**TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP THÀNH PHỐ HỒ CHÍ MINH**

**KHOA CÔNG NGHỆ THÔNG TIN**

**🙠🕮🙢**

****

***Môn học: Phát triển ứng dụng (có đồ án)***

***Topic 2:***

***IMPORT DATA BY TALEND***

**GVHD lý thuyết: Huỳnh Thái Học**

**GVHD thực hành: Đỗ Hà Phương**

**Lớp: DHCNTT11A**

**Địa chỉ GITHUB: github.com/nhom3talendPTUD/doanPTUD**

**Các thành viên trong nhóm: MSSV:**

**Nguyễn Thị Lan Anh 15032861**

**Lưu Hiền Phương Quyên 15018831**

**Mục lục**

[**I. Tổng quan:** 3](#_Toc494734415)

[**1. Tên đề tài:** 3](#_Toc494734416)

[**2. Nhiệm vụ:** 3](#_Toc494734417)

[**3. Tiến độ thực hiện:** 3](#_Toc494734418)

[**II. Nội dung:** 4](#_Toc494734419)

[**1.Talend là gì?** 4](#_Toc494734420)

[**2. Chi tiết: Xây dựng chức năng Import data qua Talend** 5](#_Toc494734421)

[2.1Tạo bảng dữ liệu csv 5](#_Toc494734422)

[2.2 Mở Talend, tạo một công việc mới: chọn job designs -> create job 5](#_Toc494734423)

[2.3Đưa file csv vào Talend: File delimited -> create file delimited. 5](#_Toc494734424)

[2.4Tạo database liên kết: Db connections -> create conncections 6](#_Toc494734425)

[2.5Dùng các công cụ trong Talend để xuất dữ liệu từ file csv sang mysql: 7](#_Toc494734426)

[2.6Chạy chương trình – kết quả: 9](#_Toc494734427)

[**III. Kết luận:** 11](#_Toc494734428)

[**Tài liệu tham khảo:** 11](#_Toc494734429)

# **I. Tổng quan:**

## **1. Tên đề tài:**

Import data by talend

## **2. Nhiệm vụ:**

Import/Export dữ liệu bằng Talend.

## **3. Tiến độ thực hiện:**

|  |  |  |  |
| --- | --- | --- | --- |
| Tuần | Công việc | Thành viên | Tiến độ công việc |
| 1  (15/8-21/8) | - Tìm hiểu Talend.  - Chuẩn bị tài liệu liên quan.  - Thảo luận, đọc tài liệu.  - Lập kế hoạch 10 tuần.  - Xác định mục tiêu công việc | - Lan Anh, Phương Quyên  - Phương Quyên  - Lan Anh, Phương Quyên  - Lan Anh, Phương Quyên  - Lan Anh | Hoàn thành |
| 2  (22/8-28/8) | - Cài đặt Talend.  - Tìm hiểu java cơ bản: Làm một số bài tập java cơ bản | - Phương Quyên  - Lan Anh, Phương Quyên | Hoàn thành |
| 3  (30/8-4/9) | - Tìm hiểu file \*.csv  - Demo đọc file csv trên Talend | - Lan Anh  - Phương Quyên | Hoàn thành |
| 4  (6/9-11/9) | - Demo: Nhận dữ liệu từ file csv xuất sang mysql.  - Làm thêm: Nhận dữ liệu từ Excel xuất sang mysql. | - Phương Quyên  - Lan Anh | Hoàn thành |
| 5  (13/9-18/9) | Insert dữ liệu | Phương Quyên | Hoàn thành |
| 6  (20/9-25/9) | Update dữ liệu | Lan Anh | Hoàn thành |
| 7  (27/9-2/10) | Update dữ liệu (tiếp) | Lan Anh |  |
| 8 | Xuất dữ liệu từ database sang file csv theo lựa chọn | Phương Quyên | Hoàn thành |
| 9 | Export talend | Phương Quyên |  |

# **II. Nội dung:**

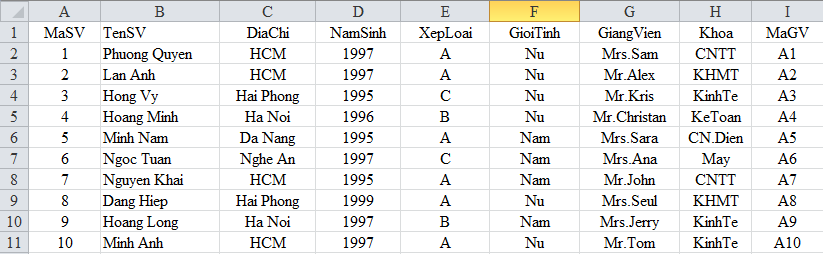
## **1.Talend là gì?**

- 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 dự 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.

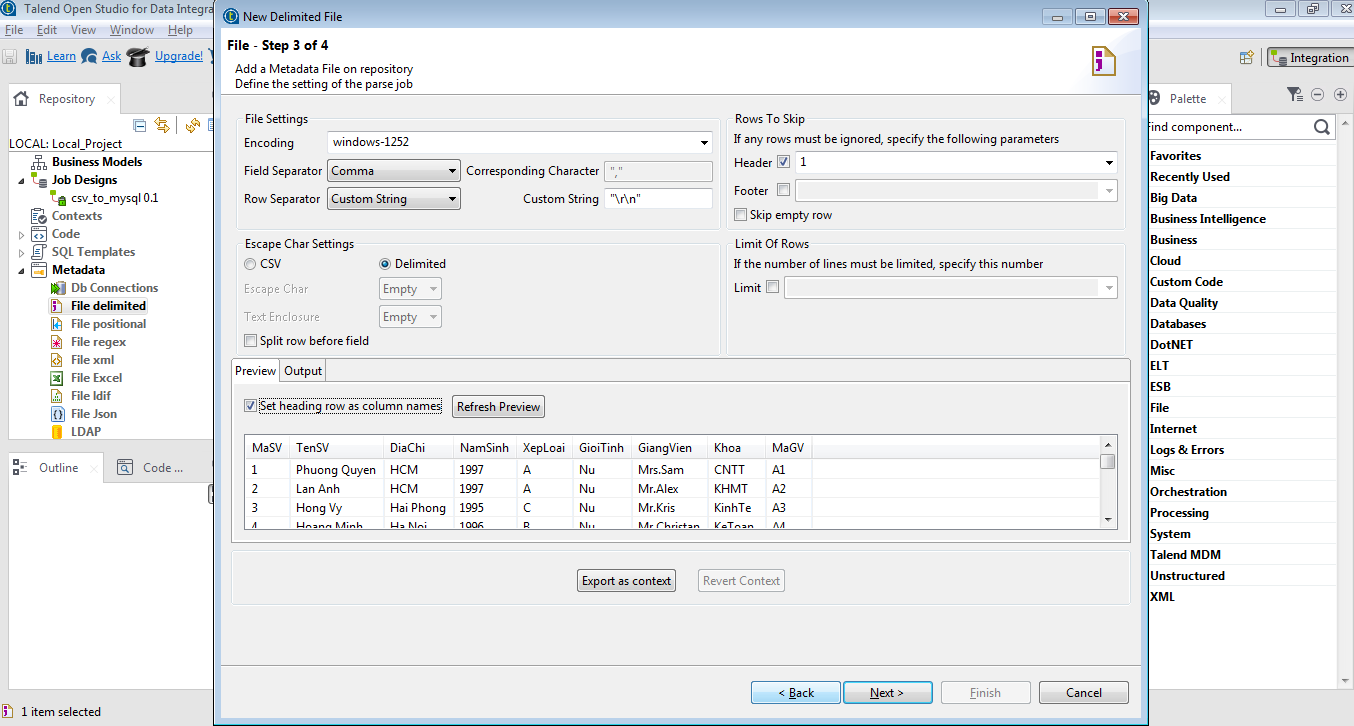
**2. Chi tiết: Xây dựng chức năng Import data qua Talend**

### 2.1. Tạo bảng dữ liệu csv

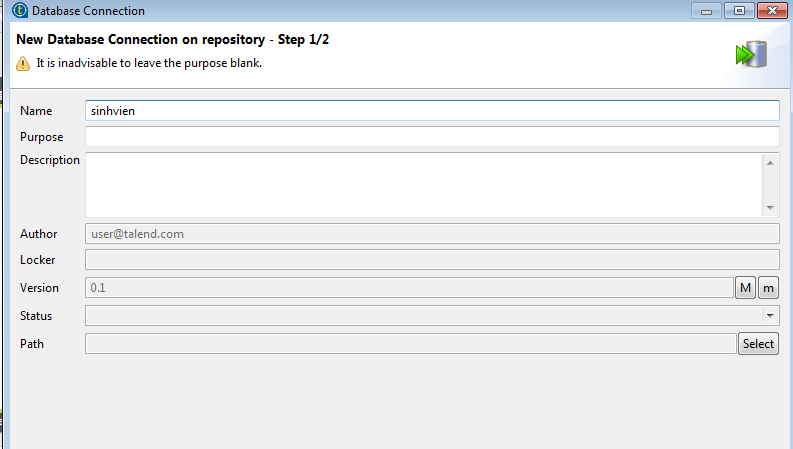


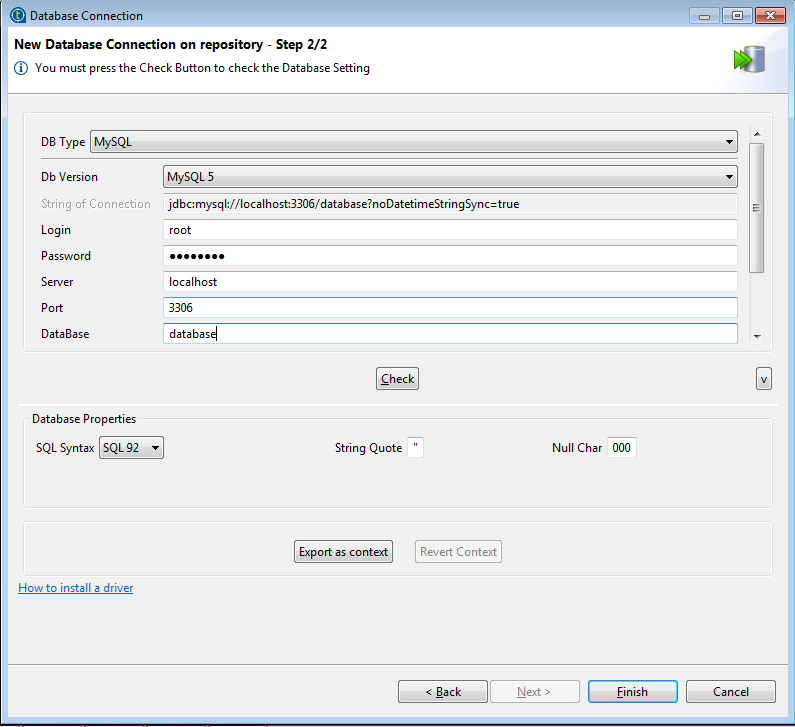
### 2.2. Mở Talend, tạo một công việc mới: chọn job designs -> create job

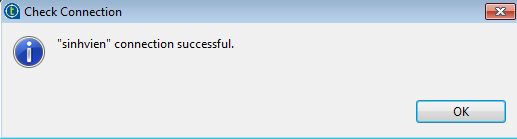
### 2.3. Đưa file csv vào Talend: File delimited -> create file delimited.



2.4Tạo database liên kết: Db connections -> create conncections





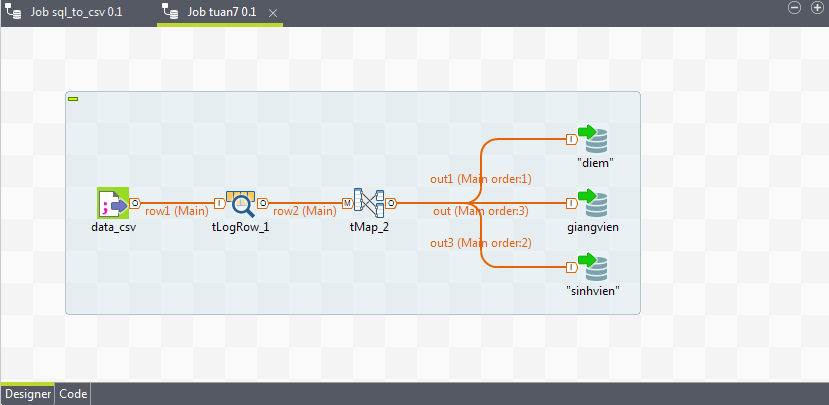


Đặt tên, tạo các kết nối đến Mysql. Hiển thị bảng như trên là việc tạo liên kết đến Mysql đã thành công. (Tạo các db giangvien, diem tương tự).

2.5.Dùng các công cụ trong Talend để xuất dữ liệu từ file csv sang mysql:

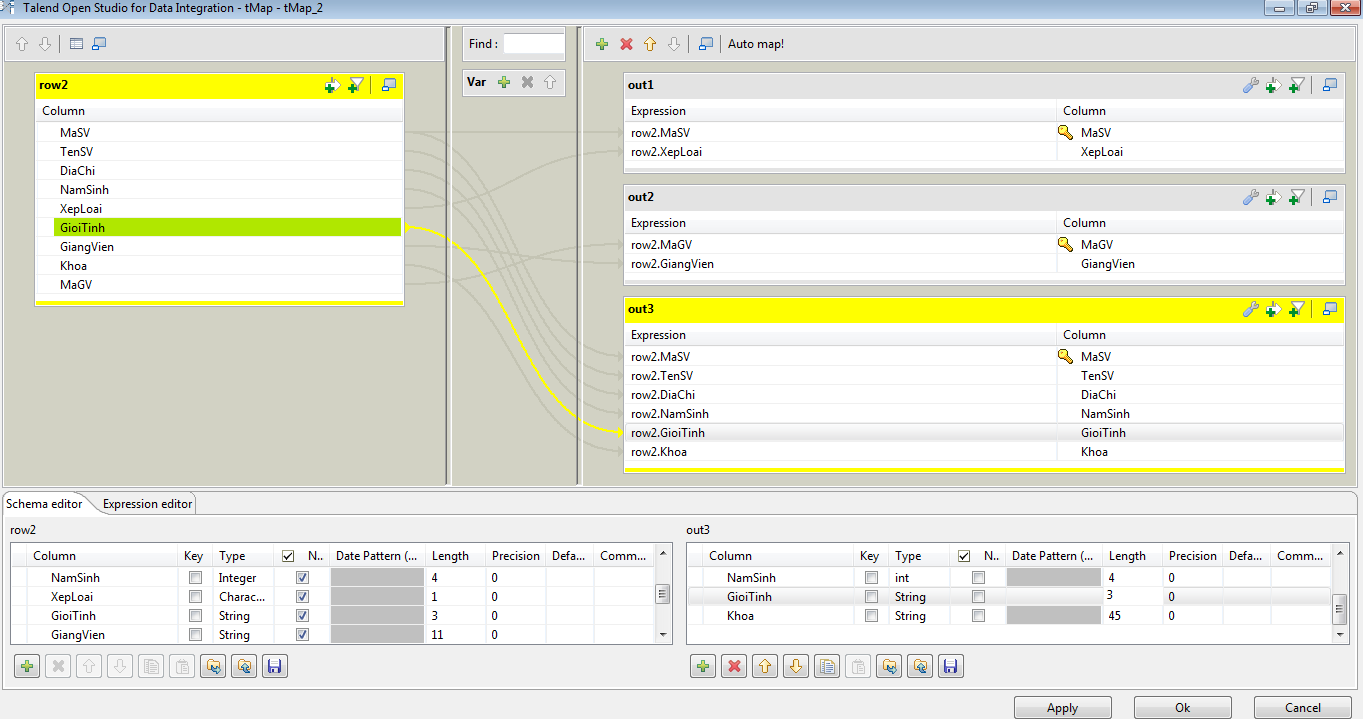
Khi kéo dữ liệu csv đã tạo trong Talend vào khung làm việc để thực hiện import, chọn tFileDelimited.

Kéo các kết nối database, chọn tMysqlOutput.



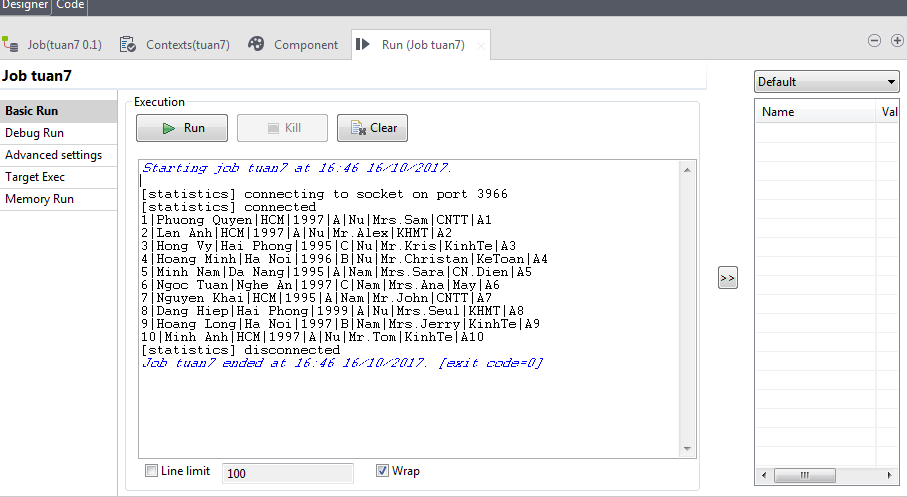
- tLogRow: hiển thị dữ liệu trên talend.

- Dùng tMap định dạng các thuộc tính. Chia các dữ liệu từ file csv vào các bảng theo mong muốn.



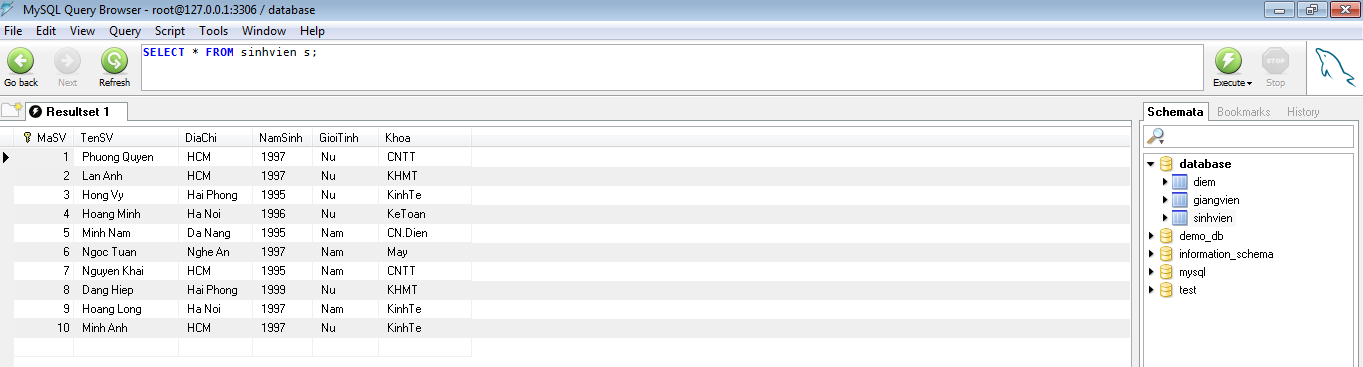
### 2.6. Chạy chương trình – kết quả:

Dữ liệu từ csv hiển thị trên Talend:

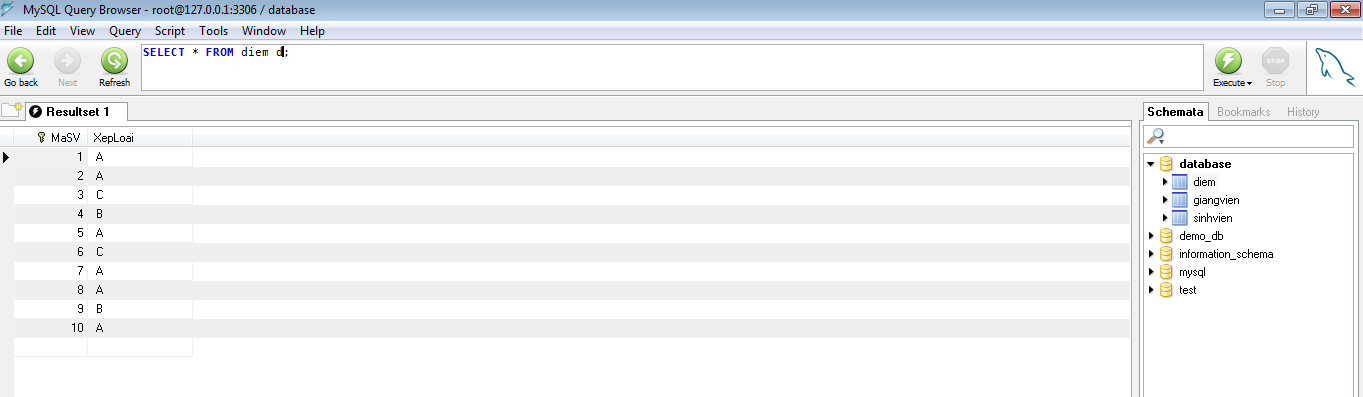


Sau khi chạy chương trình, Mysql nhận dữ liệu từ file csv. Hiển thị dữ liệu trên Mysql Query Browser:

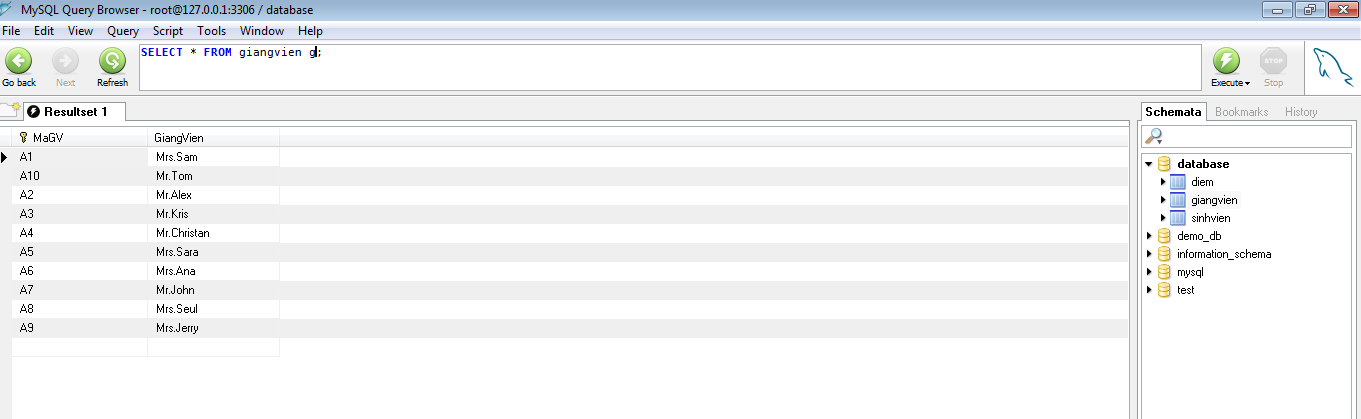
Bảng sinhvien:



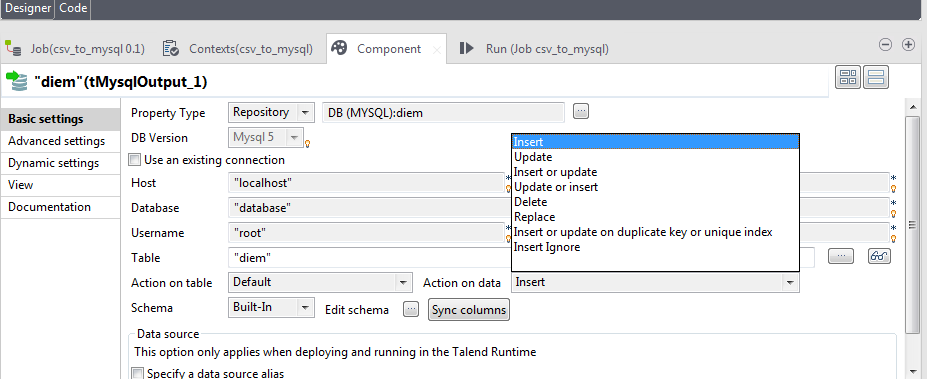
Bảng diem:



Bảng giangvien:

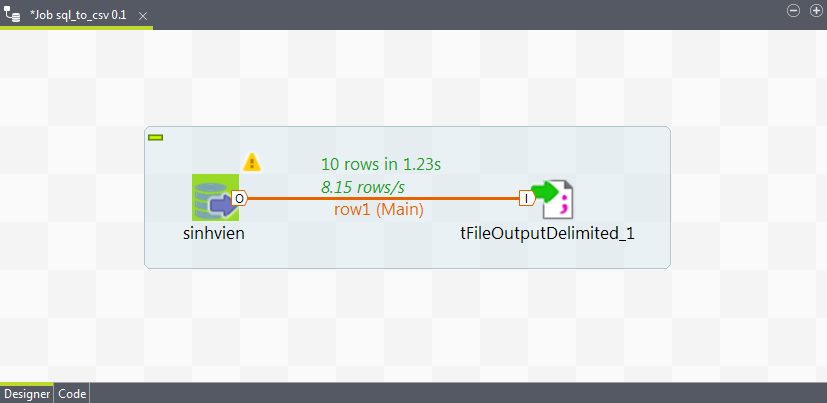


Chọn các chức năng Insert/Update/Replace/Delete chỉnh sửa dữ liệu (hình dưới):



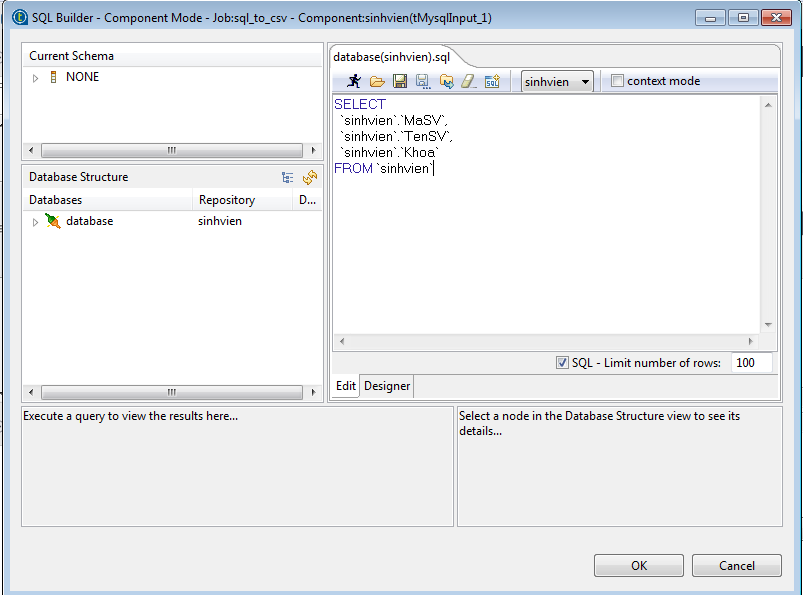
Khi chạy chương trình lần thứ 2, dữ liệu đổ vào sẽ trùng (lỗi) khóa với dữ liệu đã chạy ở lần 1, chọn chức năng Insert or update in duplicate key or unique index để không gặp lỗi về khóa.

2.7 Xuất dữ liệu chọn từ database qua file csv:



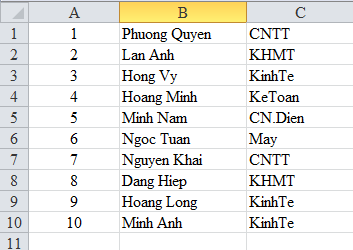
(tFileOutputDelimite: tạo 1 file csv tự động)

Ví dụ xuất dữ liệu từ bảng sinh viên, nháy chuột vào sinhvien:



Xuất dữ liệu các cột MaSV, TenSV, Khoa từ bảng sinhvien, viết câu lệnh sql (SELECT...FROM...)

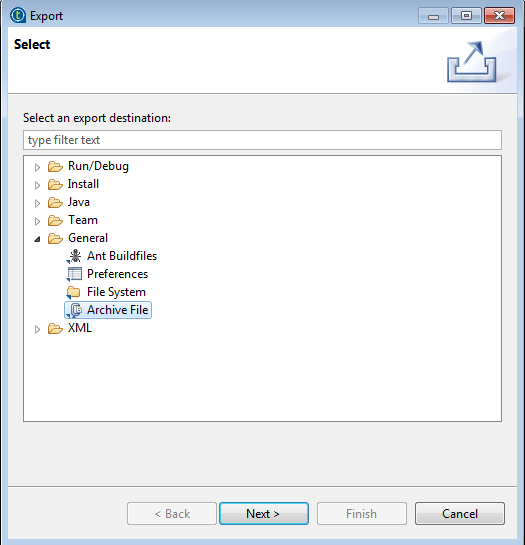
Kết quả:

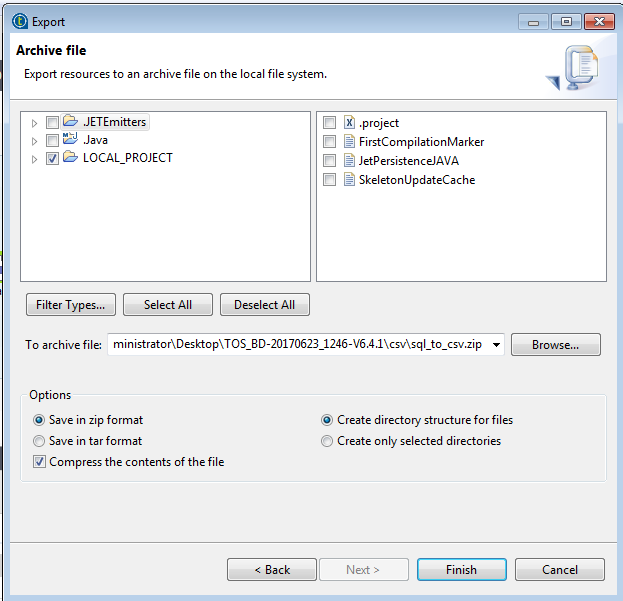


2.8 Export data:

Sau khi hoàn thành job, export job đã làm được 1 file nén .zip

Chọn file -> export.





Nhận được 1 file nén (.zip).

# **III. Kết luận:**

# **Tài liệu tham khảo:**

<https://www.youtube.com/watch?v=CXznw2YBe2I>

<https://www.youtube.com/watch?v=NIAelgGbMU4&t=205s>

<https://www.talend.com>

[http://123doc.org/document/3363726-do-an-tin-hoc-ung-dung-tim-hieu-talend.htm#](http://123doc.org/document/3363726-do-an-tin-hoc-ung-dung-tim-hieu-talend.htm)

Chú thích:

(1): tFileDelimitedInput: đọc dữ liệu cho trước.

(2): tMyslOutput: viết, cập nhật, thay đổi hoặn ngăn chặn các mục trong một cơ sở dữ liệu. tMysqlOutput thực hiện các hành động được xác định trên bảng và / hoặc trên các dữ liệu chứa trong bảng, dựa trên dòng chảy đến từ các thành phần trước trong Job.