**BỘ GIÁO DỤC VÀ ĐÀO TẠO**

**TRƯỜNG ĐẠI HỌC NÔNG LÂM THÀNH PHỐ HỒ CHÍ MINH**

**KHOA CÔNG NGHỆ THÔNG TIN**

**🙡-------🙣**



**BÁO CÁO ĐỒ ÁN**

**Môn: Data warehouse**

**ĐỀ TÀI:**  **TỶ GIÁ NGOẠI TỆ**

**Giảng viên hướng dẫn:** Nguyễn Đức Công Song

**Sinh viên thực hiện:**

20130353 - Nguyễn Thị Phương Nhi

20130340 - Nguyễn Đình Nguyên

20130352 - Trương Cao Thành Nhật

|  |
| --- |
| **Tp Hồ Chí Minh, ngày 17 tháng 12 năm 2023** |

# Lời cảm ơn

Đầu tiên nhóm xin gửi lời cảm ơn đến toàn thể quý Thầy/Cô trường Đại học Nông Lâm thành phố Hồ Chí Minh đã giúp cho nhóm có các kiến thức nền tảng để hoàn thành đề tài này.

Đặc biệt cảm ơn thầy Nguyễn Đức Công Song đã truyền đạt kiến thức cũng như góp ý trong quá trình thực hiện, giúp nhóm hoàn thành đồ án.

Trong thời gian thực hiện đề tài nhóm đã vận dụng các kiến thức được hướng dẫn cùng với các nghiên cứu kiến thức mới. Tuy nhiên trong thời gian làm đồ án ắt hẳn có những sai sót. Chính vì vậy, chúng em mong muốn nhận được sự đóng góp từ phía Thầy nhằm hoàn thiện các kiến thức đã học và biết áp dụng chúng trong tương lai.

# Thông tin liên hệ

|  |  |  |
| --- | --- | --- |
| **Họ và tên** | **Mã số sinh viên** | **Email liên hệ** |
| Trương Cao Thành Nhật | 20130352 | 20130352@st.hcmuaf.edu.vn |
| Nguyễn Thị Phương Nhi | 20130353 | 20130353@st.hcmuaf.edu.vn |
| Nguyễn Đình Nguyên | 20130340 | 20130340@st.hcmuaf.edu.vn |

# Mục lục

[Lời cảm ơn 0](#_Toc153733289)

[Thông tin liên hệ 1](#_Toc153733290)

[Mục lục 2](#_Toc153733291)

[1. Tổng quan: 4](#_Toc153733292)

[1.1 Mô tả đề tài: 4](#_Toc153733293)

[1.2 Quy trình hệ thống ETL 4](#_Toc153733294)

[1.3 Giới thiệu nguồn dữ liệu: 4](#_Toc153733295)

[1.4 Mô tả nguồn dữ liệu 4](#_Toc153733296)

[2. Kế hoạch nhóm: 5](#_Toc153733297)

[3. Kiến trúc hệ thống: 6](#_Toc153733298)

[4. Cơ sở dữ liệu: 7](#_Toc153733299)

[4.1 Database control 7](#_Toc153733300)

[4.2 Database staging 8](#_Toc153733301)

[4.3 Database warehouse 9](#_Toc153733302)

[4.4 Database mart 10](#_Toc153733303)

[5. Đặc tả dữ liệu & Dữ liệu ví dụ: 12](#_Toc153733304)

[5.1 Database Control 12](#_Toc153733305)

[5.1.1 Table data\_file 12](#_Toc153733306)

[5.1.2 Table data\_file\_configs 12](#_Toc153733307)

[5.2 Database Staging 13](#_Toc153733308)

[5.3 Database Warehouse 14](#_Toc153733309)

[5.3.1 Table Date\_dim 14](#_Toc153733310)

[5.3.2 Bảng currency dim 15](#_Toc153733311)

[5.3.3 Bảng bank\_dim 15](#_Toc153733312)

[5.3.4 Bảng exchange\_rate fact 16](#_Toc153733313)

[5.3.5 Bảng exchange\_rate\_aggregate 16](#_Toc153733314)

[5.3.6 Bảng avg\_rate\_aggregate 17](#_Toc153733315)

[5.3 Database Mart 18](#_Toc153733316)

[5.3.1 Bảng db\_rename\_exchange 18](#_Toc153733317)

[5.3.2 Bảng exchange\_rate\_aggregate 19](#_Toc153733318)

[5.3.3 Bảng db\_rename\_avg 20](#_Toc153733319)

[5.3.4 Bảng avg\_rate\_aggregate 21](#_Toc153733320)

[6. Workflow diagram 22](#_Toc153733321)

[6.1 Từ source về file 22](#_Toc153733322)

[6.2 Từ file vào database staging 24](#_Toc153733323)

[6.3 Từ database staging vào database warehouse 27](#_Toc153733324)

[6.4 Từ database warehouse vào database mart 28](#_Toc153733325)

[7. Code & liên hệ mô tả 30](#_Toc153733326)

[7.1 Từ source về file 30](#_Toc153733327)

[7.2 Từ file vào database staging 36](#_Toc153733328)

[7.3 Từ database staging vào database warehouse 41](#_Toc153733329)

[7.4 Từ database warehouse vào database mart 44](#_Toc153733330)

[8. Giao diện report demo: 48](#_Toc153733331)

[9. Tài liệu đính kèm: 48](#_Toc153733332)

# Tổng quan:

## 1.1 Mô tả đề tài:

Đề tài của Nhóm 7 là xây dựng kho dữ liệu lớn (data warehouse), quản lý và phân tích thông tin về tỷ giá ngoại tệ trên thị trường quốc tế.

Nhóm sẽ giới thiệu sơ lược về nguồn dữ liệu, mô tả dữ liệu và các quy trình để có thể quản lý và phân tích thông tin về tỷ giá ngoại tệ. Những nội dung cơ bản sẽ được nghiên cứu, phát triển và triển khai như thu thập dữ liệu, thiết kế cấu trúc dữ liệu, đặc tả dữ liệu, phát triển hệ thống phân tích và lập giao diện báo cáo dưới dạng demo. Ở phần phát triển hệ thống phân tích, để hiểu rõ và dễ dàng thực hiện hơn thì hệ thống đã được nhóm chia nhỏ ra thành các module thực hiện độc lập với nhau. Với các module sẽ có phần workflow, code và liên hệ mô tả.

## 1.2 Quy trình hệ thống ETL

Nhóm thực hiện hệ thống data warehouse bằng quy trình ETL vì thế module của nhóm gồm có:

- run.bat: Thu thập dữ liệu từ trang web của 2 ngân hàng (bidv, vietcombank) về thư mục “D:\DW\_2023\_T4\_Nhom7\file\crawl\”, dữ liệu sẽ được ghi vào tệp dưới dạng csv.

- module.jar: Lấy dữ liệu đã được thu thập từ các ngân hàng trong các tệp csv đưa vào database staging thông qua database control bằng phương pháp “flag”.

- warehouse.jar: Lọc dữ liệu sau đó chuyển đổi sang dữ liệu chuẩn cùng 1 định dạng để đưa vào database warehouse có các bảng dim và fact, sau đó thực thi các yêu cầu thông qua tạo các bảng aggregate để hỗ trợ việc sử dụng dữ liệu để báo cáo.

- mart.jar: Đẩy dữ liệu trong bảng aggregate từ database warehouse vào database mart để hiển thị dữ liệu lên trên giao diện người dùng. Ở module này sẽ liên tục cập nhật dữ liệu khi có dữ liệu mới để tránh ảnh hưởng xấu đến trải nghiệm người dùng khi dữ liệu đã thay đổi nhưng giao diện vẫn ở dữ liệu cũ.

## 1.3 Giới thiệu nguồn dữ liệu:

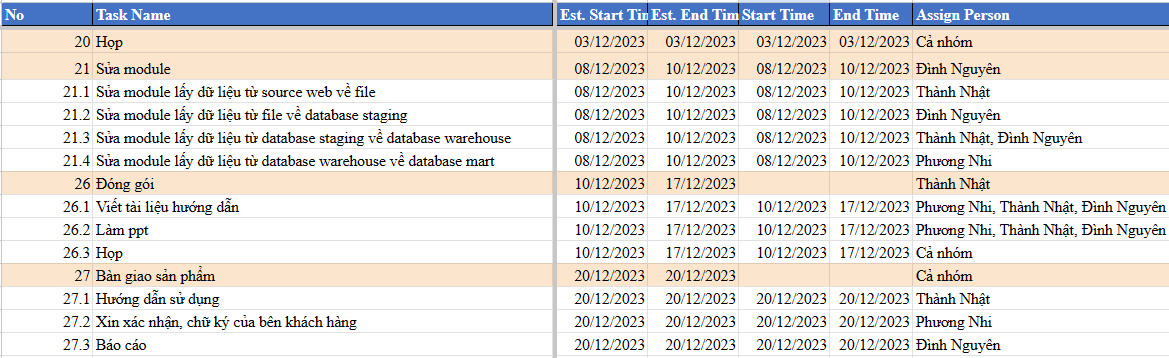
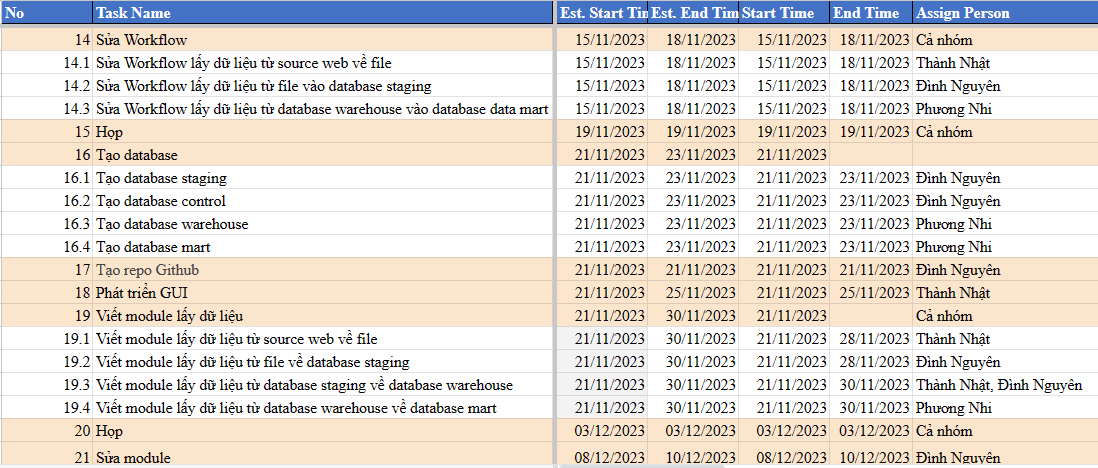
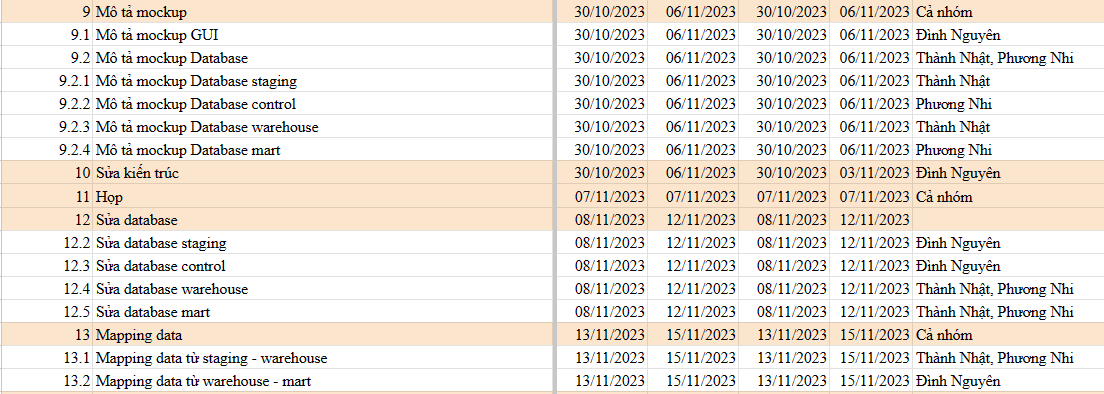
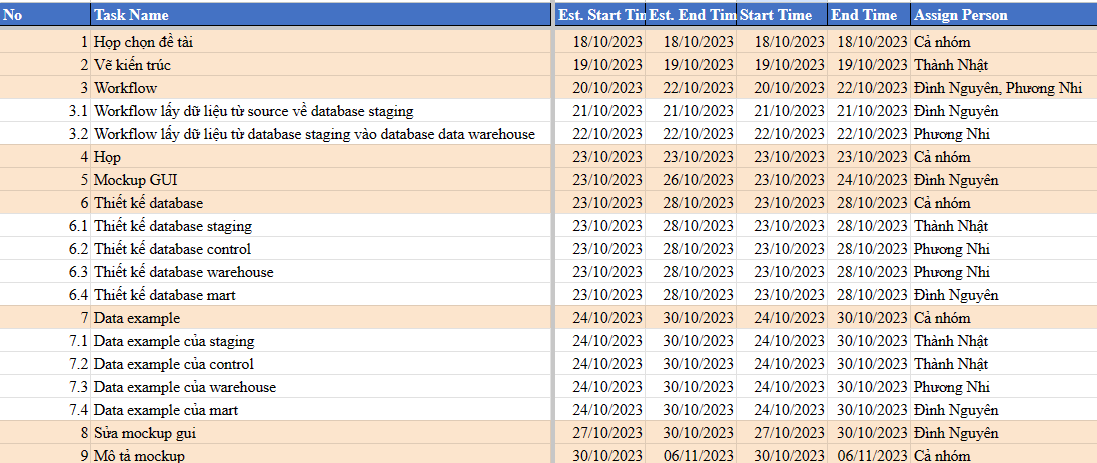
Dữ liệu lấy được từ việc thu thập dữ liệu từ nhiều nguồn website (ngân hàng bidv - [Cá nhân (bidv.com.vn)](https://www.bidv.com.vn/), ngân hàng vietcombank - [Vietcombank | Ngân hàng TMCP Ngoại thương Việt Nam](https://www.vietcombank.com.vn/)), sau đó chuyển đổi dữ liệu để có thể sử dụng cho các mục đích khác nhau như report, mining,...

## 1.4 Mô tả nguồn dữ liệu

- Dữ liệu nguồn của 2 trang bidv và vietcombank gồm 5 cột:

* Ký hiệu ngoại tệ / Mã ngoại tệ
* Tên ngoại tệ
* Mua tiền mặt và Séc / Mua tiền mặt
* Mua chuyển khoản
* Bán

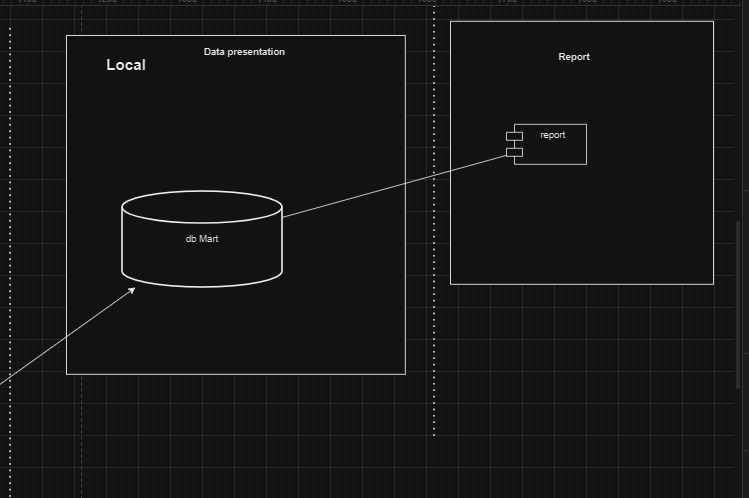
# Kế hoạch nhóm:

* Link kế hoạch chi tiết: [DW\_2023\_T4\_Nhom7](https://docs.google.com/spreadsheets/d/1xL7yAa273moRsYvRbqfqzao11JwoCoKIiFm--SWFksQ/edit?usp=sharing)

# Kiến trúc hệ thống:

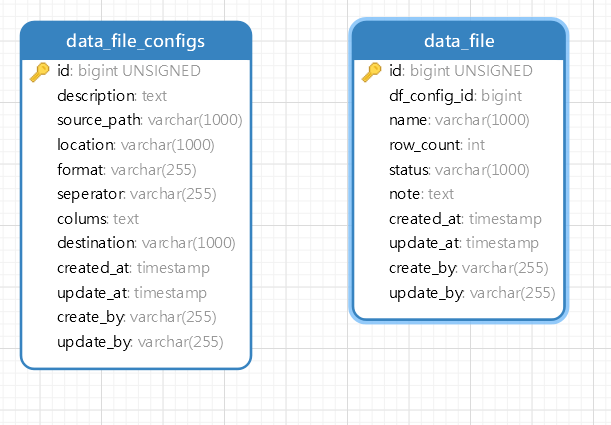
* Link kiến trúc hệ thống chi tiết:[warehouse.drawio - Kiển trúc (diagrams.net)](https://app.diagrams.net/#G1XdLJlheOkOdNm6JnVvdiHn8Rx7xPfkU7)

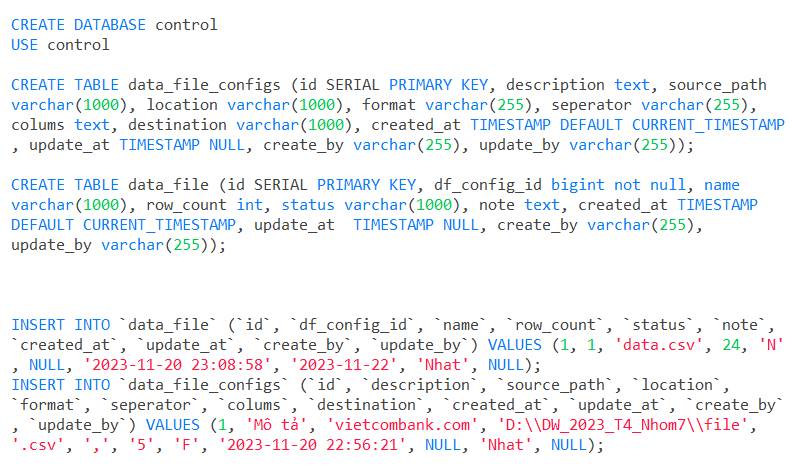




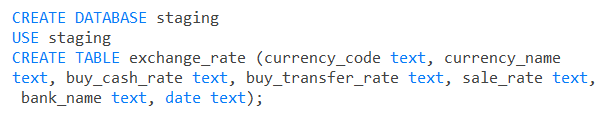
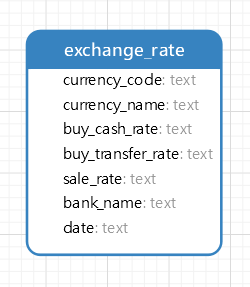
# Cơ sở dữ liệu:

## 4.1 Database control

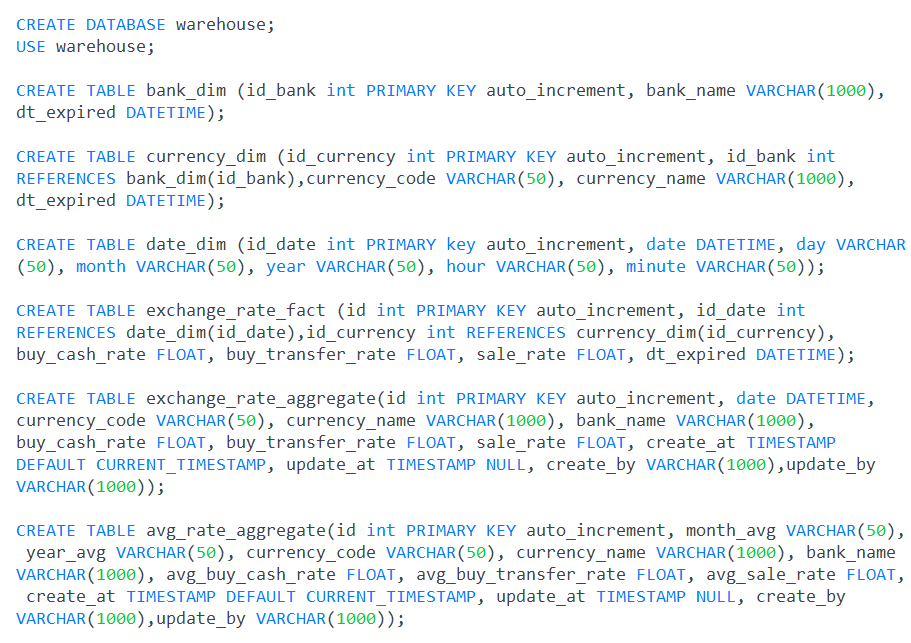
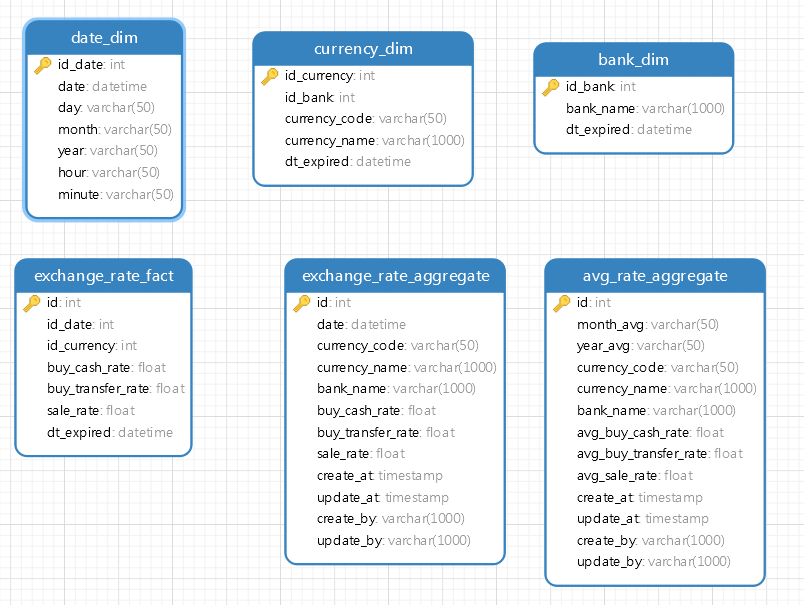




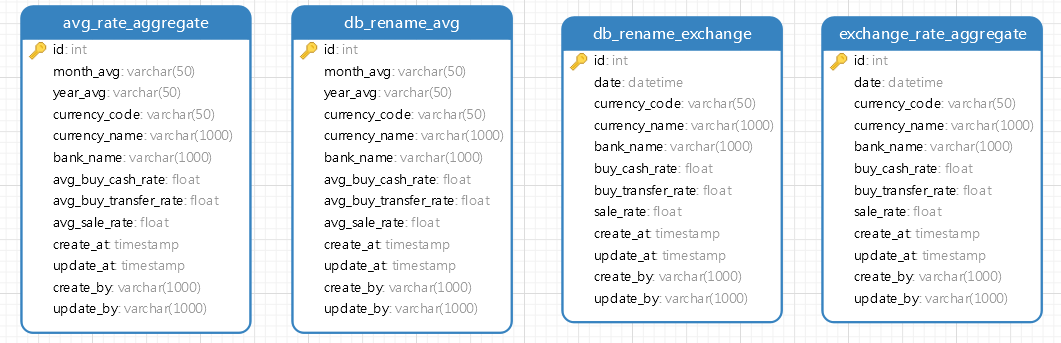
## 4.2 Database staging

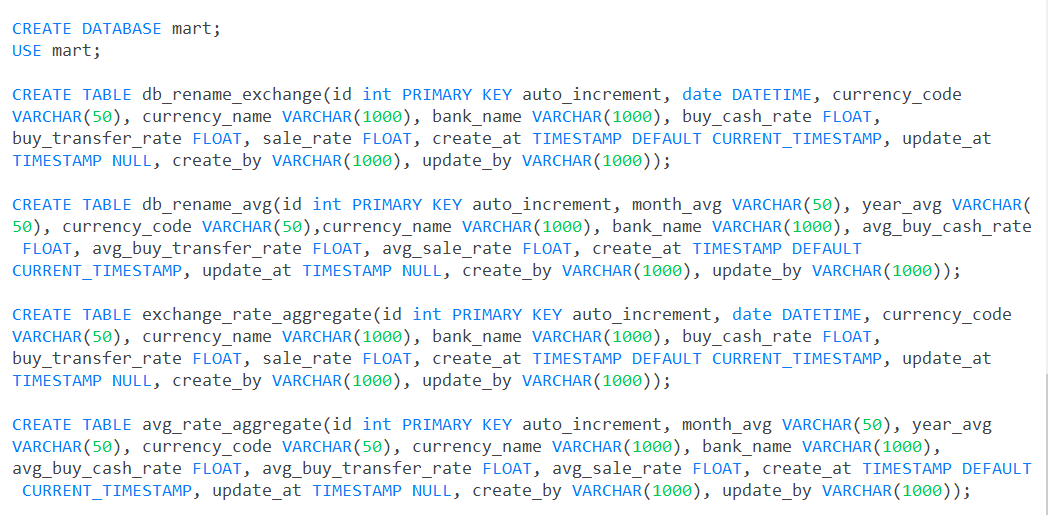


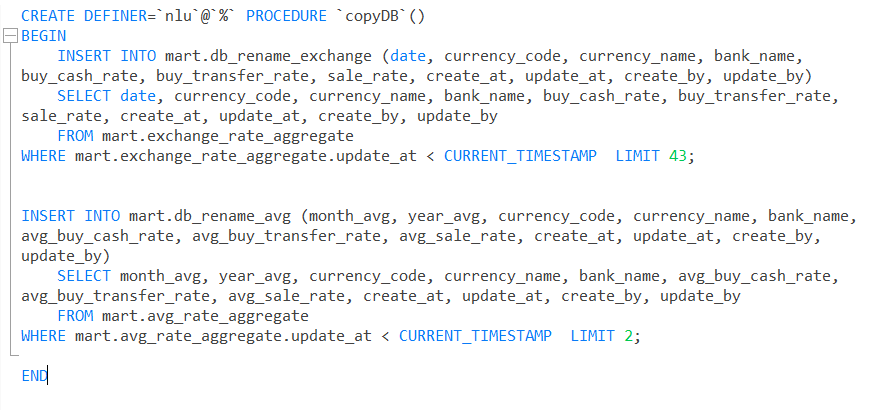
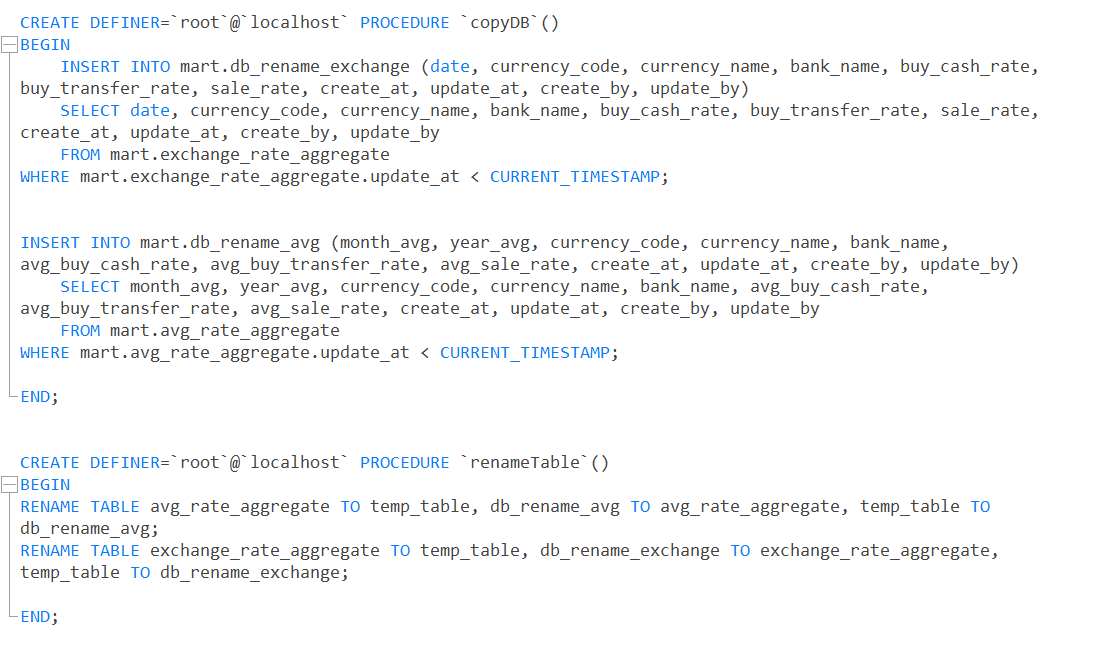
## 4.3 Database warehouse

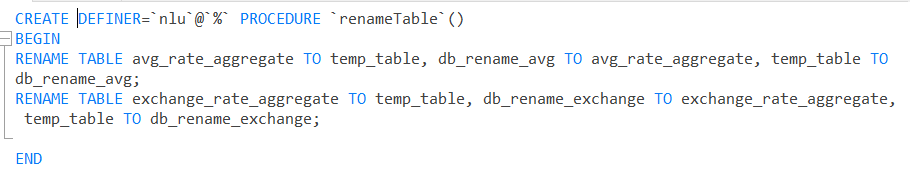


## 4.4 Database mart









Gồm 4 procedure tương ứng 2 cho local và 2 cho server.

# Đặc tả dữ liệu & Dữ liệu ví dụ:

## 5.1 Database Control

### 5.1.1 Table data\_file

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **No** | **Field Name** | **Format** | **Edit Rules** | **Description** | **Example** |
| 1 | id | bigint |  | mã của data file | 1 |
| 2 | df\_config\_id | bigint |  | mã của data\_file\_configs | 1 |
| 3 | name | varchar(1000) |  | Tên của file | vietcombank\_data\_20231210\_221055.csv |
| 4 | row\_count | bigint |  | Số lượng hàng | 24 |
| 5 | status | varchar(1000) |  | Trạng thái file | C |
| 6 | note | varchar(1000) |  | Ghi chú | Data import success |
| 7 | created\_at | timestamp |  | Ngày tạo | 2023-12-10 22:10:45 |
| 8 | updated\_at | timestamp |  | Thời gian cập nhật | 2023-12-10 00:00:00 |
| 9 | created\_by | varchar(1000) |  | Người tạo | Nhật |
| 10 | updated\_by | varchar(1000) |  | Người cập nhật | Nguyên |

### 5.1.2 Table data\_file\_configs

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **No** | **Field Name** | **Format** | **Edit Rules** | **Description** | **Example** |
| 1 | id | bigint |  | Mã data\_file\_confogs | 1 |
| 2 | description | varchar(1000) |  | Mô tả | nguồn lấy dữ liệu vietcombank |
| 3 | source\_path | varchar(1000) |  | Nguồn lấy dữ liệu | vietcombank.com |
| 4 | location | varchar(1000) |  | Vị trí lưu trữ | "D:\DW\_2023\_T4\_Nhom7\file\crawl" |
| 5 | format | varchar(255) |  | Định dạng file | .csv |
| 6 | seperator | varchar(255) |  | Phân cách | , |
| 7 | columns | varchar(255) |  | Số cột | 6 |
| 8 | destination | varchar(1000) |  | Nơi chuyển đến | F |
| 9 | created\_at | timestamp |  | Ngày tạo | 2023-12-14 13:45:04 |
| 10 | updated\_at | timestamp |  | Ngày cập nhật | 2023-12-14 13:45:04 |
| 11 | created\_by | varchar(1000) |  | Người tạo | Nhật |
| 12 | updated\_by | varchar(1000) |  | Người cập nhật | Nguyên |

## 5.2 Database Staging

Table exchange\_rate

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **No** | **Field Name** | **Format** | **Edit Rules** | **Description** | **Example** |
|  | 1 | currency\_code | text |  | Mã ngoại tệ | AUD; USD; EUR |
|  | 2 | currency\_name | text |  | Tên ngoại tệ | US DOLLAR |
|  | 3 | buy\_cash\_rate | text |  | Tỷ giá mua tiền mặt | 23,920.00 |
|  | 4 | buy\_transfer\_rate | text |  | Tỷ giá mua chuyển khoản | 23,920.00 |
|  | 5 | sale\_rate | text |  | Tỷ giá bán | 23,920.00 |
|  | 6 | name\_bank | text |  | Tên ngân hàng | VCB, BIDV |
|  | 7 | date | text |  | Ngày tháng năm | 26/12/2023 |

## 5.3 Database Warehouse

### 5.3.1 Table Date\_dim

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **No.** | **Field Name** | **Format** | **Edit Rules** | **Description** | **Example** |
| 1. | 1 | id\_date | bigint |  | id trong bảng | 1 |
|  | 2 | date | datetime |  | Ngày giờ | 2023-12-10 08:30:00 |
| 2. | 3 | day | varchar(50) |  | Ngày | 10 |
| 3. | 4 | month | varchar(50) |  | Tháng | 12 |
| 4. | 5 | year | varchar(50) |  | Năm | 2023 |
|  | 6 | house | varchar(50) |  | Giờ | 8 |
|  | 7 | minus | varchar(50) |  | Phút | 30 |

### 5.3.2 Bảng currency dim

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **No.** | **Field Name** | **Format** | **Edit Rules** | **Description** | **Example** |
| 1 | id\_currency | bigint |  | Mã id tiền tệ | 1 |
| 2 | id \_bank | bigint |  | Mã ngân hàng | 1 |
| 3 | currency\_code | varchar (50) |  | Mã tiền tệ | USD |
| 4 | currency\_name | varchar(1000) |  | Tên tiền tệ | US DOLLAR |
| 5 | dt\_expired | datetime |  | Ngày hết hạn | 2030-01-01 00:00:00 |

### 5.3.3 Bảng bank\_dim

| **No.** | **Field Name** | **Format** | **Edit Rules** | **Description** | **Example** |
| --- | --- | --- | --- | --- | --- |
| 1 | id\_bank | bigint |  | Mã id của ngân hàng | 1 |
| 2 | bank\_name | varchar(255) |  | Tên ngân hàng | BIDV |
| 3 | dt\_expired | datetime |  | Thời điểm hết hạn | 2030-01-01 00:00:00 |

### 5.3.4 Bảng exchange\_rate fact

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **No.** | **Field Name** | **Format** | **Edit Rules** | **Description** | **Example** |
| 1 | id | bigint |  | Mã id | 1 |
| 2 | id\_date | bigint |  | Mã ngày | 1 |
| 3 | id\_currency | bigint |  | Mã tiền tệ | 2 |
| 4 | buy\_cash\_rate | float |  | Giá mua tiền mặt | 23,920.00 |
| 5 | buy\_transfer\_rate | float |  | Giá mua chuyển khoản | 23,920.00 |
| 6 | sale\_rate | float |  | Tỷ giá bán | 23,920.00 |
| 7 | dt\_expired | datetime |  | Ngày hết hạn | 2023-12-10 17:15:02 |

### 5.3.5 Bảng exchange\_rate\_aggregate

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **No.** | **Field Name** | **Format** | **Edit Rules** | **Description** | **Example** |
| 1 | id | bigint |  |  | 1 |
| 2 | id\_date | bigint |  | Mã ngày | 1 |
| 3 | currency\_code | varchar(50) |  | Mã tiền tệ | USD |
| 4 | currency\_name | varchar(1000) |  | Tên mã tiền tệ | US DOLLAR |
| 5 | bank\_name | varchar(1000) |  | Tên ngân hàng | BIDV |
| 6 | buy\_cash\_rate | float |  | Tỷ giá mua bằng tiền mặt | 23,920.00 |
| 7 | buy\_transfer\_rate | float |  | Tỷ giá mua bằng chuyển khoản | 23,920.00 |
| 8 | sale\_rate | float |  | Tỷ giá bán | 23,920.00 |
| 9 | create\_at | timestamp |  | Thời gian tạo | 2023-12-14 13:45:04 |
| 10 | update\_at | timestamp |  | Thời gian cập nhật | 2023-12-14 13:45:04 |
| 11 | create\_by | vachar(1000) |  | Người tạo | Nhật |
| 11 | update\_by | vachar(1000) |  | Người cập nhật | Nhật |

### 5.3.6 Bảng avg\_rate\_aggregate

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **No.** | **Field Name** | **Format** | **Edit Rules** | **Description** | **Example** |
| 1 | id | bigint |  |  | 1 |
| 2 | month\_avg | int |  | Tháng tính trung bình | 11 |
| 3 | year\_avg | int |  | Năm tính trung bình | 2023 |
| 4 | currency\_code | varchar(50) |  | Mã tiền tệ | USD |
| 5 | currency\_name | varchar(1000) |  | Tên mã tiền tệ | US DOLLAR |
| 6 | bank\_name | varchar(1000) |  | Tên ngân hàng | BIDV |
| 7 | avg\_buy\_cash\_rate | float |  | Trung bình tỷ giá mua bằng tiền mặt (tháng) | 23,920.00 |
| 8 | avg\_buy\_transfer\_rate | float |  | Trung bình tỷ giá mua bằng chuyển khoản (tháng) | 23,920.00 |
| 9 | avg\_sale\_rate | float |  | Trung bình tỷ giá bán (tháng) | 23,920.00 |
| 10 | create\_at | timestamp |  | Thời gian tạo | 2023-12-14 13:45:04 |
| 11 | update\_at | timestamp |  | Thời gian cập nhật | 2023-12-14 13:45:04 |
| 12 | create\_by | vachar(1000) |  | Người tạo | Nhật |
| 13 | update\_by | vachar(1000) |  | Người cập nhật | Nhật |

## 5.3 Database Mart

### 5.3.1 Bảng db\_rename\_exchange

| **No.** | **Field Name** | **Format** | **Edit Rules** | **Description** | **Example** |
| --- | --- | --- | --- | --- | --- |
| 1 | id | bigint |  | Mã id | 1 |
| 2 | date | datetime |  | ngày tháng năm, thời gian | 2023-12-14 13:45:04 |
| 3 | currency\_code | varchar(50) |  | Mã tiền tệ | USD |
| 4 | currency\_name | varchar(1000) |  | Tên mã tiền tệ | US DOLLAR |
| 5 | bank\_name | varchar(50) |  | Tên ngân hàng | BIDV |
| 6 | buy\_cash\_rate | float |  | Tỷ giá mua bằng tiền mặt | 23,920.00 |
| 7 | buy\_transfer\_rate | float |  | Tỷ giá mua bằng chuyển khoản | 23,920.00 |
| 8 | sale\_rate | float |  | Tỷ giá bán | 23,920.00 |
| 9 | create\_at | timestamp |  | Người tạo | 2023-12-14 13:45:04 |
| 10 | update\_at | timestamp |  | Người cập nhật | 2023-12-14 13:45:04 |
| 11 | create\_by | vachar(1000) |  | Người tạo | Nhi |
| 12 | update\_by | vachar(1000) |  | Thời gian cập nhật | Nhi |

### 5.3.2 Bảng exchange\_rate\_aggregate

| **No.** | **Field Name** | **Format** | **Edit Rules** | **Description** | **Example** |
| --- | --- | --- | --- | --- | --- |
| 1 | id | bigint |  |  | 1 |
| 2 | date | datetime |  | ngày tháng năm, thời gian |  |
| 3 | currency\_code | varchar(50) |  | Mã tiền tệ | USD |
| 4 | currency\_name | varchar(1000) |  | Tên mã tiền tệ | US DOLLAR |
| 5 | bank\_name | varchar(50) |  | Tên ngân hàng | BIDV |
| 6 | buy\_cash\_rate | float |  | Tỷ giá mua bằng tiền mặt | 23,920.00 |
| 7 | buy\_transfer\_rate | float |  | Tỷ giá mua bằng chuyển khoản | 23,920.00 |
| 8 | sale\_rate | float |  | Tỷ giá bán | 23,920.00 |
| 9 | create\_at | timestamp |  | Người tạo | 2023-12-14 13:45:04 |
| 10 | update\_at | timestamp |  | Người cập nhật | 2023-12-14 13:45:04 |
| 11 | create\_by | vachar(1000) |  | Người tạo | Nhi |
| 12 | update\_by | vachar(1000) |  | Thời gian cập nhật | Nhi |

### 5.3.3 Bảng db\_rename\_avg

| **No.** | **Field Name** | **Format** | **Edit Rules** | **Description** | **Example** |
| --- | --- | --- | --- | --- | --- |
| 1 | id | bigint |  |  | 1 |
| 2 | month\_avg | varchar(50) |  | Tháng tính trung bình | 11 |
| 3 | year\_avg | varchar(50) |  | năm tính trung bình | 2023 |
| 4 | currency\_code | varchar(50) |  | Mã tiền tệ | USD |
| 5 | currency\_name | varchar(1000) |  | Tên mã tiền tệ | US DOLLAR |
| 6 | bank\_name | varchar(50) |  | Tên ngân hàng | BIDV |
| 7 | avg\_buy\_cash\_rate | float |  | Trung bình tỷ giá mua bằng tiền mặt (tháng) | 23,920.00 |
| 8 | avg\_buy\_transfer\_rate | float |  | Trung bình tỷ giá mua bằng chuyển khoản (tháng) | 23,920.00 |
| 9 | avg\_sale\_rate | float |  | Trung bình tỷ giá bán (tháng) | 23,920.00 |
| 10 | create\_at | timestamp |  | Thời gian tạo | 2023-12-14 13:45:04 |
| 11 | update\_at | timestamp |  | Thời gian cập nhật | 2023-12-14 13:45:04 |
| 12 | create\_by | vachar(1000) |  | Người tạo | Nhi |
| 13 | update\_by | vachar(1000) |  | Người cập nhật | Nhi |

### 5.3.4 Bảng avg\_rate\_aggregate

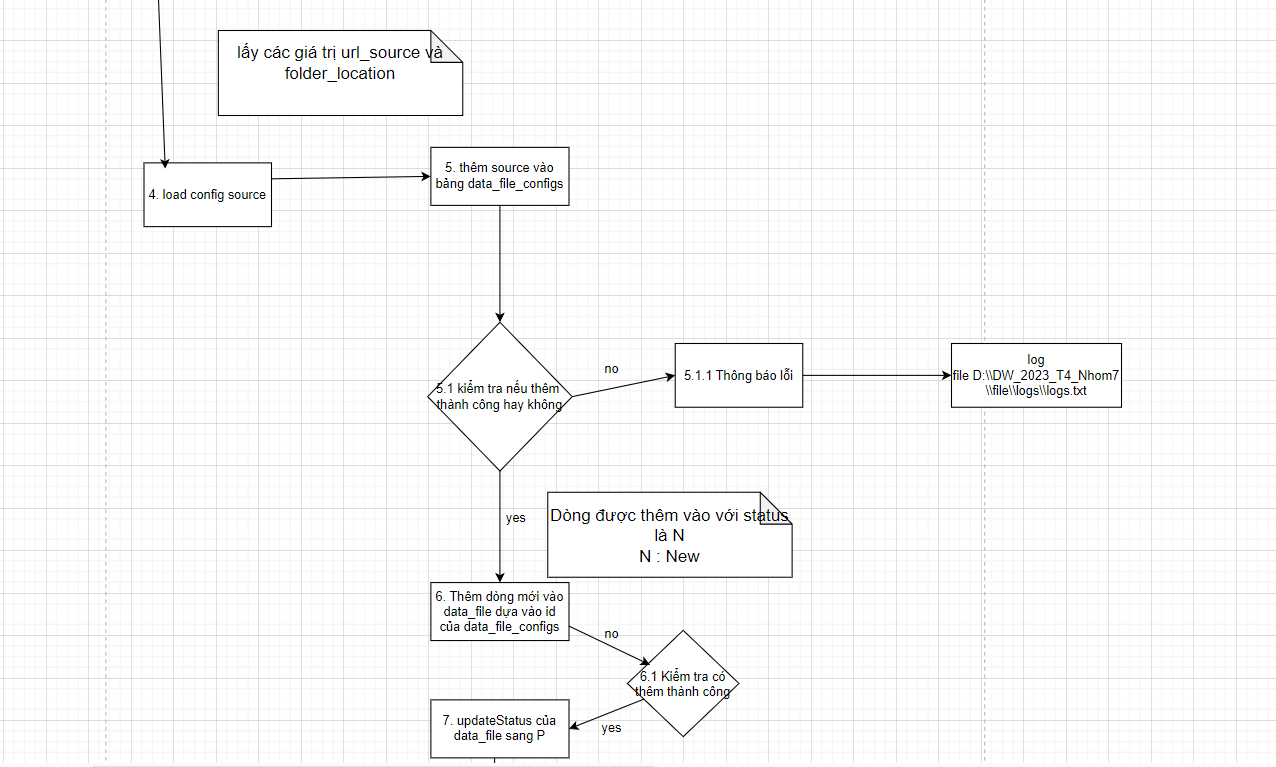
| No. | **Field Name** | **Format** | **Edit Rules** | **Description** | **Example** |
| --- | --- | --- | --- | --- | --- |
| 1 | id | bigint |  |  | 1 |
| 2 | month\_avg | varchar(50) |  | Tháng tính trung bình | 11 |
| 3 | year\_avg | varchar(50) |  | năm tính trung bình | 2023 |
| 4 | currency\_code | varchar(50) |  | Mã tiền tệ | USD |
| 5 | currency\_name | varchar(1000) |  | Tên mã tiền tệ | US DOLLAR |
| 6 | bank\_name | varchar(50) |  | Tên ngân hàng | VCB |
| 7 | avg\_buy\_cash\_rate | float |  | Trung bình tỷ giá mua bằng tiền mặt (tháng) | 23,920.00 |
| 8 | avg\_buy\_transfer\_rate | float |  | Trung bình tỷ giá mua bằng chuyển khoản (tháng) | 23,920.00 |
| 9 | avg\_sale\_rate | float |  | Trung bình tỷ giá bán (tháng) | 23,920.00 |
| 10 | create\_at | timestamp |  | Thời gian tạo | 2023-12-14 13:45:04 |
| 11 | update\_at | timestamp |  | Thời gian cập nhật | 2023-12-14 13:45:04 |
| 12 | create\_by | vachar(1000) |  | Người tạo | Nhi |
| 13 | update\_by | vachar(1000) |  | Người cập nhật | Nhi |

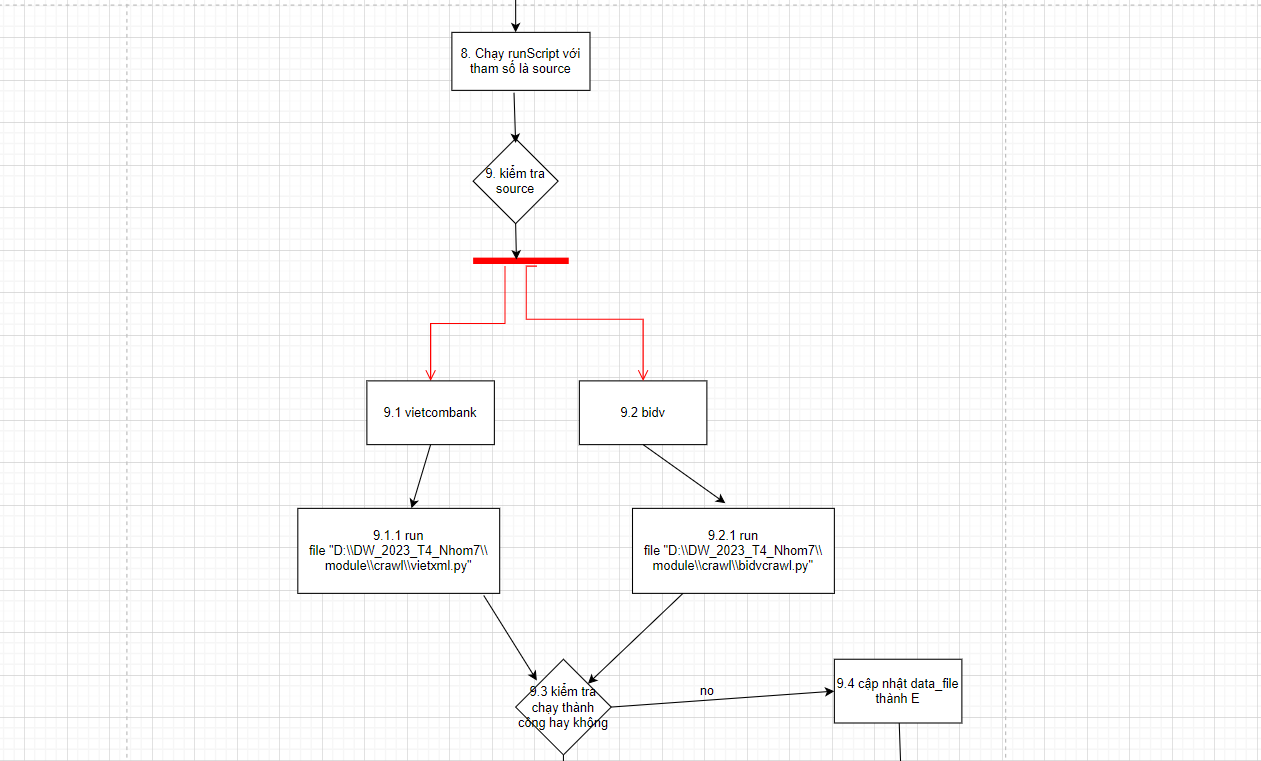
# Workflow diagram

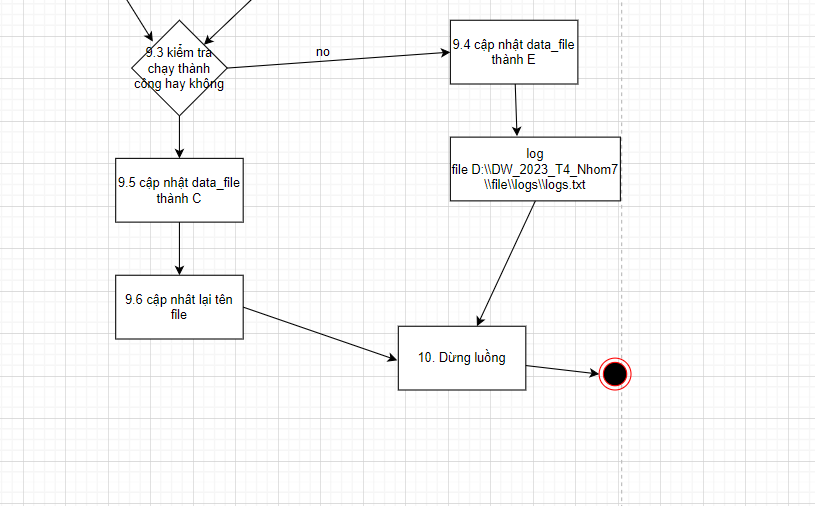
## 6.1 Từ source về file

Link : [https://app.diagrams.net/#G1XdLJlheOkOdNm6JnVvdiHn8Rx7xPfkU7#%7B%22pageId%22%3A%22nMe68txu2dav\_vV5KSd4%22%7D](https://app.diagrams.net/#G1XdLJlheOkOdNm6JnVvdiHn8Rx7xPfkU7)



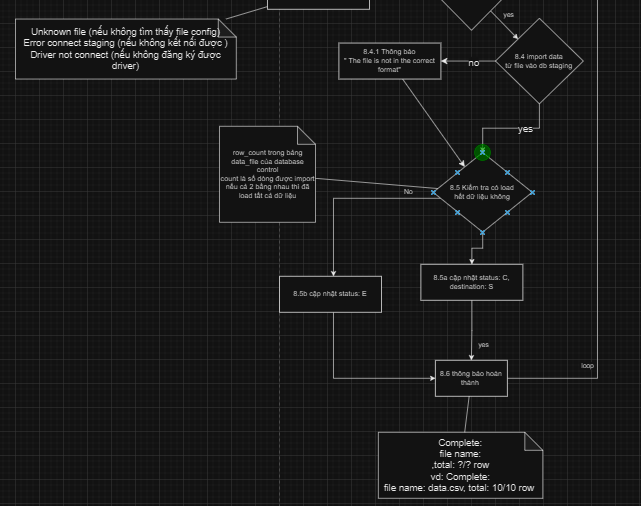
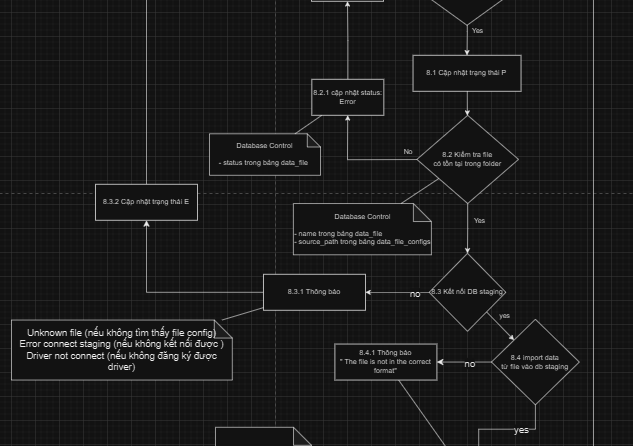
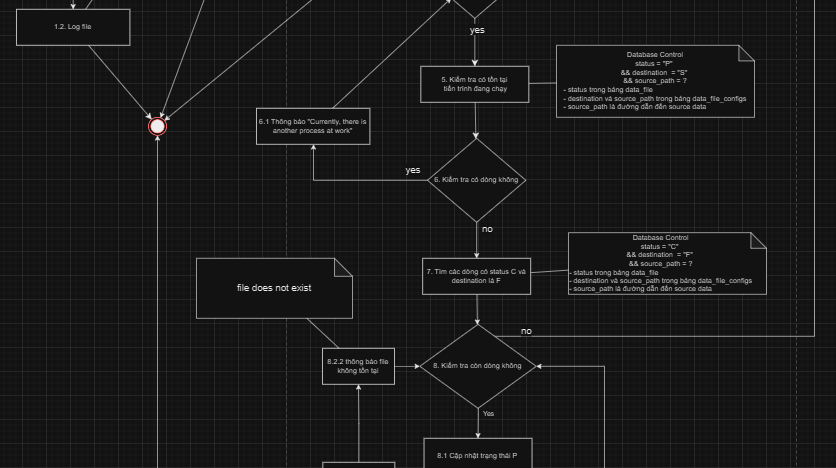
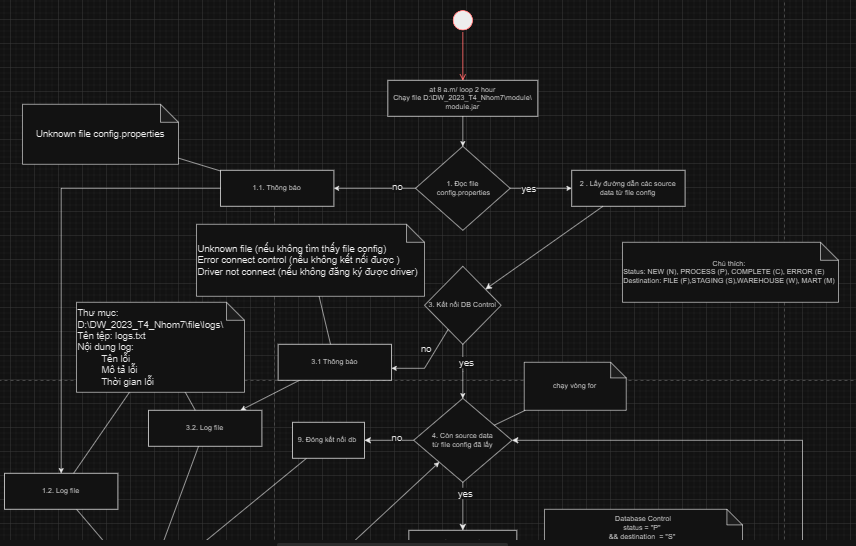






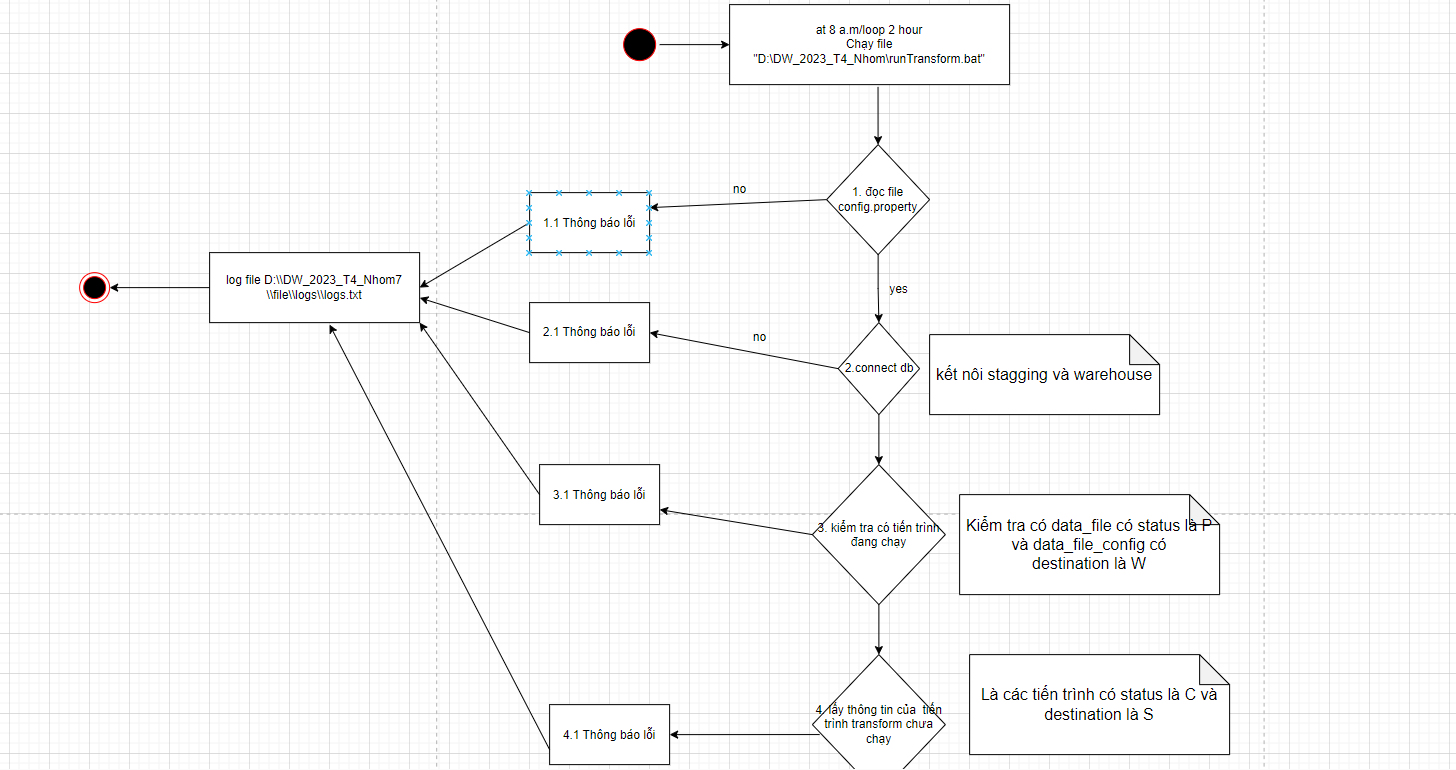
## 6.2 Từ file vào database staging

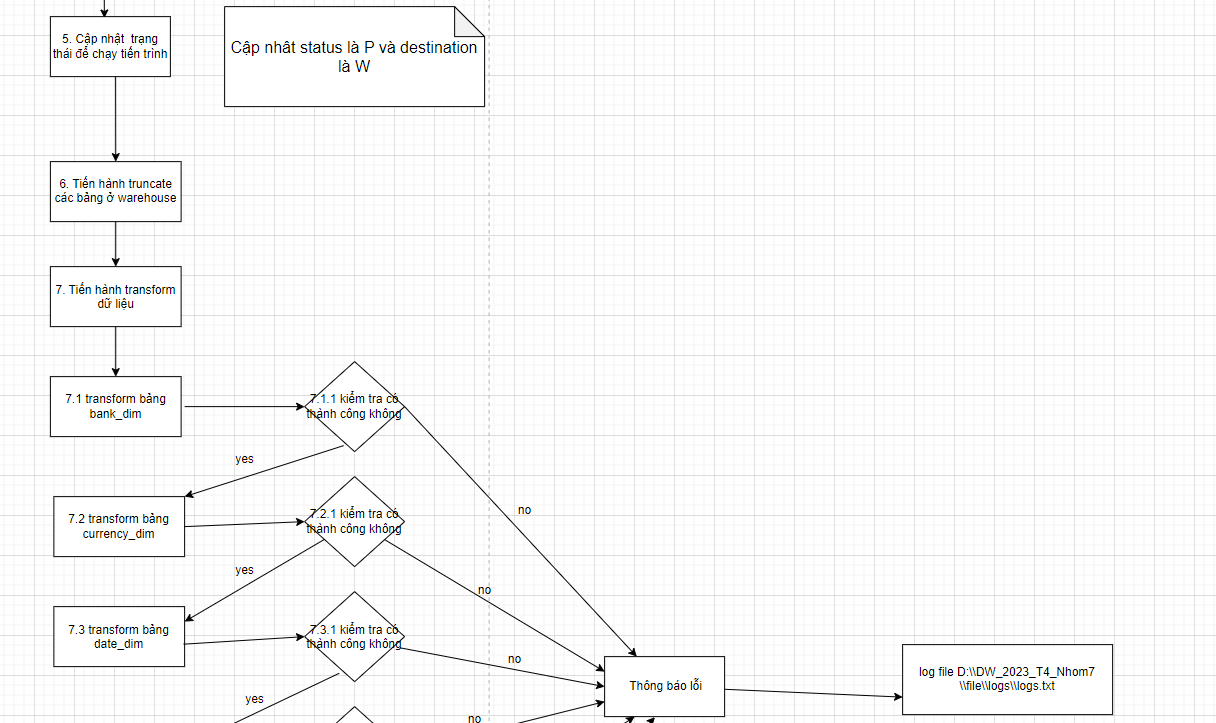
* Link xem chi tiết : [warehouse.drawio - Đẩy dữ liệu từ file vào staging (diagrams.net)](https://app.diagrams.net/?src=about#G1XdLJlheOkOdNm6JnVvdiHn8Rx7xPfkU7)

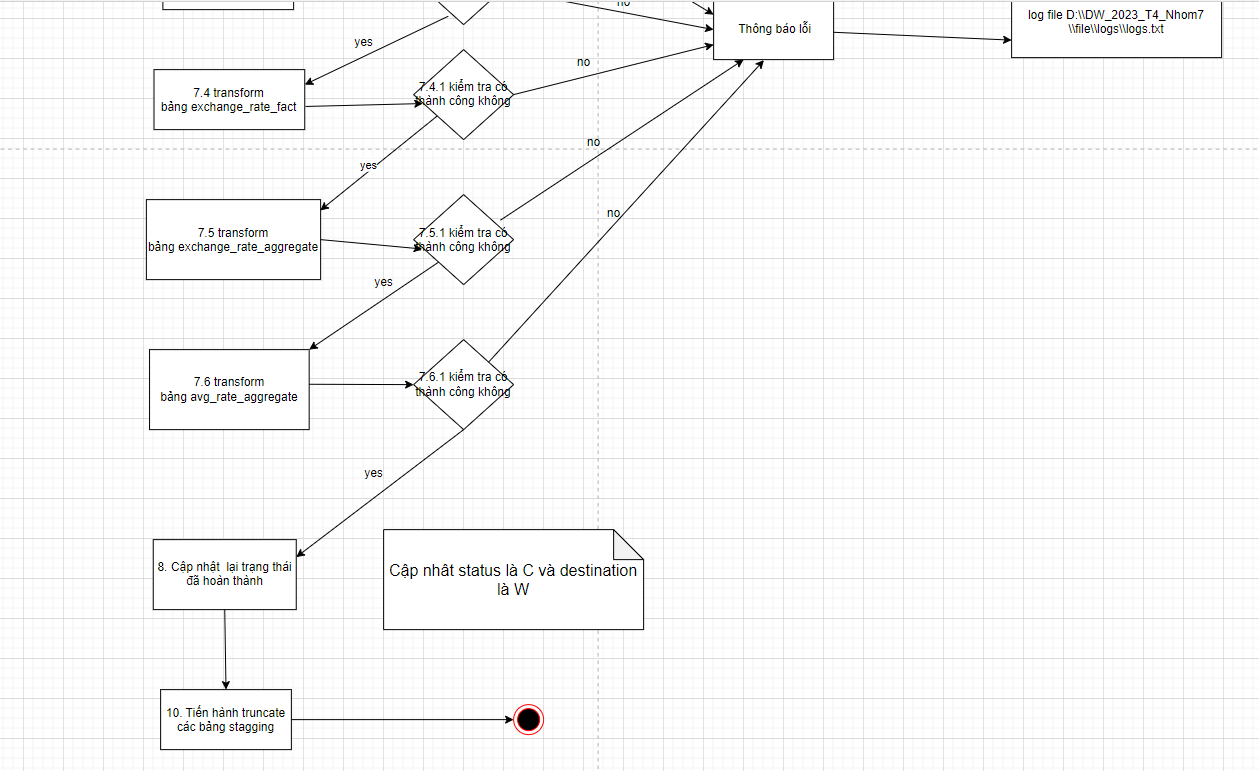


## 6.3 Từ database staging vào database warehouse

Link: [https://app.diagrams.net/#G1XdLJlheOkOdNm6JnVvdiHn8Rx7xPfkU7#%7B%22pageId%22%3A%22yDn8jKFfcJJHlvqATgxq%22%7D](https://app.diagrams.net/#G1XdLJlheOkOdNm6JnVvdiHn8Rx7xPfkU7)

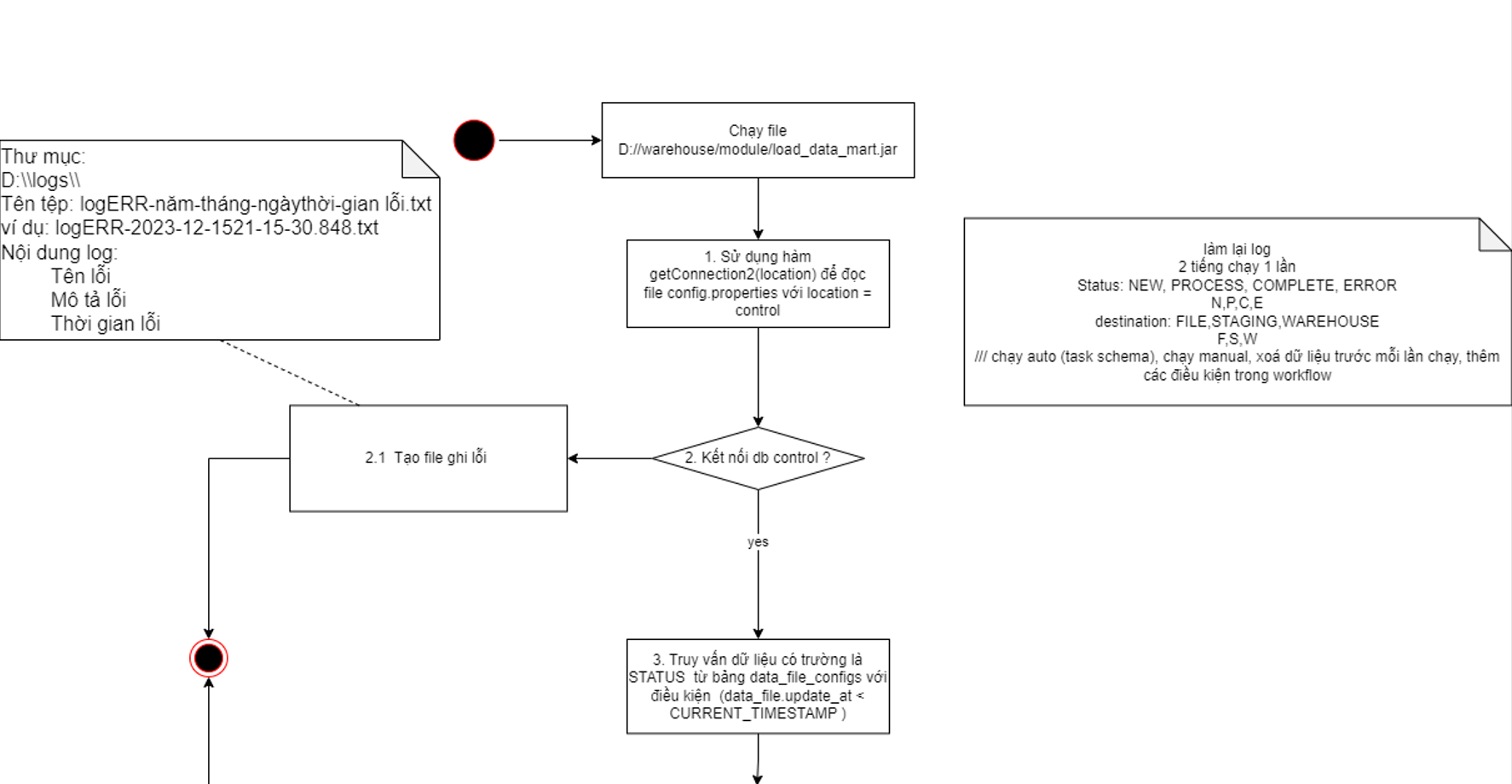


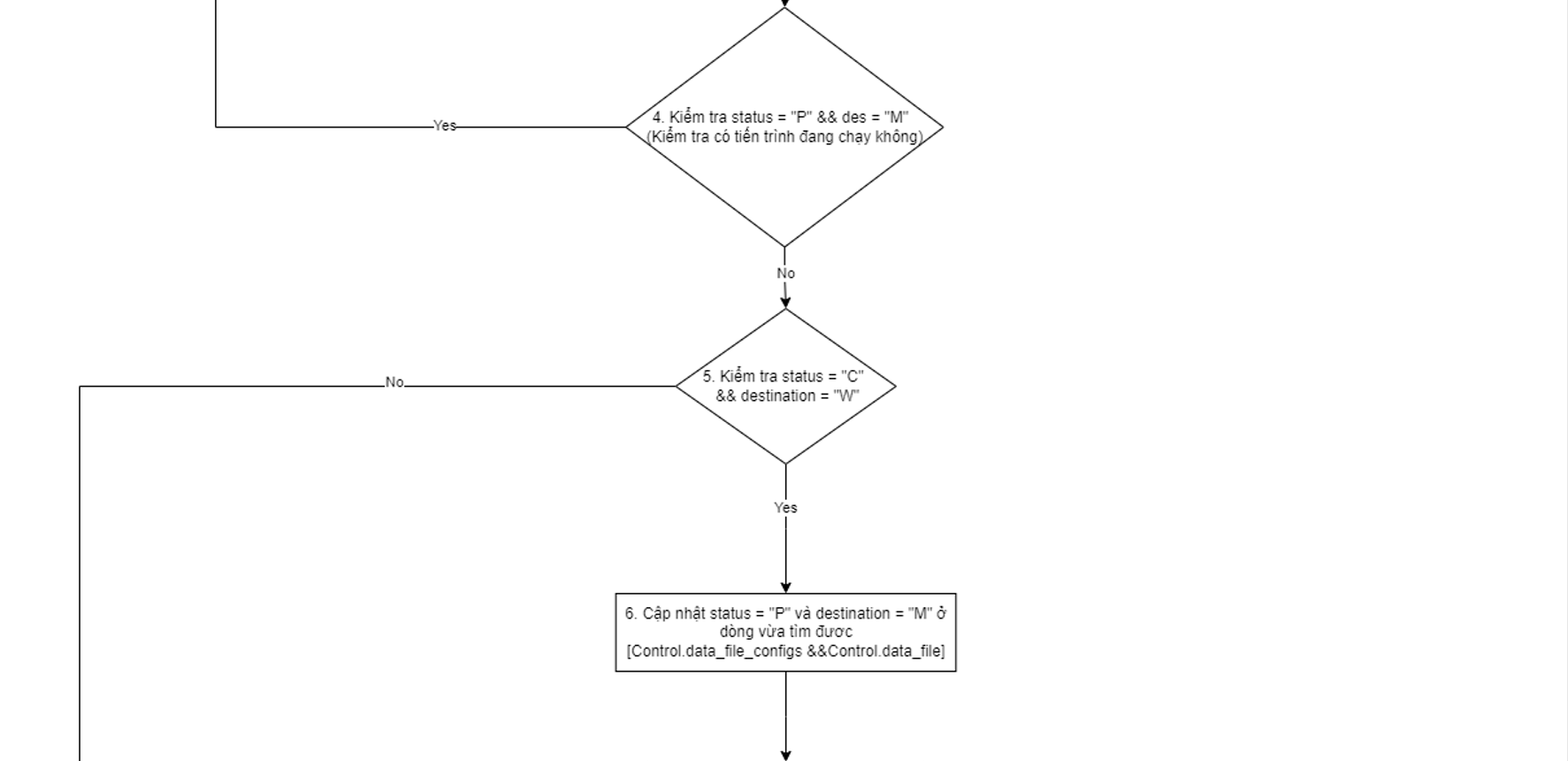


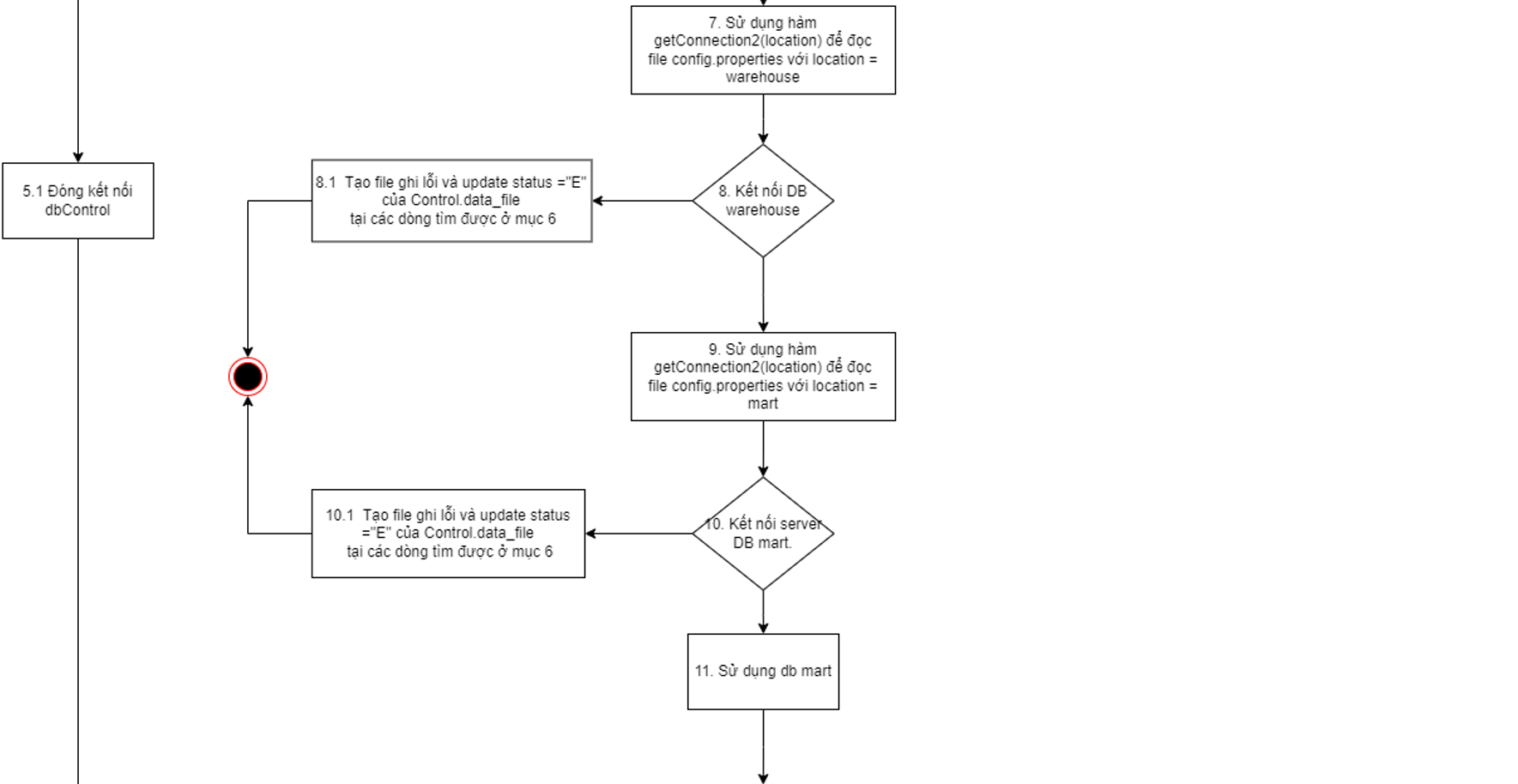


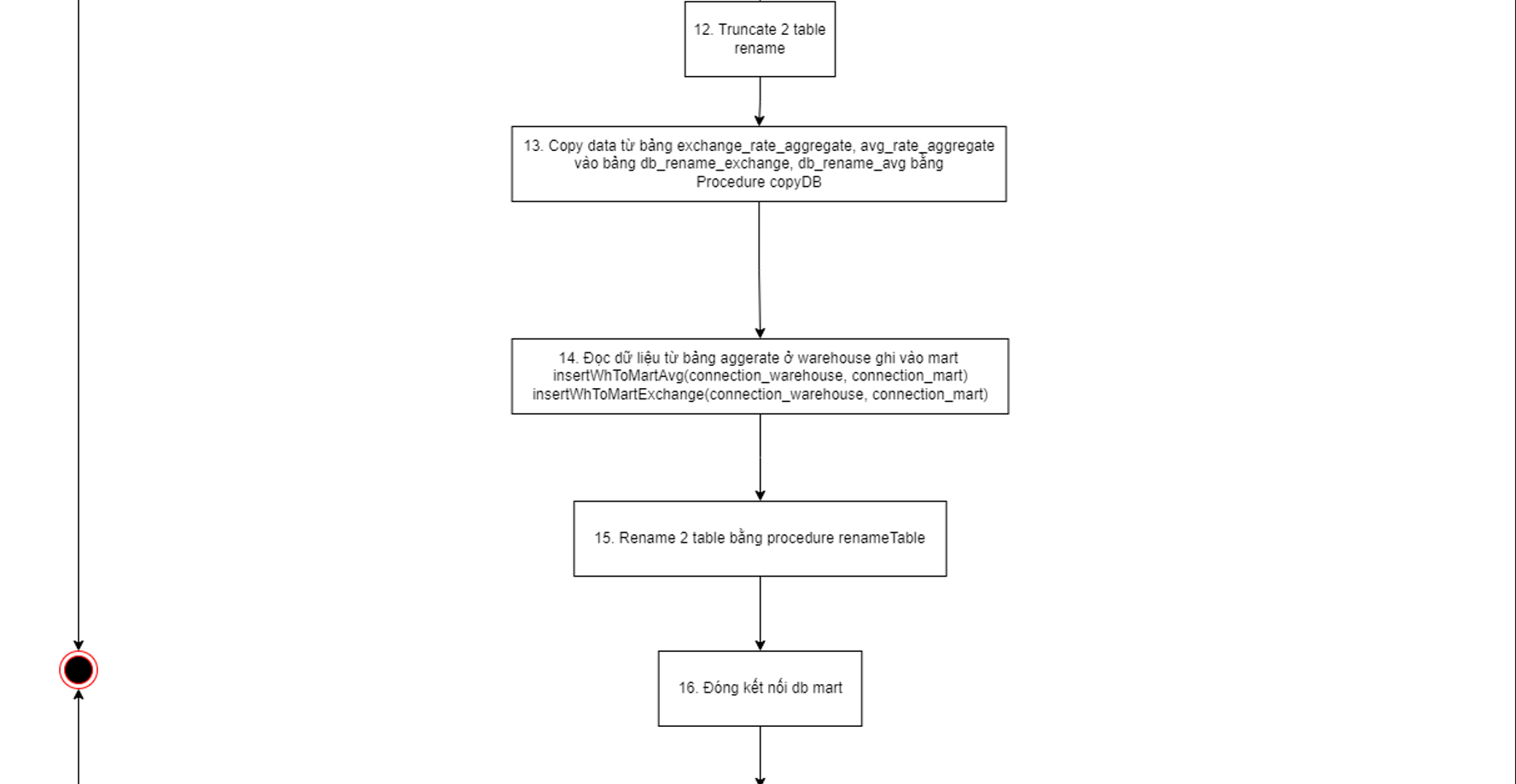
## 6.4 Từ database warehouse vào database mart

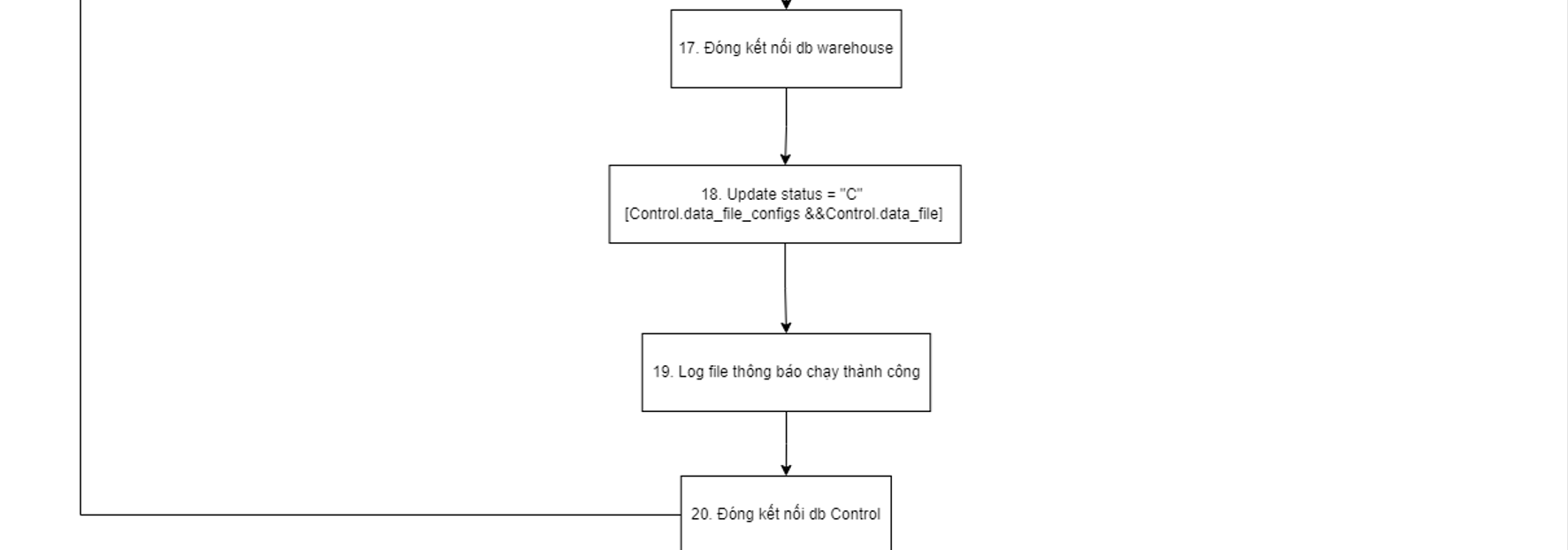
[https://app.diagrams.net/#G1XdLJlheOkOdNm6JnVvdiHn8Rx7xPfkU7#%7B%22pageId%22%3A%228FtGb0x1CQEs6DZeBxJy%22%7D](https://app.diagrams.net/#G1XdLJlheOkOdNm6JnVvdiHn8Rx7xPfkU7)





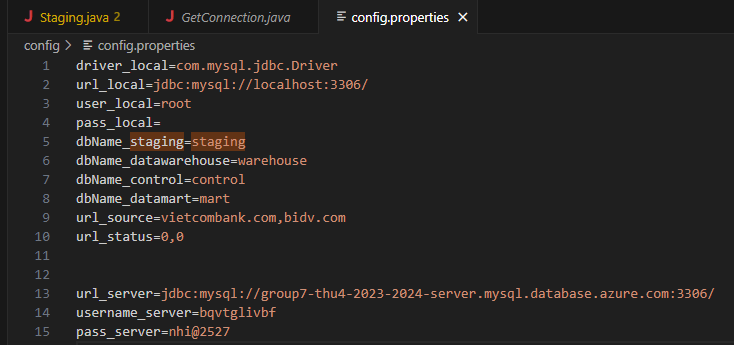






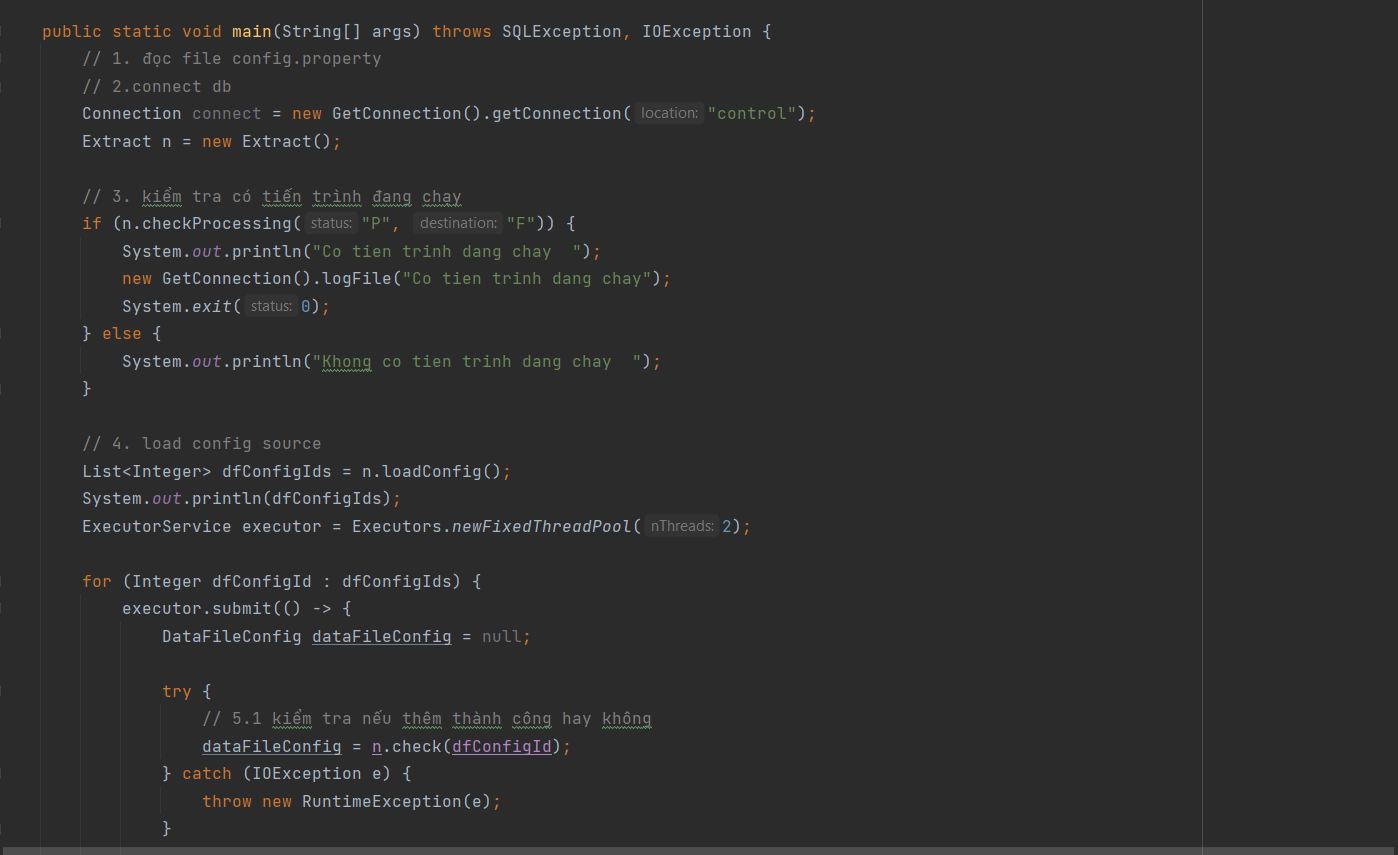
# Code & liên hệ mô tả

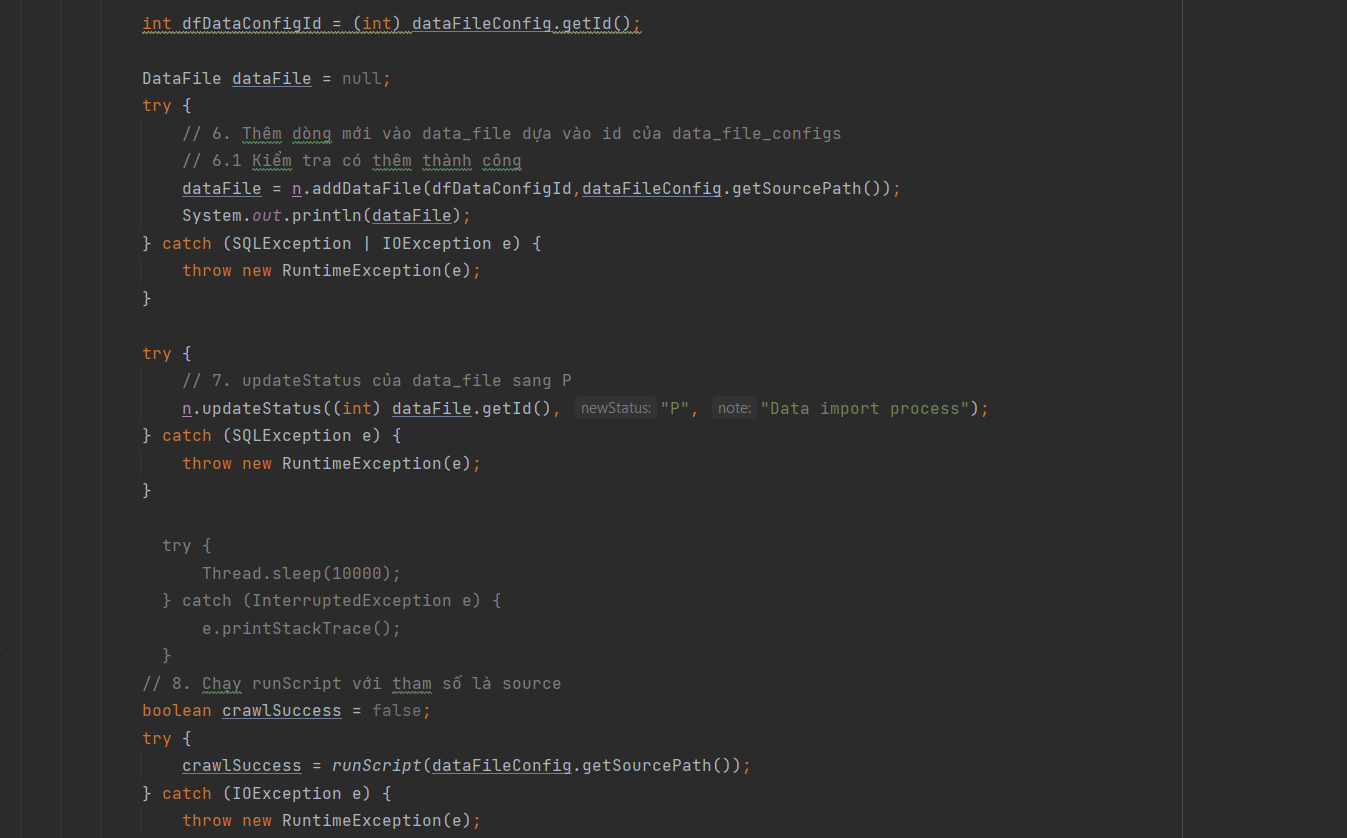
* File Config.properties

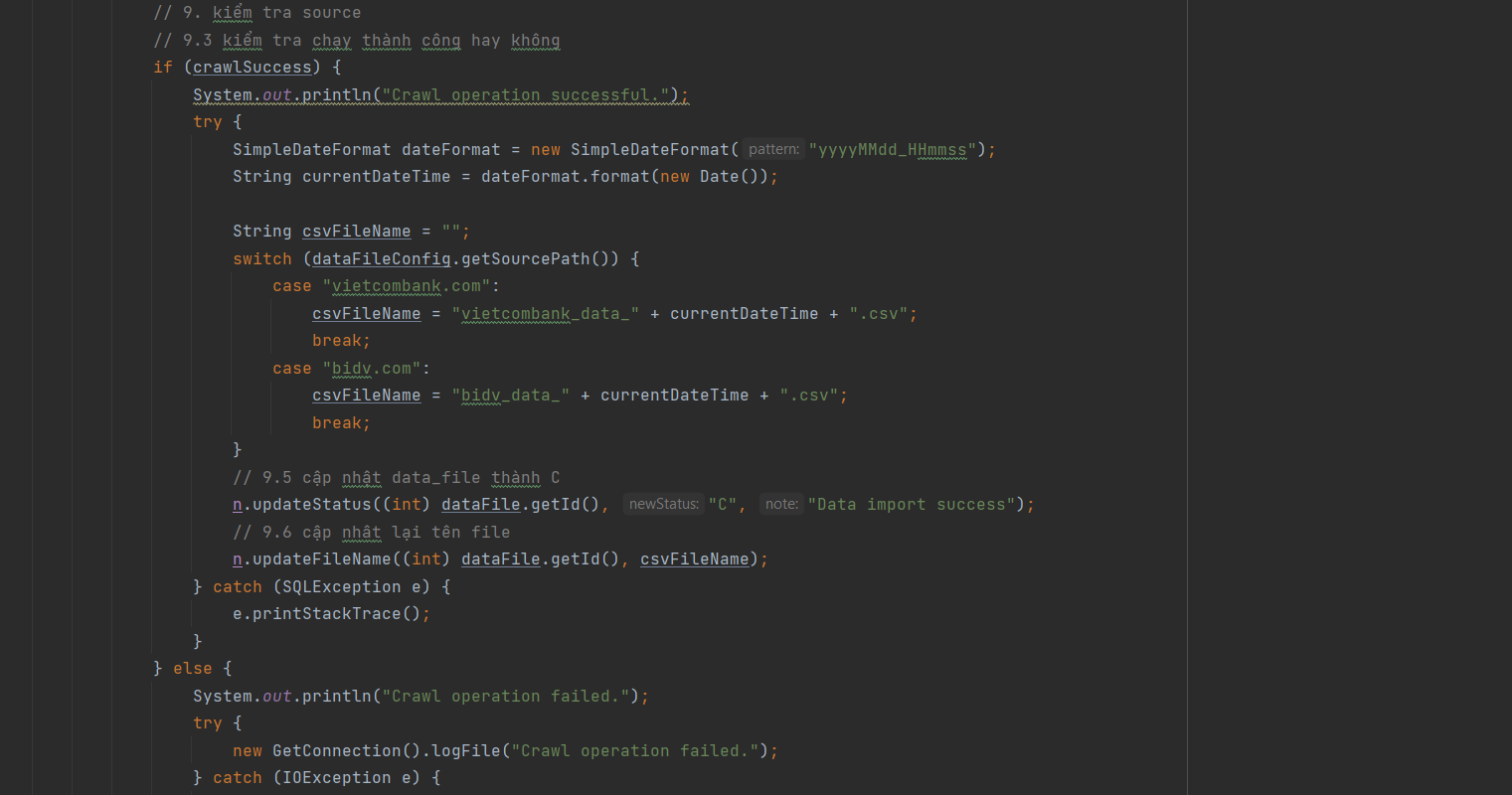


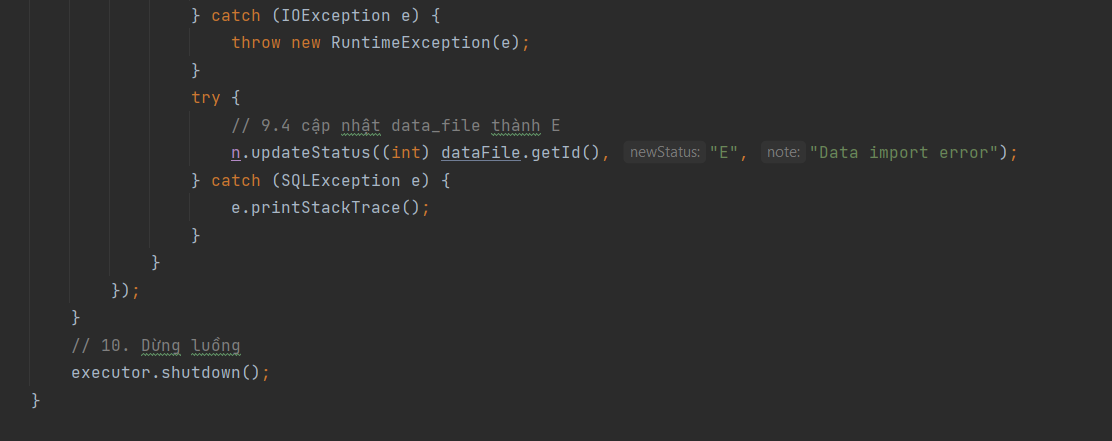
## 7.1 Từ source về file

* File Extract.java

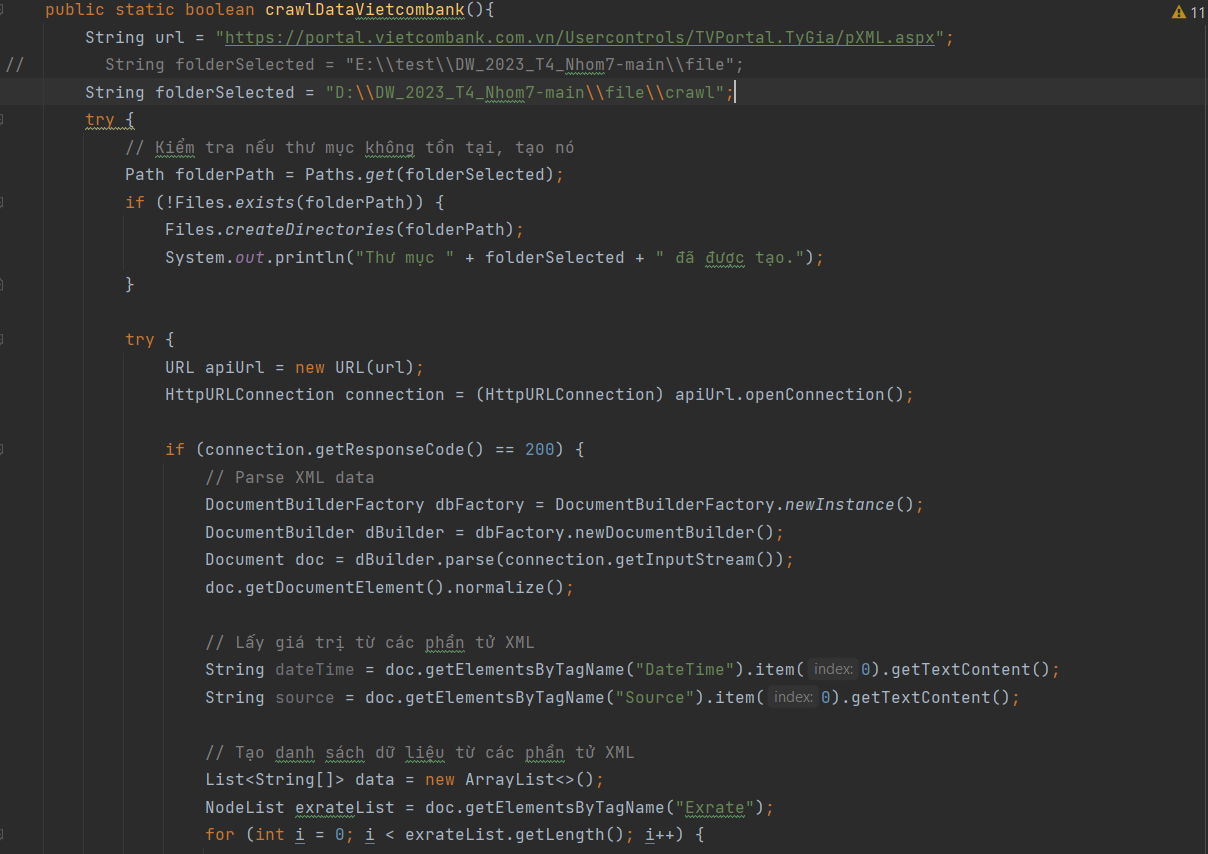


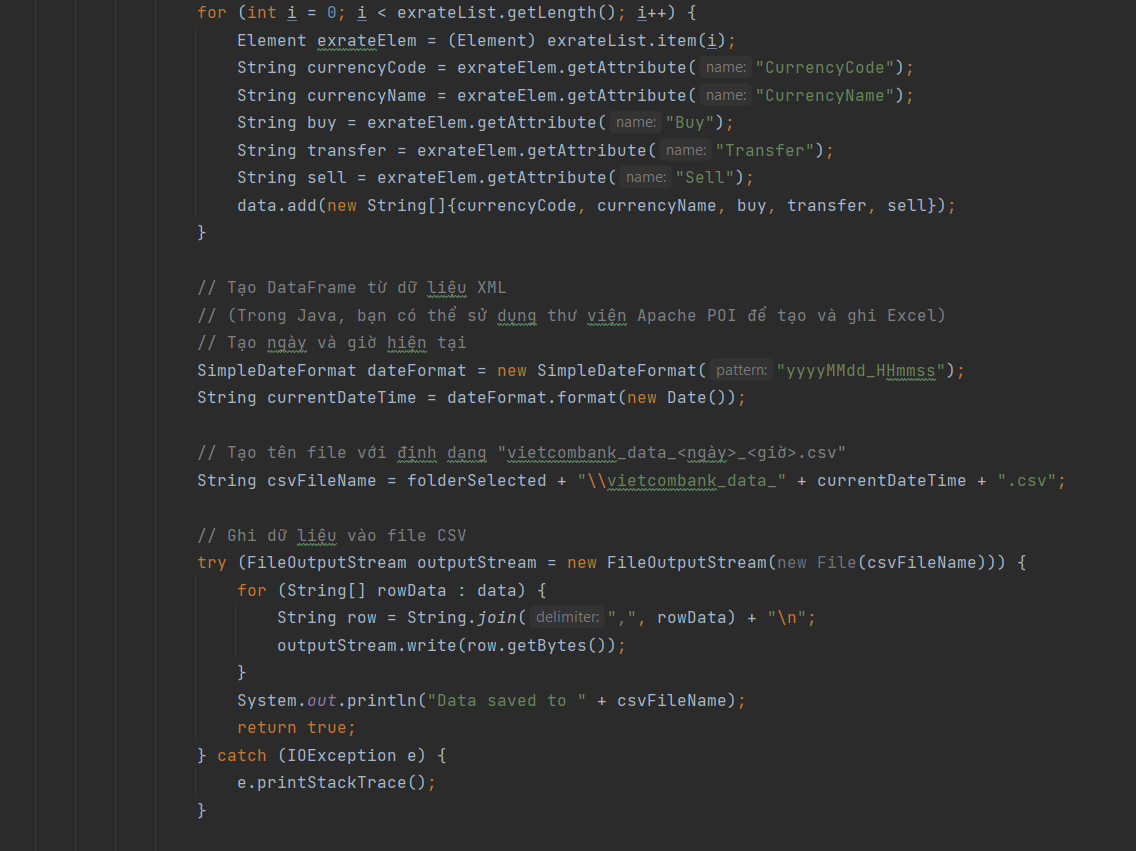


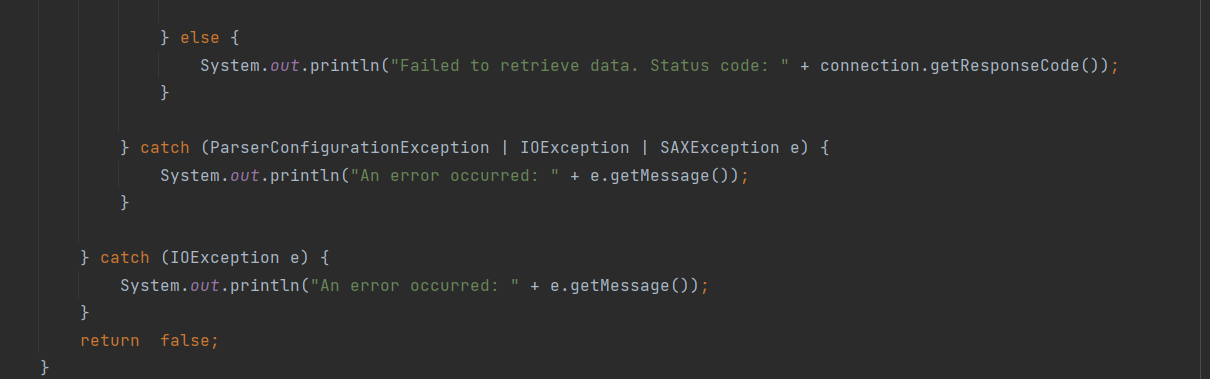




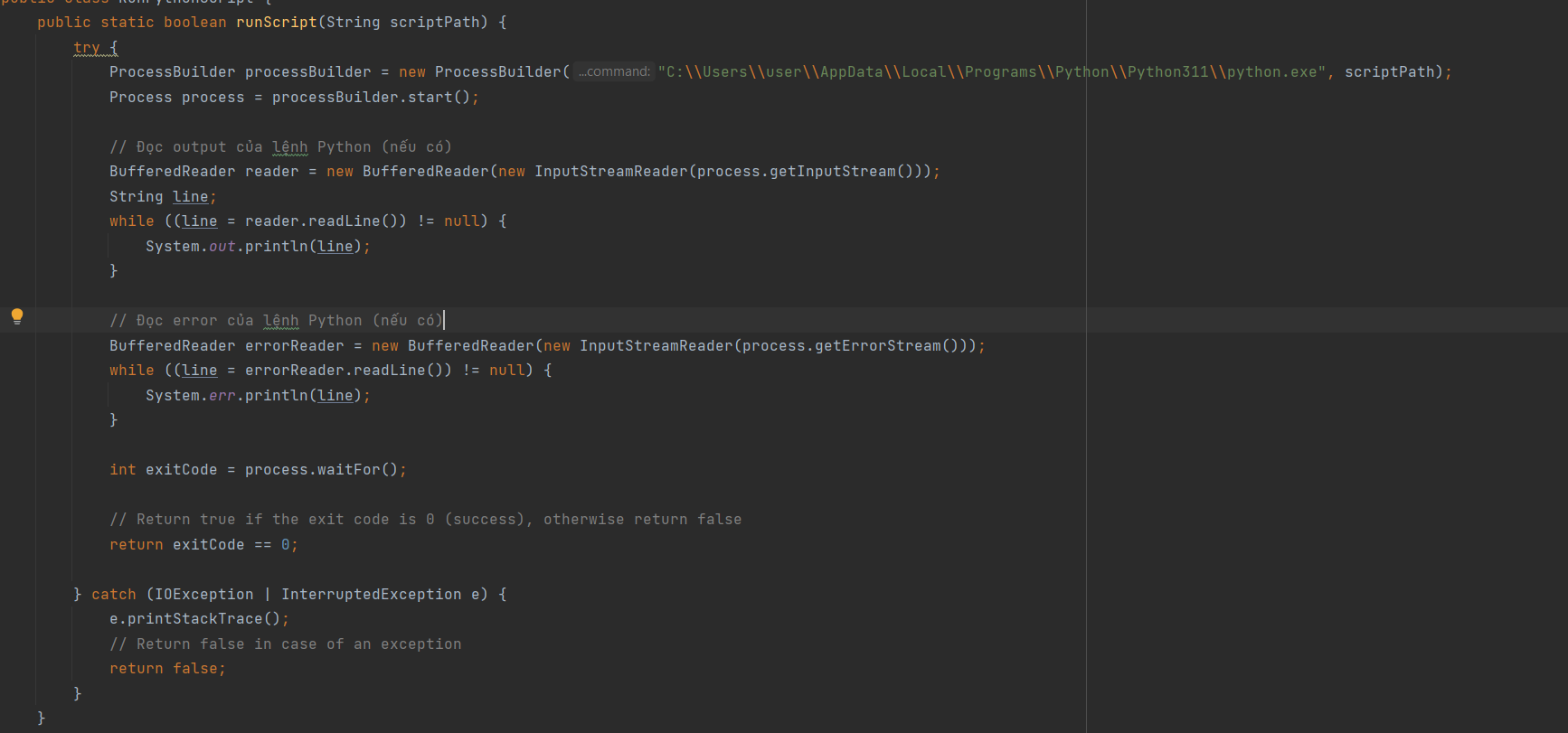
* File CrawlVietcombank.java



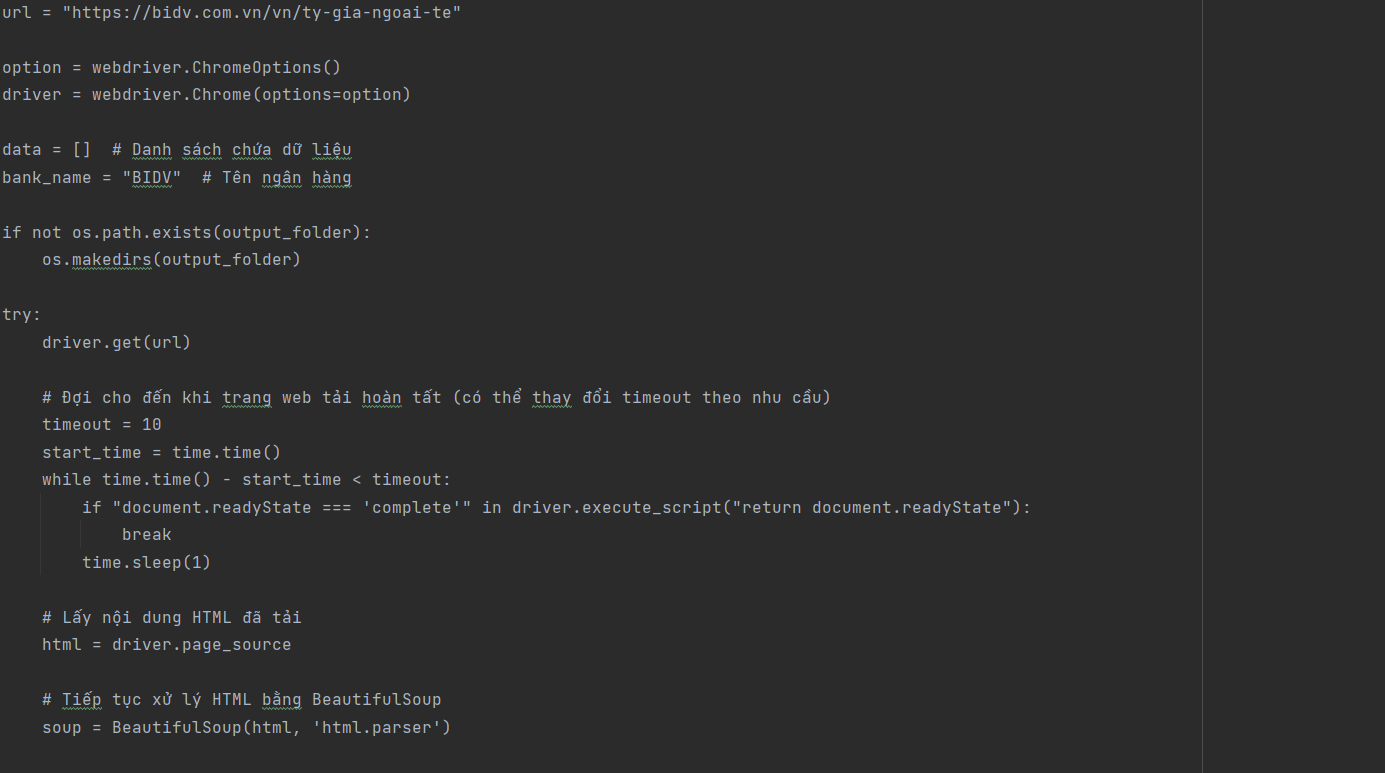


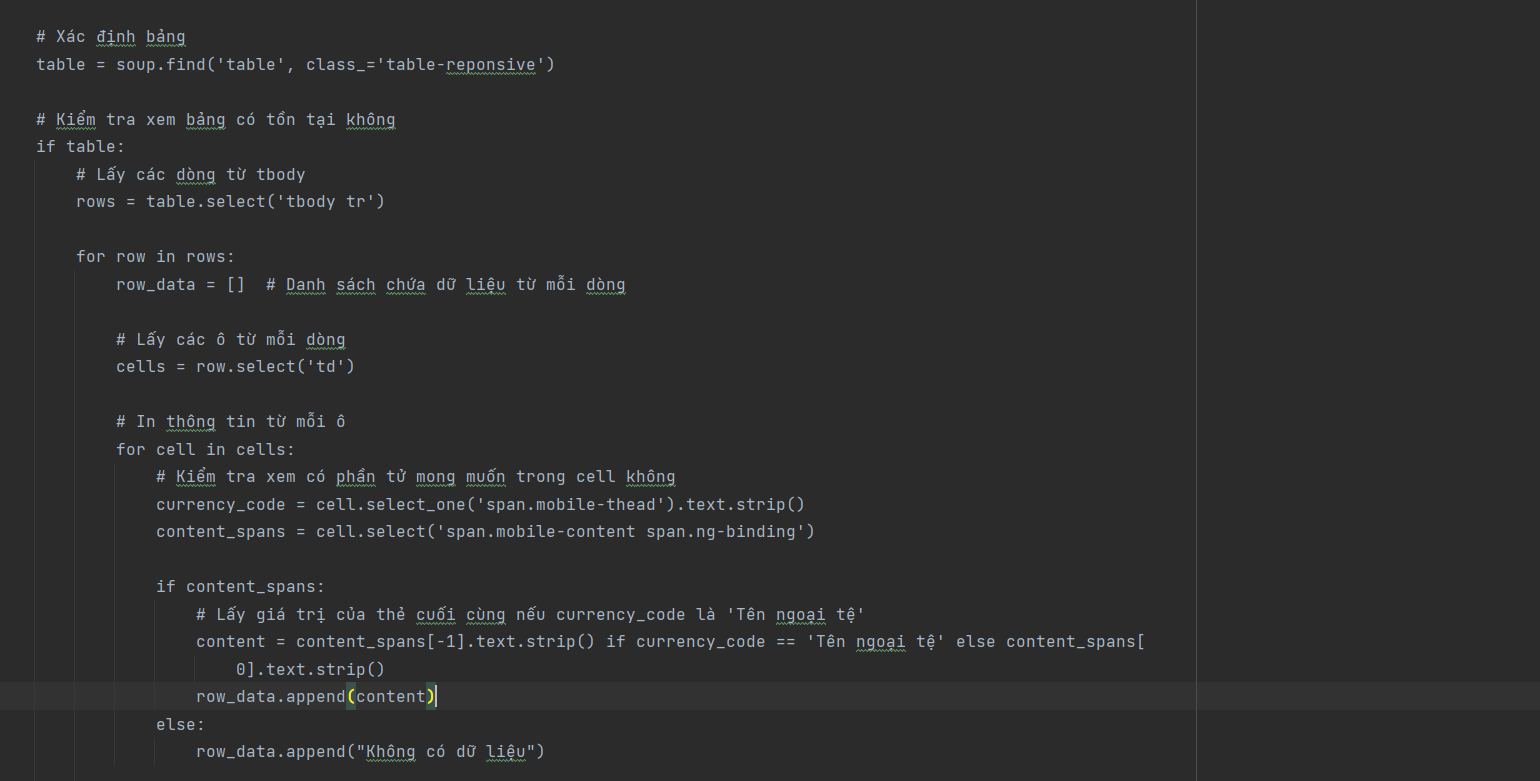


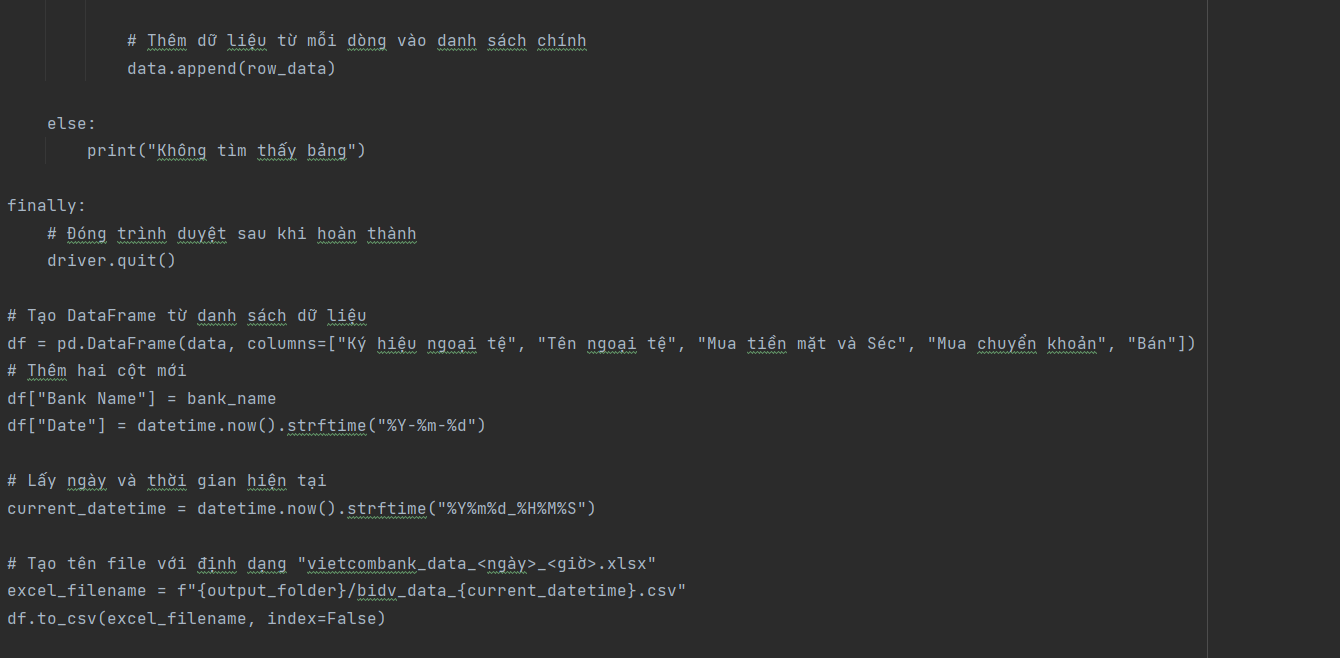
* File RunPythonScript.java



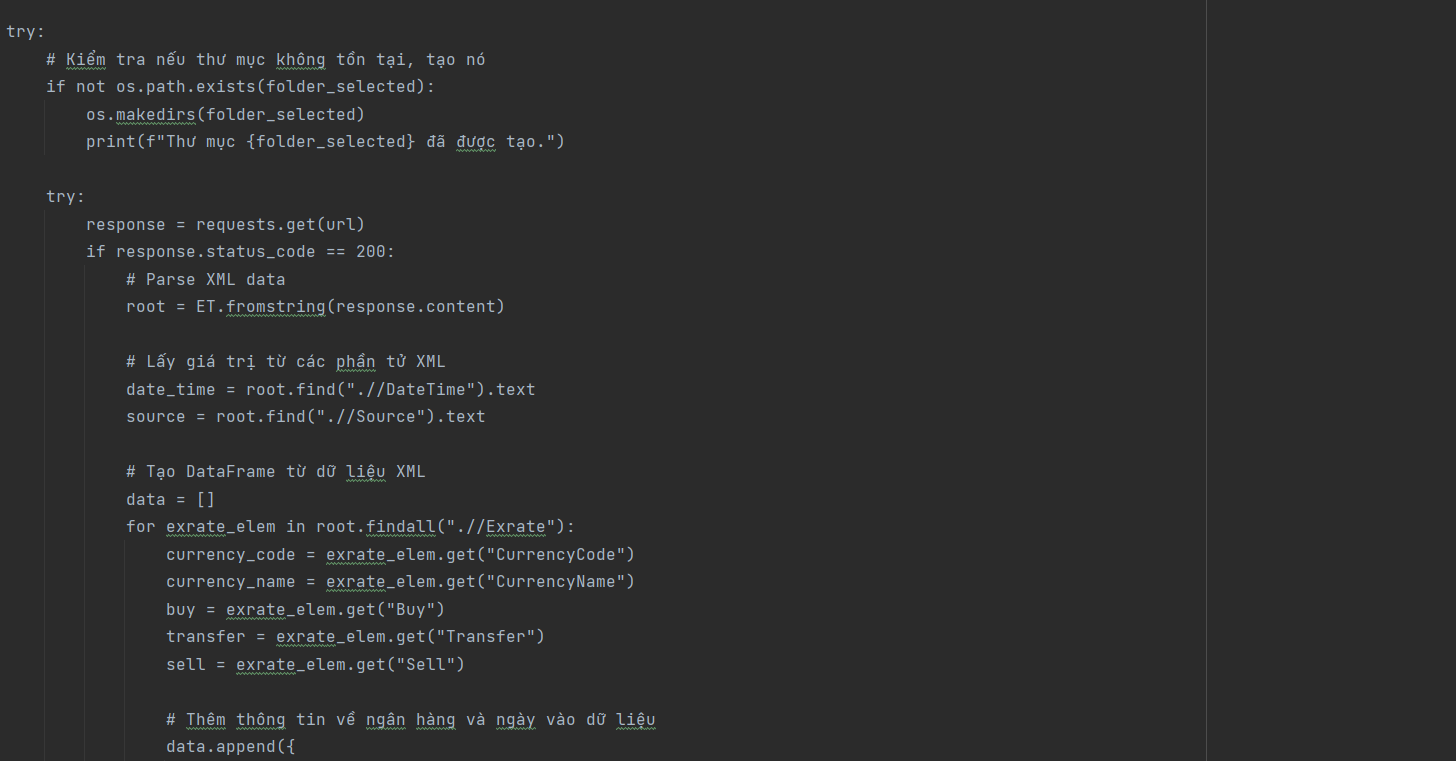
* File bidvcrawl.py

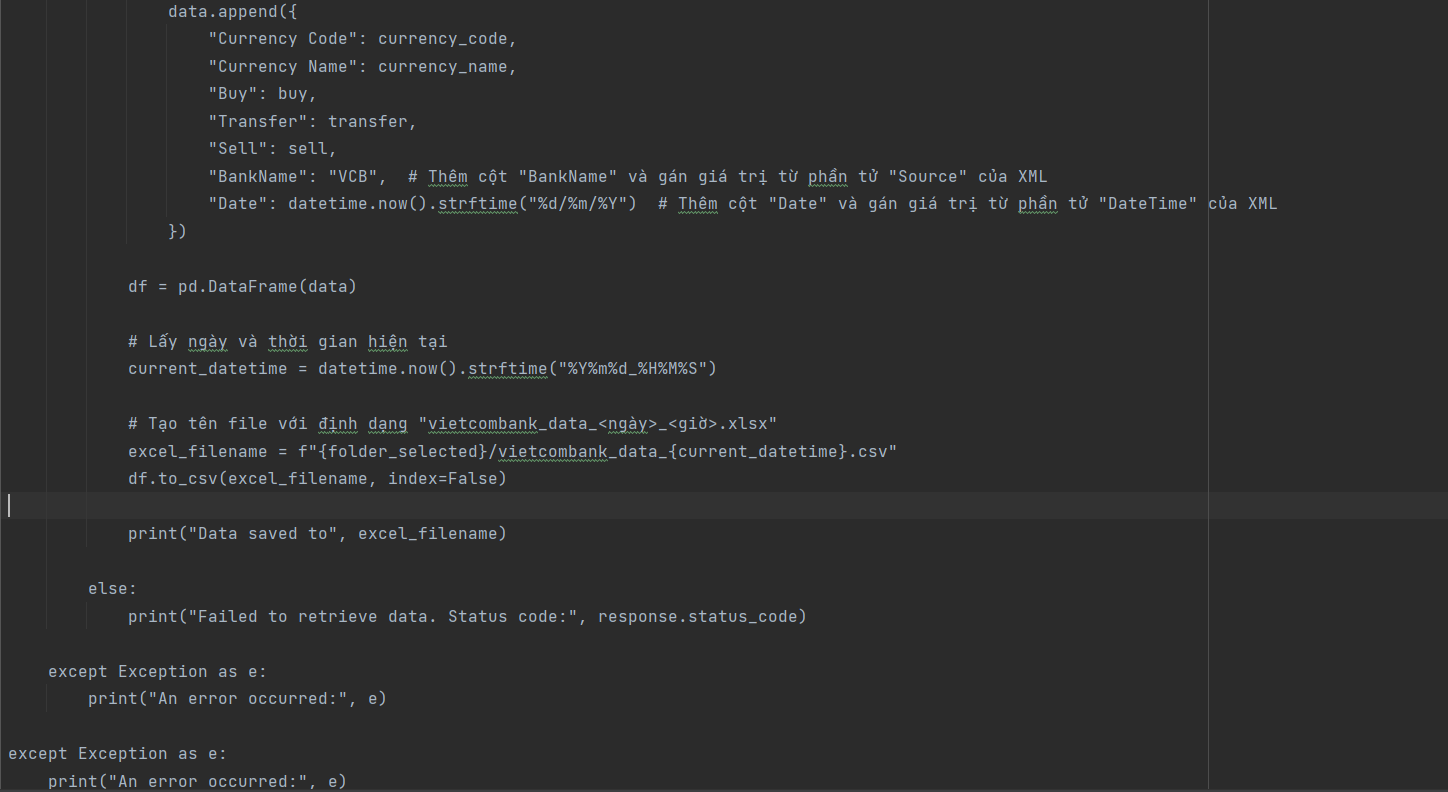






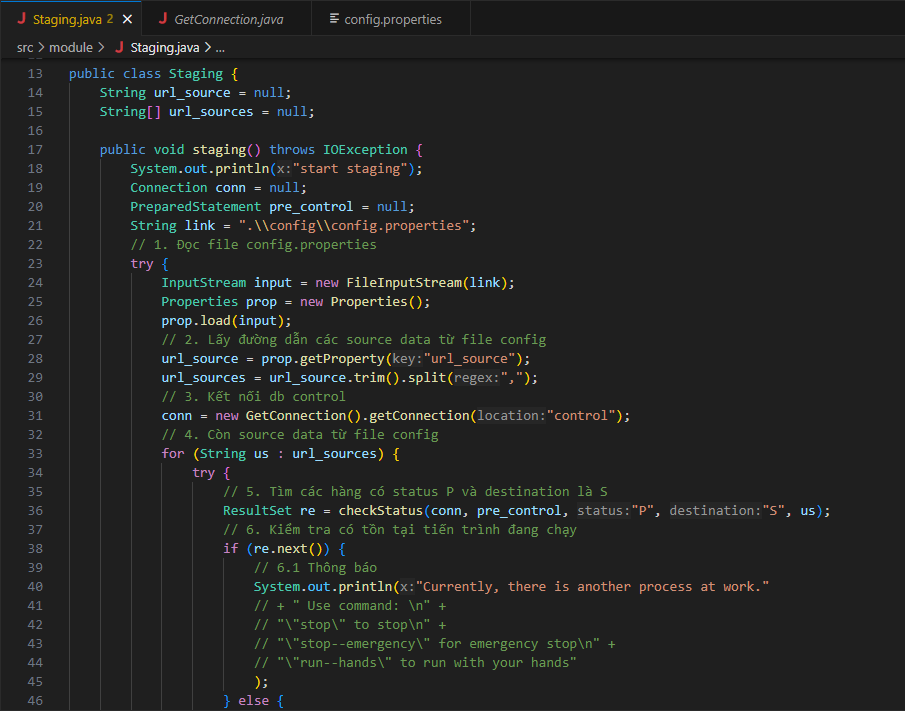
* File vietxml.py

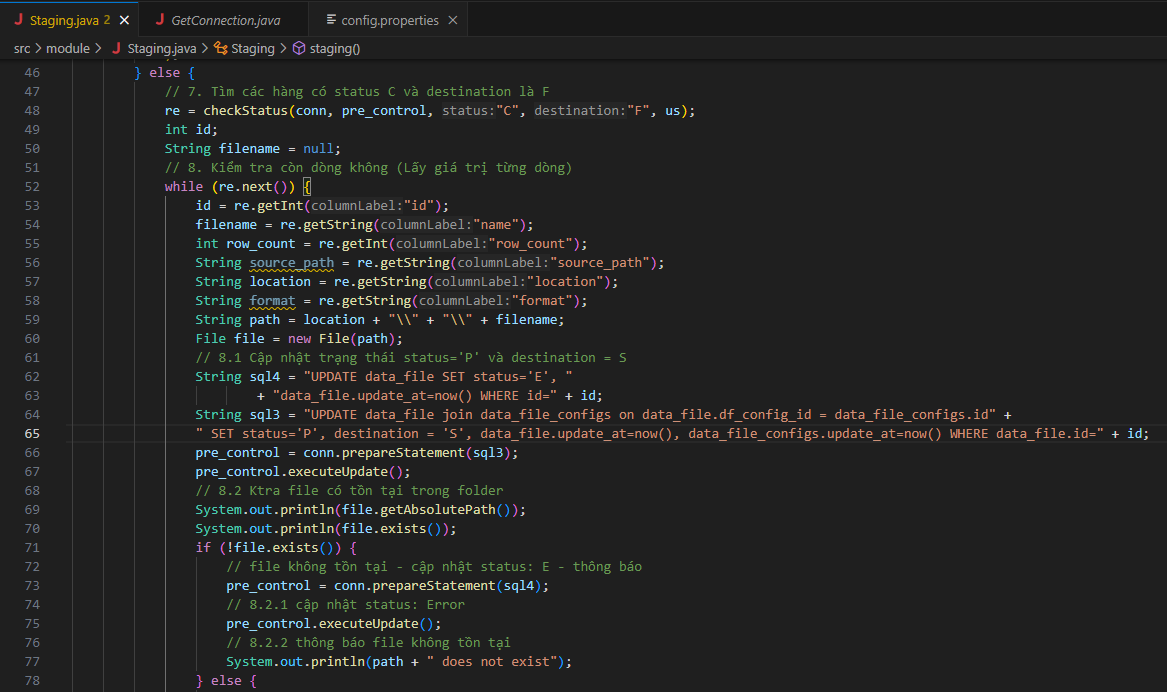


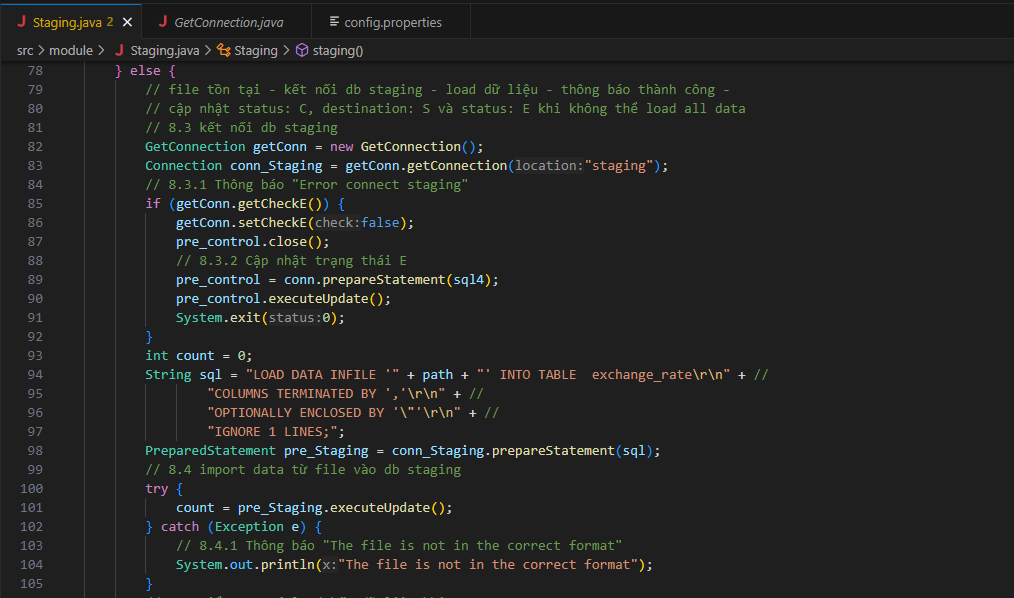


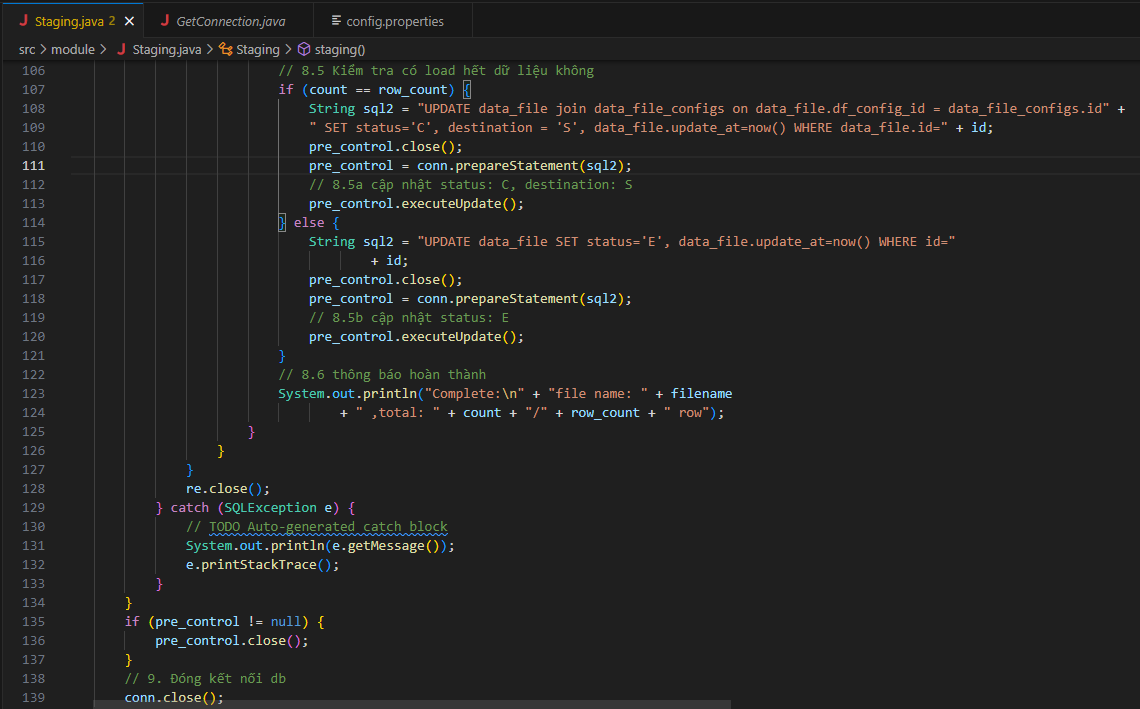
## 7.2 Từ file vào database staging

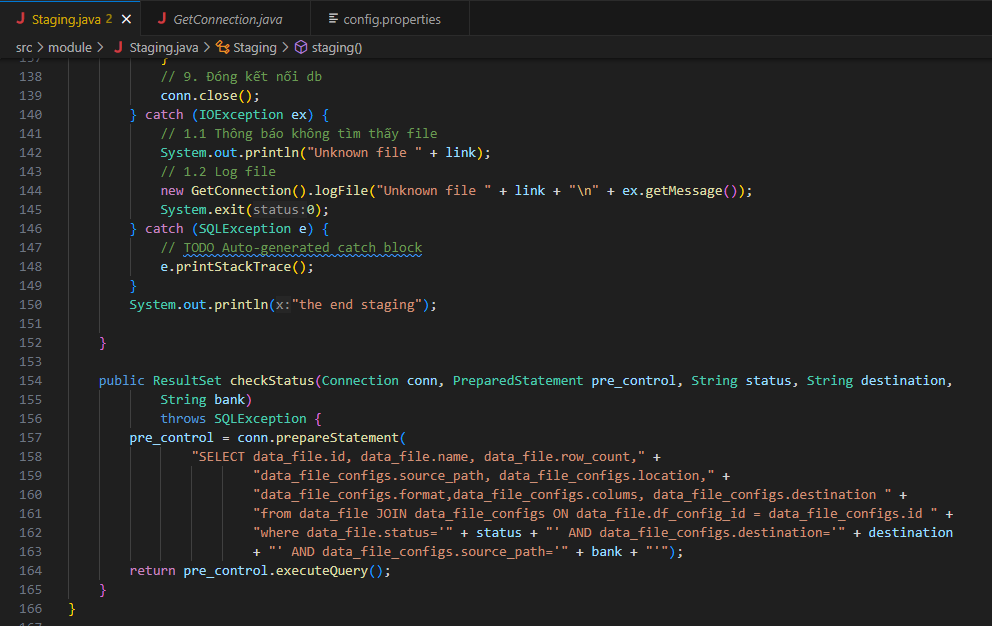
* File Staging.java



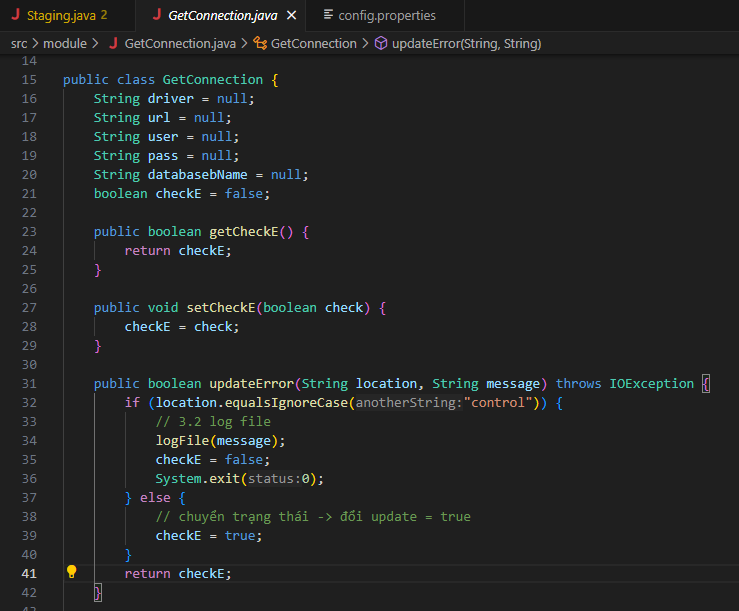


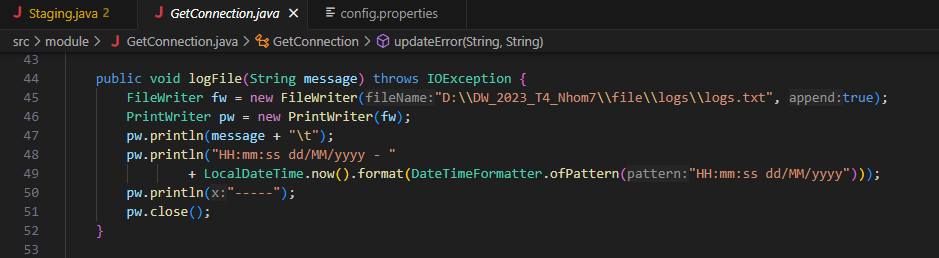


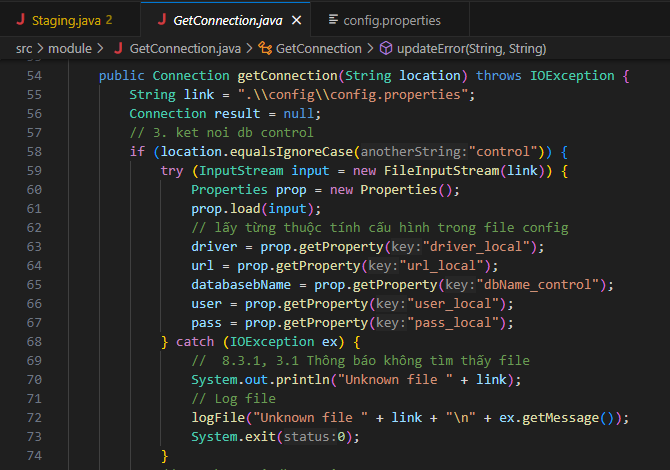


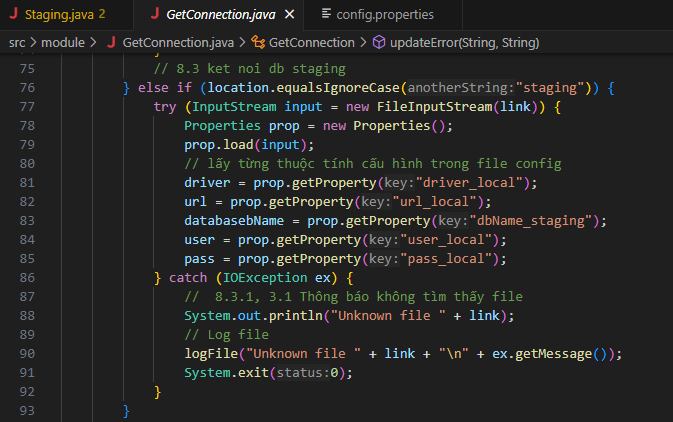


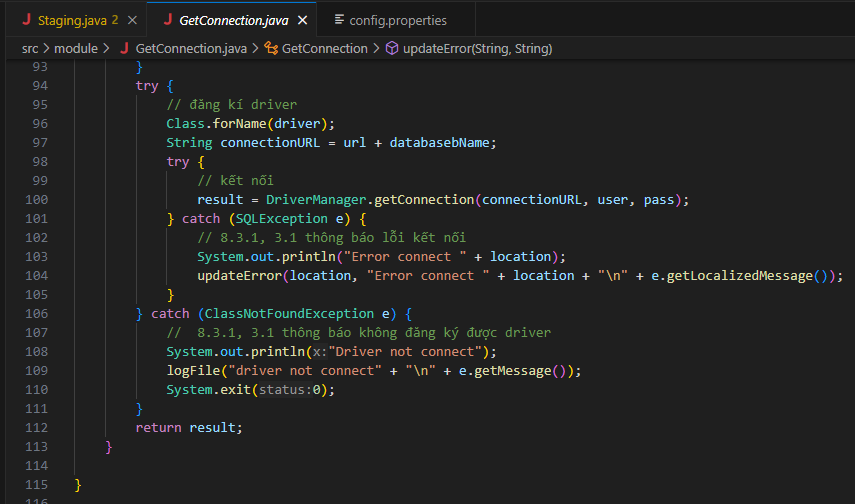
* File GetConnection.java





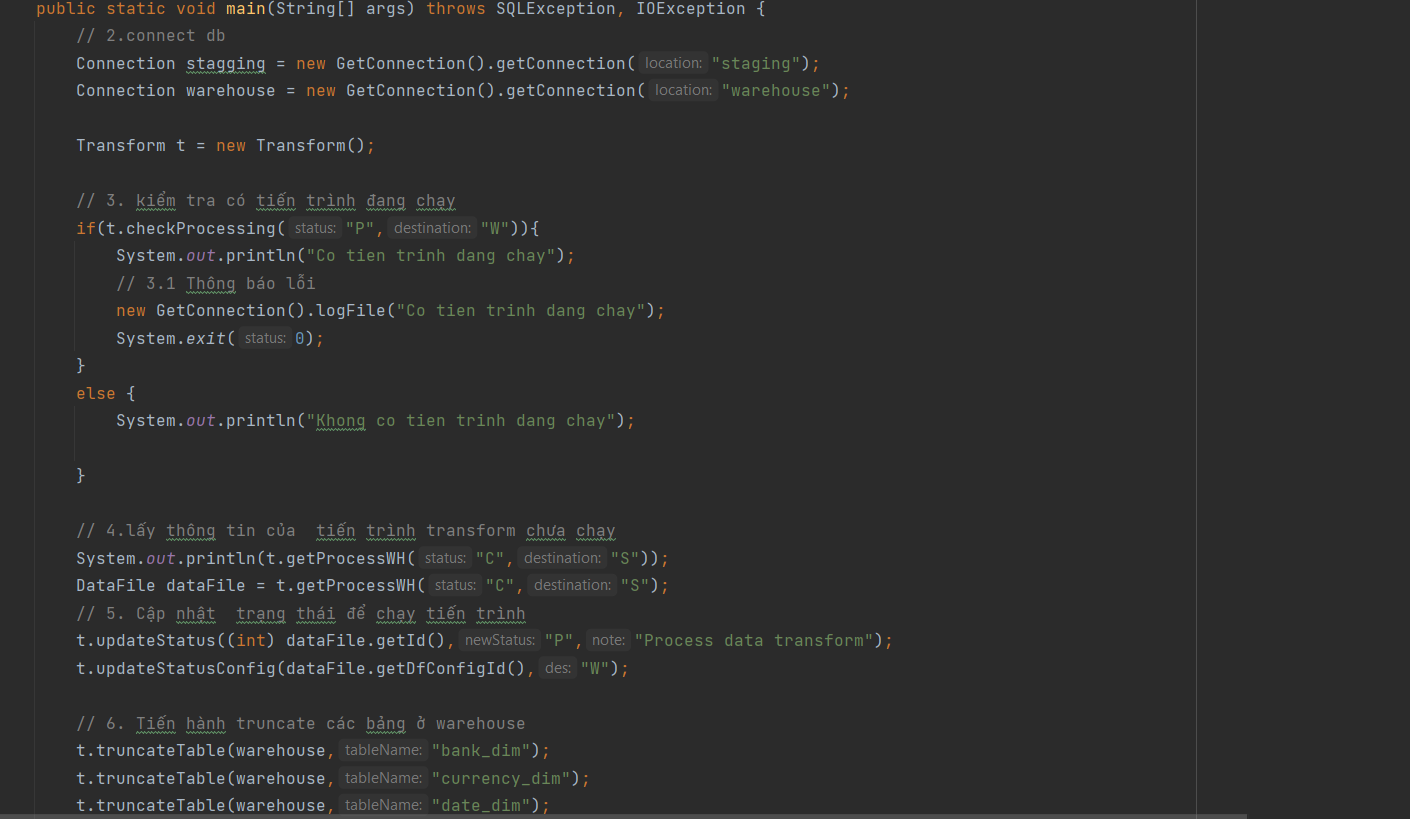


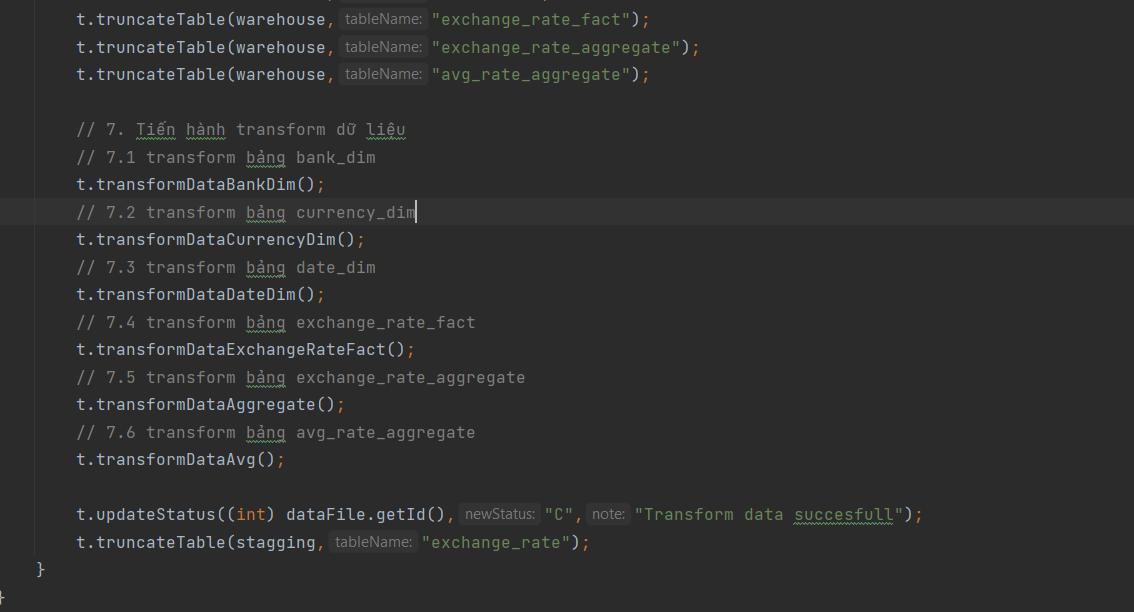




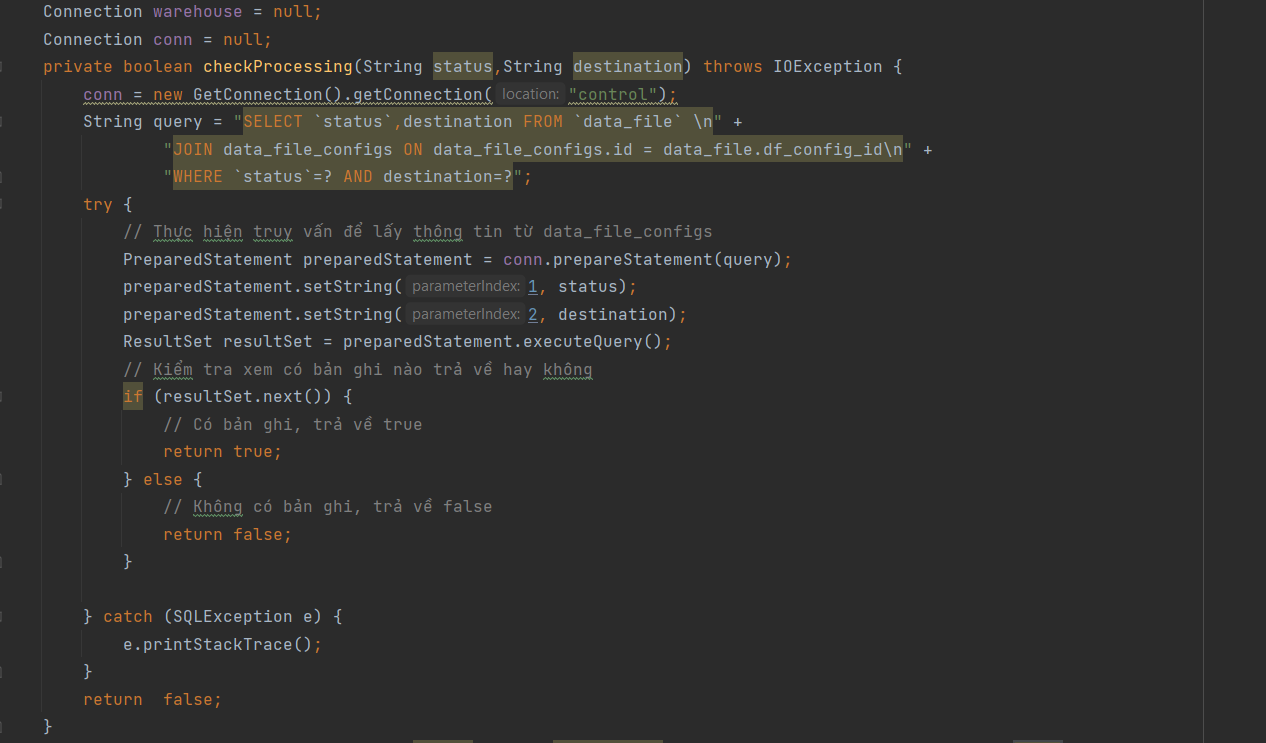
## 7.3 Từ database staging vào database warehouse

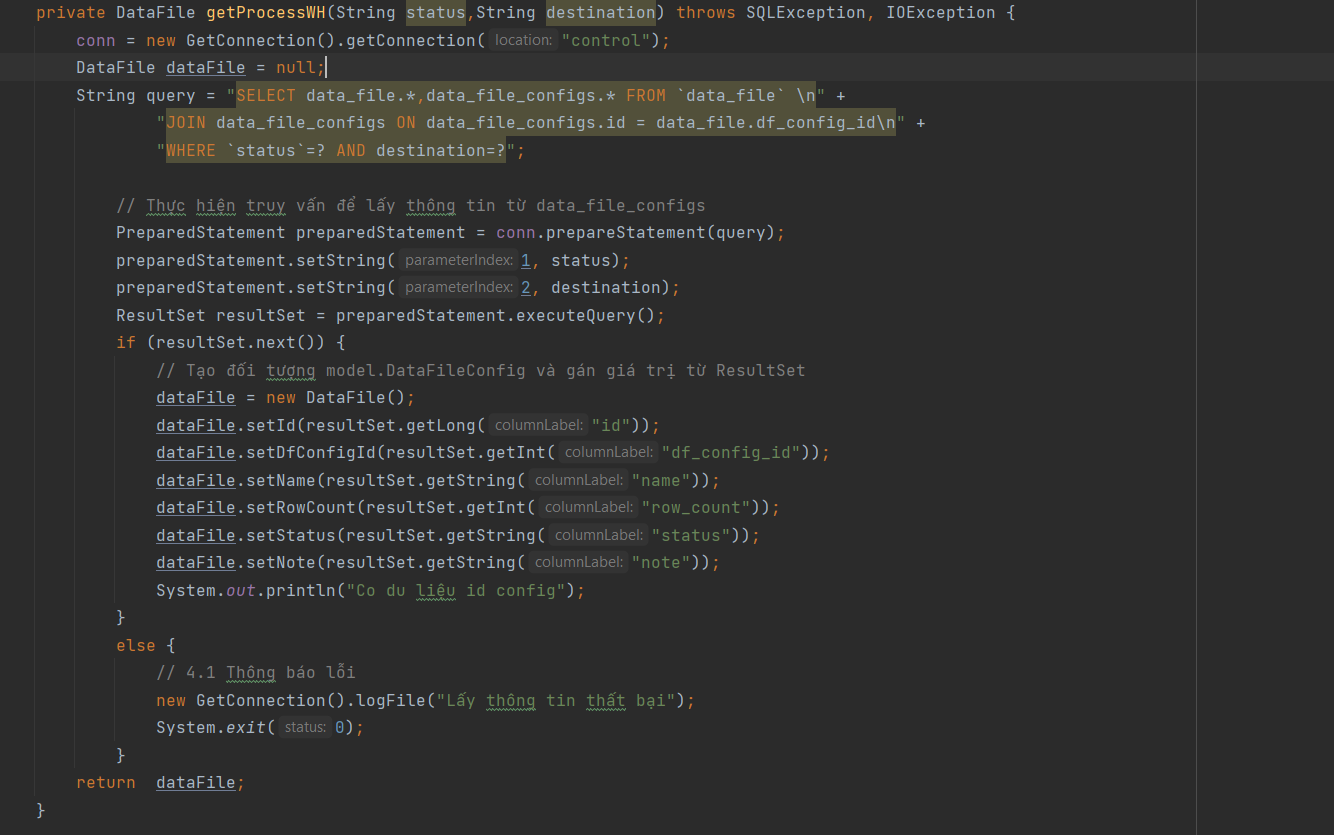
* File Transform.java

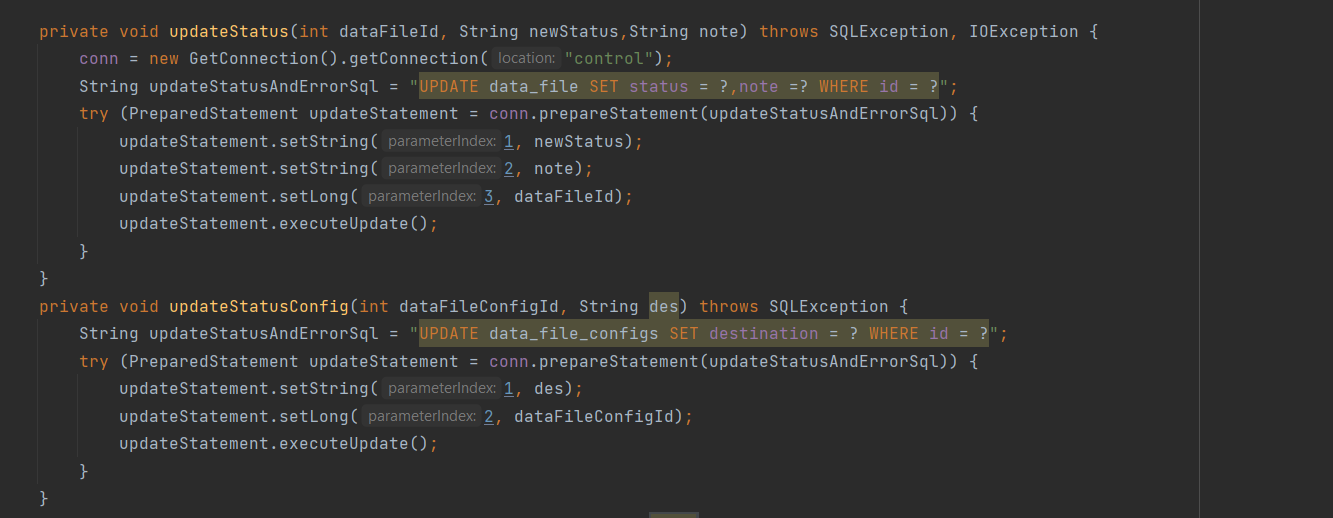


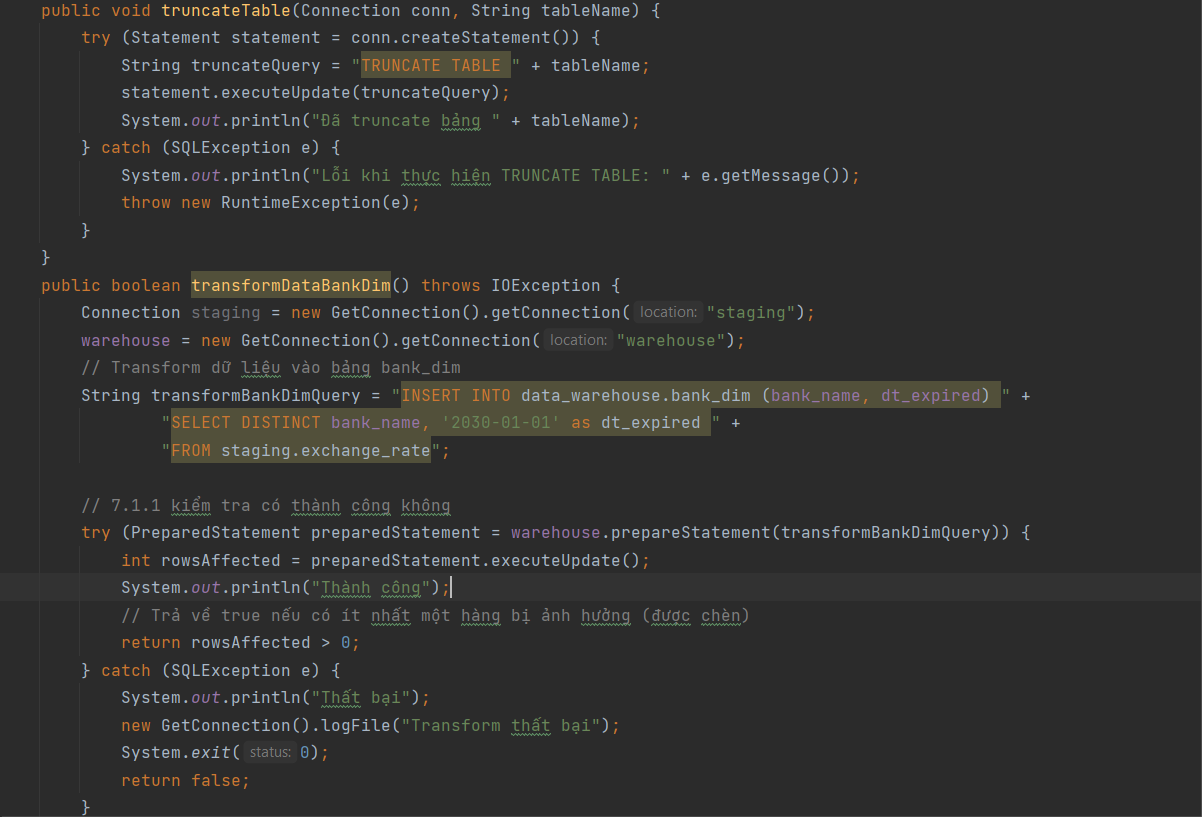


Phương thức liên quan

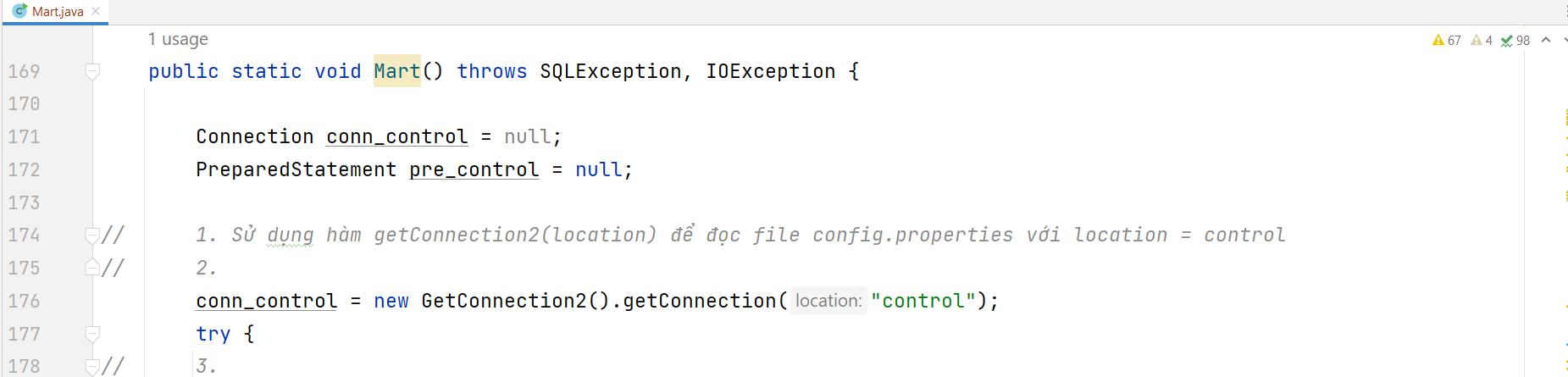




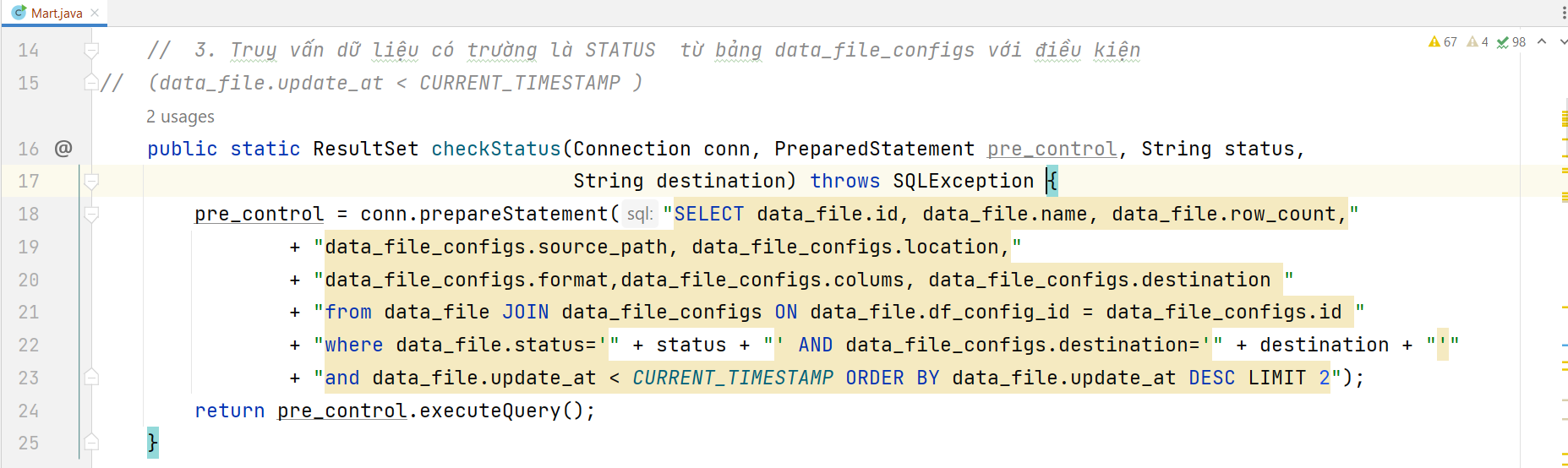




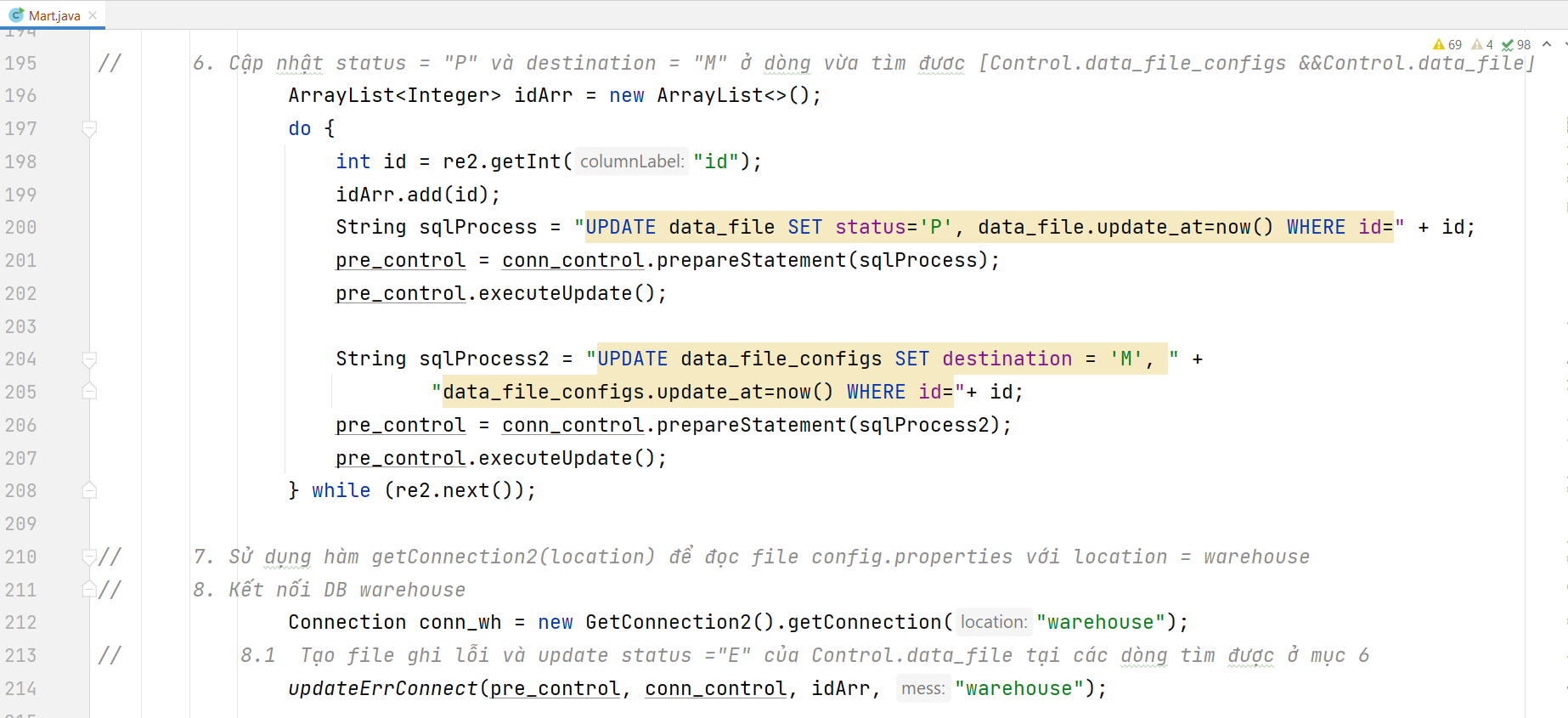
## 7.4 Từ database warehouse vào database mart



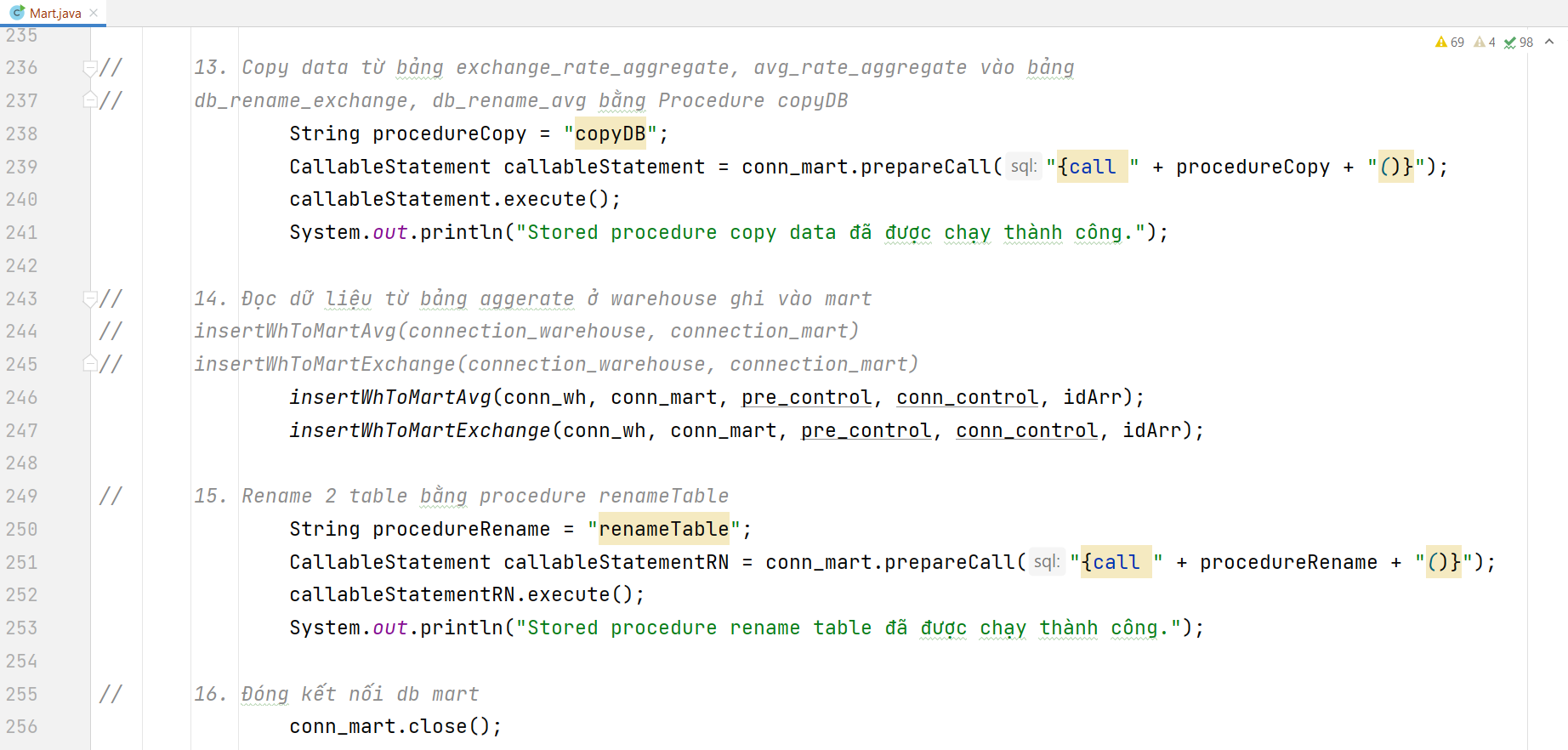


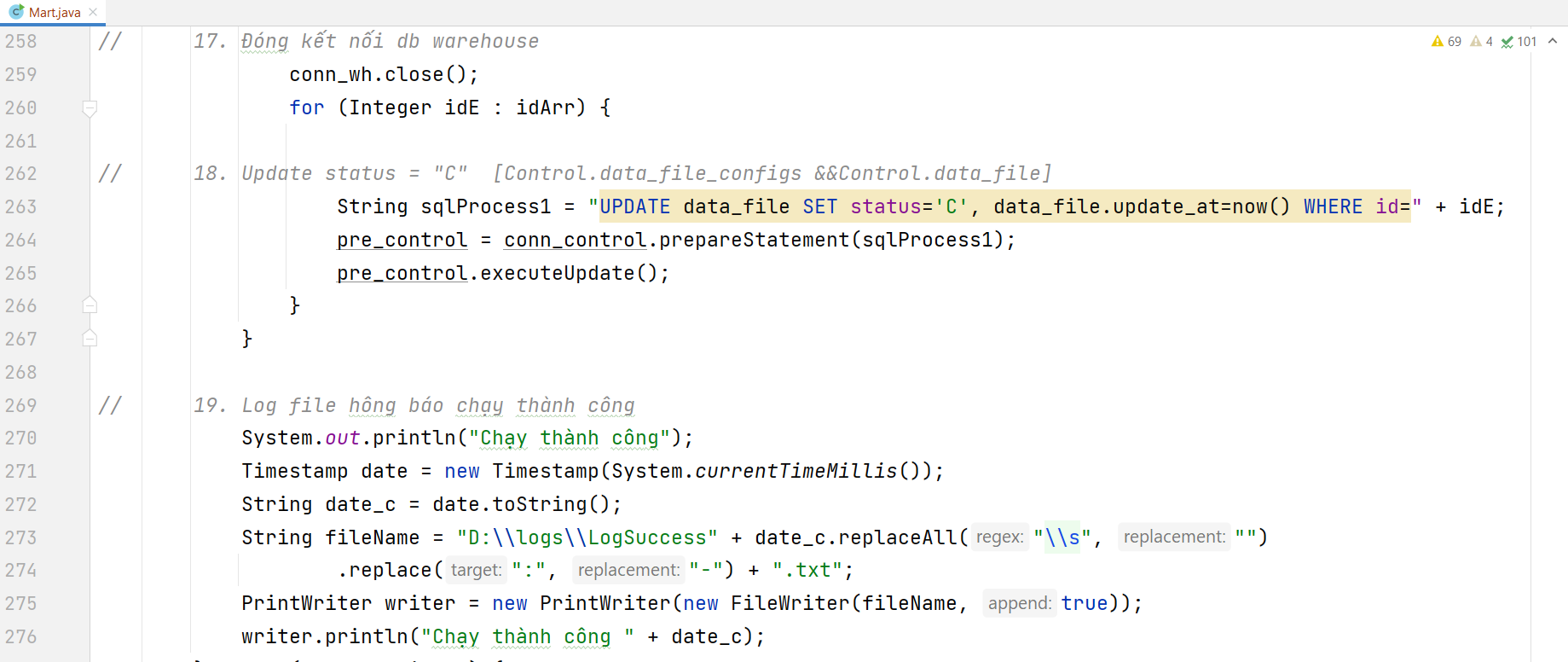


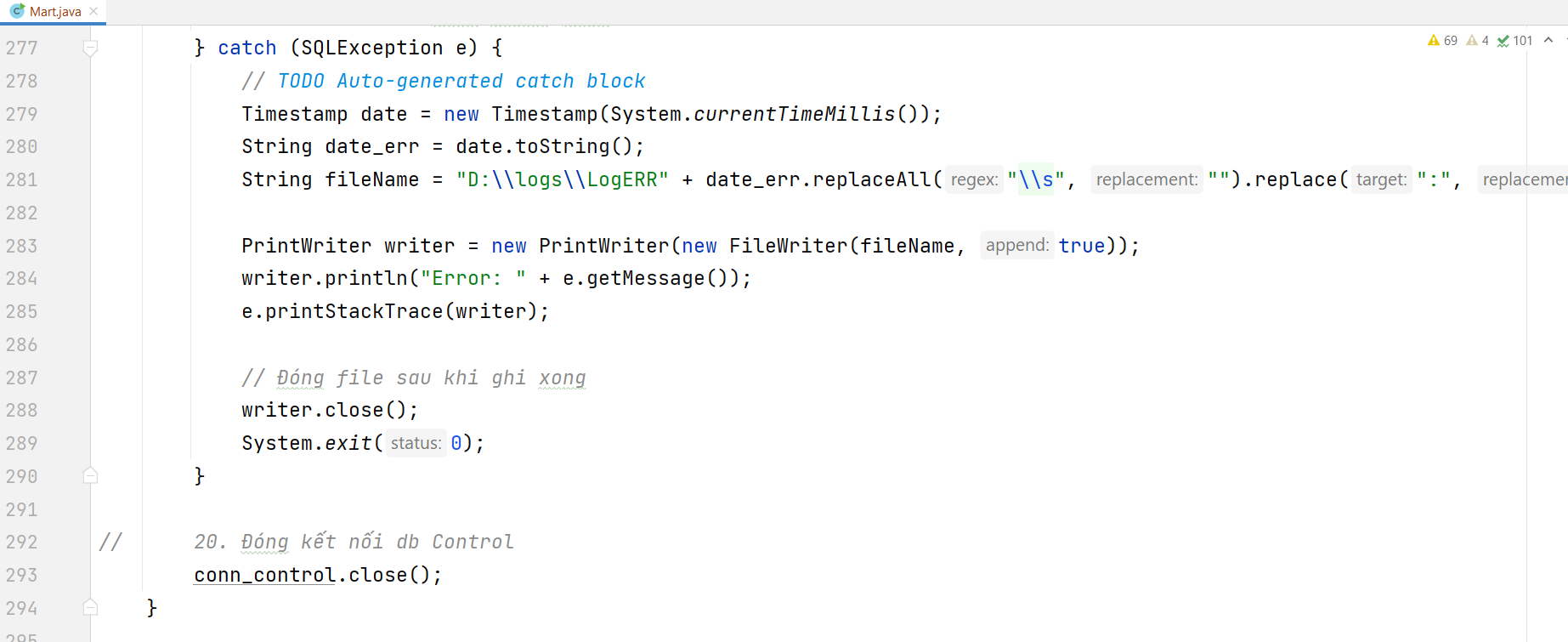




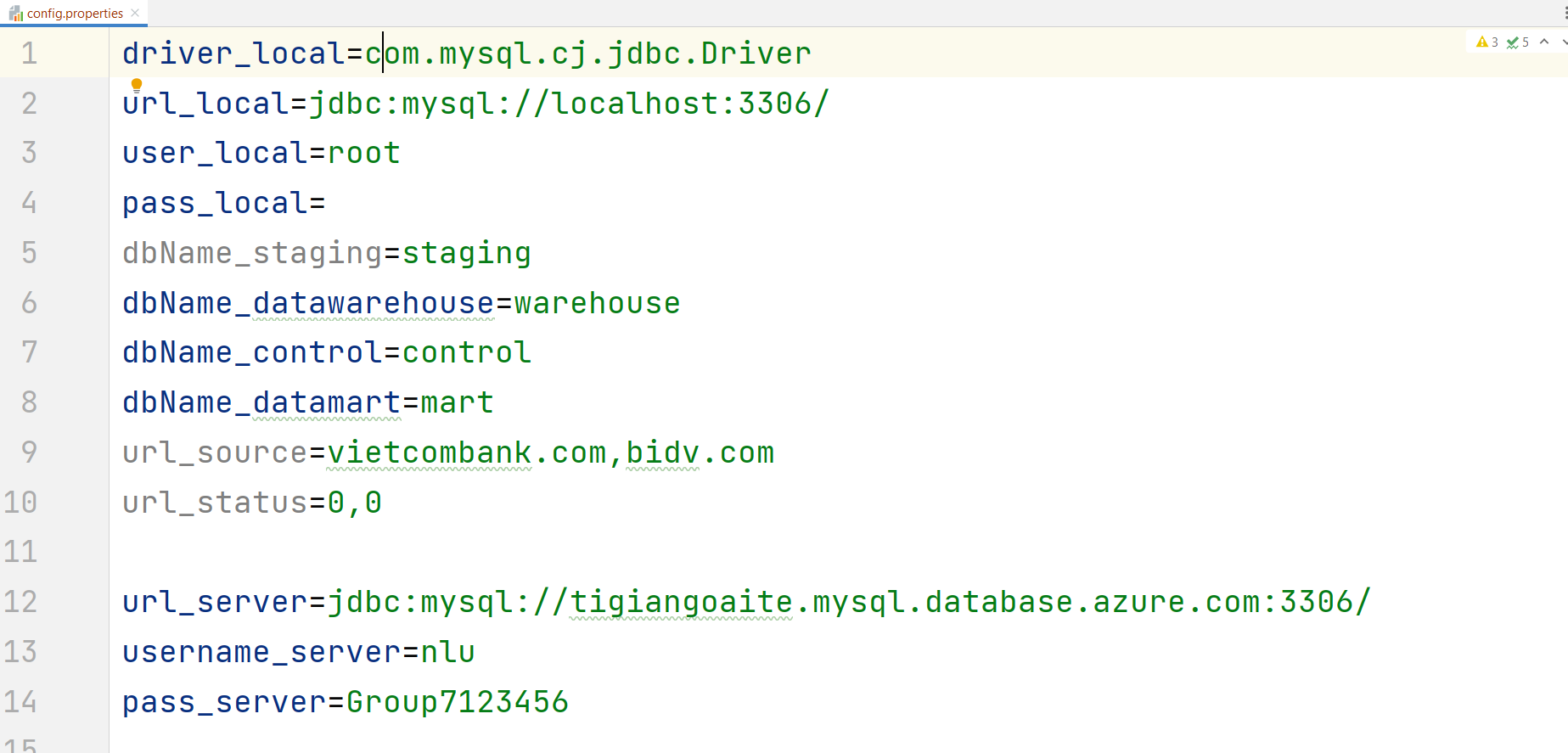




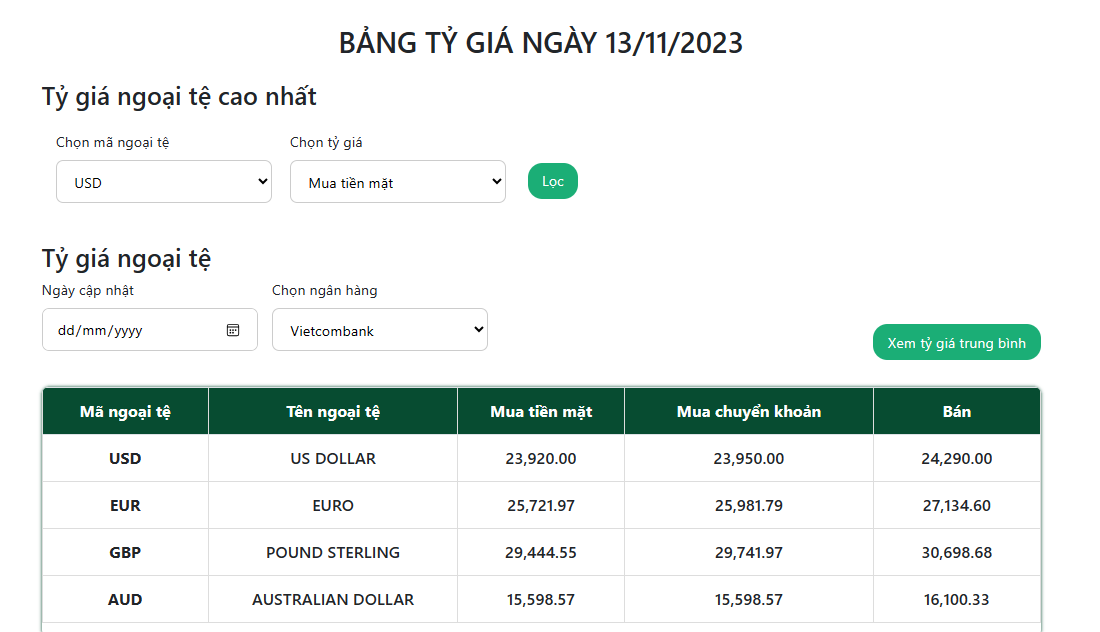




Ở phần này dùng kết nối server và local nên sẽ tương ứng với file config.properties dưới đây:



# Giao diện report demo:



# Tài liệu đính kèm:

|  |  |  |
| --- | --- | --- |
| STT | Tên tài liệu | Link |
| 1 | Trello | [Data Warehouse | Trello](https://trello.com/b/zbB3FJ7K/data-warehouse) |
| 2 | Kế hoạch nhóm | [DW\_2023\_T4\_Nhom7](https://docs.google.com/spreadsheets/d/1xL7yAa273moRsYvRbqfqzao11JwoCoKIiFm--SWFksQ/edit) |
| 3 | Data example | [Data\_example](https://docs.google.com/spreadsheets/d/1wX9RL6Der6YgKpIcZbIryGRO61wirLn79GRuk4X2-AM/edit#gid=840826166) |
| 4 | Đặc tả dữ liệu(từ sheet thứ 2 đến hết) | [DW\_2023\_T4\_Nhom7](https://docs.google.com/spreadsheets/d/1xL7yAa273moRsYvRbqfqzao11JwoCoKIiFm--SWFksQ/edit) |
| 5 | Github | [Github\_DW\_2023\_T4\_Nhom7](https://github.com/NguyenDinhNguyen0211/DW_2023_T4_Nhom7) |
| 6 | Kiến trúc (trang 1) | [draw.io](https://app.diagrams.net/?src=about#G1XdLJlheOkOdNm6JnVvdiHn8Rx7xPfkU7) |
| 7 | Workflow (trang 2-5) | [warehouse.drawio - draw.io (diagrams.net)](https://app.diagrams.net/?src=about#G1XdLJlheOkOdNm6JnVvdiHn8Rx7xPfkU7) |
| 8 | Giao diện | <https://www.figma.com/file/jyEmIxXWIX4ABf9wWKHchm/Data-warehouse?type=design&node-id=13%3A8&mode=design&t=3Cs1lwgSTPQfapvr-1> |
| 9 | Cấu hình server và đặt lịch | [CONFIGURE LOCAL & SERVER DATABASE, SETUP PROCESS LOCAL.docx - Google Tài liệu](https://docs.google.com/document/d/1f_tgK8WOedpsh0oKWDjXSzT95wcfl4Db/edit#heading=h.bn48jloggcq5) |