**BỘ GIÁO DỤC VÀ ĐÀO TẠO**

**TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT TP.HỒ CHÍ MINH**

**KHOA CÔNG NGHỆ THÔNG TIN**



**BÁO CÁO ĐỒ ÁN MÔN HỌC**

**KHO DỮ LIỆU**

**ĐỀ TÀI:**

**PHÂN TÍCH DỮ LIỆU TẤN CÔNG MẠNG**

GVHD: ThS. Nguyễn Văn Thành

Nhóm sinh viên thực hiện:

1. Lê Hồ Quốc Huy MSSV: 22133025

2. Nguyễn Duy Nam MSSV: 18133031

3. Trần Bảo Việt MSSV: 22133065

4. Lê Quỳnh Nhựt Vinh MSSV: 22133066

**TP. HỒ CHÍ MINH, 05/2025**

**NHẬN XÉT CỦA GIÁO VIÊN HƯỚNG DẪN**

*……., ngày……...tháng……năm 2025*

**Người nhận xét**

*(Ký tên và ghi rõ họ tên****)***

PHÂN CÔNG NHIỆM VỤ

| **Nhiệm vụ** | **Trần Bảo Việt**  **22133065** | **Lê Quỳnh Nhựt Vinh**  **22133066** | **Lê Hồ Quốc Huy**  **22133025** | **Nguyễn Duy Nam**  **18133031** |
| --- | --- | --- | --- | --- |
| Tìm kiếm tập dữ liệu | X | X | X | X |
| Hiểu tập dữ liệu | X | X | X | X |
| Xác định Business Process | X | X | X | X |
| Xác định bảng Dim | X |  |  | X |
| Xác định bảng Fact | X |  |  | X |
| Đẩy dữ liệu từ CSV – SQL Server | X |  |  | X |
| Tạo nguồn kết nối dữ liệu | X |  |  | X |
| Staging và load và các dim, fact | X |  |  | X |
| Nhập dữ liệu vào SSAS và tạo Data Cube |  |  | X |  |
| Phân tích SSAS |  |  | X |  |
| Đặt các câu hỏi | X | X | X | X |
| Trả lời các câu hỏi bằng SSAS |  | X | X |  |
| Trả lời các câu hỏi bằng Pivot Table |  | X | X |  |
| Report PowerBI |  | X |  |  |
| Viết báo cáo và trình bày | X | X | X | X |

**MỤC LỤC**

[**1.1. Lý do chọn đề tài. 2**](#_heading=h.eob3zx7xjtw6)

[**1.2.1. Tổng quan dataset 3**](#_heading=h.ygash6hr59m8)

[**1.3. Tiền xử lý dữ liệu 4**](#_heading=h.palddgw00dob)

[**1.4. Thiết kế kho dữ liệu 6**](#_heading=h.9aklpktwm59n)

[**1.4.1. Lược đồ kho dữ liệu 6**](#_heading=h.qgoikrkwy0zw)

[**1.4.2. Mô tả các bảng trong kho dữ liệu 7**](#_heading=h.33x505hs073)

[**1.5. Các câu truy vấn 9**](#_heading=h.3fefc21dhs5p)

[**2.1. Chuẩn bị các công cụ 10**](#_heading=h.o6t2t3y5x1dj)

[**2.2. Chuẩn bị cơ sở dữ liệu 10**](#_heading=h.9qrdjc8r4gjj)

[**2.3. Tạo mới project SSIS 11**](#_heading=h.kh571slq62kr)

[**2.4. Tạo bảng Dim và bảng Fact 11**](#_heading=h.g1cqfqo1kv1x)

[**2.4.1. Bảng Dim\_Timestamp 15**](#_heading=h.6k9ssi1p8yru)

[**2.4.2. Bảng Dim\_Address 21**](#_heading=h.1ry4xkrvevtq)

[**2.4.3. Bảng Dim\_Attack 25**](#_heading=h.fahhwunxmig6)

[**2.4.4. Bảng Dim\_Victim 31**](#_heading=h.g9xf0xfkax1u)

[**2.4.5. Bảng Fact\_CyberSecurity 37**](#_heading=h.xpg2w3jatgcj)

[**2.4.6. Tạo khoá ngoại 43**](#_heading=h.bw3h5wz4887v)

[**2.4.7. Chạy SSIS 45**](#_heading=h.b8h8b8s49nf0)

[**2.5. Kiểm tra dữ liệu các bảng 49**](#_heading=h.s2dpg7notm2a)

[**2.6. Lược đồ sau khi hoàn thành 53**](#_heading=h.vih13hati7af)

[**3.1. Tạo project SSAS mới 54**](#_heading=h.d1eaon5vsixu)

[**3.2. Xác định dữ liệu nguồn (Data Source) 55**](#_heading=h.zezdmv5j3wgq)

[**3.3. Xác định khung nhìn dữ liệu nguồn (Data Source View) 59**](#_heading=h.g81vyaanm4vj)

[**3.4. Xây dựng các khối (Cube) và deploy Cube 64**](#_heading=h.61x6sdiojqao)

[**3.4.1.1. Tạo Cube và Dimension 64**](#_heading=h.rwo1zzg00iij)

[**3.4.2. Thêm thuộc tính và chỉnh sửa property cho Dimension 69**](#_heading=h.uazggqyq6g21)

[**3.4.3. Deploy project SSAS 73**](#_heading=h.1jruk5b7ia34)

[**3.5. Xác định các độ đo (Measures) 78**](#_heading=h.8kdepwv2b3p2)

[**3.6. Phân cấp bảng chiều 80**](#_heading=h.29qj70la1frf)

[**3.7. Thực hiện các truy vấn sử dụng SSAS, Pivot table 87**](#_heading=h.kfxrs2f69bek)

[**3.7.1. Câu truy vấn 1 87**](#_heading=h.r6km723b8uq6)

[**3.7.1.1. Sử dụng SSAS 87**](#_heading=h.hs9scbvjggk)

[**3.7.1.2. Sử dụng Pivot table 88**](#_heading=h.2emq4m3jbthu)

[**3.7.2. Câu truy vấn 2 88**](#_heading=h.32yk9e3469rn)

[**3.7.2.1. Sử dụng SSAS 89**](#_heading=h.mgrp2j1y9u2l)

[**3.7.2.2. Sử dụng Pivot table 89**](#_heading=h.vi6m6ys64b5b)

[**3.7.3. Câu truy vấn 3 90**](#_heading=h.tqu7tcc56ilz)

[**3.7.3.1. Sử dụng SSAS 90**](#_heading=h.ssqtkarlqtr)

[**3.7.3.2. Sử dụng Pivot table 91**](#_heading=h.vihgmtae6qqd)

[**3.7.4. Câu truy vấn 4 92**](#_heading=h.n6z8wnthwd73)

[**3.7.4.1. Sử dụng SSAS 92**](#_heading=h.ylk5qc6dnczj)

[**3.7.4.2. Sử dụng Pivot table 93**](#_heading=h.9jdmfio8yfo4)

[**3.7.5. Câu truy vấn 5 93**](#_heading=h.lclp9vlwezdv)

[**3.7.5.1. Sử dụng SSAS 94**](#_heading=h.57fk7j7ow5ry)

[**3.7.5.2. Sử dụng Pivot table 96**](#_heading=h.gfnnwxqdffnm)

[**3.7.6. Câu truy vấn 6 96**](#_heading=h.7gelp74bicyq)

[**3.7.6.1. Sử dụng SSAS 96**](#_heading=h.th4aa6s5sbd3)

[**3.7.6.2. Sử dụng Pivot table 97**](#_heading=h.2t446oraz5or)

[**3.7.7. Câu truy vấn 7 97**](#_heading=h.8g7k7q93w3zo)

[**3.7.7.1. Sử dụng SSAS 97**](#_heading=h.izcc0mc0w2jw)

[**3.7.7.2. Sử dụng Pivot table 98**](#_heading=h.9kxalzr21bl1)

[**3.7.8. Câu truy vấn 8 99**](#_heading=h.yz6sbagzirhs)

[**3.7.8.1. Sử dụng SSAS 99**](#_heading=h.rju5y9uwbjdl)

[**3.7.8.2. Sử dụng Pivot table 99**](#_heading=h.v9th8os289on)

[**3.7.9. Câu truy vấn 9 99**](#_heading=h.rme3jxcyjet2)

[**3.7.9.1. Sử dụng SSAS 100**](#_heading=h.2kor89ze2ds5)

[**3.7.9.2. Sử dụng Pivot table 101**](#_heading=h.i56eek2wg1bi)

[**3.7.10. Câu truy vấn 10 101**](#_heading=h.bi542sd6xra2)

[**3.7.10.1. Sử dụng SSAS 102**](#_heading=h.1xcqecwkhyy7)

[**3.7.10.2. Sử dụng Pivot table 102**](#_heading=h.songqwj0pg4w)

[**3.7.11. Câu truy vấn 11 103**](#_heading=h.xpopugi4bic3)

[**3.7.11.1. Sử dụng SSAS 103**](#_heading=h.14pd70wzcqmf)

[**3.7.11.2. Sử dụng Pivot table 104**](#_heading=h.ssq50gfvvij)

[**4.1. Quá trình lập báo biểu bằng PowerBI 106**](#_heading=h.xqiw55wcxkff)

[**4.1.1. Chuẩn bị công cụ 106**](#_heading=h.cxdbg5xukipv)

[**4.1.2. Tạo mới report và kết nối 106**](#_heading=h.6w6q1deq5sm7)

[**4.1.1. Tạo report 1 108**](#_heading=h.hsb5iwqg8pao)

[**4.1.2. Tạo report 2 112**](#_heading=h.ew8d3qw7v8f2)

[**4.1.3. Tạo report 3 118**](#_heading=h.m0fzwp728jz1)

[**4.2. Quá trình lập báo biểu bằng Google Data Studio 122**](#_heading=h.rx29powq6j1p)

[**4.2.1. Tạo report 1 122**](#_heading=h.bdv4rtvbyi5o)

[**4.2.2. Tạo report 2 126**](#_heading=h.i36e039djw7f)

[**4.2.3. Tạo report 3 133**](#_heading=h.74dv0dw4yci2)

[**DANH MỤC TÀI LIỆU THAM KHẢO 138**](#_heading=h.dhmhqedcmgna)

CHƯƠNG 1. TỔNG QUAN ĐỀ TÀI

# Lý do chọn đề tài.

Lĩnh vực an ninh mạng đang ngày càng trở nên quan trọng trong thời đại số hóa hiện nay. Với sự phát triển nhanh chóng của công nghệ và sự phụ thuộc ngày càng tăng của các tổ chức, doanh nghiệp và cá nhân vào hệ thống thông tin, việc bảo vệ dữ liệu và tài sản số trở thành một ưu tiên hàng đầu. Phân tích dữ liệu về các cuộc tấn công an ninh mạng giúp chúng ta hiểu rõ hơn về các mối đe dọa, phương thức tấn công và xu hướng mới nhất trong lĩnh vực này, từ đó có thể phát triển các chiến lược phòng thủ hiệu quả hơn.

Hơn nữa, việc nghiên cứu và phân tích dữ liệu về các cuộc tấn công an ninh mạng còn mang lại giá trị to lớn cho cộng đồng và xã hội. Kết quả từ các phân tích này có thể giúp cải thiện các hệ thống bảo mật, phát triển các công cụ phòng chống mới, và nâng cao nhận thức của người dùng về an ninh mạng. Điều này không chỉ góp phần bảo vệ thông tin cá nhân và tài sản của người dùng mà còn giúp doanh nghiệp và tổ chức giảm thiểu rủi ro và tổn thất tài chính do các cuộc tấn công gây ra. Ngoài ra, trong bối cảnh các cuộc tấn công mạng ngày càng tinh vi và quy mô lớn, việc nghiên cứu sâu về chủ đề này còn có thể đóng góp vào việc xây dựng các chính sách và quy định pháp lý liên quan đến an ninh mạng, tạo ra một môi trường số an toàn và đáng tin cậy hơn cho tất cả mọi người.

* 1. **Giới thiệu về dataset.**

## Tổng quan dataset

Tên bộ dữ liệu: Cyber Security Attacks

Tác giả: Aashray Agur và Uma Venugopal

Bộ dữ liệu cung cấp thông tin về các cuộc tấn công mạng trên toàn cầu từ năm 2020 đến năm 2023

Nguồn tải dataset: https://www.kaggle.com/datasets/teamincribo/cyber-security-attacks

Bộ dữ liệu gồm có 25 cột và 40000 dòng

* + 1. **Mô tả thuộc tính**

| STT | Tên thuộc tính | Kiểu dữ liệu | Ý nghĩa |
| --- | --- | --- | --- |
| 1 | Timestamp | timestamp | Thời điểm chính xác khi sự kiện xảy ra |
| 2 | Source IP Address | varchar | Địa chỉ IP của thiết bị gửi dữ liệu |
| 3 | Destination IP Address | varchar | Địa chỉ IP của thiết bị nhận dữ liệu |
| 4 | Source Port | varchar | Số cổng trên thiết bị gửi dữ liệu |
| 5 | Destination Port | varchar | Số cổng trên thiết bị nhận dữ liệu |
| 6 | Protocol | varchar | Quy tắc truyền thông được sử dụng (TCP, UDP) |
| 7 | Packet Length | int | Kích thước gói dữ liệu |
| 8 | Packet Type | varchar | Chỉ định loại gói tin trong mạng (Data, Control) |
| 9 | Traffic Type | varchar | Phân loại lưu lượng mạng (HTTP, FTP, Email) |
| 10 | Payload Data | varchar | Nội dung của gói dữ liệu |
| 11 | Malware Indicators | varchar | Dấu hiệu của phần mềm độc hại |
| 12 | Anomaly Scores | float | Đánh giá mức độ bất thường |
| 13 | Alerts/Warnings | varchar | Cảnh báo về sự kiện nguy hiểm |
| 14 | Attack Type | varchar | Phương thức tấn công (DDoS, phishing) |
| 15 | Attack Signature | varchar | Đặc điểm nhận dạng của cuộc tấn công |
| 16 | Action Taken | varchar | Biện pháp đối phó (chặn, cách ly) |
| 17 | Severity Level | varchar | Mức độ nguy hiểm của sự cố |
| 18 | User Information | varchar | Thông tin người dùng |
| 19 | Device Information | varchar | Thông tin thiết bị liên quan |
| 20 | Network Segment | varchar | Phân đoạn mạng nơi sự kiện xảy ra |
| 21 | Geo-location Data | varchar | Vị trí địa lý của các IP liên quan |
| 22 | Proxy Information | varchar | Thông tin về proxy được sử dụng |
| 23 | Firewall Logs | varchar | Nhật ký hoạt động của tường lửa |
| 24 | IDS/IPS Alerts | varchar | Cảnh báo từ hệ thống phát hiện/ngăn chặn xâm nhập |
| 25 | Log Source | varchar | Nguồn gốc của nhật ký (Server, Firewall) |

# Tiền xử lý dữ liệu

**Bước 1**: Import các thư viện cần thiết cho việc xử lý dữ liệu và đọc file csv

import pandas as pd

import requests

from time import sleep

import xlrd

from xlutils.copy import copy

file\_path = '/Users/tranbaoviet/Downloads/file\_moi.csv'

df = pd.read\_csv(file\_path)

**Bước 2**: Tiến hành điền dữ liệu null vào các ô trống của các cột như: Malware Indicators, IDS/ÍP Alerts, Firewall Logs

df['IDS/IPS Alerts'] = df['IDS/IPS Alerts'].fillna('No data')

df['Firewall Logs'] = df['Firewall Logs'].fillna('No records')

df['Malware Indicators'].fillna('No detected')

df['Proxy Information'].fillna('No detected)

**Bước 3**: Chuyển đổi dữ liệu Source IP Address, Destination IP Address sang địa chỉ vật lí bằng cách gọi API của ipinfo.io

TOKEN = "bd75ae0f9db684"

def get\_location(ip):

    try:

        response = requests.get(f"https://ipinfo.io/{ip}?token={TOKEN}")

        data = response.json()

        city = data.get('city', '')

        region = data.get('region', '')

        country = data.get('country', '')

        return f"{city}, {region}, {country}".strip(', ')

    except:

        return "Unknown"

    finally:

        sleep(0.1)  # Để tránh vượt quá giới hạn tốc độ của API

# Kiểm tra xem cột 'Source IP Address' có tồn tại không

if 'Source IP Address' in df.columns:

    # Tạo một cột mới 'Source Address' và thêm thông tin vị trí dựa trên IP

    df['Source Address'] = df['Source IP Address'].apply(get\_location)

# Làm tương tự như Source IP Address

if 'Destination IP Address' in df.columns:

    df['Destination Address'] = df['Destination IP Address'].apply(get\_location)

**Bước 4**: Sau khi chuyển đổi IP sang địa chỉ vật lí, sẽ tồn tại những record không chứa dữ liệu vì IP bị lỗi, tiến hành làm sạch dữ liệu

df\_cleaned = df.dropna(subset=['Source Address', 'Destination Address'])

**Bước 5**: Sau khi chuyển đổi IP dữ liệu sẽ có dạng city,region,country. Nhóm em sẽ thống kê theo country nên sẽ tách cột country ra

def extract\_country(address):

    return address.split(",")[-1].strip()

df['Source Country'] = df['Source Address'].apply(extract\_country)

df['Destination Country'] = df['Destination Address'].apply(extract\_country)

**Bước 6**: Xoá các cột không có dữ liệu country.

df = df[df['Source Address'] != 'Unknown']

df = df[df['Destination Address'] != 'Unknown']

**Bước 7**: Lưu file csv.

df.to\_csv('cyber\_security\_final1.csv', index=False)

Bộ dữ liệu sau khi xử lý sẽ gồm 30 cột và 39560 dòng.

# Thiết kế kho dữ liệu

## Lược đồ kho dữ liệu

**A screenshot of a computer

Description automatically generated**

## Mô tả các bảng trong kho dữ liệu

* + - 1. **Bảng Fact\_CyberSecurity**

| STT | Tên thuộc tính | Kiểu dữ liệu | Ràng buộc | Mô tả thuộc tính |
| --- | --- | --- | --- | --- |
| 1 | FactID | int | Khoá chính | Mã fact |
| 2 | TimeStampID | int | Khoá ngoại | Mã thời gian xảy ra cuộc tấn công |
| 3 | PacketLength | int |  | Độ dài gói tin được gửi đi |
| 4 | AnomalyScores | float |  | Chỉ số bất thường |

* + - 1. **Bảng Dim\_Attack**

| STT | Tên thuộc tính | Kiểu dữ liệu | Ràng buộc | Mô tả thuộc tính |
| --- | --- | --- | --- | --- |
| 1 | AttackID | int | Khoá chính | Mã tấn công |
| 2 | Protocol | varchar |  | Quy tắc truyền thông được sử dụng (TCP, UDP) |
| 3 | PacketType | varchar |  | Chỉ định loại gói tin trong mạng (Data, Control) |
| 4 | AttackType | varchar |  | Phương thức tấn công (DDoS, phishing) |
| 5 | TrafficType | varchar |  | Phân loại lưu lượng mạng (HTTP, FTP, Email) |
| 6 | ProxyInformation | varchar |  | Thông tin về proxy được sử dụng |
| 7 | SourceAddress | int | Khoá ngoại | Mã địa chỉ người gửi gói tin |

* + - 1. **Bảng Dim\_Victim**

| STT | Tên thuộc tính | Kiểu dữ liệu | Ràng buộc | Mô tả thuộc tính |
| --- | --- | --- | --- | --- |
| 1 | VictimID | int | Khoá chính | Mã nạn nhân |
| 2 | ActionTaken | varchar |  | Biện pháp đối phó (Blocked, Ignored) |
| 3 | SeverityLevel | varchar |  | Mức độ nguy hiểm của sự cố |
| 4 | AlertsWarnings | varchar |  | Cảnh báo về sự kiện nguy hiểm |
| 5 | DeviceInformation | varchar |  | Thông tin thiết bị liên quan |
| 6 | DestionationAddress | int | Khoá ngoại | Mã địa chỉ người nhận gói tin |

* + - 1. **Bảng Dim\_Address**

| STT | Tên thuộc tính | Kiểu dữ liệu | Ràng buộc | Mô tả thuộc tính |
| --- | --- | --- | --- | --- |
| 1 | AddressID | int | Khoá chính | Mã địa chỉ |
| 2 | Country | varchar |  | Quốc gia |

* + - 1. **Bảng Dim\_Timestamp**

| STT | Tên thuộc tính | Kiểu dữ liệu | Ràng buộc | Mô tả thuộc tính |
| --- | --- | --- | --- | --- |
| 1 | TimestampID | int | Khoá chính | Mã thời gian |
| 2 | Timestamp | date |  | Ngày Tháng Năm |
| 3 | Day | int |  | Ngày |
| 4 | Month | int |  | Tháng |
| 5 | Year | int |  | Năm |

# Các câu truy vấn

1. Truy vấn drill down tổng số cuộc tấn công bằng phương thức TCP theo từng tháng năm 2022.
2. Liệt kê top 5 cuộc tấn công có điểm cao nhất và phương thức của chúng.
3. Liệt kê tổng số cuộc tấn công từng tháng trong năm với loại là mã độc.
4. Truy vấn tổng điểm Anomaly Scores của các cuộc tấn công theo từng Phương thức (Protocol) và Năm với phương thức là Malware.
5. Truy vấn tính điểm trung bình của các cuộc tấn công Malware theo từng tháng trong năm 2022.
6. Liệt kê số cuộc tấn công bị chặn theo năm.
7. Truy vấn chiều dài các gói tin đã bị hệ thống bảo mật bỏ qua theo từng năm.
8. Liệt kê số lượng các cuộc tấn công thành công theo từng phương thức.
9. Liệt kê top 10 những cuộc tấn công có độ dài lớn nhất của hai nước Nhật Bản và Mỹ.
10. Truy vấn cuộc tấn công đã được cảnh báo nhưng bị bỏ qua trong năm 2020.
11. Tổng độ dài các gói tin của từng phương thức của 20 cuộc tấn công có điểm AnomalyScore cao nhất.

CHƯƠNG 2. QUÁ TRÌNH XÂY DỰNG KHO DỮ LIỆU (SSIS)

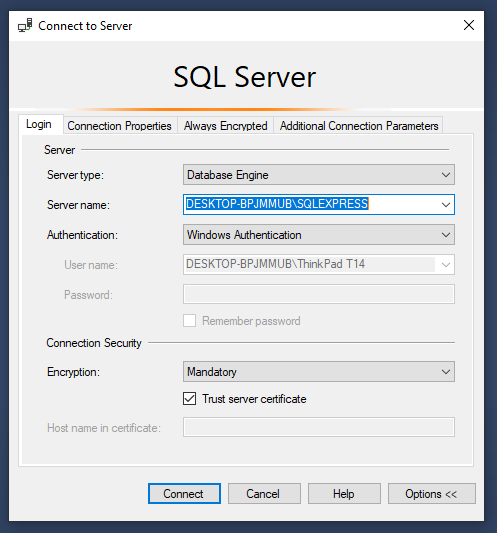
# Chuẩn bị các công cụ

Để thực hiện được quá trình SSIS ta cần chuẩn bị và cài đặt các công cụ sau:

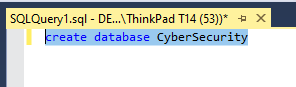
* Visual studio Community 2022
* SQL Server Integration Services Project

# Chuẩn bị cơ sở dữ liệu

**Bước 1:** Mở SQL Server v20.2 và kết nối với server bằng tài khoản user của window (Windows Authentication).

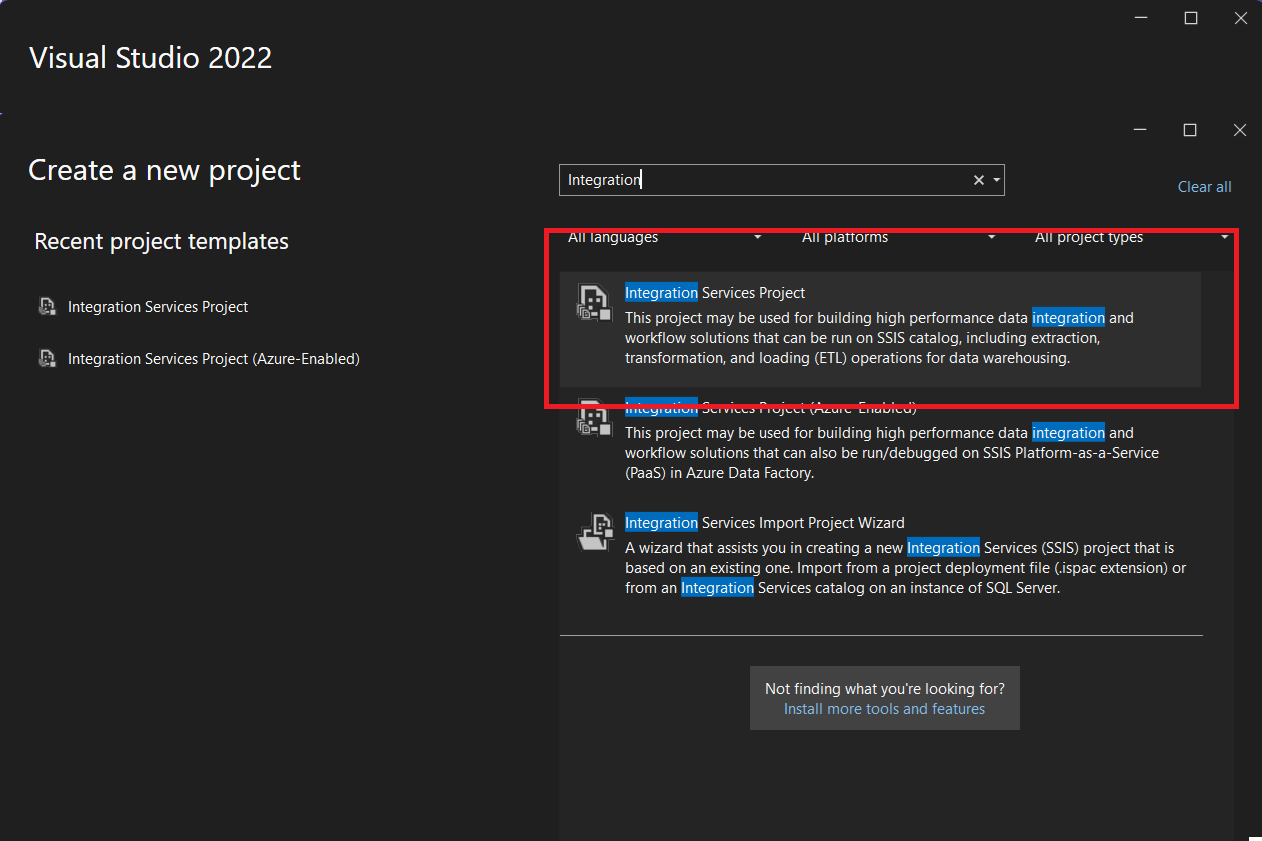


**Bước 2:** Khởi tạo cơ sở dữ liệu có tên CyberSecurity.



# Tạo mới project SSIS

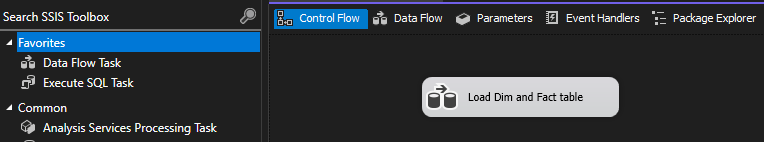
**Bước 1:** Mở Visual Studio 2022 và chọn “Create a new project”, chọn Integration Services Project và chọn Next.



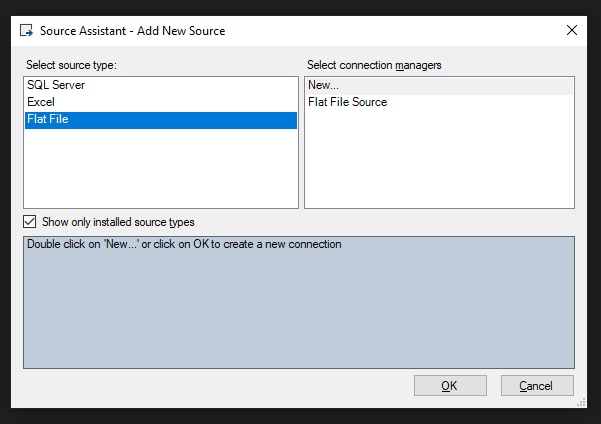
**Bước 2:** Đặt tên và thiết lập đường dẫn cho Project.

# Tạo bảng Dim và bảng Fact

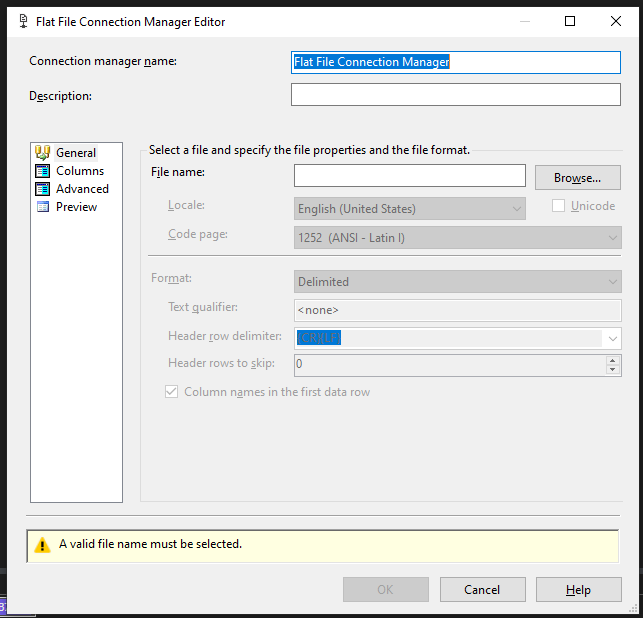
* Tạo Data Flow Task có tên là “Load Dim and Fact table” để chuẩn bị cho việc tạo các bảng Dim và Fact.



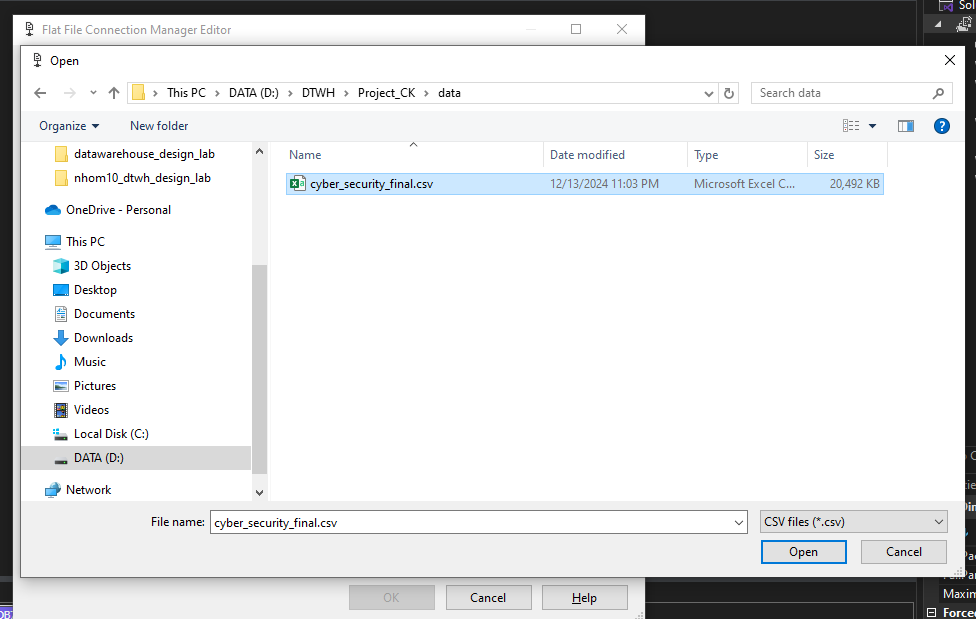
* Sang mục Data Flow, tạo mới một “Flat File Source” bằng cách add new Source Assistant từ toolbox.



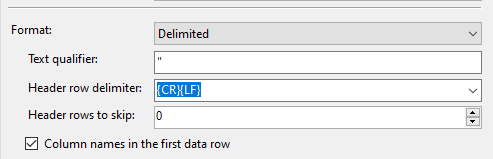
* Sau đó chọn vào “New” để tạo mới một “Flat file connection”, đọc dữ liệu từ file .csv chứa dataset.



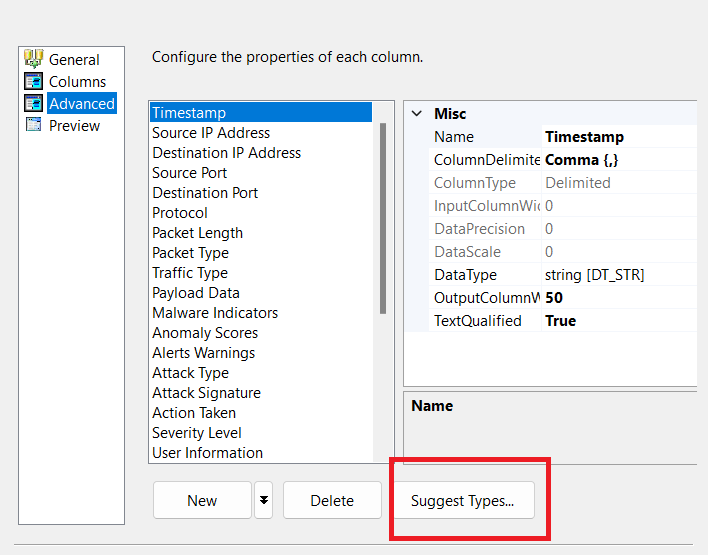
* Chọn “Browse”, chọn file .csv chứa dataset.

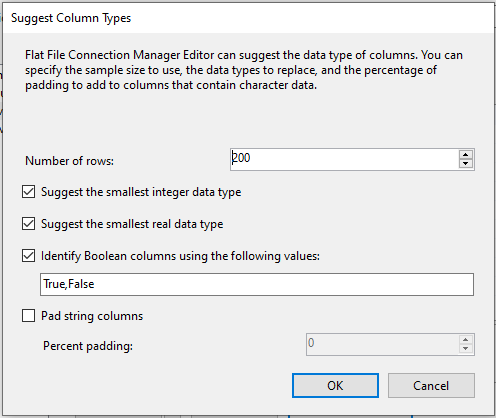


* Sau khi đã chọn file .csv, điền kí tự “ vào mục Text qualifier để tránh bị lỗi khi đọc dữ liệu

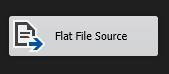


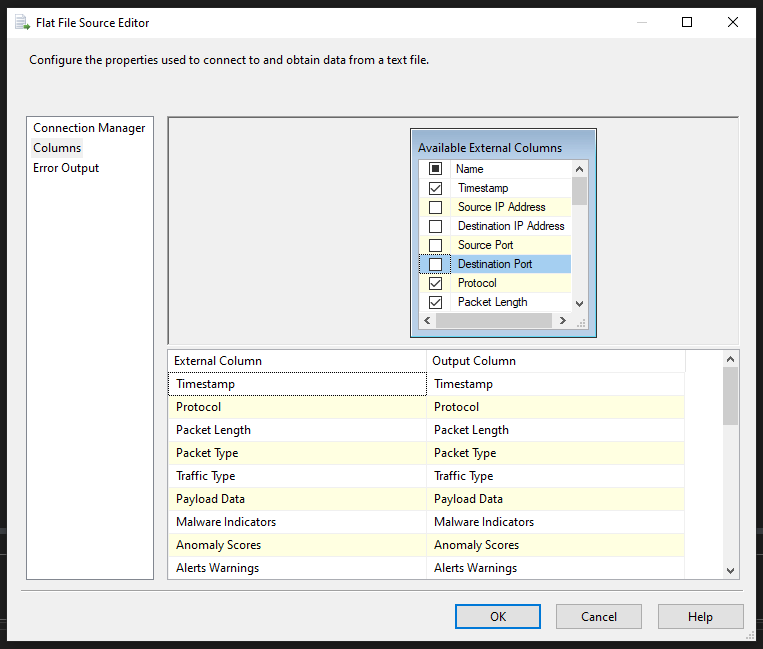
* Vào mục “Advanced” chọn “Suggest Types” để công cụ tự động chọn kiểu dữ liệu phù hợp cho từng cột và nhấn “OK”.



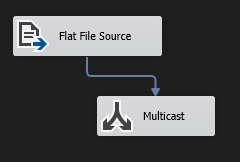


* Nhấn OK để tạo file Flat File Source. Sau đó edit vào mục “Column” Và chọn những cột trong dataset sẽ dùng.





* Chọn “Multicast” để tiến hành đổ dữ liệu vào kho.

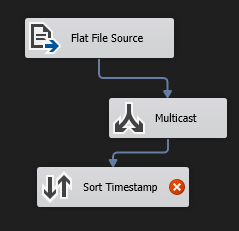


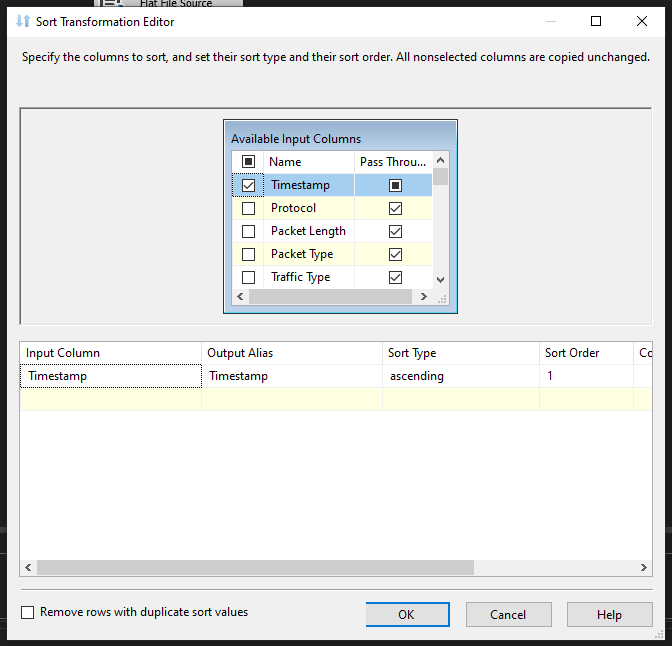
## Bảng Dim\_Timestamp

A screenshot of a computer

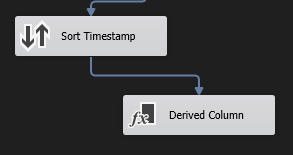
Description automatically generated

**Bước 1:** Tạo mới một “Sort” có tên là Sort Timestamp để lấy ra các cột dữ liệu cần thiết cho DIM\_Timestamp. Nhấn chuột phải và Edit để chọn Timestamp làm cột dữ liệu cho Sort Timestamp.



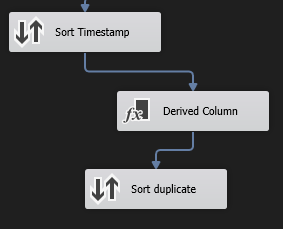


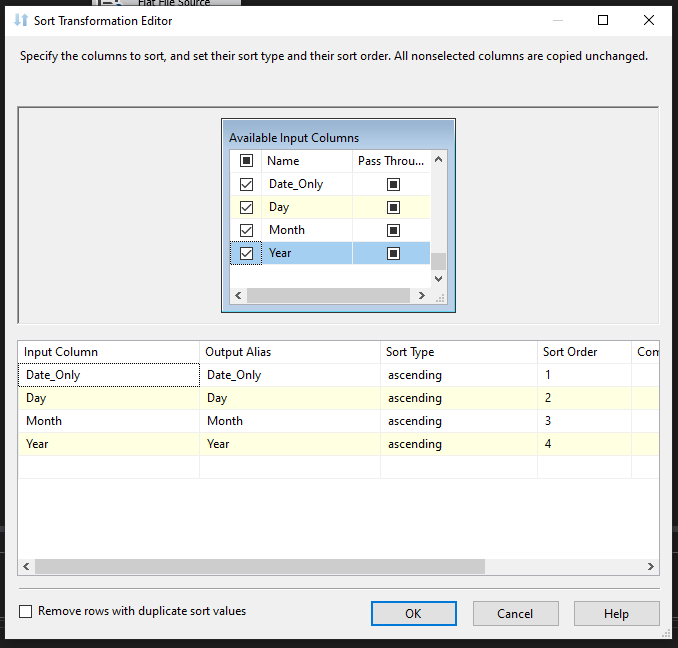
**Bước 2:** Thêm thành phần “Derived Column” và chọn Edit để chia cột dữ liệu. Ta chia Timestamp thành các cột dữ liệu Day, Month, Year và đồng thời ép kiểu dữ liệu của Timestamp thành Date



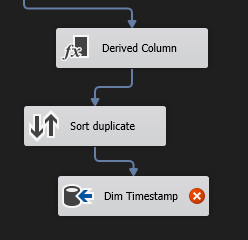


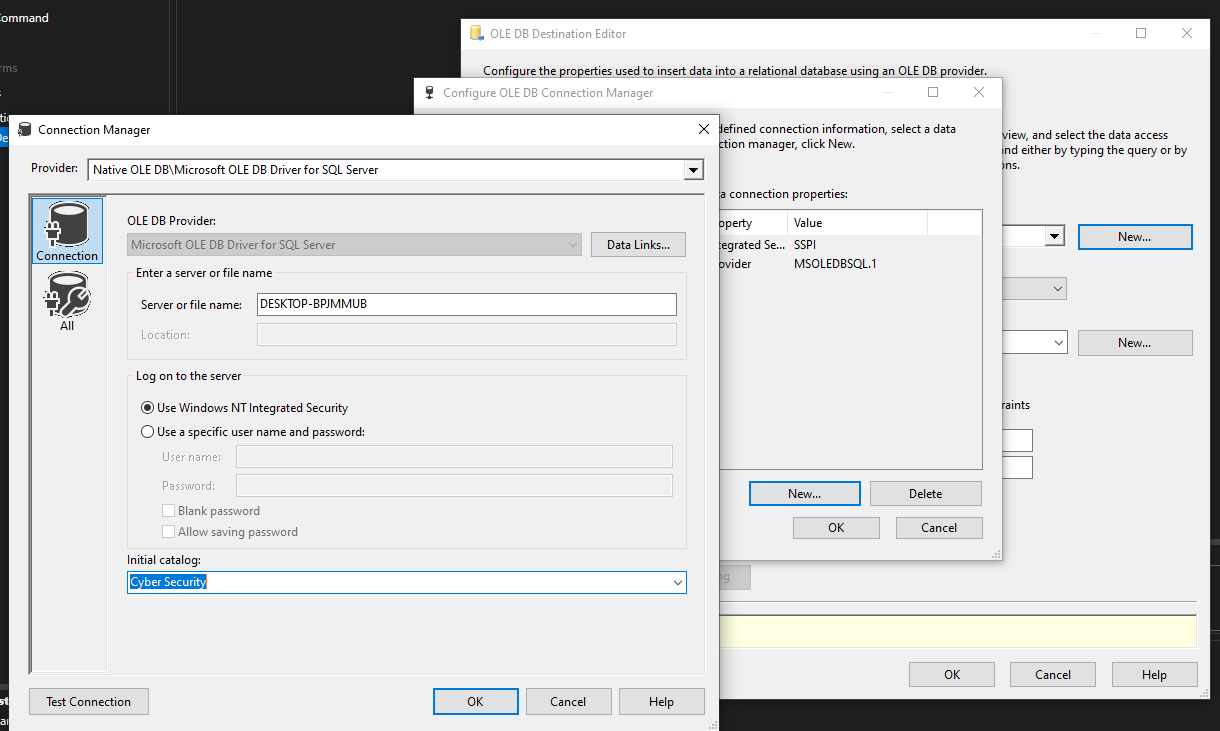
**Bước 3:** Thêm thành phần Sort để xoá các dòng dữ liệu bị trùng.



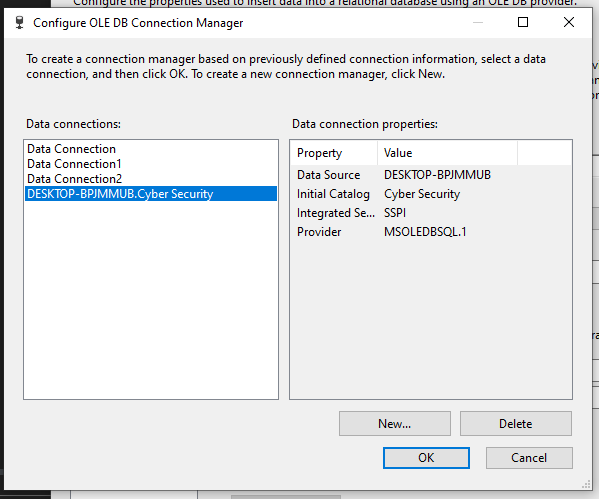


**Bước 4:** Tạo Dim\_Timestamp từ một “OLE DB Destination”. Double click vào “OLE DB Destination” để tạo một connection mới đến MS SQL Server.

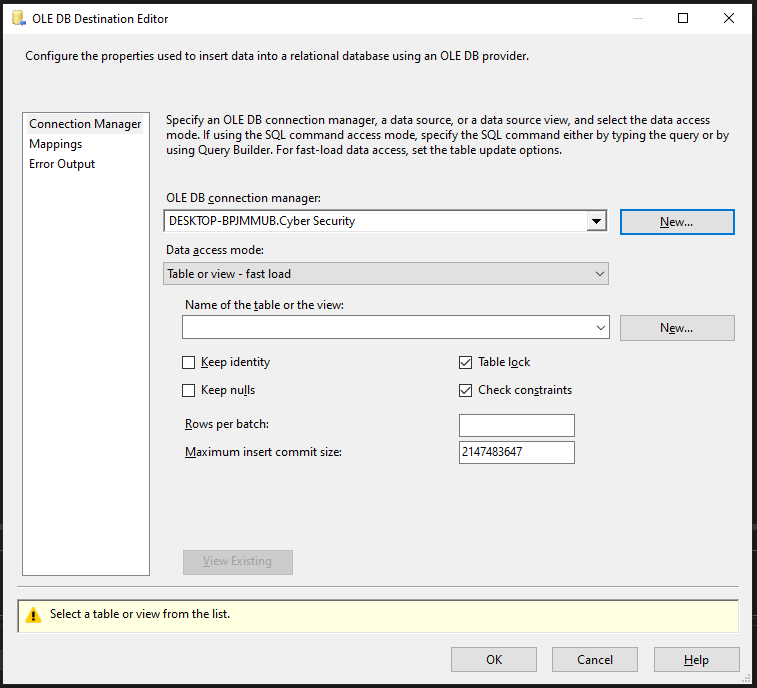


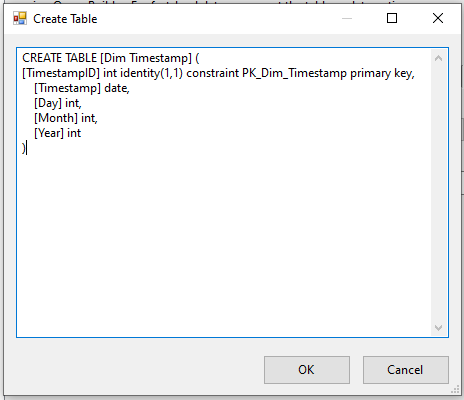


**Bước 5:** Chọn tên server name trùng với server name MS SQL Server để ta có thể kết nối đến datawarehouse CyberSecurity vừa tạo. Kết nối đến server bằng tài khoản window mặc định ( Wiondows Authentication) nhấn Test Connection để kiểm tra kết nối.

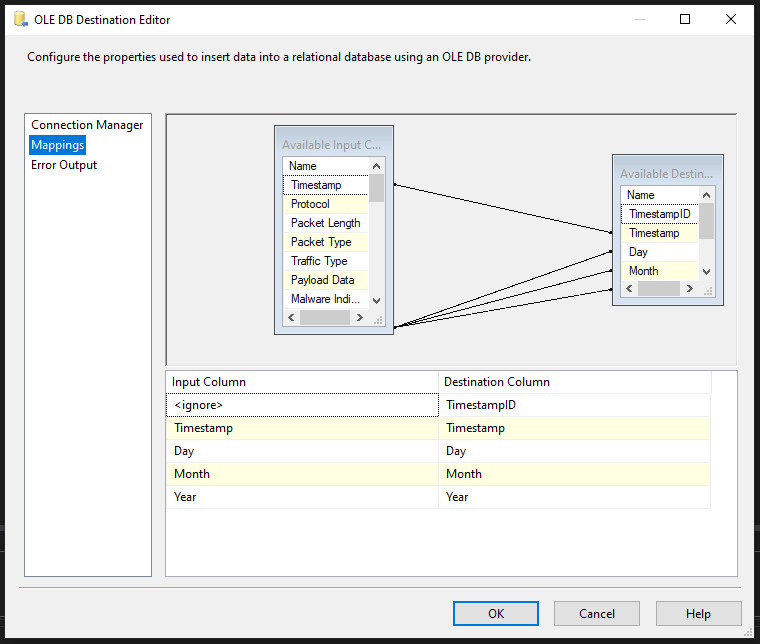


**Bước 6:** Chọn New để tạo bảng Dim\_Timestamp





**Bước 7:** Vào mục “Mappings” để kiểm tra ánh xạ cột dữ liệu. Nhấn “OK” để hoàn tất

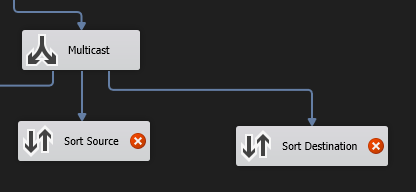


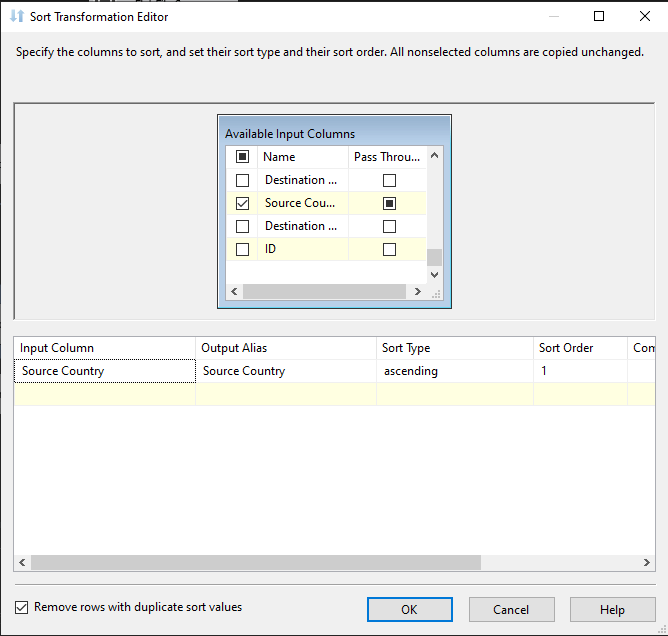
## Bảng Dim\_Address

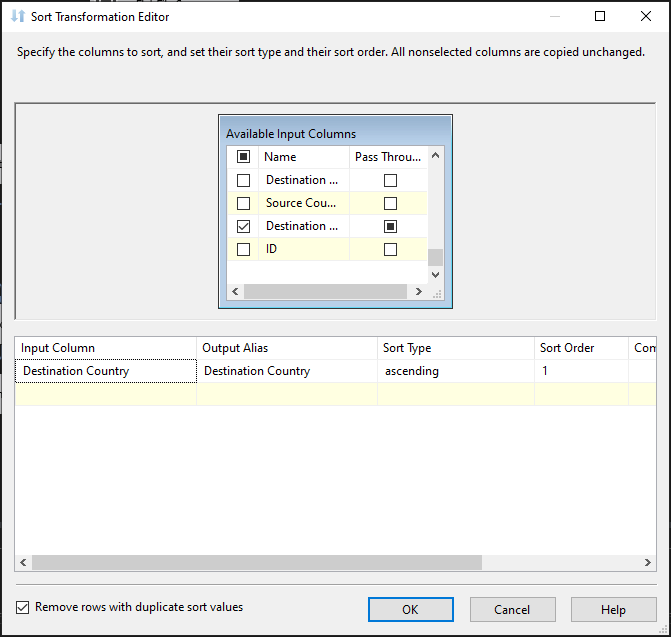
**A screenshot of a computer

Description automatically generated**

**Bước 1:** Tạo Sort Source và Sort Destination để lọc dữ liệu. Chọn xoá các dòng dữ liệu bị trùng.

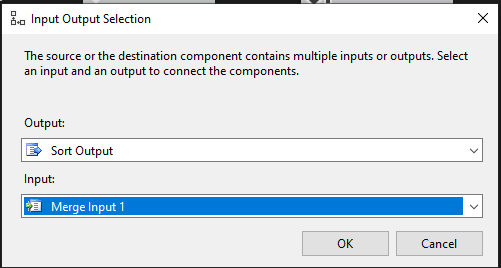




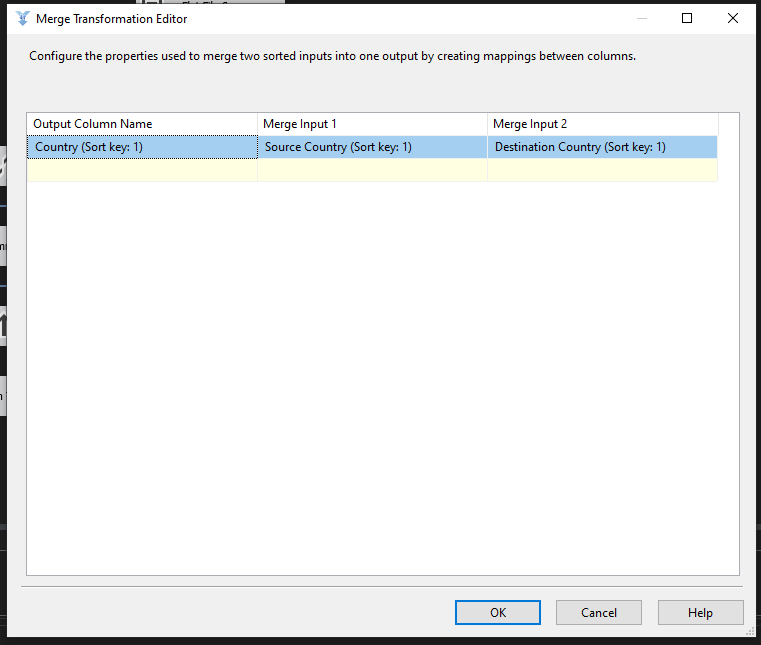


**Bước 2:** Chọn thành phần Merge để kết hợp 2 cột dữ liệu lại.

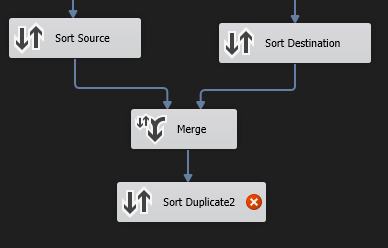
A diagram of a diagram

Description automatically generated

Chọn Merge Input 1 là Source Country và Merge Input 2 là Destination Country. Nhập tên Output là Country

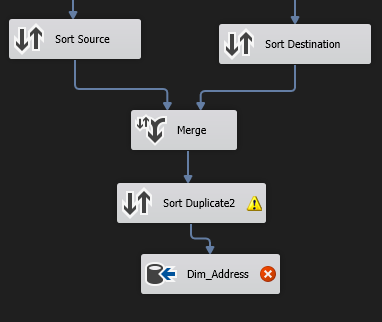


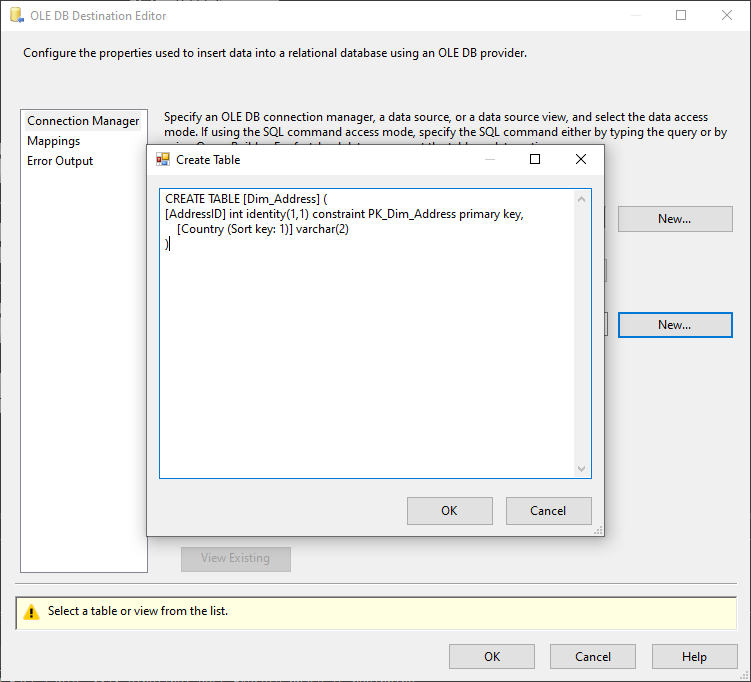
**Bước 3:** Tạo Sort để xoá các dòng dữ liệu trùng sau khi đã merge.

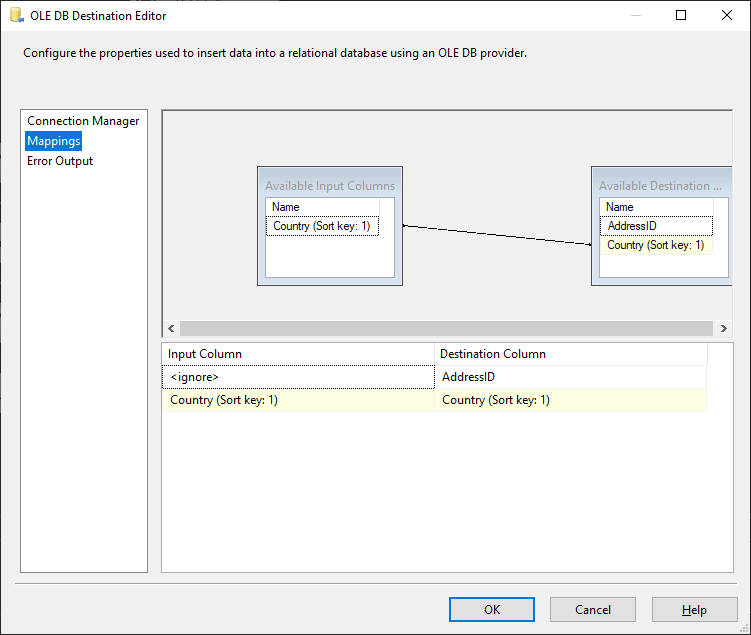




**Bước 4:** Tạo “OLE DB Destination” đặt tên là Dim\_Address để lưu kết quả. Nhấn New và tạo mới table. Vào Mappings kiểm tra ánh xạ dữ liệu và nhấn OK.





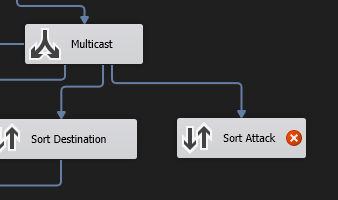


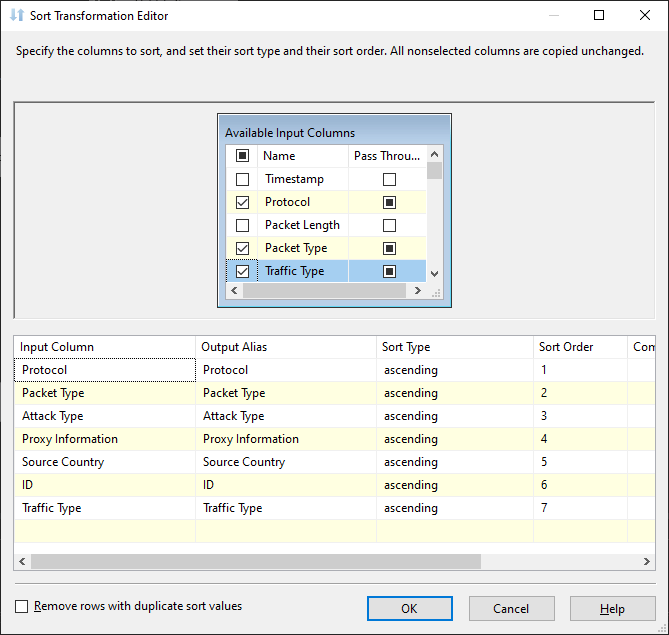
## Bảng Dim\_Attack

**A screenshot of a computer

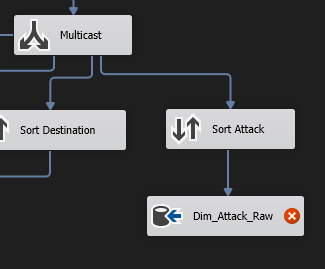
Description automatically generated**

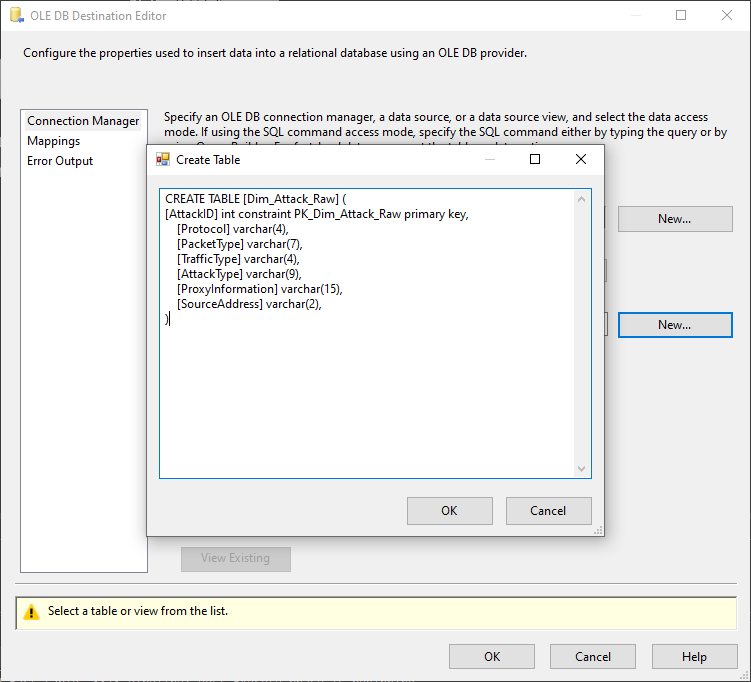
**Bước 1:** Tạo Sort để lọc các thuộc tính cần thiết cho Dim\_Attack.

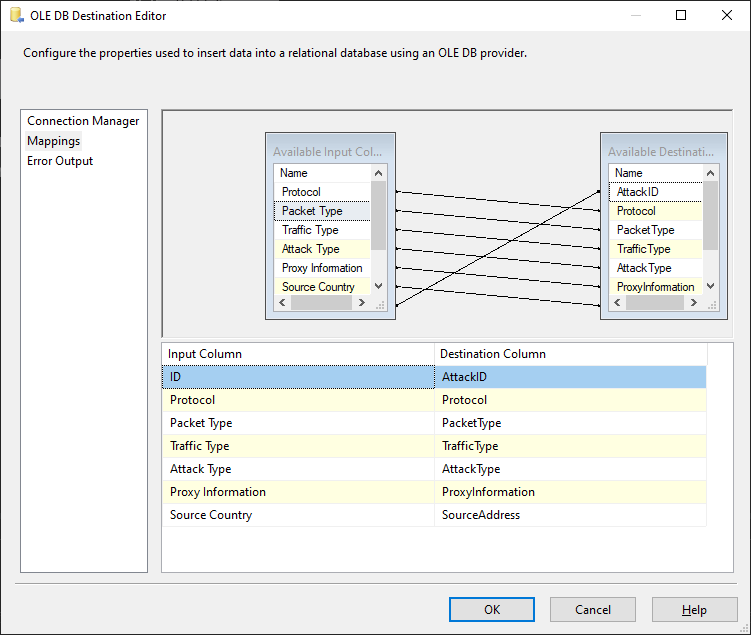




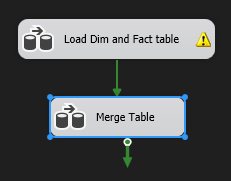
**Bước 2:** Tạo “OLE DB Destination” đặt tên là Dim\_Attack\_Raw để lưu dữ liệu. Nhấn New và tạo mới table. Vào Mappings kiểm tra ánh xạ dữ liệu và nhấn OK.



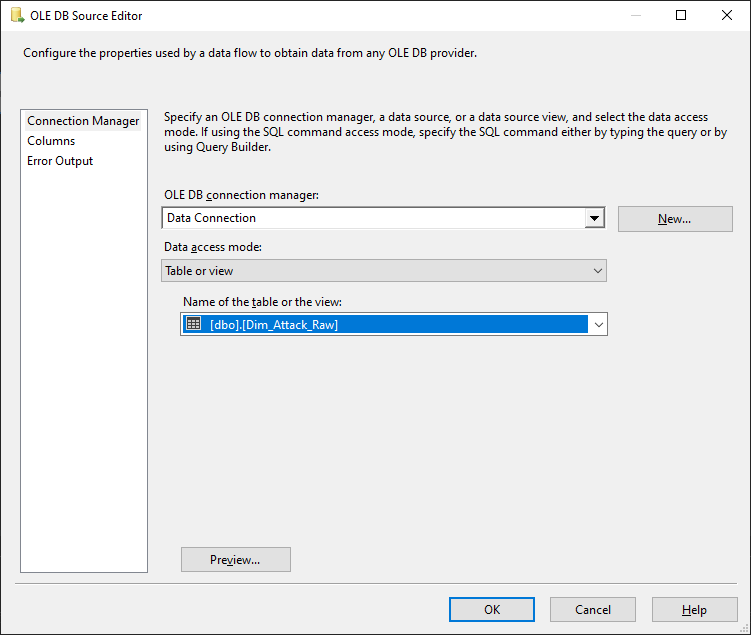


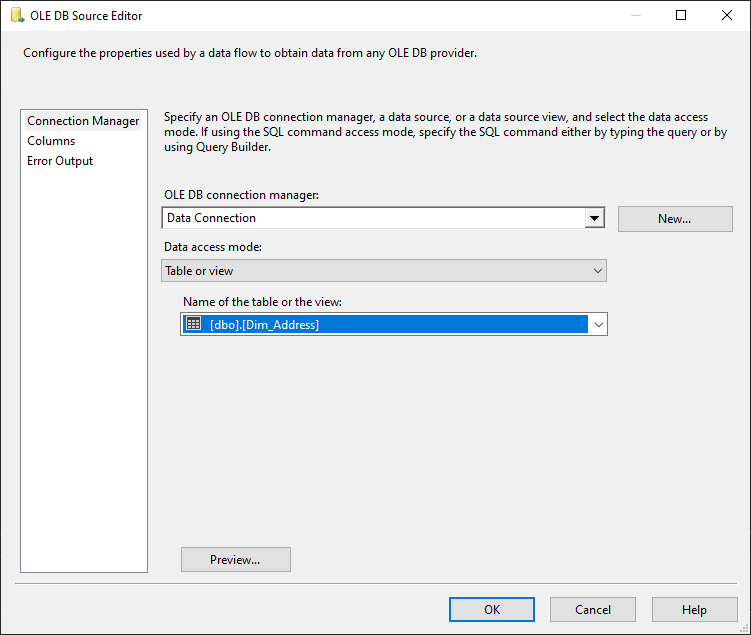


**Bước 3:** Thêm một thành phần “Data Flow Task” mới để tiến hành đổi SourceAddress lấy khoá ngoại ở bảng Dim\_Address.

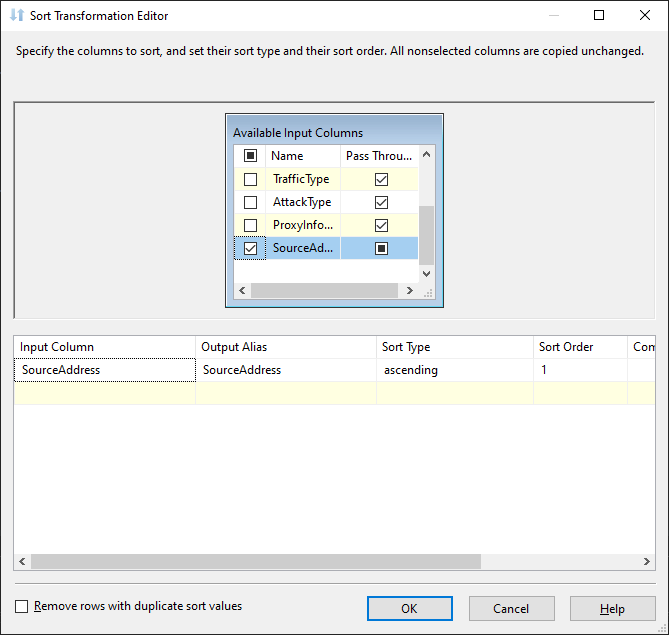


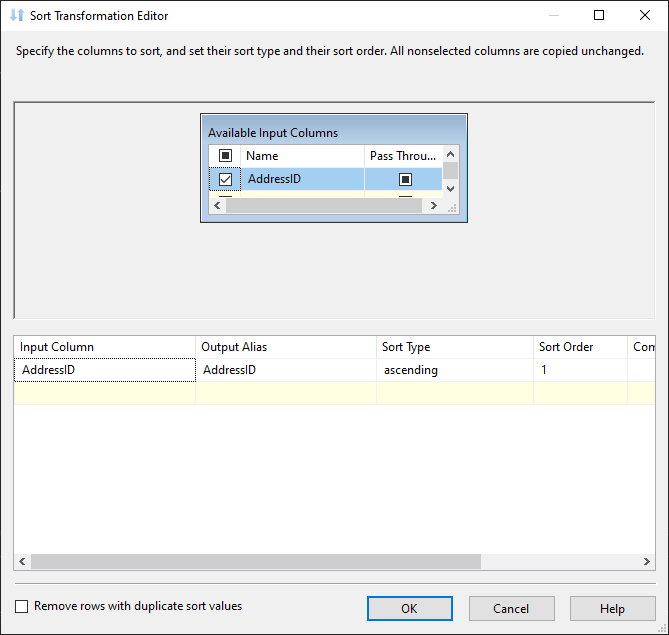
**Bước 4:** Trong Data Flow của “Merge table”, thêm 2 thành phần “OLE DB Source” của Dim\_Attack\_Raw và Dim\_Address.



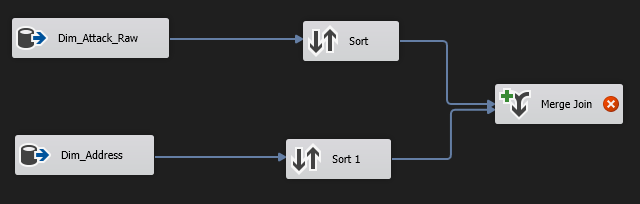


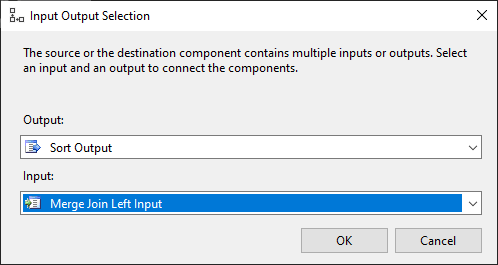
**Bước 5:** Tiến hành Sort từng table theo cột muốn đổi là SourceAddress và Country.



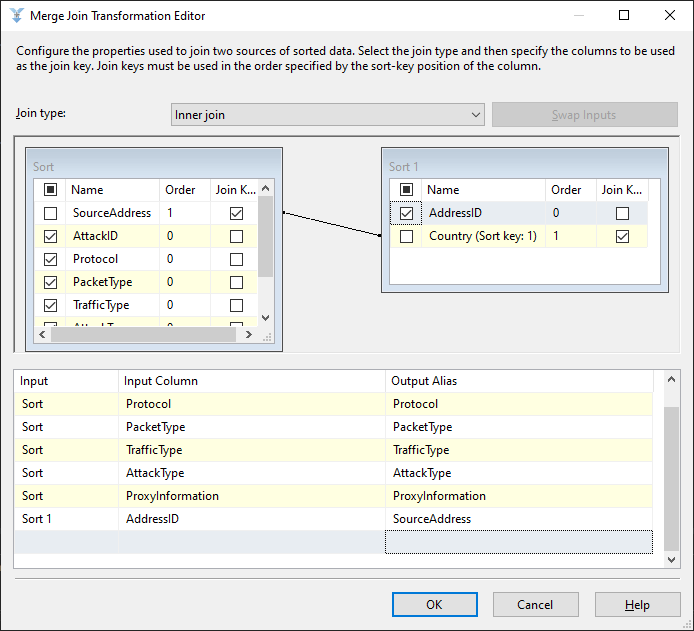


**Bước 6:** Thêm Look up để kết nối 2 nguồn dữ liệu bởi vì 2 bảng này không có chung cột thuộc tính nào.

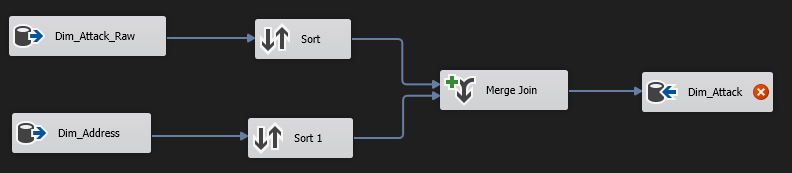


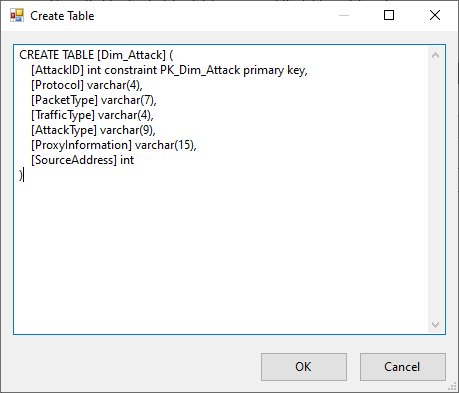


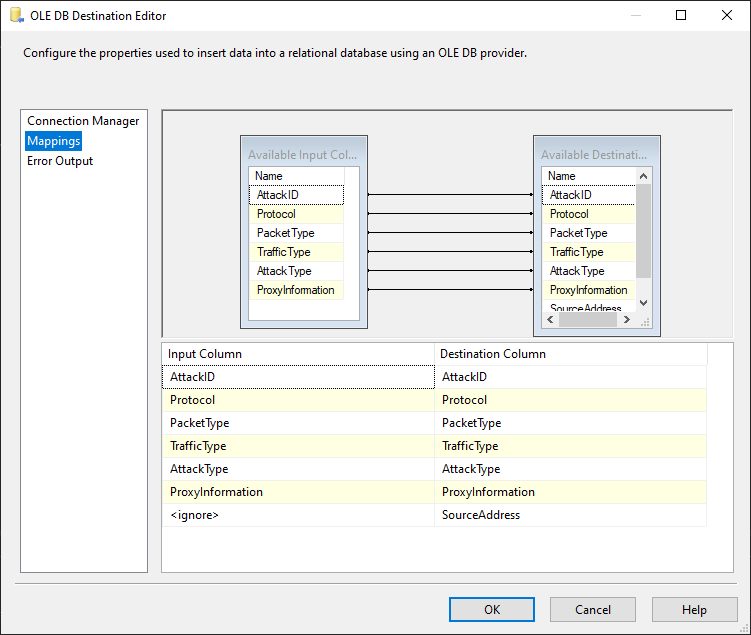
**Bước 7:** Trong Merge Join, ta chọn tất cả các cột của Sort trừ SourceAddress. Tiếp theo ta chọn AddressID của Sort1. Đặt tên cho Output của Sort1 là SourceAddress.



**Bước 8:** Thêm thành phần “OLE DB Destination” mới với tên là Dim\_Attack để lưu lại kết quả. Tạo table Dim\_Attack, kiểm tra Mappings và nhấn OK.





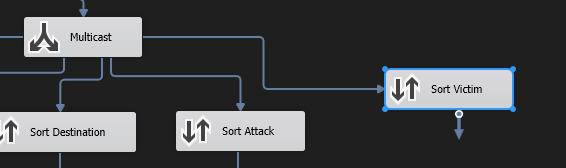


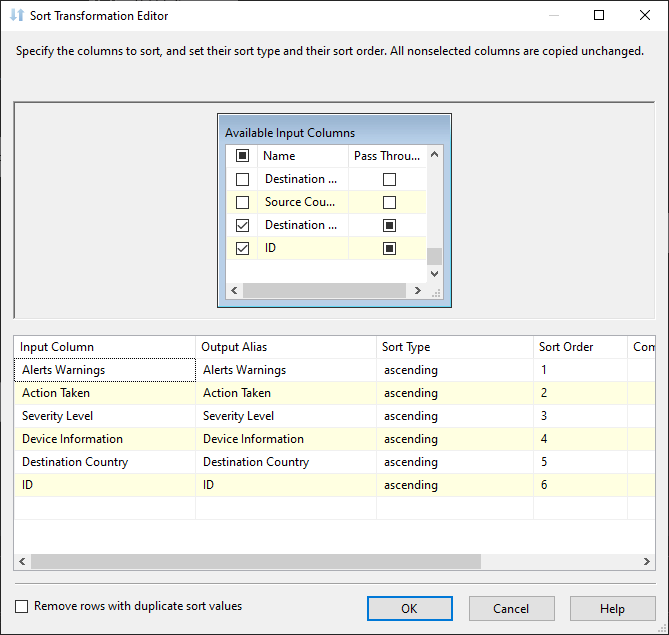
## Bảng Dim\_Victim

A screenshot of a computer

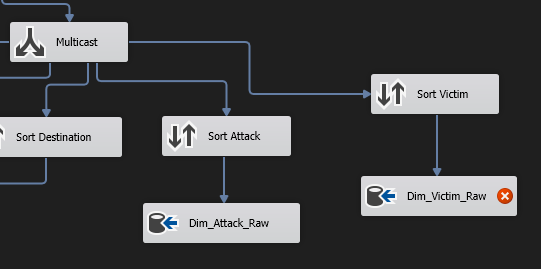
Description automatically generated

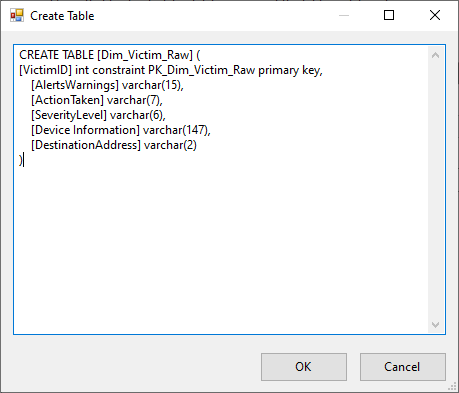
**Bước 1:** Tạo Sort để lọc các thuộc tính cần thiết cho Dim\_Victim.

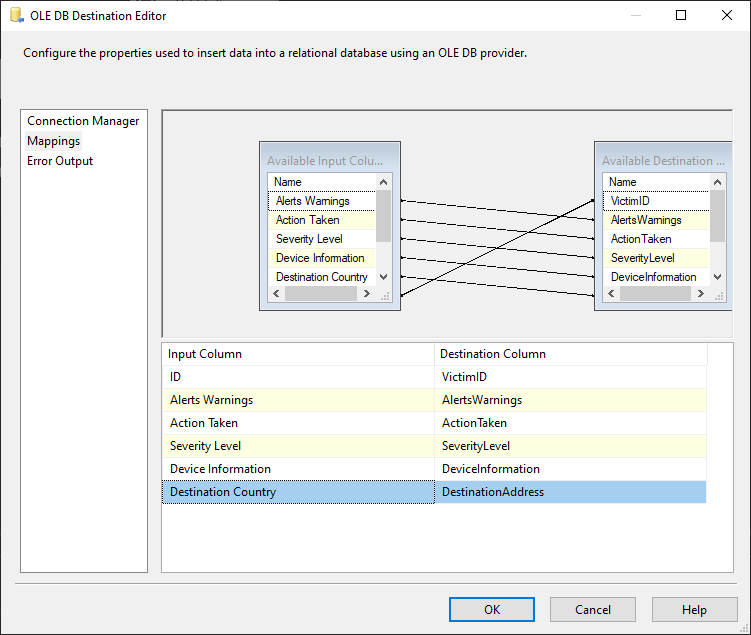




**Bước 2:** Tạo “OLE DB Destination” đặt tên là Dim\_Victim\_Raw để lưu dữ liệu. Nhấn New và tạo mới table. Vào Mappings kiểm tra ánh xạ dữ liệu và nhấn OK.







**Bước 3:** Trong Data Flow của “Merge table”, thêm 2 thành phần “OLE DB Source” của Dim\_Victim\_Raw và Dim\_Address.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

**Bước 4:** Tiến hành Sort từng table theo cột muốn đổi là DestinationAddress và Country.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

**Bước 5:** Thêm Merge Join, chọn Merge Join Left Input để giữ lại toàn bộ các dòng trong bảng Dim\_Victim\_Raw khi thực hiện phép kết trái với cột ID của bảng Dim\_Address.

A diagram of a sort

Description automatically generated with medium confidence

A screenshot of a computer

Description automatically generated

**Bước 6:** Trong Merge Join, ta chọn tất cả các cột của Sort2 trừ DestinationAddress. Tiếp theo ta chọn AddressID của Sort3. Đặt tên cho Output của Sort3 là DestinationAddress.

A screenshot of a computer

Description automatically generated

**Bước 7:** Thêm thành phần “OLE DB Destination” mới với tên là Dim\_Victim để lưu lại kết quả. Tạo table Dim\_Victim, kiểm tra Mappings và nhấn OK.

A diagram of a diagram

Description automatically generated with medium confidence

A screenshot of a computer

Description automatically generated

A screenshot of a computer

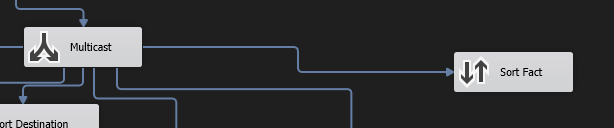
Description automatically generated

## Bảng Fact\_CyberSecurity

A screenshot of a computer

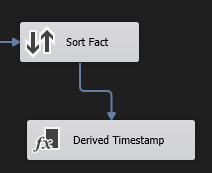
Description automatically generated

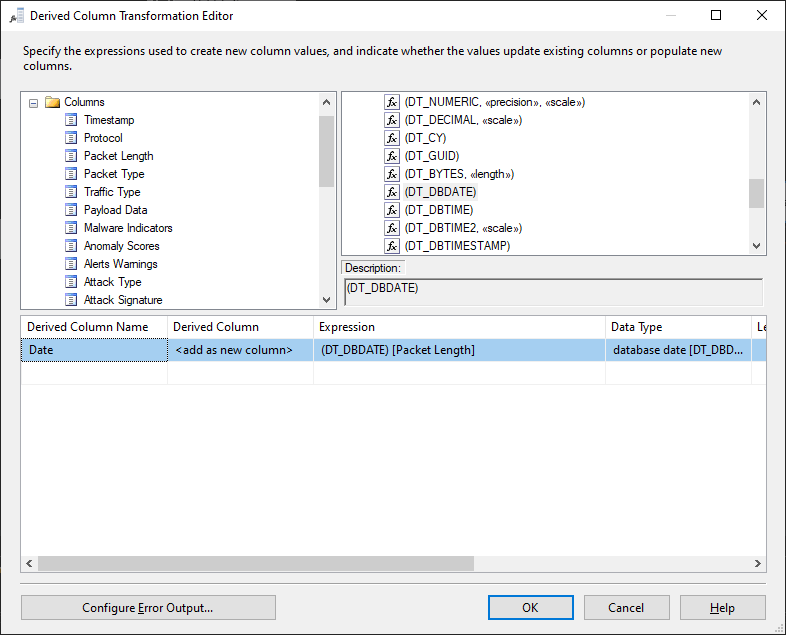
**Bước 1:** Tạo Sort để lọc các thuộc tính cần thiết cho Fact\_CyberSecurity.



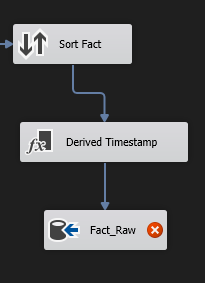


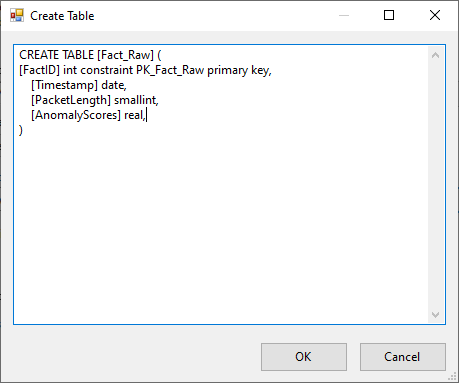
**Bước 2:** Thêm thành phần Deverived Column để ép kiểu dữ liệu Timestamp thành Date hỗ trợ cho việc đổi dữ liệu lấy khoá ngoại với Dim\_Timestamp.

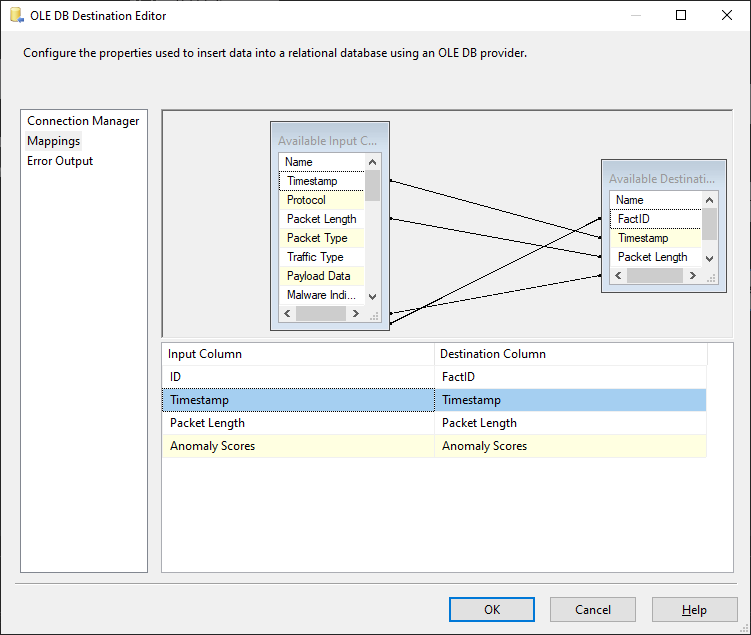




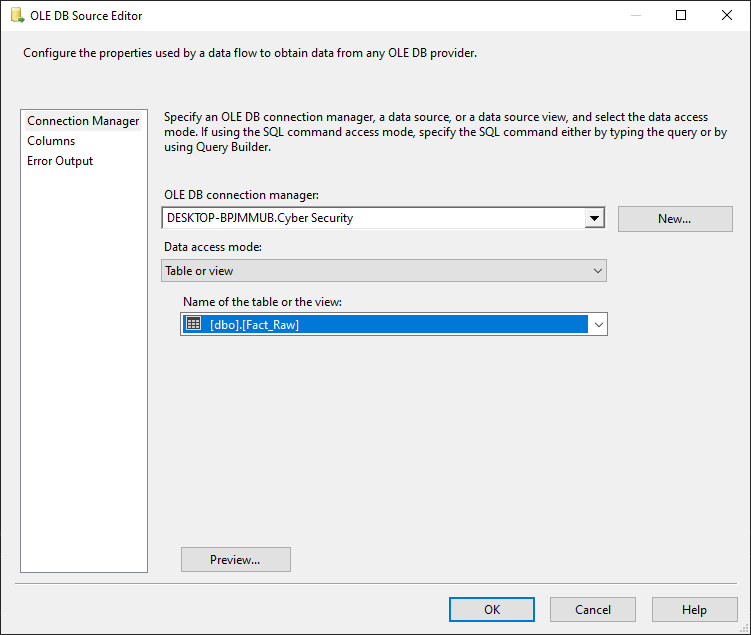
**Bước 3:** Tạo “OLE DB Destination” đặt tên là Fact\_Raw để lưu dữ liệu. Nhấn New và tạo mới table. Vào Mappings kiểm tra ánh xạ dữ liệu và nhấn OK.

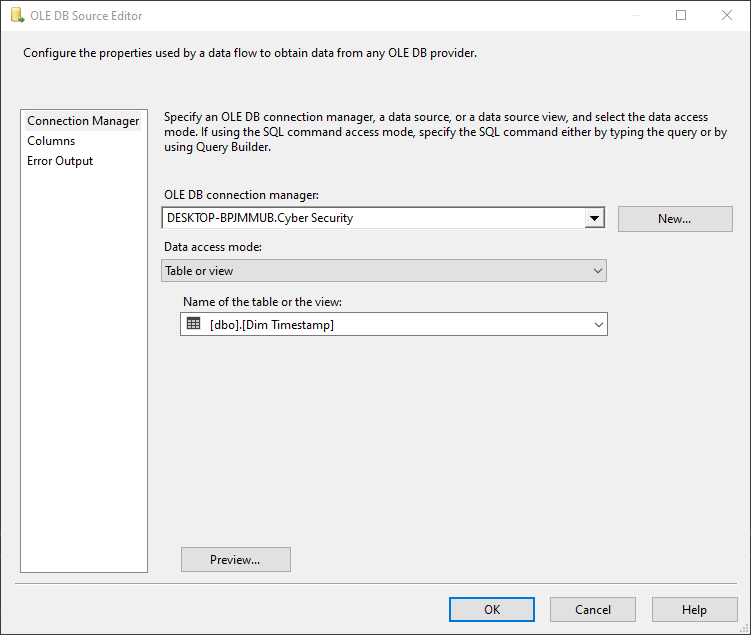




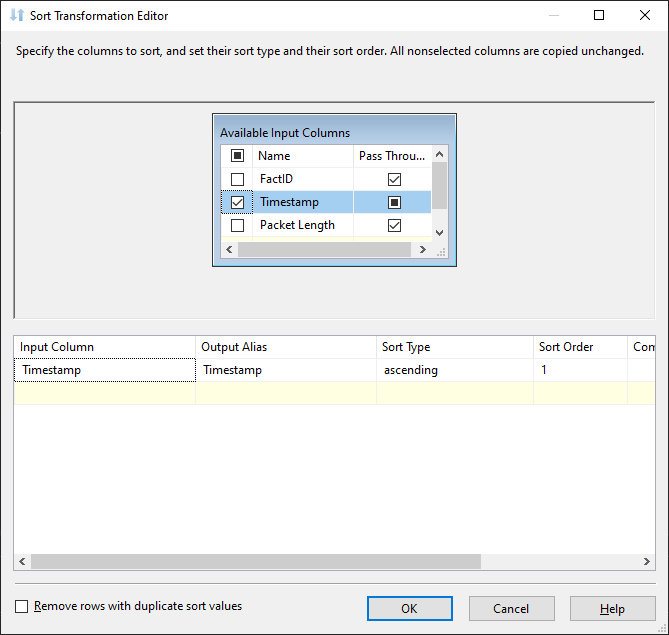


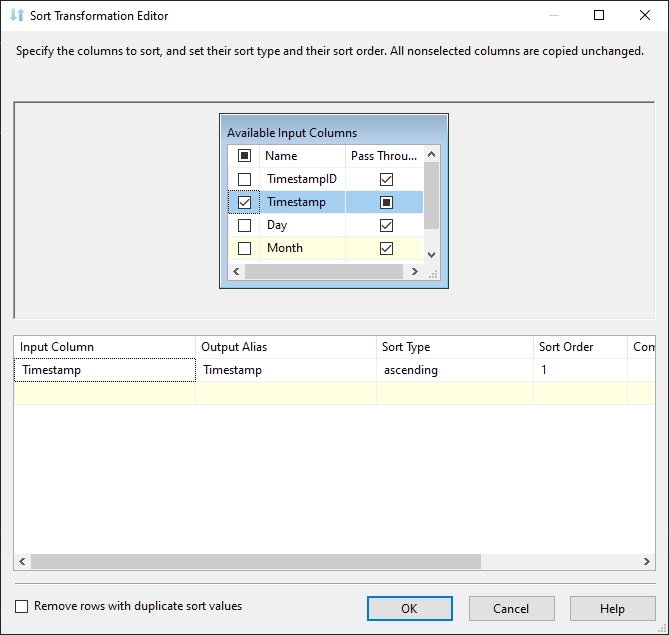
**Bước 4:** Trong Data Flow của “Merge table”, thêm 2 thành phần “OLE DB Source” của Fact\_Raw và Dim\_Timestamp.



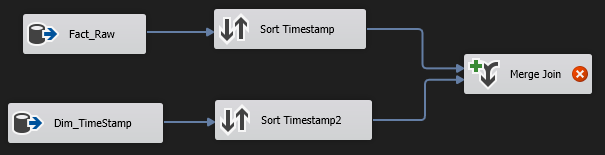


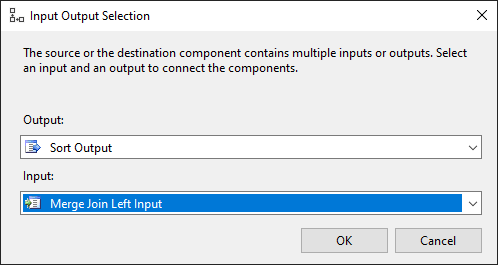
**Bước 5:** Tiến hành Sort từng table theo cột muốn đổi là Timestamp.



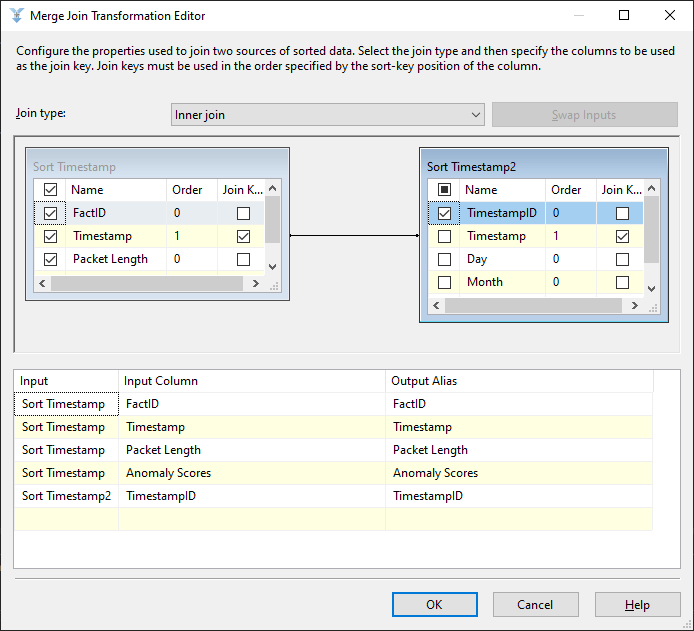


**Bước 6:** Thêm Merge Join, chọn Merge Join Left Input để giữ lại toàn bộ các dòng trong bảng Fact\_Raw khi thực hiện phép kết trái với cột ID của bảng Dim\_Timestamp

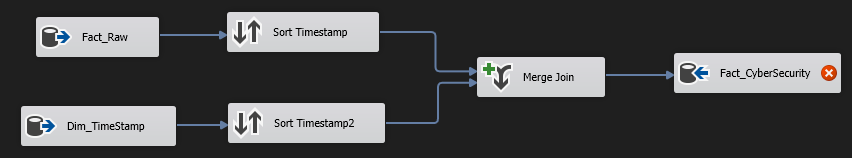


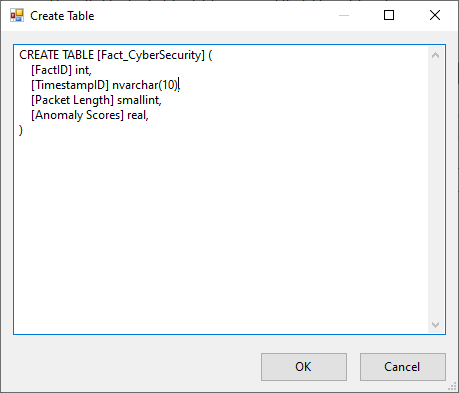


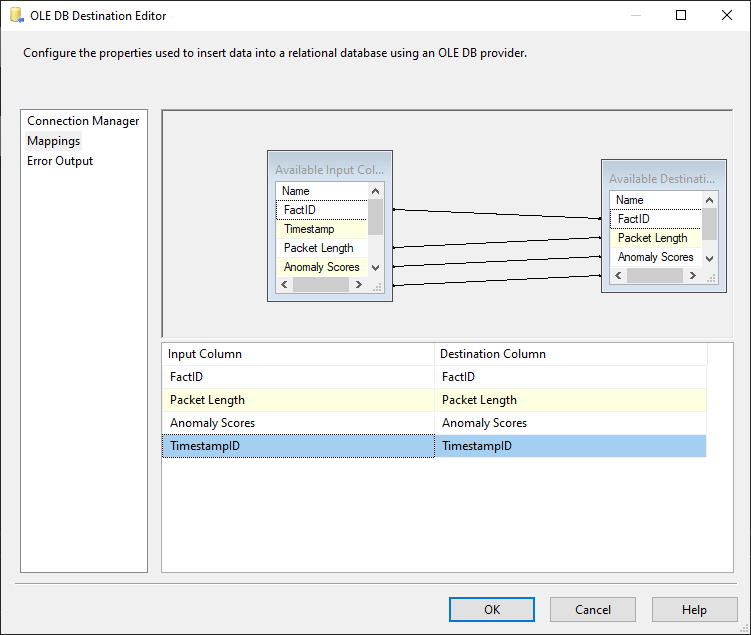
**Bước 7:** Trong Merge Join, ta chọn tất cả các cột của Sort4 trừ Timestamp. Tiếp theo ta chọn Timestamp ID của Sort5.



**Bước 8:** Thêm thành phần “OLE DB Destination” mới với tên là Fact\_CyberSecurity để lưu lại kết quả. Tạo table Fact\_CyberSecurity, kiểm tra Mappings và nhấn OK.







## Tạo khoá ngoại

**Bước 1:** Thêm một “Execute SQL Task” dưới Merge table để tạo khoá ngoại. Đặt tên là Create FK.

A screenshot of a computer

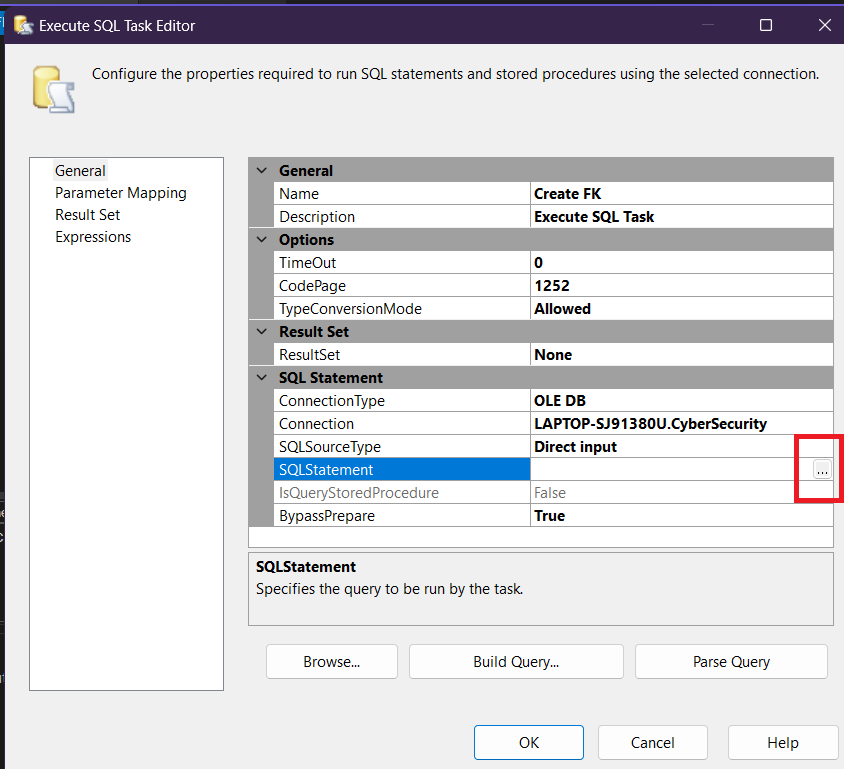
Description automatically generated

**Bước 2:** Nhấn chuột phải vào Create FK và chọn Edit. Chọn connection đã được thiết lập đến database.

A screenshot of a computer

Description automatically generated

**Bước 3:** Nhấn vào nút ba chấm ở SQLStatement và viết câu lệnh SQL để tạo khoá ngoại cho các table. Nhấn OK



A screenshot of a computer

Description automatically generated

## Chạy SSIS

**Bước 1:** Thêm vào một “Execute SQL Task” ở trước Load Dim and Fact để làm sạch dữ liệu trong table tránh trường hợp dữ liệu bị chồng chéo mỗi lần chạy lại project.

A screenshot of a computer

Description automatically generated

**Bước 2:** Chọn connection đã được kết nối và viết các câu lệnh SQL để làm sạch table.

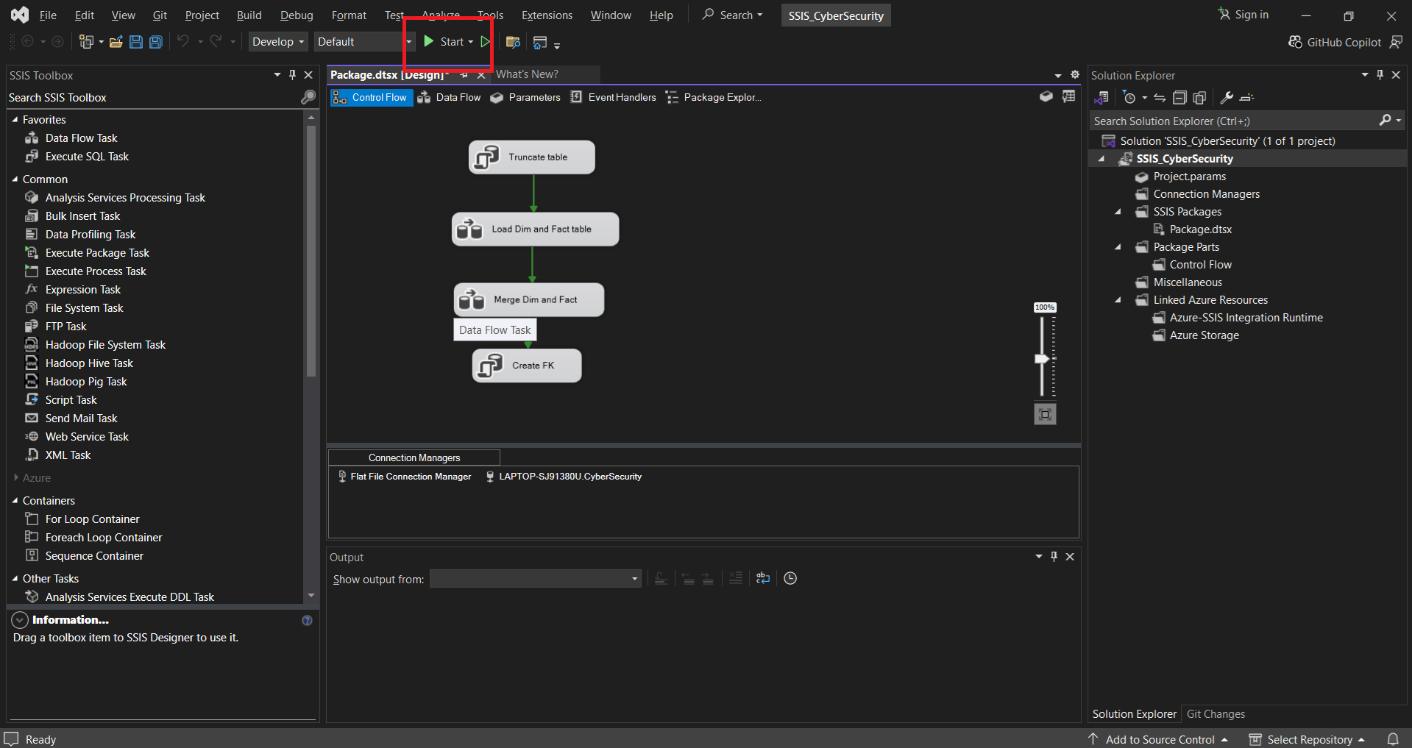
A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

**Bước 3:** Nhấn nút Start trên thanh công cụ để chạy dự án.



Kết quả chạy project:

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

**Bước 4:** Sau khi chạy project lần đầu tiên, các khoá ngoại sẽ được tạo ra. Để tránh việc vi phạm các ràng buộc khi xoá dữ liệu của table, ta sẽ thêm các câu lệnh SQL để xoá khoá ngoại trước khi xoá dữ liệu trong table trong phần Truncate table.

A screenshot of a computer

Description automatically generated

# Kiểm tra dữ liệu các bảng

Dữ liệu bảng Dim\_Timestamp:

A screenshot of a data

Description automatically generated

Dữ liệu bảng Dim\_Address:

A screenshot of a computer

Description automatically generated

Dữ liệu bảng Dim\_Attack:

A screenshot of a computer

Description automatically generated

Dữ liệu bảng Dim\_Victim:

A white background with black text

Description automatically generated

Dữ liệu bảng Fact\_CyberSecurity:

A screenshot of a table

Description automatically generated

# Lược đồ sau khi hoàn thành

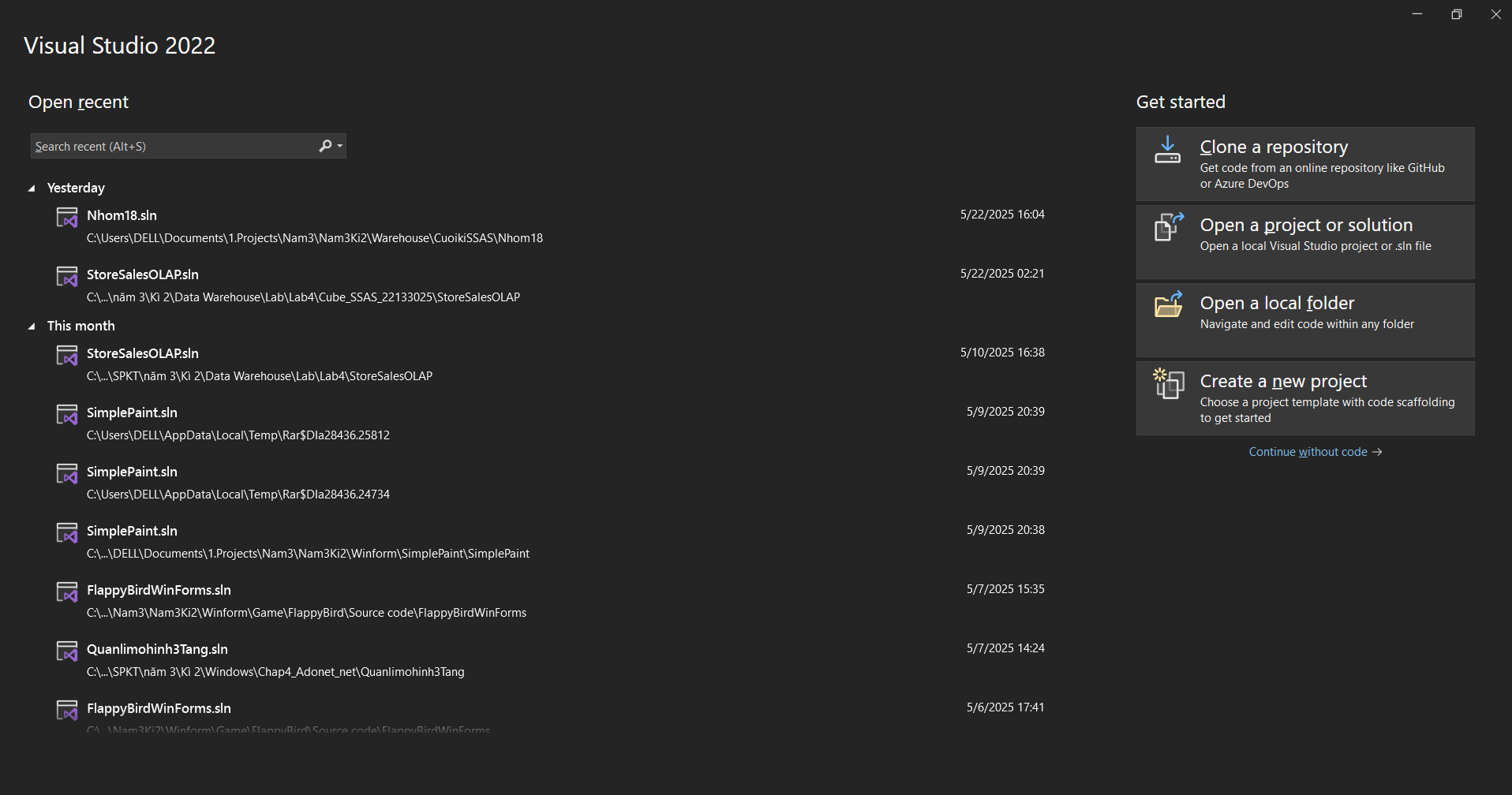
A screenshot of a computer

Description automatically generated

CHƯƠNG 3. THỰC HIỆN PHÂN TÍCH TRÊN KHO DỮ LIỆU (SSAS)

# Tạo project SSAS mới

**Bước 1:** Mở Visual Studio và chọn “Create a new project”.

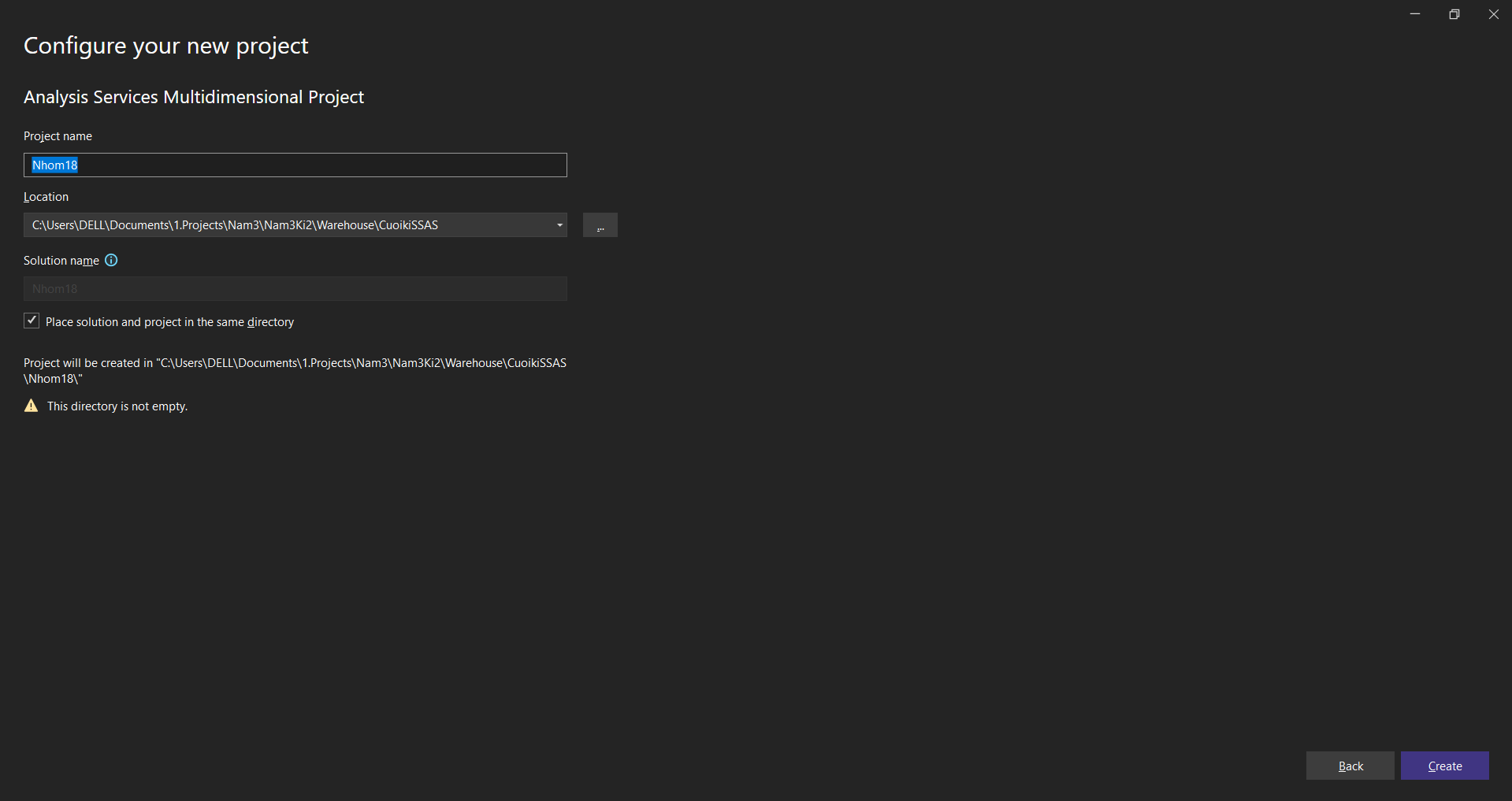
****

**Bước 2:** Chọn Analysis Services Multidimensional Project và chọn Next.

A screenshot of a computer program

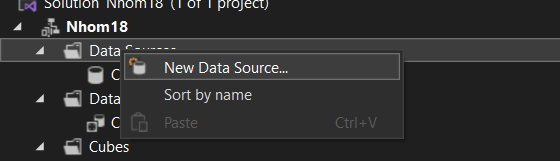
Description automatically generated

**Bước 3:** Đặt tên và thiết lập đường dẫn cho Project. Sau đó chọn Create.



# Xác định dữ liệu nguồn (Data Source)

**Bước 1:** Tại Solution Explorer, right-click vào thư mục Data Sources, chọn New Data Source.

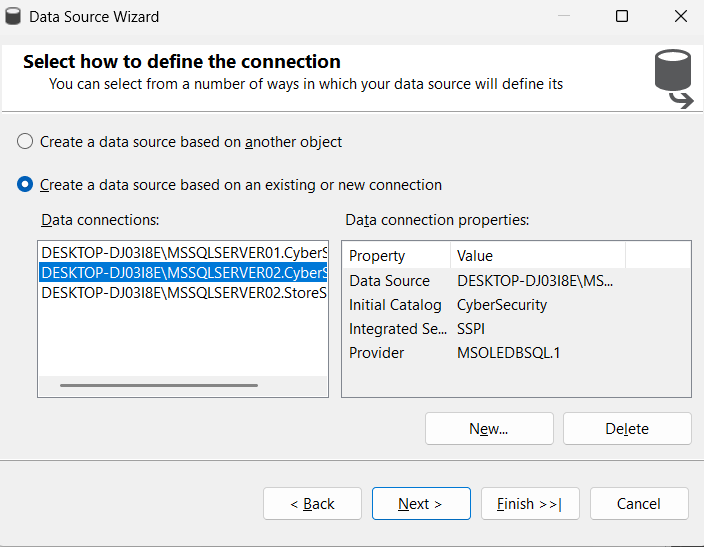


**Bước 2:** Hộp thoại Data Source Wizard xuất hiện, chọn Next để tiếp tục.

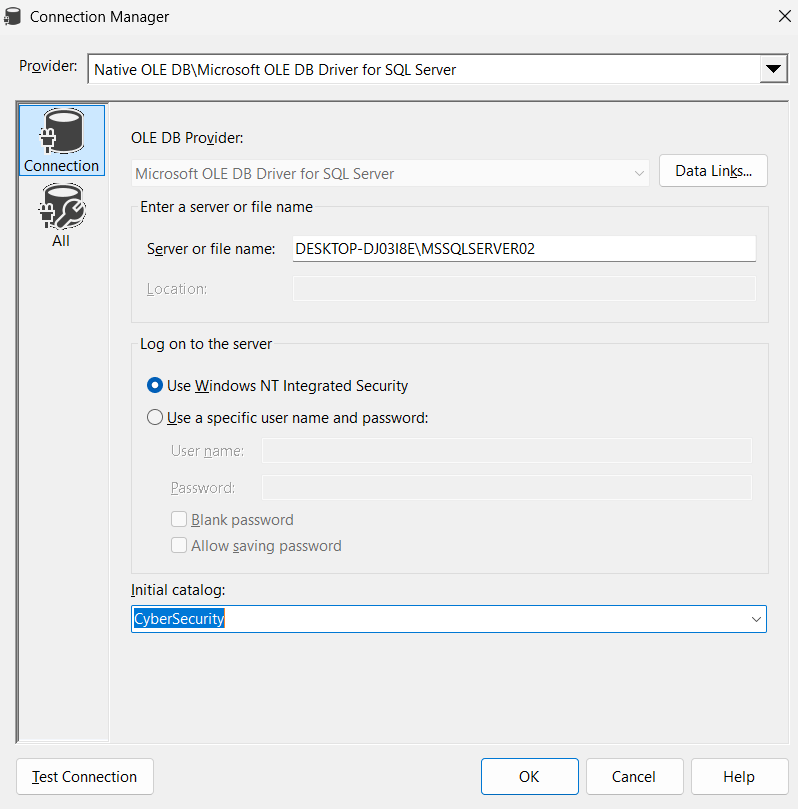
A screenshot of a data source wizard

Description automatically generated

**Bước 3:** Ở trang Select how to define the connection, chọn “Create a data source based on an existing or new connection”, chọn New để tạo mới một Data connection.



**Bước 4:** Chọn provider, server name và database để kết nối. Nhấn Next để tiếp tục.



**Bước 5:** Chọn “Use the service account”, sau đó chọn Next để tiếp tục.

A screenshot of a computer

Description automatically generated

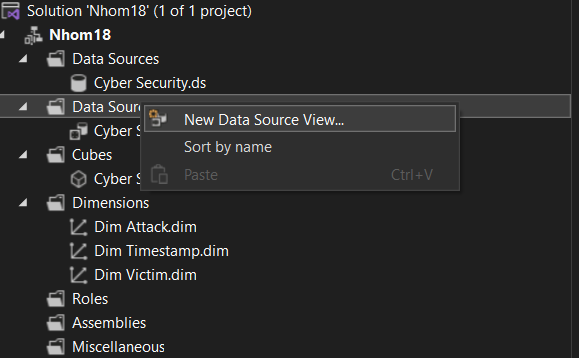
**Bước 6:** Cuối cùng ta chọn Finish để hoàn tất quy trình định nghĩa nguồn dữ liệu.

A screenshot of a computer

Description automatically generated

# Xác định khung nhìn dữ liệu nguồn (Data Source View)

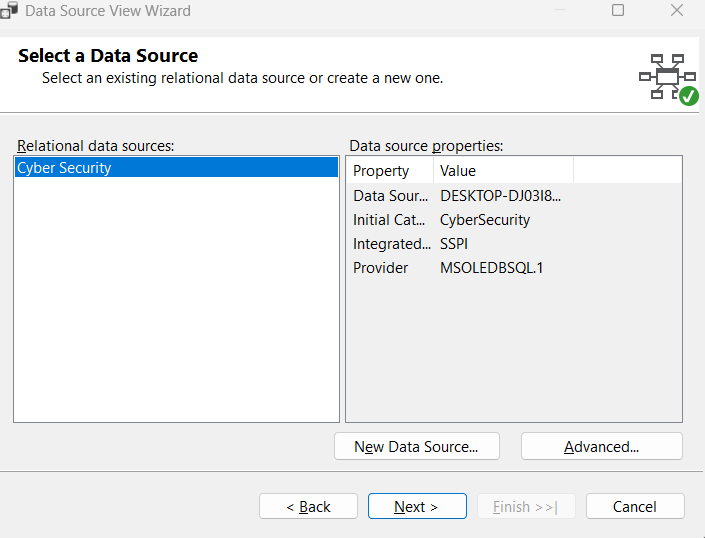
**Bước 1:** Tại Solution Explorer, right-click vào thư mục Data Source Views và chọn New Data Source View.

****

**Bước 2:** Hộp thoại Data Source View Wizard xuất hiện, chọn Next để tiếp tục. Chọn Data source vừa tạo và chọn Next để tiếp tục.

A screenshot of a computer

Description automatically generated



**Bước 3:** Chọn bảng Fact, sau đó chọn nút “>” để thêm bảng Fact vào Data source view.

A screenshot of a computer

Description automatically generated

**Bước 4:** Chọn Add Related Tables để thêm tất cả bảng Dim vào Data source view và chọn Next để tiếp tục.

A screenshot of a computer

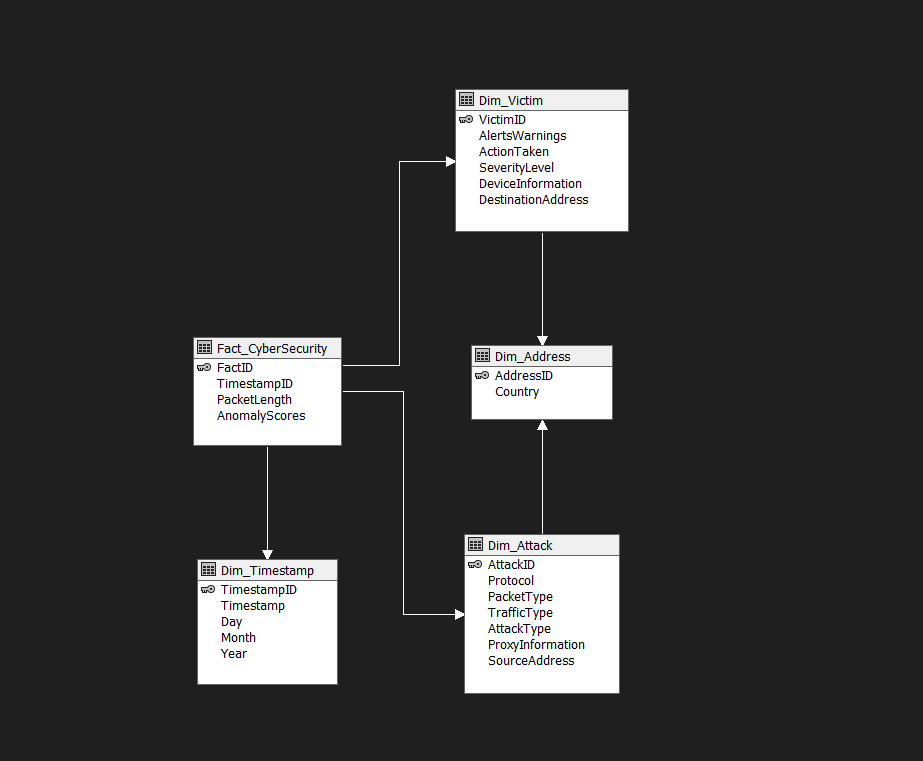
Description automatically generated

**Bước 5:** Chọn Finish để hoàn tất.

A screenshot of a computer

Description automatically generated

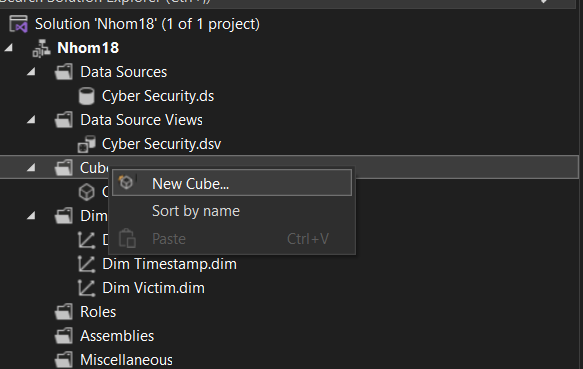
Sau khi kết thúc quá trình này, ta sẽ được data source view như hình sau:



# Xây dựng các khối (Cube) và deploy Cube

## Tạo Cube và Dimension

**Bước 1:** Tại Solution Explorer, right-click vào thư mục Cubes và chọn New Cube.



**Bước 2**: Hộp thoại Cube Wizard xuất hiện, chọn Next để tiếp tục

A screenshot of a wizard

Description automatically generated

**Bước 3:** Chọn use existing tables, sau đó chọn Next để tiếp tục.

**A screenshot of a computer program

Description automatically generated**

**Bước 4:** Chọn bảng Fact và các bảng Bridge để phân chia các measure group.

A screenshot of a computer

Description automatically generated

**Bước 5:** Chọn những độ đo đề xuất, sau đó chọn Next để tiếp tục.

A screenshot of a computer

Description automatically generated

**Bước 6:** Chọn danh sách bảng Dimension, sau đó chọn Next để tiếp tục.

A screenshot of a computer

Description automatically generated

**Bước 7:** Xem lại các độ đo được tạo, các bảng Dimension và chọn Finish để hoàn tất.

A screenshot of a computer

Description automatically generated

Sau khi kết thúc quá trình này, ta sẽ được kết quả như hình sau: Ảnh có chứa văn bản, ảnh chụp màn hình, phần mềm, Phần mềm đa phương tiện

Mô tả được tạo tự động

## Thêm thuộc tính và chỉnh sửa property cho Dimension

Ở khung Dimensions, lần lượt chọn vào từng bảng Dimension 🡪 Edit bảng Dimension để tiến hành chỉnh sửa bảng Dimension:

A screenshot of a computer

Description automatically generated

**Bảng Dim TimeStamp:**

**Bước 1:** Kéo thả thuộc tính TimeStamp, Day, Month, Year ở cột Data Source View sang cột Attributes.

A black screen with white text

Description automatically generated

**Bước 2:** Right-click tại bảng Dim TimeStamp và chọn Properties. Tại cửa sổ Properties, chỉnh sửa giá trị ErrorConfiguration 🡪 KeyDuplicate là IgnoreError.

A screenshot of a computer

Description automatically generated

**Bước 3:** Tại cửa sổ Propterties, chỉnh sửa giá trị UnknownMember là Hidden.

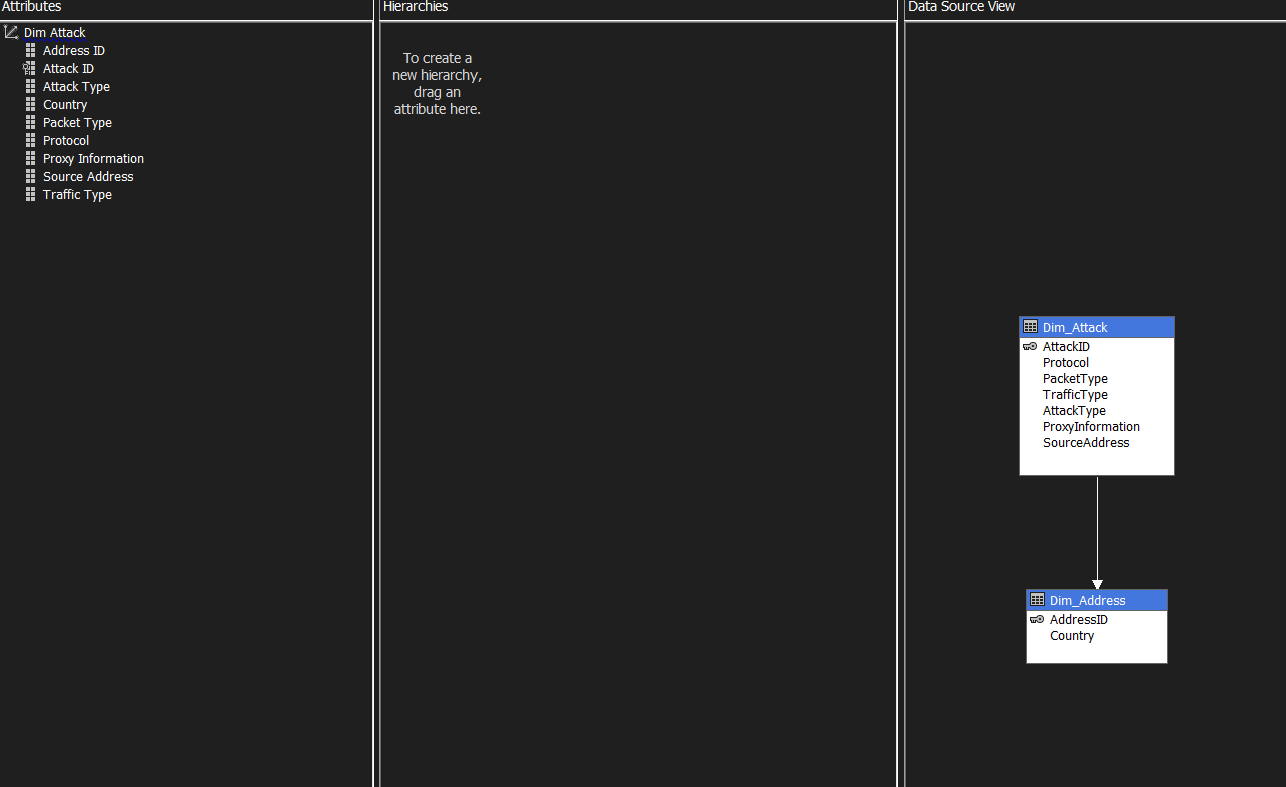
A screen shot of a computer

Description automatically generated A screenshot of a computer

Description automatically generated

**Bảng Dim Attack:**

**Bước 1:** Kéo thả thuộc tính ở cột Data Source View sang cột Attributes.



**Bước 2:** Right-click tại bảng Dim attack và chọn Properties. Tại cửa sổ Properties, chỉnh sửa giá trị ErrorConfiguration 🡪 KeyDuplicate là IgnoreError.

A screen shot of a computer

Description automatically generated

**Bước 3:** Tại cửa sổ Propterties, chỉnh sửa giá trị UnknownMember là Hidden.

A screenshot of a computer

Description automatically generated

**Bảng Dim Victim:**

**Bước 1:** Kéo thả thuộc ở cột Data Source View sang cột Attributes.

Ảnh có chứa văn bản, ảnh chụp màn hình, phần mềm, Phần mềm đa phương tiện

Mô tả được tạo tự động

**Bước 2:** Right-click tại bảng Dim\_Movie và chọn Properties. Tại cửa sổ Properties, chỉnh sửa giá trị ErrorConfiguration 🡪 KeyDuplicate là IgnoreError.

A screen shot of a computer

Description automatically generated

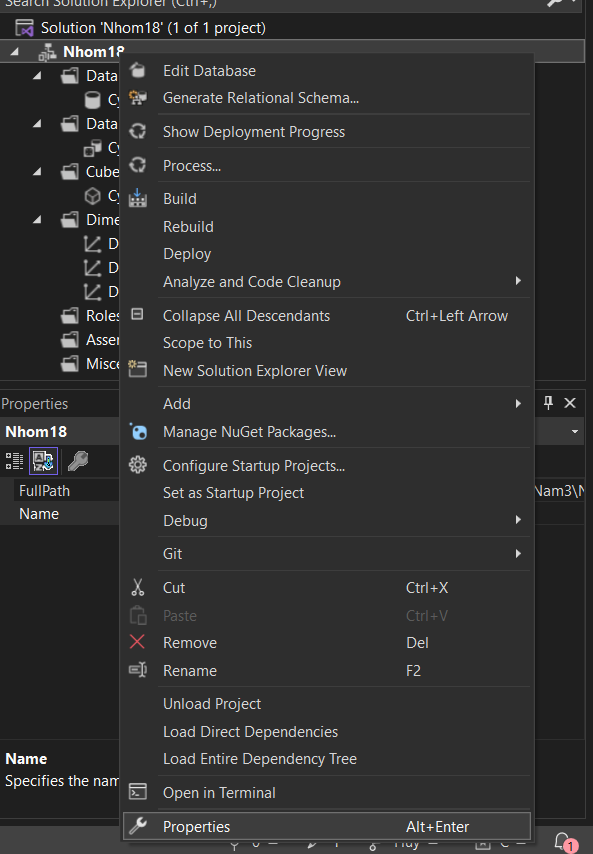
**Bước 3:** Tại cửa sổ Propterties, chỉnh sửa giá trị UnknownMember là Hidden.

A screenshot of a computer

Description automatically generated

## Deploy project SSAS

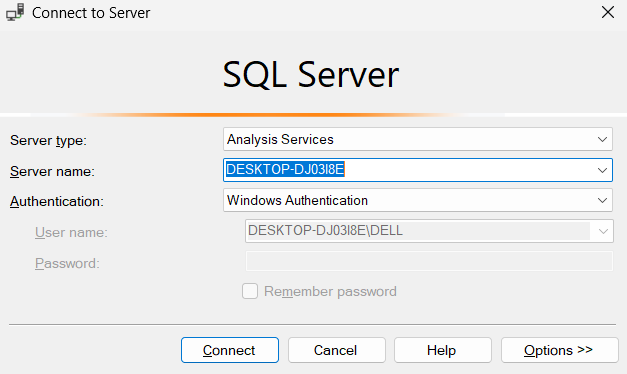
**Bước 1:** Tại Solution Explorer, right-click ở tên project và chọn Properties để chỉnh sửa kết nối đến Analysis Service của SQL Server.

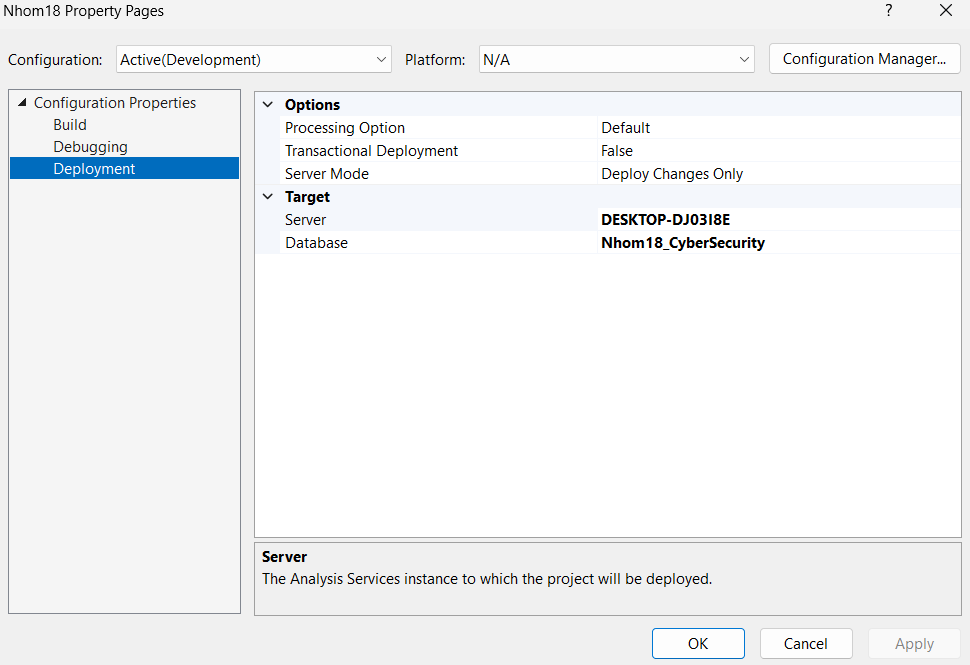


**Bước 2:** Ở cửa sổ vừa mở, đi đến Deployment và thực hiện đổi tên Server theo tên trong SQL Server.

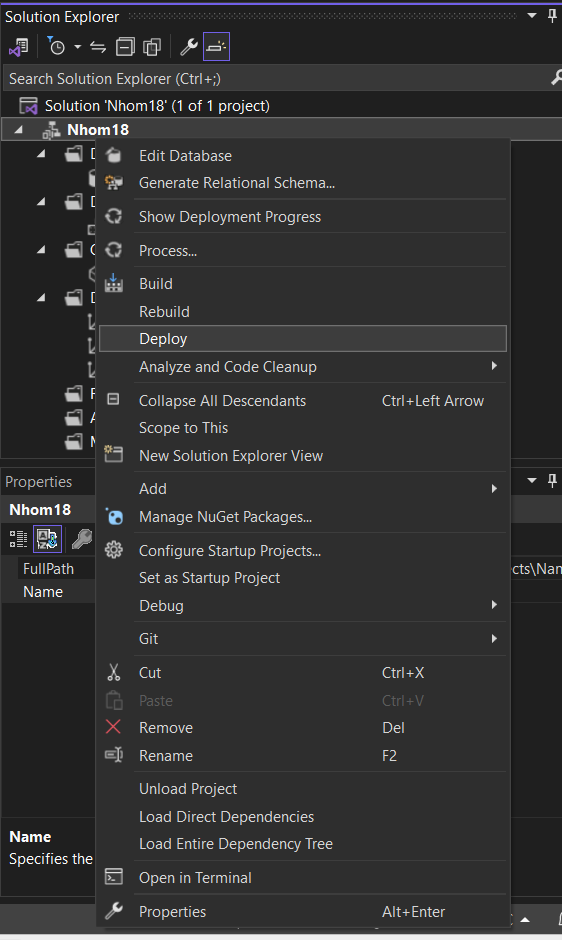
Server ban đầu là localhost:

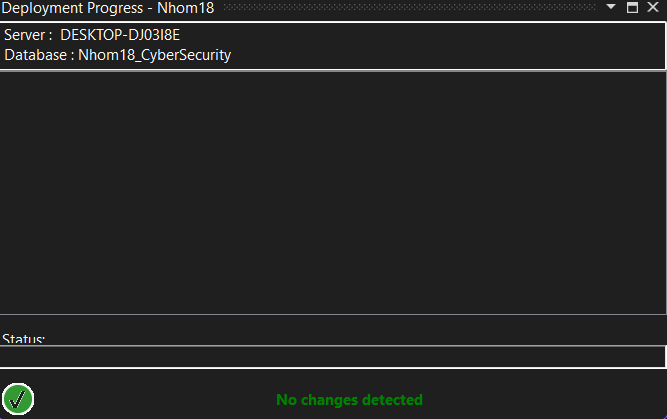
Mở SQL Server và lấy tên Server của Analysis Services như sau:





**Bước 3:** Right-click vào project đang hiện hành và chọn Deploy project.





# Xác định các độ đo (Measures)

Tại khối vừa tạo, chọn Show Measures Grid để hiện thị chi tiết các độ đo. Để tạo ra các độ đo mới, ta chọn Add new measure, sau đó tạo ra thêm một độ đo như sau:

A screenshot of a computer

Description automatically generated

Kết quả sau khi thực hiện thêm độ đo mới:

A screenshot of a computer

Description automatically generated

# Phân cấp bảng chiều

Thực hiện tạo các thuộc tính phân cấp (Hierachies) và định nghĩa Attribute Relationships cho bảng Dim Date. Tạo Hierarchy phân cấp theo Year 🡪 Month 🡪 TimeStamp:

**Bước 1:** Kéo những thuộc tính Year, Month, TimeStamp qua cửa sổ Hierarchies với thức tự từ trên xuống là phân cấp từ cao tới thấp.

A screenshot of a computer

Description automatically generated

**Bước 2:** Đổi tên Hierarchy.

A screenshot of a computer

Description automatically generated

**Bước 3:** Tại tab Attribute Relationships, định nghĩa các mối quan hệ. Thực hiện kéo thả phân cấp từ nhỏ đến lớn theo thứ tự từ trái sang phải.

A screenshot of a computer

Description automatically generated

Chỉnh sửa Relationship Type thành Rigid:

A screenshot of a computer

Description automatically generated

**Bước 4:** Chỉnh khóa cột (KeyColumns) và tên cột (Name Column) của thuộc tính TimeStamp. Vì thuộc tính này sẽ lấy khóa cột gồm chính nó và những thuộc tính cấp cao hơn.

Chuyển sang tab Dimension Structrure, cột Attributes, right-click vào thuộc tính Day và chọn Properties. Tại cửa số Properties, chọn KeyColumns. Thêm các thuộc tính cấp cao hơn vào KeyColumns, sau đó chọn OK để hoàn tất:

Screens screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Tại cửa sổ Properties, chọn NameColumn và chọn tên thuộc tính sẽ hiển thị trên Hierachy là TimeStamp:

Chỉnh thuộc tính OrderBy thành Key để Timestamp được sắp xếp theo thứ tự tăng dần:

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated

**Bước 5:** Chỉnh khóa cột (KeyColumns) và tên cột (Name Column) của thuộc tính Month. Vì thuộc tính này sẽ lấy khóa cột gồm chính nó và những thuộc tính cấp cao hơn.

Chuyển sang tab Dimension Structrure, cột Attributes, right-click vào thuộc tính Day và chọn Properties. Tại cửa số Properties, chọn KeyColumns. Thêm các thuộc tính cấp cao hơn vào KeyColumns, sau đó chọn OK để hoàn tất.

Screens screenshot of a computer

Description automatically generated

Tại cửa sổ Properties, chọn NameColumn và chọn tên thuộc tính sẽ hiển thị trên Hierachy là Month:

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Chỉnh thuộc tính OrderBy thành Key để Month được sắp xếp theo thứ tự tăng dần:

A screenshot of a computer

Description automatically generated

**Bước 6:** Thực hiện deploy project để đảm bảo không có lỗi xảy ra sau quá trình phân cấp.

Chuyển sang tab Browser để kiểm tra xem các thuộc tính đã phân cấp đúng hay chưa:

A white background with black numbers

Description automatically generated

# Thực hiện các truy vấn sử dụng SSAS, Pivot table

## Câu truy vấn 1

**Nội dung câu truy vấn:** truy vấn drill down tổng số cuộc tấn công bằng phương thức TCP theo từng tháng năm 2022

### Sử dụng SSAS

**Bước 1:** Vào mục Browser, kéo từ bảng Dim Attack thuộc tính Protocol. Tại cột Operator chọn “Equal” và Filter Expression chọn vào TCP, kéo từ bảng Dim Timestamp thuộc tính Year. Tại cột Operator chọn “Equal” và Filter Expression chọn vào 2022. Sau đó kéo các thuộc tính để hiển thị câu truy vấn, ta được kết quả:

A screenshot of a computer

Description automatically generated

### Sử dụng Pivot table



## Câu truy vấn 2

**Nội dung câu truy vấn:** Liệt kê top 5 cuộc tấn công có điểm cao nhất và phương thức của chúng.

### Sử dụng SSAS

**Bước 1:** Trong Cube, vào mục Calculation, tạo mới nameset:

A screenshot of a computer

Description automatically generated

**Bước 2**: Vào mục Browser, kéo từ bảng Dim Attack thuộc tính AttackId. Tại cột Operator chọn “In” và Filter Expression chọn vào nameset “TOP5” vừa tạo. Sau đó kéo các thuộc tính để hiển thị câu truy vấn, ta được kết quả:

A screenshot of a computer

Description automatically generated

### Sử dụng Pivot table

**A screenshot of a computer

Description automatically generated**

## Câu truy vấn 3

**Nội dung câu truy vấn:** Liệt kê tổng số cuộc tấn công từng tháng trong năm với loại là mã độc.

### Sử dụng SSAS

**Bước 1:** Vào mục Browser, kéo từ bảng Dim Attack thuộc tính Attack Type. Tại cột Operator chọn “Equal” và Filter Expression chọn vào Malware. Sau đó kéo các thuộc tính để hiển thị câu truy vấn, ta được kết quả:

A screenshot of a computer

Description automatically generated

### Sử dụng Pivot table

**A screenshot of a computer

Description automatically generated**

## 

## Câu truy vấn 4

**Nội dung câu truy vấn:** Truy vấn tổng điểm Anomaly Scores của các cuộc tấn công theo từng Phương thức (Protocol) và Năm với phương thức là Malware.

### Sử dụng SSAS

**Bước 1**: Vào mục Browser, kéo từ bảng Dim Attack thuộc tính Attack Type. Tại cột Operator chọn “Equal” và Filter Expression chọn vào Malware. Sau đó kéo các thuộc tính để hiển thị câu truy vấn, ta được kết quả:

A screenshot of a computer

Description automatically generated

### Sử dụng Pivot table

**A screenshot of a graph

Description automatically generated**

### 

## Câu truy vấn 5

**Nội dung câu truy vấn:** Truy vấn tính điểm trung bình của các cuộc tấn công Malware theo từng tháng trong năm 2022.

### Sử dụng SSAS

**Bước 1**: Tạo Calculation tính trung bình thời lượng phim của top 5 phim có điểm bình chọn cao nhất:

A screenshot of a computer

Description automatically generated

**Bước 2**: Vào mục Browser, kéo từ bảng Dim Time Stamp thuộc tính Attack Type. Tại cột Operator chọn “Equal” và Filter Expression chọn vào 2022. Sau đó kéo các thuộc tính để hiển thị câu truy vấn, ta được kết quả:

A screenshot of a computer

Description automatically generated

### Sử dụng Pivot table

**A screenshot of a computer

Description automatically generated**

## Câu truy vấn 6

**Nội dung câu truy vấn:** Liệt kê số cuộc tấn công bị chặn theo năm.

### Sử dụng SSAS

**Bước 1**: Vào mục Browser, kéo từ bảng Dim Victim thuộc tính Action Taken. Tại cột Operator chọn “Equal” và Filter Expression chọn vào Blocked. Sau đó kéo các thuộc tính để hiển thị câu truy vấn, ta được kết quả:

A screenshot of a computer

Description automatically generated

### Sử dụng Pivot table

**A screenshot of a computer

Description automatically generated**

### 

## Câu truy vấn 7

**Nội dung câu truy vấn:** Truy vấn chiều dài các gói tin đã bị hệ thống bảo mật bỏ qua theo từng năm.

### Sử dụng SSAS

**Bước 1**: Tạo Calculation tính trung bình thời lượng phim của top 5 phim có điểm bình chọn cao nhất:

A screenshot of a computer

Description automatically generated

**Bước 2** Sau đó kéo các thuộc tính để hiển thị câu truy vấn, ta được kết quả:

A screenshot of a computer

Description automatically generated

### Sử dụng Pivot table

**A screenshot of a spreadsheet

Description automatically generated**

## Câu truy vấn 8

**Nội dung câu truy vấn:** Liệt kê số lượng các cuộc tấn công thành công theo từng phương thức.

### Sử dụng SSAS

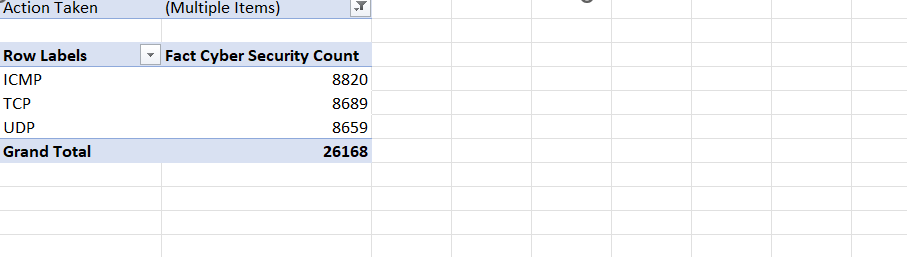
**Bước 1**: Vào mục Browser, kéo từ bảng Dim Victim thuộc tính Action Taken. Tại cột Operator chọn “Not Equal” và Filter Expression chọn vào Blocked

**Bước 2** Sau đó kéo các thuộc tính để hiển thị câu truy vấn, ta được kết quả:

A screenshot of a computer

Description automatically generated

### Sử dụng Pivot table



## Câu truy vấn 9

**Nội dung câu truy vấn:** Liệt kê top 10 những cuộc tấn công có độ dài lớn nhất của hai nước Nhật Bản và Mỹ.

### Sử dụng SSAS

**Bước 1**: Trong Cube, vào mục Calculation, tạo mới nameset:

Ảnh có chứa văn bản, ảnh chụp màn hình, phần mềm, Phần mềm đa phương tiện

Mô tả được tạo tự động

**Bước 2**: Vào mục Browser, kéo từ bảng Dim Attack thuộc tính AttackId. Tại cột Operator chọn “In” và Filter Expression chọn vào nameset “[Top 10 Attacks by Packet Length]” vừa tạo. Sau đó kéo các thuộc tính để hiển thị câu truy vấn, ta được kết quả:

Ảnh có chứa văn bản, ảnh chụp màn hình, màn hình, số

Mô tả được tạo tự động

### Sử dụng Pivot table

**Ảnh có chứa văn bản, số, phần mềm, Phông chữ

Mô tả được tạo tự động**

## Câu truy vấn 10

**Nội dung câu truy vấn:** Truy vấn cuộc tấn công đã được cảnh báo nhưng bị bỏ qua trong năm 2020

### Sử dụng SSAS

**Bước 1:** Vào mục Browser, kéo từ bảng Dim Victim thuộc tính Alert Warnings, Action Taken, Timestamp. Tại cột Operator chọn Equal Alert Triggered, Ignored, 2020

Ảnh có chứa văn bản, ảnh chụp màn hình, hàng, phần mềm

Mô tả được tạo tự động

**Bước 2:** Sau đó kéo các thuộc tính để hiển thị câu truy vấn, ta được kết quả:

Ảnh có chứa văn bản, ảnh chụp màn hình, màn hình, phần mềm

Mô tả được tạo tự động

### Sử dụng Pivot table

Ảnh có chứa văn bản, ảnh chụp màn hình, phần mềm, số

Mô tả được tạo tự động

## Câu truy vấn 11

**Nội dung câu truy vấn**: Tổng độ dài các gói tin của từng phương thức của 20 cuộc tấn công có điểm AnomalyScore cao nhất.

### Sử dụng SSAS

**Bước 1**: Trong Cube, vào mục Calculation, tạo mới nameset:

Ảnh có chứa ảnh chụp màn hình, văn bản, phần mềm, Phần mềm đa phương tiện

Mô tả được tạo tự động

**Bước 2**: Vào mục Browser, kéo từ bảng Dim Attack thuộc tính AttackId. Tại cột Operator chọn “In” và Filter Expression chọn vào nameset “[Top 20 Attacks]” vừa tạo. Sau đó kéo các thuộc tính để hiển thị câu truy vấn, ta được kết quả:

Ảnh có chứa văn bản, ảnh chụp màn hình, màn hình, phần mềm

Mô tả được tạo tự động

### Sử dụng Pivot table

**Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ, hàng

Mô tả được tạo tự động**

CHƯƠNG 4. QUÁ TRÌNH LẬP BÁO BIỂU

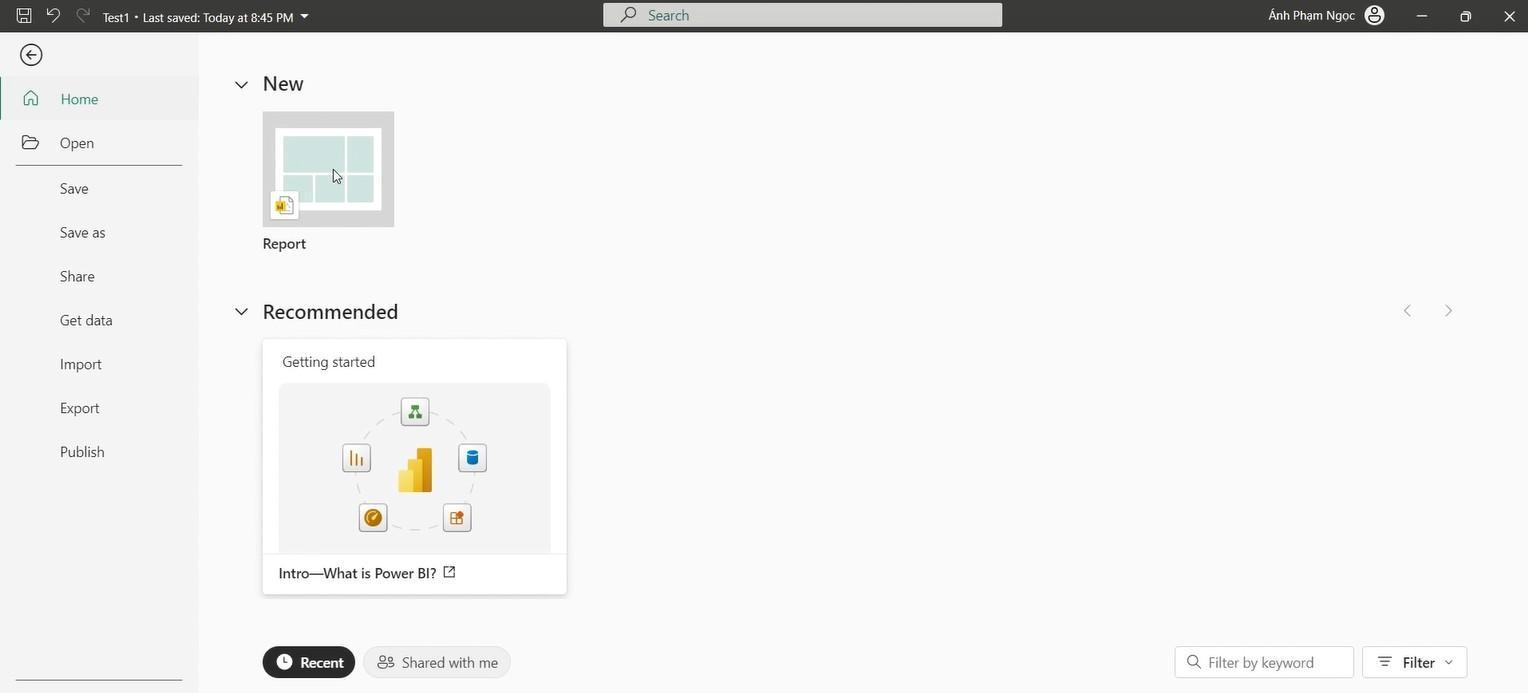
# Quá trình lập báo biểu bằng PowerBI

## Chuẩn bị công cụ

Đăng ký tài khoản và tải Power BI theo hướng dẫn tại đường link: [Sign in and download](https://learn.microsoft.com/en-us/power-bi/fundamentals/service-self-service-signup-for-power-bi) [Power BI Desktop](https://learn.microsoft.com/en-us/power-bi/fundamentals/service-self-service-signup-for-power-bi)

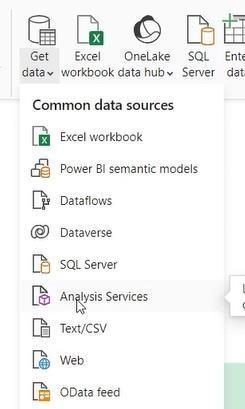
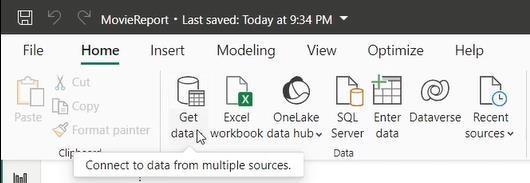
## Tạo mới report và kết nối

* Tại mục New chọn Report để tạo mới.



* Kết nối với cube trong SSAS.

**Bước 1:** Chọn Get data tại mục Home trên thanh công cụ. Chọn vào Analysis Services.



**Bước 2:** Nhập Server và Database.

Ảnh có chứa văn bản, ảnh chụp màn hình, số, Phông chữ

Mô tả được tạo tự động

* Server là server name kết nối Analysis Services trong SQL Server.

### Tạo report 1

***Câu truy vấn*:** Truy vấn drill down tổng số cuộc tấn công của từng phương thức theo từng tháng năm có kiểu tấn công là Mã độc.

**Bước 1:** Chọn biểu đồ Chart trong mục Visualizations.

Ảnh có chứa văn bản, ảnh chụp màn hình, phần mềm, Phông chữ

Mô tả được tạo tự động

**Bước 2:** Chọn từ Data các thuộc tính: Fact Cyber Security Count.

* Thêm điều kiện lọc cho AttackType: chọn vào “Malware”

Ảnh có chứa văn bản, phần mềm, ảnh chụp màn hình, số

Mô tả được tạo tự động

**Bước 3:** Thêm Legend các thuộc tính Year\_Month\_Timestamp-Year, Year\_Month\_Timestamp-Month

Ảnh có chứa văn bản, ảnh chụp màn hình

Mô tả được tạo tự động

* Kết quả với biểu đồ tròn:

Ảnh có chứa văn bản, hàng, biểu đồ, Sơ đồ

Mô tả được tạo tự động

* Drill Down xuống tháng:

Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ, Sơ đồ

Mô tả được tạo tự động

* Sơ đồ hình cột:

Ảnh có chứa văn bản, ảnh chụp màn hình, số, Phông chữ

Mô tả được tạo tự động

* Drill Down xuống tháng:

Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ, biểu đồ

Mô tả được tạo tự động

Sơ đồ hình tròn thể hiện, phương thức tấn công ta có thể chọn từng phương thức để thể hiện tổng số cuộc tấn công của phương thức đó theo từng tháng, năm.

Ảnh có chứa văn bản, ảnh chụp màn hình, biểu đồ, Phông chữ

Mô tả được tạo tự động

**Bước 4:** Tổng hợp lại report

Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ, biểu đồ

Mô tả được tạo tự động

***Nhận xét:*** từ biểu đồ ta có thể thấy:

* Phương thức được dung để tấn công nhiều nhất là ICMP
* Mỗi năm có thể có lên tới 10000 cuộc tấn công
* Tháng 3, 11, 6 là những tháng được sử dụng để tấn công nhiều nhất.

### Tạo report 2

***Câu truy vấn:*** Tổng độ dài các gói tin của từng phương thức hoặc từng năm của 20 cuộc tấn công có điểm AnomalyScore cao nhất.

**Bước 1:** Chọn biểu đồ Chart trong mục Visualizations.

Ảnh có chứa văn bản, ảnh chụp màn hình, phần mềm, Phông chữ

Mô tả được tạo tự động

**Bước 2**: Chọn từ Data các thuộc tính: Packet Length , AnomalyScore,

Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ, số

Mô tả được tạo tự động

* Thêm điều kiện lọc cho AttackId: chọn Show Items Top N , Nhập top 20, chọn thuộc tính tính toán là Anomaly Score .

Ảnh có chứa văn bản, ảnh chụp màn hình, số, Phông chữ

Mô tả được tạo tự động

**Bước 3**: Thêm Legend các thuộc tính Timestamp, Protocol,AttackId

Ảnh có chứa văn bản, ảnh chụp màn hình, phần mềm

Mô tả được tạo tự động

* Kết quả với biểu đồ bảng:

Ảnh có chứa văn bản, ảnh chụp màn hình, số, thực đơn

Mô tả được tạo tự động

* Kết quả với biểu đồ cột với tổng gói tin theo từng phương thức:

Ảnh có chứa văn bản, ảnh chụp màn hình, biểu đồ, Phông chữ

Mô tả được tạo tự động

* Kết quả với biểu đồ cột với tổng gói tin theo từng năm:

Ảnh có chứa ảnh chụp màn hình, văn bản, biểu đồ, Phông chữ

Mô tả được tạo tự động

**Bước 4:** Tổng hợp lại report

Ảnh có chứa văn bản, ảnh chụp màn hình, biểu đồ, số

Mô tả được tạo tự động

***Nhận xét:*** từ biểu đồ ta có thể thấy:

* Trong 20 cuộc tấn công có điểm bất thường cao, thì phương thức TCP độ dài lớn nhất
* Trong 20 cuộc tấn công có điểm bất thường cao, thì tổng độ dài các gói tin trong năm 2021 là lớn nhất.
* Độ chênh lệch điểm 20 gói tiên lớn nhất không nhiều từ 0.05- 0.01

### Tạo report 3

***Câu truy vấn:*** Truy vấn tổng chiều dài gói tin theo từng phương thức trong từng tháng, năm và phân bổ của nó trên thế giới

**Bước 1:** Chọn biểu đồ Chart trong mục Visualizations.

Ảnh có chứa văn bản, ảnh chụp màn hình, phần mềm, Phông chữ

Mô tả được tạo tự động

**Bước 2**: Chọn từ Data các thuộc tính: Packet Length

Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ

Mô tả được tạo tự động

**Bước 3**: Thêm Legend các thuộc tính Year\_Month\_Timestamp, Action Taken.

Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ, số

Mô tả được tạo tự động

* Kết quả với biểu đồ cột chồng

Ảnh có chứa văn bản, ảnh chụp màn hình, số, Phông chữ

Mô tả được tạo tự động

* Phân bổ gói tin theo từng quốc gia:

Ảnh có chứa văn bản, ảnh chụp màn hình, Nhiều màu sắc, hình vuông

Mô tả được tạo tự động

* Sơ đồ hình cột:

Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ, biểu đồ

Mô tả được tạo tự động

**Bước 4:** Tổng hợp lại report

Ảnh có chứa văn bản, ảnh chụp màn hình, Hình chữ nhật, biểu đồ

Mô tả được tạo tự động

***Nhận xét:*** từ biểu đồ ta có thể thấy:

* Mỹ là nước nhận được tổng độ dài gói tin nhiều nhất.
* Tháng 4 là tháng ít được sử dụng để tấn công nhất, tổng độ dài gói tin thấp nhất.
* Tổng độ dài gói tin có phương thức là TCP vào năm 2021 là cao nhất.

# Quá trình lập báo biểu bằng Google Data Studio

## Tạo report 1

***Câu truy vấn:*** Truy vấn tổng điểm Anomaly Scores của các cuộc tấn công theo từng Phương thức (Protocol) và tháng với phương thức là Malware trong năm 2021

**Bước 1**: : Chọn vào thêm biểu đồ, chọn vào biểu đồ Bảng và biểu đồ cột.

A screenshot of a computer

Description automatically generated

**Bước 2**: Chọn vào sơ đồ vừa tạo, tiến hành tạo bộ lọc cho sơ đồ

* Tạo bộ lọc mới gồm Tag là “Bao gồm”, Chọn trường là AttackType, chọn điều kiện là Bằng. Giá trị là “Malware”
* Tạo bộ lọc mới gồm Tag là “Bao gồm”, Chọn trường là Year, chọn điều kiện là Bằng. Giá trị là “2022”

A screenshot of a chat

Description automatically generated

**Bước 3**: kéo Protocol từ bảng dữ liệu vào phần phương diện, kéo AnomalyScore vào mục chỉ số.

A screenshot of a chat

Description automatically generated

* Kết quả với biểu đồ hình cột:

A blue bar graph with black text

Description automatically generated

* Kết quả với biểu đồ hình tròn:

A pie chart with numbers and a few percentages

Description automatically generated

* Sơ đồ hình cột:

Ảnh có chứa văn bản, ảnh chụp màn hình, số, Phông chữ

Mô tả được tạo tự động

Sơ đồ hình thác đổ thể hiện các quốc gia nguồn gốc của cuộc tấn công ta có thể chọn từng quốc gia để thể hiện tổng điểm Anomaly Scores của các cuộc tấn công theo từng Phương thức.

A colorful bar chart with text

Description automatically generated

**Bước 4:** Tổng hợp lại report

A screenshot of a graph

Description automatically generated

***Nhận xét:*** từ biểu đồ ta có thể thấy:

* Phương thức ICMP có khả nhận được nhận biết cao nhất
* Các phương thức có tổng điểm tương đương nhau
* Riêng nước mỹ có tổng điểm AnomalyScore cao nhất với hơn 600 ngàn điểm.

## Tạo report 2

**Câu truy vấn:** Top 10 Quốc Gia Có Nhiều Cuộc Tấn Công Mạng Nhất trong năm 2022

**Bước 1**: Chọn vào thêm biểu đồ, chọn vào biểu đồ Bảng và biểu đồ cột.

A screenshot of a computer

Description automatically generated

**Bước 2**: Chọn vào sơ đồ vừa tạo, tiến hành tạo bộ lọc cho sơ đồ

* Tạo bộ lọc mới gồm Tag là “Bao gồm”, Chọn trường là Year, chọn điều kiện là Bằng. Giá trị là “2022”

A screenshot of a computer

Description automatically generated

**Bước 3**: kéo Source Country từ bảng dữ liệu vào phần phương diện, kéo Record Count vào mục chỉ số.

A screenshot of a computer

Description automatically generated

**Bước 4**: Điều chỉnh mục sắp xếp giảm dần và điều chỉnh số hàng trên 1 trang là 10.

A screenshot of a chat

Description automatically generated

* Kết quả với biểu đồ bảng:

A screenshot of a computer

Description automatically generated

* Kết quả với biểu đồ hình cột

A graph with green bars

Description automatically generated

**Bước 5**: Sơ đồ hình tròn thể hiện các phương thức và các tháng trong năm ta có thể chọn filter theo từng phương thức và từng tháng trong năm.

A screenshot of a graph

Description automatically generated

A colorful pie chart with numbers and a number

Description automatically generated

**Bước 6:** Tổng hợp lại report

A screenshot of a graph

Description automatically generated

A screenshot of a computer

Description automatically generated

***Nhận xét:*** từ biểu đồ ta có thể thấy:

* Mỹ là quốc gia có nhiều cuộc tấn công nhất lên đến 4496 cuộc tấn công trong năm 2022.
* Tháng 11 là tháng có cao điểm của các cuộc tấn công của các nước trên thế giới
* Tại Nhật Bản, thì phương thức HTTP được sử dụng phổ biến để tấn công

## Tạo report 3

***Câu truy vấn:*** Tổng dung lượng gói tin của từng hành động khi gói tin được gửi trong từng tháng năm 2021

**Bước 1**: Chọn biểu đồ Chart trong mục Visualizations.

A screenshot of a computer

Description automatically generated

**Bước 2**: Chọn vào sơ đồ vừa tạo, tiến hành tạo bộ lọc cho sơ đồ

* Tạo bộ lọc mới gồm Tag là “Bao gồm”, Chọn trường là Year, chọn điều kiện là Bằng. Giá trị là “2021”

A screenshot of a chat

Description automatically generated

**Bước 3**: kéo Month từ bảng dữ liệu vào phần phương diện, kéo Packet Length vào mục chỉ số.

A screenshot of a graph

Description automatically generated

* Tổng dung lượng gói tính của toàn bộ tháng:

A blue rectangle with white text

Description automatically generated

* Kết quả với biểu đồ đường:

A graph with blue line

Description automatically generated

* Kết quả với biểu đồ cột chồng :

A graph with blue and pink squares

Description automatically generated

Sơ đồ hình tròn thể hiện các phương thức và các tháng trong năm ta có thể chọn filter theo từng phương thức và từng tháng trong năm.

A colorful circle with numbers and text

Description automatically generated

**Bước 4:** Tổng hợp lại report

A screenshot of a graph

Description automatically generated

***Nhận xét:*** từ biểu đồ ta có thể thấy:

* Tổng dung lượng gói tin cao điểm được gửi vào mùa hè từ thứ 4 đến tháng 8.
* Tổng dung lượng các gói tin của các gói bị Ignore của từng tháng cao hơn Blocked, và Logged
* Tổng dung lượng gói tin ICMP chiếm tỉ lệ cao nhất.

# DANH MỤC TÀI LIỆU THAM KHẢO