



Thiết kế Datawarehouse

🕒 Created time	@April 22, 2025 12:56 PM
🕒 Last updated time	@April 23, 2025 11:21 AM

Dưới đây là danh sách các bảng Dimension và Fact cần thiết cho 6 báo cáo cần làm, kèm theo các thuộc tính chi tiết và nguồn dữ liệu từ cơ sở dữ liệu Northwind gốc:

Bảng Dimension (DIM)

1. DimDate

Thuộc tính	Kiểu dữ liệu	Ý nghĩa
DateKey	INT	Khóa chính dạng số (20230101), tối ưu cho JOIN
FullDate	DATE	Ngày định dạng chuẩn cho hiển thị và lọc
Month	TINYINT	Số thứ tự tháng (1-12)
Quarter	TINYINT	Quý trong năm (1-4)
Year	SMALLINT	Năm (ví dụ: 2023)

Câu hỏi: Tại sao có FullDate rồi vẫn cần Month/Quarter/Year?

Trả lời:

- **Truy vấn nhanh hơn:** Khi lọc hoặc nhóm theo tháng/quý/năm, sử dụng trường riêng nhanh hơn nhiều so với dùng hàm trích xuất từ FullDatesql
- **Truy vấn ngắn gọn, dễ đọc**

Nguồn dữ liệu: Không trực tiếp từ bảng nào, được tạo mới bằng stored procedure.

```
CREATE PROCEDURE PopulateDimDate
    @StartDate DATE,
    @EndDate DATE
```

```

AS
BEGIN
    DECLARE @CurrentDate DATE = @StartDate;
    WHILE @CurrentDate <= @EndDate
    BEGIN
        -- Tạo DateKey theo định dạng YYYYMMDD (Ví dụ: 20252204) ⇒ T
        ối ưu hiệu suất truy vấn
        DECLARE @DateKey INT = YEAR(@CurrentDate) * 10000 +
            MONTH(@CurrentDate) * 100 +
            DAY(@CurrentDate);

        -- Chèn dữ liệu vào bảng
        INSERT INTO DimDate (
            DateKey,
            FullDate,
            Month,
            Quarter,
            Year
        )
        VALUES (
            @DateKey,
            @CurrentDate,
            MONTH(@CurrentDate),
            DATEPART(QUARTER, @CurrentDate),
            YEAR(@CurrentDate)
        );

        -- Tăng ngày lên 1
        SET @CurrentDate = DATEADD(DAY, 1, @CurrentDate);
    END;
END;

```

Dựa trên dữ liệu mẫu của Northwind:

Đơn hàng sớm nhất: 1996-07-04 (Orders.OrderDate)

Đơn hàng muộn nhất: 1998-05-06 (Orders.OrderDate)

Áp dụng mở rộng phạm vi:

StartDate = 1995-07-04 (lùi 1 năm từ đơn hàng sớm nhất)

EndDate = 2000-05-06 (thêm 2 năm từ đơn hàng muộn nhất)

⇒ EXEC PopulateDimDate '1995-07-04', '2000-05-06';

Procedure PopulateDimDate chỉ chạy một lần duy nhất khi bạn thiết lập Data \

Đây là đặc điểm quan trọng của bảng DimDate:

- Bảng tĩnh (static table): Được tạo trước và không thay đổi thường xuyên
- Dữ liệu được tạo sẵn: Tất cả các ngày trong nhiều năm được tạo cùng một lúc
- Không phụ thuộc dữ liệu giao dịch: Không cần cập nhật khi có giao dịch mới
- Duy trì định kỳ (hiếm khi): Thường chỉ cần thực hiện mỗi vài năm một lần để

2. DimCustomer

Thuộc tính	Kiểu dữ liệu	Nguồn dữ liệu	Ý nghĩa
CustomerKey	INT	Auto-generated	Khóa chính, tối ưu JOIN với FactSales
CustomerID	NCHAR(5)	Customers.CustomerID	Khóa business để tra cứu ngược
CompanyName	NVARCHAR(40)	Customers.CompanyName	Hiển thị trong báo cáo chi tiết
Country	NVARCHAR(15)	Customers.Country	Phân tích địa lý cấp quốc gia (Báo cáo 4)
Region	NVARCHAR(15)	Customers.Region	Phân tích địa lý cấp vùng (Báo cáo 4)
City	NVARCHAR(15)	Customers.City	Phân tích địa lý cấp thành phố (Báo cáo 4)

Nguồn dữ liệu: Bảng Customers

3. DimProduct

Thuộc tính	Kiểu dữ liệu	Nguồn dữ liệu	Ý nghĩa
ProductKey	INT	Auto-generated	Khóa chính
ProductID	INT	Products.ProductID	Khóa business
ProductName	NVARCHAR(40)	Products.ProductName	Tên sản phẩm

CategoryID	INT	Products.CategoryID	ID danh mục
CategoryName	NVARCHAR(15)	Categories.CategoryName	Tên danh mục (denormalized)
SupplierID	INT	Products.SupplierID	ID nhà cung cấp
SupplierName	NVARCHAR(40)	Suppliers.CompanyName	Tên nhà cung cấp (denormalized)
UnitPrice	MONEY	Products.UnitPrice	Đơn giá bán
UnitsInStock	SMALLINT	Products.UnitsInStock	Số lượng tồn kho
UnitsOnOrder	SMALLINT	Products.UnitsOnOrder	Số lượng đang đặt
ReorderLevel	SMALLINT	Products.ReorderLevel	Mức tái đặt hàng
Discontinued	BIT	Products.Discontinued	Đã ngừng kinh doanh
StockStatus	NVARCHAR(20)	Derived	Trạng thái tồn kho

Nguồn dữ liệu: Bảng Products, Categories và Suppliers.

Tại sao lại đưa **CategoryName** vào **DimProduct**:

Đây là trường hợp đưa thuộc tính từ bảng Categories vào DimProduct (denormalization) để:

1. Đơn giản hóa truy vấn báo cáo doanh số theo danh mục.
2. Hạn chế phải JOIN với bảng DimCategory trong báo cáo. ⇒ Cải thiện hiệu suất truy vấn.

4. DimCategory

Thuộc tính	Kiểu dữ liệu	Nguồn dữ liệu	Ý nghĩa
CategoryKey	INT	Auto-generated	Khóa chính, tối ưu JOIN với DimProduct/FactSales
CategoryID	INT	Categories.CategoryID	Khóa business để tra cứu ngược
CategoryName	NVARCHAR(15)	Categories.CategoryName	Tên danh mục hiển thị trong báo cáo (Báo cáo 3)
Description	NTEXT	Categories.Description	Cung cấp thông tin mô tả đầy đủ về danh mục khi cần chi tiết

Nguồn dữ liệu: Bảng Categories

5. DimEmployee

Thuộc tính	Kiểu dữ liệu	Nguồn dữ liệu	Ý nghĩa
EmployeeKey	INT	Auto-generated	Khóa chính, tối ưu JOIN với FactSales
EmployeeID	INT	Employees.EmployeeID	Khóa business để tra cứu ngược
FullName	NVARCHAR(61)	FirstName + ' ' + LastName	Họ và tên đầy đủ để hiển thị trong báo cáo
Country	NVARCHAR(15)	Employees.Country	Quốc gia - cho phân tích theo khu vực địa lý
City	NVARCHAR(15)	Employees.City	Thành phố - chi tiết hơn cho phân tích địa lý
Region	NVARCHAR(15)	Employees.Region	Vùng/bang - bổ sung cho phân tích khu vực

Nguồn dữ liệu: Bảng Employees

6. DimShipper

Thuộc tính	Kiểu dữ liệu	Nguồn dữ liệu	Ý nghĩa
ShipperKey	INT	Auto-generated	Khóa chính, tối ưu JOIN với FactOrderFulfillment
ShipperID	INT	Shippers.ShipperID	Khóa business để tra cứu ngược
CompanyName	NVARCHAR(40)	Shippers.CompanyName	Tên công ty vận chuyển hiển thị trong báo cáo

Nguồn dữ liệu: Bảng Shippers

7. DimOrder

Thuộc tính	Kiểu dữ liệu	Nguồn dữ liệu	Ý nghĩa
OrderKey	INT	Auto-generated	Khóa chính, tối ưu JOIN với Fact tables
OrderID	INT	Orders.OrderID	Khóa business để tra cứu ngược
CustomerID	NCHAR(5)	Orders.CustomerID	ID khách hàng

EmployeeID	INT	Orders.EmployeeID	ID nhân viên bán hàng
ShipVia	INT	Orders.ShipVia	ID đơn vị vận chuyển
Freight	MONEY	Orders.Freight	Chi phí vận chuyển
OrderDateKey	INT	Derived	Khóa liên kết với DimDate (ngày đặt hàng)
RequiredDateKey	INT	Derived	Khóa liên kết với DimDate (ngày yêu cầu)
ShippedDateKey	INT	Derived	Khóa liên kết với DimDate (ngày giao)
IsOnTime	BIT	Derived	Đánh dấu đơn hàng giao đúng hạn (1) hoặc trễ (0)
DelayDays	INT	Derived	Số ngày giao trễ (nếu trễ); 0 nếu giao đúng hạn

Nguồn dữ liệu: Bảng Order

8. DimOrderDetail

Thuộc tính	Kiểu dữ liệu	Nguồn dữ liệu	Ý nghĩa
OrderDetailKey	INT	Auto-generated	Khóa chính, tối ưu JOIN với Fact tables
OrderID	INT	[Order Details].OrderID	ID đơn hàng
ProductID	INT	[Order Details].ProductID	ID sản phẩm
UnitPrice	MONEY	[Order Details].UnitPrice	Đơn giá bán
Quantity	SMALLINT	[Order Details].Quantity	Số lượng bán
Discount	REAL	[Order Details].Discount	Tỷ lệ giảm giá
LineTotal	MONEY	Derived	Tổng doanh thu (UnitPrice * Quantity * (1-Discount))

Nguồn dữ liệu: Bảng OrderDetail

Bảng Fact (FACT)

1. FactSales (Fact Doanh số bán hàng)

Thuộc tính	Kiểu dữ liệu	Nguồn dữ liệu	Ý nghĩa
SalesKey	INT	Auto-generated	Khóa chính
OrderKey	INT	DimOrder.OrderKey	Khóa ngoại liên kết với DimOrder
OrderDetailKey	INT	DimOrderDetail.OrderDetailKey	Khóa ngoại liên kết với DimOrderDetail
CustomerKey	INT	DimCustomer.CustomerKey	Khóa ngoại liên kết với DimCustomer
ProductKey	INT	DimProduct.ProductKey	Khóa ngoại liên kết với DimProduct
EmployeeKey	INT	DimEmployee.EmployeeKey	Khóa ngoại liên kết với DimEmployee
ShipperKey	INT	DimShipper.ShipperKey	Khóa ngoại liên kết với DimShipper
OrderDateKey	INT	DimOrder.OrderDateKey	Khóa ngoại liên kết với DimDate (ngày đặt hàng)
SalesAmount	MONEY	DimOrderDetail.LineTotal	Tổng doanh thu
Quantity	SMALLINT	DimOrderDetail.Quantity	Số lượng bán

Nguồn dữ liệu: Bảng Orders và [Order Details]

2. FactInventory (Fact Tồn kho)

Thuộc tính	Kiểu dữ liệu	Nguồn dữ liệu	Ý nghĩa
InventoryKey	INT	Auto-generated	Khóa chính
ProductKey	INT	DimProduct.ProductKey	Khóa ngoại đến sản phẩm
DateKey	INT	DimDate.DateKey	Khóa ngoại đến ngày kiểm kê
UnitsInStock	SMALLINT	DimProduct.UnitsInStock	Số lượng tồn kho
UnitsOnOrder	SMALLINT	DimProduct.UnitsOnOrder	Số lượng đang đặt

ReorderLevel	SMALLINT	DimProduct.ReorderLevel	Mức tái đặt hàng
StockStatus	NVARCHAR(20)	DimProduct.StockStatus	Trạng thái tồn kho

Lưu ý: Vì Northwind không lưu lịch sử tồn kho, bảng này sẽ chỉ có dữ liệu tồn kho tại thời điểm hiện tại.

3. Bảng FactDeliveryReliability (Fact Độ tin cậy vận chuyển)

Thuộc tính	Kiểu dữ liệu	Nguồn dữ liệu	Ý nghĩa
DeliveryKey	INT	Auto-generated	Khóa chính
OrderKey	INT	DimOrder.OrderKey	Khóa ngoại liên kết với DimOrder
OrderDateKey	INT	DimOrder.OrderDateKey	Khóa ngoại liên kết với DimDate
RequiredDateKey	INT	DimOrder.RequiredDateKey	Khóa ngoại liên kết với DimDate
ShippedDateKey	INT	DimOrder.ShippedDateKey	Khóa ngoại liên kết với DimDate
ShipperKey	INT	DimShipper.ShipperKey	Khóa ngoại liên kết với đơn vị vận chuyển
IsOnTime	BIT	DimOrder.IsOnTime	Đánh dấu đơn hàng giao đúng hoặc sớm hơn yêu cầu
DelayDays	INT	DimOrder.DelayDays	Số ngày giao trễ

Mối liên hệ giữa các báo cáo và các Fact