**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN**

**THÔNG VIỆT – HÀN**

**KHOA KHOA HỌC MÁY TÍNH**

**----------**



**ĐỒ ÁN MÔN HỌC KHO DỮ LIỆU**

**ĐỀ TÀI: PHÂN TÍCH DỮ LIỆU CHUỖI CUNG ỨNG**

**Sinh viên thực hiện: Trần Thị Thủy – 22IT.B204**

**Huỳnh Thể Thúy An – 22IT.B003**

**Lê Văn Trung - 22IT.B229**

**Giảng viên hướng dẫn: ThS. Mai Lam**

***Đà Nẵng, tháng 10 năm 2024***

**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN**

**THÔNG VIỆT – HÀN**

**KHOA KHOA HỌC MÁY TÍNH**

**----------**



**ĐỒ ÁN MÔN HỌC KHO DỮ LIỆU**

**ĐỀ TÀI: PHÂN TÍCH DỮ LIỆU CHUỖI CUNG ỨNG**

**Sinh viên thực hiện: Trần Thị Thủy – 22IT.B204**

**Huỳnh Thể Thúy An – 22IT.B003**

**Lê Văn Trung - 22IT.B229**

**Giảng viên hướng dẫn: ThS. Mai Lam**

***Đà Nẵng, tháng 10 năm 2024***

# 

# LỜI CẢM ƠN

Trong quá trình thực hiện đề tài “Phân tích dữ liệu chuỗi cung ứng” này, chúng em muốn gửi lời cảm ơn chân thành nhất của mình tới tất cả những người đã giúp đỡ, hỗ trợ nhóm cả về kiến thức lẫn tinh thần trong quá trình thực hiện.

Trước hết, chúng em xin chân thành cảm ơn thầy giáo – ThS. Mai Lam - Giảng viên Khoa Công Nghệ Thông Tin, Trường Đại học Công nghệ thông tin và Truyền thông Việt - Hàn, người đã trực tiếp hướng dẫn, nhận xét, giúp đỡ nhóm em trong suốt quá trình thực hiện đồ án.

Do thời gian thực hiện đề tài có hạn, kiến thức còn nhiều hạn chế nên đề tài thực hiện chắc chắn không thể tránh khỏi những thiếu sót. Nhóm em rất mong nhận được ý kiến đóng góp của các thầy cô để có thêm kinh nghiệm và tiếp tục hoàn thiện dự án của mình.

Nhóm chúng em xin chân thành cảm ơn!

**Nhóm sinh viên thực hiện**

Trần Thị Thủy

Huỳnh Thể Thúy An

Lê Văn Trung

# NHẬN XÉT CỦA GIẢNG VIÊN

# MỤC LỤC

[LỜI CẢM ƠN i](#_Toc182157990)

[NHẬN XÉT CỦA GIẢNG VIÊN ii](#_Toc182157991)

[MỤC LỤC 1](#_Toc182157992)

[DANH MỤC HÌNH ẢNH 2](#_Toc182157993)

[DANH MỤC BẢNG 3](#_Toc182157994)

[CHƯƠNG 1 : TỔNG QUAN VỀ ĐỀ TÀI 4](#_Toc182157995)

[1. Lý do chọn đề tài 4](#_Toc182157996)

[2. Các công cụ sử dụng 4](#_Toc182157997)

[3. Mô tả dataset 4](#_Toc182157998)

[3.1. Danh sách thuộc tính được phân tích 5](#_Toc182157999)

[4. Xây dựng kho dữ liệu 6](#_Toc182158000)

[4.1. Lược đồ hình sao (Star schema) 6](#_Toc182158001)

[CHƯƠNG 2: PHÂN TÍCH VÀ THIẾT KẾ KHO DỮ LIỆU 10](#_Toc182158002)

[1. Trước khi xử lý dữ liệu 10](#_Toc182158003)

[1.1. Những thuộc tính cần thiết cho bảng fact chính 10](#_Toc182158004)

[1.2. Sau khi xử lý dữ liệu 11](#_Toc182158005)

[CHƯƠNG 3: TRIỂN KHAI KHO DỮ LIỆU 15](#_Toc182158006)

[1. SSIS 15](#_Toc182158007)

[1.1. Quá trình đổ dữ liệu vào SSIS 16](#_Toc182158008)

[2. SSAS 16](#_Toc182158009)

[2.1. Danh sách các câu truy vấn 17](#_Toc182158010)

[KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN 31](#_Toc182158011)

[1. Kết quả đạt được 31](#_Toc182158012)

[2. Hướng phát triển 31](#_Toc182158013)

[KẾT LUẬN 33](#_Toc182158014)

[1. Kết quả đạt được 33](#_Toc182158015)

[TÀI LIỆU THAM KHẢO 34](#_Toc182158016)

# DANH MỤC HÌNH ẢNH

[Hình 1: Dữ liệu DataCoSupplyChainDataset 5](#_Toc182160277)

[Hình 2: Logical 8](#_Toc182160278)

[Hình 3: Physical 9](#_Toc182160279)

[Hình 4: DIM PRODUCT 11](#_Toc182160280)

[Hình 5: DIM CITY 12](#_Toc182160281)

[Hình 6: DIM COUNTRY 12](#_Toc182160282)

[Hình 7: DIM CUSTOMER 13](#_Toc182160283)

[Hình 8: DIM MARKET 13](#_Toc182160284)

[Hình 9: DIM STATE 14](#_Toc182160285)

[Hình 10: Bảng FACT 14](#_Toc182160286)

[Hình 11: Mô hình SSIS 15](#_Toc182160287)

[Hình 12: Data Flow 16](#_Toc182160288)

[Hình 13: Mô hình SSAS 17](#_Toc182160289)

[Hình 14: Dữ liệu truy vấn số lượng đơn hàng và tổng doanh thu theo khách hàng 18](#_Toc182160290)

[Hình 15: Số lượng đơn hàng và tổng doanh thu theo khách hàng 19](#_Toc182160291)

[Hình 16: Dữ liệu truy vấn Tổng lợi nhuận theo từng thành phố 20](#_Toc182160292)

[Hình 17: Tổng lợi nhuận theo từng thành phố 21](#_Toc182160293)

[Hình 18: Dữ liệu truy vấn Phân tích doanh thu theo thị trường 22](#_Toc182160294)

[Hình 19: Phân tích doanh thu theo thị trường 22](#_Toc182160295)

[Hình 20: Dữ liệu truy vấn Tổng doanh thu theo bang 23](#_Toc182160296)

[Hình 21: Tổng doanh thu theo bang 24](#_Toc182160297)

[Hình 22: Dữ liệu truy vấn Tổng doanh thu theo từng quốc gia 25](#_Toc182160298)

[Hình 23: Tổng doanh thu theo từng quốc gia 26](#_Toc182160299)

[Hình 24: Dữ liệu truy vấn Tổng số lượng sản phẩm bán ra theo danh mục sản phẩm 27](#_Toc182160300)

[Hình 25: Tổng số lượng sản phẩm bán ra theo danh mục sản phẩm 28](#_Toc182160301)

[Hình 26: Dữ liệu truy vấn Doanh thu trung bình mỗi đơn hàng theo Product Name 29](#_Toc182160302)

[Hình 27: Doanh thu trung bình mỗi đơn hàng theo Product Name 30](#_Toc182160303)

# DANH MỤC BẢNG

[Bảng 1: . Danh sách thuộc tính 6](#_Toc182068660)

[Bảng 2: Thuộc tính cần thiết cho bảng Fact 9](#_Toc182068661)

# CHƯƠNG 1 : TỔNG QUAN VỀ ĐỀ TÀI

## 1. Lý do chọn đề tài

Việc chọn đề tài phân tích dữ liệu chuỗi cung ứng xuất phát từ tầm quan trọng ngày càng tăng của chuỗi cung ứng trong bối cảnh nền kinh tế toàn cầu hóa và cạnh tranh khốc liệt hiện nay. Các doanh nghiệp luôn đối mặt với yêu cầu tối ưu hóa quy trình sản xuất, giảm chi phí, và nâng cao hiệu quả vận hành. Phân tích dữ liệu chuỗi cung ứng giúp các doanh nghiệp đưa ra quyết định chính xác hơn trong việc dự báo nhu cầu, quản lý tồn kho, tối ưu hóa lộ trình vận chuyển và quản lý rủi ro.

Hơn nữa, với sự phát triển mạnh mẽ của công nghệ và dữ liệu lớn, khả năng thu thập và phân tích dữ liệu trong chuỗi cung ứng trở nên dễ dàng và hiệu quả hơn, giúp doanh nghiệp phản ứng linh hoạt với những thay đổi bất ngờ và duy trì lợi thế cạnh tranh.

Do đó, nghiên cứu về phân tích dữ liệu chuỗi cung ứng không chỉ góp phần nâng cao hiệu quả hoạt động của doanh nghiệp mà còn mang lại những giá trị thiết thực trong việc xây dựng và duy trì một chuỗi cung ứng bền vững và tối ưu.

## 2. Các công cụ sử dụng

* Microsoft Visual Studio 2022.
* Microsoft SQL Server Management Studio 2022.
* Công cụ SQL Server Data Tools cho phiên bản Visual Studio 2022.

## 3. Mô tả dataset

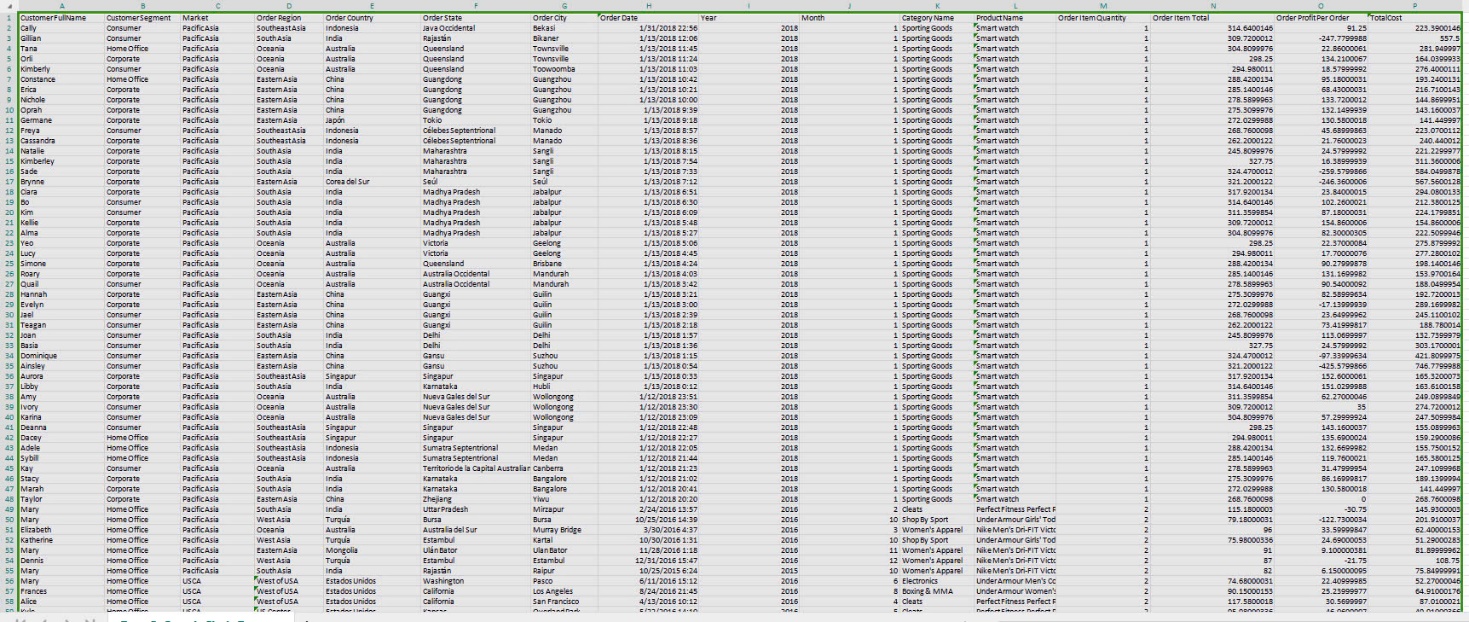
**Nơi lấy dữ liệu:** Kaggle

Kaggle là một nền tảng trực tuyến cho phép người dùng chia sẻ ý tưởng, tìm cảm hứng và tham gia vào các cuộc thi cùng với các nhà khoa học dữ liệu khác. Đây cũng là nơi để học hỏi các thông tin mới và các thủ thuật lập trình, đồng thời cung cấp các ví dụ ứng dụng khoa học dữ liệu trong thực tế. Kaggle có rất nhiều bộ dữ liệu từ các lĩnh vực khác nhau, từ dữ liệu đơn giản như doanh số bán trò chơi điện tử cho đến dữ liệu phức tạp như ô nhiễm không khí. Các dữ liệu này có nguồn gốc đáng tin cậy, tạo cơ hội để người dùng đào tạo và thử nghiệm các mô hình trên các dự án có tính ứng dụng cao.

**Tên dataset:** Gồm có file DataCoSupplyChainDataset

**Link dataset:** [DataCoSupplyChainDataset](https://www.kaggle.com/datasets/shashwatwork/dataco-smart-supply-chain-for-big-data-analysis/data)

Dữ liệu gồm có 79.993 dòng, 16 cột



Hình 1: Dữ liệu DataCoSupplyChainDataset

### 3.1. Danh sách thuộc tính được phân tích

|  |  |  |
| --- | --- | --- |
| **STT** | **Tên thuộc tính** | **Mô tả** |
| 1 | Customer FullName | Tên khách hàng |
| 2 | Customer Segment | Phân khúc khách hàng |
| 3 | Market | Thị trường |
| 4 | Order Region | Vùng đặt hàng |
| 5 | Order Country | Quốc gia đặt hàng |
| 6 | Order State | Trạng thái đơn hàng |
| 7 | Order City | Thành phố đặt hàng |
| 8 | Order Date | Ngày đặt hàng |
| 9 | Year | Năm đặt hàng |
| 10 | Month | Tháng đặt hàng |
| 11 | Category Name | Tên danh mục sản phẩm |
| 12 | Product Name | Tên sản phẩm |
| 13 | Order Item Quantity | Số lượng mặt hàng trong đơn hàng |
| 14 | Order Item Total | Tổng giá trị mặt hàng trong đơn hàng |
| 15 | Order Profit Per Order | Lợi nhuận trên mỗi đơn hàng |
| 16 | TotalCost | Tổng chi phí |

Bảng 1: . Danh sách thuộc tính

## 4. Xây dựng kho dữ liệu

Lược đồ đa chiều (multidimensional schema) là một cấu trúc dữ liệu được sử dụng để tổ chức các thông tin trong các hệ thống quản lý cơ sở dữ liệu, nhằm giúp cho việc truy xuất và phân tích dữ liệu trở nên dễ dàng và nhanh chóng hơn.

Lược đồ đa chiều được sử dụng trong các hệ thống OLAP (Online Analytical Processing) để tổ chức dữ liệu theo các kết nối giữa các thuộc tính và các số liệu có liên quan.

Có 2 loại lược đồ đa chiều: Lược đồ hình sao (star schema) và lược đồ hình bông tuyết (snowflake schema).

### 4.1. Lược đồ hình sao (Star schema)

Lược đồ hình sao (Star schema) là một loại lược đồ đa chiều được sử dụng trong hệ thống quản lý cơ sở dữ liệu phân tích (OLAP) và các ứng dụng Business Intelligence (BI). Nó được xây dựng dựa trên mô hình hóa dữ liệu hướng tài nguyên (Resource-oriented data modeling) và được sử dụng để tổ chức dữ liệu cho việc thực hiện các phân tích dữ liệu.

Trong lược đồ hình sao, dữ liệu được tổ chức thành một bảng trung tâm (fact table) và các bảng chi tiết (dimension tables) liên quan đến bảng trung tâm. Bảng trung tâm chứa các số liệu định lượng, trong khi các bảng chi tiết chứa các thông tin phân loại (thuộc tính). Mỗi bảng chi tiết đại diện cho một chiều của dữ liệu và được kết nối với bảng trung tâm thông qua một khóa ngoại.

* **Ưu điểm:**

*Tốc độ truy vấn nhanh*: Do dữ liệu được tổ chức thành các bảng trung tâm và các bảng chi tiết liên quan, việc truy vấn và phân tích dữ liệu trở nên nhanh chóng và hiệu quả hơn => xử lý khối lượng dữ liệu lớn dễ dàng hơn.

*Thích hợp cho phân tích dữ liệu*: Lược đồ hình sao được thiết kế để phục vụ cho các mục đích phân tích dữ liệu và đưa ra quyết định. Các bảng chi tiết đại diện cho các thuộc tính của dữ liệu, giúp cho người dùng có thể thực hiện các phân tích đa chiều và tìm kiếm các mối quan hệ giữa các thuộc tính khác nhau.

*Linh hoạt và dễ dàng mở rộng*: Lược đồ hình sao linh hoạt và có thể dễ dàng mở rộng khi cần thiết. Nếu cần thêm các thuộc tính mới vào hệ thống, người dùng chỉ cần thêm các bảng chi tiết mới và kết nối chúng với bảng trung tâm.

*Dễ dàng hiểu và sử dụng*: Lược đồ hình sao đơn giản và dễ hiểu. Các bảng trung tâm và bảng chi tiết được tổ chức theo cấu trúc phẳng, giúp cho người dùng dễ dàng hình dung và phân tích dữ liệu.

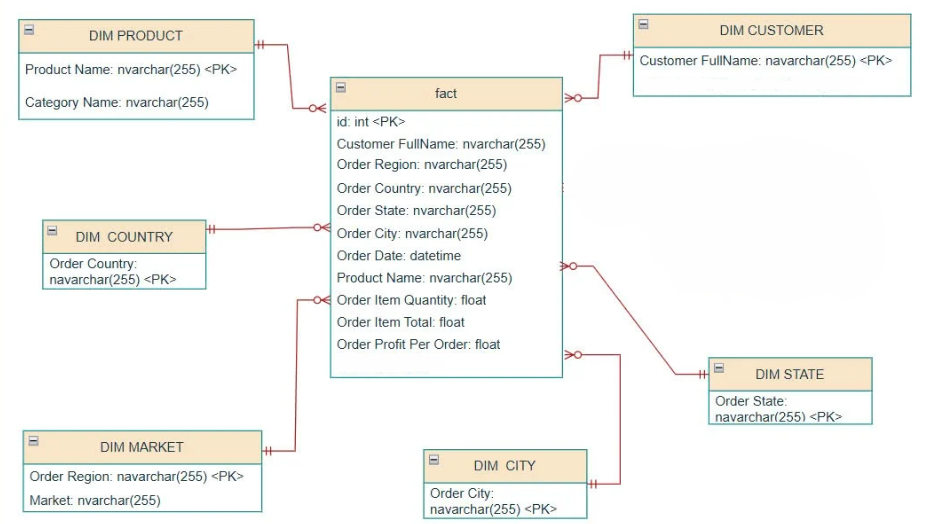
* **Nhược điểm:**

*Dữ liệu trùng lặp*: Một số dữ liệu có thể bị trùng lặp trong các bảng chi tiết. Điều này có thể xảy ra khi các bảng chi tiết có liên kết với nhiều bảng trung tâm khác nhau => làm tăng khối lượng dữ liệu, ảnh hưởng đến hiệu suất.

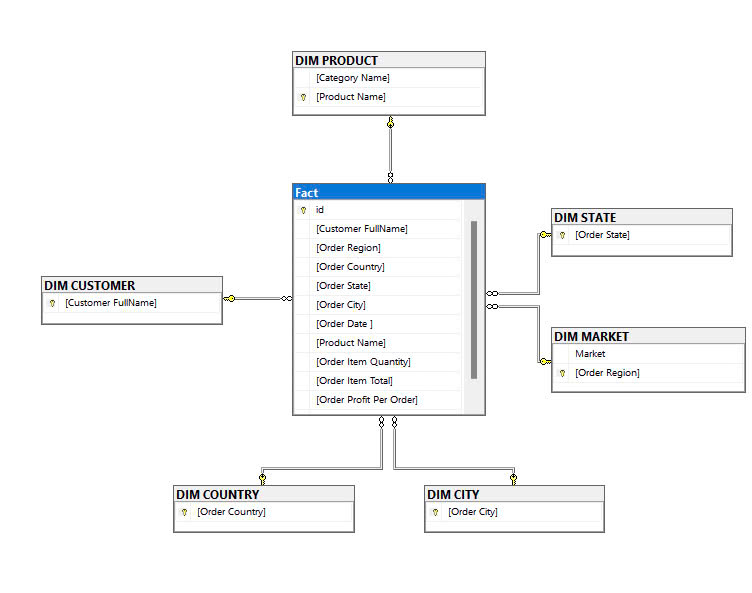
*Không phù hợp cho các tình huống phức tạp*: Lược đồ hình sao là một mô hình đơn giản, chỉ phù hợp với các tình huống phân tích dữ liệu đơn giản.

*Không linh hoạt cho các loại truy vấn khác nhau*: Khi cần thực hiện các loại truy vấn như truy vấn thống kê hoặc truy vấn văn bản đầy đủ, lược đồ hình sao có thể không phù hợp.

* **Thiết kế lược đồ:** Trong đó bảng Fact bao gồm Fact là bảng sự kiện chứa các khóa ngoại của các bảng Dimension và các thuộc tính khác. DIM PRODUCT, DIM CUSTOMER, DIM DATE, DIM SATE, DIM CITY, DIM COUNTRY, DIM MARKET là các bảng chiều chứa thuộc tính khóa chính và các thuộc tính khác của bảng.



Hình 2: Logical

****

Hình 3: Physical

# CHƯƠNG 2: PHÂN TÍCH VÀ THIẾT KẾ KHO DỮ LIỆU

## 1. Trước khi xử lý dữ liệu

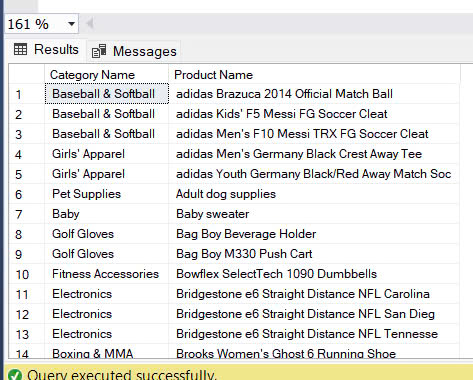
### 1.1. Những thuộc tính cần thiết cho bảng fact chính

|  |  |
| --- | --- |
| Order City | Thành phố đặt hàng |
| Order Country | Quốc gia đặt hàng |
| Customer FullName | Tên khách hàng |
| Order Region | Vùng đặt hàng |
| Product Name | Tên sản phẩm |
| Order Date | Ngày đặt hàng |
| Order State | Trạng thái đơn hàng |
| Order Item Quantity | Số lượng mặt hàng trong đơn hàng |
| Order Item Total | Tổng giá trị mặt hàng trong đơn hàng |
| Order Profit Per Order | Lợi nhuận trên mỗi đơn hàng |

Bảng 2: Thuộc tính cần thiết cho bảng Fact

### 1.2. Sau khi xử lý dữ liệu

#### 1.2.1. Các bảng Dim



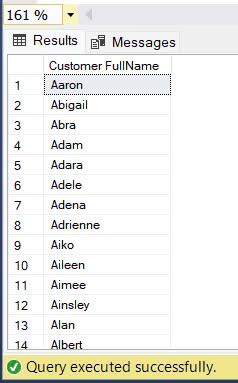
Hình 4: DIM PRODUCT



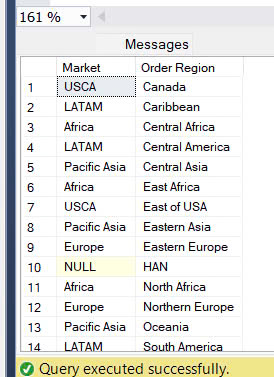
Hình 5: DIM CITY



Hình 6: DIM COUNTRY



Hình 7: DIM CUSTOMER

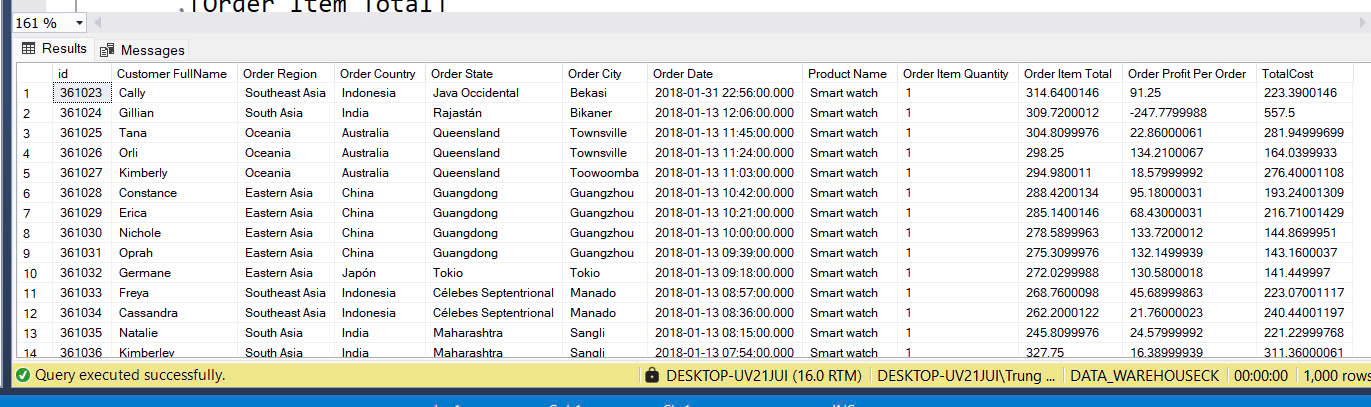


Hình 8: DIM MARKET



Hình 9: DIM STATE

#### 1.2.2. Bảng Fact



Hình 10: Bảng FACT

# CHƯƠNG 3: TRIỂN KHAI KHO DỮ LIỆU

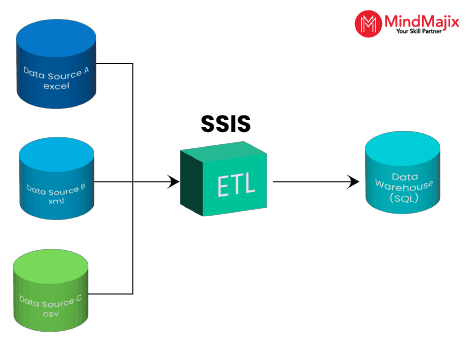
## 1. SSIS

SSIS là một phần của Microsoft SQL Server, được sử dụng để xây dựng các giải pháp tích hợp dữ liệu và xử lý dữ liệu trong môi trường Data Warehouse.

Data Warehouse là một hệ thống lưu trữ dữ liệu được tổ chức một cách logic và có cấu trúc, nhằm hỗ trợ việc phân tích dữ liệu và ra quyết định kinh doanh. Mục tiêu của một Data Warehouse là cung cấp cho người dùng cuối thông tin dễ dàng, tổ chức và có ý nghĩa từ các nguồn dữ liệu khác nhau.

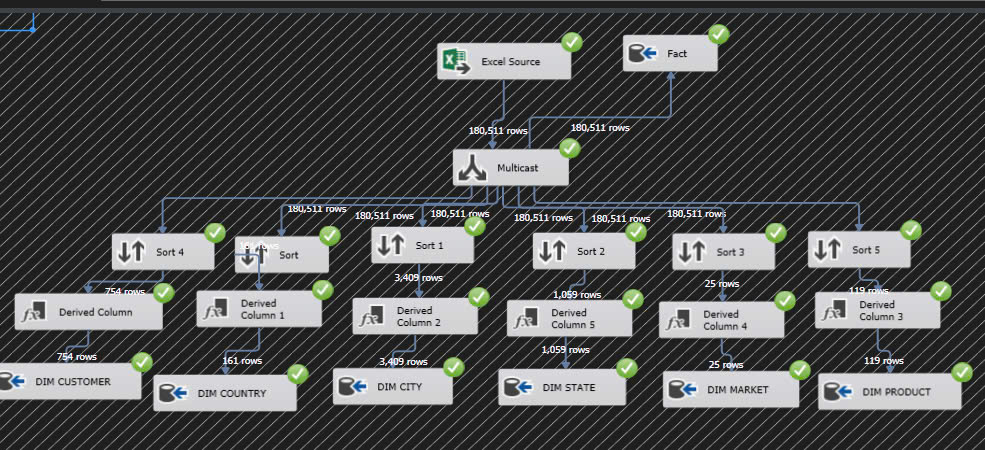
SSIS cung cấp một giao diện đồ họa và các công cụ để xây dựng quy trình tích hợp dữ liệu. Nó cho phép bạn trích xuất dữ liệu từ các nguồn khác nhau như cơ sở dữ liệu SQL Server, Oracle, Excel, CSV, và nhiều nguồn dữ liệu khác. Bạn có thể biến đổi, làm sạch và chuyển đổi dữ liệu theo yêu cầu của mình, sau đó xuất dữ liệu đã xử lý vào Data Warehouse hoặc hệ thống khác để phân tích.

SSIS cung cấp các công cụ để xây dựng các gói dữ liệu, trong đó có các thành phần như nguồn dữ liệu (data source), bước xử lý dữ liệu (data transformation), và đích đến dữ liệu (data destination). Bạn có thể tạo các luồng dữ liệu, áp dụng biến đổi dữ liệu, thực hiện kiểm tra và kiểm soát, và lên lịch thực hiện các gói SSIS.



Hình 11: Mô hình SSIS

### Quá trình đổ dữ liệu vào SSIS



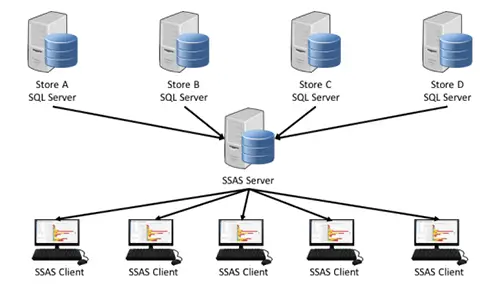
Hình 12: Data Flow

## 2. SSAS

SSAS là một công cụ phân tích và máy chủ OLAP đa chiều trong Microsoft SQL Server, hỗ trợ xử lý khối lượng dữ liệu lớn với phân tích đa chiều. SSAS có hai dạng chính là Đa chiều và Dạng bảng. Công cụ này cung cấp khả năng tương thích ngược cấp API, sử dụng OLEDB cho OLAP làm API truy cập và MDX làm ngôn ngữ truy vấn chính. SSAS hỗ trợ ba kiến trúc phân tích chính là MOLAP, HOLAP và ROLAP, cho phép linh hoạt trong việc lưu trữ và xử lý dữ liệu.

SSAS có thể hoạt động trong cả chế độ máy khách-máy chủ và chế độ ngoại tuyến, tạo sự linh hoạt trong việc truy cập dữ liệu. Công cụ này cung cấp các trình hướng dẫn và công cụ thiết kế, giúp việc tạo và quản lý mô hình dữ liệu dễ dàng hơn. Khả năng tùy chỉnh ứng dụng cũng được hỗ trợ rộng rãi, cho phép đáp ứng nhu cầu cụ thể của doanh nghiệp. Hơn nữa, SSAS cung cấp các tính năng bảo mật, báo cáo chi tiết, cấu trúc động và siêu dữ liệu được chia sẻ.

Quá trình xử lý dữ liệu trong SSAS bao gồm việc tổng hợp dữ liệu từ RDBMS và đưa vào các khối dữ liệu SSAS thông qua các dự án dịch vụ phân tích. Khi các khối dữ liệu này được xây dựng thành cơ sở dữ liệu phân tích, chúng có thể được sử dụng để thực hiện nhiều phân tích khác nhau, giúp doanh nghiệp tận dụng thông tin cho các mục đích báo cáo và ra quyết định.



Hình 13: Mô hình SSAS

### 2.1. Danh sách các câu truy vấn

* **Số lượng đơn hàng và tổng doanh thu theo khách hàng:**

SELECT dc.[Customer FullName],

COUNT(f.id) AS OrderCount,

SUM(f.[Order Item Total]) AS TotalRevenue

FROM [Fact] f

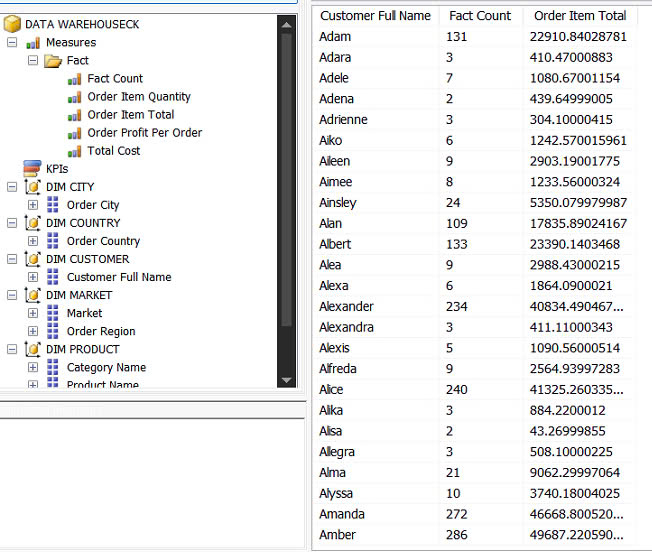
JOIN [DIM CUSTOMER] dc ON f.[Customer FullName] = dc.[Customer FullName]

GROUP BY dc.[Customer FullName]

ORDER BY TotalRevenue DESC;



Hình 14: Dữ liệu truy vấn số lượng đơn hàng và tổng doanh thu theo khách hàng



Hình 15: Số lượng đơn hàng và tổng doanh thu theo khách hàng

* **Tổng lợi nhuận theo từng thành phố:**

SELECT dci.[Order City],

SUM(f.[Order Profit Per Order]) AS TotalProfit

FROM [Fact] f

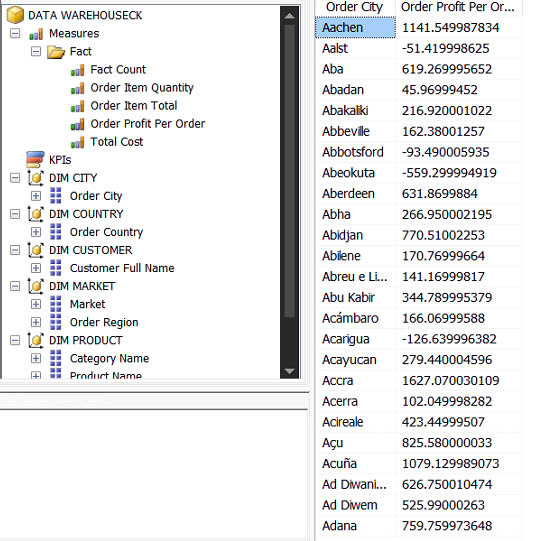
JOIN [DIM CITY] dci ON f.[Order City] = dci.[Order City]

GROUP BY dci.[Order City]

ORDER BY TotalProfit DESC;



Hình 16: Dữ liệu truy vấn Tổng lợi nhuận theo từng thành phố



Hình 17: Tổng lợi nhuận theo từng thành phố

* **Phân tích doanh thu theo thị trường:**

SELECT dm.[Market],

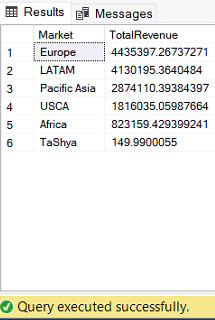
SUM(f.[Order Item Total]) AS TotalRevenue

FROM [Fact] f

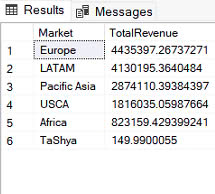
JOIN [DIM MARKET] dm ON f.[Order Region] = dm.[Order Region]

GROUP BY dm.[Market]

ORDER BY TotalRevenue DESC;​



Hình 18: Dữ liệu truy vấn Phân tích doanh thu theo thị trường



Hình 19: Phân tích doanh thu theo thị trường

* **Tổng doanh thu theo bang:**

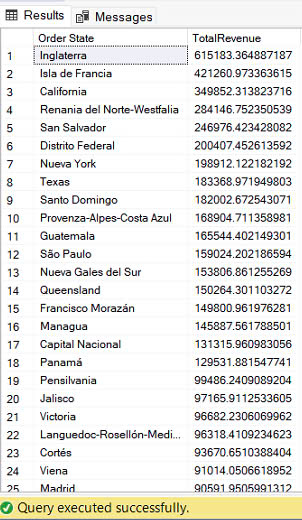
SELECT ds.[Order State], SUM(f.[Order Item Total]) AS TotalRevenue

FROM Fact f

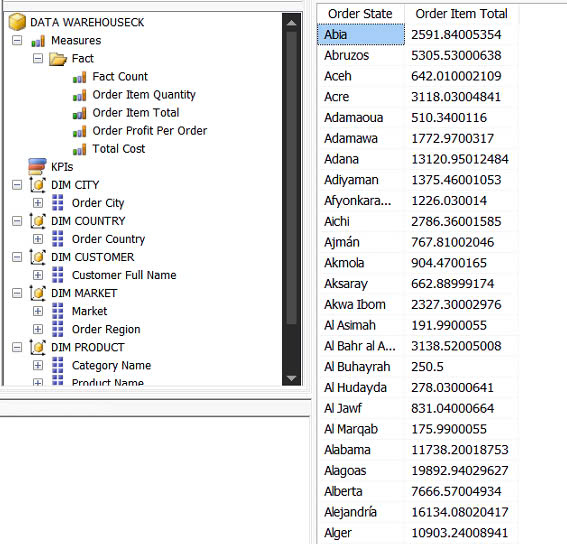
JOIN [DIM STATE] ds ON f.[Order State] = ds.[Order State]

GROUP BY ds.[Order State]

ORDER BY TotalRevenue DESC;



Hình 20: Dữ liệu truy vấn Tổng doanh thu theo bang



Hình 21: Tổng doanh thu theo bang

* **Tổng doanh thu theo từng quốc gia:**

SELECT dc.[Order Country],

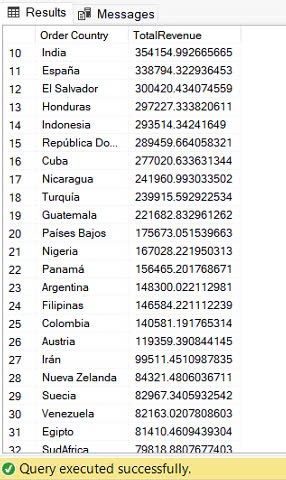
SUM(f.[Order Item Total]) AS TotalRevenue

FROM [Fact] f

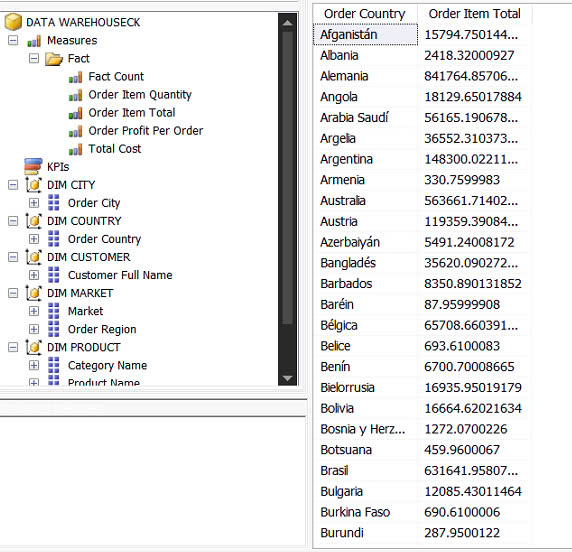
JOIN [DIM COUNTRY] dc ON f.[Order Country] = dc.[Order Country]

GROUP BY dc.[Order Country]

ORDER BY TotalRevenue DESC;



Hình 22: Dữ liệu truy vấn Tổng doanh thu theo từng quốc gia



Hình 23: Tổng doanh thu theo từng quốc gia

* **Tổng số lượng sản phẩm bán ra theo danh mục sản phẩm:**

SELECT dp.[Category Name], SUM(f.[Order Item Quantity]) AS TotalQuantity

FROM Fact f

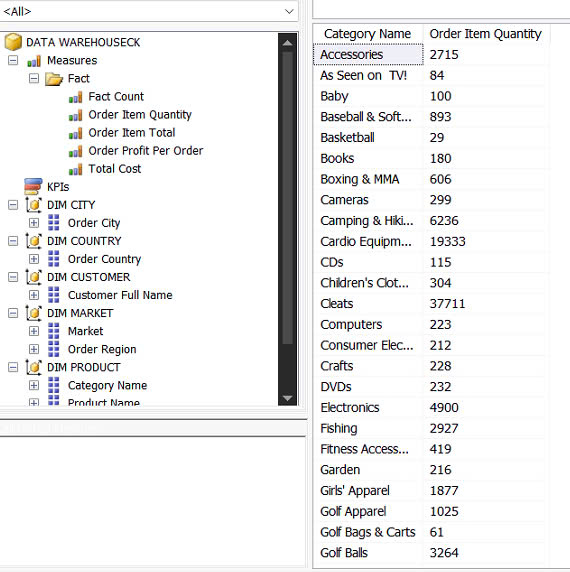
JOIN [DIM PRODUCT] dp ON f.[Product Name] = dp.[Product Name]

GROUP BY dp.[Category Name]

ORDER BY TotalQuantity DESC;



Hình 24: Dữ liệu truy vấn Tổng số lượng sản phẩm bán ra theo danh mục sản phẩm



Hình 25: Tổng số lượng sản phẩm bán ra theo danh mục sản phẩm

* **Doanh thu trung bình mỗi đơn hàng theo Product Name:**

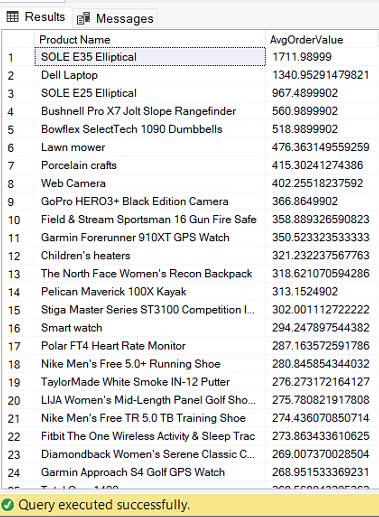
SELECT f.[Product Name],

AVG(f.[Order Item Total]) AS AvgOrderValue

FROM [Fact] f

GROUP BY f.[Product Name]

ORDER BY AvgOrderValue DESC;



Hình 26: Dữ liệu truy vấn Doanh thu trung bình mỗi đơn hàng theo Product Name



Hình 27: Doanh thu trung bình mỗi đơn hàng theo Product Name

# KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

## 1. Kết quả đạt được

Dự án xây dựng kho dữ liệu cho mục tiêu phân tích dữ liệu chuỗi cung ứng đã hoàn thành các bước cơ bản, từ thu thập, làm sạch và chuyển đổi dữ liệu đến việc xây dựng cấu trúc kho dữ liệu, phục vụ cho việc phân tích hiệu quả. Các công cụ như SSIS (SQL Server Integration Services), SSAS (SQL Server Analysis Services) đã mang lại nhiều lợi ích thiết thực trong quá trình phát triển:

SSIS đã hỗ trợ hiệu quả việc kết nối và xử lý dữ liệu từ nhiều nguồn khác nhau trong chuỗi cung ứng, bao gồm dữ liệu từ các nhà cung cấp, kho hàng, hệ thống vận chuyển, và các bộ phận liên quan. Điều này đảm bảo chất lượng và tính chính xác của dữ liệu trước khi lưu trữ trong kho dữ liệu, giúp các bộ phận quản lý dễ dàng theo dõi và xử lý thông tin liên quan đến hàng hóa, tồn kho, và vận chuyển.

SSAS giúp tạo các mô hình phân tích đa chiều và các cubes, làm nền tảng cho việc phân tích chi tiết về hiệu suất chuỗi cung ứng, từ tỷ lệ giao hàng đúng hạn, thời gian vận chuyển trung bình, đến mức độ tồn kho và yêu cầu sản xuất. Những mô hình phân tích này cung cấp cái nhìn sâu sắc về các yếu tố ảnh hưởng đến hiệu quả chuỗi cung ứng, giúp các nhà quản lý đưa ra các quyết định chính xác và nhanh chóng.

Kho dữ liệu này đã nâng cao khả năng phân tích và dự báo, từ đó hỗ trợ các bộ phận quản lý chuỗi cung ứng đưa ra các quyết định tối ưu, cải thiện quy trình vận hành, tối ưu hóa lượng tồn kho, và nâng cao hiệu quả tổng thể trong quá trình cung cấp sản phẩm.

## Hướng phát triển

Tích hợp thêm nguồn dữ liệu thời gian thực: Mở rộng kho dữ liệu để tích hợp thêm các nguồn dữ liệu thời gian thực từ các cảm biến IoT và hệ thống theo dõi GPS trên xe vận chuyển. Điều này giúp theo dõi hàng hóa, giám sát điều kiện vận chuyển, và phát hiện các vấn đề nhanh chóng, cải thiện khả năng phản ứng tức thì của chuỗi cung ứng.

Áp dụng các mô hình Machine Learning và AI để cải thiện khả năng dự báo nhu cầu, phát hiện các xu hướng tiêu thụ, dự đoán lượng tồn kho tối ưu và lập kế hoạch sản xuất. Điều này giúp giảm chi phí lưu kho, giảm thiểu hàng tồn kho dư thừa hoặc thiếu hụt và tăng cường hiệu quả chuỗi cung ứng.

Tích hợp mô hình phân tích rủi ro: Thiết lập các chỉ số rủi ro để phân tích và đánh giá các yếu tố ảnh hưởng đến chuỗi cung ứng, chẳng hạn như các vấn đề về cung ứng, môi trường, địa chính trị và dịch bệnh. Mô hình này sẽ hỗ trợ lập kế hoạch dự phòng và tăng cường tính linh hoạt của chuỗi cung ứng.

Xây dựng bảng điều khiển (Dashboard) phân tích tương tác: Phát triển các bảng điều khiển phân tích trực quan, tương tác, cho phép các nhà quản lý theo dõi các chỉ số hiệu suất (KPI) như tỷ lệ giao hàng đúng hạn, mức tồn kho, và chi phí vận chuyển theo thời gian thực và có thể tùy chỉnh theo nhu cầu phân tích cụ thể.

Nâng cấp hệ thống phân quyền và bảo mật: Tăng cường hệ thống bảo mật dữ liệu, đảm bảo quyền truy cập theo vai trò và trách nhiệm trong chuỗi cung ứng, tránh tình trạng rò rỉ thông tin nhạy cảm, bảo vệ dữ liệu khách hàng và thông tin hợp tác của nhà cung cấp.

# KẾT LUẬN

## Kết quả đạt được

Tìm hiểu và áp dụng kiến thức về xây dựng kho dữ liệu đạt được những kết quả sau:

* Hiểu rõ các khái niệm cơ bản về kho dữ liệu cũng như các tính chất cần thiết mà một kho dữ liệu phải có.
* Nắm vững kiến thức và có khả năng xây dựng một kho dữ liệu hoàn chỉnh để phục vụ cho việc khai thác dữ liệu.
* Xây dựng thành công một kho dữ liệu hoàn chỉnh.
* Tối ưu hóa các câu truy vấn để đạt hiệu quả cao.

Áp dụng kiến thức về SSIS, SSAS để:

* Phân tích doanh thu trung bình mỗi đơn hàng theo sản phẩm
* Phân tích tổng số lượng sản phẩm bán ra theo danh mục sản phẩm
* Phân tích tổng doanh thu theo từng quốc gia
* Phân tích tổng lợi nhuận theo từng thành phố
* Phân tích số lượng đơn hàng và tổng doanh thu theo khách hàng
* Phân tích doanh thu theo thị trường
* Tổng doanh thu theo bang

# TÀI LIỆU THAM KHẢO

[1]. DataCo SMART SUPPLY CHAIN

<https://www.kaggle.com/datasets/shashwatwork/dataco-smart-supply-chain-for-big-data-analysis/data>

[2]. ETL project with SSIS:

[ETL project with SSIS: Lookup](https://www.youtube.com/watch?app=desktop&v=8JyOFgrhjL0)

[4]. ETL with SSIS: Merge Join Table and Derived Column,

[ETL with SSIS: Merge Join Table and Derived Column](https://www.youtube.com/watch?v=smZT4aoaKMU&list=PL2vTwB9LeGfDndnObrtpGHr2_sQLV41oq&index=13)

[5]. Analysis Services - SSAS [[https://learn.microsoft.com/en-us/analysis-services/multidimensional-models/create-an-analysis-services-project-ssdt?view=asallproducts-allversions](https://learn.microsoft.com/en-us/analysis-services/multidimensional-models/create-an-analysis-services-project-ssdt?view=asallproducts-allversions%20)]