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

**KHOA CÔNG NGHỆ THÔNG TIN**

****

**Môn học: Data Warehouse**

**Xây Dựng Hệ Thống Data Warehouse Báo Cáo Giá Sản Phẩm Tủ Lạnh**

GV hướng dẫn: ThS.Nguyễn Đức Công Song

Nhóm thực hiện: Nhóm 5

Thành viên:

21130279 – Trần Quí Bằng

20130257 – Nguyễn Thị Hiền

21130382 - Nguyễn Văn Huy

21130410 – Đặng Tấn Kiệt

**Mục Lục**

[1 Kiến trúc hệ thống 3](#_Toc180403401)

[2 Mô tả dữ liệu đầu vào (Datafeed Specification) 3](#_Toc180403402)

[2.1 Nguồn dữ liệu đầu vào 3](#_Toc180403403)

[2.2 Đặc điểm nguồn cung cấp (Feed Characteristics) 4](#_Toc180403404)

[3 Cở sở dữ liệu 5](#_Toc180403405)

[3.1 Database Control 5](#_Toc180403406)

[3.2 Data Staging 6](#_Toc180403407)

[3.3 Database Data Warehouse 8](#_Toc180403408)

[3.4 Data Mart 10](#_Toc180403409)

[4 Sơ đồ luồng thực hiện công việc 11](#_Toc180403410)

[4.1 Sơ đồ luồng thực hiện việc lấy dữ liệu vào hệ thống. 11](#_Toc180403411)

**Reversion History**

|  |  |  |  |
| --- | --- | --- | --- |
| Ngày thực hiện | Ngày hoàn thành | Mô tả | Thực hiện |
|  |  |  |  |
|  |  |  |  |
| 7/10/2024 | 19/10/2024 | Thiết kế kiến trúc hệ thống | Bằng |
| 7/10/2024 | 19/10/2024 | Mô tả dữ liệu đầu vào | Hiền |
| 7/10/2024 | 19/10/2024 | Tạo file ví dụ dữ liệu đầu vào | Hiền |
| 7/10/2024 | 19/10/2024 | Xây dựng cơ sở dữ liệu cho hệ thống | Huy |
| 7/10/2024 | 19/10/2024 | Thiết kế sơ đồ luồng thực hiện việc lấy dữ liệu vào hệ thống | Kiệt |

# Kiến trúc hệ thống

A diagram of a diagram

Description automatically generated

Hình 1.1 [Kiến trúc hệ thống Data Warehouse báo cáo giá tủ lạnh](https://app.diagrams.net/#G1nIijAAs55gAscq6kSZzddv7brRIlMX9j)

# Mô tả dữ liệu đầu vào (Datafeed Specification)

## Nguồn dữ liệu đầu vào

* Nguồn dữ liệu từ website: **https:dienmayxanh.com/tu-lanh**

Các thuộc tính của dữ liệu và element HTML cần lấy:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **No** | **Field Name** | **Format** | **Edit Rules** | **Description** | **Specification** | **Element** |
| 1 | name | text |  | Tên sản phẩm | Tên sản phẩm cung cấp mã model và thương hiệu của tủ lạnh | (“h1.product-name”).text(); |
| 2 | image | text |  | Hình ảnh minh họa | Đường dẫn URL của hình ảnh sản phẩm, giúp người dùng xem mẫu mã. | ("picture > img").attr("src"); |
| 3 | brand\_name | text |  | Tên thương hiệu | Tên thương hiệu của sản phẩm tủ lạnh (ví dụ: LG, Samsung, Toshiba, v.v.). | (“ul.text-specifi active>li”).text(); |
| 4 | description | text |  | Mô tả thông tin sản phẩm | Mô tả chi tiết về các tính năng nổi bật của tủ lạnh gồm:Tổng quan;Mức tiêu thụ điện năng;Công nghệ bảo quản và làm lạnh;Tiện ích;Thông tin lắp đặt.. | (“div.specification-item-236996”).text(); |
| 5 | price | Decimal |  | Giá bán | Giá bán gốc của sản phẩm trước khi có giảm giá hoặc khuyến mãi. | (“div.price-one   >p.box-price-old”); |
| 6 | price\_sale | Decimal |  | Giá khuyến mãi | Giá bán sau khi đã giảm trong các đợt khuyến mãi. | (“div.price-one >p.box-price-present”); |
| 7 | discount | Decimal |  | Phần trăm giảm giá | Phần trăm giảm giá so với giá gốc. | (“div.price-one >p.box-price-percent”); |
| 8 | Payment priority | Decimal |  | Ưu đãi thanh toán online | Ưu đải giảm thêm  khi thanh toán qua thẻ ngân hàng online | (“div.sliderX”); |

## Đặc điểm nguồn cung cấp (Feed Characteristics)

* Log file: C:\Users\DW\data\dienmayxanh\_dd\_MM\_YYYY.json
* Frequency: 1 ngày/ 1 lần
* Content: Chứa dữ liệu là thông tin các sản phẩm tủ lạnh.
* Estimated Table size:
* Pre-processing: extract zip
* Estimated monthly size req:

# Cở sở dữ liệu

## Database Control

A screenshot of a computer

Description automatically generated

Hình 3.1. Diagram Database Control

Table file\_logs

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| No | Field Name | Format | Edit Rules | Description | Example |
| 1 | log\_id | INT |  | id của bản ghi log | 1 |
| 2 | config\_id | INT |  | id file cấu hình | 1 |
| 3 | status | VARCHAR(50) |  | Trạng thái quá trình | success |
| s | timestamp | DATETIME |  | Thời gian ghi lại trạng thái | 2024-10-16 10:05:00 |
| 5 | error\_message | TEXT |  | Thông tin chi tiết về lỗi (nếu có) | NULL |

Table file\_config

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| No | Field Name | Format | Edit Rules | Description | Example |
| 1 | config\_id | INT |  | id cấu hình | 1 |
| 2 | file\_name | VARCHAR(255) |  | Tên file nguồn dữ liệu | https://dienmayxanh.com/tu-lanh |
| 3 | last\_extracted | DATETIME |  | Thời gian lần cuối lấy dữ liệu | 2024-10-16 10:00:00 |
| 4 | schedule\_id | INT |  | id lịch trình lấy dữ liệu | 1 |

Table schedule

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| No | Field Name | Format | Edit Rules | Description | Example |
| 1 | schedule\_id | INT |  | id lịch trình | 1 |
| 2 | frequency | VARCHAR(50) |  | Tần suất lấy dữ liệu | daily |
| 3 | next\_run | DATETIME |  | Thời gian lấy dữ liệu lần tiếp theo | 2024-10-18 10:00:00 |

## Data Staging

A screenshot of a computer

Description automatically generated

Hình 3.2. Diagram Database Staging

Table staging\_fridges

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| No | Field Name | Format | Edit Rules | Description | Example |
| 1 | fridge\_id | INT |  | id của tủ lạnh | 1 |
| 2 | fridge\_name | VARCHAR(255) |  | Tên tủ lạnh | Tủ lạnh Samsung Inverter 208 lít |
| 3 | brand\_name | VARCHAR(100) |  | Tên thương hiệu tủ lạnh | Samsung |
| 4 | image\_url | VARCHAR(500) |  | url ảnh tủ lạnh | https://cdn.dienmayxanh.com/tu-lanh-samsung.jpg |
| 5 | price | DECIMAL(10, 2) |  | Giá bán hiện tại của tủ lạnh | 5990000 |
| 6 | original\_price | DECIMAL(10, 2) |  | Giá gốc | 6990000 |
| 7 | discount\_percentage | DECIMAL(5, 2) |  | Phần trăm giảm giá | |  | | --- | |  |  |  | | --- | | 14.29 | |
| 8 | description | TEXT |  | Mô tả của tủ lạnh | |  | | --- | |  |  |  | | --- | | Tủ lạnh dung tích lớn, tiết kiệm điện. | |

Table fridges

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| No | Field Name | Format | Edit Rules | Description | Example |
| 1 | fridge\_id | INT |  | id của tủ lạnh | 1 |
| 2 | fridge\_name | VARCHAR(255) |  | Tên tủ lạnh | Tủ lạnh Samsung Inverter 208 lít |
| 3 | brand\_id | INT |  | id của thương hiệu tủ lạnh | 1 |
| 4 | image\_url | VARCHAR(500) |  | url ảnh tủ lạnh | https://cdn.dienmayxanh.com/tu-lanh-samsung.jpg |
| 5 | price | DECIMAL(10, 2) |  | Giá bán hiện tại của tủ lạnh | 5990000 |
| 6 | original\_price | DECIMAL(10, 2) |  | Giá gốc | |  | | --- | |  |  |  | | --- | | 6990000 | |
| 7 | discount\_percentage | DECIMAL(5, 2) |  | Phần trăm giảm giá | |  | | --- | |  |  |  | | --- | | 14.29 | |
| 8 | description | TEXT |  | Mô tả của tủ lạnh | Tủ lạnh dung tích lớn, tiết kiệm điện. |

Table brands

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| No | Field Name | Format | Edit Rules | Description | Example |
| 1 | brand\_id | INT |  | id của tủ lạnh | 1 |
| 2 | brand\_name | VARCHAR(100) |  | Tên tủ lạnh | Samsung |

## Database Data Warehouse

A screenshot of a data warehouse

Description automatically generated

Hình 3.3. Diagram Database Data Warehouse

Table dim\_product

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| No | Field Name | Format | Edit Rules | Description | Example |
| 1 | product\_id | INT |  | id của tủ lạnh | 1 |
| 2 | product\_name | VARCHAR(255) |  | Tên tủ lạnh | Tủ lạnh Samsung Inverter 208 lít |
| 3 | brand\_id | INT |  | id của thương hiệu tủ lạnh | |  | | --- | |  |  |  | | --- | | 1 | |
| 4 | image\_url | VARCHAR(500) |  | url ảnh tủ lạnh | https://cdn.dienmayxanh.com/tu-lanh-samsung.jpg |
| 5 | description | TEXT |  | Mô tả của tủ lạnh | Tủ lạnh dung tích lớn, tiết kiệm điện. |

Table dim\_time

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| No | Field Name | Format | Edit Rules | Description | Example |
| 1 | time\_id | INT |  | id thời gian | 1 |
| 2 | date | DATE |  | Ngày cụ thể | 2024-10-16 |
| 3 | month | INT |  | Tháng | 10 |
| 4 | year | INT |  | Năm | 2024 |

Table dim\_brand

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| No | Field Name | Format | Edit Rules | Description | Example |
| 1 | brand\_id | INT |  | id thương hiệu | 1 |
| 2 | brand\_name | VARCHAR(100) |  | Tên thương hiệu | Samsung |

Table dim\_price

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| No | Field Name | Format | Edit Rules | Description | Example |
| 1 | fact\_id | INT |  | id thương hiệu | 1 |
| 2 | product\_id | VARCHAR(100) |  | id thương hiệu | 1 |
| 3 | time\_id | INT |  | id thời gian | 1 |
| 4 | current\_price | DECIMAL(10, 2) |  | Giá hiện tại | 5990000 |
| 5 | original\_price | DECIMAL(10, 2) |  | Giá gốc | 6990000 |
| 6 | discount\_percentage | DECIMAL(5, 2) |  | Phần trăm giảm giá | 14.29 |

## Data Mart

A screenshot of a computer

Description automatically generated

Hình 3.4. Diagram Database Data Mart

Table present\_price

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| No | Field Name | Format | Edit Rules | Description | Example |
| 1 | product\_id | INT |  | id của tủ lạnh | 1 |
| 2 | product\_name | VARCHAR(255) |  | Tên tủ lạnh | Tủ lạnh Samsung Inverter 208 lít |
| 3 | brand\_name | INT |  | Tên thương hiệu tủ lạnh | Samsung |
| 4 | Current\_price | DECIMAL(10, 2) |  | Giá hiện tại | 5990000 |
| 5 | Discount\_percentage | DECIMAL(5, 2) |  | Phần trăm giảm giá | 14.29 |
| 6 | Last\_updated | DATETIME |  | Ngày cập nhật giá gần nhất | 2024-10-16 10:00:00 |

# Sơ đồ luồng thực hiện công việc

## Sơ đồ luồng thực hiện việc lấy dữ liệu vào hệ thống.

A diagram of a company

Description automatically generated

Hình 4.1. Sơ đồ luồng việc thực hiện lấy dữ liệu vào hệ thống