**BỘ GIÁO DỤC VÀ ĐÀO TẠO**

**TRƯỜNG ĐẠI HỌC NÔNG LÂM TPHCM**

**KHOA CÔNG NGHỆ THÔNG TIN**



**Môn: Data Warehouse**

**GVHD: Th.S Nguyễn Đức Công Song**

**Đề tài: Báo cáo giá xe đạp**

**Sinh viên thực hiện:**

21130558 – Mai Xuân Thức

20130238 – Hoàng Nhựt Duy

21130340 – Lê Ngọc Hà

21130448 – Cao Thành Nam

21130168 – Võ Vũ Minh Tâm

**Mục lục**

[**Lời mở đầu** 2](#_Toc182426804)

[**LỜI CÁM ƠN** 3](#_Toc182426805)

[**NHẬN XÉT** 4](#_Toc182426806)

[**Revision History** 5](#_Toc182426807)

[**I.** **KIẾN TRÚC HỆ THỐNG** 5](#_Toc182426808)

[**II.** **MÔ TẢ DỮ LIỆU ĐẦU VÀO (DATAFEED SPECIFICATION)** 6](#_Toc182426809)

[**1.** **MÔ TẢ DỮ LIỆU ĐẦU VÀO** 6](#_Toc182426810)

[**III.** **CẤU TRÚC CSDL CONTROL** 6](#_Toc182426811)

[**1.** **CẤU TRÚC CƠ BẢN, KIỂU DỮ LIỆU** 6](#_Toc182426812)

[**a.** **Config** 6](#_Toc182426813)

[**b.** **logs** 6](#_Toc182426814)

[**c.** **Status** 7](#_Toc182426815)

[**d.** **dateDim** 7](#_Toc182426816)

[**e.** **monthDim** 8](#_Toc182426817)

[**2.** **CLASS DIAGRAM** 8](#_Toc182426818)

[**IV.** **CẤU TRÚC CSDL STAGING** 9](#_Toc182426819)

[**1.** **CẤU TRÚC CƠ BẢN, KIỂU DỮ LIỆU** 9](#_Toc182426820)

[**a.** **bikes** 9](#_Toc182426821)

[**2.** **CLASS DIAGRAM** 10](#_Toc182426822)

[**V.** **CẤU TRÚC CSDL CỦA DATAWAREHOUSE** 10](#_Toc182426823)

[**1.** **CẤU TRÚC CƠ BẢN, KIỂU DỮ LIỆU** 10](#_Toc182426824)

[**f.** **Brand** 10](#_Toc182426825)

[**g.** **Bicycle** 10](#_Toc182426826)

[**h.** **Category** 11](#_Toc182426827)

[**i.** **BicycleDetails** 11](#_Toc182426828)

[**2.** **CLASS DIAGRAM** 11](#_Toc182426829)

[**3.** **SƠ ĐỒ LUỒNG THỰC HIỆN VIỆC LẤY DỮ LIỆU VÀO HỆ THỐNG** 12](#_Toc182426830)

# **Lời mở đầu**

Trong thời đại công nghệ số ngày nay, việc quản lý và phân tích dữ liệu đóng vai trò quan trọng trong việc ra quyết định chiến lược cho các doanh nghiệp. Đặc biệt trong ngành bán lẻ, dữ liệu về sản phẩm, giá cả, và hành vi người tiêu dùng trở thành nguồn tài nguyên quý giá giúp nâng cao hiệu quả kinh doanh. Đề tài “Báo cáo giá xe đạp” ra đời nhằm giải quyết nhu cầu lưu trữ, quản lý, và phân tích dữ liệu giá cả từ các nguồn khác nhau, hỗ trợ doanh nghiệp trong việc cung cấp thông tin chính xác và kịp thời. Kho dữ liệu (Data Warehouse) không chỉ giúp lưu trữ dữ liệu từ nhiều nguồn mà còn cung cấp các công cụ phân tích mạnh mẽ để dự đoán xu hướng giá, quản lý hàng tồn kho, và tối ưu hóa chiến lược kinh doanh. Bằng cách tổng hợp dữ liệu giá từ các website bán xe đạp, hệ thống sẽ cho phép doanh nghiệp theo dõi biến động giá cả, so sánh giữa các thương hiệu, và phát hiện những cơ hội kinh doanh tiềm năng. Báo cáo này sẽ trình bày quá trình xây dựng hệ thống Data Warehouse cho dữ liệu giá xe đạp, bao gồm các bước thu thập dữ liệu, thiết kế hệ thống kho dữ liệu, và ứng dụng phân tích nhằm hỗ trợ doanh nghiệp ra quyết định dựa trên dữ liệu. Thông qua nghiên cứu này, chúng tôi kỳ vọng sẽ giúp các doanh nghiệp bán lẻ xe đạp không chỉ nâng cao năng lực quản lý dữ liệu mà còn phát triển chiến lược cạnh tranh hiệu quả trong thị trường.

**LỜI CÁM ƠN**

Chúng em xin chân thành cảm ơn Thầy đã tận tình hướng dẫn và truyền đạt những kiến thức quý báu trong suốt khóa học Data Warehouse. Nhờ sự chỉ dẫn tận tâm của Thầy, chúng em đã có cơ hội tiếp cận với những khái niệm và kỹ thuật mới mẻ, cũng như thực hành và phát triển các dự án thực tế. Khóa học này đã trang bị cho chúng em không chỉ kiến thức chuyên môn mà còn cả những kỹ năng quan trọng trong công việc và cuộc sống. Những bài học và kinh nghiệm quý báu mà chúng em đã nhận được sẽ là hành trang vững chắc trên con đường phát triển sự nghiệp trong lĩnh vực công nghệ thông tin. Song cũng không tránh khỏi những thiếu sót, mong thầy sẽ tận tình chỉ bảo thêm cho chúng em.

**NHẬN XÉT**

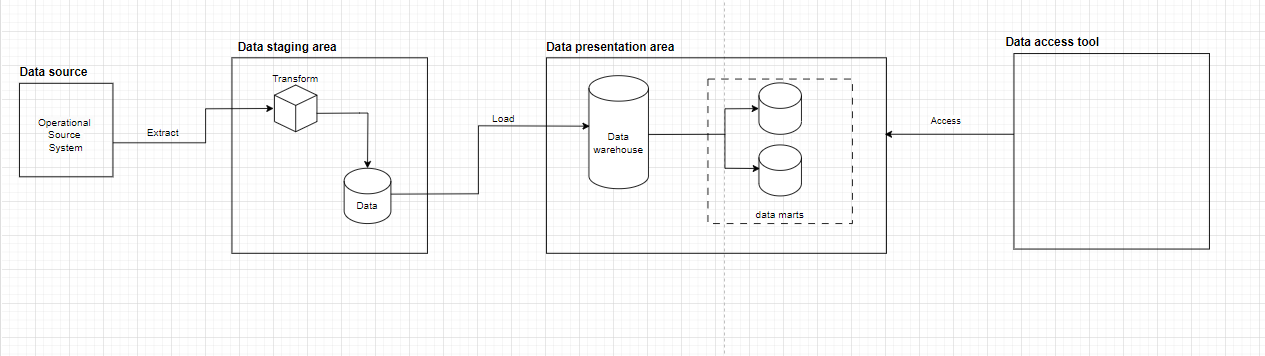
Ngày 02 tháng 06 năm 2024

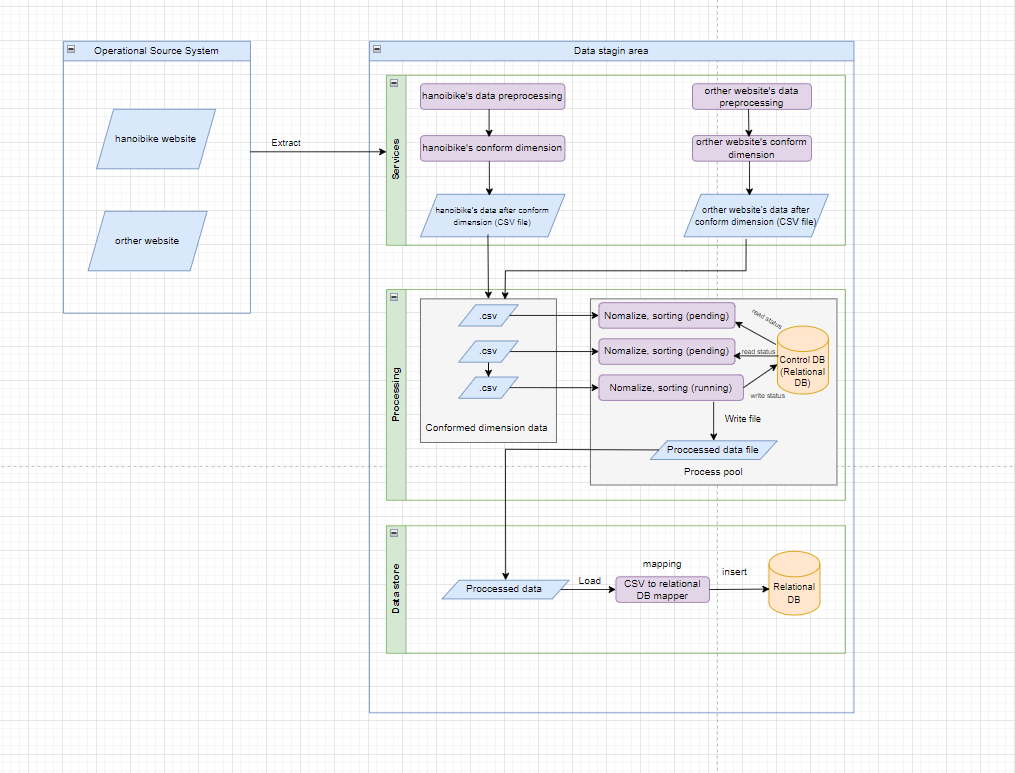
(**Ký tên**)

**Revision History**

|  |  |  |
| --- | --- | --- |
| **Ngày** | **Tác giả** | **Nội dung** |
| 13/11/2024 | C.T.Nam | Sửa lại data feed, thêm mô tả cho CSDL staging và control |

# **KIẾN TRÚC HỆ THỐNG**





# **MÔ TẢ DỮ LIỆU ĐẦU VÀO (DATAFEED SPECIFICATION)**

## **MÔ TẢ DỮ LIỆU ĐẦU VÀO**

|  |  |
| --- | --- |
| **Trường** | **Kiểu dữ liệu** |
| id | String |
| name | String |
| price | String |
| price\_sale | String |
| brand | String |
| color | String |
| size | String |
| status | String |
| description\_part1 | String |
| description\_part2 | String |
| description\_part3 | String |
| timeStartScrape | LocalDateTime |
| timeEndScrape | LocalDateTime |
| timeStartInsert | LocalDateTime |

# **CẤU TRÚC CSDL CONTROL**

## **CẤU TRÚC CƠ BẢN, KIỂU DỮ LIỆU**

### **Config**

|  |  |  |
| --- | --- | --- |
| **Trường** | **Kiểu dữ liệu** | **Mô tả** |
| id | INT(6) PRIMARY KEY AUTO\_INCREMENT | Định danh duy nhất cho từng website. |
| website | VARCHAR(255) | Tên của trang web. |
| saveFolder | VARCHAR(255) | Đường dẫn đến thư mục lưu trữ sau khi lấy dữ liệu. |
| keyword | VARCHAR(255) | Từ khóa của trang web để hỗ trợ quá trình tìm kiếm dữ liệu. |
| scrapeTimes | INT(6) | Số lần đã lấy dữ liệu từ trang web. |

### **logs**

|  |  |  |
| --- | --- | --- |
| **Trường** | **Kiểu dữ liệu** | **Mô tả** |
| id | INT(6) PRIMARY KEY AUTO\_INCREMENT | Định danh duy nhất cho từng log. |
| message | VARCHAR(255) | Ghi chú hoặc mô tả cho log. |
| quantity | INT(6) | Số lượng dữ liệu đã lấy. |
| statusId | INT(6) | Trạng thái của log (có thể tham chiếu đến bảng trạng thái nếu cần). |
| websiteId | INT(6) | ID của website, liên kết với bảng "websites". |

|  |  |  |
| --- | --- | --- |
| timeStart | DATETIME | Thời gian bắt đầu truy cập vào trang web. |
| timeEnd | DATETIME | Thời gian kết thúc truy cập vào trang web. |

### **Status**

|  |  |  |
| --- | --- | --- |
| **Trường** | **Kiểu dữ liệu** | **Mô tả** |
| id | INT(6) PRIMARY KEY AUTO\_INCREMENT | Định danh duy nhất cho từng trạng thái. |
| name | VARCHAR(255) | Tên của trạng thái (ví dụ: active, inactive).  **RUNNING**: hệ thống đang trong quá trình lấy dữ liệu  **WAITING**: thể hiện trong ngày hôm nay website này chưa được lấy dữ liệu và đang chờ để lấy dữ liệu  **FAILED**: thể hiện nếu số lượng dòng lấy dữ liệu bằng không  **COMPLETE**: đã hoàn thành lấy dữ liệu thành công |

### **dateDim**

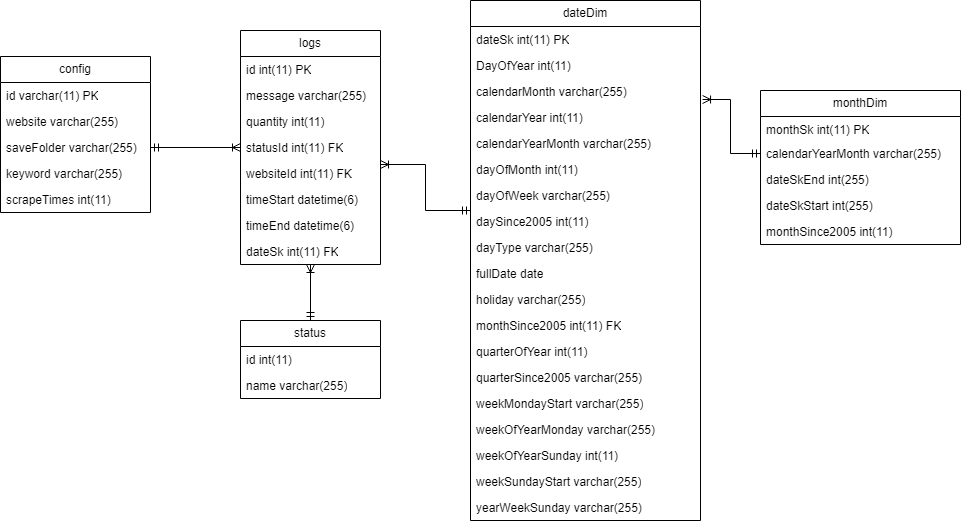
|  |  |  |
| --- | --- | --- |
| **Trường** | **Kiểu dữ liệu** | **Mô tả** |
| dateSk | INT(6) PRIMARY KEY | Khóa chính, duy nhất định danh từng ngày. |
| DayOfYear | INT(6) | Ngày thứ i trong năm (1, 2, 3,...). |
| calendarMonth | VARCHAR(255) | Tháng được lưu trữ ở dạng chữ (January, February,...). |
| calendarYear | YEAR | Năm (2005, 2006, 2007,...). |
| calendarYearMonth | VARCHAR(255) | Năm và tháng (2005-Jan, 2005-Feb,...). |
| dayOfMonth | INT(6) | Ngày thứ i trong tháng (1, 2, 3,...). |
| dayOfWeek | VARCHAR(255) | Thứ trong tuần lưu ở dạng chữ (Monday, Tuesday,...). |
| daySince2005 | INT(6) | Ngày thứ i tính từ ngày 01/01/2005. |
| dayType | VARCHAR(255) | Ngày trong tuần hay cuối tuần (weekday, weekend). |
| fullDate | DATE | Ngày tháng năm đầy đủ (YYYY/mm/dd). |
| holiday | VARCHAR(255) | Ngày lễ hay không (non-holiday, holiday). |
| monthSince2005 | INT(6) | Tháng thứ i tính từ ngày 01/01/2005. |
| quarterOfYear | INT(6) | Quý của năm (1, 2, 3, 4). |

|  |  |  |
| --- | --- | --- |
| quarterSince2005 | VARCHAR(255) | Quý của năm (ví dụ: 2005-Q1, 2005-Q2). |
| weekMondayStart | DATE | Ngày thứ 2 của tuần (định dạng dd/mm/YYYY). |
| weekOfYearMonday | INT(6) | Tuần thứ i(1…53) của năm, đầu tuần tính từ thứ hai. |
| weekOfYearSunday | INT(6) | Tuần thứ i(1...52) của năm, đầu tuần tính từ Chủ nhật. |
| weekSundayStart | DATE | Ngày Chủ nhật của tuần (định dạng dd/mm/YYYY). |
| yearWeekSunday | VARCHAR(255) | Năm và tuần, đầu tuần tính từ Chủ nhật (2005-W01, 2005-W02). |

### **monthDim**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| |  | | --- | | **Trường** | | |  | | --- | | **Trường** | | |  | | --- | | **Trường** | |
| monthSk | INT(6) PRIMARY KEY | Khóa ngoại, liên kết với cột monthSince2005 trong dateDim. |
| calendarYearMonth | VARCHAR(255) | Năm và tháng (2005-Jan, 2005-Feb,...). |
| dateSkEnd | INT(6) | Khóa ngoại liên kết đến cột dateSk trong dateDim, thể hiện ngày kết thúc của tháng. |
| dateSkStart | INT(6) | Khóa ngoại liên kết đến cột dateSk trong dateDim, thể hiện ngày bắt đầu của tháng. |
| monthSince2005 | INT(6) | Tháng thứ i(1, 2, 3, 4…) tính từ ngày 01/01/2005. |

1. **CLASS DIAGRAM**

****

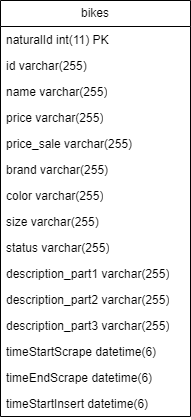
1. **CẤU TRÚC CSDL STAGING**

## **CẤU TRÚC CƠ BẢN, KIỂU DỮ LIỆU**

### **bikes**

|  |  |  |
| --- | --- | --- |
| **Trường** | **Kiểu dữ liệu** | **Mô tả** |
| naturalId | INT(11) PRIMARY KEY | Định danh duy nhất cho từng xe đạp (có thể là mã tự nhiên hoặc định danh ngoài). |
| id | VARCHAR(255) | Mã xe đạp (có thể là SKU hoặc mã do hệ thống quản lý). |
| name | VARCHAR(255) | Tên của xe đạp. |
| price | VARCHAR(255) | Giá bán của xe đạp. |
| price\_sale | VARCHAR(255) | Giá bán đã giảm (nếu có giảm giá). |
| brand | VARCHAR(255) | Thương hiệu của xe đạp. |
| color | VARCHAR(255) | Màu sắc của xe đạp. |
| size | VARCHAR(255) | Kích thước xe đạp. |
| status | VARCHAR(255) | Tình trạng xe đạp. |
| description\_part1 | VARCHAR(255) | Các thông số kỹ thuật (phần 1). |
| description\_part2 | VARCHAR(255) | Các thông số kỹ thuật (phần 2). |
| description\_part3 | VARCHAR(255) | Các thông số kỹ thuật (phần 3). |
| timeStartScrape | DATETIME | Thời gian bắt đầu lấy dữ liệu từ nguồn. |
| timeEndScrape | DATETIME | Thời gian kết thúc lấy dữ liệu từ nguồn. |
| timeStartInsert | DATETIME | Thời gian thêm vào cơ sở dữ liệu. |

1. **CLASS DIAGRAM**

****

1. **CẤU TRÚC CSDL CỦA DATAWAREHOUSE**

## **CẤU TRÚC CƠ BẢN, KIỂU DỮ LIỆU**

### **Brand**

|  |  |  |
| --- | --- | --- |
| **Trường** | **Kiểu dữ liệu** | **Mô tả** |
| BrandID | INT(6) (PK) | Khóa chính |
| BrandName | VARCHAR | Tên thương hiệu |

### **Bicycle**

|  |  |  |
| --- | --- | --- |
| **Trường** | **Kiểu dữ liệu** | **Mô tả** |
| BicycleID | INT(6) (PK) | Khóa chính |
| Name | VARCHAR | Tên sản phẩm |
| BrandID | INT(6) (FK) | Liên kết với thương hiệu |
| Price | DECIMAL | Giá |
| CategoryID | INT(6) (FK) | Liên kết với danh mục sản phẩm |
| Image | VARCHAR | Đường dẫn hình ảnh sản phẩm |
| Description | TEXT | Mô tả chi tiết sản phẩm |
| Price\_sale | DECIMAL | Giá đã giảm |

|  |  |  |
| --- | --- | --- |
| Size | INT(6) | Kích thước |
| Status | VARCHAR | Tình trạng |

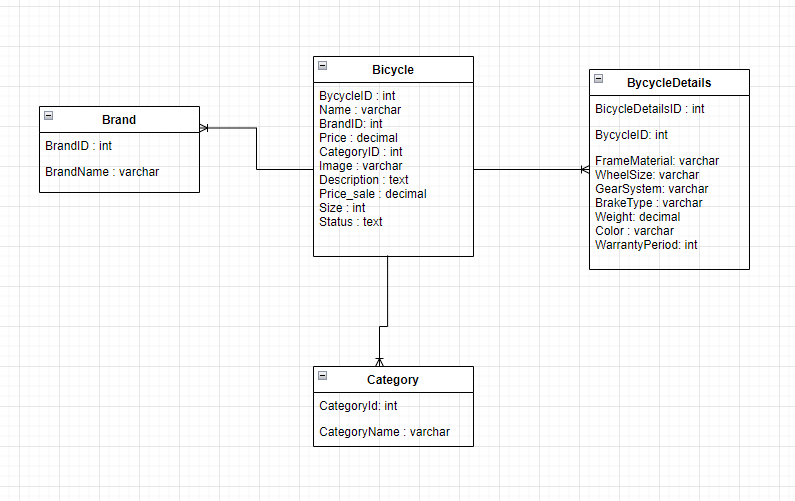
### **Category**

|  |  |  |
| --- | --- | --- |
| **Trường** | **Kiểu dữ liệu** | **Mô tả** |
| CategoryID | INT(6) (PK) | |  | | --- | |  |  |  | | --- | | Khóa chính | |
| CategoryName | VARCHAR | |  | | --- | |  |  |  | | --- | | Tên danh mục | |

### **BicycleDetails**

|  |  |  |
| --- | --- | --- |
| **Trường** | **Kiểu dữ liệu** | **Mô tả** |
| BicycleDetailsID | INT(6)(PK) | |  | | --- | | Khóa chính | |
| BicycleID | INT(6)(FK) | Liên kết với bicycle |
| FrameMaterial | VARCHAR | Chất liệu khung xe |
| WheelSize | VARCHAR | Kích cỡ bánh xe |
| GearSystem | VARCHAR | Hệ thống bánh răng |
| BrakeType | VARCHAR | Loại phanh |
| Weight | DECIMAL | Trọng lượng xe (kg) |
| Color | VARCHAR | Màu sắc |
| WarrantyPeriod | INT(6) | Thời gian bảo hành (tháng) |

## **CLASS DIAGRAM**



1. **SƠ ĐỒ LUỒNG THỰC HIỆN VIỆC LẤY DỮ LIỆU VÀO HỆ THỐNG**