



Phân tích dữ liệu nguồn (Northwind)

🕒 Created time	@April 22, 2025 10:29 AM
🕒 Last updated time	@April 22, 2025 12:47 PM

1. Tổng quan về bộ dữ liệu Northwind

Northwind là cơ sở dữ liệu mẫu do Microsoft phát triển, mô phỏng một công ty thương mại có tên "Northwind Traders", chuyên về việc nhập khẩu và bán các loại thực phẩm đặc sản từ khắp nơi trên thế giới.

Cơ sở dữ liệu gồm các thông tin về:

- Sản phẩm và danh mục sản phẩm
- Khách hàng và đơn đặt hàng
- Nhân viên và phân công lãnh thổ
- Nhà cung cấp và đơn vị vận chuyển

Quy mô dữ liệu:

1. **Categories** (8 bản ghi)

- 8 danh mục sản phẩm bao gồm: Beverages, Condiments, Confections, Dairy Products, Grains/Cereals, Meat/Poultry, Produce, Seafood

2. **Customers** (91 bản ghi)

- 91 khách hàng từ nhiều quốc gia khác nhau
- Mỗi khách hàng có ID gồm 5 ký tự

3. **Employees** (9 bản ghi)

- 9 nhân viên với thông tin cá nhân và nghiệp vụ
- Có cấu trúc phân cấp quản lý [ReportsTo - Giá trị của ReportsTo là EmployeeID của một nhân viên khác (người quản lý)]
- Đây là một ví dụ về **quan hệ tự tham chiếu** (self-referencing relationship)
- Trong trường hợp này chỉ có 9 nhân viên nên có thể mô phỏng cấu trúc phân cấp như sau:

```

Andrew Fuller (Vice President, Sales)
├── Nancy Davolio (Sales Representative)
├── Janet Leverling (Sales Representative)
├── Margaret Peacock (Sales Representative)
├── Laura Callahan (Inside Sales Coordinator)
├── Steven Buchanan (Sales Manager)
│   ├── Michael Suyama (Sales Representative)
│   ├── Robert King (Sales Representative)
│   └── Anne Dodsworth (Sales Representative)

```

4. **EmployeeTerritories** (49 bản ghi)

- Liên kết nhiều-nhiều giữa Employees và Territories
- Phân công các nhân viên phụ trách các vùng lãnh thổ

5. **Order Details** (2,155 bản ghi)

- Chi tiết các sản phẩm trong mỗi đơn hàng
- Trung bình mỗi đơn hàng có khoảng 2.6 sản phẩm

6. **Orders** (830 bản ghi)

- 830 đơn hàng từ tháng 7/1996 đến tháng 5/1998
- Bao gồm thông tin về khách hàng, nhân viên và vận chuyển

7. **Products** (77 bản ghi)

- 77 sản phẩm thuộc 8 danh mục
- Có thông tin về giá, tồn kho và nhà cung cấp

8. **Region** (4 bản ghi)

- 4 vùng: Eastern, Western, Northern, Southern
- Dùng để phân chia lãnh thổ bán hàng

9. **Shippers** (3 bản ghi)

- 3 công ty vận chuyển: Speedy Express, United Package, Federal Shipping
- Sử dụng để giao hàng cho khách hàng

10. **Suppliers** (29 bản ghi)

- 29 nhà cung cấp từ nhiều quốc gia
- Cung cấp các sản phẩm cho Northwind

11. **Territories** (53 bản ghi)

- 53 lãnh thổ bán hàng thuộc 4 vùng

- Được phân công cho các nhân viên

Tỷ lệ quan trọng trong dữ liệu:

- Trung bình mỗi khách hàng có khoảng 9.1 đơn hàng (830 đơn hàng / 91 khách hàng)
- Trung bình mỗi nhân viên bán hàng xử lý khoảng 92.2 đơn hàng (830 đơn hàng / 9 nhân viên)
- Trung bình mỗi đơn hàng có 2.6 sản phẩm (2,155 chi tiết đơn hàng / 830 đơn hàng)
- Mỗi sản phẩm xuất hiện trung bình 28 lần trong các đơn hàng (2,155 chi tiết đơn hàng / 77 sản phẩm)

2. Danh sách các bảng trong Northwind Database

Bảng Categories

Lưu trữ thông tin về các danh mục sản phẩm (như đồ uống, gia vị, hải sản...). Giúp phân loại sản phẩm theo nhóm.

Tên cột	Kiểu dữ liệu	Khóa	Null/Rỗng	Mô tả	Lưu ý
CategoryID	Integer	Primary	Không	ID danh mục sản phẩm	Auto-increment, giá trị bắt đầu từ 1
CategoryName	NVarChar(15)		Không	Tên danh mục	Bắt buộc phải có giá trị
Description	NText		Có	Mô tả chi tiết về danh mục	Có thể NULL
Picture	Image		Có	Hình ảnh đại diện cho danh mục	Có thể NULL, dữ liệu nhị phân lớn

Bảng Customers

Chứa thông tin về khách hàng của công ty như tên, địa chỉ, thông tin liên hệ. Đây là bảng quan trọng để theo dõi các đối tác mua hàng.

Tên cột	Kiểu dữ liệu	Khóa	Null/Rỗng	Mô tả	Lưu ý
CustomerID	NChar(5)	Primary	Không	Mã khách hàng	Mã cố định 5 ký tự
CompanyName	NVarChar(40)		Không	Tên công ty khách hàng	Bắt buộc phải có giá trị
ContactName	NVarChar(30)		Có	Tên người liên hệ	Có thể NULL
ContactTitle	NVarChar(30)		Có	Chức danh người liên hệ	Có thể NULL
Address	NVarChar(60)		Có	Địa chỉ công ty	Có thể NULL
City	NVarChar(15)		Có	Thành phố	Có thể NULL
Region	NVarChar(15)		Có	Vùng/Tiểu bang	Thường NULL với khách hàng nước ngoài

Tên cột	Kiểu dữ liệu	Khóa	Null/Rỗng	Mô tả	Lưu ý
PostalCode	NVarChar(10)		Có	Mã bưu điện	Có thể NULL, định dạng khác nhau theo quốc gia
Country	NVarChar(15)		Có	Quốc gia	Có thể NULL nhưng thường có giá trị
Phone	NVarChar(24)		Có	Số điện thoại	Có thể NULL, định dạng không đồng nhất
Fax	NVarChar(24)		Có	Số fax	Thường NULL, ít được sử dụng

Bảng Employees

Lưu trữ dữ liệu nhân viên của công ty, bao gồm thông tin cá nhân, ngày thuê, chức vụ và quan hệ quản lý (ai báo cáo cho ai).

Tên cột	Kiểu dữ liệu	Khóa	Null/Rỗng	Mô tả	Lưu ý
EmployeeID	Integer	Primary	Không	ID nhân viên	Auto-increment
LastName	NVarChar(20)		Không	Họ nhân viên	Bắt buộc phải có giá trị
FirstName	NVarChar(10)		Không	Tên nhân viên	Bắt buộc phải có giá trị
Title	NVarChar(30)		Có	Chức danh công việc	Có thể NULL
TitleOfCourtesy	NVarChar(25)		Có	Danh xưng (Mr., Ms., v.v.)	Có thể NULL
BirthDate	DateTime		Có	Ngày sinh	Có thể NULL
HireDate	DateTime		Có	Ngày thuê	Thường có giá trị, có thể NULL
Address	NVarChar(60)		Có	Địa chỉ nhà	Có thể NULL
City	NVarChar(15)		Có	Thành phố	Có thể NULL
Region	NVarChar(15)		Có	Vùng/Tiểu bang	Thường NULL với nhân viên nước ngoài
PostalCode	NVarChar(10)		Có	Mã bưu điện	Có thể NULL
Country	NVarChar(15)		Có	Quốc gia	Có thể NULL
HomePhone	NVarChar(24)		Có	Số điện thoại nhà	Có thể NULL
Extension	NVarChar(4)		Có	Số máy nhánh	Có thể NULL
Photo	Image		Có	Ảnh nhân viên	Có thể NULL, dữ liệu nhị phân lớn
Notes	NText		Có	Ghi chú về nhân viên	Có thể NULL, văn bản dài

Tên cột	Kiểu dữ liệu	Khóa	Null/Rỗng	Mô tả	Lưu ý
ReportsTo	Integer	Foreign	Có	ID của quản lý trực tiếp	NULL cho người đứng đầu, FK đến EmployeeID
PhotoPath	NVarChar(255)		Có	Đường dẫn đến file ảnh	Có thể NULL

Bảng Orders

Lưu trữ thông tin tổng quan về đơn hàng như mã khách hàng, nhân viên phụ trách, ngày đặt hàng, ngày giao hàng, địa chỉ giao hàng.

Tên cột	Kiểu dữ liệu	Khóa	Null/Rỗng	Mô tả	Lưu ý
OrderID	Integer	Primary	Không	ID đơn hàng	Auto-increment
CustomerID	NChar(5)	Foreign	Có	Mã khách hàng	FK đến Customers.CustomerID, hiếm khi NULL
EmployeeID	Integer	Foreign	Có	ID nhân viên xử lý đơn hàng	FK đến Employees.EmployeeID, hiếm khi NULL
OrderDate	DateTime		Có	Ngày đặt hàng	Có thể NULL, nhưng thường có giá trị
RequiredDate	DateTime		Có	Ngày yêu cầu giao hàng	Có thể NULL
ShippedDate	DateTime		Có	Ngày thực tế giao hàng	NULL nếu chưa giao hàng
ShipVia	Integer	Foreign	Có	ID đơn vị vận chuyển	FK đến Shippers.ShipperID
Freight	Money		Có	Phí vận chuyển	Mặc định = 0 nếu NULL
ShipName	NVarChar(40)		Có	Tên người nhận hàng	Thường lấy từ thông tin khách hàng
ShipAddress	NVarChar(60)		Có	Địa chỉ giao hàng	Có thể NULL
ShipCity	NVarChar(15)		Có	Thành phố giao hàng	Có thể NULL
ShipRegion	NVarChar(15)		Có	Vùng/Tiểu bang giao hàng	Thường NULL với địa chỉ nước ngoài
ShipPostalCode	NVarChar(10)		Có	Mã bưu điện nơi giao hàng	Có thể NULL
ShipCountry	NVarChar(15)		Có	Quốc gia giao hàng	Có thể NULL

Bảng Order Details

Chứa chi tiết của từng đơn hàng: sản phẩm nào, số lượng, giá bán, chiết khấu. Đây là bảng liên kết giữa Orders và Products.

Tên cột	Kiểu dữ liệu	Khóa	Null/Rỗng	Mô tả	Lưu ý
OrderID	Integer	Primary/Foreign	Không	ID đơn hàng	FK đến Orders.OrderID
ProductID	Integer	Primary/Foreign	Không	ID sản phẩm	FK đến Products.ProductID
UnitPrice	Money		Không	Đơn giá	Không được < 0, không phải lúc nào cũng bằng giá trong Products
Quantity	SmallInt		Không	Số lượng	Mặc định = 1, không được ≤ 0
Discount	Real		Không	Tỷ lệ chiết khấu (0-1)	Mặc định = 0, giới hạn từ 0 đến 1

Bảng Products

Chứa thông tin về các sản phẩm: tên, nhà cung cấp, danh mục, giá, số lượng tồn kho, mức tồn kho cần đặt hàng lại.

Tên cột	Kiểu dữ liệu	Khóa	Null/Rỗng	Mô tả	Lưu ý
ProductID	Integer	Primary	Không	ID sản phẩm	Auto-increment
ProductName	NVarChar(40)		Không	Tên sản phẩm	Bắt buộc phải có giá trị
SupplierID	Integer	Foreign	Có	ID nhà cung cấp	FK đến Suppliers.SupplierID
CategoryID	Integer	Foreign	Có	ID danh mục sản phẩm	FK đến Categories.CategoryID
QuantityPerUnit	NVarChar(20)		Có	Quy cách đóng gói	VD: "10 boxes x 20 bags", có thể NULL
UnitPrice	Money		Có	Đơn giá	Mặc định = 0 nếu NULL, ≥ 0
UnitsInStock	SmallInt		Có	Số lượng tồn kho	Mặc định = 0 nếu NULL, ≥ 0
UnitsOnOrder	SmallInt		Có	Số lượng đang đặt	Mặc định = 0 nếu NULL, ≥ 0
ReorderLevel	SmallInt		Có	Ngưỡng cần đặt hàng lại	Mặc định = 0 nếu NULL, ≥ 0
Discontinued	Bit		Không	Đã ngừng kinh doanh	Mặc định = 0 (false)

Bảng Shippers

Lưu trữ thông tin về các công ty vận chuyển đối tác của Northwind.

Tên cột	Kiểu dữ liệu	Khóa	Null/Rỗng	Mô tả	Lưu ý
ShipperID	Integer	Primary	Không	ID đơn vị vận chuyển	Auto-increment
CompanyName	NVarChar(40)		Không	Tên công ty vận chuyển	Bắt buộc phải có giá trị
Phone	NVarChar(24)		Có	Số điện thoại	Có thể NULL

Bảng Suppliers

Chứa dữ liệu về các nhà cung cấp sản phẩm cho công ty, bao gồm tên công ty, người liên hệ, địa chỉ và thông tin liên lạc.

Tên cột	Kiểu dữ liệu	Khóa	Null/Rỗng	Mô tả	Lưu ý
SupplierID	Integer	Primary	Không	ID nhà cung cấp	Auto-increment
CompanyName	NVarChar(40)		Không	Tên công ty nhà cung cấp	Bắt buộc phải có giá trị
ContactName	NVarChar(30)		Có	Tên người liên hệ	Có thể NULL
ContactTitle	NVarChar(30)		Có	Chức danh người liên hệ	Có thể NULL
Address	NVarChar(60)		Có	Địa chỉ công ty	Có thể NULL
City	NVarChar(15)		Có	Thành phố	Có thể NULL
Region	NVarChar(15)		Có	Vùng/Tiểu bang	Thường NULL với nhà cung cấp nước ngoài
PostalCode	NVarChar(10)		Có	Mã bưu điện	Có thể NULL
Country	NVarChar(15)		Có	Quốc gia	Có thể NULL
Phone	NVarChar(24)		Có	Số điện thoại	Có thể NULL
Fax	NVarChar(24)		Có	Số fax	Có thể NULL
HomePage	NText		Có	Website nhà cung cấp	Có thể NULL, văn bản dài

Bảng Region

Chứa thông tin về các vùng địa lý (thường là để phân loại thị trường).

Tên cột	Kiểu dữ liệu	Khóa	Null/Rỗng	Mô tả	Lưu ý
RegionID	Integer	Primary	Không	ID vùng	Không phải auto-increment
RegionDescription	NChar(50)		Không	Mô tả vùng	Bắt buộc phải có giá trị

Bảng Territories

Lưu trữ các khu vực kinh doanh nhỏ hơn, thuộc về các vùng lớn.

Tên cột	Kiểu dữ liệu	Khóa	Null/Rỗng	Mô tả	Lưu ý
TerritoryID	NVarChar(20)	Primary	Không	ID lãnh thổ	Không phải auto-increment
TerritoryDescription	NChar(50)		Không	Mô tả lãnh thổ	Bắt buộc phải có giá trị
RegionID	Integer	Foreign	Không	ID vùng	FK đến Region.RegionID

Bảng EmployeeTerritories

Bảng liên kết giữa nhân viên và khu vực, cho biết nhân viên nào phụ trách khu vực nào.

Tên cột	Kiểu dữ liệu	Khóa	Null/Rỗng	Mô tả	Lưu ý
EmployeeID	Integer	Primary/Foreign	Không	ID nhân viên	FK đến Employees.EmployeeID
TerritoryID	NVarChar(20)	Primary/Foreign	Không	ID lãnh thổ	FK đến Territories.TerritoryID

3. Northwind Database Diagram

