**BỘ THÔNG TIN VÀ TRUYỀN THÔNG**

**HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG**

**---------------------------------------**



**BÁO CÁO KHO DỮ LIỆU**

|  |  |  |
| --- | --- | --- |
| **Giảng viên:** | **Nguyễn Quỳnh Chi** |  |
| **Nhóm:** | **04** |  |
| **Sinh viên thực hiện:** | **Nguyễn Hải Anh** | **B15DCCN037** |
|  | **Trần Đức Lâm** | **B15DCCN301** |
|  | **Lê Minh Tuấn** | **B15DCCN609** |
|  | **Trần Văn Vũ** | **B15DCCN653** |

**HÀ NỘI – 2019**

**MỤC LỤC**

[**I.** **GIỚI THIỆU** 2](#_Toc514184098)

[**1.** **Mục Tiêu** 2](#_Toc514184099)

[**2.** **Phạm Vi Dự Án** 2](#_Toc514184100)

[**II.** **YÊU CẦU NGHIỆP VỤ** 2](#_Toc514184101)

[**III.** **ĐẶC TẢ CHỨC NĂNG** 3](#_Toc514184102)

[**1.** **Đặc Tả Đầu Vào** 3](#_Toc514184103)

[**2.** **Đặc Tả Đầu Ra** 3](#_Toc514184104)

[**IV.** **THIẾT KẾ KHO DỮ LIỆU** 5](#_Toc514184105)

[**1.** **Chuyển Đổi Lược Đồ Dữ Liệu Giữa Các Mô Hình** 5](#_Toc514184106)

[**2.** **Thao Tác Xử Lý Phân Tích Trực Tuyến** 14](#_Toc514184107)

[**V.** **BÁO CÁO PHÂN TÍCH TRỰC TUYẾN** 16](#_Toc514184108)

1. **GIỚI THIỆU**
2. **Mục Tiêu**

Mục tiêu của bài toán là xây dựng kho dữ liệu tích hợp các dữ liệu của doanh nghiệp dùng cho phân tích các bài toán kinh doanh để đảm bảo các tiêu chí:

* Đáp ứng đầy đủ các yêu cầu của khách hàng từ những cửa hàng nằm trong thành phố mà khách hàng sinh sống. Nếu những yêu cầu đó không thể đáp ứng thì công ty sẽ chuyển yêu cầu tới các thành phố khi đến khi mặt hàng có đủ để cung cấp.
* Xử lý đặt hàng của khách hàng của doanh nghiệp.
* Nâng cao chất lượng kinh doanh hiện doanh nghiệp đang cung cấp.
* Đặc biệt xác định được tương đối chính xác thị phần của đối thủ cạnh tranh.
* Hỗ trợ để các nhân viên thực hiện tốt, hiệu quả công việc của mình, như có những quyết định hợp lý, nhanh và bán được nhiều hàng hơn, năng suất cao hơn, thu được lợi nhuận cao hơn v.v.
* Kiểm soát được tình hình, số lượng các mặt hàng, cửa hàng và khách hàng.

1. **Phạm Vi Bài Tập Lớn**

* Quản lý việc kinh doanh của một doanh nghiệp gồm nhiều cửa hàng nằm rải rác ở nhiều thành phố và bang khác nhau nhằm đáp ứng được nhu cầu của khách hàng.
* Nguồn dữ liệu của hệ thống cơ sở dữ liệu quan hệ hiện tại của doanh nghiệp.
* Ứng dụng kỹ thuật OLAP để xây dựng khối dữ liệu phân tích sản lượng hàng hóa, khách hàng …của doanh nghiệp theo hướng đa chiều.

1. **YÊU CẦU NGHIỆP VỤ**

Để đáp ứng được nhu cầu của khách hàng, hệ thống kho dữ liệu trích lọc dữ liệu từ hai cơ sở dữ liệu hiện có để cho vào kho dữ liệu và cung cấp các xử lý phân tích trực tuyến với các thao tác cuộn lên, khoan xuống, chọn và chiếu dựa trên những yêu cầu chọn lựa của khách hàng và các bảng theo chiều để đáp ứng yêu cầu của người sử dụng. Thiết lập một chiều thời gian, hệ thống sẽ sinh một báo cáo Phân tích trực tuyến cho những yêu cầu sau:

1. Tìm tất cả các cửa hàng cùng với thành phố, bang, số điện thoại, mô tả, kích cỡ, trọng lượng và đơn giá của tất cả các mặt hàng được bán ở kho đó.

2. Tìm tất cả các đơn đặt hàng với tên khách hàng và ngày đặt hàng được thực hiện bởi khách hàng đó

3. Tìm tất cả các cửa hàng cùng với tên thành phố và số điện thoại mà có bán các mặt hàng được đặt bởi một khách hàng nào đó

4. Tìm địa chỉ văn phòng đại diện với tên thành phố, bang của tất cả các cửa hàng lưu kho một mặt hàng nào đó với số lượng trên mức cụ thể.

5. Với mỗi một đơn đặt hàng của khách, liệt kê các mặt hàng được đặt cùng với mô tả, mã cửa hàng, tên thành phố và các cửa hàng có bán mặt hàng đó.

6. Tìm thành phố và bang mà một khách hàng nào đó sinh sống

7. Tìm mức độ tồn kho của một mặt hàng cụ thể tại tất cả các cửa hàng ở một thành phố cụ thể nào đó

8. Tìm các mặt hàng, số lượng đặt, khách hàng, cửa hàng và thành phố của một đơn đặt hàng.

9. Tìm các khách hàng du lịch, khách hàng đặt theo đường bưu điện và khách hàng thuộc cả hai loại

1. **ĐẶC TẢ CHỨC NĂNG**

Công cụ sử dụng là SQL Server 2016 và Oracle

1. **Đặc Tả Đầu Vào**

* Cơ sở dữ liệu văn phòng đại diện gồm: các thông tin của khách hàng.
* Cơ sở dữ liệu bán hàng gồm các thông tin của văn phòng đại diện, cửa hàng, mặt hàng, đơn hàng, mặt hàng được lưu trong kho và mặt hàng được đặt.
* Phương thức quản lý tại các cửa hàng ,văn phòng đại diện, doanh nghiệp

1. **Đặc Tả Đầu Ra**

Đầu ra của kho dữ liệu là các bảng dữ liệu động để hỗ trợ cho người sử dụng ra quyết định thực hiện các phép chiếu chọn, các thao tác cơ bản như cuộn lên, khoan xuống…

1. **THIẾT KẾ KHO DỮ LIỆU**
2. **Chuyển Đổi Lược Đồ Dữ Liệu Giữa Các Mô Hình**

*Bước 1: Phân loại các quan hệ và thuộc tính*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Relation name | Reltype | Primary key | KAP | FKA | NKA |
| Khách hàng | PR1 | Mã KH |  |  | Tên KH  Mã Thành phố  Ngày đặt hàng đầu tiên |
| Khách hàng du lịch | PR2 | Mã KH | Mã KH |  | Hướng dẫn viên du lịch  Thời gian |
| Khách hàng bưu điện | PR2 | Mã KH | Mã KH |  | Địa chỉ bưu điện  Thời gian |
| Văn phòng đại diện | PR1 | Mã thành phố |  |  | Tên Thành phố  Địa chỉ VP  Bang  Thời gian |
| Cửa hàng | PR1 | Mã cửa hàng |  | Mã thành phố | Số điện thoại  Thời gian |
| Mặt hàng | PR1 | Mã MH |  |  | Mô tả  Kích cỡ  Trọng lượng  Giá  Thời gian |
| Mặt hàng được lưu trữ | SR1 | Mã cửa hàng  Mã mặt hàng | Mã cửa hàng  Mã mặt hàng |  | Số lượng trong kho  Thời gian |
| Đơn đặt hàng | PR1 | Mã đơn |  |  | Ngày đặt hàng  Mã Khách hàng |
| Mặt hàng được đặt | SR1 | Mã đơn  Mã mặt hàng | Mã đơn  Mã mặt hàng |  | Số lượng đặt  Giá đặt  Thời gian |

*Bước 2: Ánh xạ PR1 thành thực thể*

* Mã Cửa Hàng
* Mã Thành Phố
* Số Điện Thoại
* Thời Gian

Cửa Hàng

* Mã KH
* Tên Khách Hàng
* Mã Thành Phố
* Ngày Đặt Hàng Đầu Tiên

Khách Hàng

* Mã Thành Phố
* Tên Thành Phố
* Địa Chỉ VP
* Bang
* Thời Gian

Văn Phòng Đại Diện

* Mã Đơn
* Ngày Đặt Hàng
* Mã Khách Hàng

Đơn Đặt Hàng

* Mã MH
* Mô Tả
* Kích Cỡ
* Trọng Lượng
* Giá
* Thời Gian

Mặt Hàng

*Bước 3: Ánh xạ PR2 sang thực thể yếu*

* Mã KH
* Hướng Dẫn Viên Du Lịch
* Thời Gian

Khách Hàng

Khách Hàng Bưu Điện

* Mã KH
* Tên Khách Hàng
* Mã Thành Phố
* Ngày Đặt Hàng Đầu Tiên
* Mã KH
* Địa Chỉ Bưu Điện
* Thời Gian

Khách Du Lịch

*Bước 4: Ánh xạ SR1 thành quan hệ 2 ngôi hoặc nhiều ngôi*

Khách Hàng

Khách Du Lịch

R1

* Mã KH
* Tên Khách Hàng
* Mã Thành Phố
* Ngày Đặt Hàng Đầu Tiên
* Mã KH
* Hướng Dẫn Viên Du Lịch
* Thời Gian

Khách Hàng

Khách Bưu Điện

R2 ewt

* Mã KH
* Tên Khách Hàng
* Mã Thành Phố
* Ngày Đặt Hàng Đầu Tiên
* Mã KH
* Địa Chỉ Bưu Điện
* Thời Gian

Mặt Hàng

Cửa Hàng

Mặt Hàng Được Lưu Trữ

* Mã MH
* Mô Tả
* Kích Cỡ
* Trọng Lượng
* Giá
* Thời Gian
* Mã Cửa Hàng
* Mã Thành Phố
* Số Điện Thoại
* Thời Gian
* Số Lượng Trong Kho
* Thời gian

N

M

Mặt Hàng

Đơn Đặt Hàng

Mặt Hàng Được Đặt

* Mã MH
* Mô Tả
* Kích Cỡ
* Trọng Lượng
* Giá
* Thời Gian
* Mã Đơn
* Ngày Đặt Hàng
* Mã Khách Hàng
* Số Lượng Đặt
* Giá Đặt
* Thời Gian

M

N

M

*Bước 5: Ánh xạ mỗi FKA thành một quan hệ*

Văn Phòng Đại Diện

Cửa Hàng

Quản Lý

* Mã Thành Phố
* Tên Thành Phố
* Địa Chỉ VP
* Bang
* Thời Gian
* Mã Cửa Hàng
* Mã Thành Phố
* Số Điện Thoại
* Thời Gian

1

N

*Bước 6: Ánh xạ phụ thuộc bao hàm sang ngữ nghĩa*

|  |  |
| --- | --- |
| **Các phụ thuộc bao hàm phát sinh** | **Ngữ nghĩa suy ra** |
| Cửa hàng.Mã thành phố ⊆VP đại diện.Mã thành phố | Quan hệ một – nhiều giữa thực thể VP đại diện và Cửa hàng |
| MH\_được lưu trữ.Mã cửa hàng ⊆Cửa hàng. Mã cửa hàng  MH\_được lưu trữ.Mã mặt hàng ⊆Mặt hàng.Mã mặt hàng | Quan hệ nhiều – nhiều giữa thực thể Cửa hàng và Mặt hàng |
| MH\_được đặt.Mã mặt hàng ⊆Mặt hàng.Mã mặt hàng MH\_được đặt.Mã đơn⊆Đơn đặt hàng.Mã đơn | Quan hệ nhiều – nhiều giữa thực thể Đơn đặt hàng và Mặt hàng |
| KH bưu điện.Mã KH⊆Khách hàng.Mã KH  KH du lịch.Mã KH⊆Khách hàng.Mã KH | Quan hệ khái quát hóa giao nhau giữa thực thể khách hàng và KH bưu điện, KH du lịch |

*Bước 7: Lược đồ EER*

Mặt Hàng Được Đặt

Đơn Đặt Hàng

Mặt Hàng Được Lưu Trữ

Mặt Hàng

Văn Phòng Đại Diện

Quản Lý

Cửa Hàng

* Mã Đơn
* Ngày Đặt Hàng
* Mã Khách Hàng
* Mã Thành Phố
* Tên Thành Phố
* Địa Chỉ VP
* Bang
* Thời Gian

1

M

* Số Lượng Đặt
* Giá Đặt
* Thời Gian

* Mã MH
* Mô Tả
* Kích Cỡ
* Trọng Lượng
* Giá
* Thời Gian
* Mã Cửa Hàng
* Mã Thành Phố
* Số Điện Thoại
* Thời Gian

N

M

N

N

* Số Lượng Trong Kho
* Thời gian

**Tạo Cơ Sở Dữ Liệu Tích Hợp Từ EER (IDB)**

*Trộn các tập thực thể bằng các quan hệ 2 ngôi và khái quát hóa*

Vì Mã thành phố xuất hiện ở thực thể Văn Phòng đại diện và thực thể Khách hàng, Mã khách hàng xuất hiện ở thực thể Khách Hàng và thực thể Đơn đặt hàng nên chúng được thể hiện như hình vẽ

Khách Hàng

Đơn Đặt Hàng

Đặt

* Mã KH
* Tên Khách Hàng
* Mã Thành Phố
* Ngày Đặt Hàng Đầu Tiên
* Mã Đơn
* Ngày Đặt Hàng
* Mã Khách Hàng

N

1

N

1

Văn Phòng Đại Diện

Khách Hàng

Quản Lý

* Mã Thành Phố
* Tên Thành Phố
* Địa Chỉ VP
* Bang
* Thời Gian
* Mã KH
* Tên Khách Hàng
* Mã Thành Phố
* Ngày Đặt Hàng Đầu Tiên

**Tích hợp 2 lược đồ trên thành lược đồ EER như sau:**

N

N

1

* Mã Cửa Hàng
* Mã Thành Phố
* Số Điện Thoại
* Thời Gian
* Số Lượng Trong Kho
* Thời gian
* Mã MH
* Mô Tả
* Kích Cỡ
* Trọng Lượng
* Giá
* Thời Gian
* Số Lượng Đặt
* Giá Đặt
* Thời Gian
* Mã Đơn
* Ngày Đặt Hàng
* Mã Khách Hàng
* Mã KH
* Tên Khách Hàng
* Mã Thành Phố
* Ngày Đặt Hàng Đầu Tiên

Quản Lý

Khách Hàng

Khách Hàng Du Lịch

Đặt

Văn Phòng Đại Diện

Đơn Đặt Hàng

Quản Lý

Cửa Hàng

Mặt hàng được đặt

Mặt Hàng Được Lưu trữ

Mặt Hàng

Khách Hàng Bưu Điện

* Mã Thành Phố
* Tên Thành Phố
* Địa Chỉ VP
* Bang
* Thời Gian
* Mã KH
* Địa Chỉ Bưu Điện
* Thời Gian
* Mã KH
* Hướng Dẫn Viên Du Lịch
* Thời Gian

1

N

1

N

N

M

M

**Lược Đồ Hình Sao**

Time

* Time\_key
* Day
* Day\_of\_the\_week
* Month
* Quarter
* year

Mặt hàng

* Mã MH
* Mô tả
* Kích cỡ
* Trọng lượng
* Giá

Location

* Location\_key
* Tên thành phố
* Địa chỉ VP
* Bang

Cửa Hàng

* Mã cửa hàng
* Số điện thoại

Khách Hàng

* Mã KH
* Tên khách hàng
* Ngày đặt hàng đầu tiên
* Loại khách hàng

|  |
| --- |
| Sales Fact Table |
| Time\_key |
| Mã MH |
| Mã Cửa Hàng |
| Location\_key |
| Mã KH |
| Units\_sold |
| Dollars\_sold |

|  |
| --- |
| Measures |

**Lược đồ bông tuyết**

Time

* Time\_key
* Day
* Day\_of\_the\_week
* Month
* Quarter
* year

Mặt hàng

* Mã MH
* Mô tả
* Kích cỡ
* Trọng lượng
* Giá

Location

* Location\_key
* Địa chỉ VP
* Mã thành phố

Cửa Hàng

* Mã cửa hàng
* Số điện thoại

Khách Hàng

* Mã KH
* Tên khách hàng
* Ngày đặt hàng đầu tiên
* Mã loại khách hàng

Văn Phòng Đại Diện

* Location\_key
* Địa chỉ VP
* Mã thành phố

Loại Khách Hàng

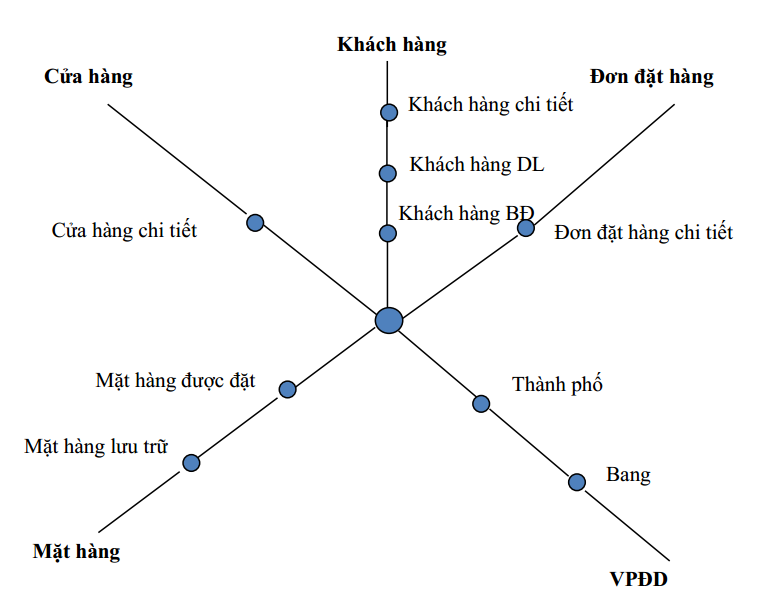
* Mã loại khách hàng
* Khách hàng du lịch
* Khách hàng bưu điện

|  |
| --- |
| Sales Fact Table |
| Time\_key |
| Mã MH |
| Mã Cửa Hàng |
| Location\_key |
| Mã KH |
| Units\_sold |
| Dollars\_sold |

|  |
| --- |
| Measures |

1. **Thao Tác Xử Lý Phân Tích Trực Tuyến**

Mô hình mạng truy vấn:



* Tìm tất cả các cửa hàng cùng với thành phố, bang, số điện thoại, mô tả, kích cỡ, trọng lượng và đơn giá của tất cả các mặt hàng được bán ở kho đó.

Câu truy vấn này lấy dữ liệu theo 3 chiều: Cửa hàng, VPĐD và Mặt hàng. Ở chiều Mặt hàng là Mặt hàng lưu trữ. Ở chiều VPĐD là Thành phố, Bang. Ở chiều Cửa hàng là Cửa hàng chi tiết.

* Tìm tất cả các đơn đặt hàng với tên khách hàng và ngày đặt hàng được thực hiện bởi khách hàng đó

Câu truy vấn này lấy dữ liệu theo 2 chiều là: Đơn đặt hàng và Khách hàng.

Ở chiều Đơn đặt hàng là Đơn đặt hàng chi tiết. Ở chiều Khách hàng là Khách hàng chi tiết.

* Tìm tất cả các cửa hàng cùng với tên thành phố và số điện thoại mà có bán các mặt hàng được đặt bởi một khách hàng nào đó.

Câu truy vấn này lấy dữ liệu theo 4 chiều là: Cửa hàng, VPĐD, Khách hàng, Mặt hàng. Ở chiều Cửa hàng là Cửa hàng chi tiết.Ở chiều VPĐD là Thành phố. Ở chiều Khách hàng là Khách hàng chi tiết. Ở chiều Mặt hàng là Mặt hàng được đặt.

* Tìm địa chỉ văn phòng đại diện với tên thành phố, bang của tất cả các cửa hàng lưu kho một mặt hàng nào đó với số lượng trên mức cụ thể.

Câu truy vấn này lấy dữ liệu theo 3 chiều: VPĐD, Cửa hàng, Mặt hàng. Ở chiều VPĐD là Thành phố, Bang. Ở Cửa hàng là Cửa hàng chi tiết. Ở Mặt hàng là Mặt hàng lưu trữ.

* Với mỗi một đơn đặt hàng của khách, liệt kê các mặt hàng được đặt cùng với mô tả, mã cửa hàng, tên thành phố và các cửa hàng có bán mặt hàng đó.

Câu truy vấn này lấy dữ liệu theo 4 chiều: Đơn đặt hàng, VPĐD, Cửa hàng, Mặt hàng. Ở VPĐD là Thành phố. Ở chiều Cửa hàng là Cửa hàng chi tiết. Ở Mặt hàng là Mặt hàng được đặt. Ở chiều Đơn đặt hàng là Đơn đặt hàng chi tiết.

* Tìm thành phố và bang mà một khách hàng nào đó sinh sống.

Câu truy vấn này lấy dữ liệu theo 2 chiều: VPĐD và Khách hàng. Ở chiều VPĐD là Bang và Thành phố. Ở chiều Khách hàng là Khách hàng chi tiết.

* Tìm mức độ tồn kho của một mặt hàng cụ thể tại tất cả các cửa hàng ở một thành phố cụ thể nào đó.

Câu truy vẫn này lấy dữ liệu theo 3 chiều: VPĐD, Cửa hàng, Mặt hàng. Ở chiều VPĐD là Thành phố. Ở chiều Cửa hàng là Cửa hàng chi tiết. Ở chiều Mặt hàng là Mặt hàng lưu trữ.

* Tìm các mặt hàng, số lượng đặt, khách hàng, cửa hàng và thành phố của một đơn đặt hàng.

Câu truy vấn này lấy dữ liệu theo 5 chiều: Mặt hàng, Khách hàng, Cửa hàng, VPĐD, Đơn đặt hàng.

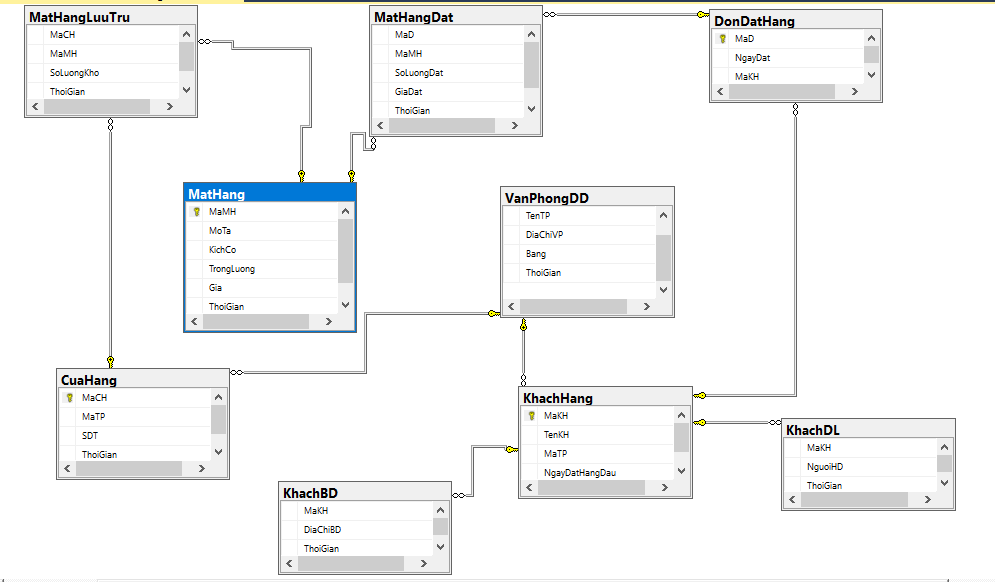
Ở chiều Mặt hàng là Mặt hàng được đặt. Ở chiều Khách hàng là Khách hàng chi tiết. Ở Cửa hàng là Cửa hàng chi tiết. Ở VPĐD là Thành phố. Ở Đơn đặt hàng là Đơn đặt hàng chi tiết.

* Tìm các khách hàng du lịch, khách hàng đặt theo đường bưu điện và khách hàng thuộc cả hai loại.

Câu truy vấn này lấy dữ liệu theo 1 chiều: Khách hàng. Chiều Khách hàng là: Khách hàng du lịch, Khách hàng bưu điện.

1. **BÁO CÁO PHÂN TÍCH TRỰC TUYẾN**

* Khởi động SQL Server Management Studio và tạo CSDL có tên là **KhoDL** như sau và nhập vào các bảng các bản ghi để phân tích.

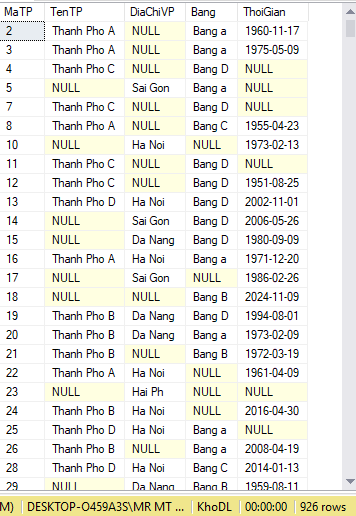


*Sinh Dữ Liệu:*

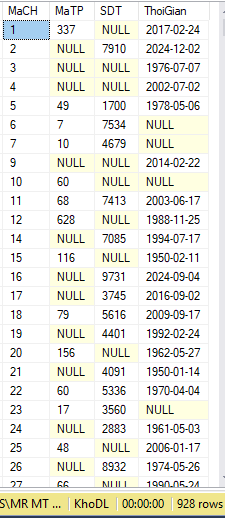
* Sử dụng tool sinh dữ liệu DTM Data Generator

*Dữ liệu các bản sau khi sinh:*

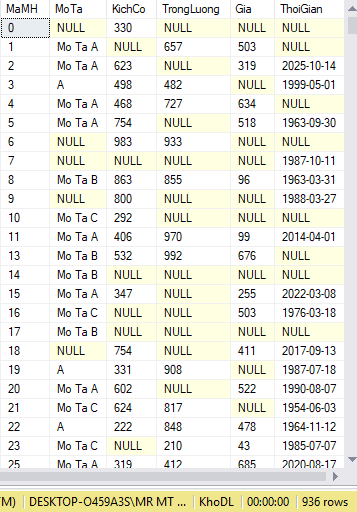
**Bảng VanPhongDD: 926 bản ghi**



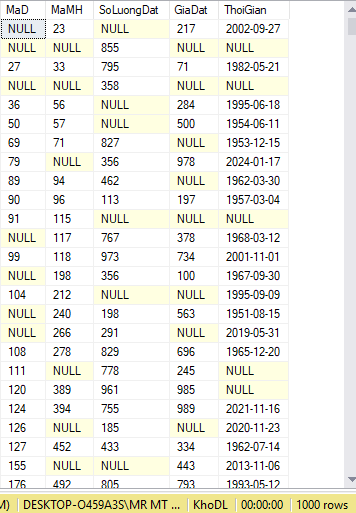
**Bảng CuaHang: 928 bản ghi**



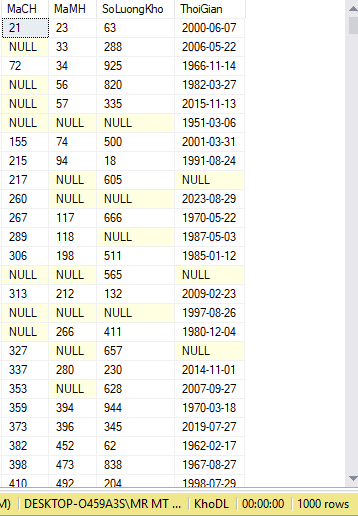
**Bảng MatHang: 936 bản ghi**



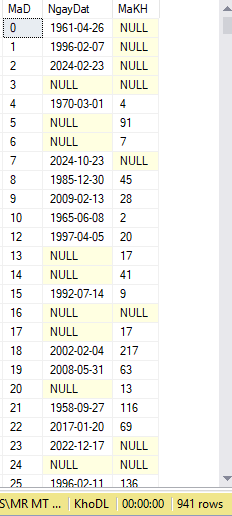
**Bảng MatHangDat: 1000 bản ghi**



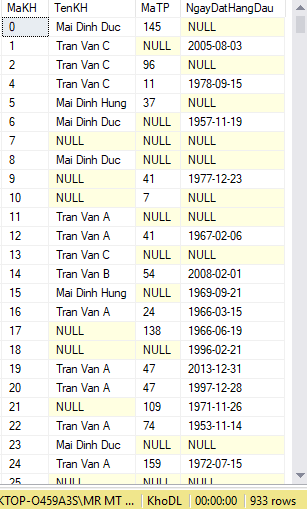
**Bảng MatHangLuuTru: 1000 bản ghi**



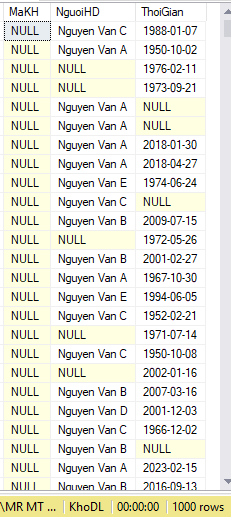
**Bảng DonDatHang: 941 bản ghi**



**Bảng KhachHang: 933 bản ghi**

****

**Bảng KhachDL: 1000 bản ghi**



**Bảng KhachBD: 1000 bản ghi**

