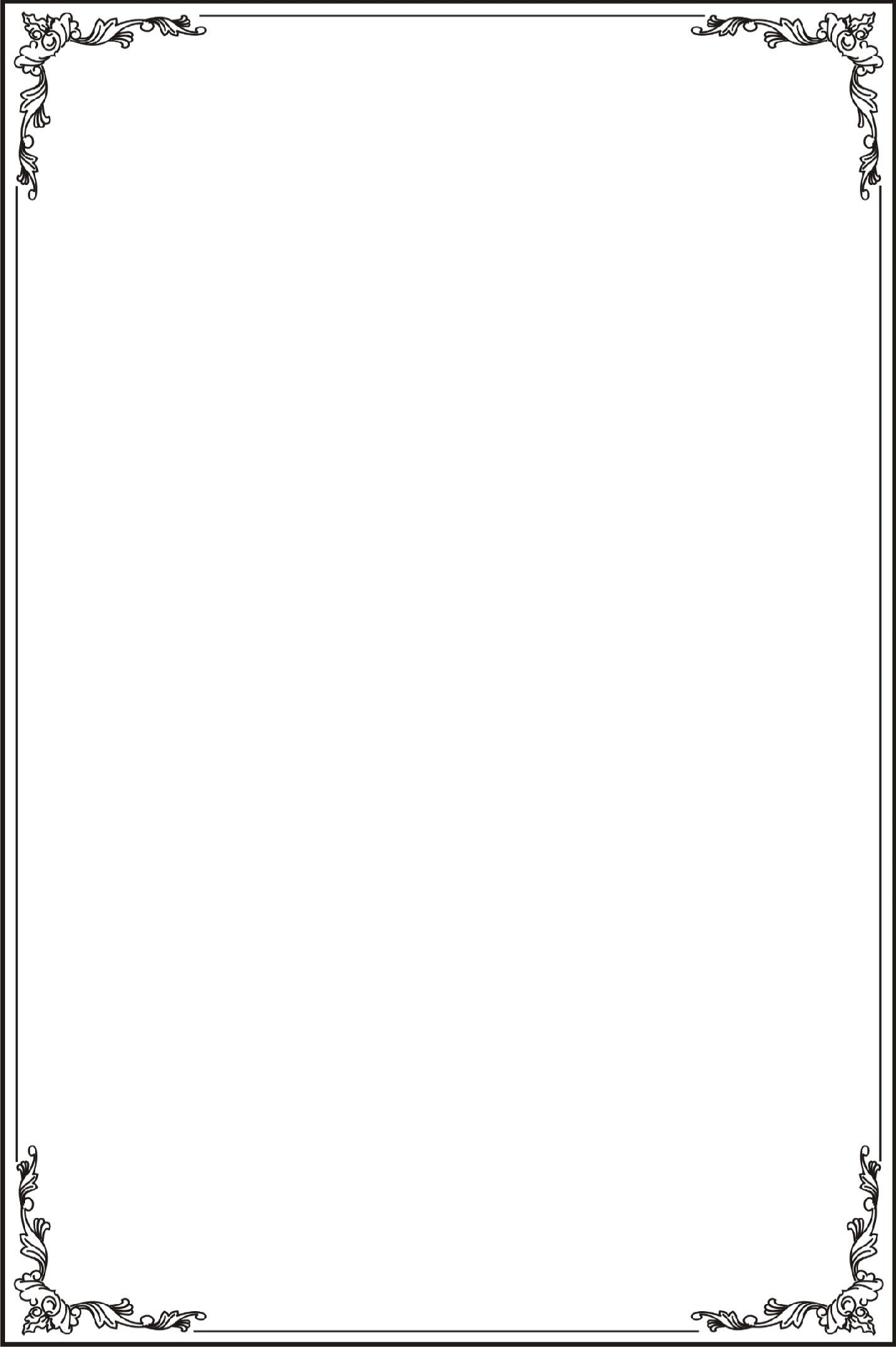
**HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG**



**KHOA CÔNG NGHỆ THÔNG TIN I**

-----🙞🙜🕮🙞🙜-----

**BÁO CÁO BÀI TẬP LỚN**

**KHO DỮ LIỆU VÀ KHAI PHÁ DỮ LIỆU**

**Đề tài: Kho dữ liệu**

**Giảng viên: Nguyễn Quỳnh Chi**

**Nhóm lớp môn học: 04**

**Nhóm bài tập lớn: 06**

**Sinh viên:**

1. **Lê Văn Thắng – B19DCCN658**
2. **Nguyễn Thành Đạt – B19DCCN177**
3. **Nguyễn Đức Việt – B19DCCN713**
4. **Nguyễn Hữu Huy – B19DCCN311**
5. **Ngô Thế Long – B19DCCN395**

**Hà Nội, tháng 4 năm 2023**

**MỤC LỤC**

[LỜI MỞ ĐẦU 2](#_Toc134099438)

[I. Giới thiệu chung 3](#_Toc134099439)

[II. Yêu cầu nghiệp vụ 3](#_Toc134099440)

[III. Đặc tả chức năng 4](#_Toc134099441)

[IV. Thiết kế kho dữ liệu 4](#_Toc134099442)

[V. Cài đặt các khối dữ liệu 11](#_Toc134099443)

[VI. Các báo cáo phân tích trực tuyến 25](#_Toc134099444)

[VII. Kiểm tra tính đúng đắn của dữ liệu 30](#_Toc134099445)

[VIII. Kết luận 30](#_Toc134099446)

# 

# LỜI MỞ ĐẦU

Hiện nay, với sự phát triển của công nghệ, việc lưu trữ và quản lý dữ liệu ngày càng trở nên quan trọng hơn bao giờ hết. Các doanh nghiệp và tổ chức cần có những giải pháp hiệu quả để quản lý dữ liệu của mình, đồng thời đáp ứng nhu cầu lưu trữ và truy xuất thông tin một cách nhanh chóng và chính xác.

Kho dữ liệu là một khái niệm quan trọng trong việc quản lý và lưu trữ dữ liệu. Nó được xây dựng với mục đích tối ưu hóa việc truy xuất thông tin và phân tích dữ liệu. Kho dữ liệu cung cấp một không gian lưu trữ dữ liệu đồng thời cung cấp các công cụ phân tích dữ liệu để giúp người dùng dễ dàng truy xuất và xử lý thông tin một cách nhanh chóng và chính xác. Kho dữ liệu thường được sử dụng trong các doanh nghiệp hay tổ chức, nơi mà nhu cầu lưu trữ và truy xuất thông tin là rất lớn.

Trong bài tập lớn này, nhóm mình sẽ thiết kế và cài đặt một kho dữ liệu cho một hệ thống xử lý đặt hàng của khách hàng cho một doanh nghiệp.

|  |  |
| --- | --- |
| **Thành viên** | **Công việc** |
| Lê Văn Thắng | Tạo phân cấp để tính khối cho OLAP, Thiết kế các khối dữ liệu để thực hiện OLAP |
| Nguyễn Thành Đạt | Tích hợp đúng dữ liệu từ hai nguồn thành một nguồn, Thiết kế mô hình dữ liệu được tích hợp, Thiết kế đúng mô hình dữ liệu cho kho |
| Nguyễn Đức Việt | Tích hợp đúng dữ liệu từ hai nguồn thành một nguồn, Thiết kế mô hình dữ liệu được tích hợp, Sinh dữ liệu cho các nguồn dữ liệu, Thiết kế đúng mô hình dữ liệu cho kho, Cách ánh xạ để đổ dữ liệu từ CSDL tích hợp vào kho, Tài liệu |
| Nguyễn Hữu Huy | Tạo phân cấp để tính khối cho OLAP, Cách ánh xạ để đổ dữ liệu từ CSDL tích hợp vào kho, powerpoint |
| Ngô Thế Long | Tài liệu và powerpoint, Sinh dữ liệu cho các nguồn dữ liệu |

## Giới thiệu chung

* + - 1. Mục tiêu

Xây dựng kho dữ liệu với các dữ liệu của doanh nghiệp để:

* Tạo lập các báo cáo kinh doanh nhằm hỗ trợ người quản lý đưa ra quyết định hợp lý
* Tích hợp toàn bộ dữ liệu của doanh nghiệp
* Trình bày thông tin nhất quán vì dữ liệu đến từ nhiều nơi khác nhau
* Giúp cho doanh nghiệp điều hành các dự án, các nghiệp vụ hiệu quả
  + - 1. Phạm vi của 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 để đáp ứng được nhu cầu của khách hàng.

## 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

## Đặc tả chức năng

* Đặc tả đầu vào: Lược đồ quan hệ của hệ thống cơ sở dữ liệu quan hệ hiện tại của doanh nghiệp này như sau:
* Cơ sở dữ liệu Văn phòng đại diện bao gồm những quan hệ với các lược đồ như sau:
* Khách hàng (Mã KH, Tên KH, Mã Thành phố, Ngày đặt hàng đầu tiên)
* Khách hàng du lịch (\* Mã KH, Hướng dẫn viên du lịch, Thời gian)
* Khách hàng bưu điện (\* Mã KH, Địa chỉ bưu điện, Thời gian)
* Cơ sở dữ liệu Bán hàng gồm những quan hệ với lược đồ như sau:
* Văn phòng đại diện (Mã Thành phố, Tên Thành phố, Địa chỉ VP, Bang, Thời gian)
* Cửa hàng (Mã cửa hàng, \* Mã Thành phố, Số điện thoại, Thời gian)
* Mặt hàng (Mã MH, Mô tả, Kích cỡ, Trọng lượng, Giá, Thời gian)
* Mặt hàng\_được lưu trữ (\* Mã cửa hàng, \* Mã mặt hàng, Số lượng trong kho, Thời gian)
* Đơn đặt hàng (Mã đơn, Ngày đặt hàng, Mã Khách hàng)
* Mặt hàng được đặt (\* Mã đơn, \* Mã mặt hàng, Số lượng đặt, Giá đặt, Thời gian)
* Đặc tả đầu ra: Các bảng dữ liệu động cho phép người quản lý thực hiện trực tuyến thao tác roll up, drill down, slice, dice … để hỗ trợ người quản lý ra quyết định.

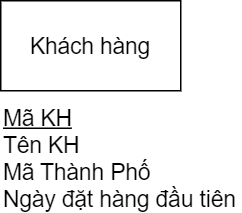
## Thiết kế kho dữ liệu

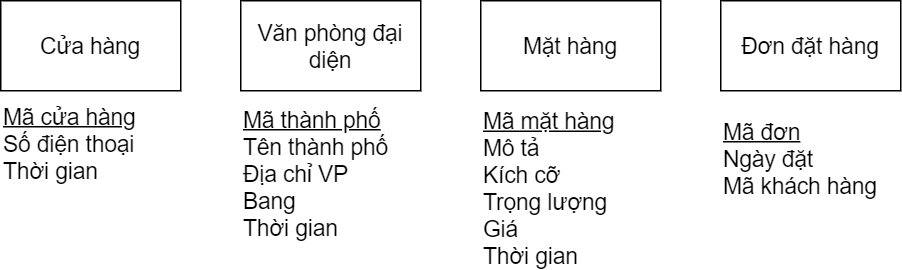
1. Tích hợp lược đồ thành mô hình thực thể liên kết mở rộng

**Bước 1: Xác định từng quan hệ, khóa và trường**

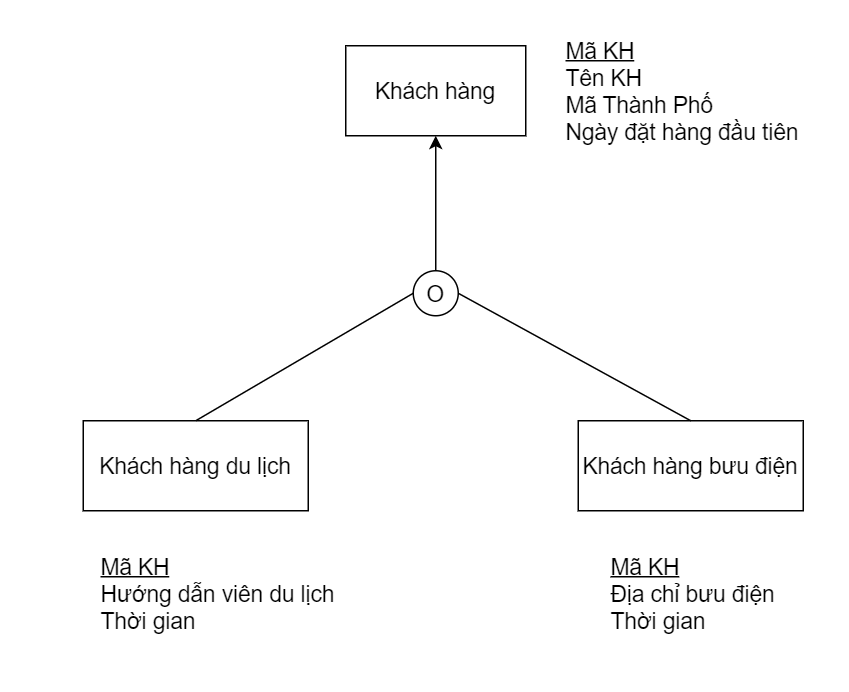
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Relation Name** | **Rel type** | **Primary-Key** | **KAP** | **KAG** | **KFA** | **NKA** |
| Database1: Văn phòng đại diện | | | | | | |
| Khách hàng | PR1 | Mã KH |  |  |  | Tên KH  Mã Thành phố  Ngày đặt hàng đầu tiên |
| Khách 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 |
| Database2: Bán hàng | | | | | | |
| 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 mặt hàng |  |  |  | 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 |

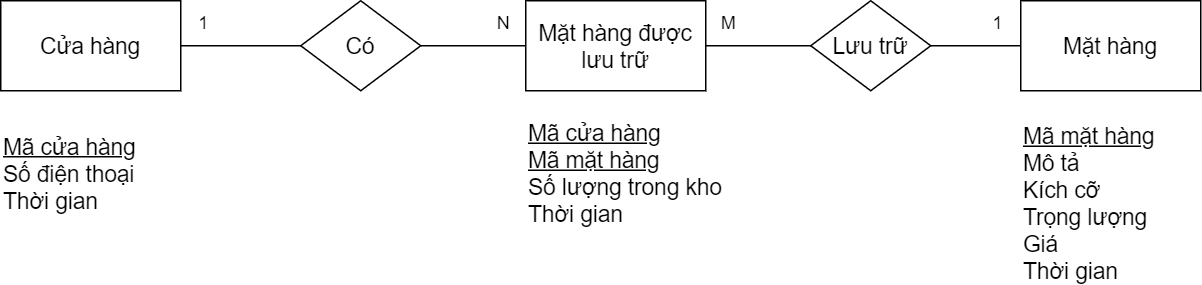
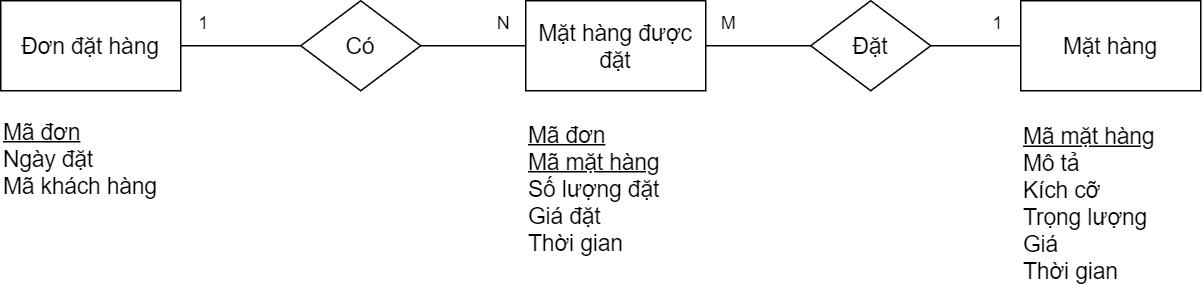
**Bước 2: Ánh xạ PR1 thành thực thể**

DB1: Văn phòng đại diện

DB2: Bán hàng

**Bước 3: Ánh xạ PR2 sang thực thể lớp con**

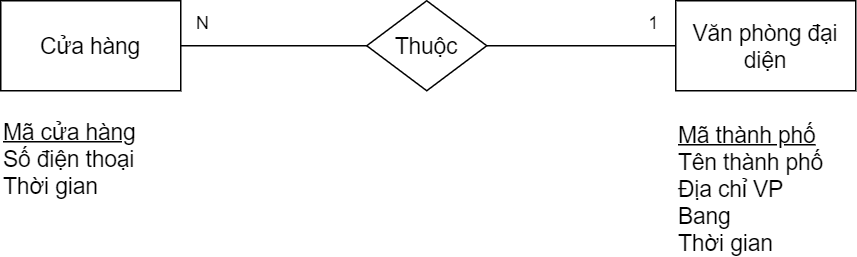
DB1: Văn phòng đại diện

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

DB2: Bán hàng

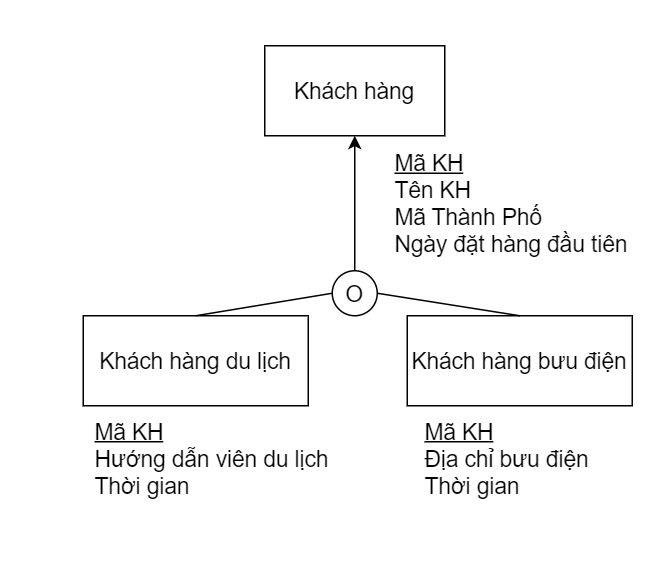
**Bước 5:** **Ánh xạ SR2 thành quan hệ 2 ngôi hoặc nhiều ngôi**

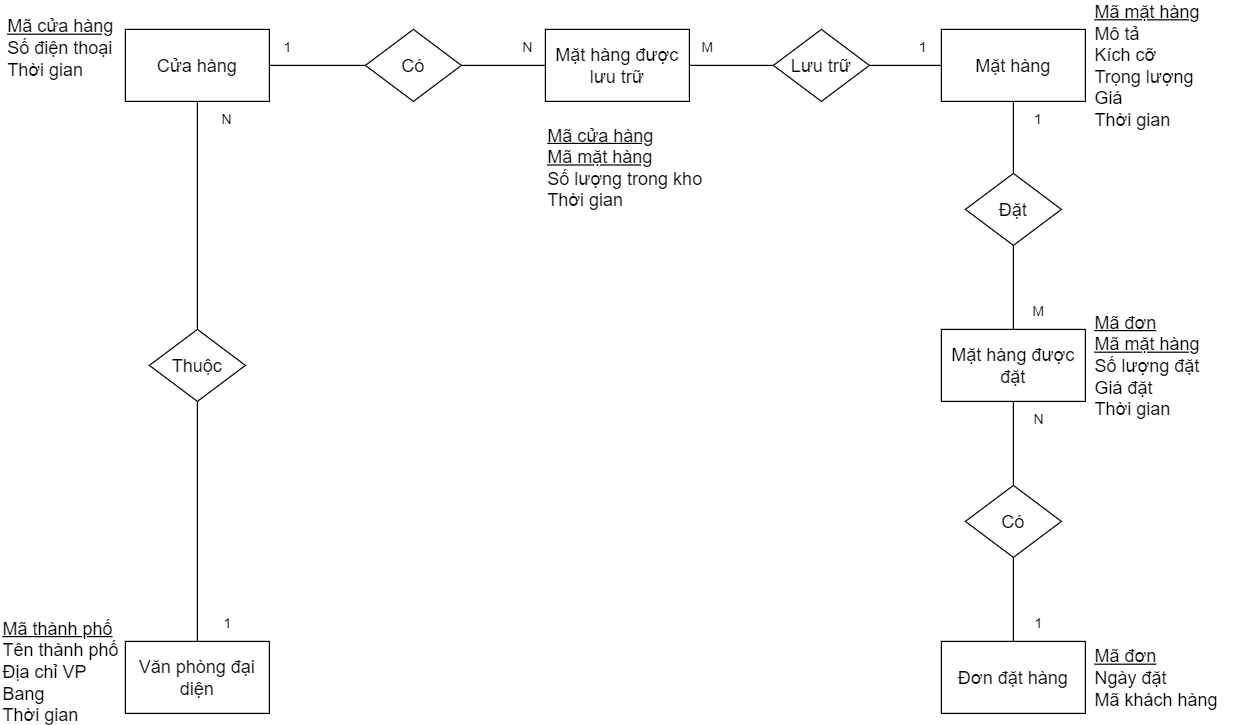
**Bước 6:** **Ánh xạ mỗi FKA thành một quan hệ**



**Bước 7: Ánh xạ phụ thuộc bao hàm ngữ nghĩa**

|  |  |
| --- | --- |
| **Các phụ thuộc bao hàm phái sinh** | **Ngữ nghĩa suy ra** |
| MHDuocLT.MaMH ⊆ MatHang.MaMH  MHDuocLT.MaCH ⊆ CuaHang.MaCH | Quan hệ nhiều-một giữa thực thể Mặt hàng được lưu trữ và Mặt hàng, Mặt hàng được lưu trữ và Cửa hàng |
| MHDuocDat.MaMH ⊆ MatHang.MaMH  MHDuocDat.MaDon ⊆ DonDatHang.MaDon | Quan hệ nhiều-một giữa thực thể Mặt hàng được đặt và Mặt hàng, Mặt hàng được đặt và Đơn đặt hàng |
| CuaHang.MaTP ⊆ VPDaiDien.MaTP | Quan hệ nhiều-một giữa thực thể Cửa hàng và Văn phòng đại diện |

**Bước 8: Từ kết quả 7 bước trên ta thu được 2 mô hình EER1 và EER2 sau đây:**



**Bước 9: Giải quyết xung đột giữa các lược đồ EER**

Giải quyết xung đột về đồng âm và đồng nghĩa giữa các thực thể:

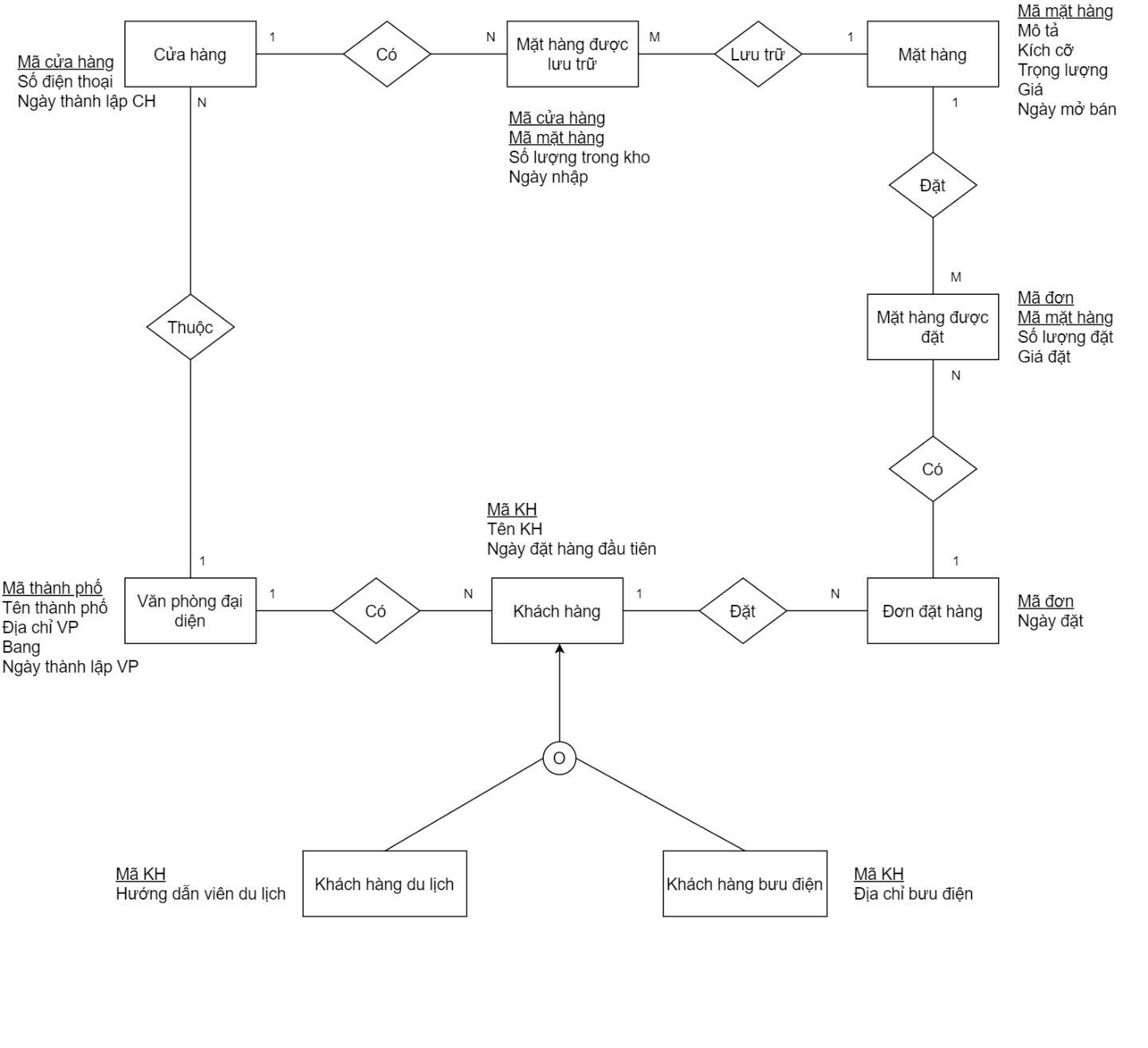
* Văn phòng đại diện: Thời gian ⇒ Ngày thành lập văn phòng
* Cửa hàng: Thời gian ⇒ Ngày thành lập cửa hàng
* Mặt hàng: Thời gian ⇒ Ngày mở bán
* Mặt hàng được lưu trữ: Thời gian ⇒ Ngày nhập
* Mặt hàng được đặt: Thời gian ⇒ Xóa bỏ
* Khách hàng du lịch: Thời gian ⇒ Xóa bỏ
* Khách hàng bưu điện: Thời gian ⇒ Xóa bỏ

**Bước 10: Tích hợp 2 lược đồ EER1 và EER2 thành IEER**

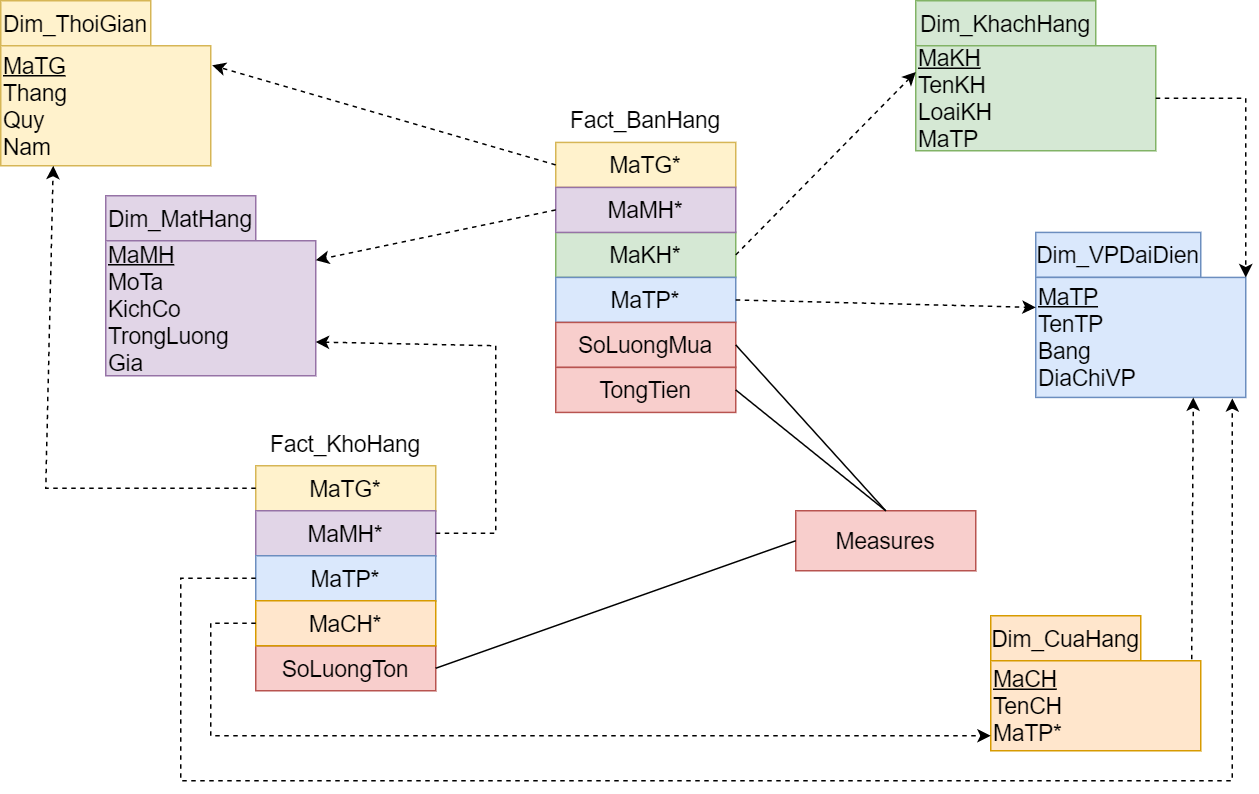
Thực thể Khách hàng chứa Mã TP là khóa chính của thực thể Văn Phòng đại diện và thực thể Đơn đặt hàng chữa Mã KH là khóa chính của thực thể Khách hàng

|  |  |
| --- | --- |
| **Các phụ thuộc bao hàm phái sinh** | **Ngữ nghĩa suy ra** |
| KhachHang.MaTP ⊆ VPDaiDien.MaTP | Quan hệ nhiều-một giữa thực thể Khách hàng và Văn phòng đại diện |
| DonDatHang.MaKH ⊆ KhachHang.MaKH | Quan hệ nhiều-một giữa thực thể Đơn đặt hàng và Khách hàng |

Ta tích hợp được lược đồ IEER sau:

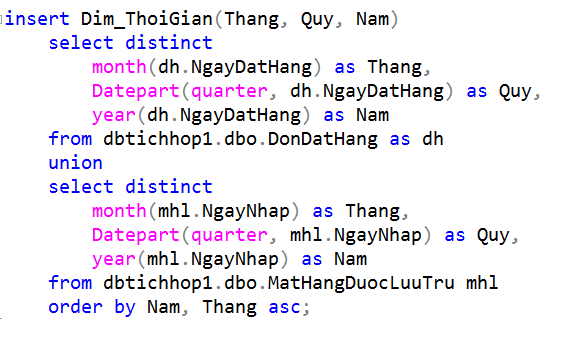


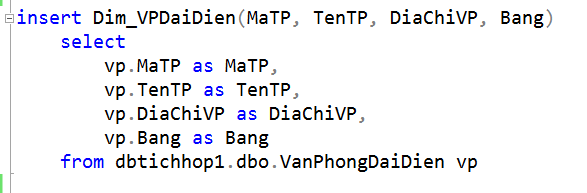
1. Thiết kế lược đồ



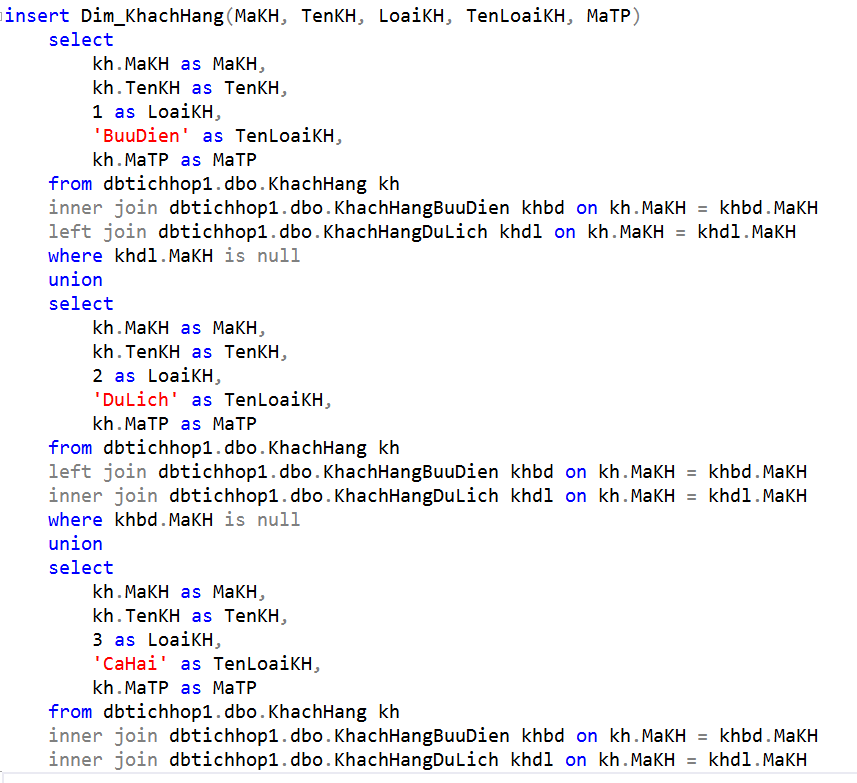
## Cài đặt các khối dữ liệu

Xây dựng các bảng dimension, fact và đổ dữ liệu từ cơ sở dữ liệu tích hợp vào kho dữ liệu

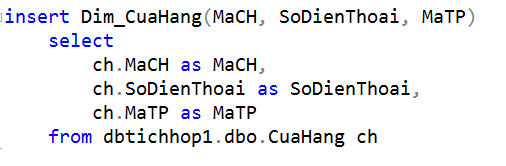
* Dim\_ThoiGian:
* Dim\_VPDaiDien:

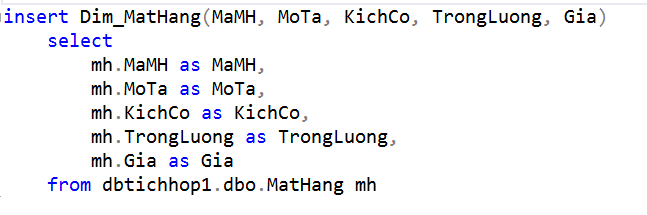
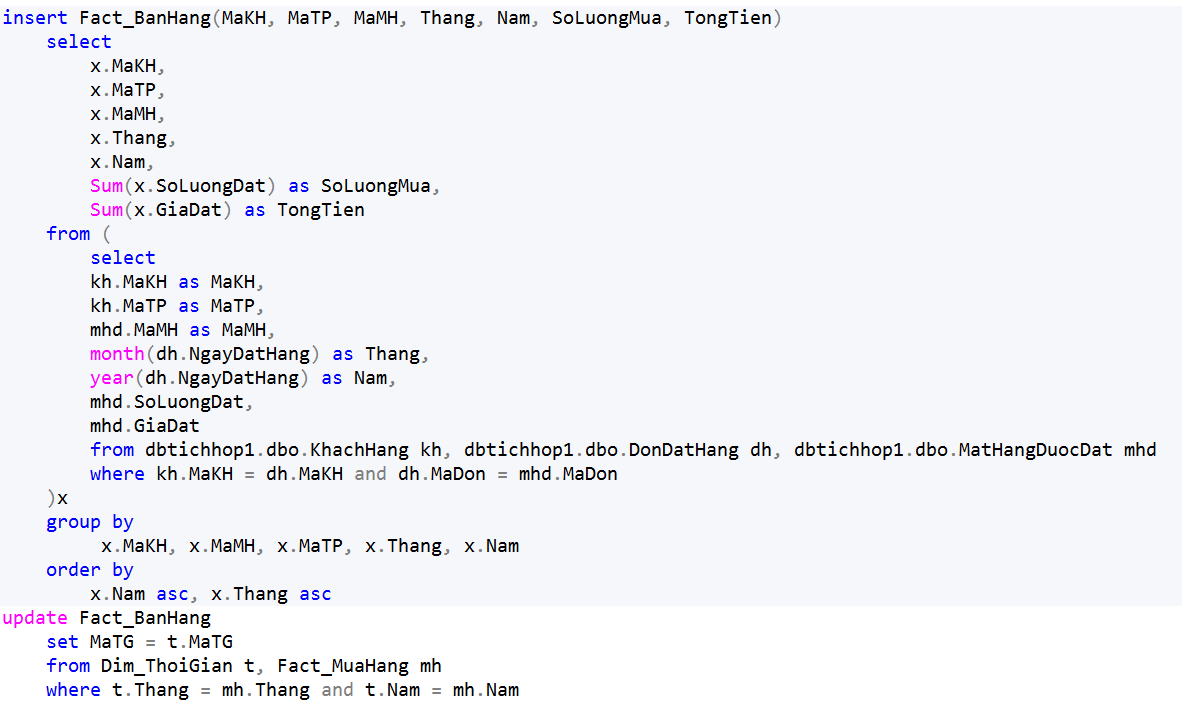
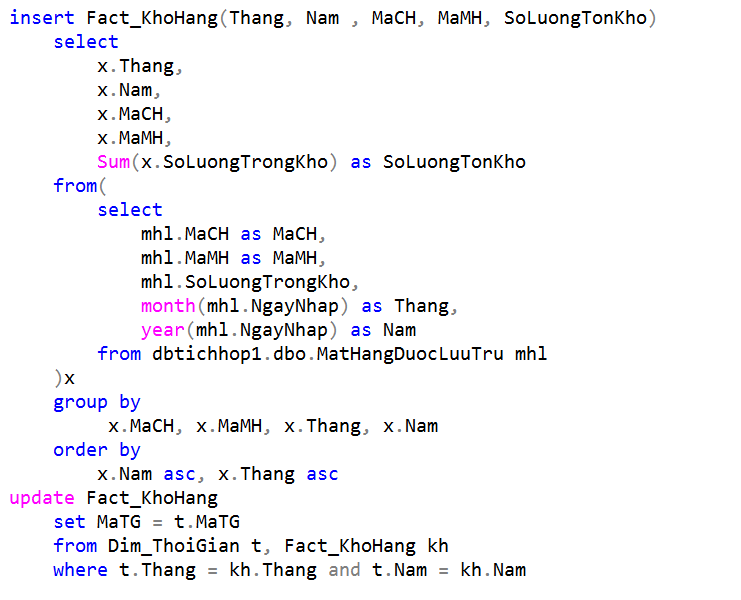


* Dim\_KhachHang:



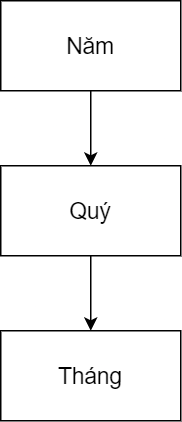
* Dim\_CuaHang:



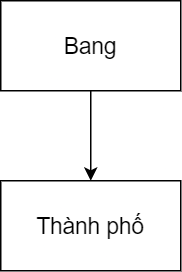
* Dim\_MatHang:
* Fact\_BanHang:
* Fact\_KhoHang: 

Phân cấp các chiều dữ liệu

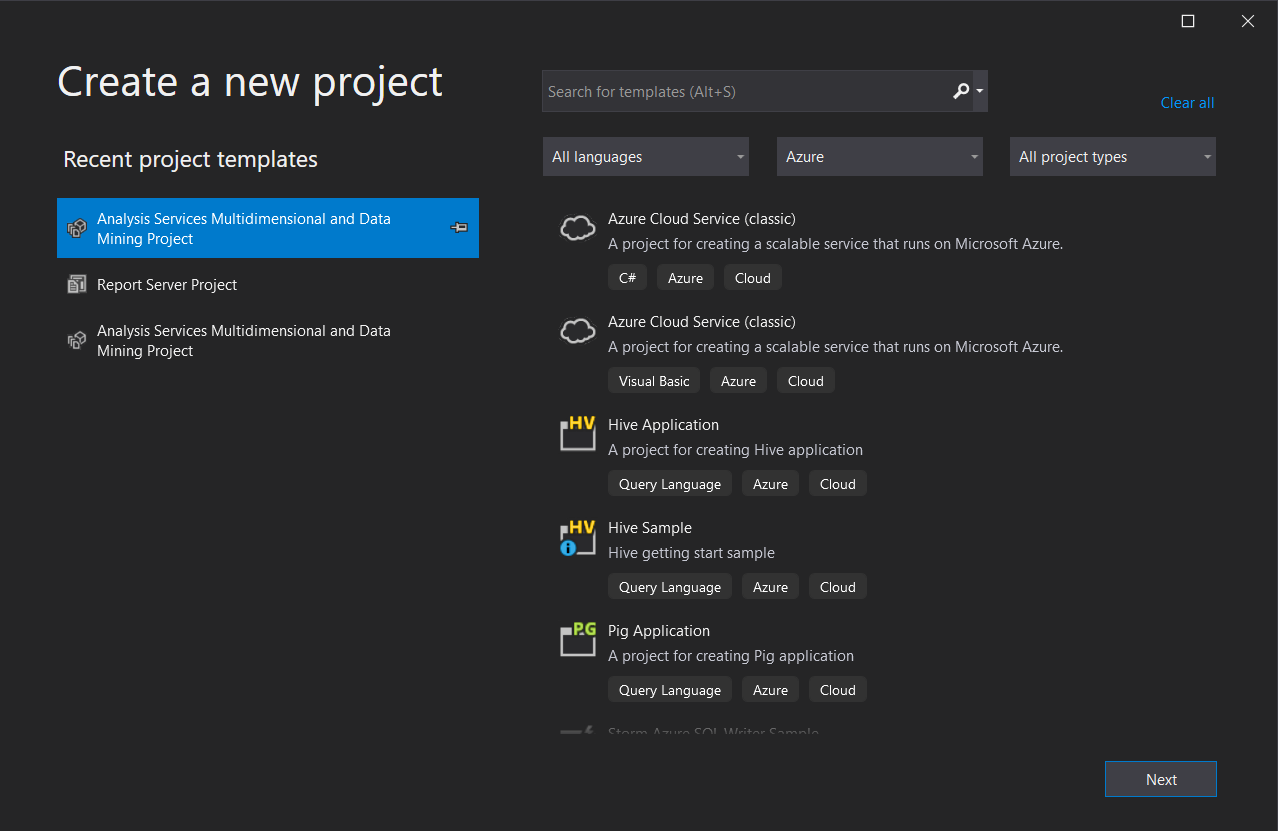
1. Chiều thời gian

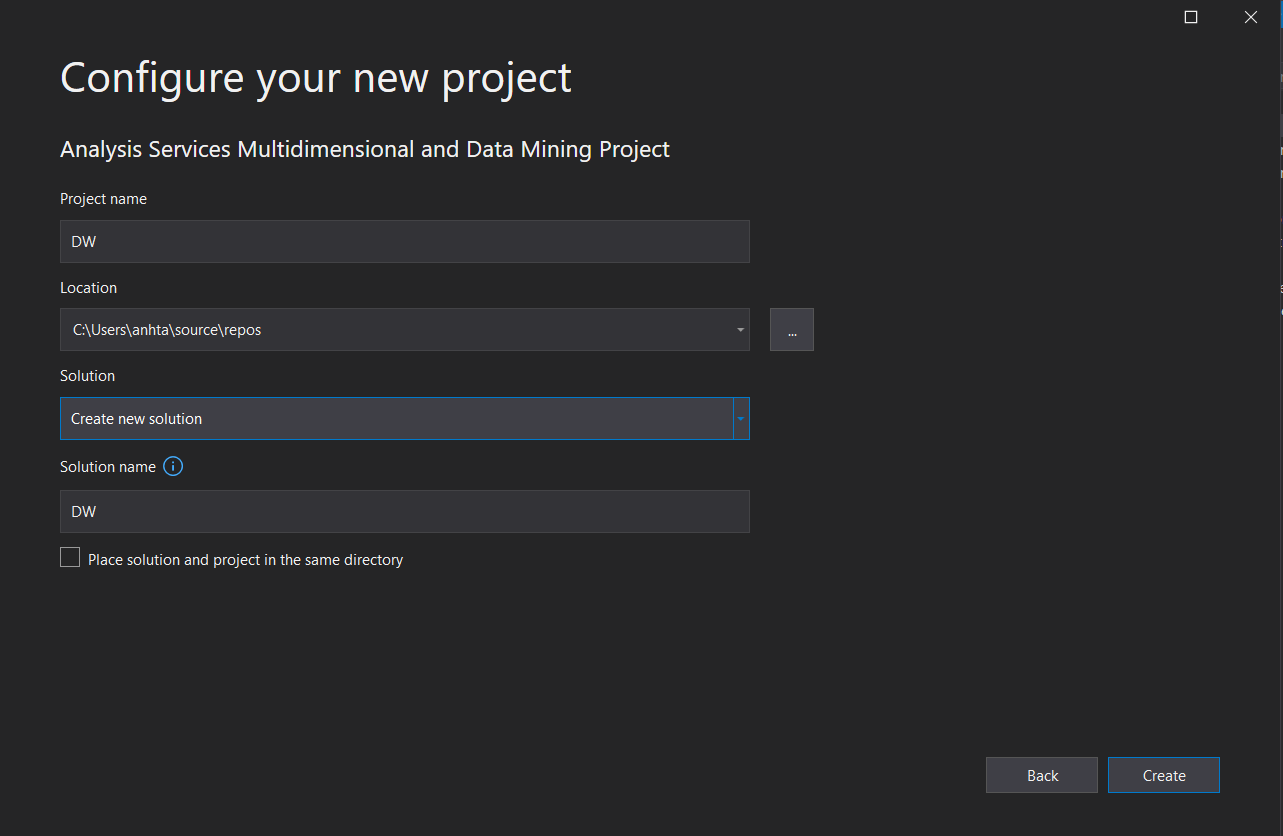


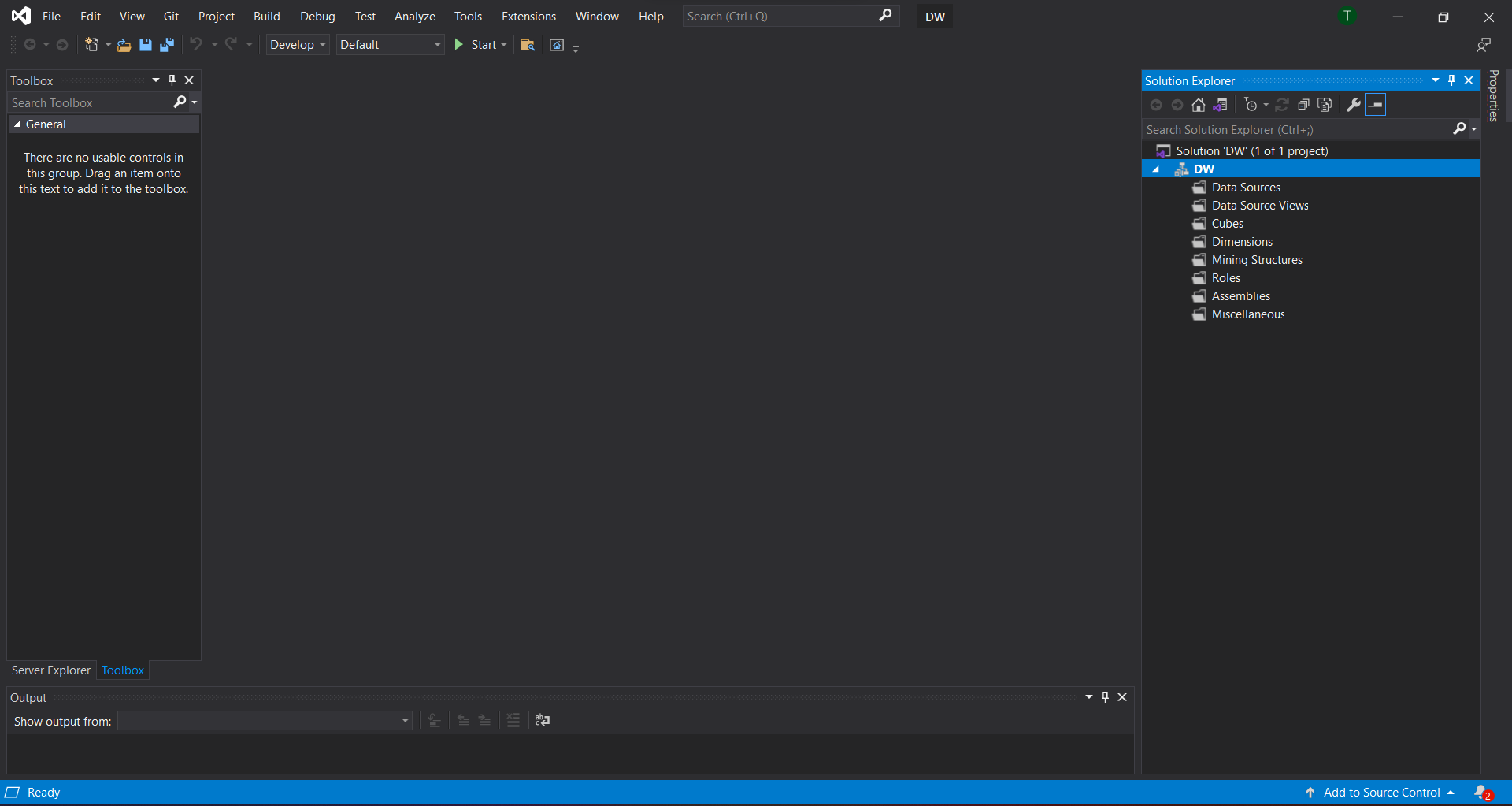
1. Chiều văn phòng đại diện

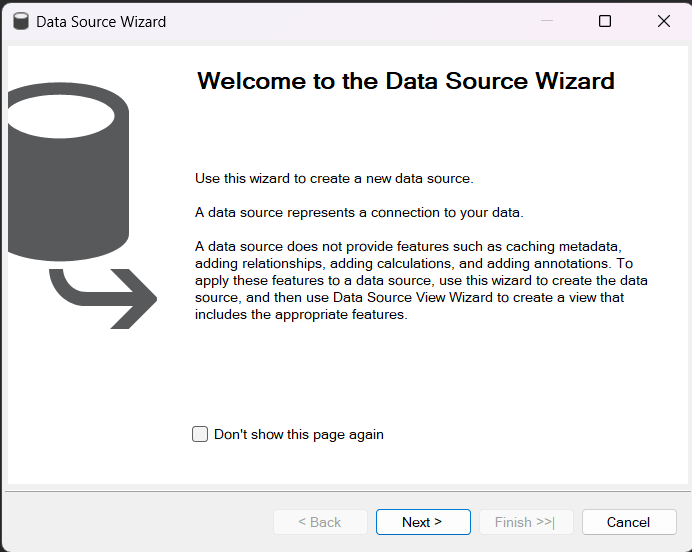


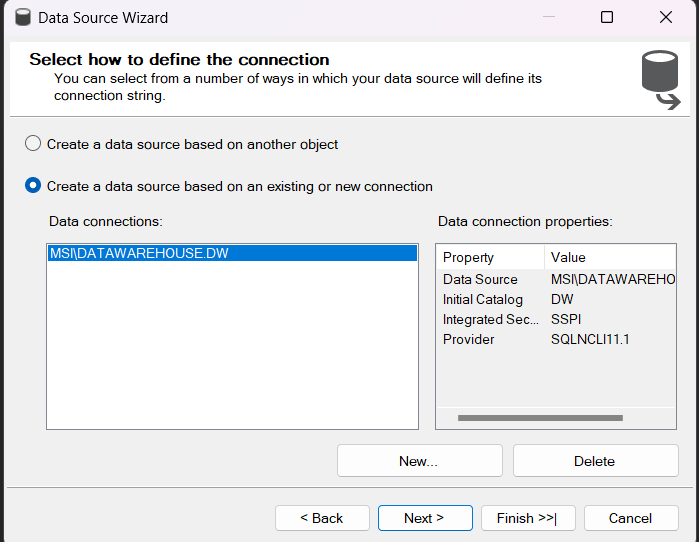
Xây dựng OLAP với Visual Studio:

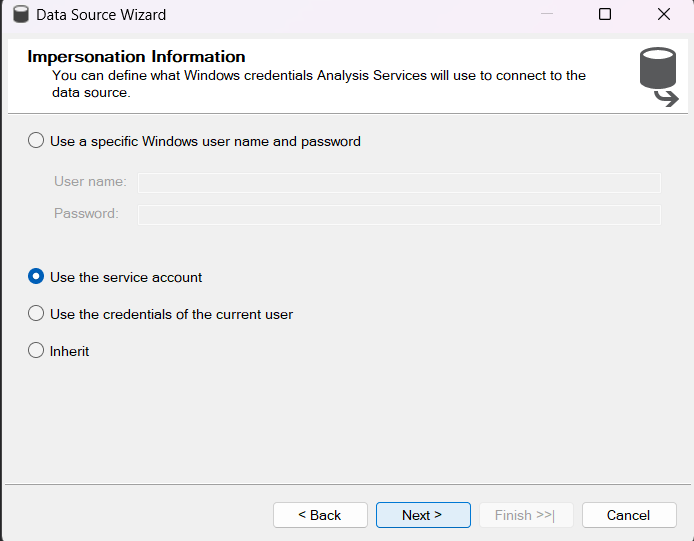


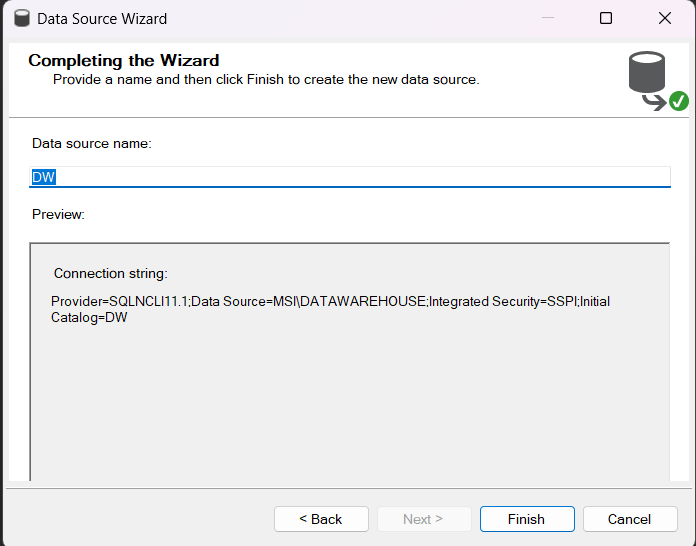


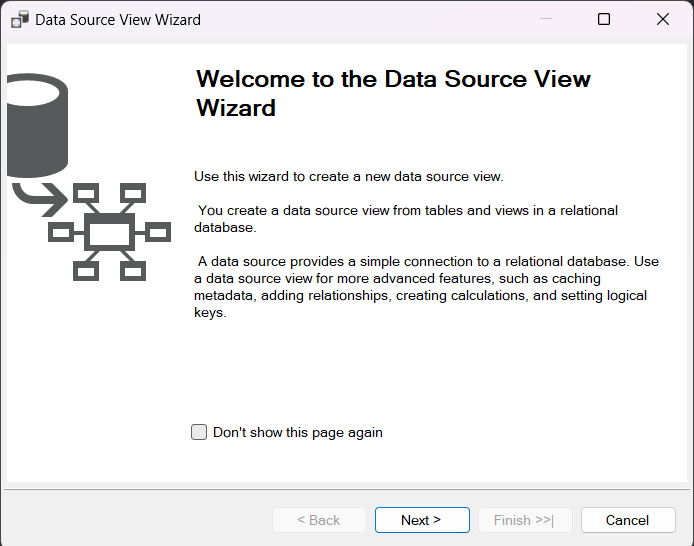


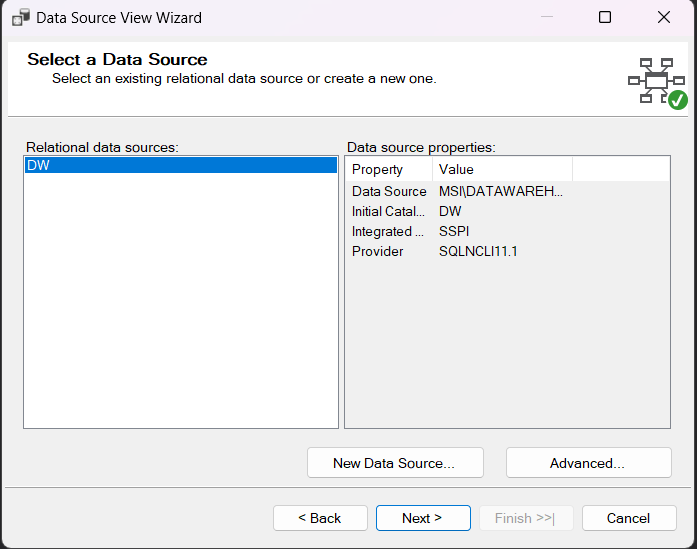


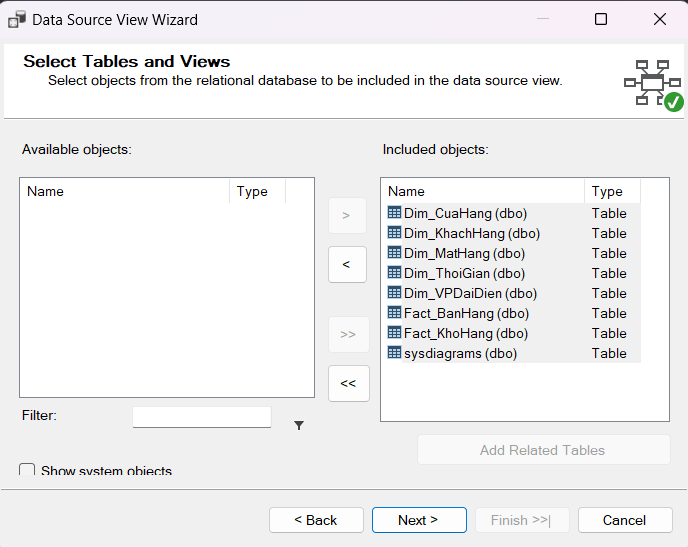


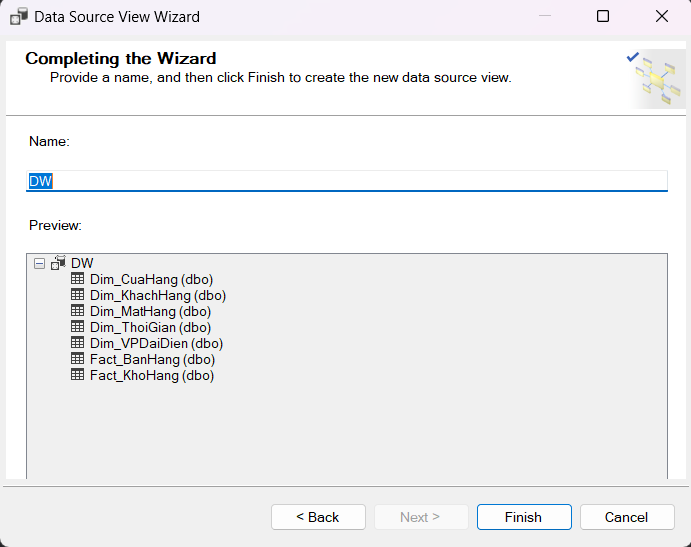


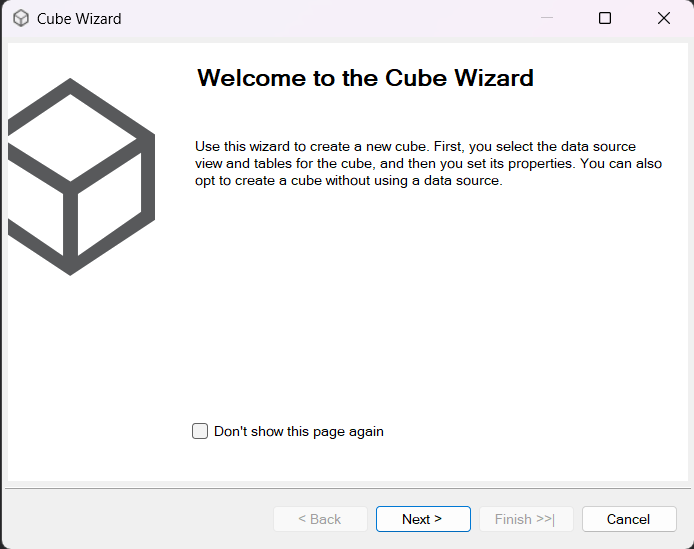


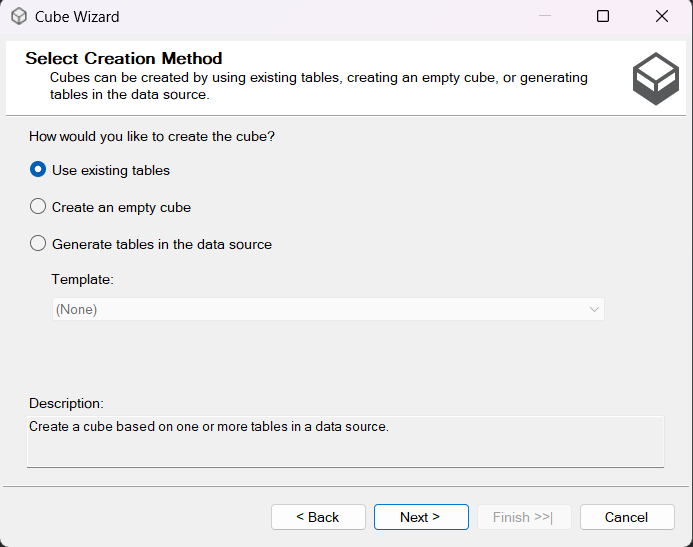


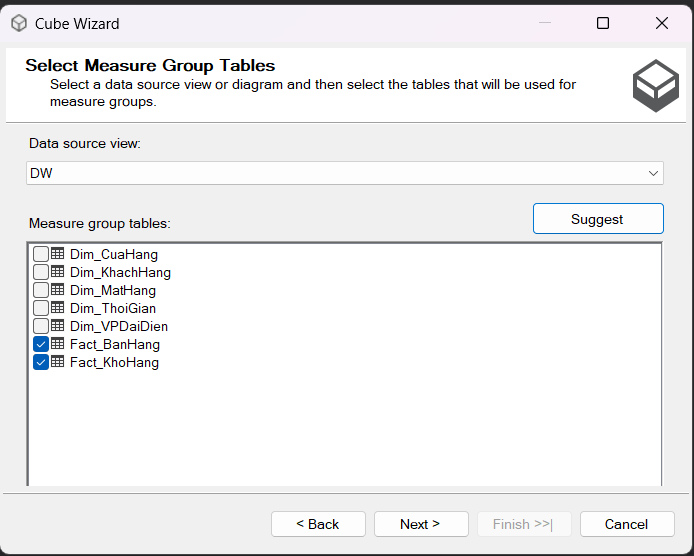


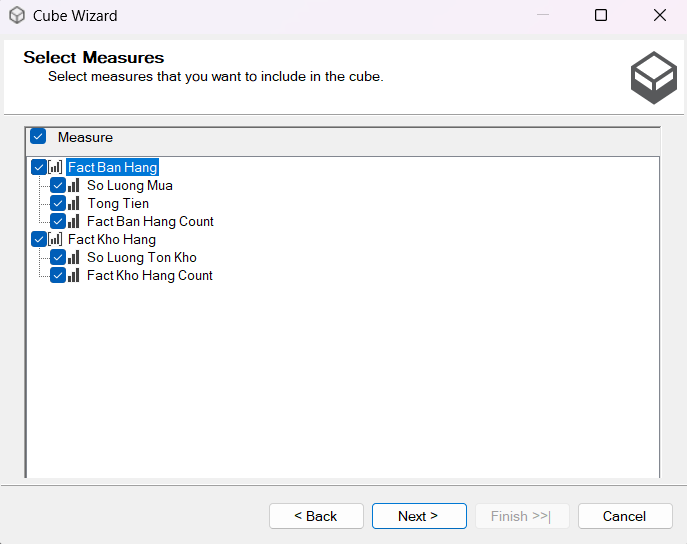


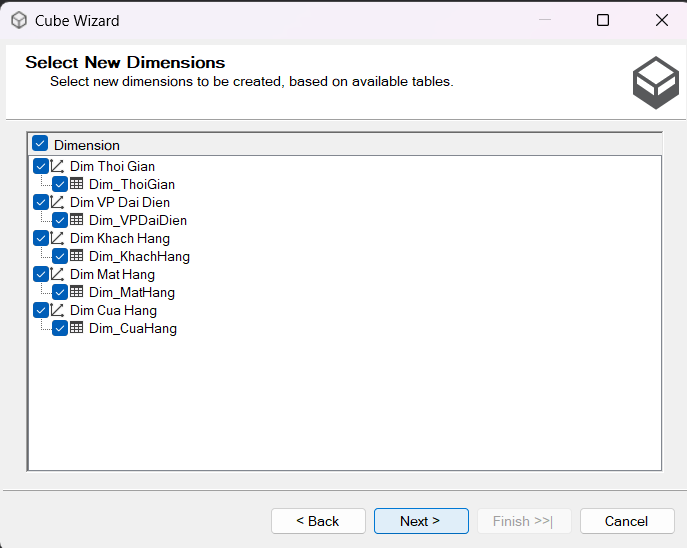
\

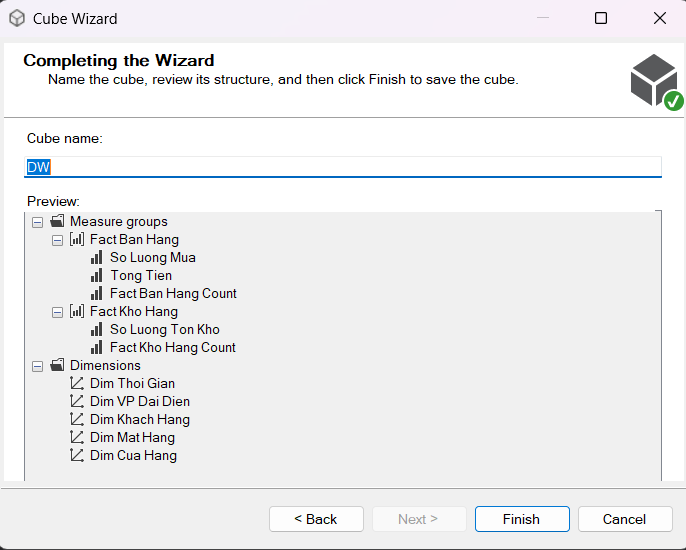


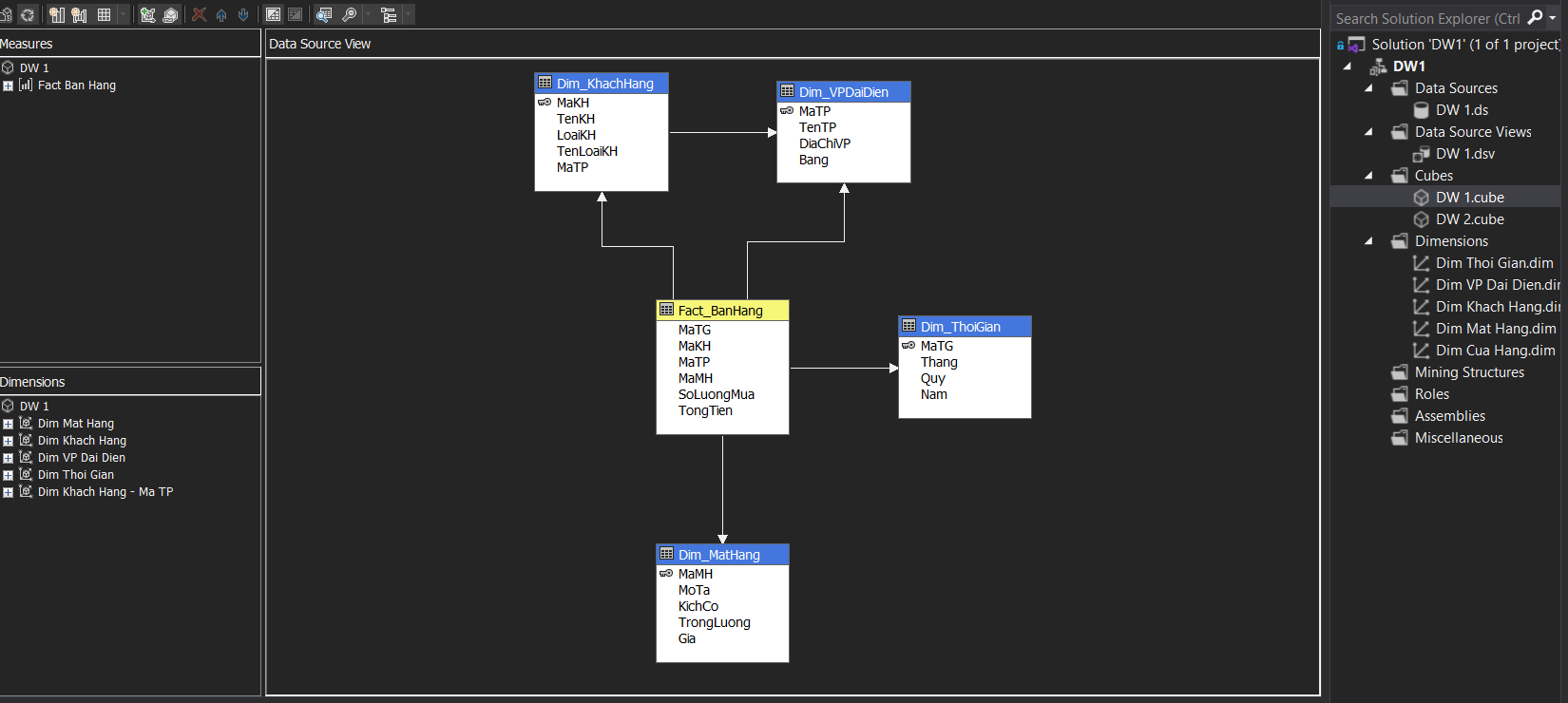


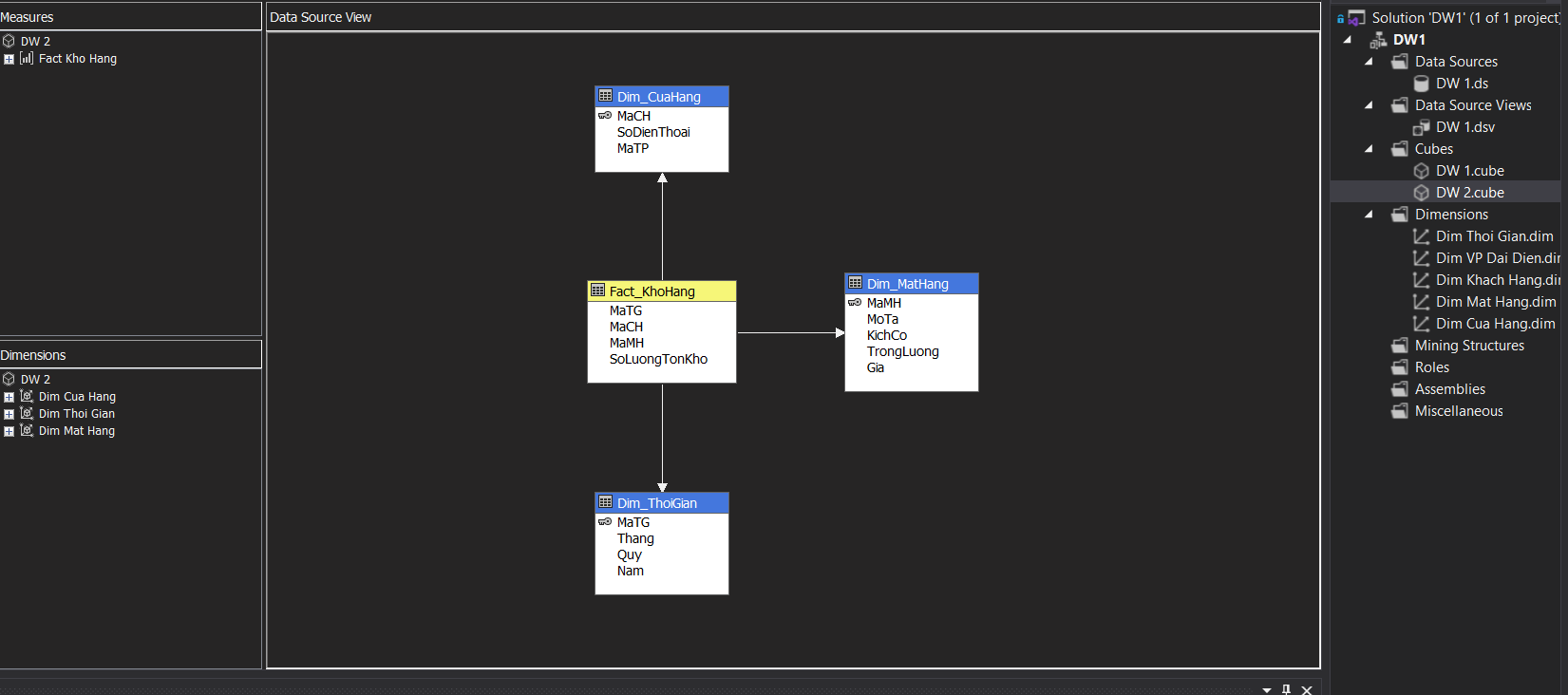


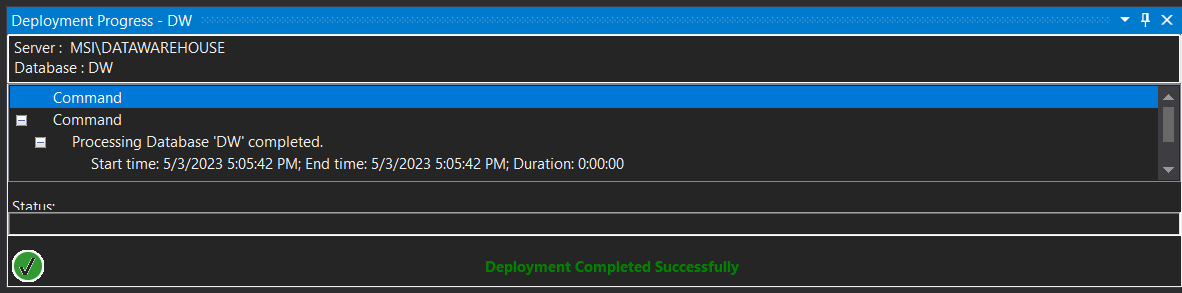


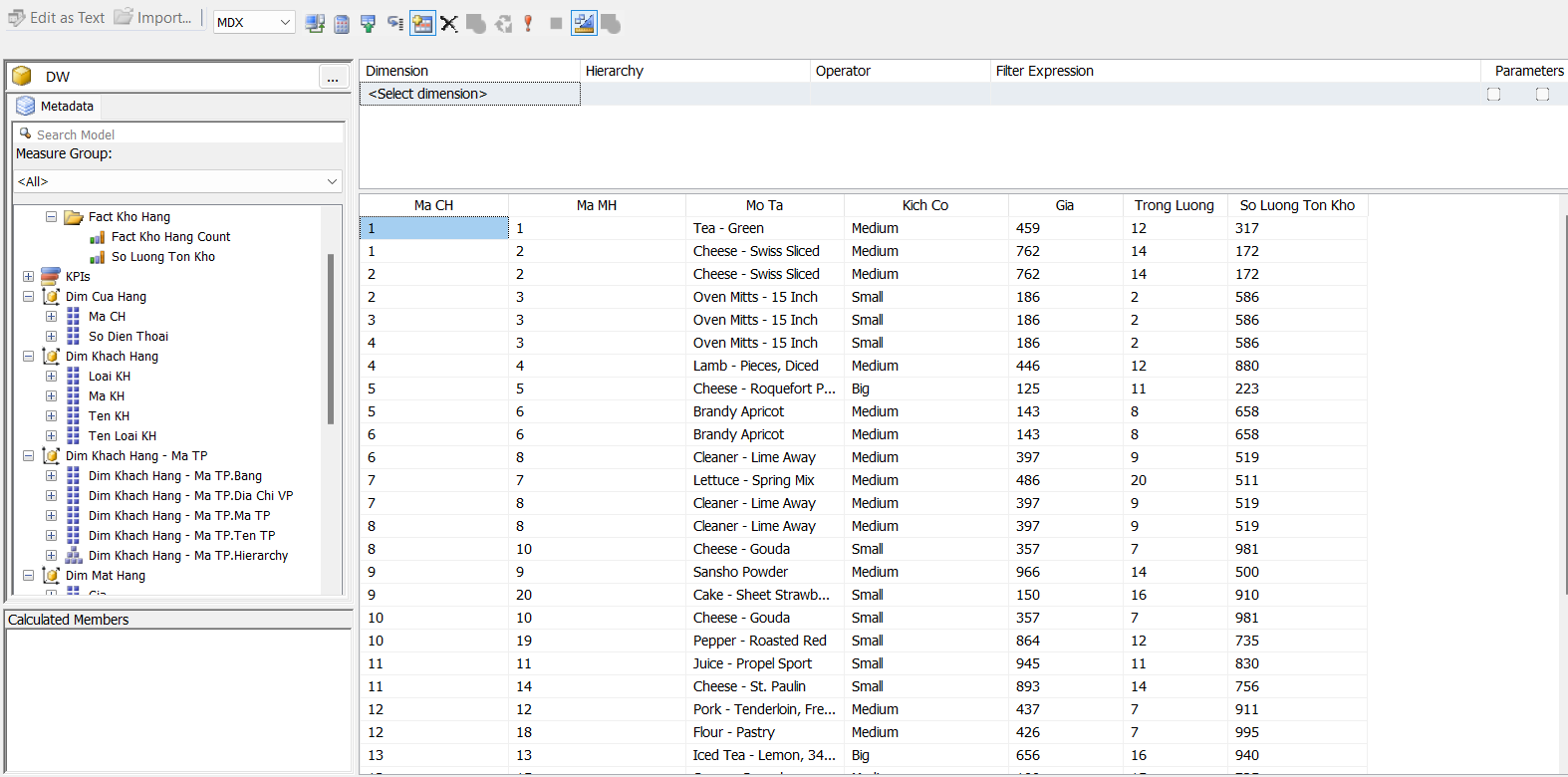




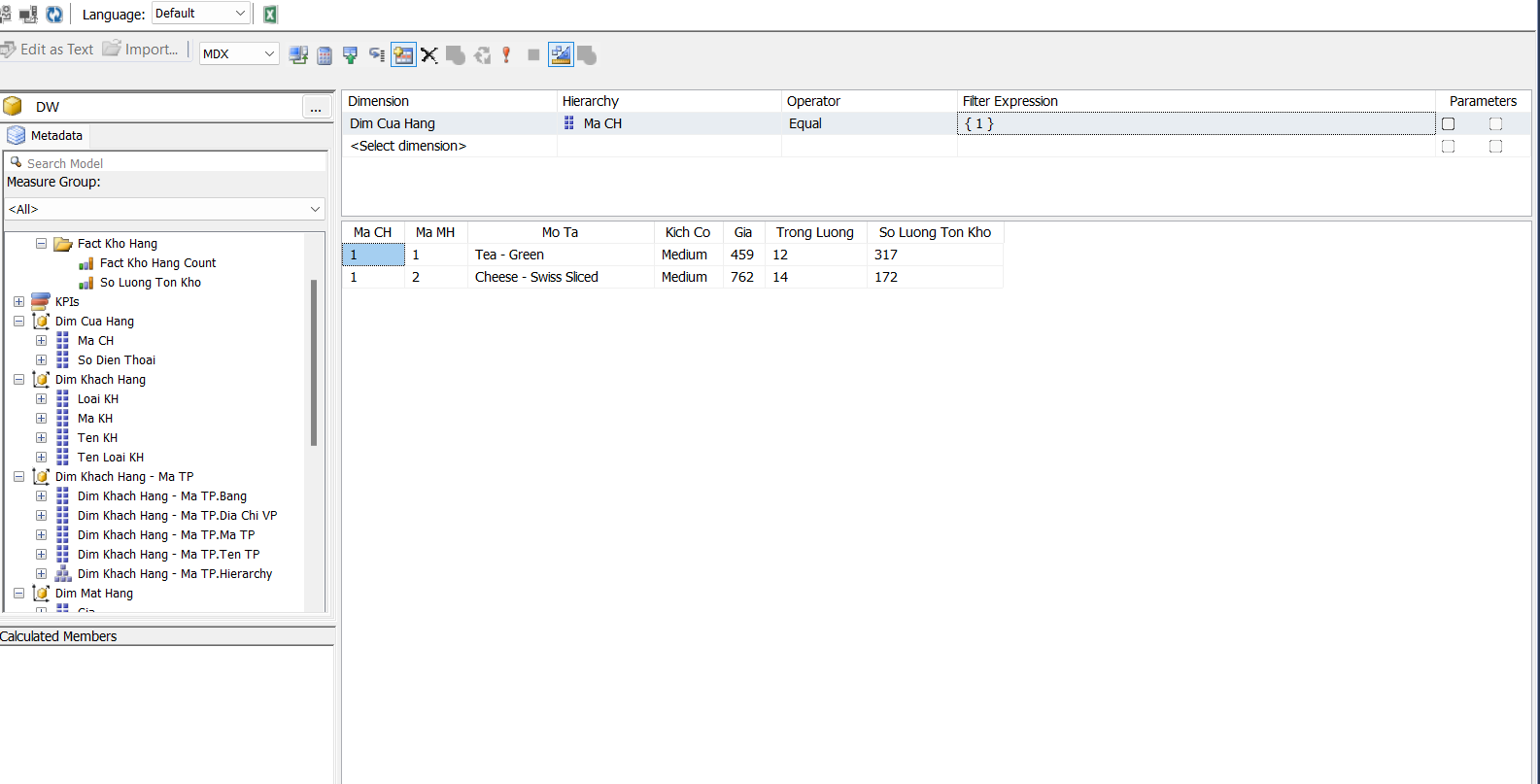




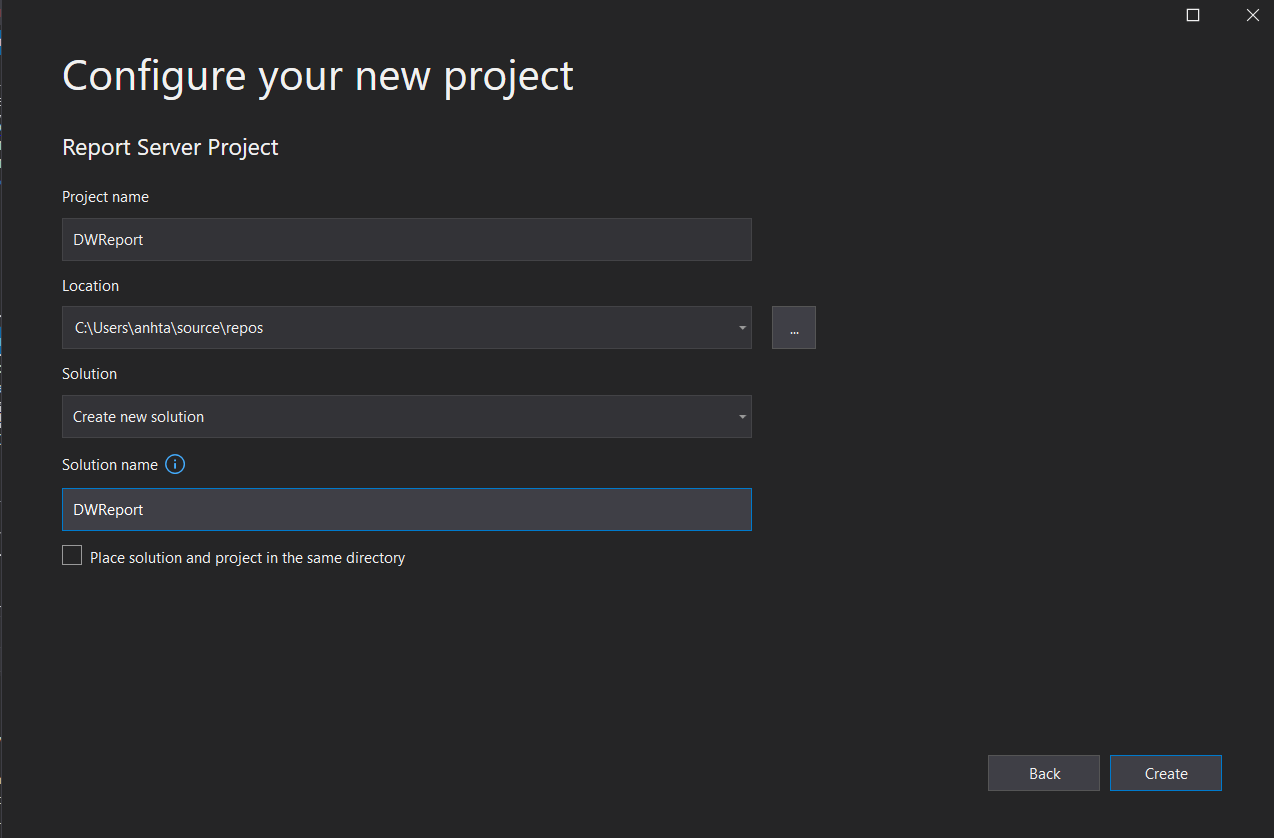


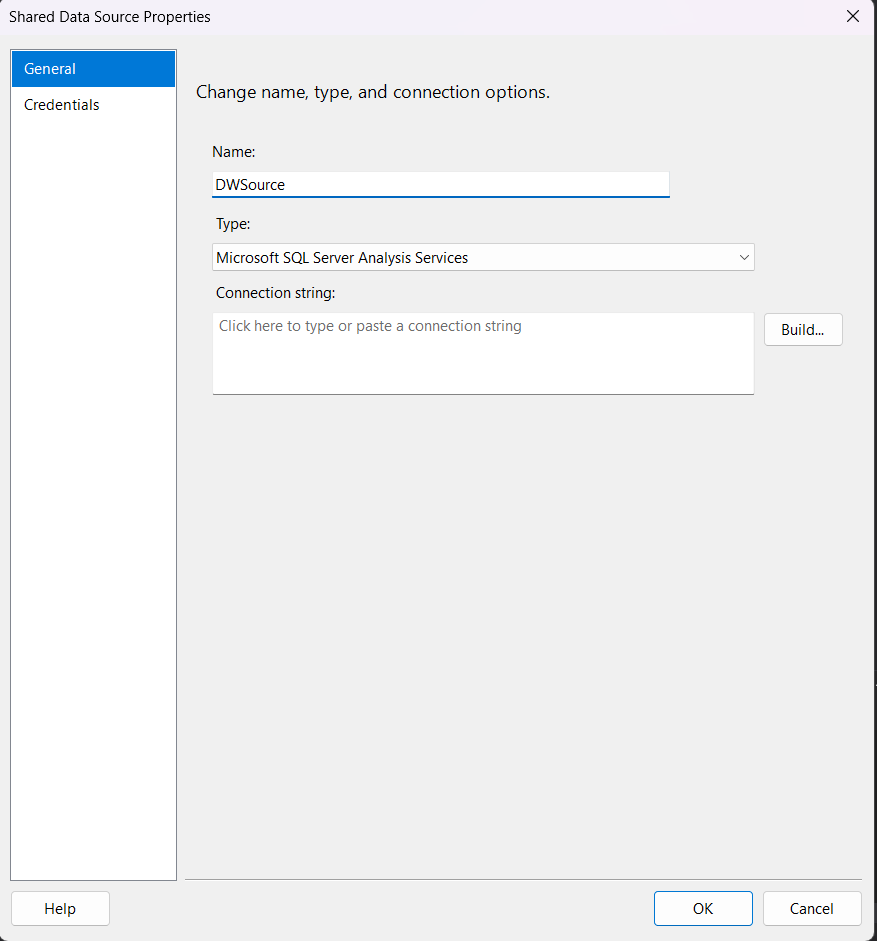


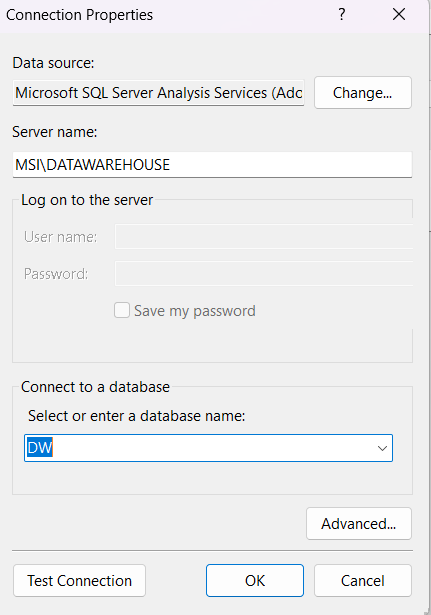
Tìm kiếm số lượng các mặt hàng trong kho của cửa hàng có mã 1 :

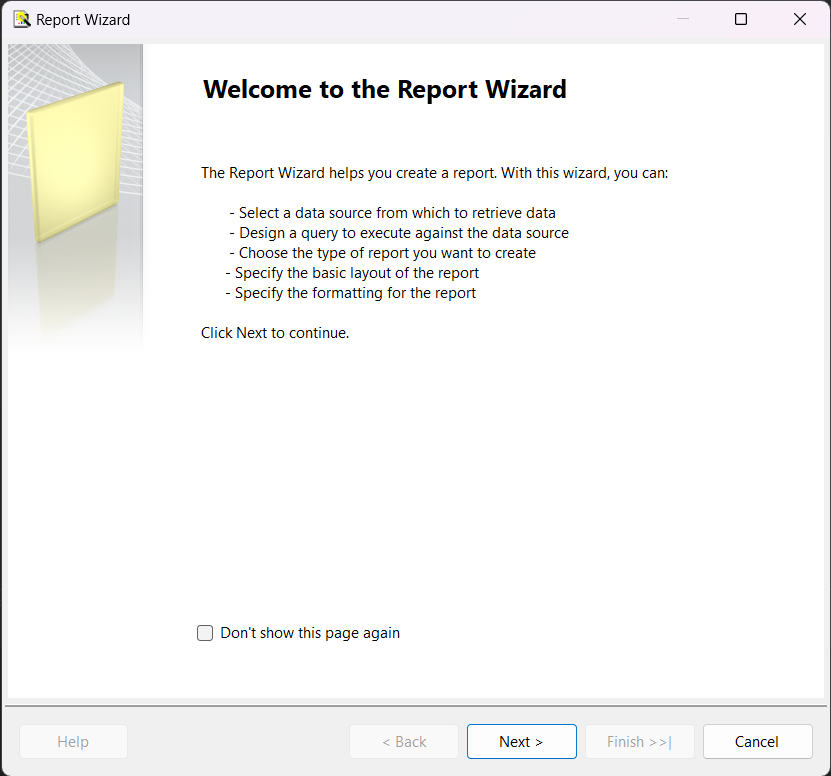


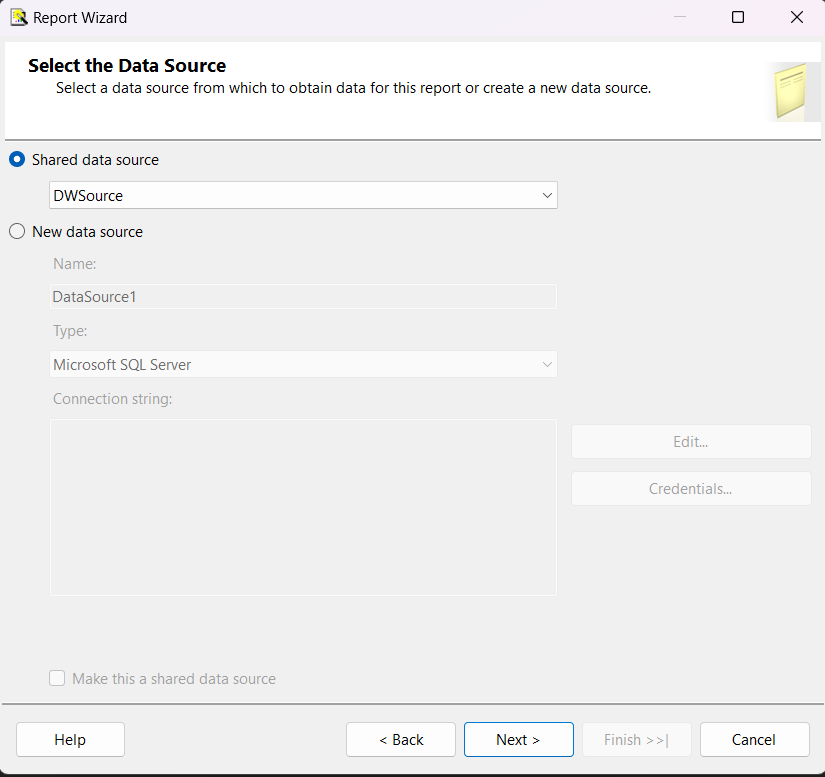
## Các báo cáo phân tích trực tuyến

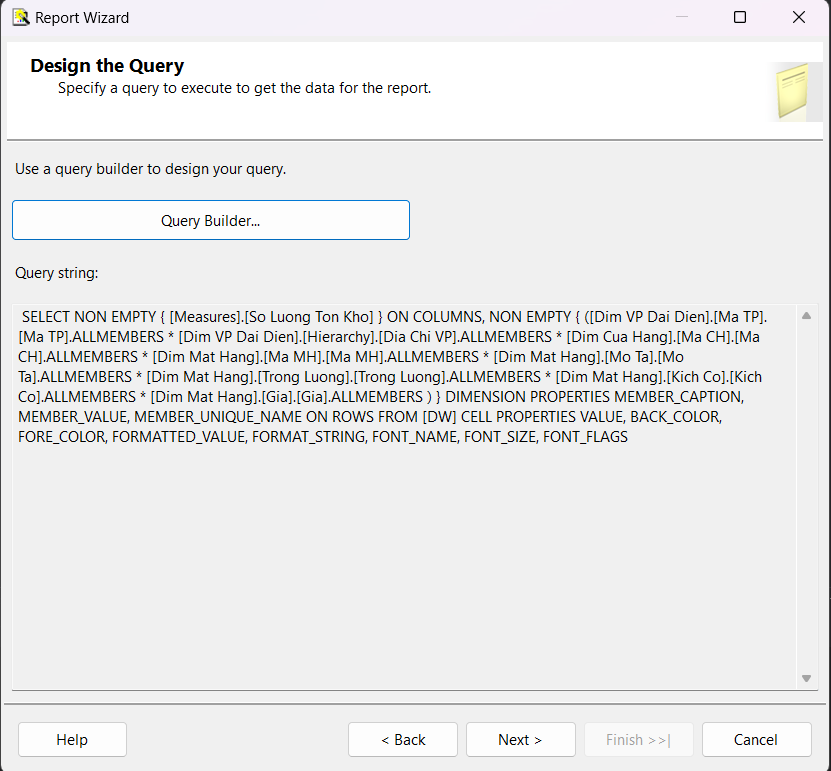






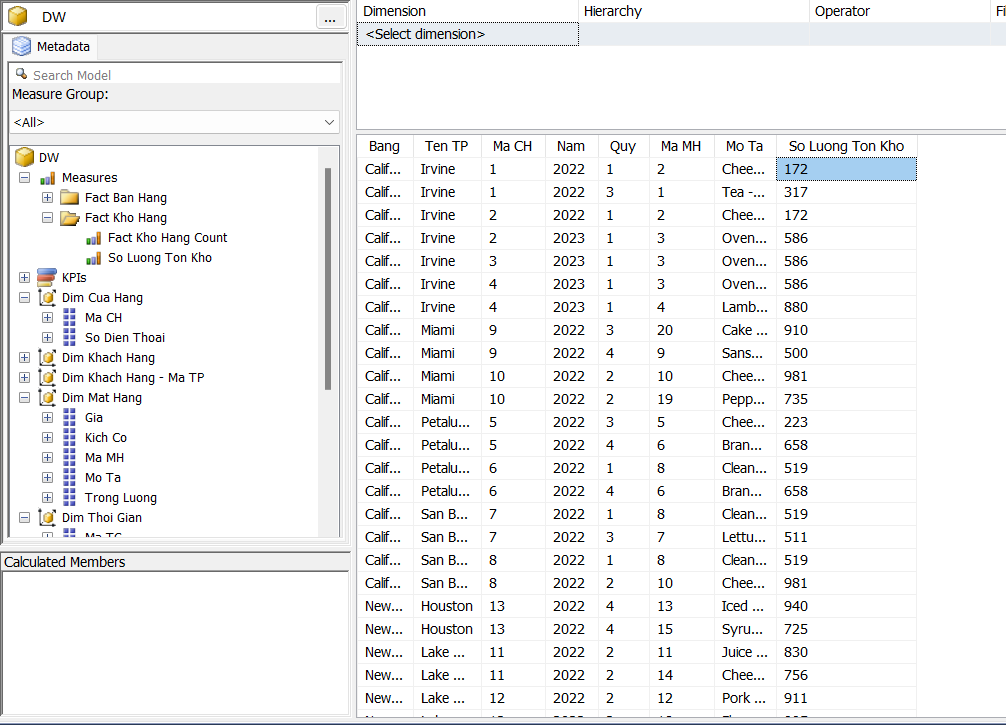








## Kiểm tra tính đúng đắn của dữ liệu



## Kết luận

Những công việc đã hoàn thành:

* Tích hợp đúng dữ liệu từ hai nguồn thành một nguồn
* Thiết kế mô hình dữ liệu được tích hợp
* Sinh dữ liệu cho các nguồn dữ liệu
* Thiết kế mô hình dữ liệu cho kho
* Đổ dữ liệu từ CSDL tích hợp vào kho
* Tạo phân cấp các chiều dữ liệu để tính khối cho OLAP
* Thiết kế các khối dữ liệu để thực hiện OLAP
* Đưa OLAP lên web thực hiện được thao tác khoan sâu xuống, lên
* Viết tài liệu đầy đủ