**BỘ GIÁO DỤC VÀ ĐÀO TẠO**

**TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT TP. HCM**

**KHOA CÔNG NGHỆ THÔNG TIN**

✩✩✩✩✩

****

**MÔN HỌC: KHO DỮ LIỆU**

**BÁO CÁO CUỐI KỲ**

**ĐỀ TÀI: XÂY DỰNG KHO DỮ LIỆU VỂ**

**SUPERMART GROCERY SALES**

**GVHD: Nguyễn Văn Thành**

**SVTH:**   **MSSV**

Phạm Trung Nghĩa 20133073

Ngô Hoàng Khắc Tường 20133112

Nguyễn Thị Tường Vi 20133113

**Lớp thứ 5\_Tiết 1-4**

**Tp.Hồ Chí Minh, tháng 5 năm 2023**

**ĐÁNH GIÁ & CHẤM ĐIỂM**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **STT** | **THÀNH VIÊN** | **ĐÁNH GIÁ** | **ĐIỂM** | **KÝ TÊN** |
| 1 | Phạm Trung Nghĩa |  |  |  |
| 2 | Nguyễn Thị Tường Vi |  |  |  |
| 3 | Ngô Hoàng Khắc Tường |  |  |  |

**NHẬN XÉT**

*Ký tên*

**Nguyễn Văn Thành**

**LỜI CẢM ƠN**

*Đầu tiên, nhóm em xin chân thành gửi lời cảm ơn với Thầy Nguyễn Văn Thành – Giảng viên phụ trách môn Khoi Dữ Liệu – Trường Đại học Sư phạm kỹ thuật TP.HCM.*

*Trong thời gian học tập, nhóm em nhận được sự nhiều sự giúp đỡ từ Thầy. Thầy đã cung cấp cho chúng em đầy đủ kiến thức, chỉ bảo và đóng góp những ý kiến quý báu giúp chúng em có thêm hiểu biết để hoàn thành được báo cáo cuối kì cho môn học. Thầy đã cung cấp tài liệu và hướng dẫn tận tình cho chúng em trong suốt quá trình học tập. Thầy luôn nhiệt tình vui vẻ chỉ dạy, chia sẻ giúp chúng em trong môn học. Chúng em rất trân quý sự tâm huyết và trách nhiệm của Thầy trong công việc giảng dạy và truyền đạt kiến thức.*

*Trong quá trình thực hiện đồ án, dựa trên kiến thức được Thầy cung cấp qua các buổi học lý thuyết cũng như thực hành trên lớp, kết hợp với việc tự tìm hiểu những công cụ và kiến thức mới , nhóm đã cố gắng thực hiện đồ án một cách tốt nhất .Tuy nhiên, đồ án còn chưa được hoàn hiện và có thể còn có sai sót.*

*Nhóm rất mong nhận được sự góp ý từ Thầy nhằm rút ra những kinh nghiệm quý báu và hoàn thiện vốn kiến thức để nhóm có thể hoàn thành những đồ án, dự án khác trong tương lai .*

***Nhóm chúng em xin chân thành cảm ơn Thầy!***

MỤC LỤC

[CHƯƠNG 1: LÝ DO CHỌN ĐỀ TÀI VÀ GIỚI THIỆU VỀ DATASET 1](#_Toc135251076)

[1.1. Lý do chọn đề tài 1](#_Toc135251077)

[1.2. Giới thiệu về dataset 1](#_Toc135251078)

[1.3. Giới thiệu về công cụ và thư viện sử dụng 2](#_Toc135251079)

[1.3.1 Visual Studio 2019 có tích hợp SQL Server Datatool 2](#_Toc135251080)

[1.3.2 SQL Server 3](#_Toc135251081)

[1.3.3 Power Bi 4](#_Toc135251082)

[CHƯƠNG 2: THIẾT KẾ XÂY DỰNG CƠ SỞ DỮ LIỆU TÁC NGHIỆP(OLAP) 5](#_Toc135251083)

[2.1. Business process 5](#_Toc135251084)

[2.2. Dữ liệu kế bảng Dim và Fact 5](#_Toc135251085)

[CHƯƠNG 3: TÍCH HỢP DỮ LIỆU VÀO KHO DỮ LIỆU (SSIS) 8](#_Toc135251086)

[3.1. Tạo mới Project 8](#_Toc135251087)

[3.2. Thiết lập các connection 10](#_Toc135251092)

[3.3. Đổ dữ liệu từ Excel vào kho dữ liệu 12](#_Toc135251093)

[3.4. Thực thi project 22](#_Toc135251094)

[CHƯƠNG 4: PHÂN TÍCH DỮ LIỆU (SSAS) 23](#_Toc135251095)

[4.1 Tạo mới project 23](#_Toc135251096)

[4.2 Tạo datasource 24](#_Toc135251097)

[4.3 Tạo datasource view 28](#_Toc135251098)

[4.4 Tạo datacubes 30](#_Toc135251099)

[CHƯƠNG 5: KHAI PHÁ DỮ LIỆU (POWER BI) 39](#_Toc135251100)

[5.1. Dash board 39](#_Toc135251101)

[5.2. Biểu đồ 41](#_Toc135251102)

[5.2.1 Sales By Category 41](#_Toc135251103)

[5.2.2 Sales By SubCategory 42](#_Toc135251104)

[5.2.3 Profit By Category 43](#_Toc135251105)

[5.2.4 Profit By SubCategory 43](#_Toc135251106)

[5.2.5 Profit And Sales By Region 44](#_Toc135251107)

[5.2.6 Sales Of Region By Year 45](#_Toc135251108)

[5.2.7 Discount and Sales per City 45](#_Toc135251109)

[5.2.8 Discount and Sales per Region 46](#_Toc135251110)

[5.2.9 Total sales of Category per Year 46](#_Toc135251111)

[5.2.10 Total sales by Customer per Month 47](#_Toc135251112)

[TÀI LIỆU THAM KHẢO 48](#_Toc135251113)

[KẾT LUẬN 49](#_Toc135251114)

# CHƯƠNG 1: LÝ DO CHỌN ĐỀ TÀI VÀ GIỚI THIỆU VỀ DATASET

## Lý do chọn đề tài

Trong thời đại ngày, các hình thức mua bán, kinh doanh là các hoạt động phổ biến – có thể bắt gặp ờ mọi nơi trong đời sống của chúng ta. Nhưng việc phân tích các dữ liệu từ hoạt động này chỉ mới được quan tâm trong những năm gần đây. Nó đem lại nhiều lợi ích vô giá, từ việc quản lí các hoạt động mua bán, quản lý doanh thu, lợi nhuận đến việc đưa ra các chiến lược kinh doanh mới,… Từ dó chúng ta thấy việc quản lý và phân tích dữ liệu mua bán là hết sức cần thiết. Do do nhóm em đã quyết định chọn đề tài xây dựng kho dữ liệu quản lý dữ liệu bán hàng.

## Giới thiệu về dataset

Nhóm lấy nguồn dữ liệu từ trang: [Supermart Grocery Sales](https://www.kaggle.com/datasets/mohamedharris/supermart-grocery-sales-retail-analytics-dataset)

Đây là một tập dữ liệu hư cấu được tạo ra để giúp các nhà phân tích dữ liệu thực hành phân tích dữ liệu khám phá và trực quan hóa dữ liệu. Tập dữ liệu có dữ liệu về các đơn đặt hàng của khách hàng trên một ứng dụng giao hàng tạp hóa. Tập dữ liệu được thiết kế với giả định rằng các đơn đặt hàng được đặt bởi khách hàng sống ở bang Tamil Nadu, Ấn Độ.

Tập dữ liệu gồm: 11 cột, 9994 dòng

|  |  |
| --- | --- |
| OrderID | ID đơn hàng |
| CustomerName | Tên khách hàng |
| Category | Tên danh mục sản phẩm |
| SubCategory | Tên danh mục phụ |
| City | Thành phố |
| Order Date | Ngày đặt hàng |
| Region | Khu vực |
| Sales | Tổng tiền hoá đơn |
| Discount | Giảm giá |
| Profit | Lợi nhuận thu được |
| State | Bang |

## Giới thiệu về công cụ và thư viện sử dụng

### Visual Studio 2019 có tích hợp SQL Server Datatool

Visual Studio 2019 là một môi trường phát triển tích hợp (Integrated Development Environment - IDE) mạnh mẽ được phát triển bởi Microsoft. Ngoài việc hỗ trợ phát triển ứng dụng trên nhiều nền tảng và ngôn ngữ, Visual Studio 2019 cũng tích hợp SQL Server Data Tools (SSDT) để hỗ trợ quản lý cơ sở dữ liệu.

* SQL Server Data Tools (SSDT) là một bộ công cụ cho phép bạn phát triển cơ sở dữ liệu trong môi trường Visual Studio. Khi tích hợp SSDT vào Visual Studio 2019, bạn có thể thực hiện các công việc sau:
* Thiết kế cơ sở dữ liệu: Bạn có thể sử dụng các công cụ trực quan của SSDT để tạo và chỉnh sửa cấu trúc cơ sở dữ liệu. Các đối tượng như bảng, view, stored procedure và function có thể được tạo và quản lý dễ dàng.
* Quản lý phiên bản và hợp nhất: Visual Studio 2019 tích hợp với các hệ thống quản lý phiên bản như Git và Team Foundation Server, cho phép bạn quản lý lịch sử thay đổi của cơ sở dữ liệu và hợp nhất các phiên bản khác nhau. Điều này giúp đảm bảo rằng các thay đổi cơ sở dữ liệu được kiểm soát và duy trì đồng bộ trong quá trình phát triển ứng dụng.
* Triển khai cơ sở dữ liệu: SSDT cho phép bạn triển khai cấu trúc cơ sở dữ liệu và dữ liệu tương ứng từ môi trường phát triển sang môi trường sản xuất. Bạn có thể tự động triển khai các phiên bản mới nhất của cơ sở dữ liệu và đảm bảo rằng sự thay đổi được áp dụng một cách đáng tin cậy.
* Kiểm tra và gỡ lỗi cơ sở dữ liệu: SSDT cung cấp các công cụ kiểm tra và gỡ lỗi giúp bạn xác nhận tính đúng đắn của mã SQL và tìm ra lỗi trong quá trình phát triển. Bạn có thể thực hiện kiểm tra cú pháp, kiểm tra logic và kiểm tra hiệu năng để đảm bảo chất lượng của cơ sở dữ liệu.

Tổng quan về tích hợp SQL Server Data Tools trong Visual Studio 2019 cho thấy Visual Studio không chỉ là một môi trường phát triển ứng dụng mạnh mẽ, mà còn cung cấp công cụ và tính năng

### SQL Server

SQL Server là một hệ quản trị cơ sở dữ liệu quan hệ (Relational Database Management System (RDBMS) ) sử dụng câu lệnh SQL (Transact-SQL) để trao đổi dữ liệu giữa máy Client và máy cài SQL Server. Một RDBMS bao gồm databases, database engine và các ứng dụng dùng để quản lý dữ liệu và các bộ phận khác nhau trong RDBMS.

SQL Server được tối ưu để có thể chạy trên môi trường cơ sở dữ liệu rất lớn (Very Large Database Environment) lên đến Tera-Byte và có thể phục vụ cùng lúc cho hàng ngàn user. SQL Server có thể kết hợp “ăn ý” với các server khác như Microsoft Internet Information Server (IIS), E-Commerce Server, Proxy Server….

Một vài ấn bản SQL Server:

* Enterprise : chứa tất cả cá đặc điểm nổi bật của SQL Server, bao gồm nhân bộ máy cơ sở dữ liệu và các dịch vụ đi kèm cùng với các công cụ cho tạo và quản lý phân cụm SQL Server. Nó có thể quản lý các CSDL lớn tới 524 petabytes và đánh địa chỉ 12 terabytes bộ nhớ và hỗ trợ tới 640 bộ vi xử lý(các core của cpu)
* Standard : Rất thích hợp cho các công ty vừa và nhỏ vì giá thành rẻ hơn nhiều so với Enterprise Edition, nhưng lại bị giới hạn một số chức năng cao cấp (advanced features) khác, edition này có thể chạy tốt trên hệ thống lên đến 4 CPU và 2 GB RAM.
* Developer : Có đầy đủ các tính năng của Enterprise Edition nhưng được chế tạo đặc biệt như giới hạn số lượng người kết nối vào Server cùng một lúc…. Ðây là phiên bản sử dụng cho phát triển và kiểm tra ứng dụng. Phiên bản này phù hợp cho các cá nhân, tổ chức xây dựng và kiểm tra ứng dụng
* Workgroup: ấn bản SQL Server Workgroup bao gồm chức năng lõi cơ sở dữ liệu nhưng không có các dịch vụ đi kèm. Chú ý phiên bản này không còn tồn tại ở SQL Server 2012.
* Express : SQL Server Express dễ sử dụng và quản trị cơ sở dữ liệu đơn giản. Được tích hợp với Microsoft Visual Studio, nên dễ dàng để phát triển các ứng dụng dữ liệu, an toàn trong lưu trữ, và nhanh chóng triển khai. SQL Server Express là phiên bản miễn phí, không giới hạn về số cơ ở dữ liệu hoặc người sử dụng, nhưng nó chỉ dùng cho 1 bộ vi xử lý với 1 GB bộ nhớ và 10 GB file cơ sở dữ liệu. SQL Server Express là lựa chọn tốt cho những người dùng chỉ cần một phiên bản SQL Server 2005 nhỏ gọn, dùng trên máy chủ có cấu hình thấp, những nhà phát triển ứng dụng không chuyên hay những người yêu thích xây dựng các ứng dụng nhỏ.

### Power Bi

Microsoft Power BI là một sản phẩm phần mềm trực quan hóa dữ liệu tương tác được phát triển bởi Microsoft với trọng tâm chính là kinh doanh thông minh (BI). Power BI là một phần của Microsoft Power Platform. Power BI là tập hợp các dịch vụ phần mềm, ứng dụng và trình kết nối, hoạt động cùng nhau để biến các nguồn dữ liệu không liên quan thành thông tin chi tiết, mạch lạc, trực quan. Dữ liệu có thể được nhập bằng cách đọc trực tiếp từ cơ sở dữ liệu, trang web hoặc các tệp có cấu trúc như bảng tính, CSV, XML và JSON.

Power BI cung cấp các dịch vụ kinh doanh thông minh dựa trên điện toán đám mây, được gọi là "Dịch vụ Power BI", cùng với giao diện dựa trên máy tính để bàn, được gọi là "Power BI Desktop". Microsoft Power BI cung cấp khả năng kho dữ liệu bao gồm chuẩn bị dữ liệu, khai phá dữ liệu và bảng điều khiển tương tác. Vào tháng 3 năm 2016, Microsoft đã phát hành một dịch vụ bổ sung có tên là Power BI Embedded trên nền tảng Azure của mình. Một điểm khác biệt chính của sản phẩm là khả năng tải hình ảnh trực quan tùy chỉnh.

# CHƯƠNG 2: THIẾT KẾ XÂY DỰNG CƠ SỞ DỮ LIỆU TÁC NGHIỆP(OLAP)

## Business process

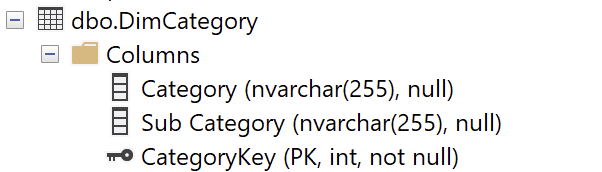
Sale reporting:

* Doanh số của cửa hàng theo thời gian, khu vực, danh mục?
* Lợi nhuận của cửa hàng theo thời gian, khu vực, danh mục?
* Giảm giá theo thời gian, khu vực, danh mục?
* Danh mục sản phẩm nào bán nhiều nhất, ít nhất?
* Khu vực nào có doanh số lợi nhuận nhiều nhất?
* Khách hàng nào mua nhiều nhất ở từng khu vực?
* Khu vực nào có tổng giảm giá nhiều nhất, so sánh giảm giá có ảnh hưởng nhiều đến lợi nhuận của khu vực đó ( giảm giá có giúp khu vực bán chạy hơn hay không?)
* Các yếu tố ảnh hưởng đến lợi nhuận ( vị trí, giảm giá?)

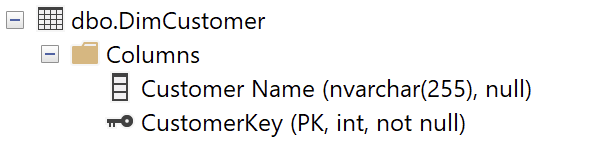
## Dữ liệu kế bảng Dim và Fact

Dùng SQL server để tạo các bảng Dim và Fact

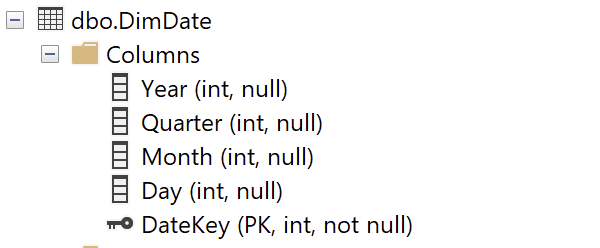
Dim Category



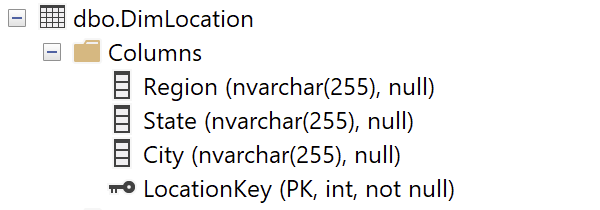
DimCustomer



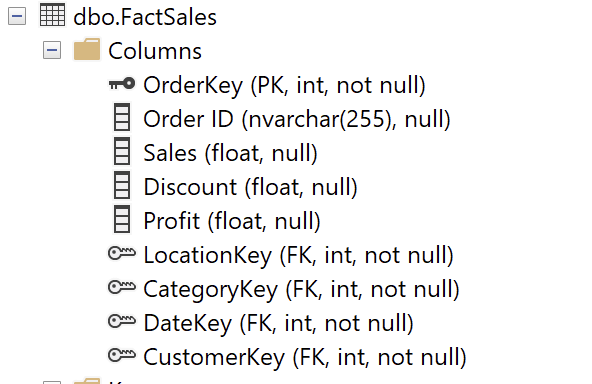
DimDate

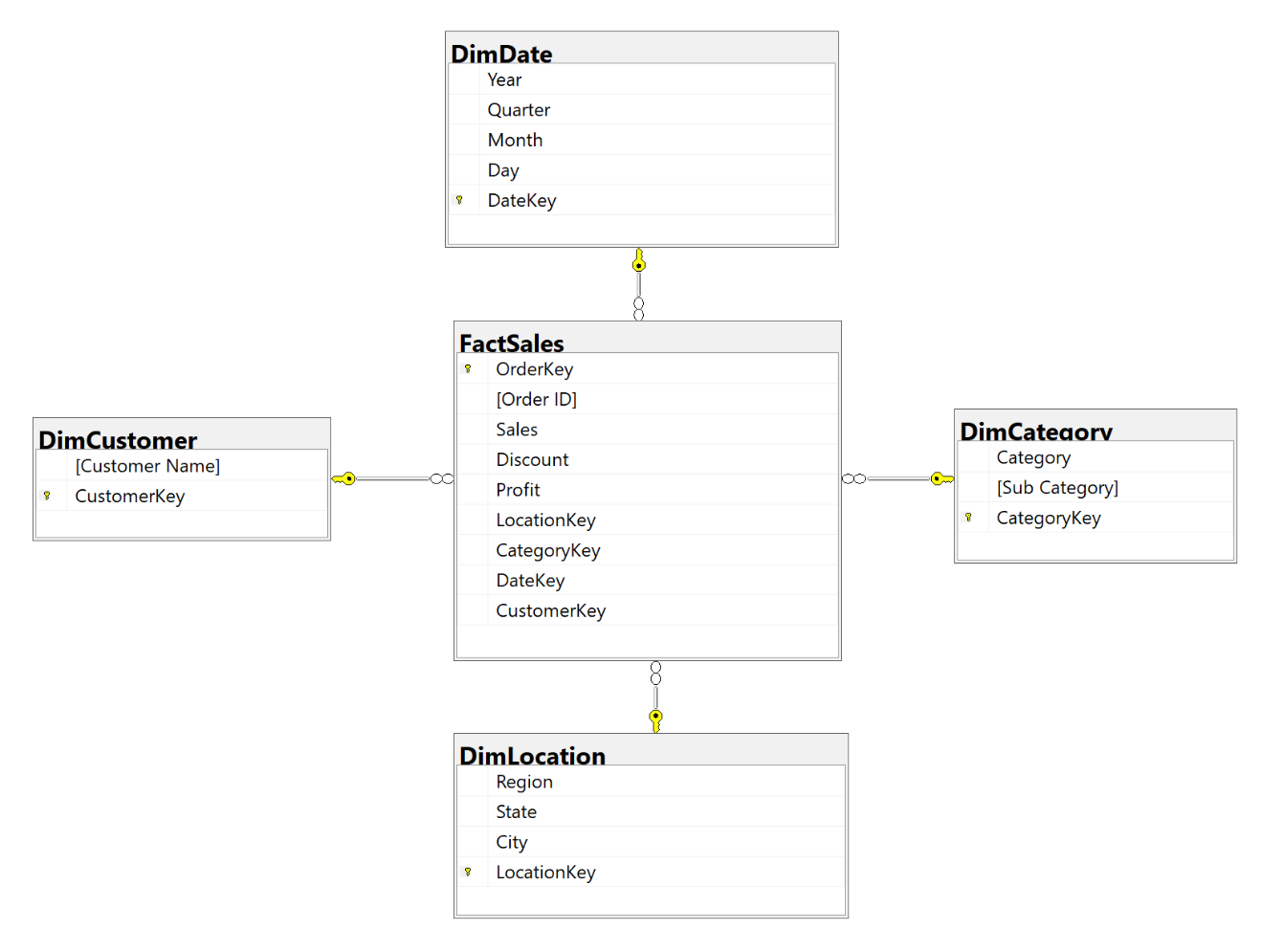


DimLocation



FactSales

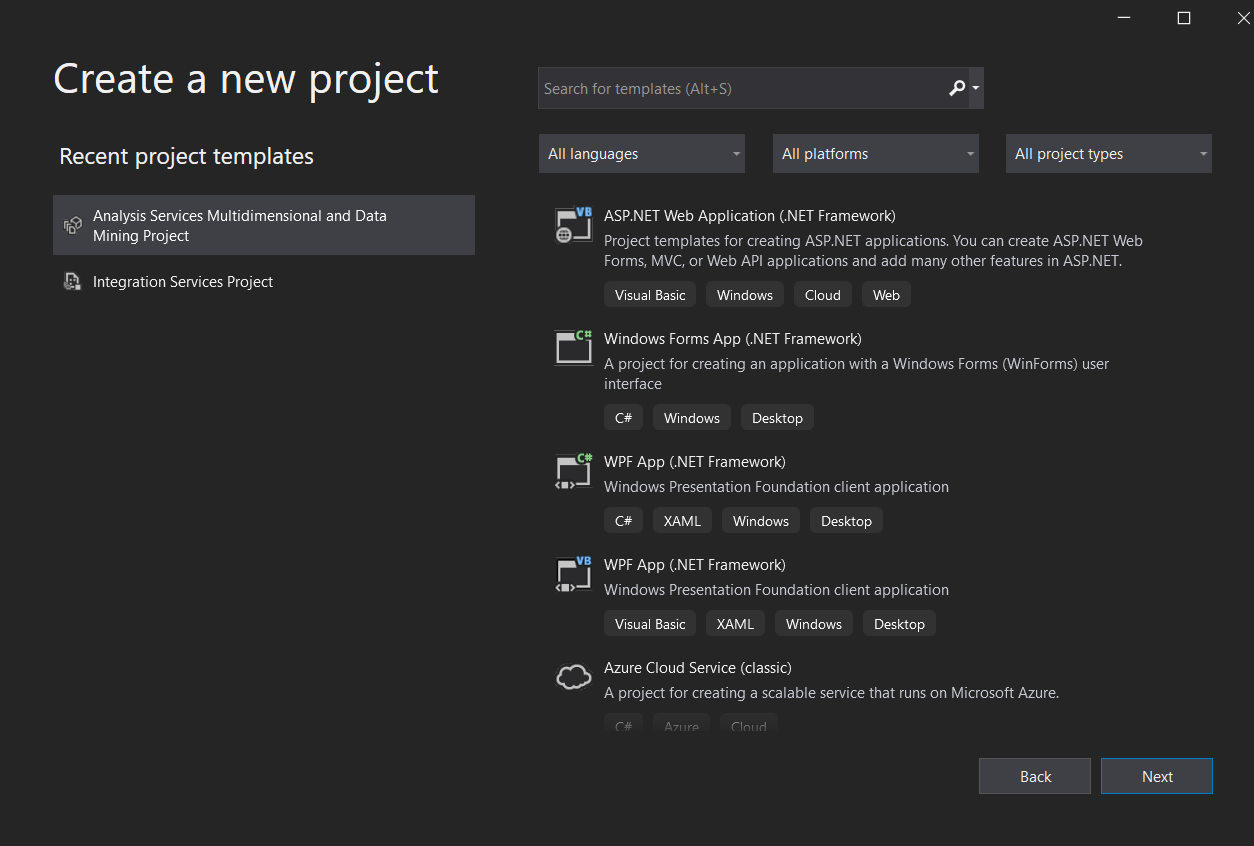


Lược đồ hình sao

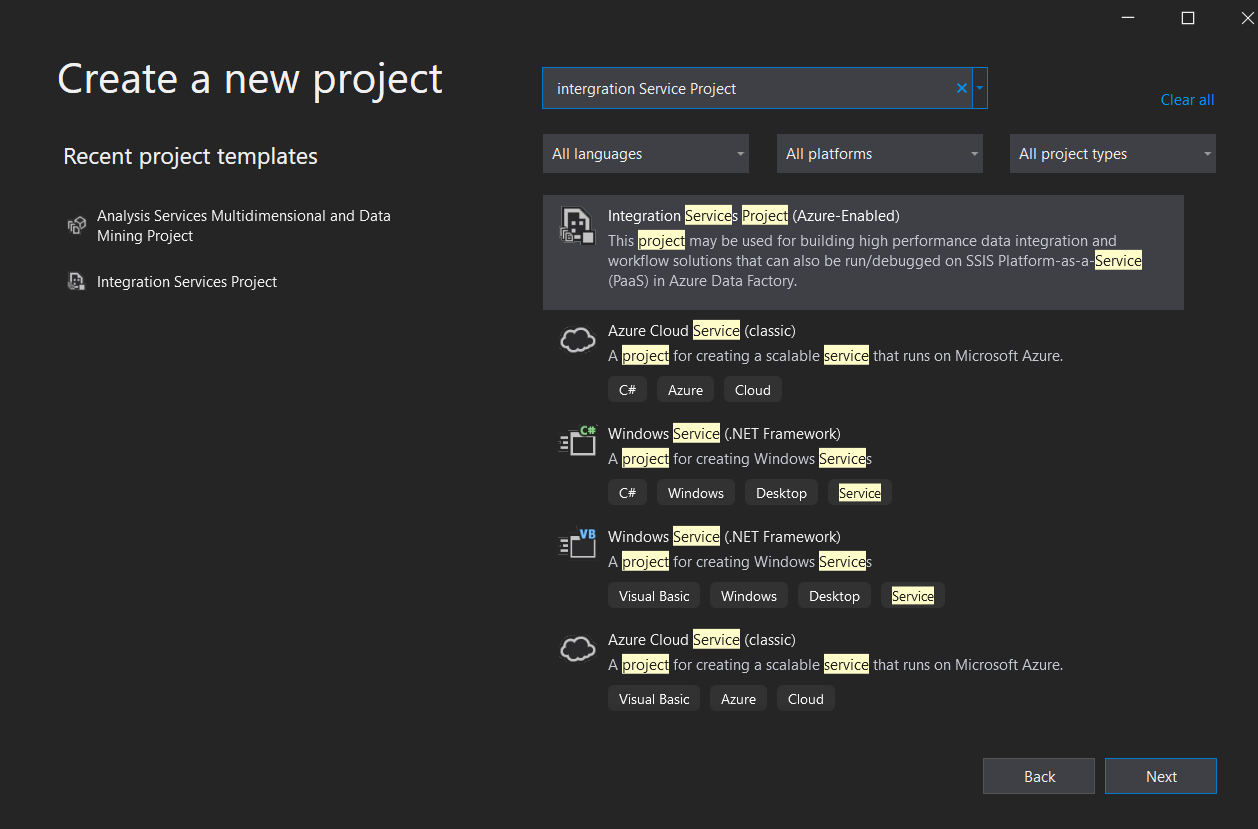
# CHƯƠNG 3: TÍCH HỢP DỮ LIỆU VÀO KHO DỮ LIỆU (SSIS)

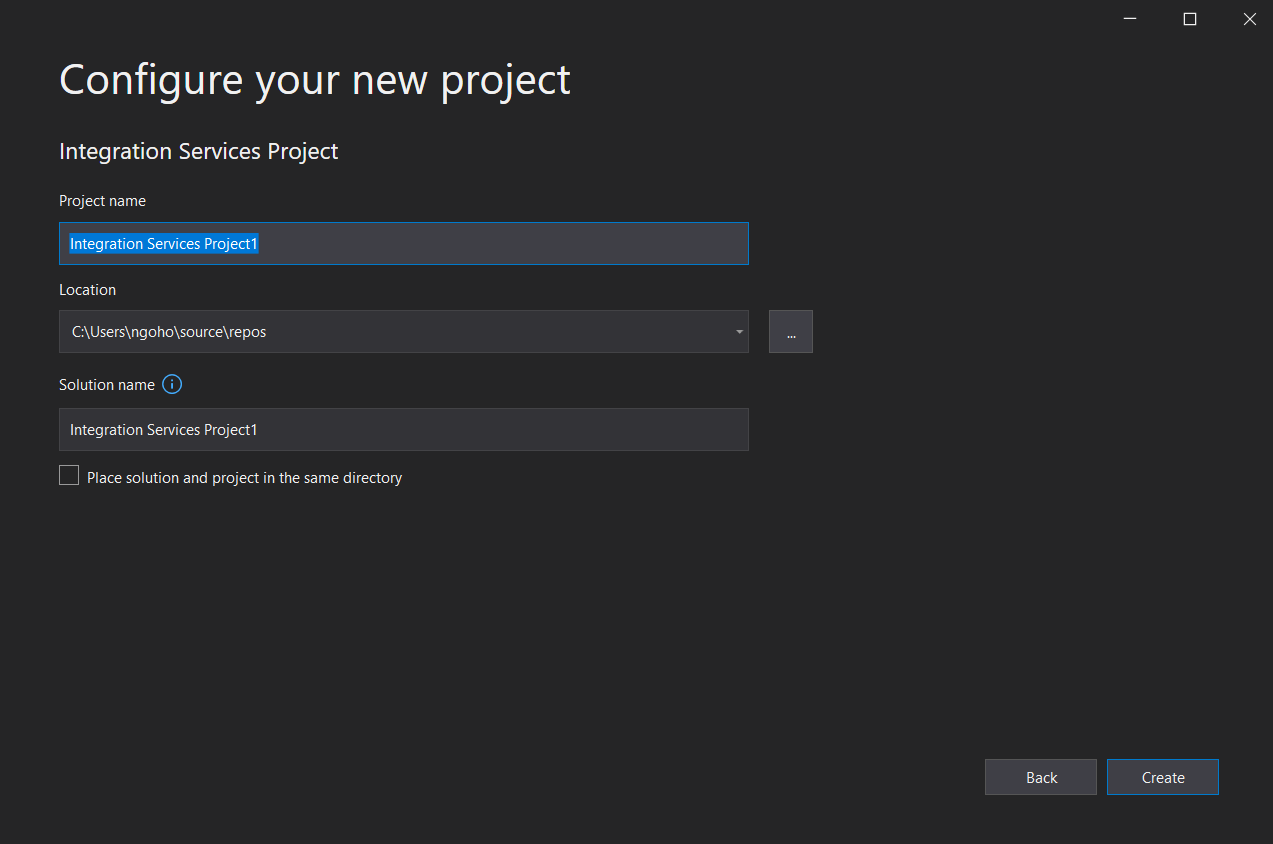
## 3.1. Tạo mới Project

Đầu tiên: File -> New -> Project



Tiếp theo thì tạo một Integration Services Project bằng cách search Integration Services Project (Điều kiện tiên quyết là IDE đã cài extension SQL Server Integration Services Project), sau đó chọn Next -> Create



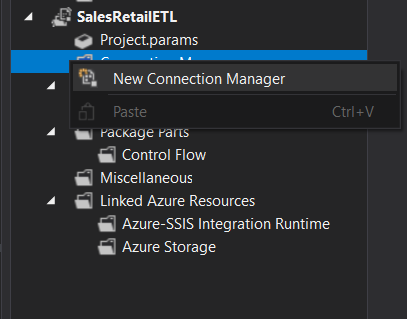




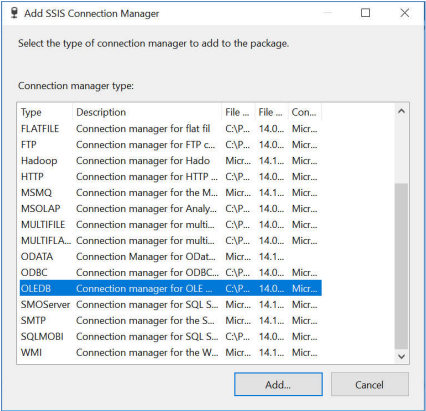
## Thiết lập các connection

Để có thể truy xuất dữ liệu cho quá trình ETL, ta cần thiết lập các connection tới nguồn (source) và đích (target). Ta có thể tạo các connection này trong mỗi package. Tuy nhiên, để có thể sử dụng lại các connection trong các package khác nhau, ta sẽ tạo các connection một cách toàn cục (globally).

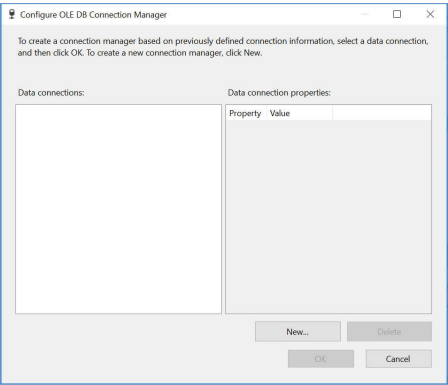
Bước 1: Tại solution explorer, right-click vào Connection Managers, chọn New Connection Manager.



Bước 2: Chọn OLEDB từ Add SSIS Connection Manager dialog và click Add…



Bước 3: Configure OLE DB Connection Manager dialog xuất hiện, click New…

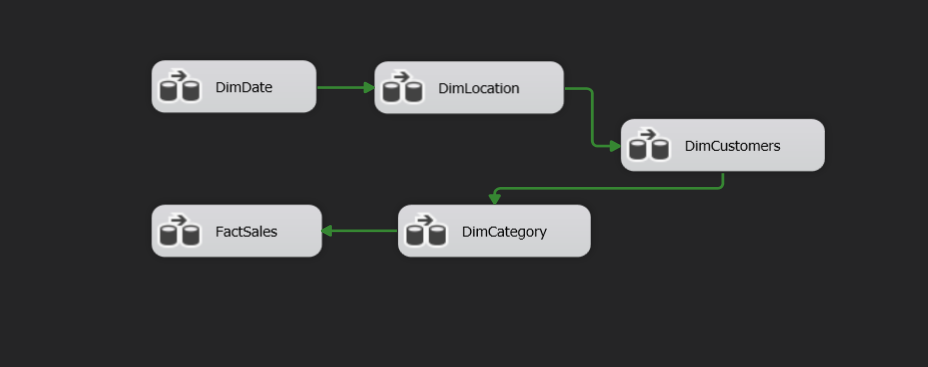


Bạn sẽ thấy Connection Manager dialog. Đặt tên cho các thông tin như bên dưới:

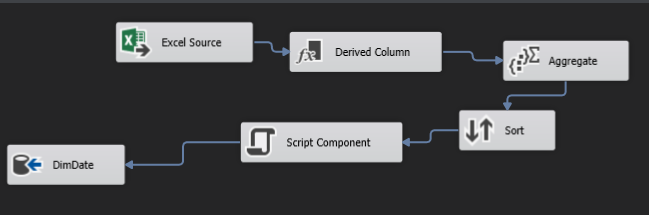
* Provider: SQL Server Native Client
* Server Name: <Tên server>
* Log on to the Server: Use Windows Authentication
* Connect to a database: <Tên database muốn connect>

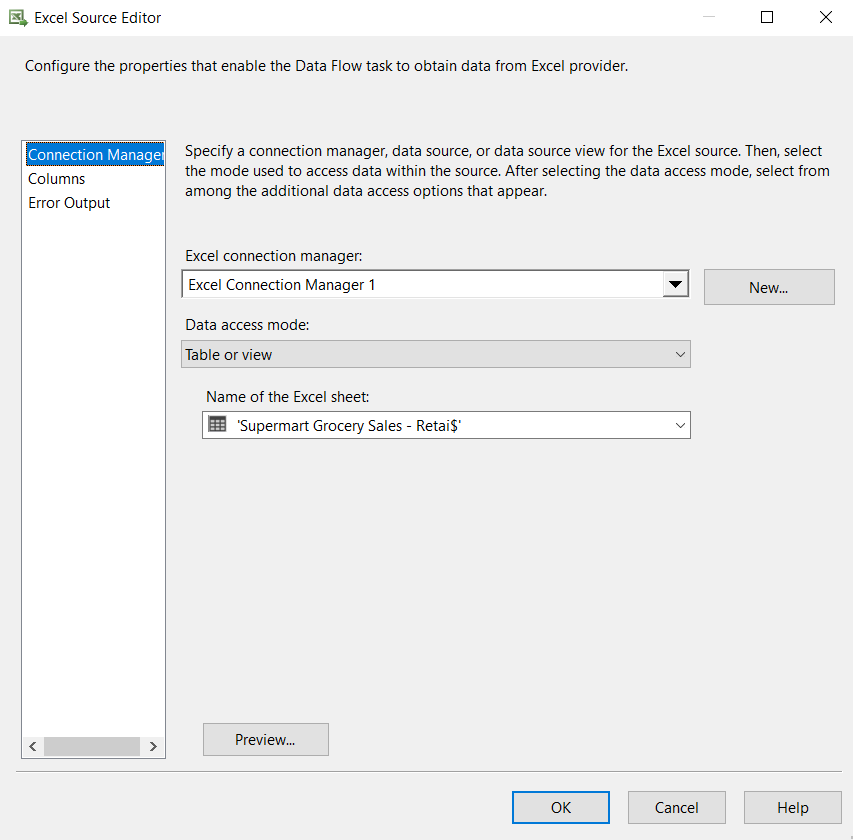
## Đổ dữ liệu từ Excel vào kho dữ liệu

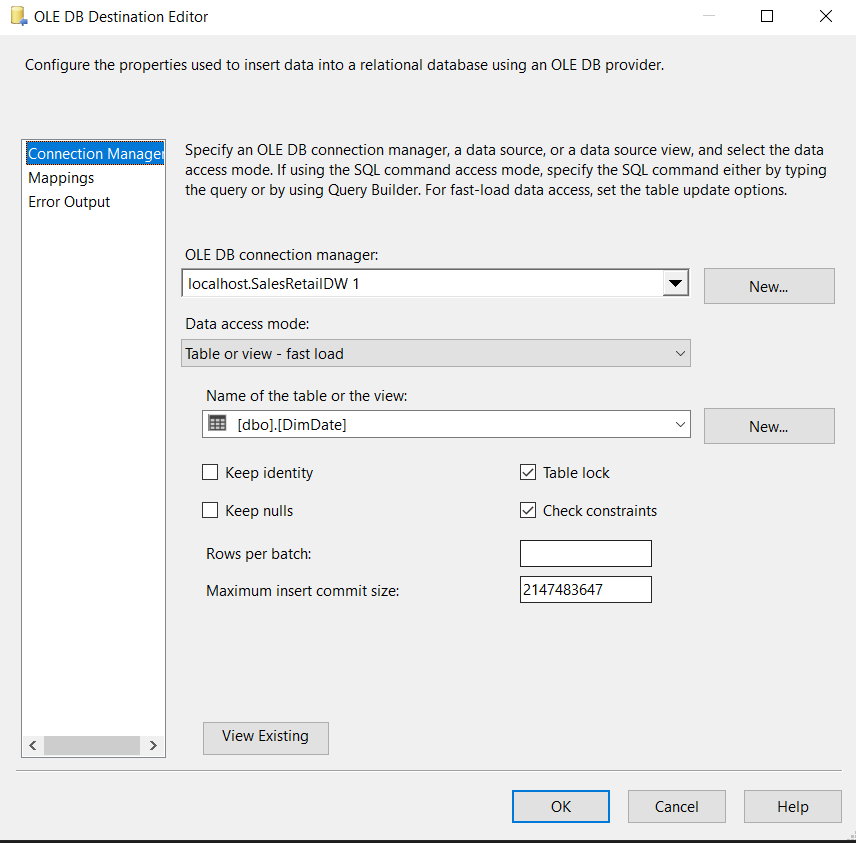
Tạo Control Flow:



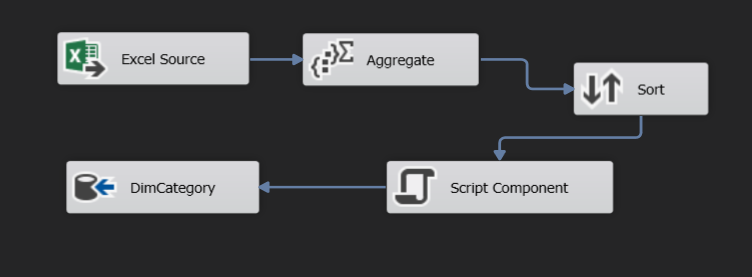
DimDate : với table source là ‘Supermart Grocery Sales – Retail’ từ file excel và destination là DimDate của database SalesRetailDW – table này được tạo sẵn và đọc từ SQL server.

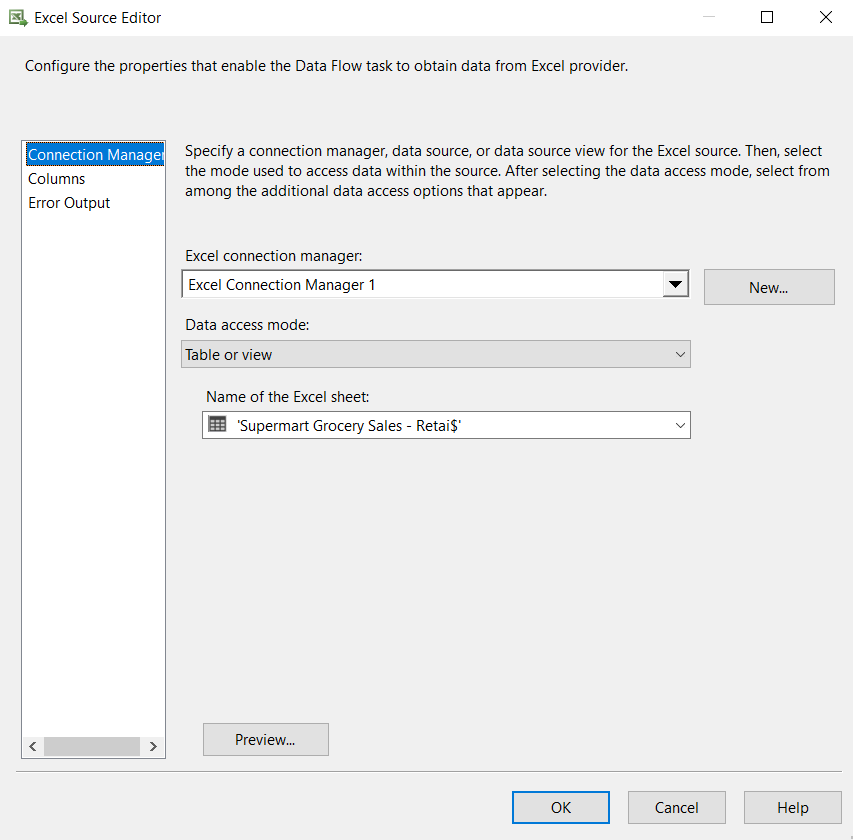


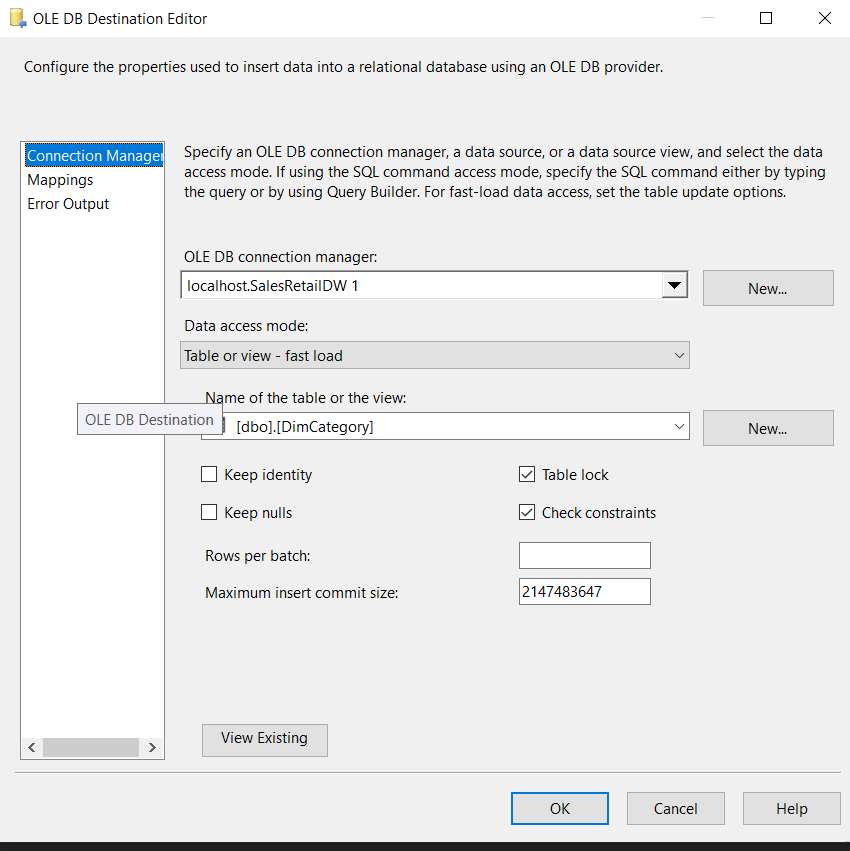




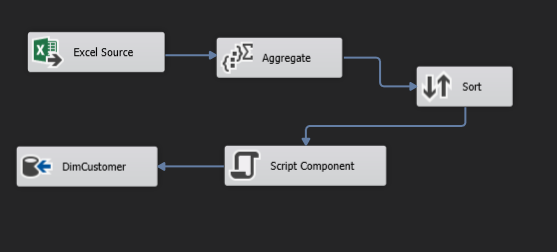
DimCategory : với table source là ‘Supermart Grocery Sales – Retail’ từ file excel và destination là DimCategory của database SalesRetailDW – table này được tạo sẵn và đọc từ SQL server.

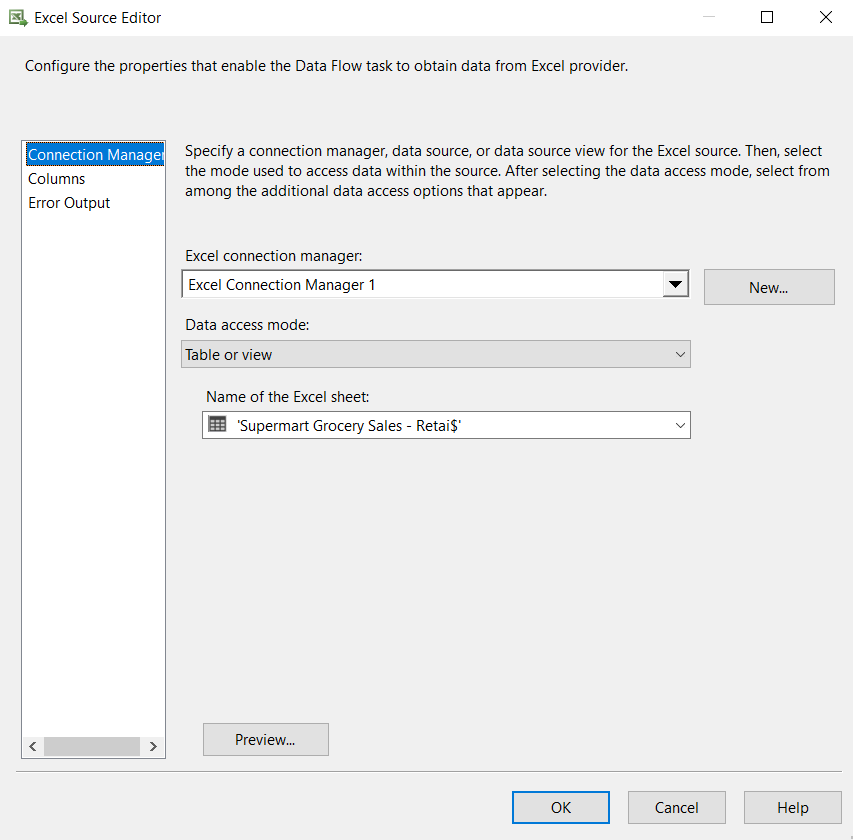


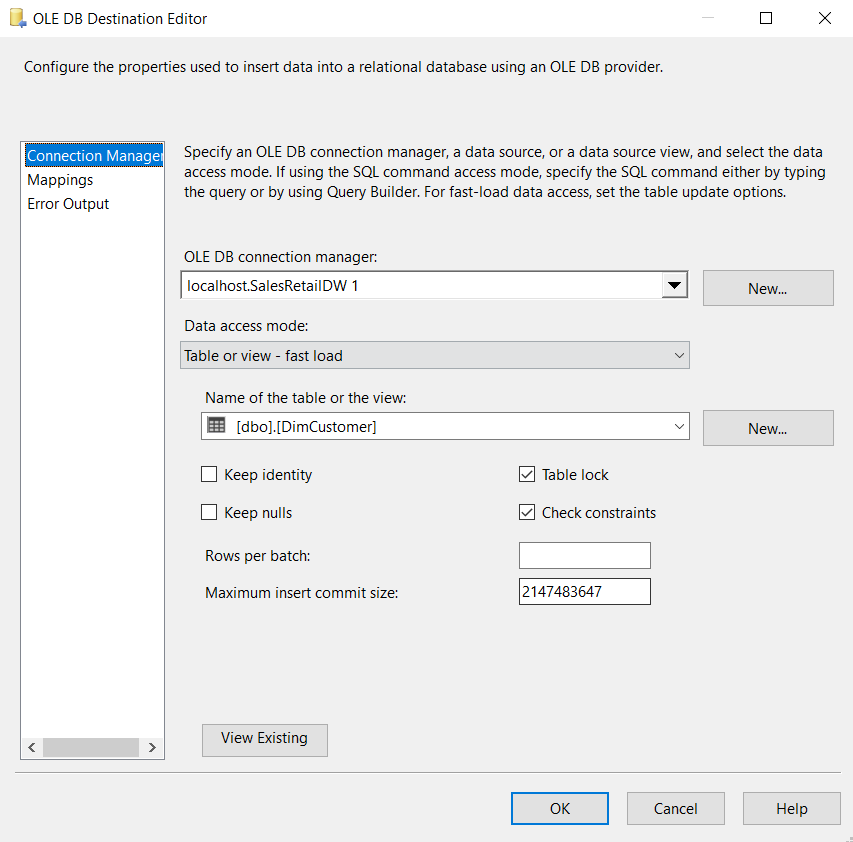




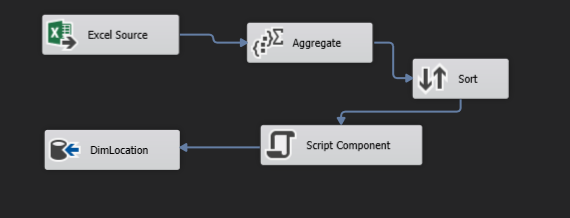
DimCustomers : với table source là ‘Supermart Grocery Sales – Retail’ từ file excel và destination là DimCustomer của database SalesRetailDW – table này được tạo sẵn và đọc từ SQL server.

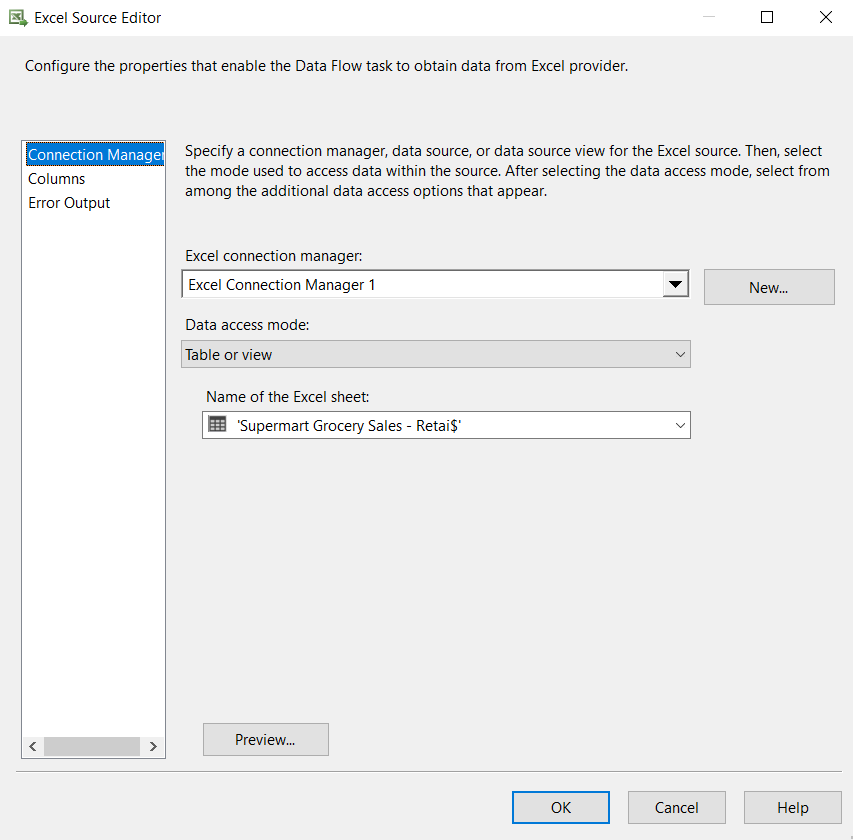






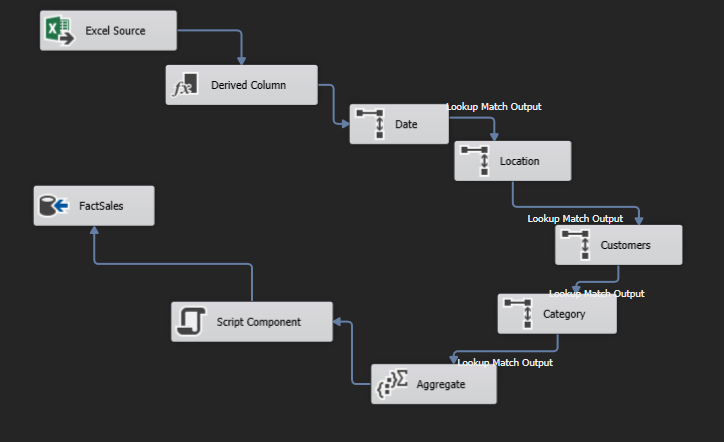
DimLocation : với table source là ‘Supermart Grocery Sales – Retail’ từ file excel và destination là DimLocation của database SalesRetailDW – table này được tạo sẵn và đọc từ SQL server.

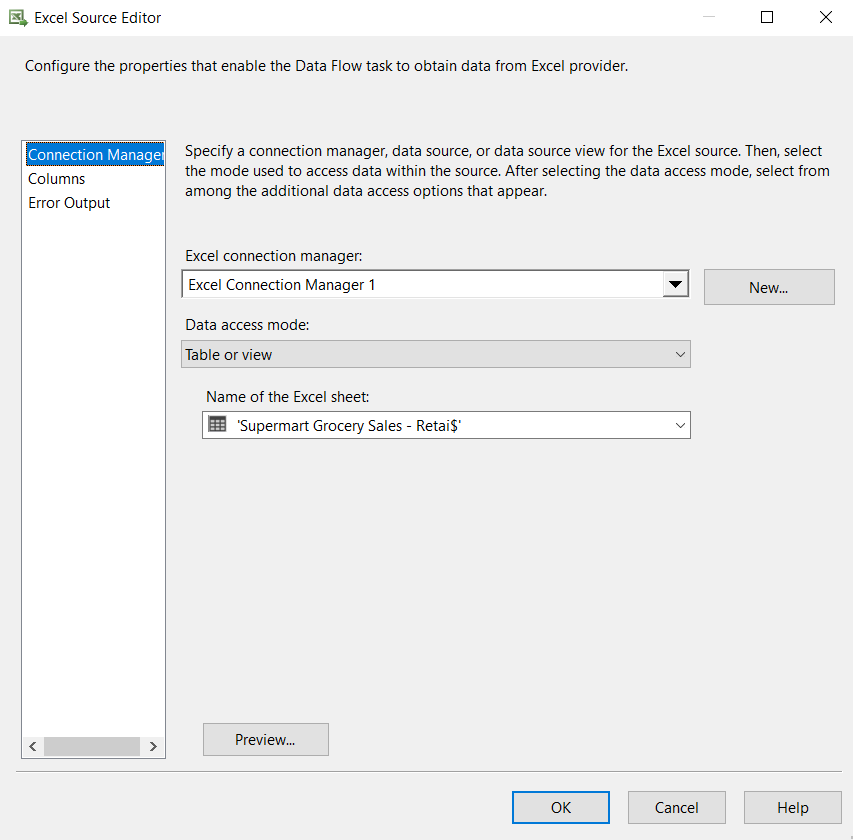


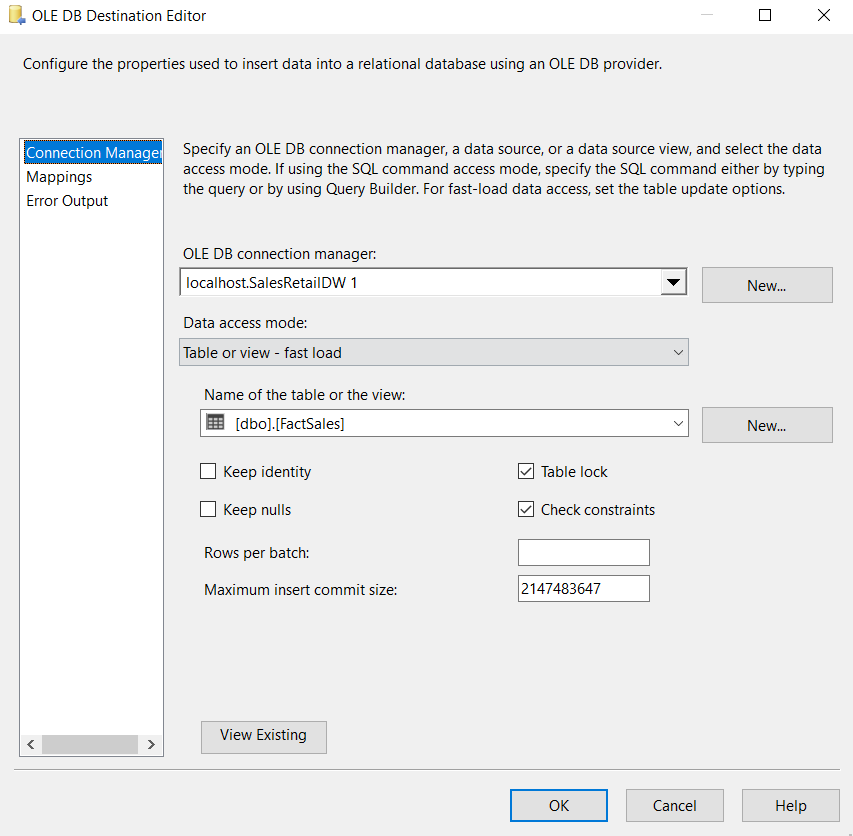




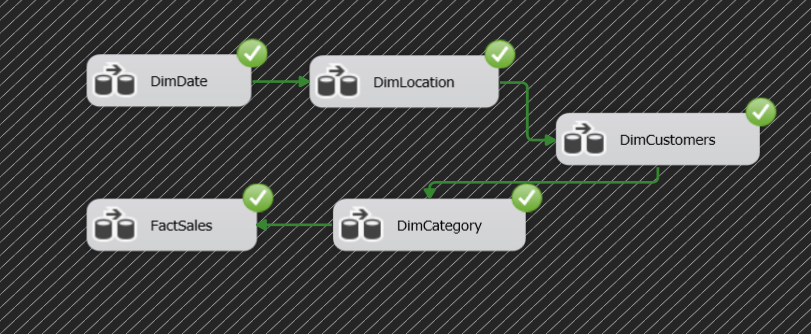
FactSales : với table source là ‘Supermart Grocery Sales – Retail’ từ file excel và destination là FactSales của database SalesRetailDW – table này được tạo sẵn và đọc từ SQL server.







## Thực thi project



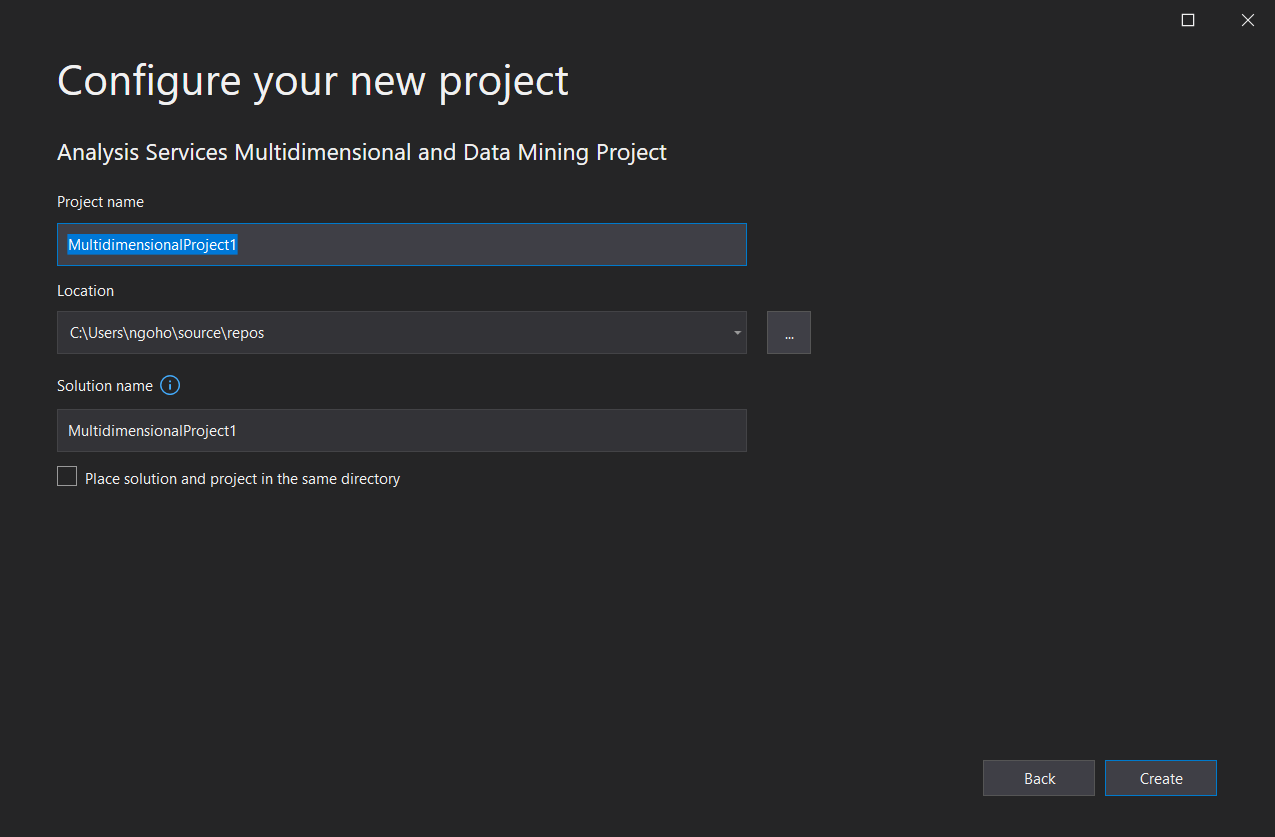
# CHƯƠNG 4: PHÂN TÍCH DỮ LIỆU (SSAS)

## Tạo mới project

Đầu tiên: File -> New -> Project

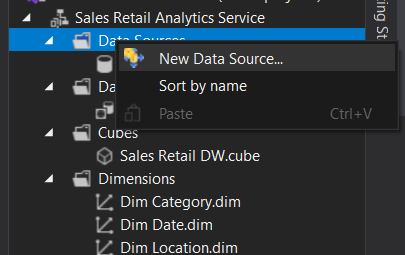


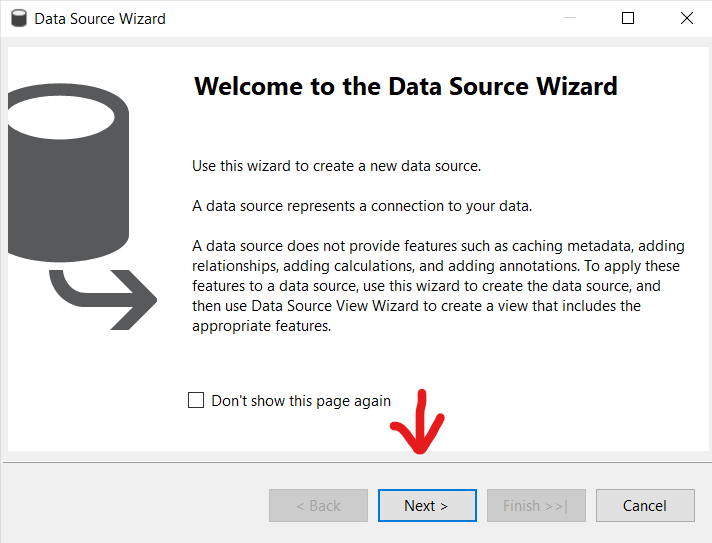
Tiếp theo thì tạo một Analysis Services Multidimensional and Data Mining Project bằng cách search Analysis Services Multidimensional and Data Mining Project (Điều kiện tiên quyết là IDE đã cài extension SQL Server Analysis Services), sau đó chọn Next -> Create

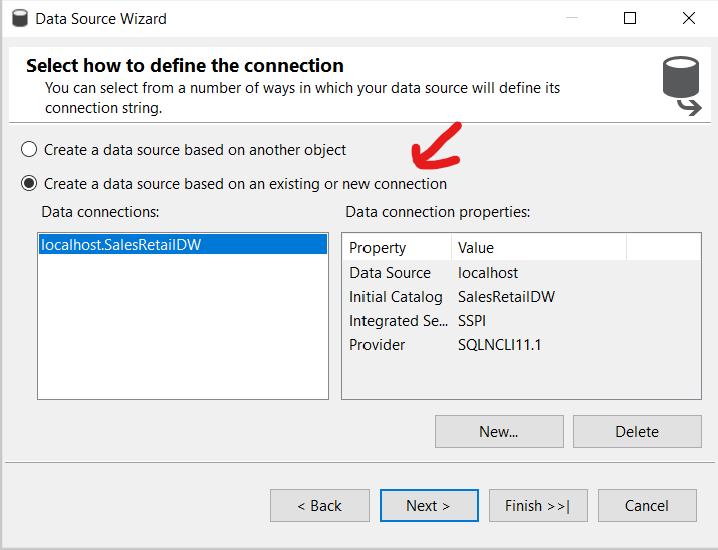
**

## Tạo datasource

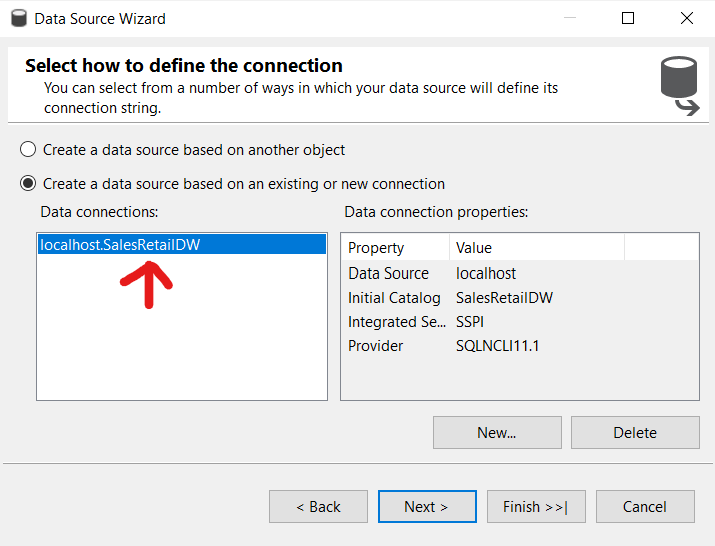
Ta tiến hành tạo datasource cho project theo từng bước sau :





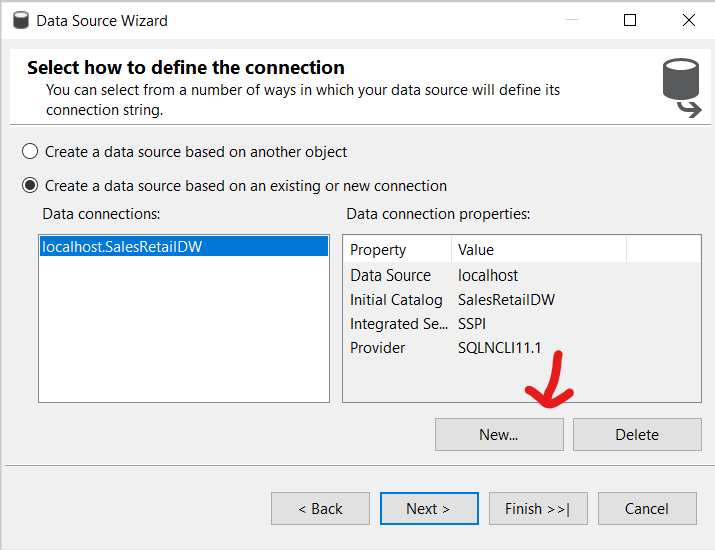


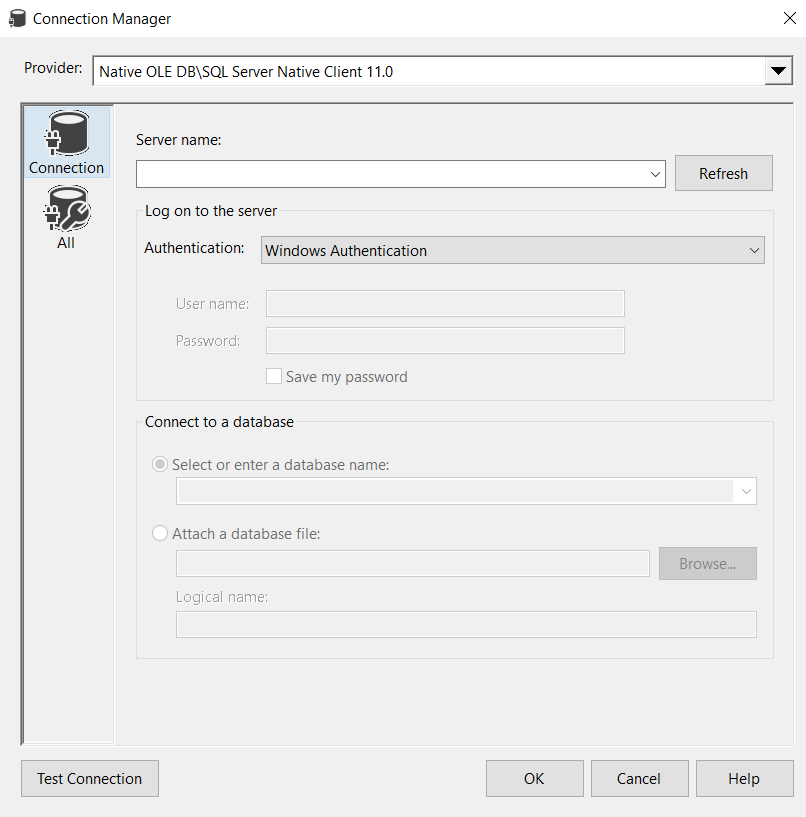
Nếu đã tồn tại connection



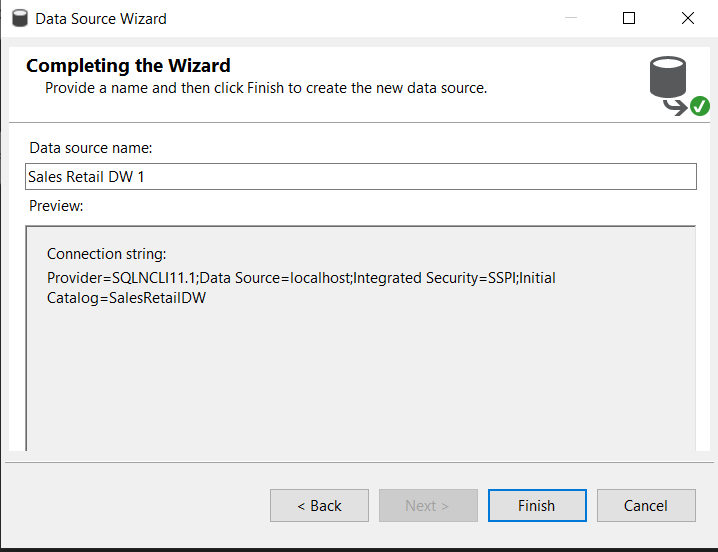
Và click next để tiếp tục.

Nếu chưa tồn tại connection :





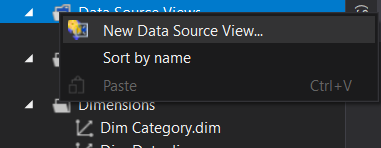
Điền những thông tin cần thiết



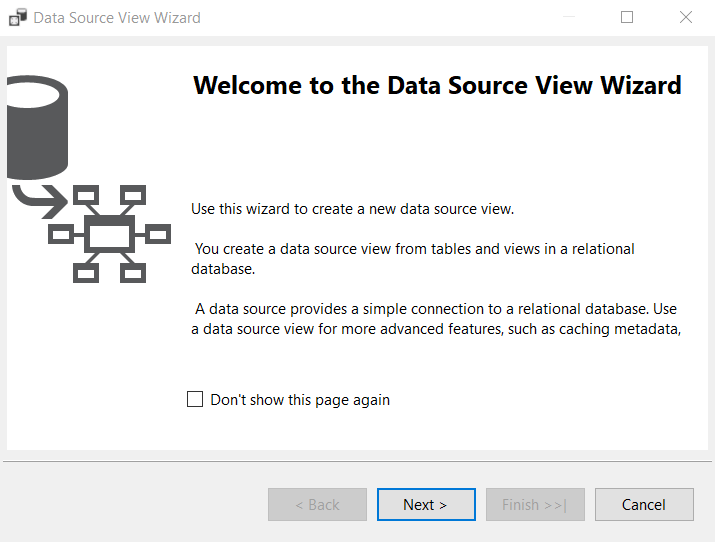
Sau khi tạo connection thành công.Click Finish để hoàn tất.

## Tạo datasource view

Để tiến hành tạo datasource view , ta thực hiện các bước sau :

****

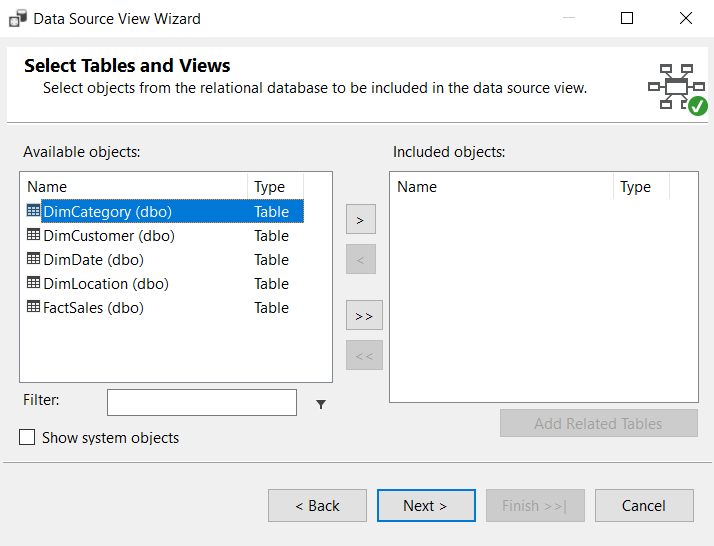
Click ‘Next’



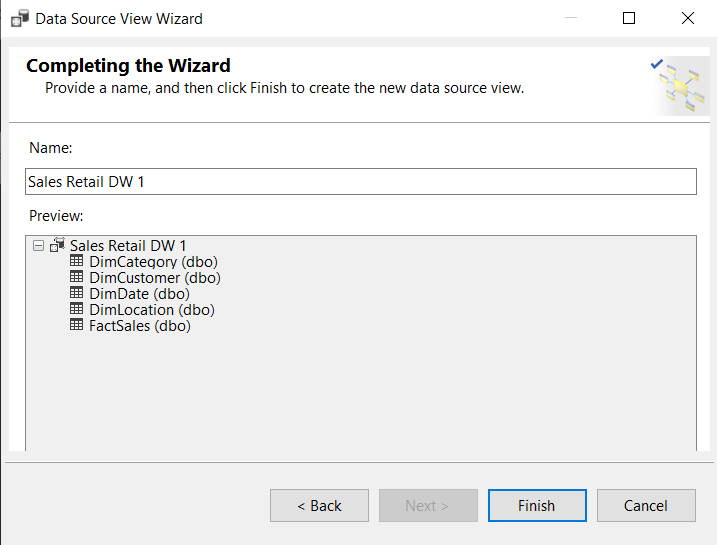
Chọn data source ‘Sales Retail DW’ có sẵn và click ‘Next’ để tiếp tục

****

Chọn những object phù hợp hoặc click ‘>>’ để chọn tất cả ( ở đây nhóm chọn tất cả ). Click ‘Next’ để tiếp tục.

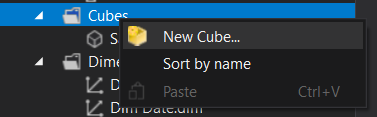
****

Click ‘Finish’ để hoàn tất.

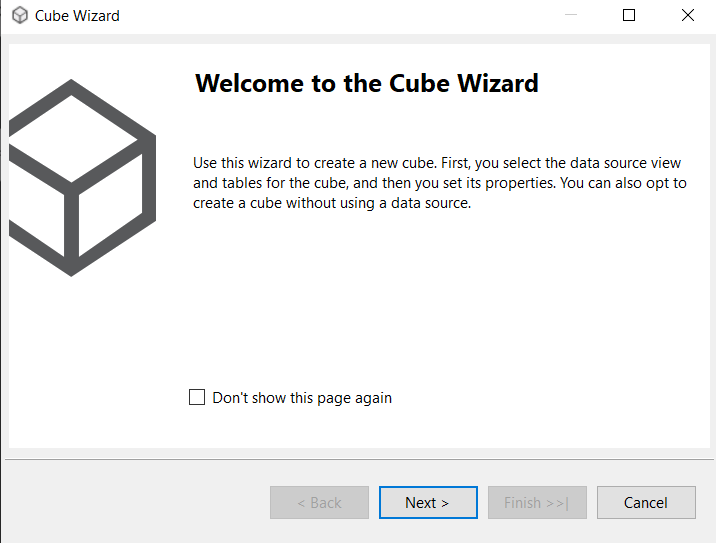


## Tạo datacubes

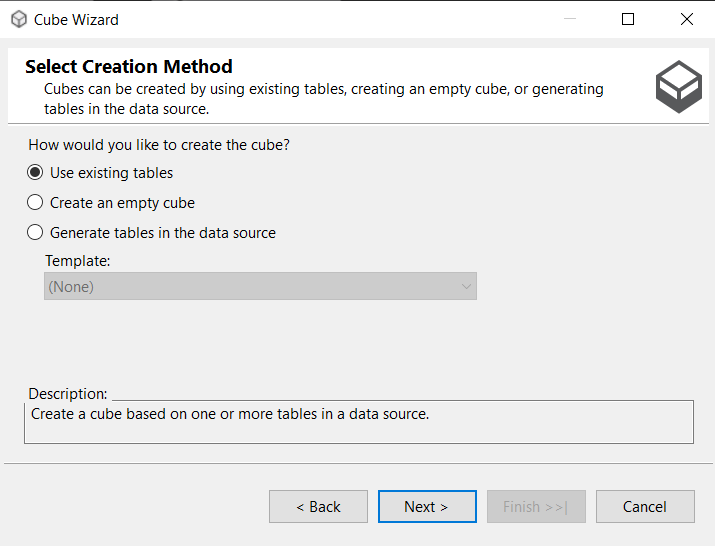
Để tiến hành tạo Datacubes ta thực hiện các bước sau :



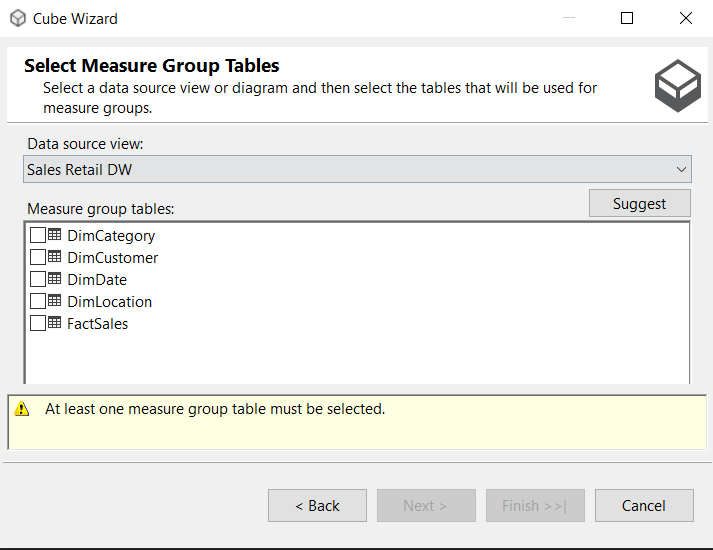
Click ‘Next’ để tiếp tục .



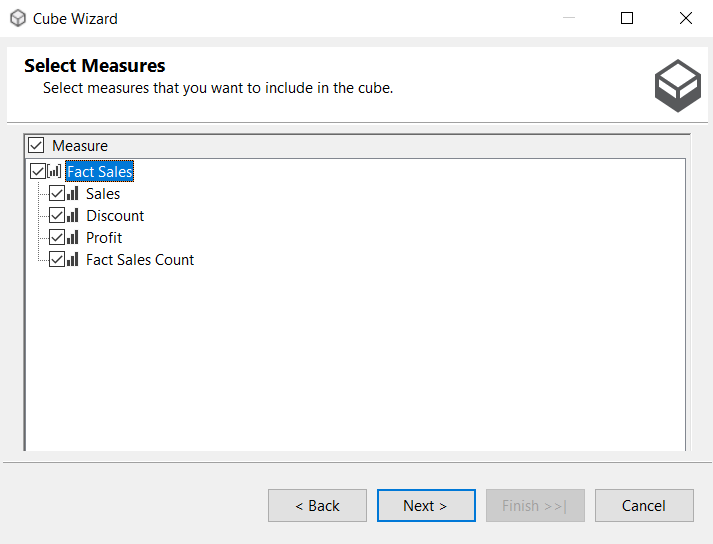
Chọn những lựa chọn thích hợp ( ở đây nhóm chọn ‘Use existing table’ ) và click ‘Next’ để tiếp tục.



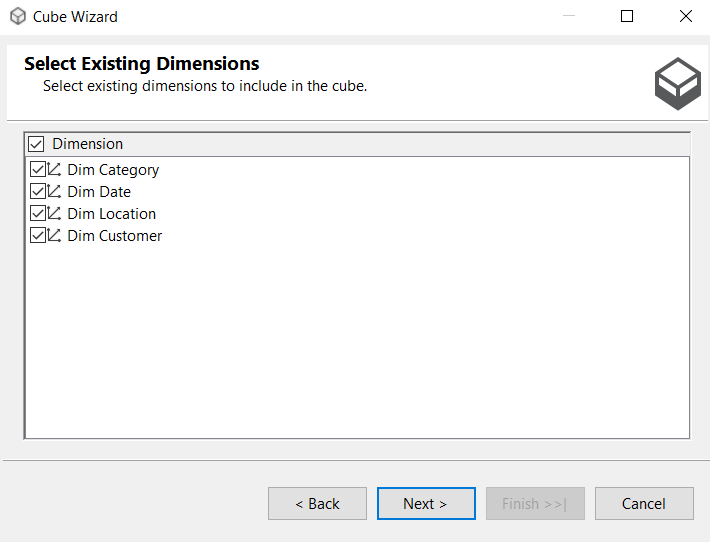
Chọn measure muốn tạo cube ( ở đây nhóm chọn FactSales) và click ‘Next’ để tiếp tục.



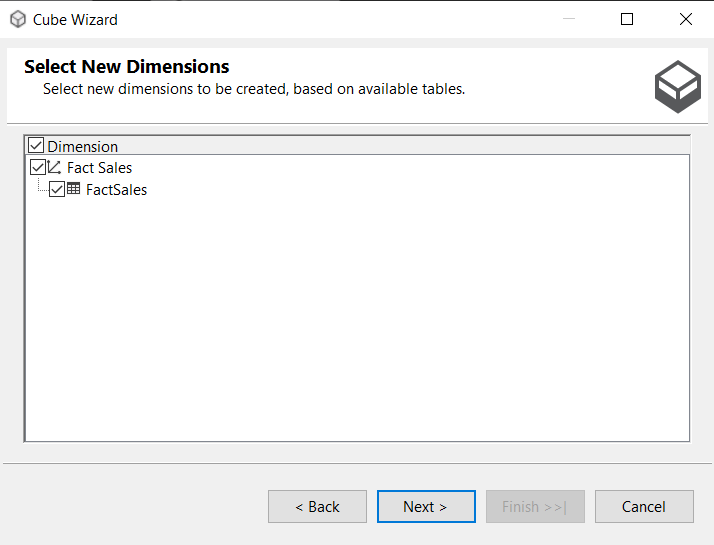
Click ‘Next’ để tiếp tục.

****

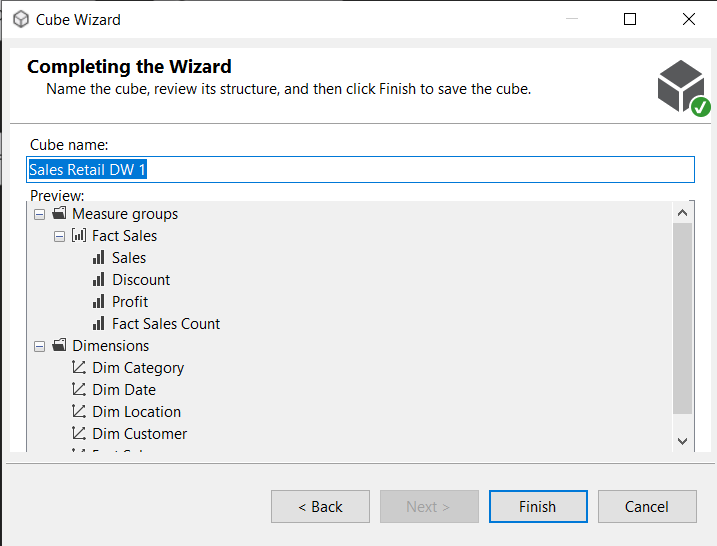
Click ‘Next’ để tiếp tục.

****

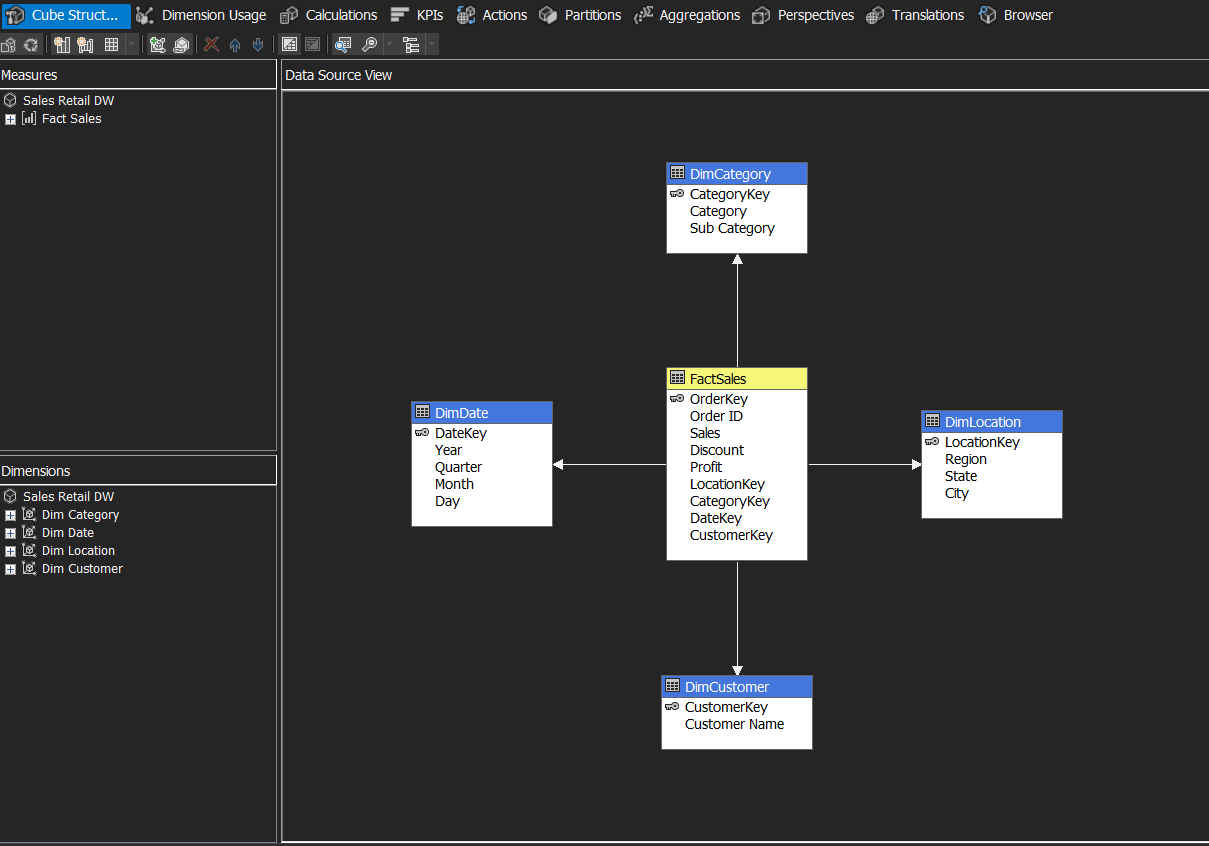
Click ‘Next’ để tiếp tục.

****

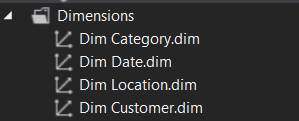
Click ‘Finish’ để hoàn tất .

****

Giao diện Data Cube sau khi tạo

****

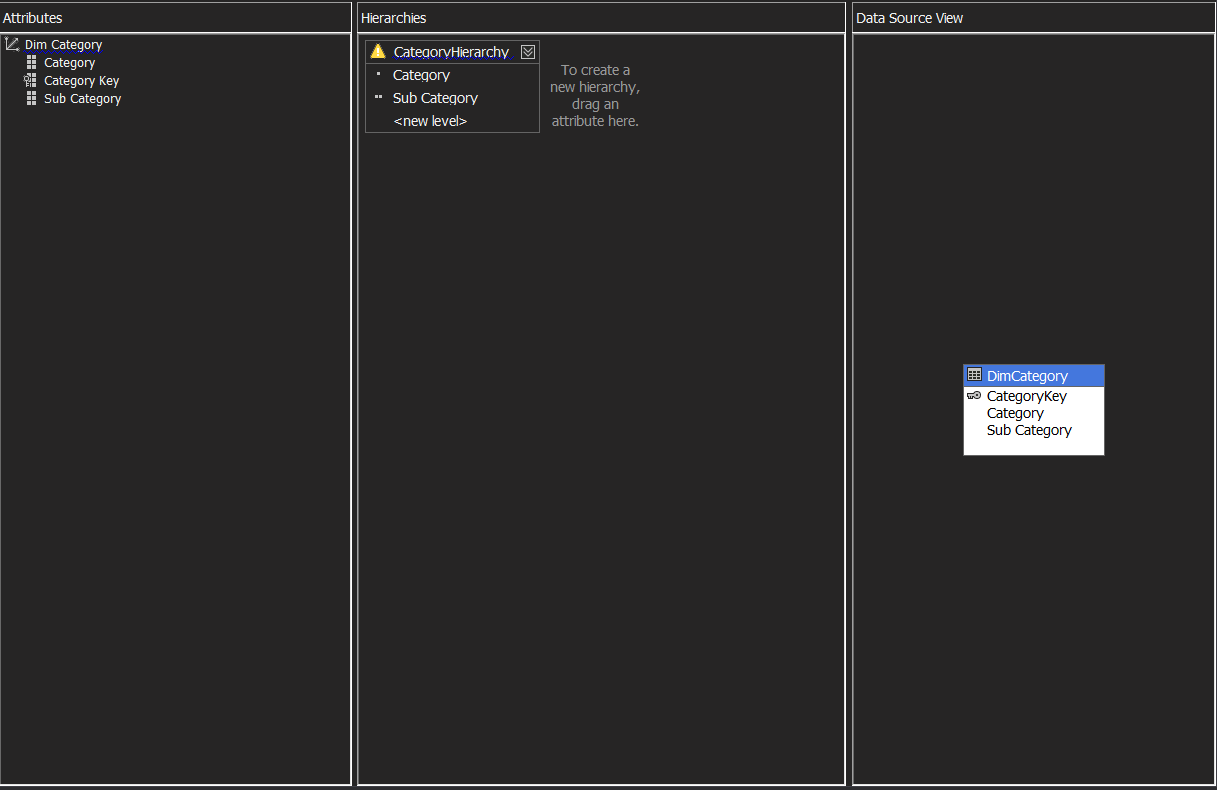
Các dimension tables được tạo :

****

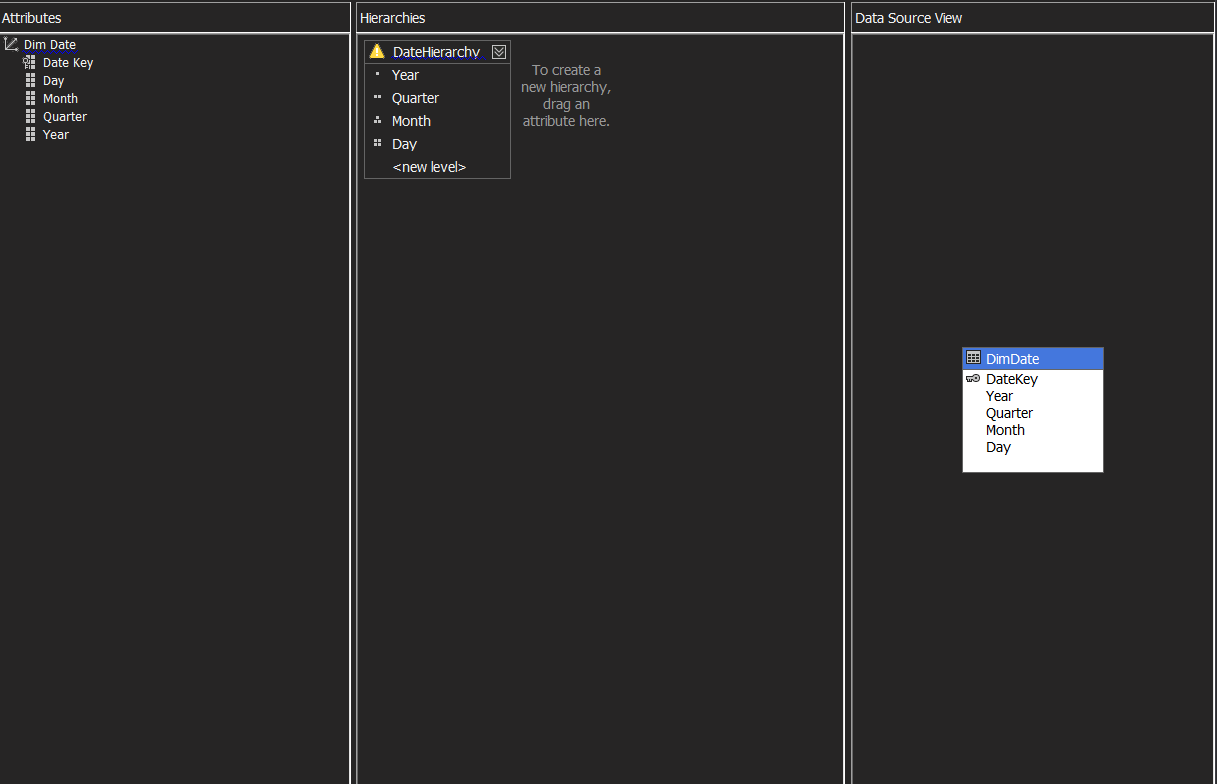
Tạo hierarchy

* Nhóm thực hiện tạo hierarchy đối với các dimension table như sau :

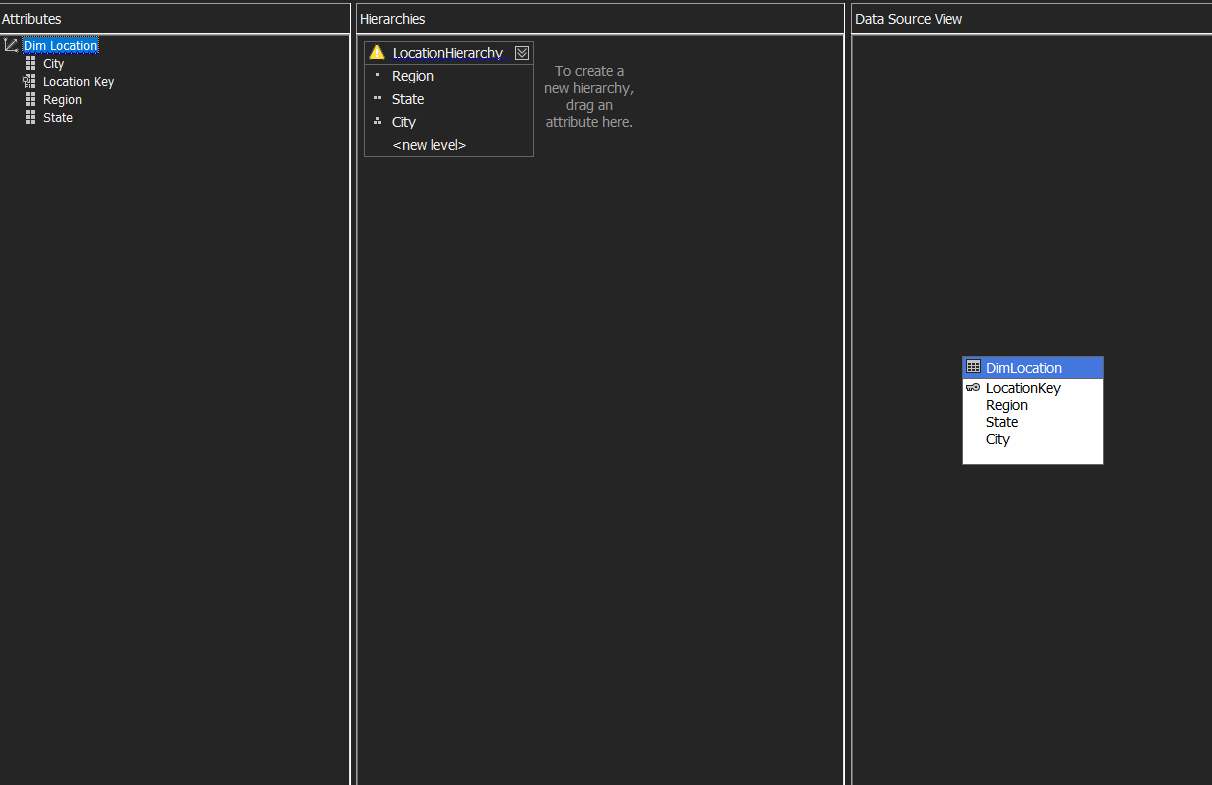
DimCategory : Category -> SubCategory.

****

DimDate : Year -> Quarter -> Month -> Day.

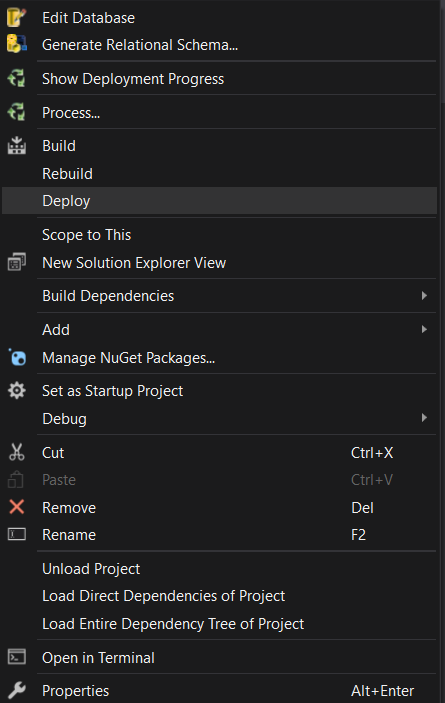
****

DimLocation : Region -> State -> City

****

Deploy project

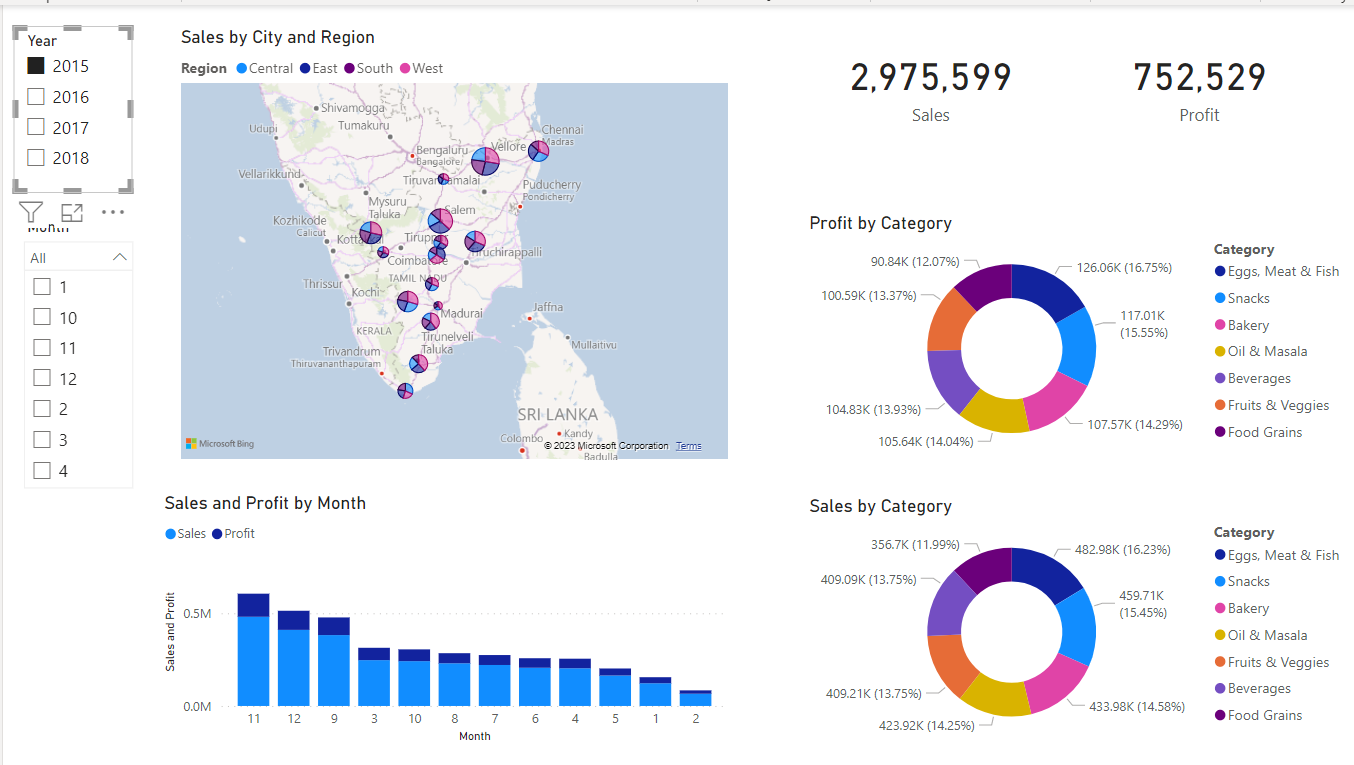
* Nhấp chuột phải vào project và chọn ‘Deploy’



* Nếu hiển thị như hình sau thì đã deploy thành công

# CHƯƠNG 5: KHAI PHÁ DỮ LIỆU (POWER BI)

## Dash board



Dash board gồm:

* Hai combobox cho phép chọn năm (từ 2015 - 2018) và tháng (từ 1 đến 12)

A screenshot of a computer

Description automatically generated with medium confidenceA screenshot of a computer

Description automatically generated with medium confidence

* Hai giá trị hiển thị tổng doanh thu và lợi nhuận

A picture containing text, font, screenshot, white

Description automatically generated

Bốn biểu đồ thống kê dữ liệu:

* Biểu đồ Map: thống kê doanh số bán hàng theo khu vực

A picture containing text, map, screenshot, atlas

Description automatically generated

* Biểu đồ Cột: thống kê doanh số bán hàng và lợi nhuận theo từng tháng trong năm được chọn.

A picture containing text, screenshot, plot, font

Description automatically generated

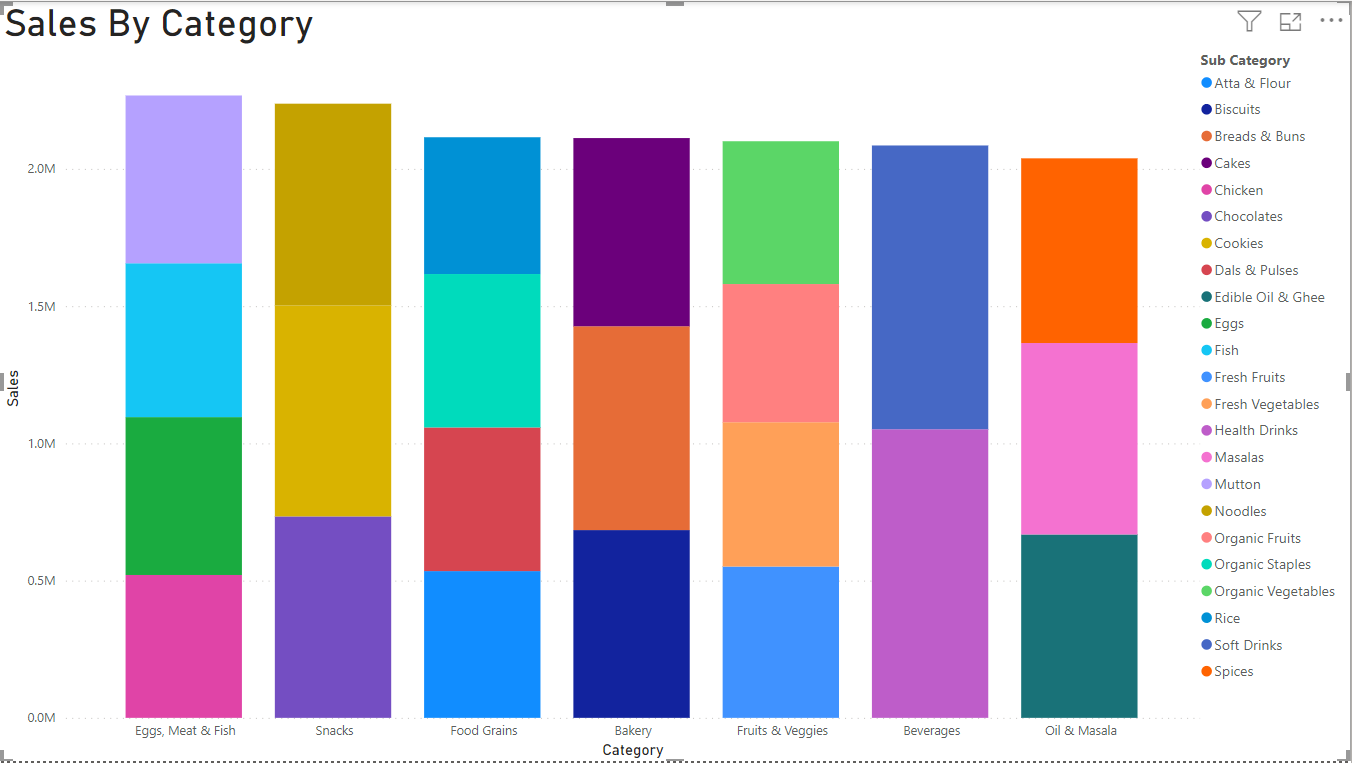
* Hai biểu đồ tròn thống kê doanh số và lợi nhuận theo từng danh mục.

A picture containing text, screenshot, circle, colorfulness

Description automatically generated

## Biểu đồ

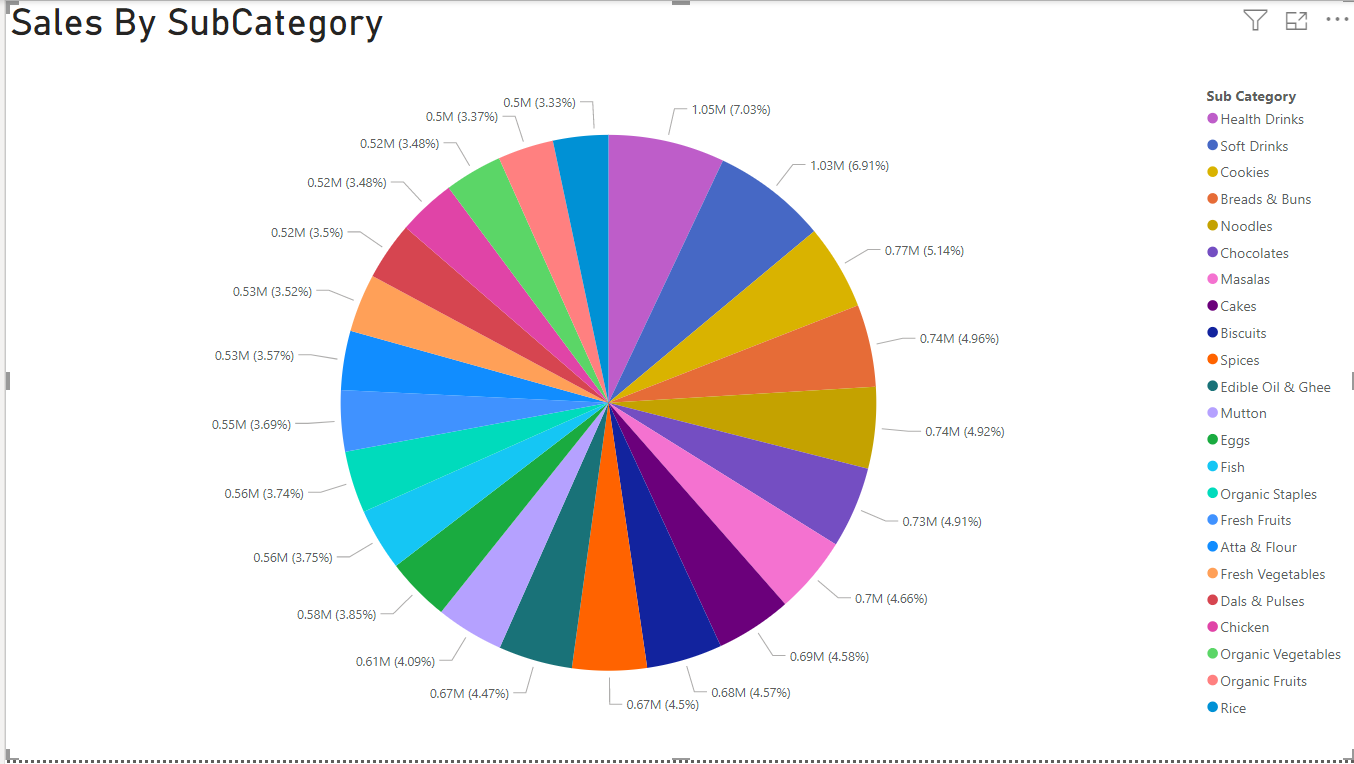
### Sales By Category



Biểu đồ phân tích doanh số bán hàng theo danh mục

Dựa vào biểu đồ ta có thể thấy doanh số bán hàng của danh mục “Eggs, Meat & Fish” là cao nhất trong đó danh mục con “Mutton” chiếm tỉ lệ nhiều nhất, trong khi đó danh mục “Oil & Masala” có doanh số thấp nhất và các danh mục con chiếm tỉ lệ gần như bằng nhau

### Sales By SubCategory



Biểu đồ phân tích doanh số theo danh mục con

Dựa vào biểu đồ ta có thể thấy doanh số của danh mục con “Health Drinks” là cao nhất với 1.05M chiếm 7.03% trên tổng doanh số bán hàng, trong khi đó doanh số thấp nhất thuôc về danh mục con “Rice” với 0.5M chiếm 3.33%

### Profit By Category

A screenshot of a computer screen

Description automatically generated with low confidence

Biểu đồ phân tích lợi nhuận theo danh mục

Dựa vào biểu đồ ta có thể thấy lợi nhuận bán hàng của danh mục “Snacks” là cao nhất trong đó danh mục con “Noodles” chiếm tỉ lệ nhiều nhất, trong khi đó danh mục “Oil & Masala” có lợi nhuận thấp nhất và các danh mục con chiếm tỉ lệ gần như bằng nhau

### Profit By SubCategory

A picture containing text, screenshot, colorfulness, diagram

Description automatically generated

Biểu đồ phân tích lợi nhuận theo danh mục con

Dựa vào biểu đồ ta có thể thấy lợi nhuận của danh mục con “Health Drinks” là cao nhất với 267,47K chiếm 7.14% trên tổng lợi nhuận, trong khi đó lợi nhuận thấp nhất thuộc về danh mục con “Chicken” với 124,05K chiếm 3.31% trên tổng lợi nhuận.

### Profit And Sales By Region

A picture containing text, screenshot, rectangle, electric blue

Description automatically generated

Biểu đồ phân tích doanh số và lợi nhuận trên từng khu vực.

Quan sát biểu đồ trên ta nhận thấy rằng tổng doanh số và lợi nhuận trên từng khu vực xếp theo thứ tự giảm dần lần lượt là Tây, Đông, Trung Tâm, Nam và cuối cùng là phía Bắc

Khu vực phía Bắc hầu như không có thực hiện một cuộc giao dịch nào.

### Sales Of Region By Year

A picture containing text, screenshot, line, plot

Description automatically generated

Biểu đồ thống kê doanh số bán hàng của từng khu vực qua từng năm

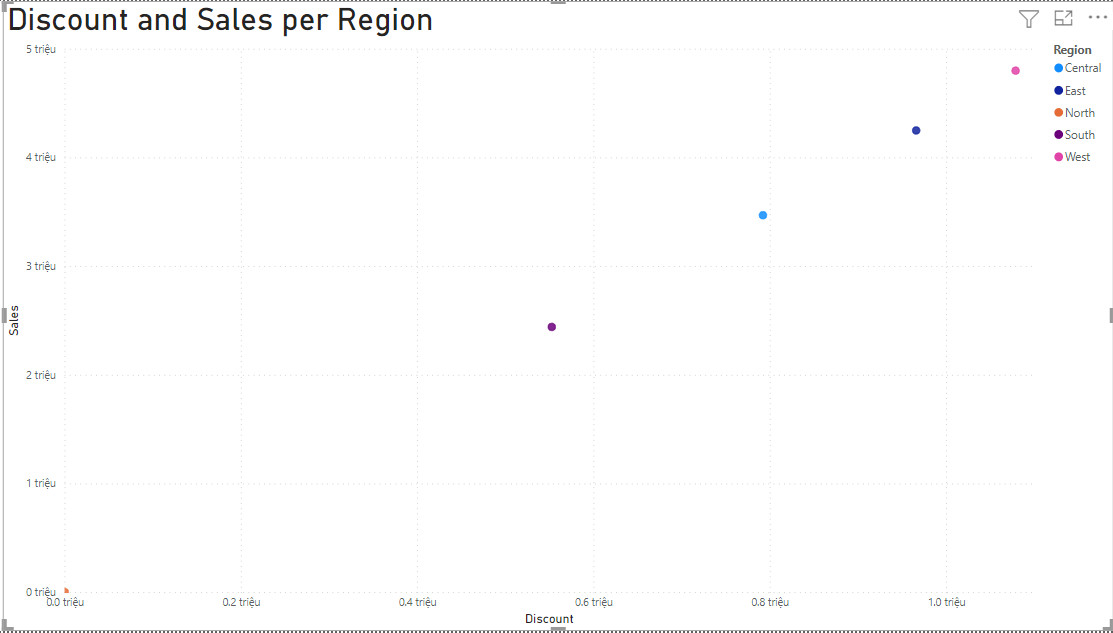
Dựa vào biểu đồ ta nhận doanh số bán hàng của 3 khu vực là Tây, khu Trung Tâm và phía Nam có xu hướng giảm nhẹ từ năm 2015 đến năm 2016 và tăng mạnh từ năm 2016 đến năm 2018. Trong khi đó thì khu vực phía Đông có doanh số tăng lên qua từng năm.

### Discount and Sales per City



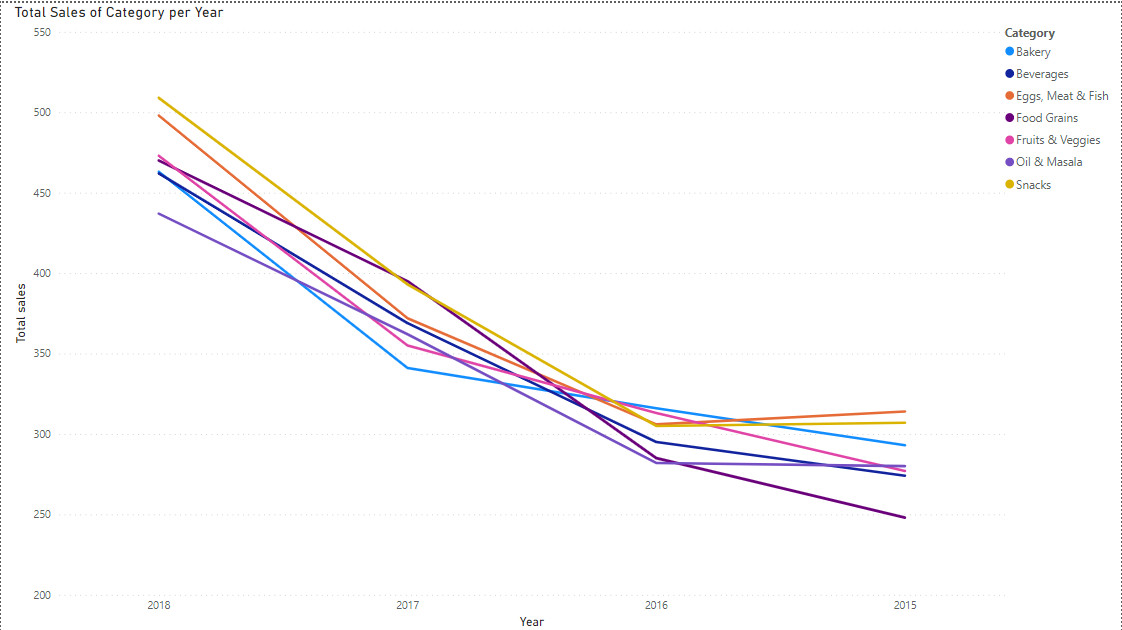
Biểu đồ giảm giá và doanh thu theo thành phố. Ở đây ta thấy thành phố Kanyakumari có doanh thu và giảm giá cao nhất

### Discount and Sales per Region



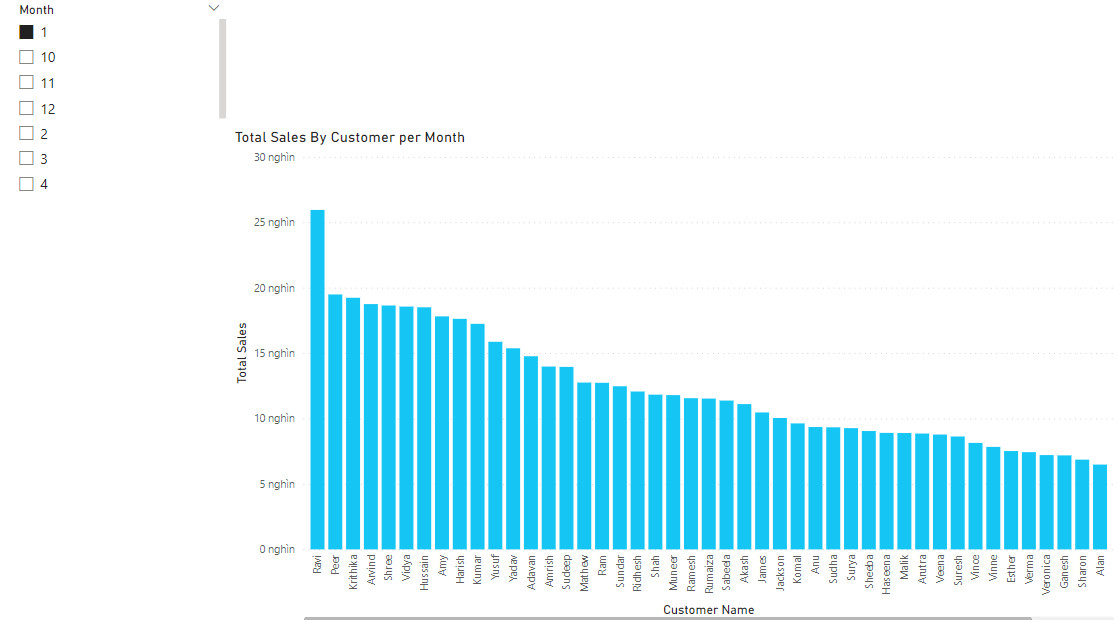
Biểu đồ thể hiện doanh số và giảm giá theo khu vực. Ta thấy Central là khu vực có doanh số và giảm giá cao nhất, thấp nhất là East.

### Total sales of Category per Year



Biểu đồ thể hiện số category bán theo năm. Từ đó biết được category bán nhiều nhất theo năm. Năm 2018 snack bán được nhiều nhất, năm 2017 oil& Masala bán nhiều nhất, năm 2016 Bakery bán được nhiều nhất, năm 2015 Eggs, Meat & Fish bán được nhiều nhất.

### Total sales by Customer per Month



Biểu đồ thể hiện doanh số khách hàng đem lại theo tháng. Để biết được khách nào đống góp cho doanh số nhiều nhất theo tháng (khách vip). Theo biểu đồ ta thấy khách hàng có tên Ravi đóng góp nhiều nhất ở tháng 1.

# TÀI LIỆU THAM KHẢO

Microsoft, 12/10/2022, Multidimensional Modeling (Adventure Works Tutorial), <https://learn.microsoft.com/en-us/analysis-services/multidimensional-tutorial/multidimensional-modeling-adventure-works-tutorial?view=asallproducts-allversions>

Microsoft, 06/15/2022, Get started with Power BI Desktop, <https://learn.microsoft.com/en-us/power-bi/fundamentals/desktop-getting-started>

Microsoft, 03/01/2023, Integration Services Tutorials, <https://learn.microsoft.com/en-us/power-bi/fundamentals/desktop-getting-started>

# KẾT LUẬN

Sau bài nghiên cứu giúp chúng ta hiểu rõ quá trình tạo ra kho dữ liệu, phân tích dữ liệu, sử dụng Power Bi để trự quan hoá dữ liệu bằng cách tạo dash board, và các biểu đồ sinh động. Cụ thể ở tập dữ liệu nhóm đã chọn là phân tích được doanh thu, lợi nhuận theo năm, tháng, quí, khu vực… của một cửa hiền tiện lợi. Từ đó có thể đưa ra các chiến lược kinh doanh phù hợp, mở rộng thị trường kinh doanh ở khu vực tiềm năng, chọn những mặt hàng ưu chuộng với khách hàng. Kết quả nghiên cứu có thể làm tiền đề cho các nghiên cứu tiếp theo. Tuy nhiên bài nghiên cứu còn hạn chế do tập dữ liệu còn hạn chế, còn thiếu về thời gian và chưa chi tiết các loại sản phẩm.