**Đề Tài: Import Data By Talend.**

|  |  |  |
| --- | --- | --- |
| STT | Thành Viên | MSSV |
| 1 | Nguyễn Công Thịnh | 15024631 |
| 2 | Võ Đông Lâm | 15014921 |

Mục đích lập nhóm: Nghiên cứu và hoàn thành đề tài nhóm đã chọn.

Nền tảng thực hiện: Talend

**Kế hoạch thực hiện đồ án**

|  |  |
| --- | --- |
| **Buổi thực hành** | **Nội Dung** |
| 1 | Tìm hiểu Talend là gì? Download, cài đặt về Talend. |
| 2 | Metadata và Schemas |
| 3 | Mapping data |
| 4 | Sử dụng java trong Talend |
| 5 | Làm việc với Database |
| 6 | Managing files |
| 7 | Làm việc với XML, Queues và Wed Services |
| 8 | Tổng kết những gì nhóm làm được |
| 9 |  |

**Buổi thực hành 1( Ngày 15/08/2017): Tìm hiểu Talend là gì? Download, cài đặt về Talend.**

* 1. **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 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.

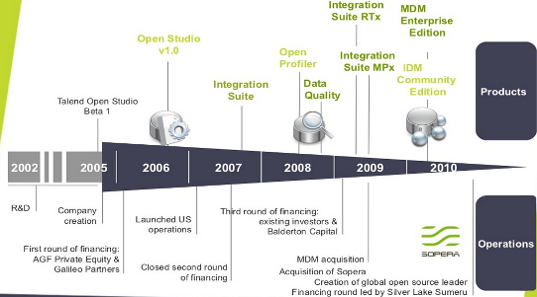
Talend - nhà lãnh đạo toàn cầu trong hội nhập mã nguồn mở, cung cấp giải pháp tích hợp mà thực sự mở rộng cho bất kỳ loại thách thức tích hợp, bất kỳ khối lượng dữ liệu và bất kỳ phạm vi dự án dù đơn giản hay phức tạp. Chỉ có số liệu đánh giá cao khả năng mở rộng của Talend, ứng dụng và quy trình kinh 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 .

* 1. **Lịch sự ra dời 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.**

**1.3. Các sản phẩm của Talend.**

Talend hiện có 3 dòng sản phẩm:

1. **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.

1. **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.

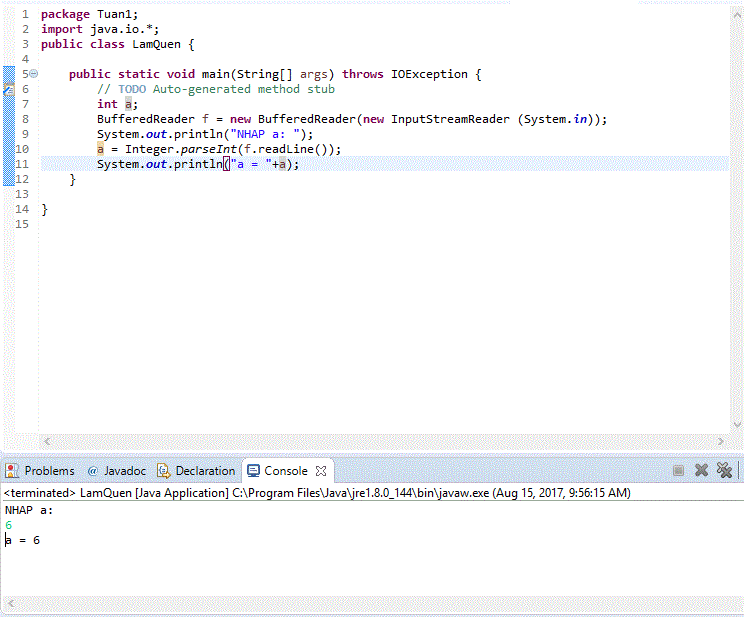
1. **Dòng sản phẩm tích hợp ứng dụng dựa trên nền điện toán đám mây.**

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.

**Tìm hiểu về java:**

Ví dụ do giảng viên cung cấp:

Tập làm quen với phần nhập xuất cơ bản của java thông qua ví dụ



Công việc của các thành viên trong nhóm:

* Tìm tài liệu về Talend: Nguyễn Công Thịnh
* Sắp xếp và chỉnh sửa báo cáo thực hành: Võ Đông Lâm

\***Tài liệu tham khảo** : 1.“*Đồ án tin học ứng dụng tìm hiểu talend*”

2.”*UG\_EN\_DI\_TOSDI*”

***Tiến độ buổi thực hành 1(15/08/217) : đảm bảo đúng tiến độ thực hành. Thực hiện đầy đủ các nội dung công việc được đề ra.***