**ĐẠI HỌC QUỐC GIA TP.HCM**

**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**

**KHOA HỆ THỐNG THÔNG TIN**

🙞🙜🕮🙞🙜

****

**ĐỒ ÁN MÔN HỌC**

**KHO DỮ LIỆU VÀ OLAP**

Đề Tài:

**Earthquake Data of Tukey in 1910 - 2017**

* Các thành viên:
* Trần Minh Sang - 16521023
* Huỳnh Văn Tín - 16521827

TP.HCM, Tháng năm 2019



ĐẠI HỌC QUỐC GIA TP.HCM

TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN

KHOA HỆ THỐNG THÔNG TIN

🙞🙜🕮🙞🙜

****

ĐỒ ÁN MÔN HỌC

KHO DỮ LIỆU VÀ OLAP

Đề Tài:

Earthquake Data of Tukey in 1910 - 2017

* Các thành viên:
* Trần Minh Sang - 16521023
* Huỳnh Văn Tín - 16521827

Giảng viên hướng dẫn:

Cô Nguyễn Thị Kim Phụng

TP.HCM, Tháng năm 2019

**NHẬN XÉT CỦA GIẢNG VIÊN**

………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………

**LỜI CẢM ƠN**

Quá trình học tập và rèn luyện tại khoa Hệ thống thông tin - Trường Đại học Công nghệ Thông tin, chúng em đã được trang bị các kiến thức cơ bản, các kỹ năng thực tế để có thể áp dụng vào thực hiện các Project, các bài tập lớn. Chúng em xin gửi lời cảm ơn chân thành đến quý Thầy Cô khoa Hệ thống thông tin đã truyền đạt cho chúng em những kiến thức bổ ích ấy để chúng em có thể hoàn thành tốt đồ án môn học của mình.

Đặc biệt, chúng em xin gửi lời cảm ơn sâu sắc nhất tới Cô Nguyễn Thị Kim Phụng, là người trực tiếp hướng dẫn, truyền đạt kiến thức môn Kho dữ liệu và OLAP để chúng em có thể hoàn thành tốt nhất đồ án môn học này.

Trong quá trình thực hiện đồ án không khỏi mắc phải những sai sót, kính mong nhận được sự góp ý của thầy và các bạn để nhóm có thể hoàn thiện đồ án của mình hơn nữa, đó cũng là kinh nghiệm quý giá để mỗi người trong nhóm vận dụng vào những đồ án sau này và công việc trong tương lai. Xin chân thành cảm ơn!

Thay mặt nhóm thực hiện

Trần Minh Sang

**Mục lục**

[Chương 1: GIỚI THIỆU CHUNG VỀ ĐỀ TÀI 4](#_Toc21377122)

[1. Giới thiệu đề tài: Earthquake Data of Tukey in 1910 – 2017 4](#_Toc21377123)

[2. Tiền xử lý dữ liệu 4](#_Toc21377124)

[2.1 Dữ liệu thô chưa qua xử lý 4](#_Toc21377125)

[2.2 Dữ liệu đã qua xử lý 4](#_Toc21377126)

[Chương 2: XÂY DỰNG KHO DỮ LIỆU (SSIS) 5](#_Toc21377127)

[1. Tạo kho dữ liệu 5](#_Toc21377128)

[1.1 Thiết lập data source 5](#_Toc21377129)

[1.2 Thiết lập data destination 8](#_Toc21377130)

[1.3 Thiết lập data source và data destination hoàn chỉnh 11](#_Toc21377131)

[1.4 Thiết lập Execute SQL Task 12](#_Toc21377132)

[1.5 Run project 13](#_Toc21377133)

[1.6 Kiểm tra dữ liệu 15](#_Toc21377134)

[1.7 Kết quả sau khi tạo kho dữ liệu 16](#_Toc21377135)

[2. Danh sách các bảng dữ liệu 16](#_Toc21377136)

[2.1 Table areaDim – 5161 records 16](#_Toc21377137)

[2.2 Table cityDim – 81 records 16](#_Toc21377138)

[2.3 Table directionDim – 8 records 17](#_Toc21377139)

[2.4 Table latlongDim – 9260 records 17](#_Toc21377140)

[2.5 Table FACT – 10062 records 17](#_Toc21377141)

[2.6 Table timeDim 18](#_Toc21377142)

# **Chương 1: GIỚI THIỆU CHUNG VỀ ĐỀ TÀI**

## **1. Giới thiệu đề tài: Earthquake Data of Tukey in 1910 – 2017**

Dữ liệu thống kê về các trận động đất ở Thổ Nhĩ Kì được ghi lại ở các vĩ độ từ 25 - 50; kinh độ 15 - 60. Vì các trạm đo sáng được đặt ở Thổ Nhĩ Kỳ, hầu hết các trận động đất được ghi nhận đều ở vĩ độ từ 35 - 45; kinh độ 25 - 45. Bộ lọc thời gian tìm kiếm cơ sở dữ liệu được đặt thành ngày 27/09/1910 đến 27/09/2017. Vì có quá nhiều trận động đất có cường độ nhỏ hơn 4.0, bộ lọc cường độ được đặt thành 3,5 đến 9.0 (không có trận động đất nào được ghi nhận lớn hơn 9.0 cường độ).

Dữ liệu cho chúng ta một cái nhìn trực quan về động đất ở Thổ Nhĩ Kì từ 1910 – 2017. Từ đó, chúng ta có thể tìm ra quy luật xảy ra động đất ở đây và có cách ứng phó phù hợp. Dữ liệu cũng cung cấp nguồn tài nguyên quan trọng cho những nhà nghiên cứu về động đất nói riêng và những nhà khoa học nói chung.

Nguồn dữ liệu: <https://www.kaggle.com/caganseval/earthquake>

## **2. Tiền xử lý dữ liệu**

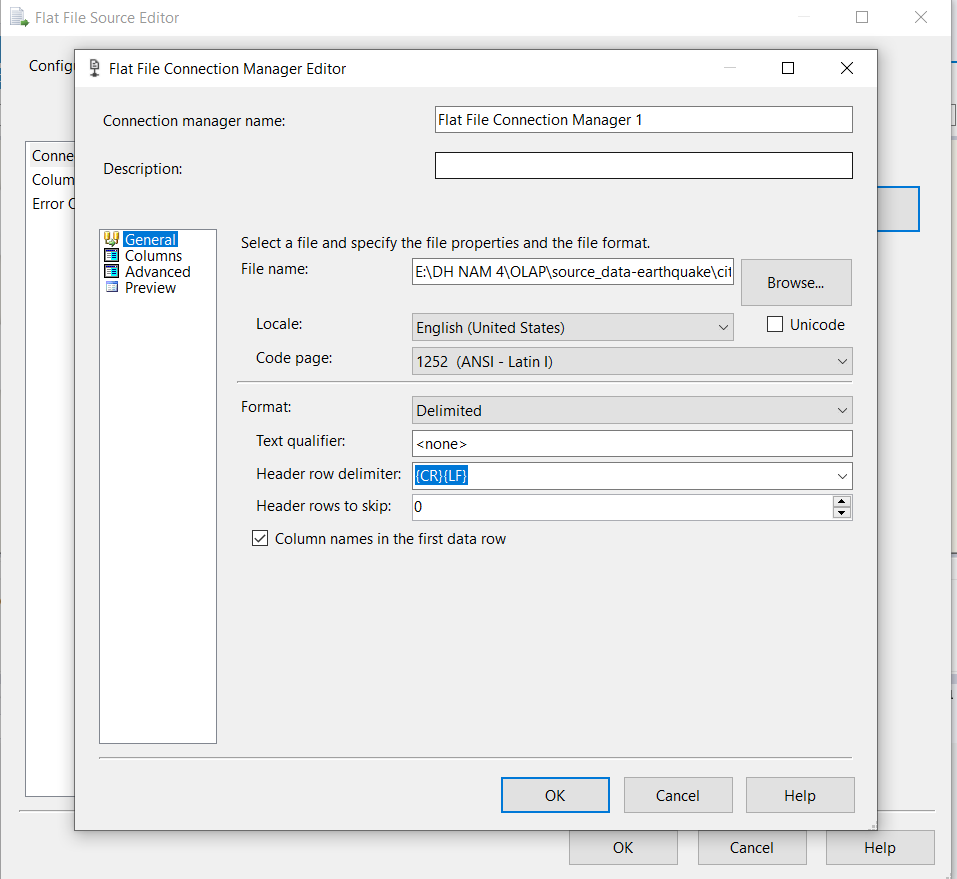
### **2.1 Dữ liệu thô chưa qua xử lý**

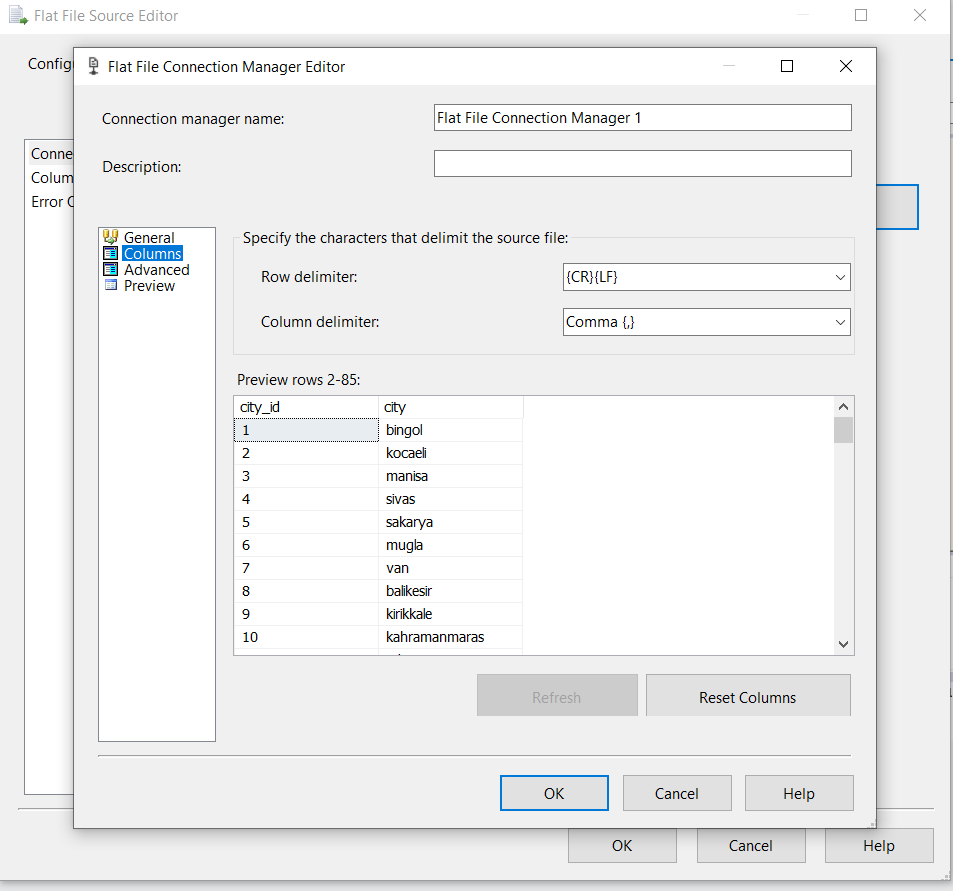
### **2.2 Dữ liệu đã qua xử lý**

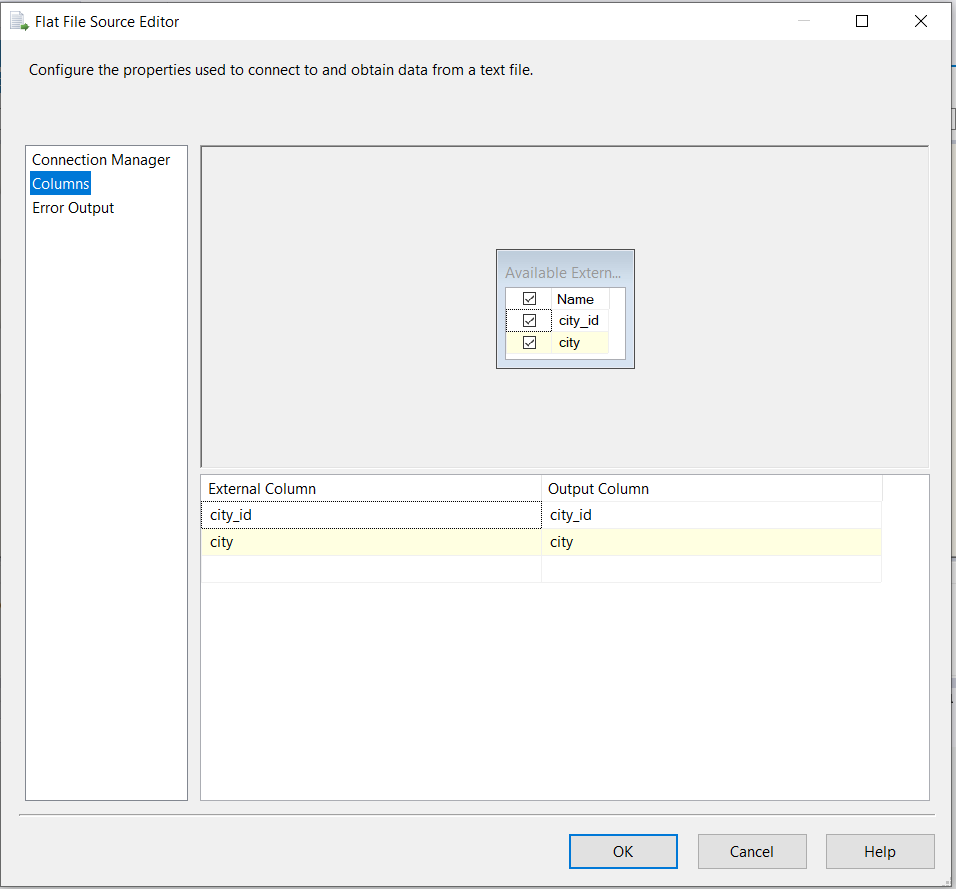
# **Chương 2: XÂY DỰNG KHO DỮ LIỆU (SSIS)**

## **1. Tạo kho dữ liệu**

### **1.1 Thiết lập data source**

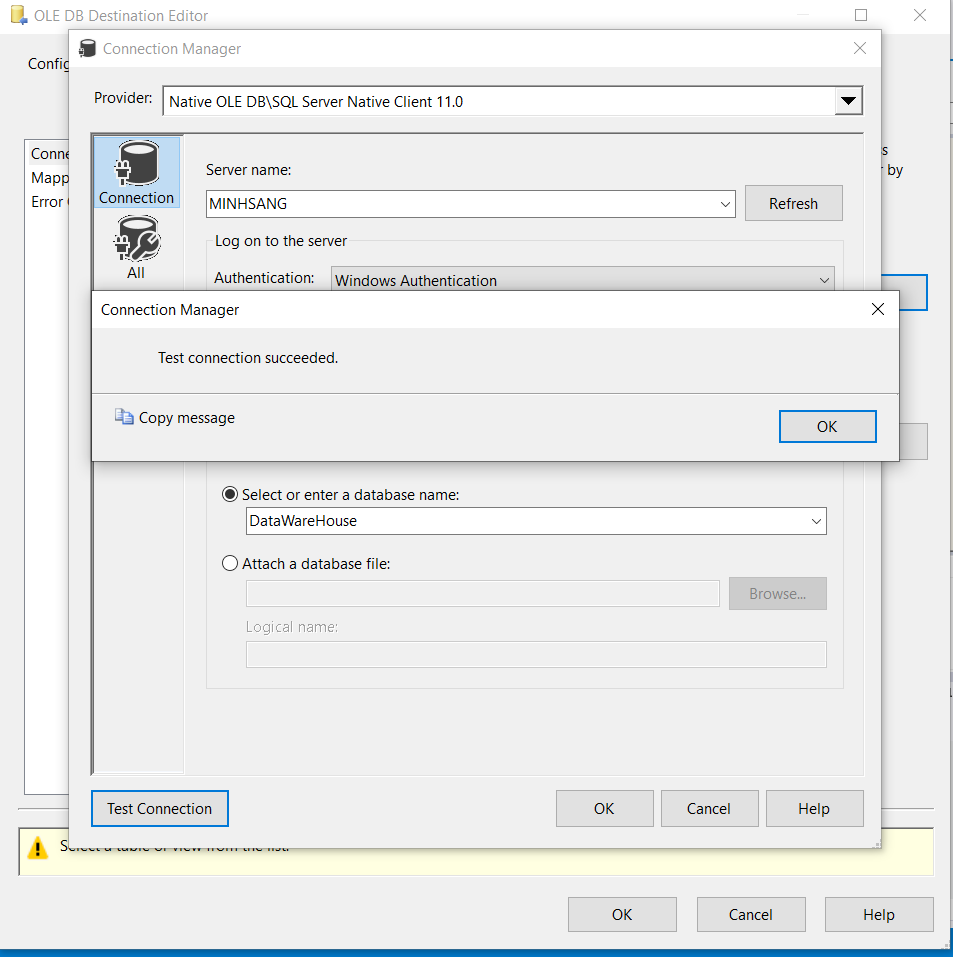




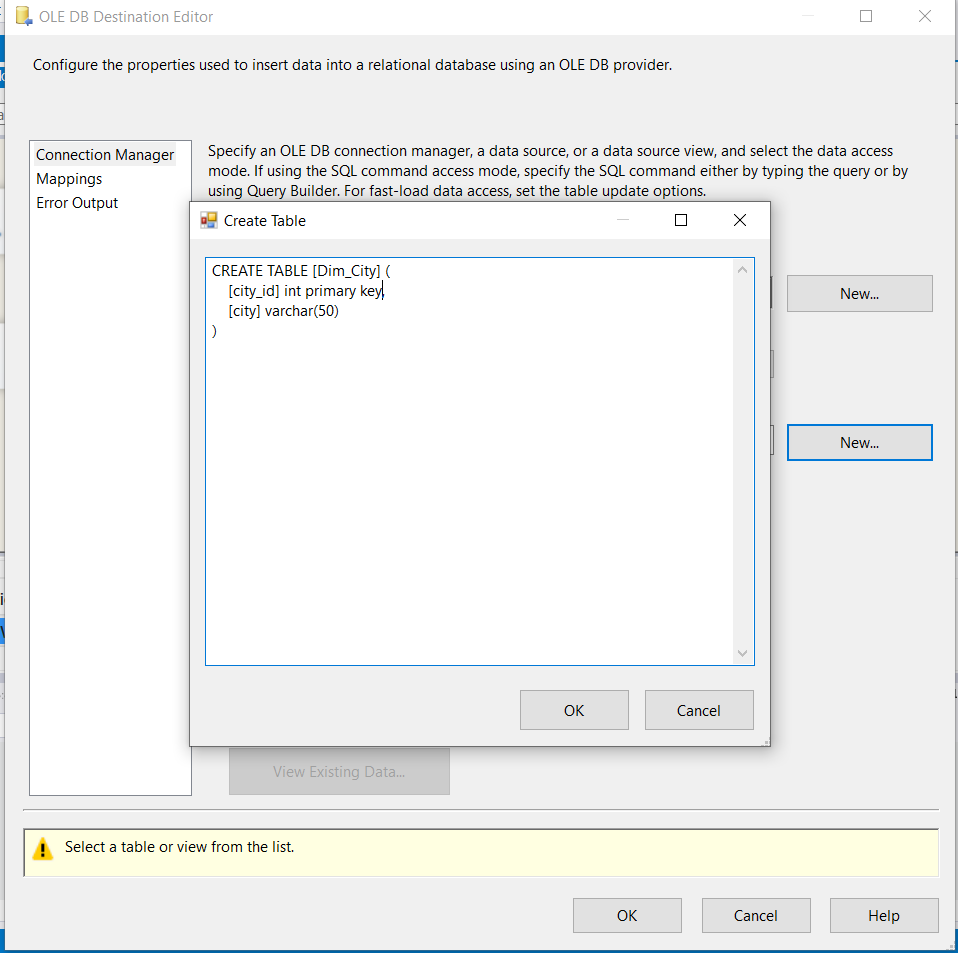


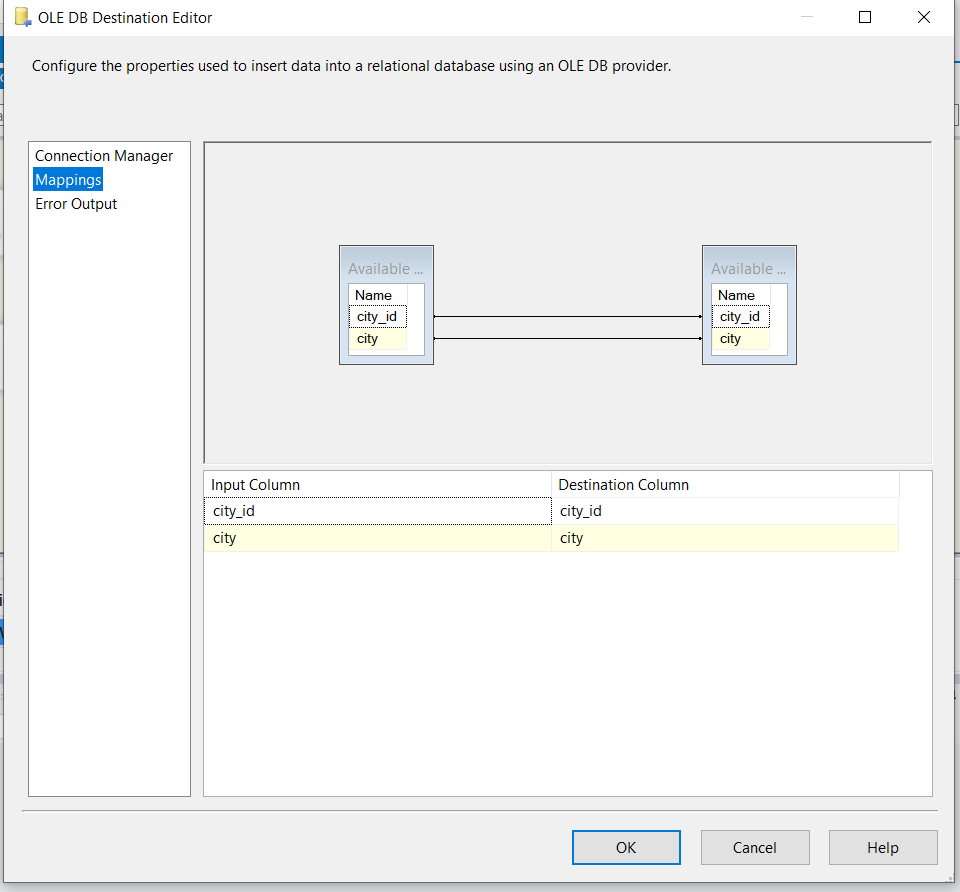
### **1.2 Thiết lập data destination**

- Tạo data connection

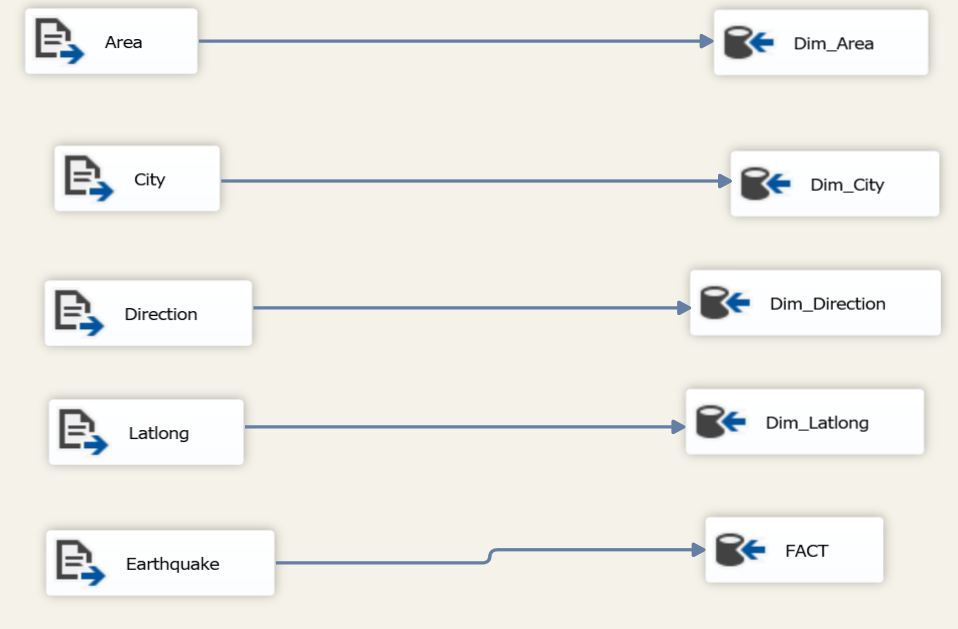


- Tạo bảng

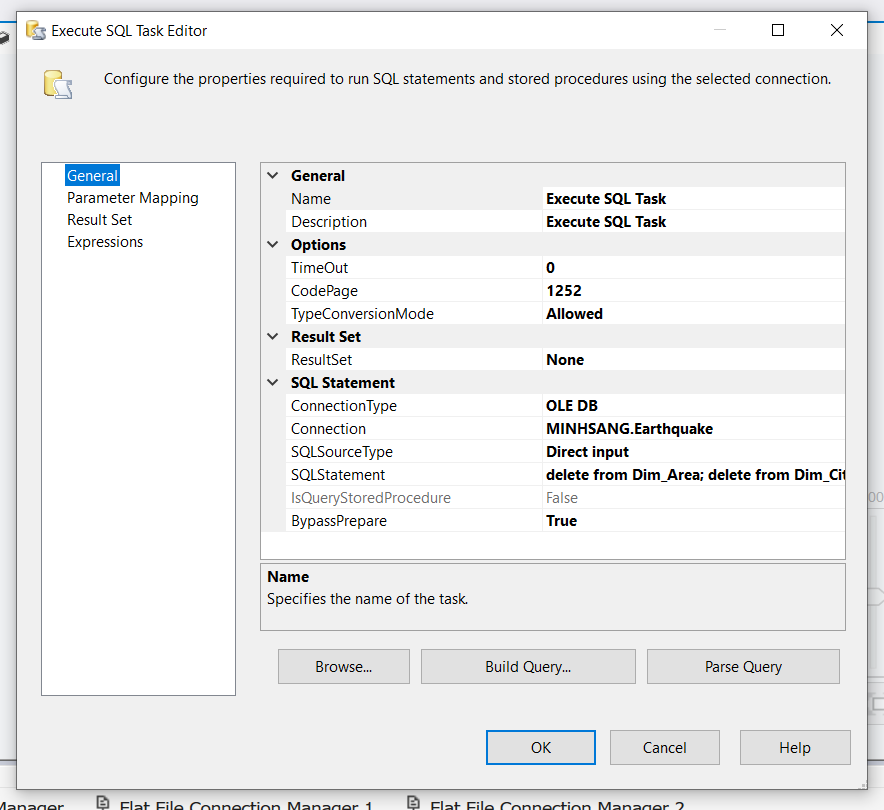




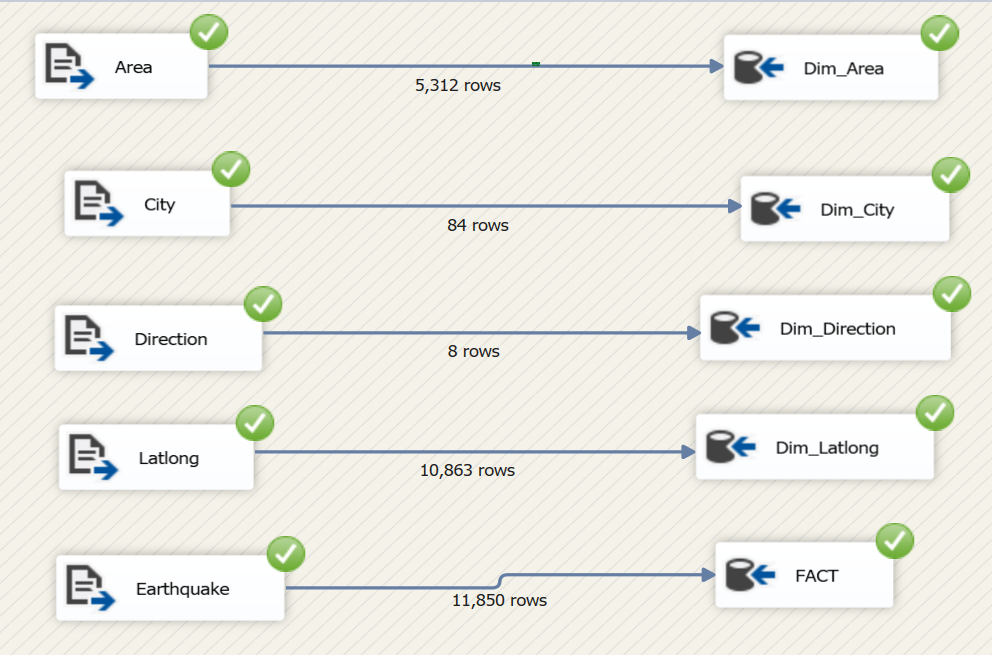
### **1.3 Thiết lập data source và data destination hoàn chỉnh**

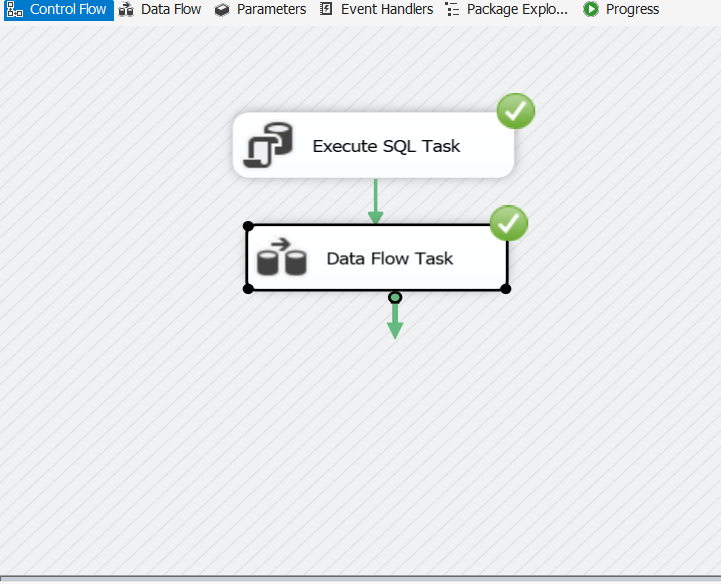


### **1.4 Thiết lập Execute SQL Task**

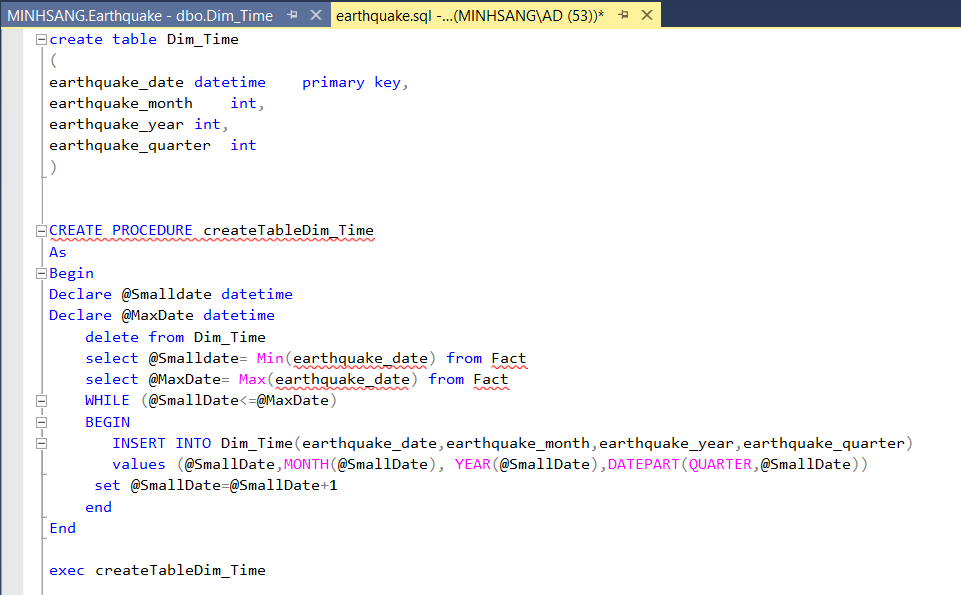


### **1.5 Run project**

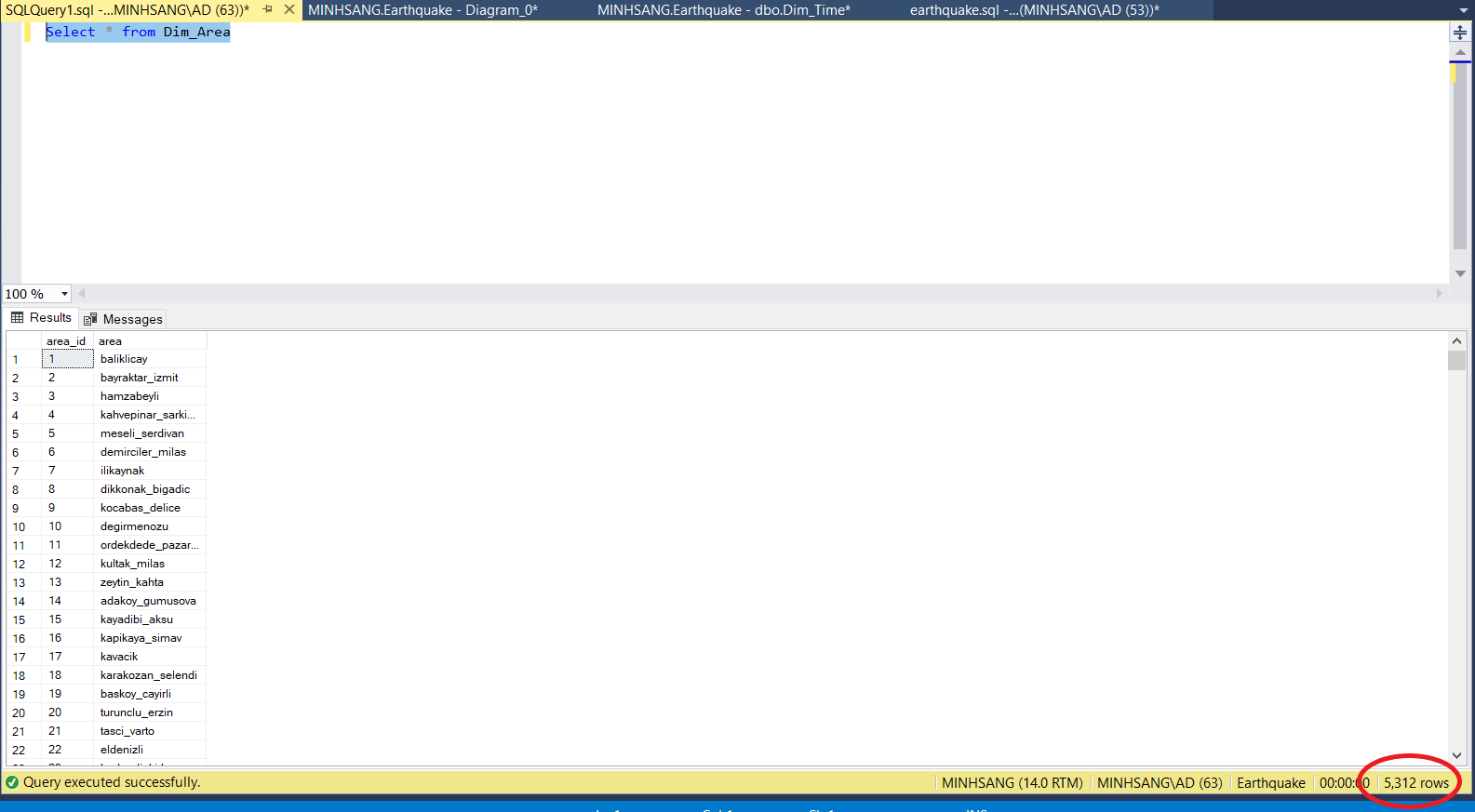


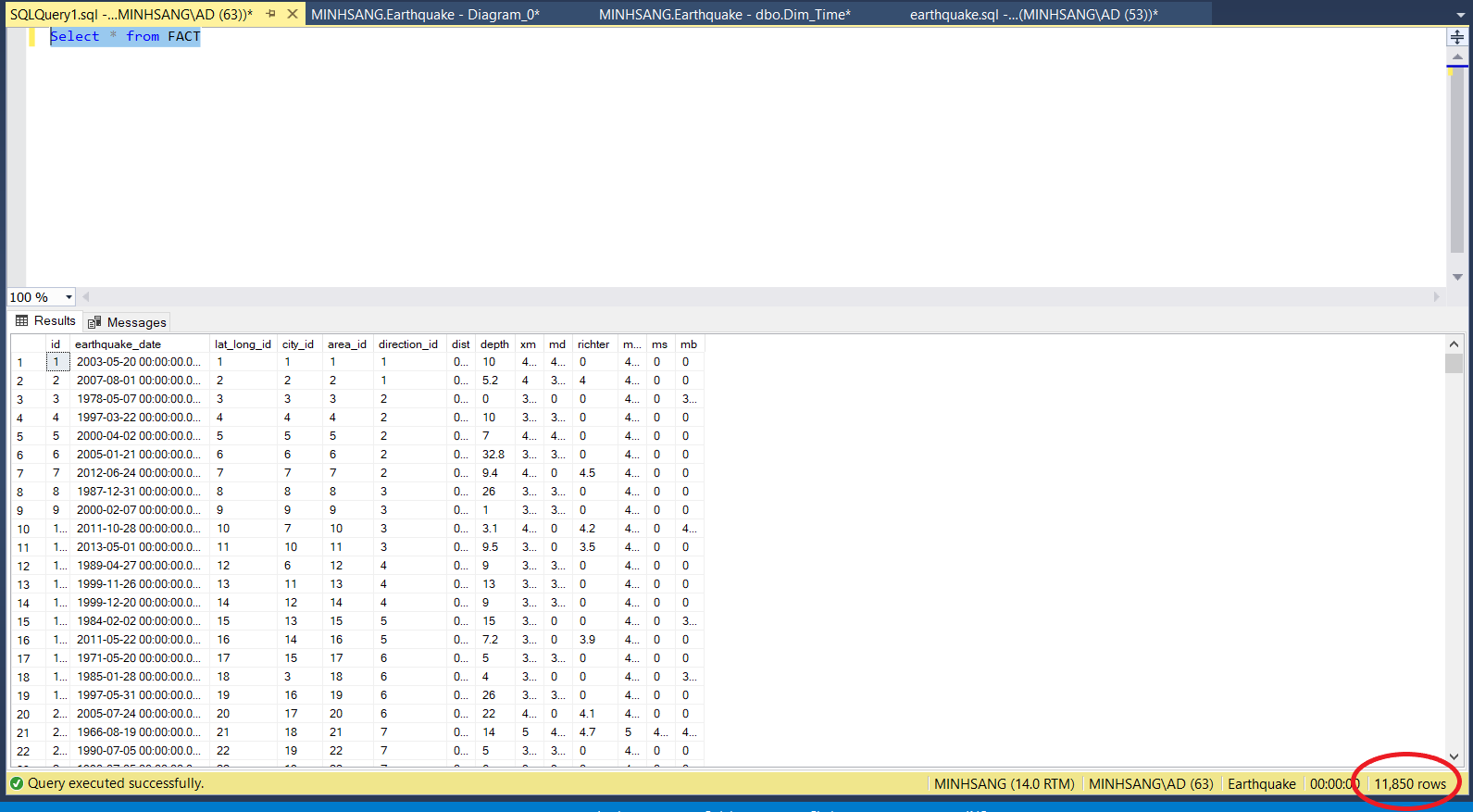


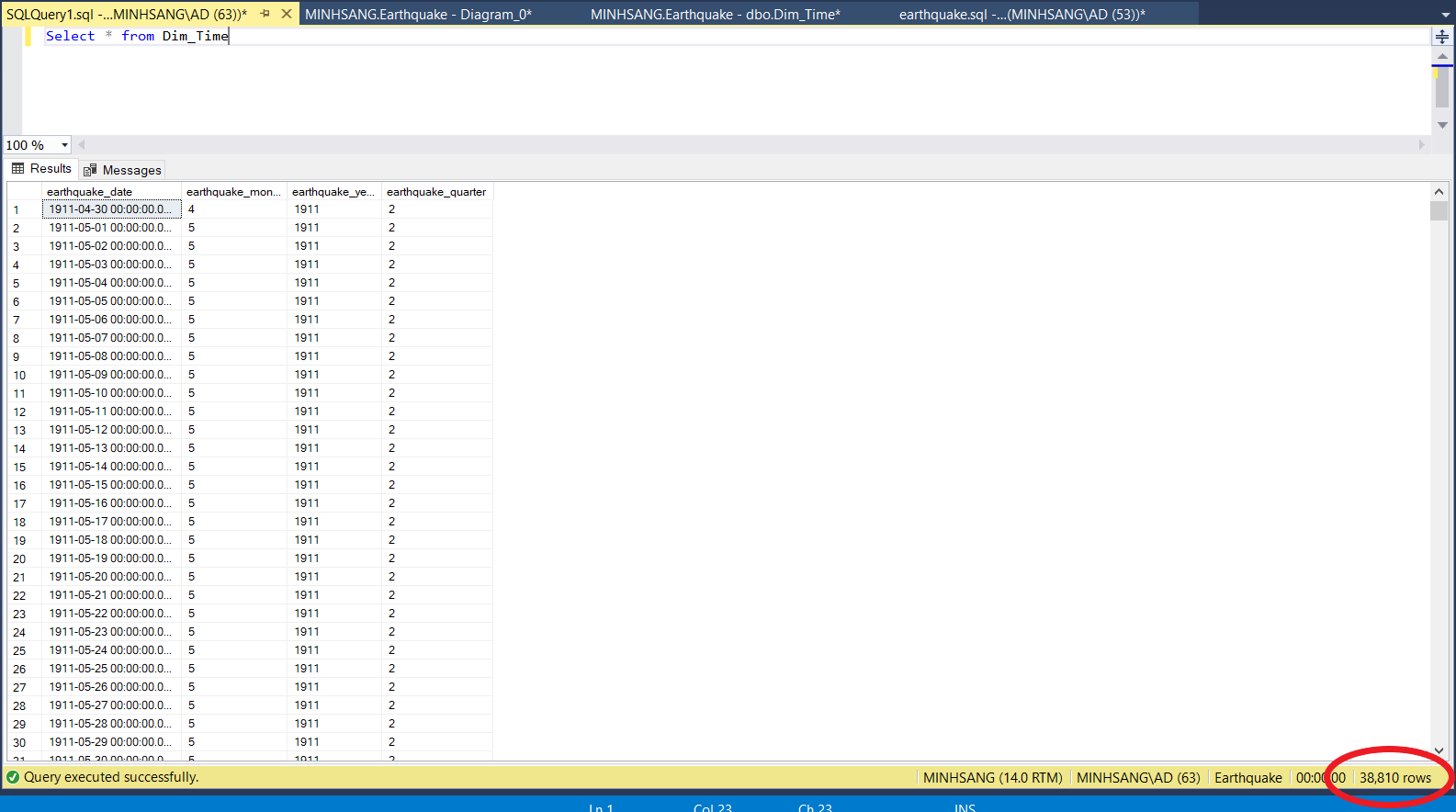
**1.6 Tạo bảng Dim\_Time**



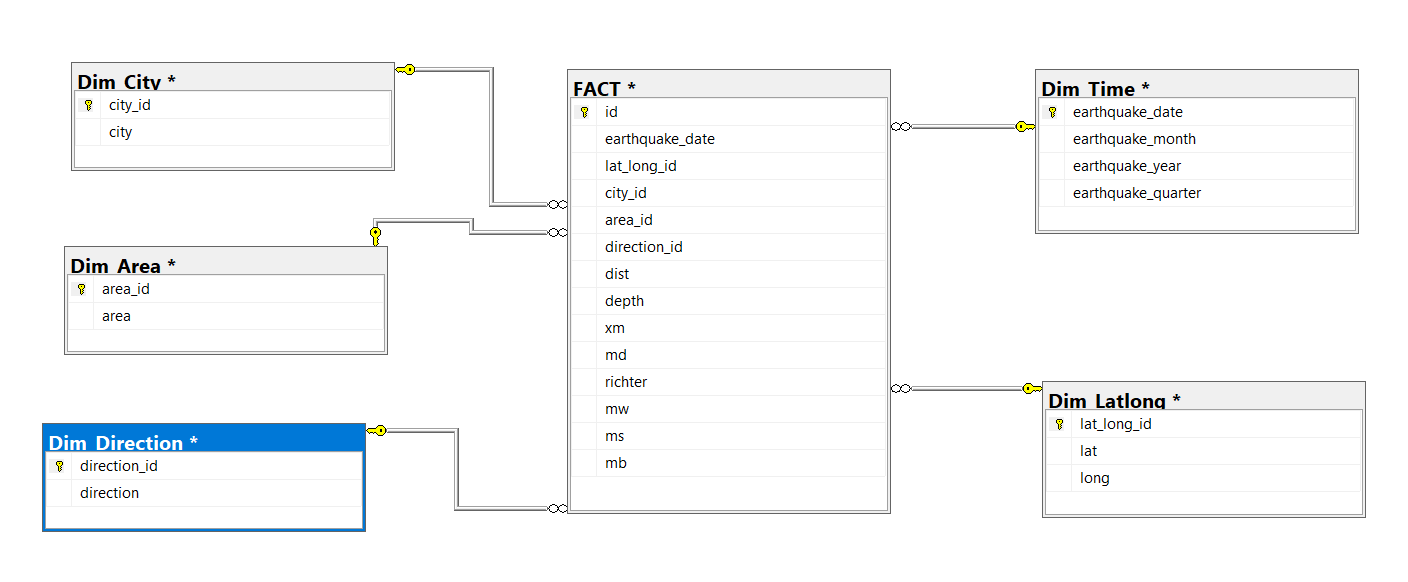
### **1.6 Kiểm tra dữ liệu**







### **1.7 Kết quả sau khi tạo kho dữ liệu**



## **2. Danh sách các bảng dữ liệu**

### **2.1 Table Dim\_Area – 5312 records**

|  |  |  |  |
| --- | --- | --- | --- |
| STT | Tên thuộc tính | Kiểu dữ liệu | Ý nghĩa |
| 1 | area\_id | int | Mã khu vực |
| 2 | are | varchar(50) | Tên khu vực |

### **2.2 Table Dim\_City – 84 records**

|  |  |  |  |
| --- | --- | --- | --- |
| STT | Tên thuộc tính | Kiểu dữ liệu | Ý nghĩa |
| 1 | city\_id | int | Mã thành phố |
| 2 | city | varchar(50) | Tên thành phố |

### **2.3 Table Dim\_Direction – 8 records**

|  |  |  |  |
| --- | --- | --- | --- |
| STT | Tên thuộc tính | Kiểu dữ liệu | Ý nghĩa |
| 1 | direction\_id | int | Mã hướng |
| 2 | direction | varchar(50) | Tên hướng |

### **2.4 Table Dim\_Latlong – 10863 records**

|  |  |  |  |
| --- | --- | --- | --- |
| STT | Tên thuộc tính | Kiểu dữ liệu | Ý nghĩa |
| 1 | lat\_long\_id | int | Mã kinh độ vĩ độ |
| 2 | lat | varchar(50) | Vĩ độ |
| 3 | long | varchar(50) | Kinh độ |

### **2.5 Table Dim\_Time – 38810 records**

|  |  |  |  |
| --- | --- | --- | --- |
| STT | Tên thuộc tính | Kiểu dữ liệu | Ý nghĩa |
| 1 | earthquake\_date | datetime | Ngày |
| 2 | earthquake\_month | int | Tháng |
| 3 | earthquake\_year | int | Năm |
| 4 | earthquake\_quarter | int | Quý |

### **2.6 Table FACT – 11850 records**

|  |  |  |  |
| --- | --- | --- | --- |
| STT | Tên thuộc tính | Kiểu dữ liệu | Ý nghĩa |
| 1 | id | int | Mã trận động đất |
| 2 | earthquake\_date | datetime | Ngày |
| 4 | lat\_long\_id | int | Mã kinh độ vĩ độ |
| 5 | city\_id | int | Mã thành phố |
| 6 | area\_id | int | Mã vùng |
| 7 | direction\_id | int | Mã hướng |
| 8 | dist | float | Bề rộng |
| 9 | depth | float | Bề sâu |
| 10 | xm | float | Thông số lớn nhất của trận động đất |
| 11 | md | float | Thời lượng |
| 12 | richter | float | Độ richter |
| 13 | mw | float | Cường độ mô men |
| 14 | ms | float | Cường độ sóng bề mặt |
| 15 | mb | float | Cường độ sóng cơ thể |