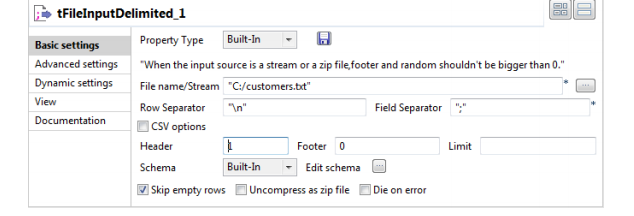
B3. Để thu hẹp tìm kiếm, nhập trường văn bản tên của thành phần mà bạn muốn thêm hoặc một phần của nó, hoặc một cụm từ mô tả chức năng của thành phần nếu bạn không biết tên của nó, và sau đó nhấp đúp vào thành phần tFileOutputDelimited trong ví dụ này, trên danh sách thành phần để thêm nó vào không gian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



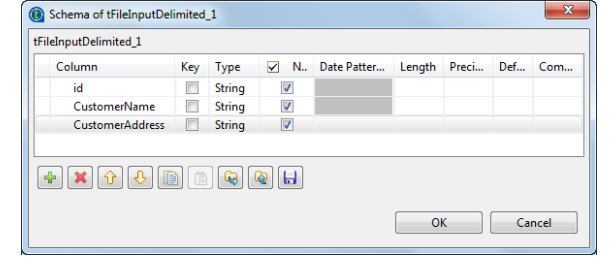
**+CẤU HÌNH CÁC THÀNH PHẦN**

\*Cấu hình thành phần **tFileInputDelimited**

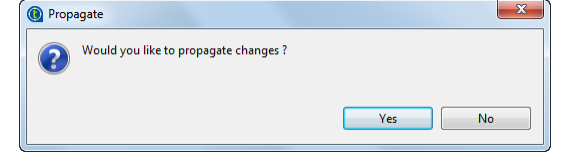
1. Nhấp đúp vào thành phần **tFileInputDelimited** để mở khung nhìn **Basic settings.**



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

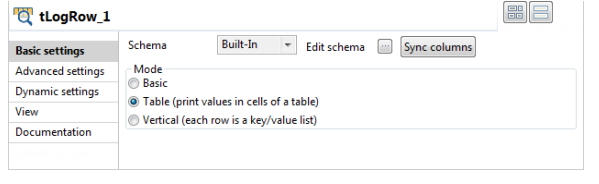


Cấu hình thành phần **tLogRow**

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

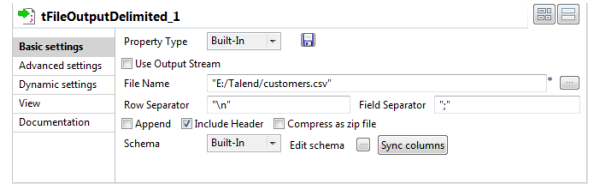
2. Trong vùng **Mode**, chọn **Table (print values in cells of a table).**

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.

**Thực hiện công việc**

1. Nhấn **Ctrl + S** để lưu job

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

Ví dụ minh họa

