**BỘ GIÁO DỤC VÀ ĐẠO TẠO**

**TRƯỜNG ĐẠI HỌC NÔNG LÂM TP.HCM**

**KHOA CÔNG NGHỆ THÔNG TIN**

**----🙣🕮🙡----**

Logo

Description automatically generated

BỘ MÔN DATA WAREHOUSE

Đề tài : BÁO CÁO GIÁ VÀNG

GVHD: Ths.Nguyễn Đức Công Song.

Lớp: Thứ 3 – Ca 2.

Nhóm: 6.

*TP.HCM 08, tháng 10 năm 2024*

Mục lục

[REVISION HISTORY 3](#_Toc181549212)

[1. Kiến trúc hệ thống. 4](#_Toc181549213)

[2. Mô tả dữ liệu đầu vào (Datafeed Specification). 4](#_Toc181549214)

[2.1. Atribute Specification. 4](#_Toc181549215)

[2.2. Data/Process Controls. 6](#_Toc181549216)

[3. Cấu trúc cơ sở dữ liệu hệ thống. 6](#_Toc181549217)

[3.1. Cấu trúc cơ sở dữ liệu control. 6](#_Toc181549218)

[3.2. Cấu trúc cơ sở dữ liệu staging. 6](#_Toc181549219)

[3.3. Cấu trúc cơ sở dữ liệu data warehouse. 6](#_Toc181549220)

[3.4. Cấu trúc cơ sở dữ liệu data mart. 6](#_Toc181549221)

[4. Sơ đồ luồng thực hiện. 6](#_Toc181549222)

[4.1. Lấy data từ source về file. 6](#_Toc181549223)

[4.2. Lấy data từ file lên staging. 6](#_Toc181549224)

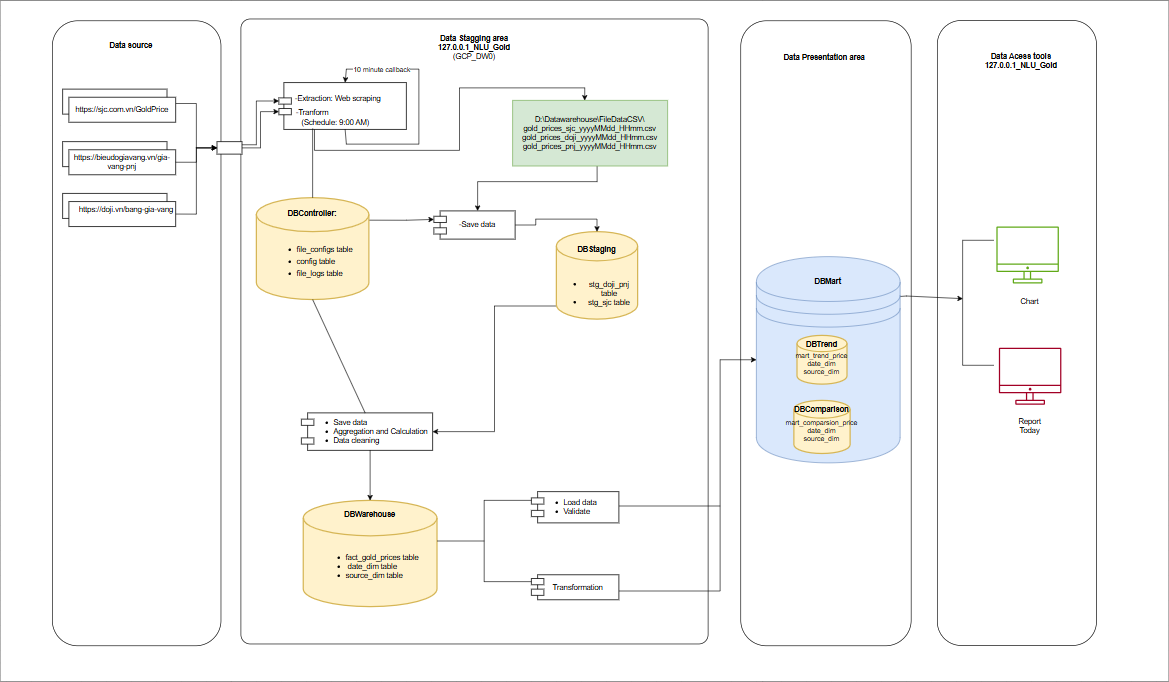
[4.3. Lấy data từ staging lên data warehouse. 6](#_Toc181549225)

[4.4. Lấy data từ data warehouse lên data mart. 6](#_Toc181549226)

REVISION HISTORY

|  |  |  |  |
| --- | --- | --- | --- |
| **Date** | **Version** | **Author** | **Changes** |
| 20/10/2024 | 1.0 | Nguyễn Trung Kiên | Initial |
| 1/11/2024 | 1.1 | Nguyễn Hoàng Ánh Dương | 1 |
| 4/11/2024 | 1.2 | Lê Quốc Trung | 2.1.1 , 2.1.2 |
| 5/11/2024 | 1.3 | Nguyễn Trung Kiên | 2.1.3, 2.1.4, 3.1, 3.2, 3.3, 3.4 |
| 9/12/2024 | 1.4 | Nguyễn Hoàng Ánh Dương | 3 |

# Kiến trúc hệ thống.



# Mô tả dữ liệu đầu vào (Datafeed Specification).

## Atribute Specification.

Phần này cung cấp các định nghĩa ngắn gọn về từng thuộc tính, bao gồm định dạng trường và tiêu chí chỉnh sửa. Để biết thêm thông tin về các thuật ngữ được sử dụng để định nghĩa thuộc tính.

Đặc điểm kỹ thuật thuộc tính dữ liệu giá vàng.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Field Name | Format | Mask | Edit Rules | Description | Example | Unit |
| id | Integer | N/A |  | Mã định danh duy nhất cho mỗi bản ghi giá vàng | 1 | N/A |
| date | Date | YYYY-MM-DD |  | Ngày cập nhật giá vàng | 2024-11-03 | N/A |
| type\_name | Varchar | N/A |  | Tên loại vàng | Vàng SJC 1L, 10L, 1KG | N/A |
| region | Varchar | N/A |  | Khu vực, thị trường áp dụng giá vàng | Hà Nội | N/A |
| buy\_price | Decimal |  |  | Giá mua vàng | 87500000 | VNĐ |
| sell\_price | Decimal |  |  | Giá bán vàng | 89500000 | VNĐ |
| source | Varchar |  |  | Nguồn lấy giá vàng | SJC |  |
| path | Varchar |  |  | Đường dẫn lấy file | D:/Datawarehouse/ |  |
| server | Text |  |  | WAP site hostname |  |  |
| ip | Text |  |  | Địa chỉ IP lưu trữ dữ liệu | 192.168.1.1 |  |

### File lưu trữ dữ liệu lấy từ nguồn.

**Data size**: ~9KB/ngày.

**Data Location**: D:/Datawarehouse/FileDataCSV/

**Data Frequency**: Hàng Ngày.

**Name Convention**: [Nguồn]\_[ngày]-[tháng]-[năm].csv.

\***Nguồn**: gold\_prices\_sjc, gold\_prices\_pnj, gold\_prices\_doji.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **No** | **Field Name** | **Type** | **Source (PNJ[[1]](#footnote-1), SJC[[2]](#footnote-2), Doji[[3]](#footnote-3))** | **Example** | **Description** |
| 1 | Id | Text | PNJ, SJC, Doji | 1 | Mã định danh cho loại vàng |
| 2 | TypeName | Text | PNJ, Doji, SJC | SJC-HN bán lẻ | Tên loại vàng, khối lượng |
| 3 | Company | Text | PNJ, Doji | Trong nước | Khu vực, thị trường áp dụng giá vàng. |
| 4 | BuyValue | Text | PNJ, SJC, Doji | 8700000 | Giá mua vào |
| 5 | SellValue | Text | PNJ, SJC, Doji | 8900000 | Giá bán ra |
| 6 | BranchName | Text | SJC | Hồ Chí Minh | Khu vực, thị trường áp dụng giá vàng |
| 7 | Buy | Text | SJC | 8700 | Giá mua vào (đơn vị ngàn) |
| 9 | Sell | Text | SJC | 8900 | Giá bán ra (đơn vị ngàn) |
| 11 | BuyDiffer | Text | SJC | null | Sự chênh lệch giá mua |
| 12 | BuyDifferValue | Text | SJC | 0 | Giá trị chênh lệch giá mua |
| 13 | SellDiffer | Text | SJC | null | Sự chênh lệch giá bán |
| 14 | SellDifferValue | Text | SJC | 0 | Giá trị chênh lệch giá bán |
| 15 | GroupDate | Text | SJC | /Date(2135596800000)/ | Ngày giờ nhóm dữ liệu. Giá trị mặc định khi không có dữ liệu là ngày 1/1/0001. |

File Format:

* **Dạng tệp tin:** Dữ liệu nguồn được lưu dưới 2 dạng tệp tin là CSV và XLSX.
* **Phân cách dữ liệu:** Các giá trị dữ liệu được phân cách bằng dấu phẩy (,) giữa các trường dữ liệu.
* **Hàng đầu tiên là header:** Thông thường, hàng đầu tiên trong tệp tin CSV được sử dụng để định rõ tên của các trường dữ liệu (column headers). Điều này giúp xác định ý nghĩa của mỗi trường dữ liệu.

### Staging.

Input: D:/Datawarehouse/FileDataCSV/[Nguồn]\_[ngày]-[tháng]-[năm].csv.

Naming Convention: stg

Delivery Schedule: hàng ngày.

staging schema: stg\_doji\_pnj, stg\_sjc.

**stg\_doji\_pnj:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Field Name** | **Format** | **Length** | **Mask** | **Edit Rules** | **Sample Value** |
| Id | auto |  |  |  | 1 |
| TypeName | varchar | 255 |  |  | SJC HN – Bán lẻ |
| Company | varchar | 100 |  |  | Trong nước |
| BuyValue | decimal | 10,2 |  |  | 87000000,00 |
| SellValue | decimal | 10,2 |  |  | 89000000,00 |
| Date | datetime |  |  |  | 2024-11-03 |
| Source | varchar | 10 |  |  | PNJ |

**stg\_sjc:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Field Name** | **Format** | **Length** | **Mask** | **Edit Rules** | **Sample Value** |
| Id | auto |  |  |  | 1 |
| TypeName | varchar | 255 |  |  | Vàng SJC 1L, 10L, 1KG |
| BranchName | varchar | 100 |  |  | Trong nước |
| BuyValue | decimal | 10,2 |  |  | 87000000,00 |
| BuyDifferValue | decimal | 10,2 |  |  | 0 |
| SellValue | decimal | 10,2 |  |  | 89000000,00 |
| SellDifferValue | decimal | 10,2 |  |  | 0 |
| GroupDate | datetime |  |  |  | 2024-11-03 |
| Source | varchar | 10 |  |  | PNJ |

### Data Warehouse

**Input**: stg\_doji\_pnj.db, stg\_sjc.db

**Naming Convention:** datawarehouse

**Delivery Schedule:** Hàng ngày

**Data warehouse schema:** fact\_gold\_prices

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Field Name** | **Format** | **Length** | **Mask** | **Edit Rules** | **Sample Value** |
| id | auto |  |  |  | 1 |
| id\_date | auto |  |  |  | id của ngày trong bảng datamart.date\_dim |
| id\_source | int |  |  |  | id của tên nguồn trong bảng datamart.source\_dim |
| type\_name | varchar | 255 |  |  | SJC HN - Bán lẻ |
| region | varchar | 100 |  |  | Trong nước |
| buy\_value | decimal | 10,2 | VND |  | 87000000,00 |
| sell\_value | decimal | 10,2 | VND |  | 89000000,00 |
| comparison\_buy | int |  |  |  | -1: giảm, 0: không đổi, 1: tăng |
| comparison\_sell | int |  |  |  | -1: giảm, 0: không đổi, 1: tăng |
| percentage\_buy | decimal | 5,2 | % |  | 2 |
| percentage\_sell | datetime | 5,2 | % |  | 1 |

**datawarehouse schema: date\_dim**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Field Name** | **Format** | **Length** | **Mask** | **Edit Rules** | **Sample Value** |
| id | auto |  |  |  | 1 |
| full\_date | Date |  |  |  | 2024-11-05 |
| day\_of\_week | varchar | 20 |  |  | Tuesday |
| month | varchar | 20 |  |  | 11 |
| year | varchar | 4 |  |  | 2024 |

**datawarehouse schema: source\_dim**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Field Name** | **Format** | **Length** | **Mask** | **Edit Rules** | **Sample Value** |
| id | auto |  |  |  | 1 |
| source | varchar | 10 |  |  | SJC |

### Datamart

**Input**: datawarehouse.db

**Naming convention:** datamart

**Delivery Schedule:** Hàng ngày

**Datamart schema:** mart\_trend\_price, mart \_comparison\_price.

**Datamart schema: mart\_trend\_price:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Field Name** | **Format** | **Length** | **Mask** | **Edit Rules** | **Sample Value** |
| id | auto |  |  |  | 1 |
| id\_date | auto |  |  |  | id của ngày trong bảng datamart.date\_dim |
| id\_source | int |  |  |  | id của tên nguồn trong bảng datamart.source\_dim |
| type\_name | varchar | 255 |  |  | SJC HN - Bán lẻ |
| region | varchar | 100 |  |  | Trong nước |
| buy\_value | decimal | 10,2 | VND |  | 87000000,00 |
| sell\_value | decimal | 10,2 | VND |  | 89000000,00 |

date\_dim

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Field Name** | **Format** | **Length** | **Mask** | **Edit Rules** | **Sample Value** |
| id | auto |  |  |  | 1 |
| full\_date | Date |  |  |  | 2024-11-05 |
| day\_of\_week | varchar | 20 |  |  | Tuesday |
| month | varchar | 20 |  |  | 11 |
| year | varchar | 4 |  |  | 2024 |

source\_dim

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Field Name** | **Format** | **Length** | **Mask** | **Edit Rules** | **Sample Value** |
| id | auto |  |  |  | 1 |
| source | varchar | 10 |  |  | SJC |

**Datamart schema: mart\_comparison\_price:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Field Name** | **Format** | **Length** | **Mask** | **Edit Rules** | **Sample Value** |
| id | auto |  |  |  | 1 |
| id\_date | auto |  |  |  | id của ngày trong bảng datamart.date\_dim |
| id\_source | int |  |  |  | id của tên nguồn trong bảng datamart.source\_dim |
| type\_name | varchar | 255 |  |  | SJC HN - Bán lẻ |
| region | varchar | 100 |  |  | Trong nước |
| buy\_value | decimal | 10,2 | VND |  | 87000000,00 |
| sell\_value | decimal | 10,2 | VND |  | 89000000,00 |
| comparison\_buy | int |  |  |  | -1: giảm, 0: không đổi, 1: tăng |
| comparison\_sell | int |  |  |  | -1: giảm, 0: không đổi, 1: tăng |
| percentage\_buy | decimal | 5,2 | % |  | 2 |
| percentage\_sell | datetime | 5,2 | % |  | 1 |

date\_dim

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Field Name** | **Format** | **Length** | **Mask** | **Edit Rules** | **Sample Value** |
| id | auto |  |  |  | 1 |
| full\_date | Date |  |  |  | 2024-11-05 |

source\_dim

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Field Name** | **Format** | **Length** | **Mask** | **Edit Rules** | **Sample Value** |
| id | auto |  |  |  | 1 |
| source | varchar | 10 |  |  | SJC |

## Data/Process Controls.

Kiểm soát dữ liệu và quy trình được đưa ra để đảm bảo rằng nguồn cấp dữ liệu giá vàng là chính xác, đầy đủ và kịp thời. Các điều khiển chính là:

| **Control** | **Description** |
| --- | --- |
| Start | Bắt đầu thực hiện 1 tiến trình |
| Success | Thực hiện thành công 1 tiến trình |
| Run | Đang thực hiện 1 tiến trình |
| Fail | Thất bại khi thực hiện 1 sự kiện |
| Can not run | Không thể chạy 1 tiến trình |
| Should be after 17h | Chạy 1 tiến trình sau 17h |
| Data was got | Dữ liệu đã được lấy |
| Data already exists | Dữ liệu đã tồn tại |

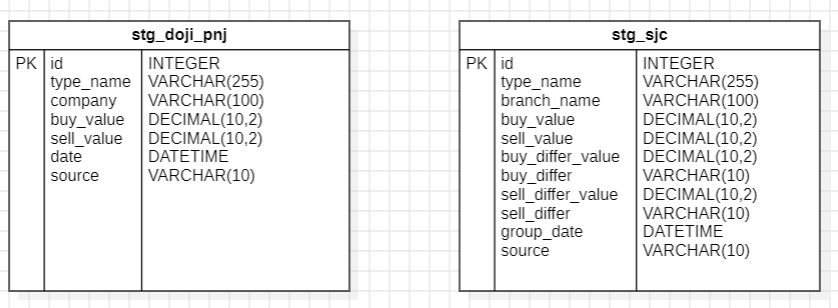
# Cấu trúc cơ sở dữ liệu hệ thống.

## Cấu trúc cơ sở dữ liệu control.

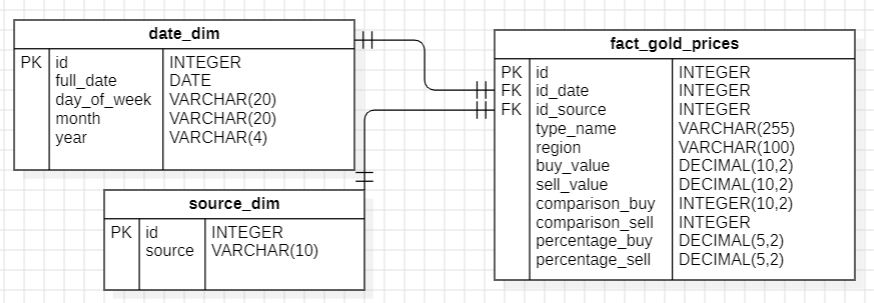
Ảnh có chứa văn bản, ảnh chụp màn hình, số, Phông chữ

Mô tả được tạo tự động

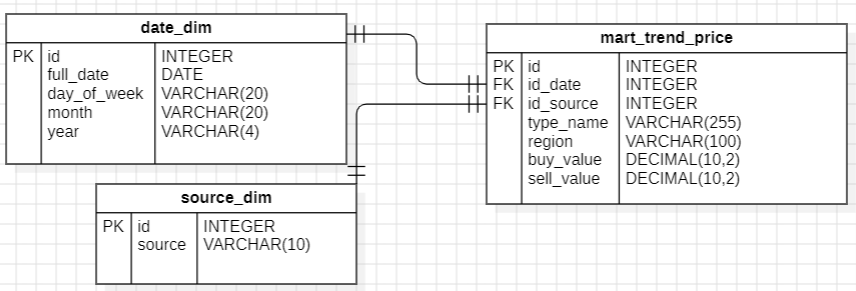
## Cấu trúc cơ sở dữ liệu staging.

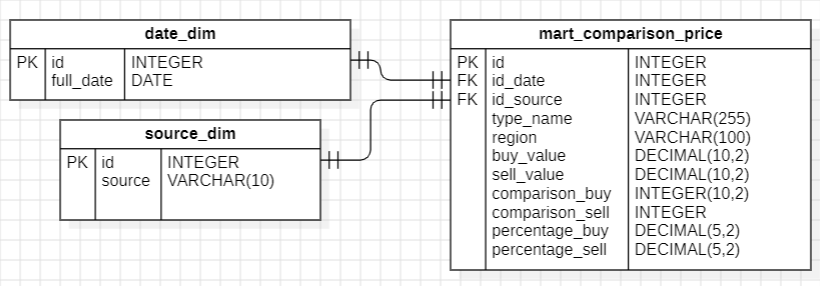


## Cấu trúc cơ sở dữ liệu data warehouse.



## Cấu trúc cơ sở dữ liệu data mart.





# Sơ đồ luồng thực hiện.

## Lấy data từ source về file.

## Lấy data từ file lên staging.

## Lấy data từ staging lên data warehouse.

## Lấy data từ data warehouse lên data mart.

1. PNJ: https://bieudogiavang.vn/gia-vang-pnj [↑](#footnote-ref-1)
2. SJC: https://sjc.com.vn/GoldPrice/Services/PriceService.ashx [↑](#footnote-ref-2)
3. Doji: https://doji.vn/bang-gia-vang/ [↑](#footnote-ref-3)