BÁO CÁO DATA WAREHOUSE

**<< Quy định viết Tên dự án**

**Project Title (Tên dự án):** Top Video Games

**Data Warehouse Accident Analysis in the US (2016–2021)**

Môn học: Data Warehouse

Lớp: DP20301

Nhóm : 3  
Sinh viên 1: Vũ Đình An MSSV: PS45462

Sinh viên 2: Võ Thành Phúc MSSV: …………………  
Sinh viên 3: Phạm Minh Tình MSSV: …………………  
Sinh viên 4: Trương Minh Vũ MSSV: …………………  
GVHD: Trần Văn Huy

Báo cáo đồ án nhóm được nộp nhằm hoàn thành một phần yêu cầu của học phần  
“Data WareHouse – DAT204”.

© Copyright by Group 1 2026

<<Tên Dự án>>

TÓM TẮT ĐỀ TÀI

Dự án tập trung xây dựng kho dữ liệu (Data Warehouse) phục vụ phân tích tai nạn giao thông tại Hoa Kỳ giai đoạn 2016–2021. Dữ liệu được xử lý qua quy trình ETL và tổ chức theo mô hình Star Schema nhằm hỗ trợ phân tích xu hướng, mức độ nghiêm trọng và các yếu tố ảnh hưởng đến tai nạn.

<<Sử dụng style “Abstract text” - double spaced>>

« **Mục lục (TOC) có thể được chèn tự động tại đây nếu bạn đã sử dụng đúng các style cho tiêu đề chương, mục và tiểu mục.  
Trong Microsoft Word, bạn có thể kiểm tra nội dung sẽ hiển thị trong Mục lục bằng cách chuyển sang chế độ “Outline” và ẩn phần nội dung văn bản (body text).** »

**Mục lục**

[1. GIỚI THIỆU 6](#_heading=h.7mk67bmjtekt)

[2. MỤC TIÊU & PHẠM VI 6](#_heading=h.naiz7o4jgq7c)

[3. DATASET 6](#_heading=h.67g6ehn7rsff)

[4. PHÂN TÍCH YÊU CẦU 6](#_heading=h.2fmoz2hkzbep)

[5. THIẾT KẾ DATA WAREHOUSE 6](#_heading=h.8f86rgogp44s)

[6. ETL 6](#_heading=h.bxktysii90n)

[7. CÔNG NGHỆ SỬ DỤNG 7](#_heading=h.w7je8zqvxw37)

[8. PHÂN TÍCH & DASHBOARD 7](#_heading=h.cvjl88sm6g6h)

[9. ĐÁNH GIÁ KẾT QUẢ 7](#_heading=h.4mlvcc9xh8o1)

[10. HẠN CHẾ & HƯỚNG PHÁT TRIỂN 7](#_heading=h.htmmy3xre2mx)

[TÀI LIỆU THAM KHẢO 7](#_heading=h.y5t5eioy4y3e)

**Mục Lục Hình ảnh(nếu có)**

**Mục Lục hình ảnh(nếu có)**

# GIỚI THIỆU

**Phần nội dung chính báo cáo bắt đầu từ đây.**

« **Sử dụng các Heading styles từ Heading 1 đến Heading 9 cho nội dung luận văn.  
Tiêu đề chương sử dụng style Heading 1.** »

« **Tiêu đề mục cấp 1 sử dụng style Heading 2.** »

« **Phần nội dung văn bản sử dụng style Body Text hoặc Normal.  
Style mặc định là Times New Roman, cỡ chữ 12 pt, giãn dòng 1.5, canh đều hai lề (block style), không thụt dòng đầu đoạn, và có khoảng cách 12 pt phía trên mỗi đoạn văn.** »

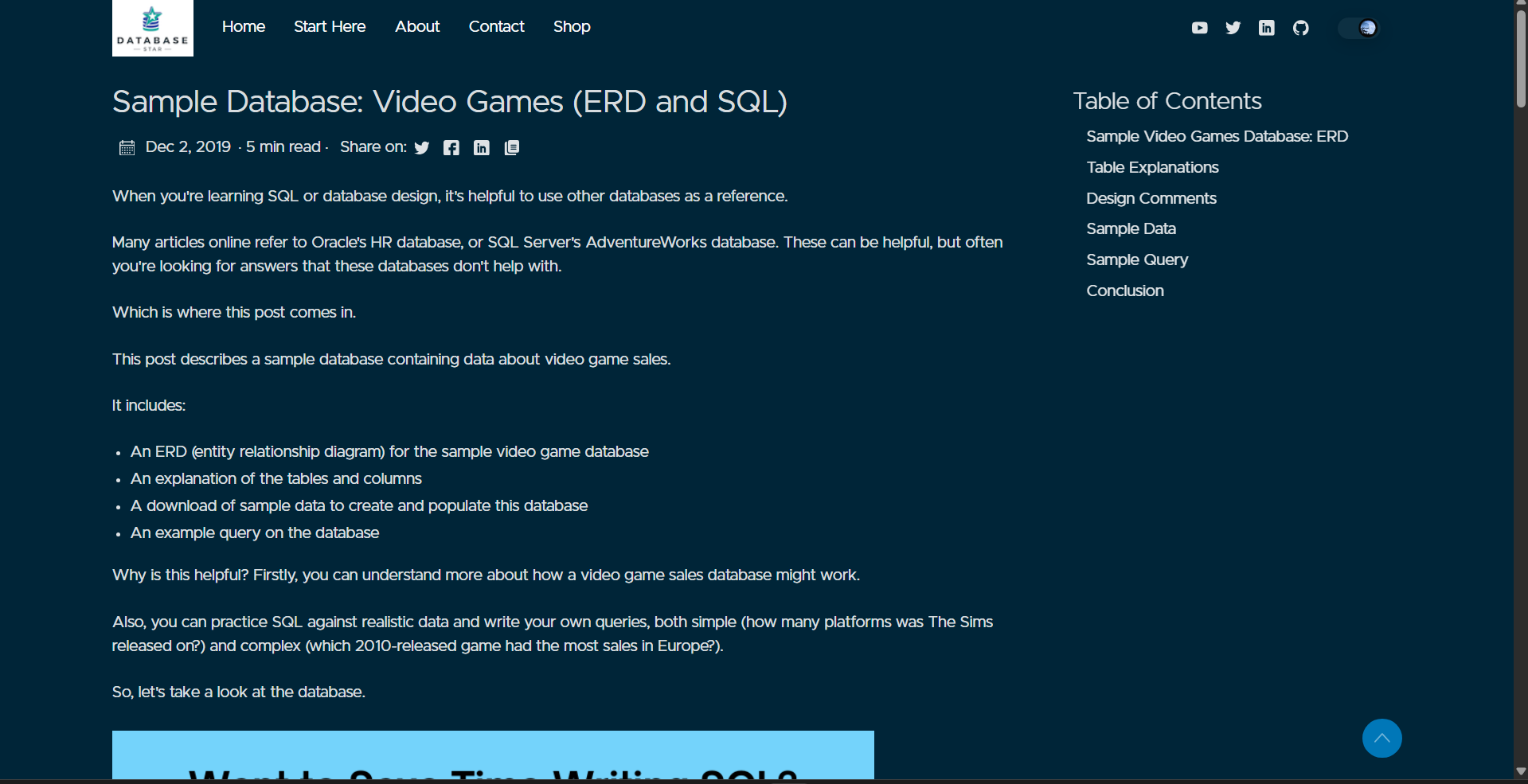
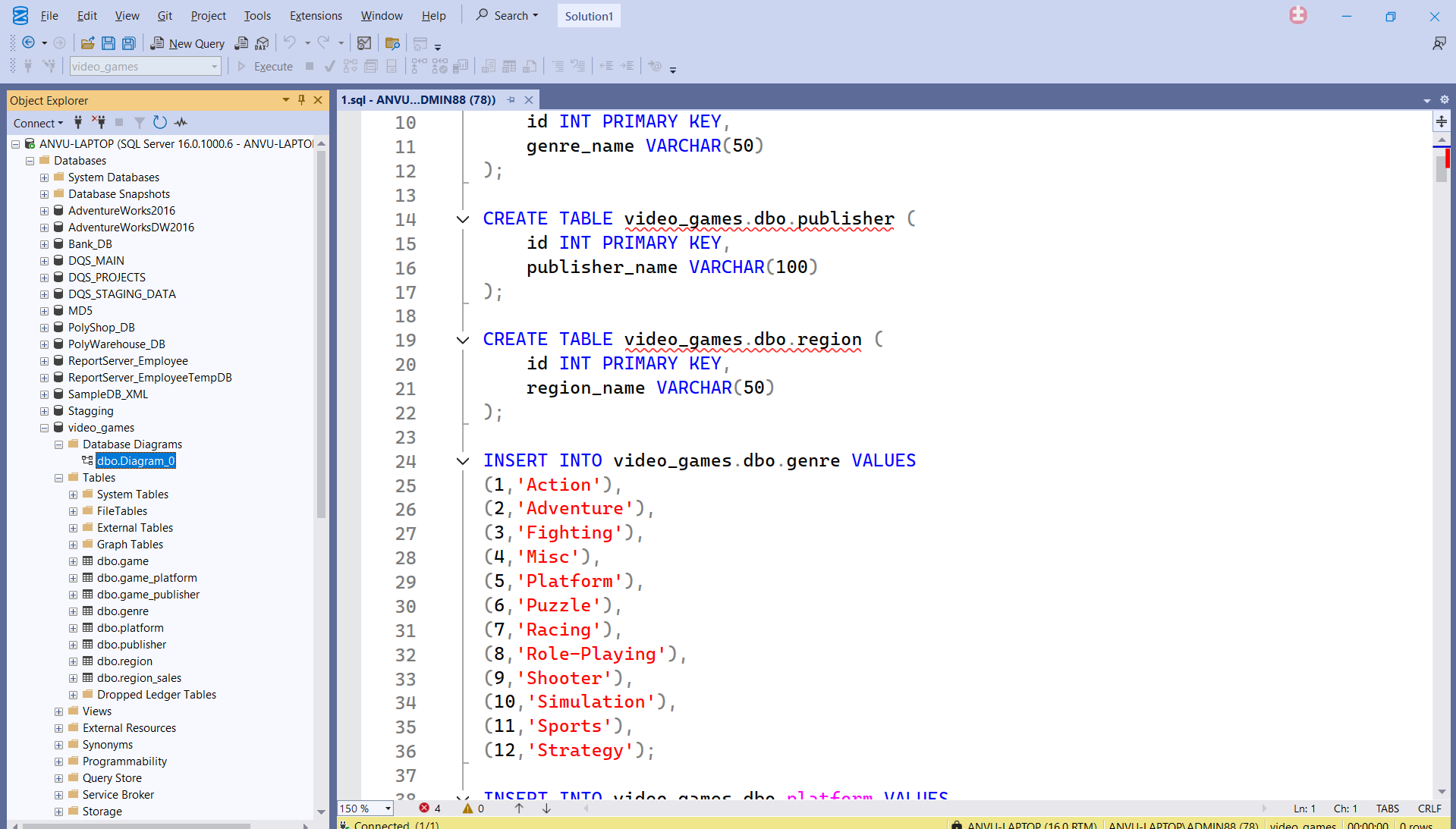
Tai nạn giao thông là vấn đề nghiêm trọng ảnh hưởng đến an toàn và kinh tế xã hội. Việc ứng dụng Data Warehouse giúp tổng hợp và phân tích dữ liệu hiệu quả, hỗ trợ ra quyết định.

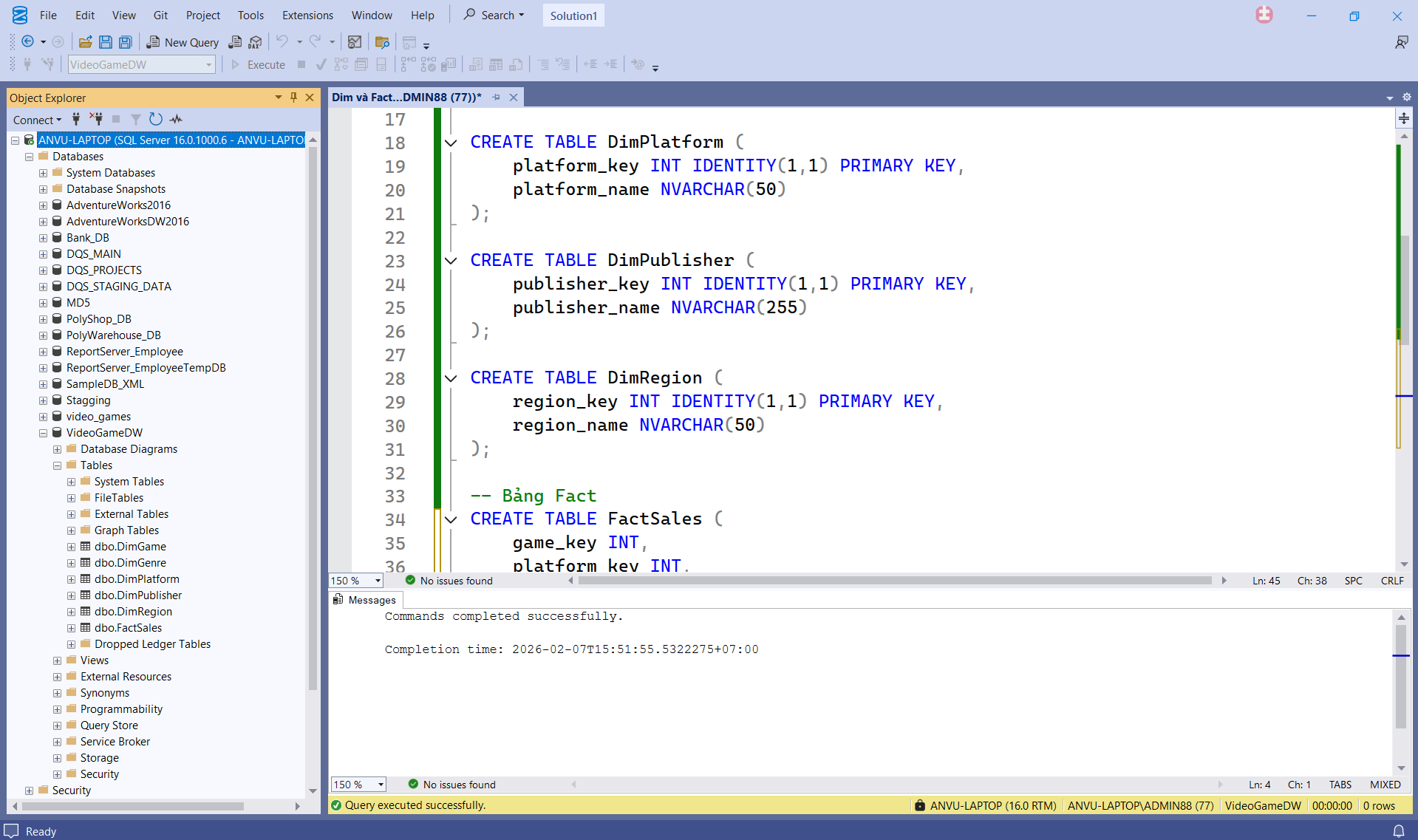
# 2. MỤC TIÊU & PHẠM VI

- Thiết kế Data Warehouse theo mô hình Star Schema  
- Thực hiện quy trình ETL từ dữ liệu thô  
- Xây dựng báo cáo và dashboard phân tích  
  
Phạm vi: Dữ liệu tai nạn giao thông tại Hoa Kỳ từ 2016 đến 2021.

# 3. DATASET

Dữ liệu được lấy từ Kaggle – US Accidents Dataset. Dataset bao gồm thông tin thời gian, vị trí địa lý, điều kiện thời tiết, giới hạn tốc độ và mức độ nghiêm trọng của tai nạn.

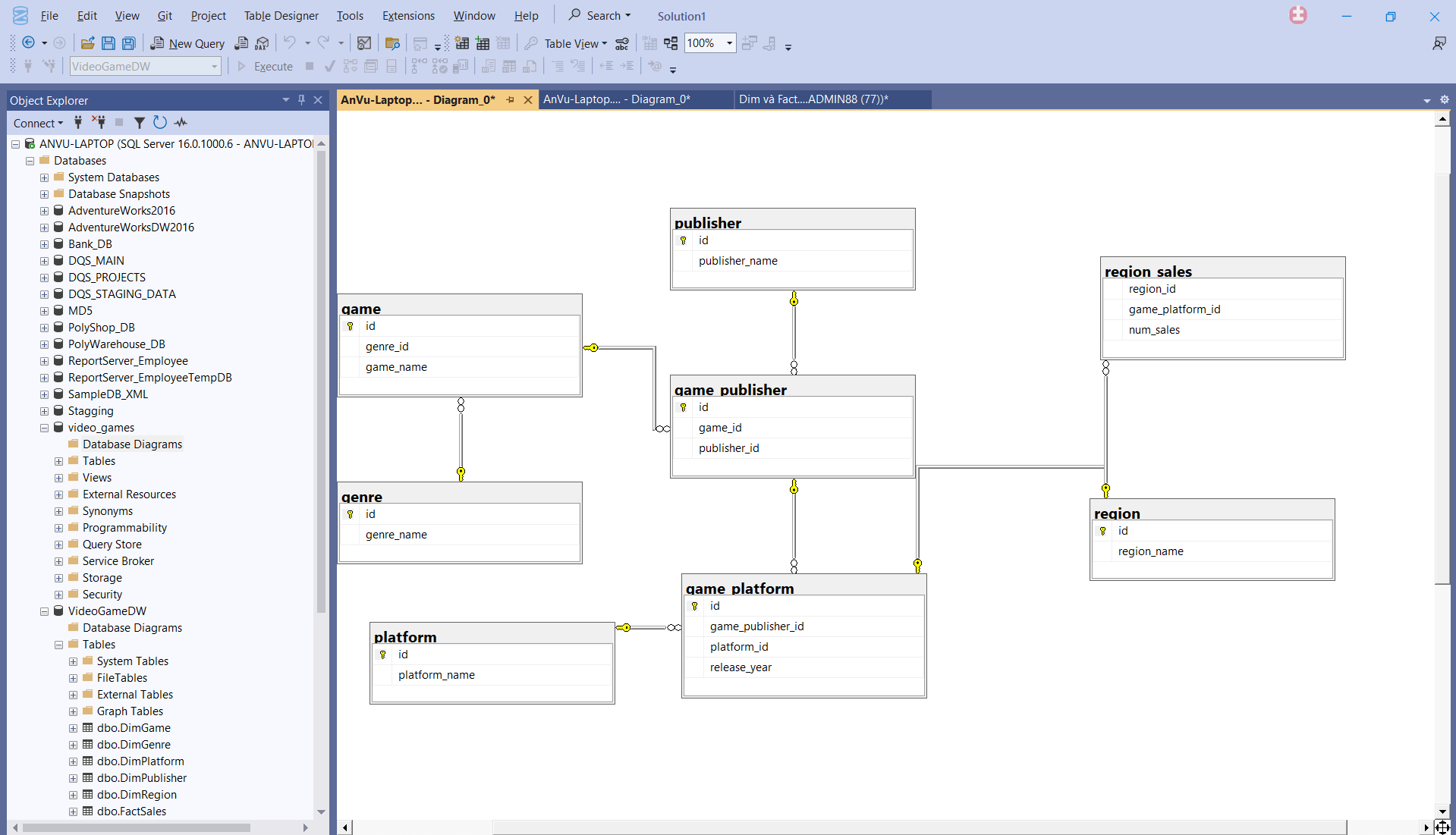
Database: [https://www.databasestar.com/sample-database-video-games/#documentTop](https://www.databasestar.com/sample-database-video-games/#documentTop)

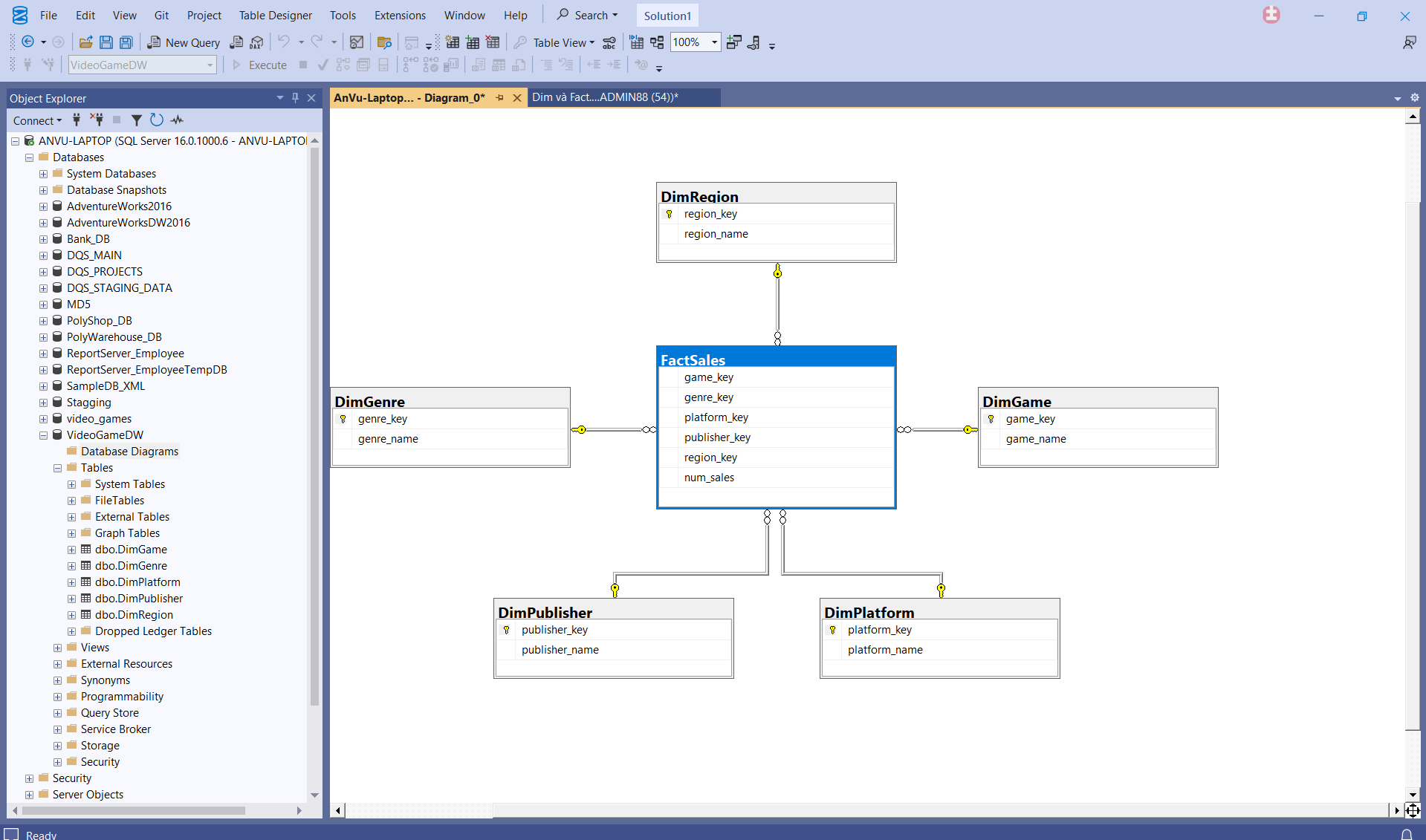
Tạo bảng Dim và Fact

# 4. PHÂN TÍCH YÊU CẦU

Các câu hỏi phân tích chính:  
- Bang nào xảy ra nhiều tai nạn nhất?  
- Tai nạn thường xảy ra vào thời điểm nào trong ngày?  
- Điều kiện thời tiết ảnh hưởng như thế nào đến mức độ nghiêm trọng?

# 5. THIẾT KẾ DATA WAREHOUSE

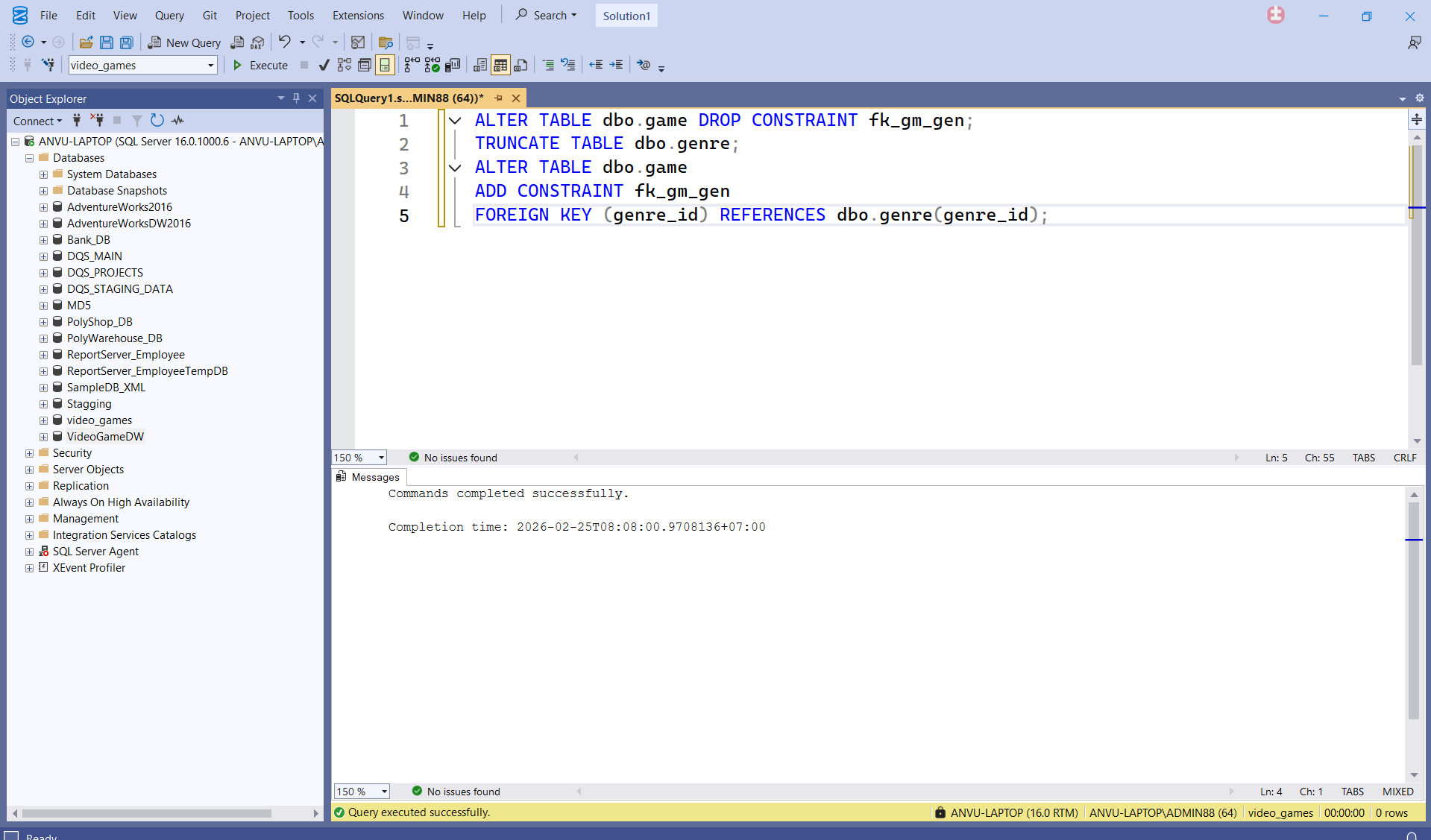
Data Warehouse được thiết kế theo mô hình Star Schema gồm:  
- Fact\_Accident  
- Dim\_Date  
- Dim\_Location  
- Dim\_Weather  
- Dim\_SpeedLimit  
  
DB video\_games

DB VideoGameDW

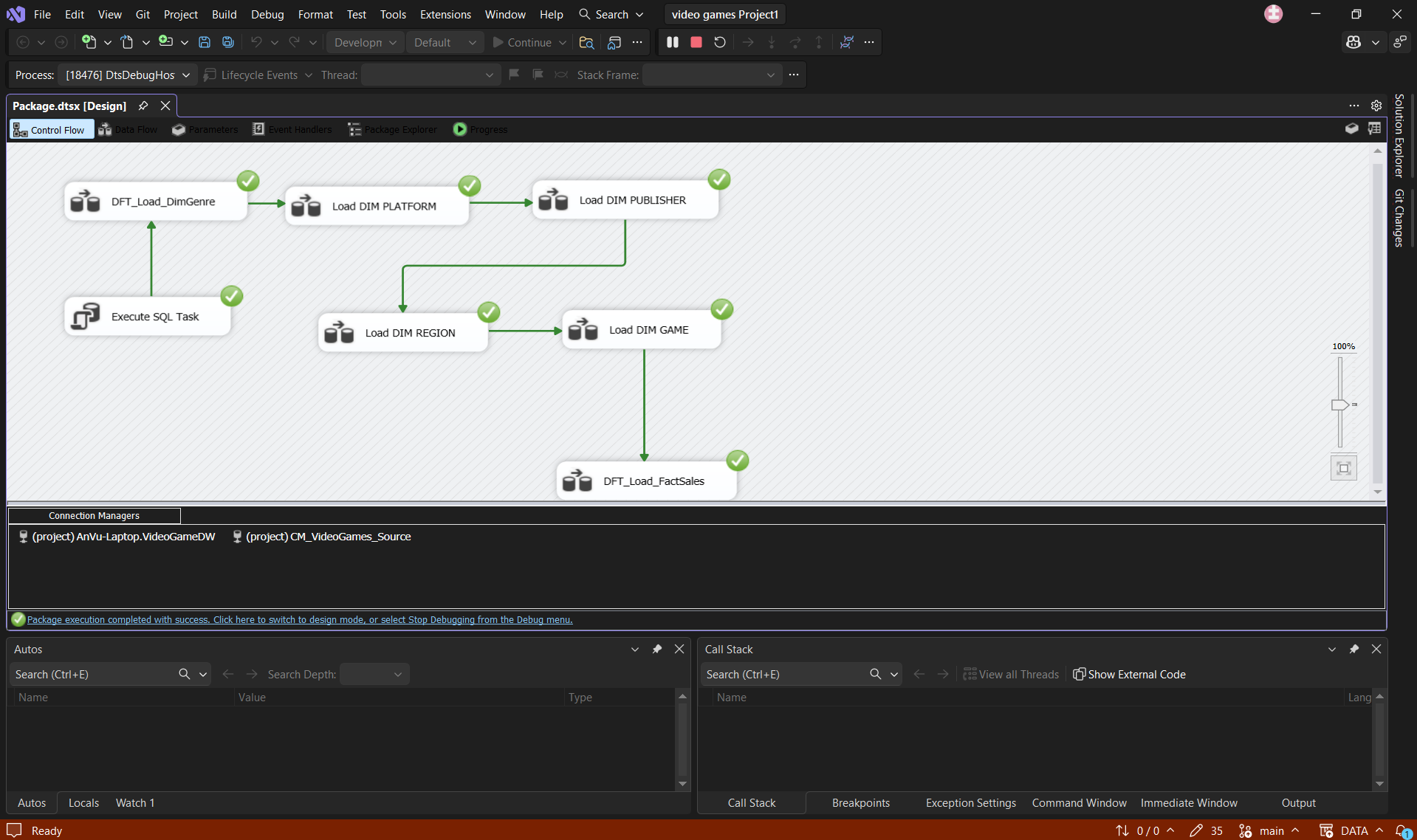
# 6. ETL

Extract: Lấy dữ liệu từ Database  
Transform: Làm sạch, chuẩn hóa dữ liệu, xử lý missing values  
Load: Nạp dữ liệu vào các bảng Dimension và Fact

* Truncate Table trong SQL



* Kết quả



# 7. CÔNG NGHỆ SỬ DỤNG

- SQL (sql server)

- Visual Studio

- Power BI

# 8. PHÂN TÍCH & DASHBOARD

Sinh viên xây dựng dashboard thể hiện:  
- Số vụ tai nạn theo bang  
- Tai nạn theo thời gian  
- Severity theo điều kiện thời tiết

# 9. ĐÁNH GIÁ KẾT QUẢ

Hệ thống Data Warehouse hỗ trợ tốt cho việc phân tích dữ liệu tai nạn và đưa ra insight có giá trị.

# 10. HẠN CHẾ & HƯỚNG PHÁT TRIỂN

Hạn chế: Dữ liệu thiếu một số thông tin chi tiết.  
Hướng phát triển: Bổ sung dữ liệu mới và mô hình dự báo.

# TÀI LIỆU THAM KHẢO

Kaggle US Accidents Dataset  
GitHub: DATA-WAREHOUSE-ACCIDENT-US-2016-2021