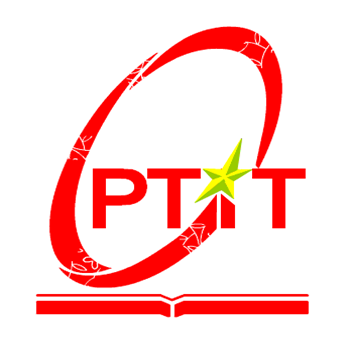
**HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG**

**KHOA CÔNG NGHỆ THÔNG TIN**

-----🙞🙜🕮🙞🙜-----



**KHO DỮ LIỆU VÀ KHAI PHÁ DỮ LIỆU**

**BÁO CÁO KHO DỮ LIỆU**

**Giảng viên : Nguyễn Quỳnh Chi**

**Nhóm lớp học: 04**

**Nhóm BTL : 07**

**Thành viên :**

Lê Tiến Thành - B20DCCN644

**Hà Nội - 2024**

**LỜI NÓI ĐẦU**

Lời đầu tiên, chúng em xin gửi lời cảm ơn đến Học viện Công nghệ Bưu chính Viễn Thông đã tạo điều kiện cho chúng em được học môn Kho dữ liệu và khai phá dữ liệu. Đặc biệt, chúng em xin gửi lời cảm ơn chân thành và sâu sắc nhất tới cô Nguyễn Quỳnh Chi, giảng viên bộ môn đã hướng dẫn và truyền đạt những kiến thức hết sức bổ ích và quý báu trong suốt thời gian học tập vừa qua. Với vốn hiểu biết sâu rộng và kinh nghiệm nhiều năm giảng dạy cũng như làm việc trong môi trường công nghệ thông tin, cô khiến chúng em thật sự ấn tượng, khâm phục và “ngỡ ngàng” trước những hiểu biết của cô.

Kho dữ liệu và khai phá dữ liệu là một môn học thật sự rất hay và bổ ích tuy nhiên đây là môn có khối lượng kiến thức tương đối nhiều và khó có thể hiểu rõ, hiểu sâu nhanh chóng khi thời lượng học trên lớp có hạn. Mặc dù cô đã truyền đạt nhiệt tình và tận tâm nhưng do năng lực có hạn, khả năng tư duy và khả năng tiếp thu không nhanh nên trong báo cáo bài tập lớn này chúng em khó có thể tránh khỏi được những sai sót. Do đó, chúng em kính mong cô xem xét và bổ sung giúp đỡ chúng hoàn thiện bài báo cáo này một cách đầy đủ và đúng đắn hơn. Một lần nữa, chúng em xin chân thành cảm ơn cô!

**BẢNG PHÂN CÔNG**

|  |  |  |  |
| --- | --- | --- | --- |
| STT | Họ và Tên | Mã sinh viên | Công việc |
| 1 |  |  | - Sinh dữ liệu cho các nguồn dữ liệu  - Tạo phân cấp để tính khối cho OLAP  - Tạo metadata  - Viết báo cáo |
| 2 |  |  | -Sinh dữ liệu cho các nguồn dữ liệu  - Cách ánh xạ để đổ dữ liệu từ CSDL tích hợp vào kho  - Tạo phân cấp để tính khối cho OLAP  - Cài đặt khối dữ liệu  - Viết báo cáo |
| 3 |  |  | - Thiết kế các khối dữ liệu để thực hiện OLAP  - Tạo metadata  - Slide |
| 4 | Lê Tiến Thành |  | -Tích hợp đúng dữ liệu từ hai nguồn thành một nguồn  - Thiết kế mô hình dữ liệu được tích hợp  -Thiết kế mô hình dữ liệu cho kho  - Thiết kế các khối dữ liệu để thực hiện OLAP  - Tính toán các khối cho OLAP, và lưu trữ  - Viết báo cáo |
| 5 |  |  | -Tích hợp đúng dữ liệu từ hai nguồn thành một nguồn  - Thiết kế mô hình dữ liệu được tích hợp  - Thiết kế mô hình dữ liệu cho kho  - Viết báo cáo |
| 6 |  |  | - Viết ứng dụng để đưa OLAP lên Web thực hiện các thao tác khoan sâu xuống, lên, chiếu, chọn  - Viết báo cáo |

**MỤC LỤC**

[**I. GIỚI THIỆU 4**](#_ip55tgtubvpt)

[1. Mục tiêu bài tập lớn 4](#_om7uwhivqnnu)

[2. Phạm vi bài tập lớn 4](#_40glpwr4yhb4)

[**II. PHÂN TÍCH YÊU CẦU 4**](#_ru2q5ug299gl)

[1. Yêu cầu nghiệp vụ 4](#_gtts9isdj8o1)

[2. Đặc tả chức năng 5](#_q8yoyhwm0zho)

[**III. THIẾT KẾ KHO DỮ LIỆU 6**](#_1ss96zh31gqo)

[1. Tích hợp đúng dữ liệu từ hai nguồn thành một nguồn (đổi sang mô hình thực thể liên kết rồi mới tích hợp mô hình ER) 6](#_d6y4ob6j55xb)

[1.1. Chuyển đổi ngược từ các bảng CSDL quan hệ sang mô hình thực thể liên kết 6](#_4oxns4odsm52)

[1.2. Tích hợp 2 mô hình thực thể liên kết mở rộng 9](#_gp8g7uqcezbe)

[2. Thiết kế mô hình dữ liệu cho kho 11](#_wm08e4wa12ve)

[2.1. Sale\_Fact 12](#_l3bkdd3uhqdk)

[2.2. Time\_Dim 12](#_qgc1c07zygzn)

[2.3. Customer\_Dim 13](#_jrsmw521qxz)

[2.4. Item\_Dim 13](#_ndxkkumddrdh)

[2.5. Store\_Dim 14](#_ueg9j5ej01o0)

[2.6. Chiều thứ cấp Location\_Dim 14](#_vodadhxf2okh)

[2.7. Kho 14](#_trw2yfarp1be)

[3. Tạo phân cấp 15](#_9f0thavkaerk)

[3.1. Phân cấp cho chiều Thời gian 15](#_enf9n7c73b80)

[3.2. Phân cấp cho chiều Location 15](#_th6akm5t1lpe)

[4. Sinh dữ liệu cho IDB 16](#_w6sy3dwh3fdl)

[4.1. Code sinh dữ liệu 16](#_o3o3wc254bms)

[4.2. Kết quả sau khi sinh dữ liệu 17](#_inq6rssa9ixo)

[5. Cách ánh xạ dữ liệu từ IDB sang kho 19](#_ygx5wrqe3y4c)

[5.1. Ánh xạ dữ liệu vào các bảng Dimension 19](#_eaygg688mm7t)

[5.2. Ánh xạ dữ liệu vào bảng Sale\_Fact: 21](#_cta6lop4vxot)

[**IV. CÀI ĐẶT , TÍNH TOÁN VÀ LƯU TRỮ CÁC KHỐI DỮ LIỆU 22**](#_qm7pt49x3w7c)

[1. Cài đặt 22](#_lt5ybuybol3)

[2. Tính toán và lưu trữ khối dữ liệu 27](#_1jm1267wuq1o)

[**V .METADATA 32**](#_rupzlwat690x)

[**VI. XÂY DỰNG ỨNG DỤNG BÁO CÁO PHÂN TÍCH TRỰC TUYẾN 36**](#_dmfmtadqf01s)

[**VII. KIỂM TRA TÍNH ĐÚNG ĐẮN 37**](#_q7goffkupjiu)

[**VIII. KẾT LUẬN 38**](#_lqhmlecgt919)

## 

## GIỚI THIỆU

### Mục tiêu bài tập lớn

* Phát triển kỹ năng thực hành của sinh viên trong việc xây dựng và quản lý cơ sở dữ liệu, bao gồm thiết kế, triển khai, và tối ưu hóa.
* Đặt ra các vấn đề thực tế trong việc lưu trữ và truy xuất dữ liệu, và yêu cầu sinh viên tìm ra giải pháp hiệu quả.
* Tạo được một kho dữ liệu nhiều chiều lưu trữ thông tin của toàn bộ hệ thống cửa hàng
* Tận dụng nguồn dữ liệu này để sử dụng cho những mục đích hỗ trợ cho công việc kinh doanh
* Tạo thuận lợi tối đa khi xem xét dữ liệu với mục đích phân tích kinh doanh và ra quyết định.

### Phạm vi bài tập lớn

Phạm vi xây dựng kho dữ liệu là quản lý tổng số lượng mỗi mặt hàng và đơn đặt hàng được lưu trữ tại mỗi tỉnh, thành phố với thời gian cụ thể.

* Quản lý việc kinh doanh của một doanh nghiệp gồm nhiều cửa hàng nằm rải rác ở nhiều thành phố và bang khác nhau nhằm đáp ứng được nhu cầu của khách hàng.
* Nguồn dữ liệu: từ hệ thống cơ sở dữ liệu quan hệ hiện tại của doanh nghiệp.
* Sử dụng MS SQL Server và Oracle phù hợp với công ty khách hàng
* Áp dụng kỹ thuật OLAP (Online Analytical Processing) để xây dựng khối dữ liệu phân tích về khách hàng, mặt hàng và các yếu tố khác của doanh nghiệp theo nhiều chiều.

## II. PHÂN TÍCH YÊU CẦU

### 1. Yêu cầu nghiệp vụ

Để đáp ứng được nhu cầu của khách hàng, hệ thống kho dữ liệu trích lọc dữ liệu từ hai cơ sở dữ liệu hiện có để cho vào kho dữ liệu và cung cấp các xử lý phân tích trực tuyến với các thao tác cuộn lên, khoan xuống, chọn và chiếu dựa trên những yêu cầu chọn lựa của khách hàng và các bảng theo chiều để đáp ứng yêu cầu của người sử dụng. Thiết lập một chiều thời gian, hệ thống sẽ sinh một báo cáo Phân tích trực tuyến cho những yêu cầu sau:

1. Tìm tất cả các cửa hàng cùng với thành phố, bang, số điện thoại, mô tả, kích cỡ, trọng lượng và đơn giá của tất cả các mặt hàng được bán ở kho đó.
2. Tìm tất cả các đơn đặt hàng với tên khách hàng và ngày đặt hàng được thực hiện bởi khách hàng đó
3. Tìm tất cả các cửa hàng cùng với tên thành phố và số điện thoại mà có bán các mặt hàng được đặt bởi một khách hàng nào đó
4. Tìm địa chỉ văn phòng đại diện với tên thành phố, bang của tất cả các cửa hàng lưu kho một mặt hàng nào đó với số lượng trên mức cụ thể.
5. Với mỗi một đơn đặt hàng của khách, liệt kê các mặt hàng được đặt cùng với mô tả, mã cửa hàng, tên thành phố và các cửa hàng có bán mặt hàng đó.
6. Tìm thành phố và bang mà một khách hàng nào đó sinh sống
7. Tìm mức độ tồn kho của một mặt hàng cụ thể tại tất cả các cửa hàng ở một thành phố cụ thể nào đó
8. Tìm các mặt hàng, số lượng đặt, khách hàng, cửa hàng và thành phố của một đơn đặt hàng.
9. Tìm các khách hàng du lịch, khách hàng đặt theo đường bưu điện và khách hàng thuộc cả hai loại

### 2. Đặc tả chức năng

1. **Đặc tả đầu vào**

- Lược đồ quan hệ của hệ thống cơ sở dữ liệu quan hệ hiện tại của doanh nghiệp này như sau:

* Cơ sở dữ liệu Văn phòng đại diện bao gồm những quan hệ với các lược đồ như sau:

Khách hàng (Mã KH, Tên KH, Mã Thành phố, Ngày đặt hàng đầu tiên)

Khách hàng du lịch (\*Mã KH, Hướng dẫn viên du lịch, Thời gian)

Khách hàng bưu điện (\* Mã KH, Địa chỉ bưu điện, Thời gian)

* Cơ sở dữ liệu Bán hàng gồm những quan hệ với lược đồ như sau:

Văn phòng đại diện (Mã Thành phố, Tên Thành phố, Địa chỉ VP, Bang, Thời gian)

Cửa hàng (Mã cửa hàng, \* Mã Thành phố, Số điện thoại, Thời gian)Mặt hàng (Mã MH, Mô tả, Kích cỡ, Trọng lượng, Giá, Thời gian)

Mặt hàng được lưu trữ (\* Mã cửa hàng, \* Mã mặt hàng, Số lượng trong kho, Thời gian)

Đơn đặt hàng (Mã đơn, Ngày đặt hàng, Mã Khách hàng)

Mặt hàng được đặt (\* Mã đơn, \* Mã mặt hàng, Số lượng đặt, Giá đặt, Thời gian)

1. **Đặc tả đầu ra**

- Các bảng dữ liệu động để hỗ trợ cho người sử dụng ra quyết định thực hiện các phép chiếu chọn, các thao tác cơ bản như roll up, drill down, slice, dice...

- Kho dữ liệu được chuyển hóa, phân ra thành những mục tiêu như khách hàng, bán hàng, sản phẩm... nhờ công cụ OLAP( Online Analyze Processing) để hỗ trợ việc truy vấn.

- Bản báo cáo thống kê theo khách hàng, cửa hàng, khách du lịch…

## III. THIẾT KẾ KHO DỮ LIỆU

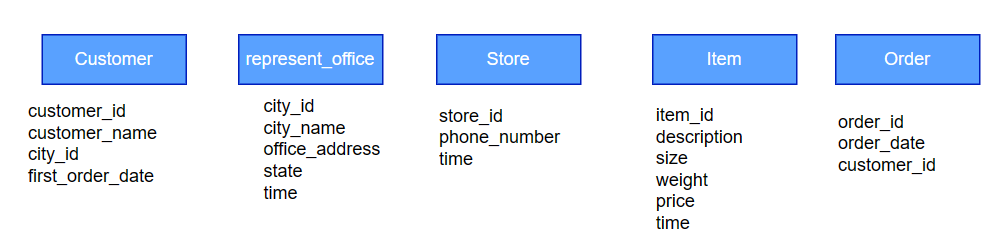
### Tích hợp đúng dữ liệu từ hai nguồn thành một nguồn (đổi sang mô hình thực thể liên kết rồi mới tích hợp mô hình ER)

#### **1.1. Chuyển đổi ngược từ các bảng CSDL quan hệ sang mô hình thực thể liên kết**

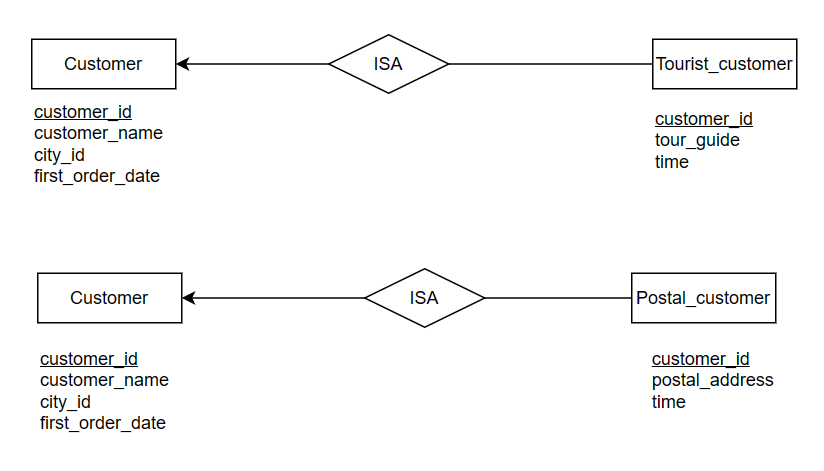
1. **Bảng phân loại các quan hệ và các thuộc tính theo định nghĩa**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Relation Name | RelType | Primary Key | KAP | KAG | FKA | NKA |
| customer | PR1 | customer\_id |  |  |  | customer\_name, city\_id, first\_order\_date |
| tourist\_customer | PR2 | customer\_id | customer\_id |  |  | tour\_guide, time |
| postal\_customer | PR2 | customer\_id | customer\_id |  |  | postal\_address, time |
| represent\_office | PR1 | city\_id |  |  |  | city\_name, office\_address, state, time |
| store | PR1 | store\_id |  |  | city\_id | phone\_number, time |
| item | PR1 | item\_id |  |  |  | description, size, weight, price, time |
| stored\_item | SR1 | store\_id item\_id | store\_id item\_id |  |  | stored\_quantity, time |
| order | PR1 | order\_id |  |  |  | order\_date  customer\_id |
| ordered\_item | SR1 | order\_id  item\_id | order\_id  item\_id |  |  | order\_quantity, order\_price, time |

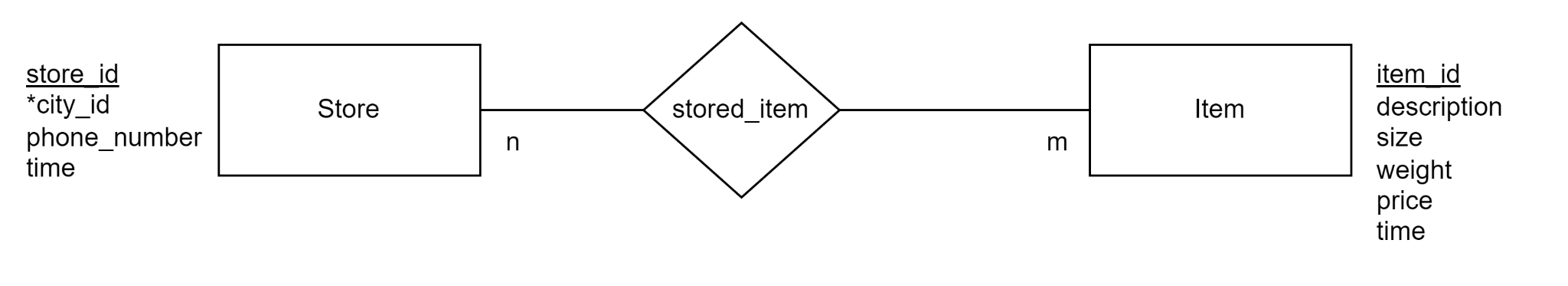
1. **Ánh xạ PR1 thành thực thể**

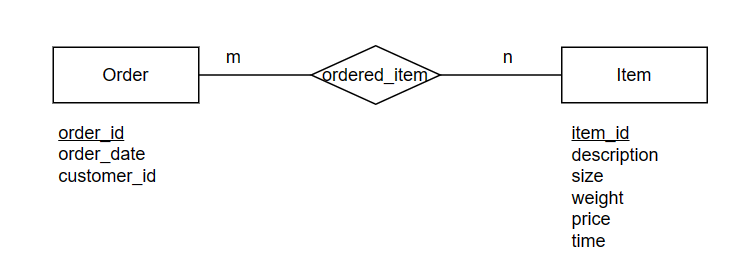


1. **Ánh xạ PR2 sang thực thể yếu**

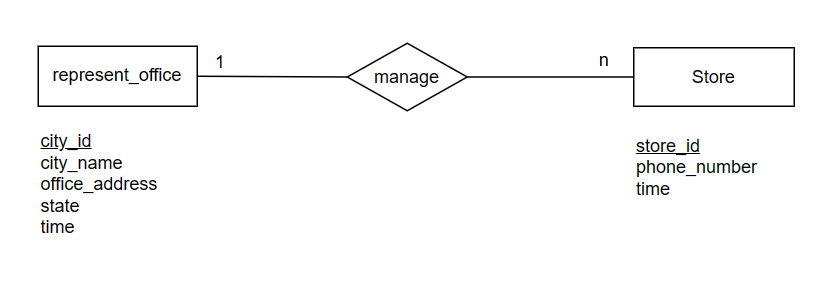


1. **Ánh xạ SR1 thành quan hệ hai ngôi hoặc nhiều ngôi**

* 



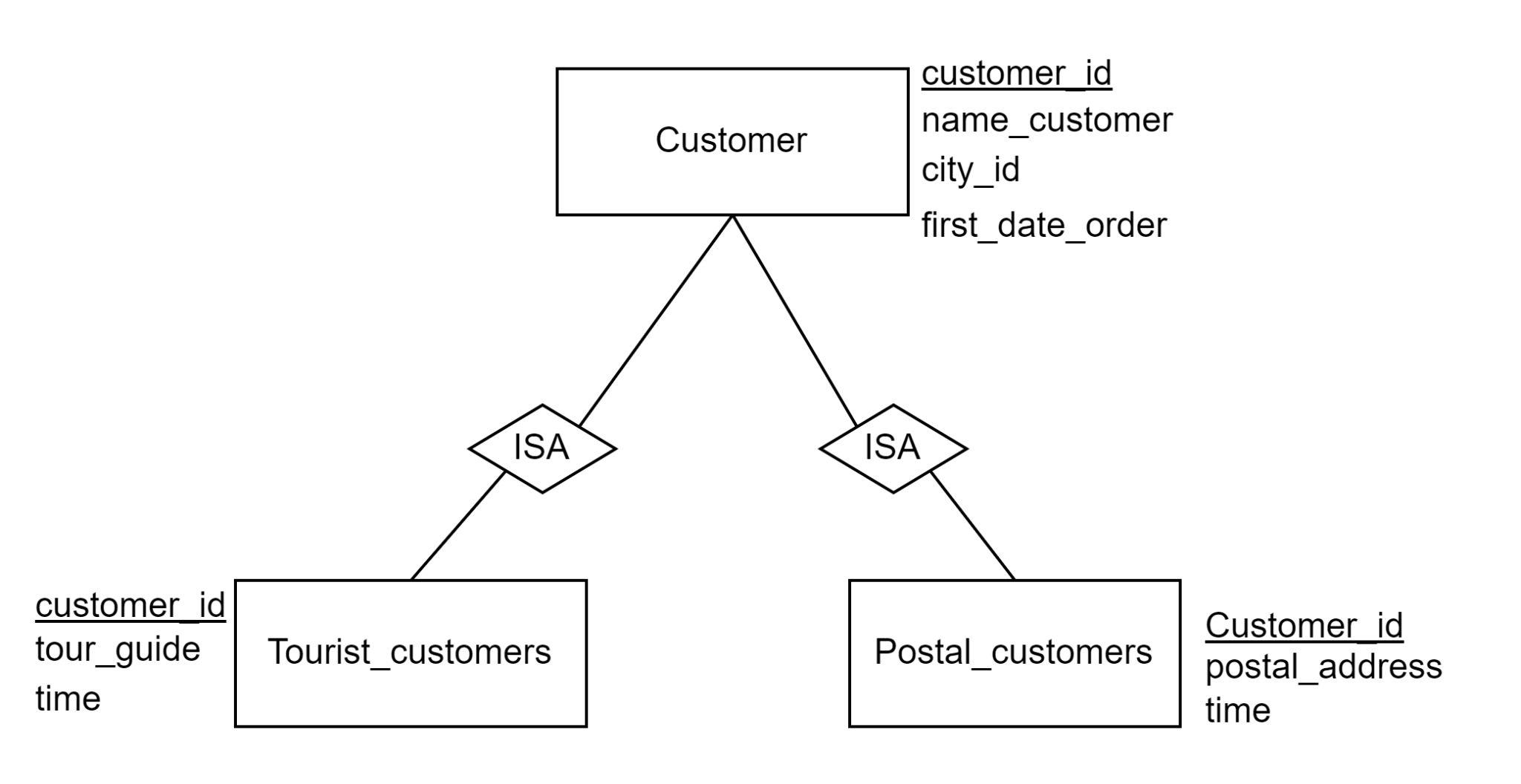
1. **Ánh xạ FKA thành một quan hệ**



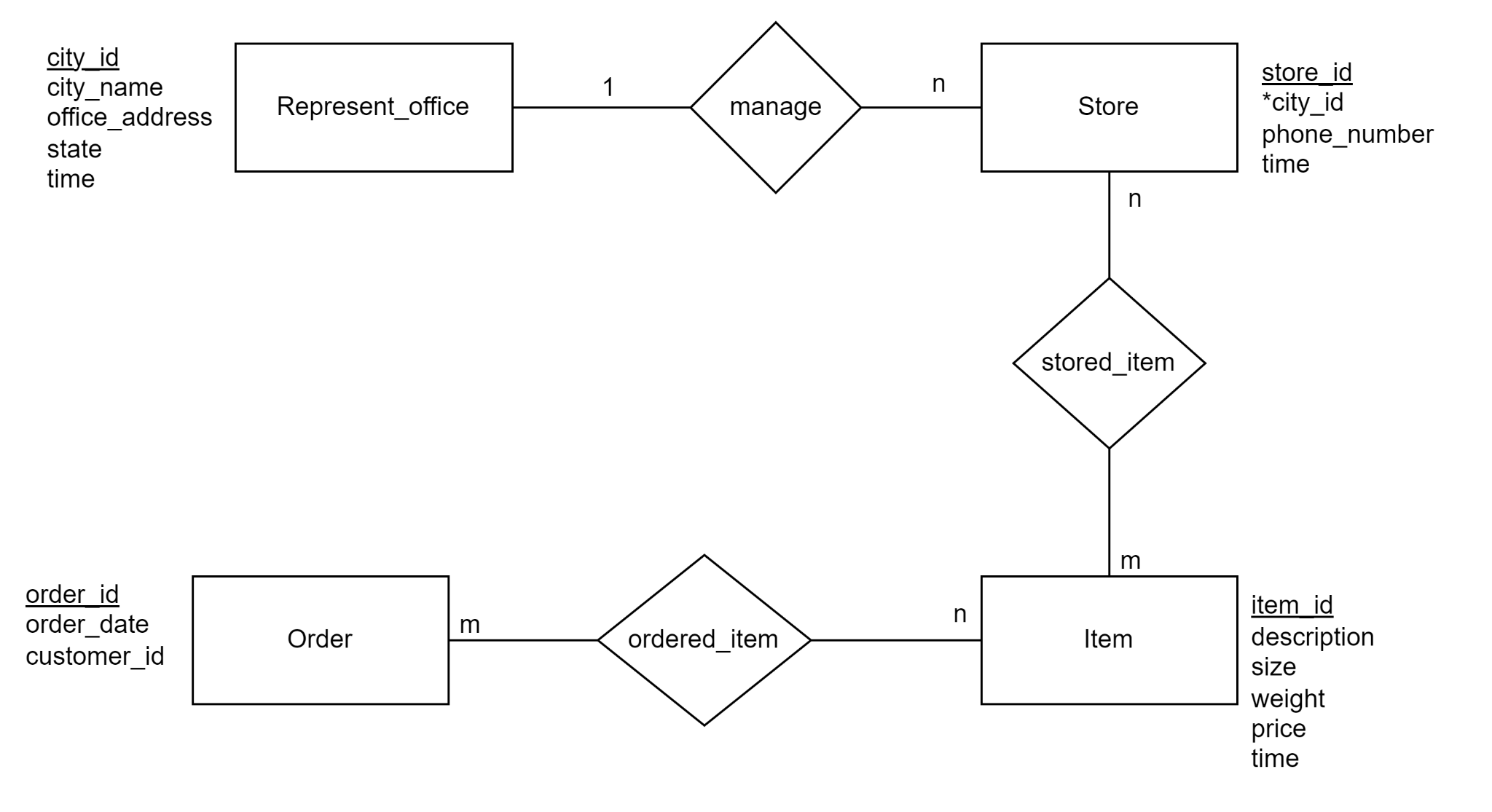
1. **Ánh xạ phụ thuộc bao hàm sang ngữ nghĩa**

|  |  |
| --- | --- |
| Các phụ thuộc bao hàm phát sinh | Ngữ nghĩa sinh ra |
| Store.city\_idRepresent\_office.city\_id | Quan hệ một nhiều giữa thực thể Represent\_office và Store |
| Stored\_item.store\_id  Store.store\_id  Stored\_item.item\_id  Item.item\_id | Quan hệ nhiều nhiều giữa thực thể Store và Item |
| Order\_item.item\_id  Item.item\_id  Order\_item.order\_id  Item.order\_id | Quan hệ nhiều nhiều giữa thực thể Order và Item |

1. **EER Văn phòng đại diện**



1. **EER Bán hàng**



#### **1.2. Tích hợp 2 mô hình thực thể liên kết mở rộng**

1. **Giải quyết các xung đột giữa các lược đồ**

TH1: Giải quyết xung đột về đồng âm và đồng nghĩa

* Thuộc tính “time” của 2 thực thể tourist\_customer và postal\_customer đồng nghĩa với thuộc tính “first\_order\_date” của thực thể Customer

=> Thực hiện loại bỏ thuộc tính “time” của thực thể tourist\_customer và postal\_customer

* Thuộc tính ‘time” của các thực thể represent\_office, store, item, stored\_item, ordered\_item là từ đồng âm có ý nghĩa khác nhau

=> Thực hiện đổi tên:

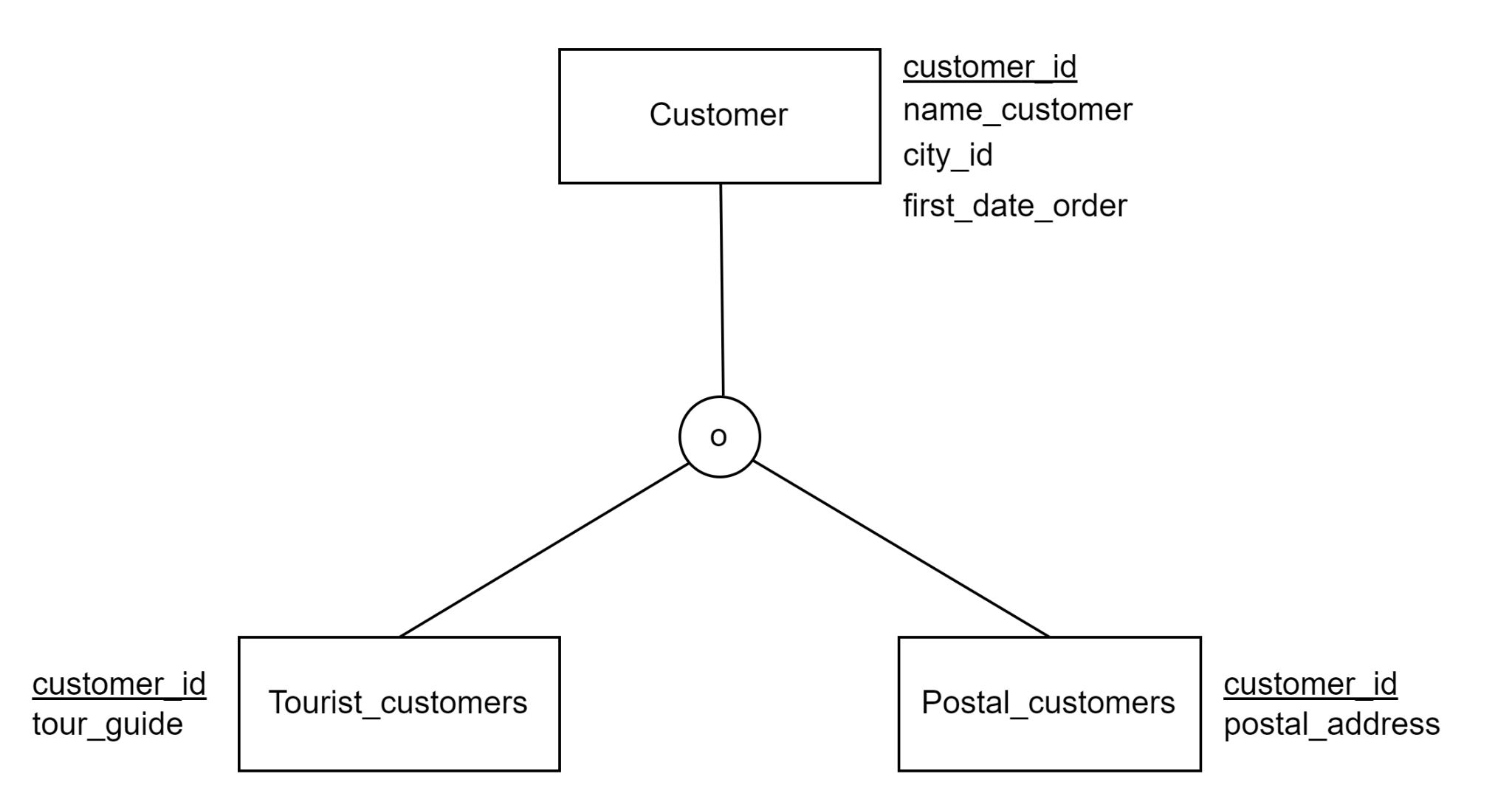
|  |  |  |
| --- | --- | --- |
| Tên thực thể | Tên thuộc tính ban đầu | Tên thuộc tính sau khi đổi |
| Represent\_office | time | office\_created\_time |
| Store | time | store\_created\_time |
| Item | time | item\_created\_time |
| Stored\_item | time | stored\_item\_time |

* Thuộc tính “time” của 2 thực thể ordered\_item và Order là đồng nghĩa, đều mang nghĩa ngày đặt hàng

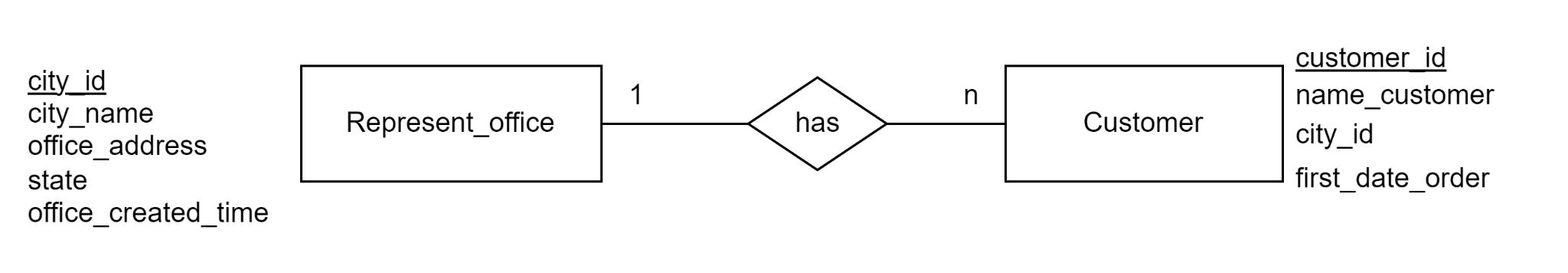
=> Thực hiện loại bỏ thuộc tính “time” của thực thể ordered\_item

1. **Trộn các thực thể**

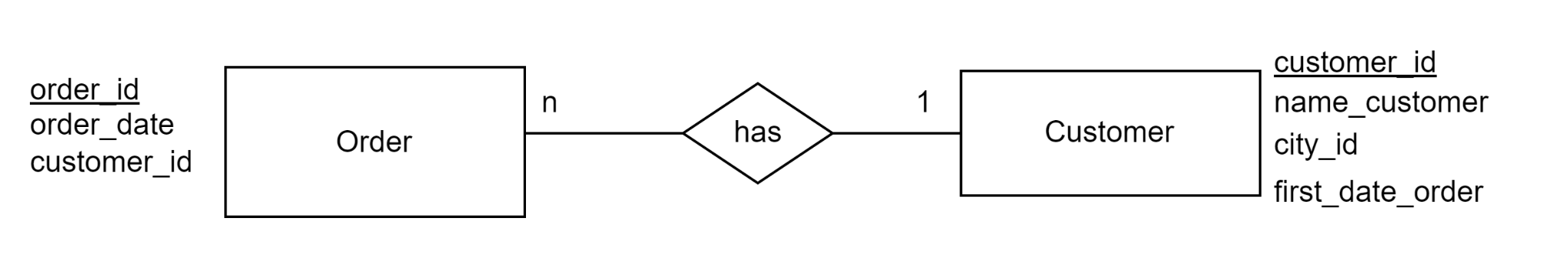
* Customer, Tourist\_customer, Postal\_customer đều có khóa chính là Customer\_id nên trộn 3 thực thể này:



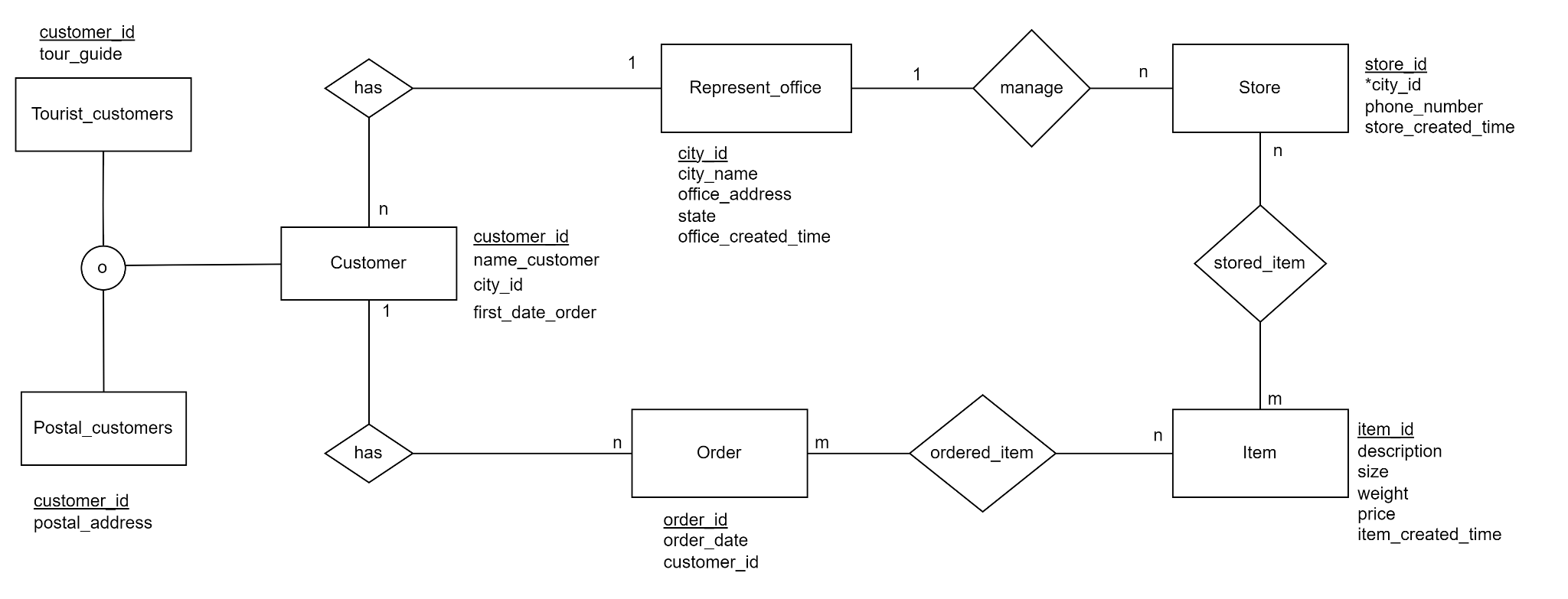
* Thực thể Customer có thuộc tính “city\_id” và thực thể represent\_office có khóa chính là “city\_id” nên thực hiện trộn 2 thực thể Represent\_office và Customer



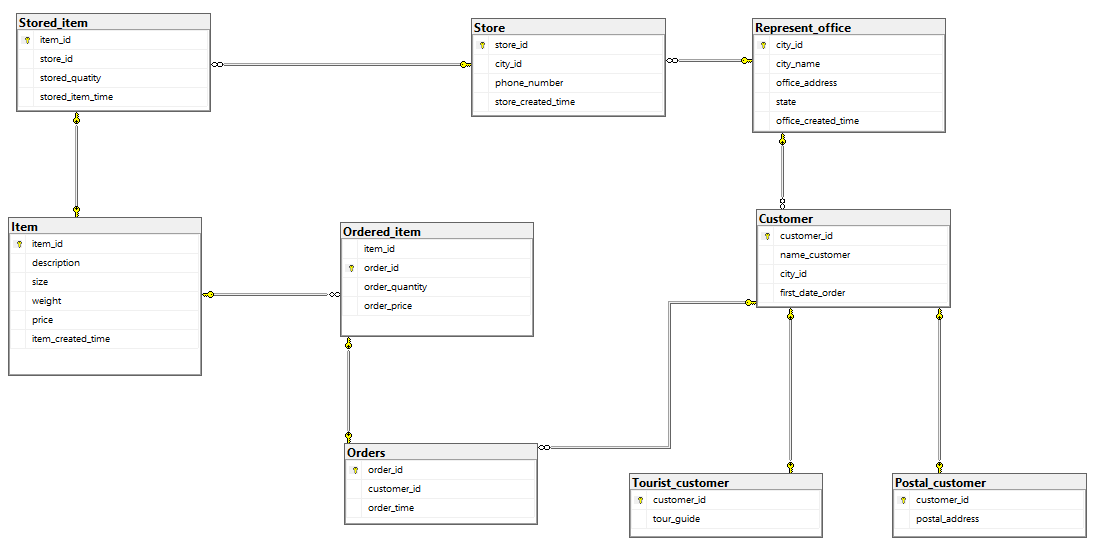
* Thực thể Customer có khóa chính là “customer\_id” và thực thể Order có thuộc tính “customer\_id” nên thực hiện trộn 2 thực thể Order và Customer



1. **Tích hợp 2 EER thành IEER:**



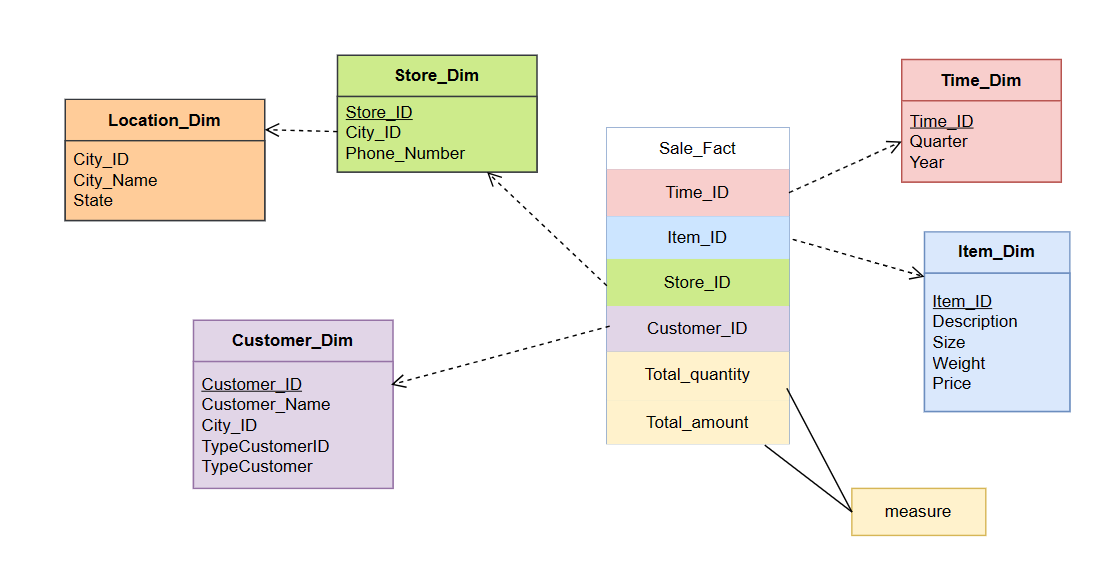
1. **Lược đồ cơ sở dữ liệu tích hợp**



### 2. Thiết kế mô hình dữ liệu cho kho

Kho được thiết kế theo mô hình bông tuyết gồm:

* 1 bảng fact : Sale\_Fact
* 4 Chiều chính : Customer\_Dim, Item\_Dim, Store\_Dim, Time\_Dim
* 1 Chiều thứ cấp: Location\_Dim



#### **2.1. Sale\_Fact**

* **Các độ đo trong bảng:**
* Total\_quantity: Tổng số lượng hàng mà khách hàng Customer\_ID đã mua sản phẩm Item\_ID tại cửa hàng Store\_ID trong một quý.
* Total\_amount: Tổng tiền hàng mà khách hàng Customer\_ID đã bỏ ra để amua sản phẩm Item\_ID tại cửa hàng Store\_ID trong một quý.

|  |  |  |
| --- | --- | --- |
| Thuộc tính | Mô tả | Ví dụ |
| Time\_ID | mã thời gian | 3 |
| Customer\_ID | mã khách hàng | 1223 |
| Item\_ID | mã mặt hàng | 50 |
| Store\_ID | mã cửa hàng | 15 |
| Total\_quantity | tổng số lượng đã mua | 25 |
| Total\_amount | tổng chi phí | 1097.17 |

#### **2.2. Time\_Dim**

Các bảng ghi chứa thông tin về thời gian

|  |  |  |
| --- | --- | --- |
| Thuộc tính | Mô tả | Ví dụ |
| Time\_ID | Mã thời gian - Khóa chính | 2 |
| Quarter | Quý trong năm | 1 |
| Year | Năm | 2020 |

#### **2.3. Customer\_Dim**

Các bản ghi chứa thông tin về khách hàng

TypeCustomerID mang một trong 2 giá trị (“1”, “2”) tương ứng với TypeCustomer: (Postal, Tourist). Hai trường này được sử dụng để phân biệt loại khách hàng.

|  |  |  |
| --- | --- | --- |
| Thuộc tính | Mô tả | Ví dụ |
| Customer\_ID | Khóa của bảng - Khóa chính | 563 |
| Customer\_Name | Tên khách hàng | Thanh Trang |
| City\_id | Mã thành phố | 14 |
| TypeCustomer | Tên loại khách hàng | 1 |
| TypeCustomerID | Mã loại khách hàng | Postal |

#### **2.4. Item\_Dim**

Các bản ghi chứa thông tin về mặt hàng

|  |  |  |
| --- | --- | --- |
| Thuộc tính | Mô tả | Ví dụ |
| Item\_ID | Mã mặt hàng - Khóa chính | 3 |
| Description | Mô tả về mặt hàng | Chất liệu từ cotton |
| Size | Kích cỡ mặt hàng | X |
| Weight | Trọng lượng | 0,45kg |
| Price | Giá cả | 145000 |

#### **2.5. Store\_Dim**

Tất cả các bản ghi về cửa hàng của hệ thống

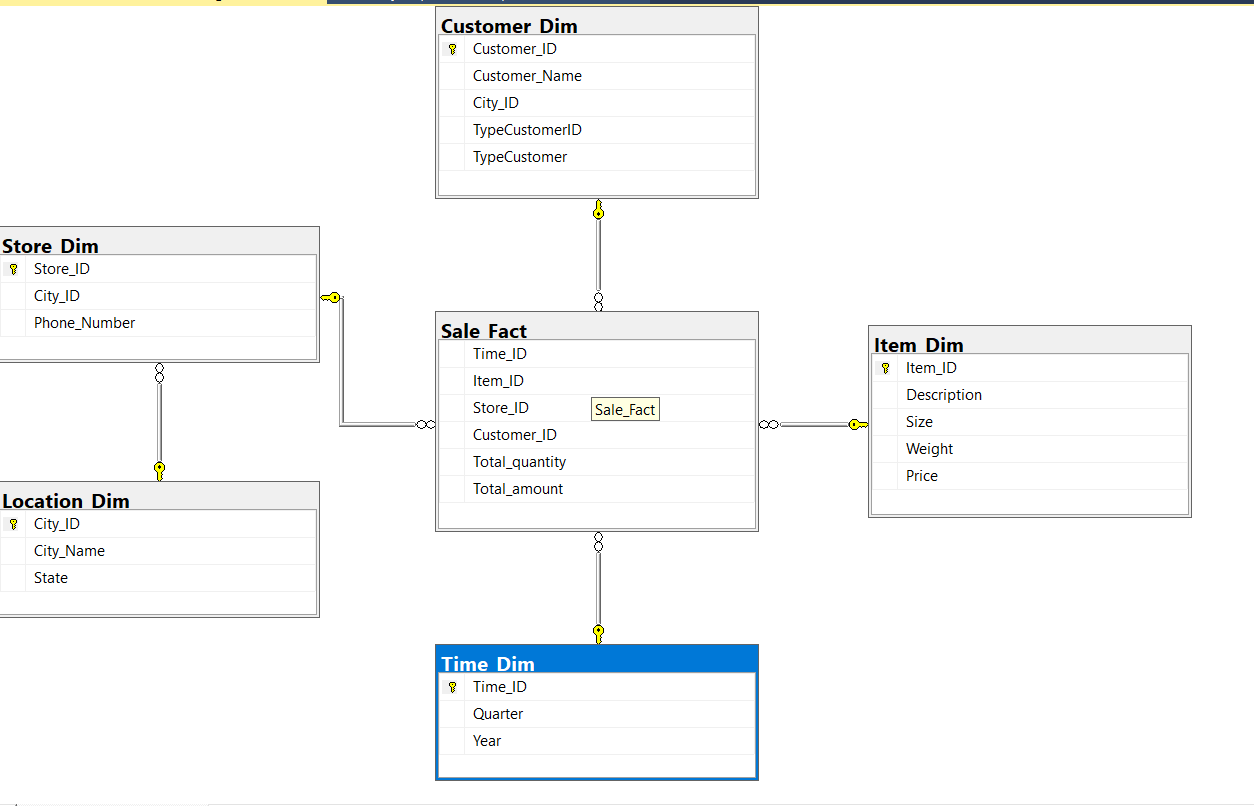
|  |  |  |
| --- | --- | --- |
| Thuộc tính | Mô tả | Ví dụ |
| Store\_ID | Mã cửa hàng-Khóa chính | 24 |
| City\_ID | mã thành phố | 15 |
| Phone\_Number | Số hotline cửa hàng | 0989654321 |

#### **2.6. Chiều thứ cấp Location\_Dim**

Tất cả các bản ghi về địa chỉ của hệ thống

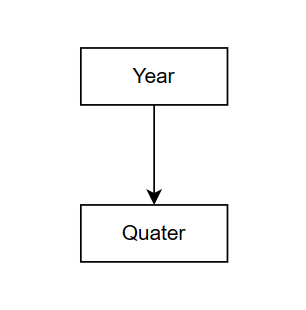
|  |  |  |
| --- | --- | --- |
| Thuộc tính | Mô tả | Ví dụ |
| City\_ID | Mã thành phố-Khóa chính | 1 |
| City\_Name | Tên thành phố | Aurora |
| State | Tên Bang | California |

#### **2.7. Kho**

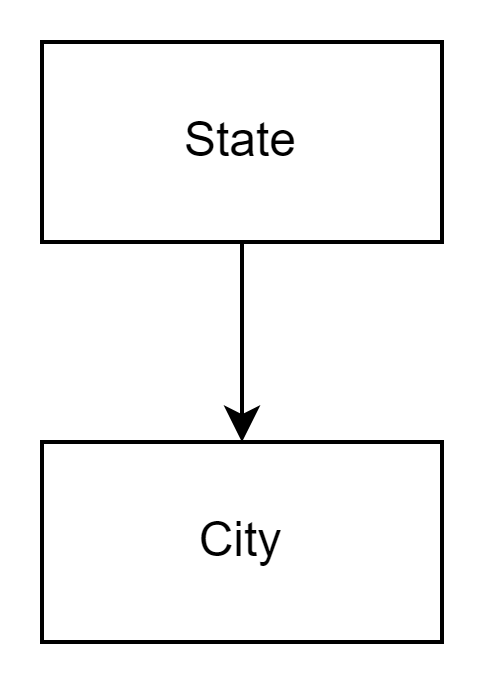
****

### 3. Tạo phân cấp

#### **3.1. Phân cấp cho chiều Thời gian**

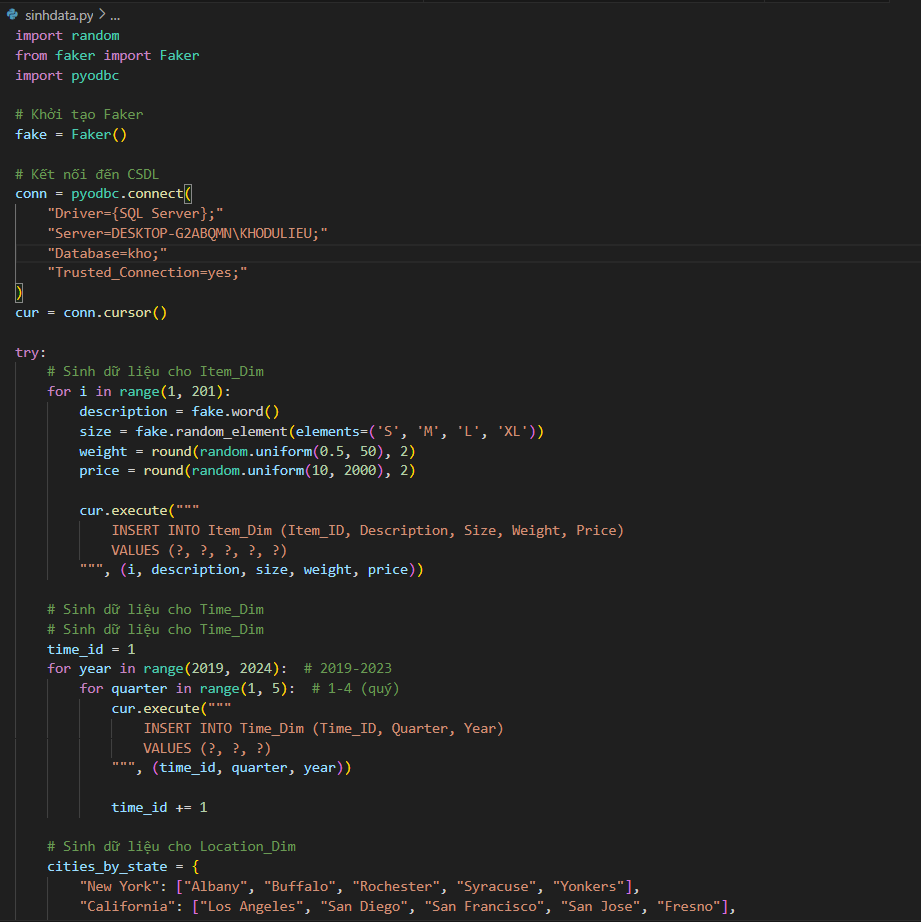


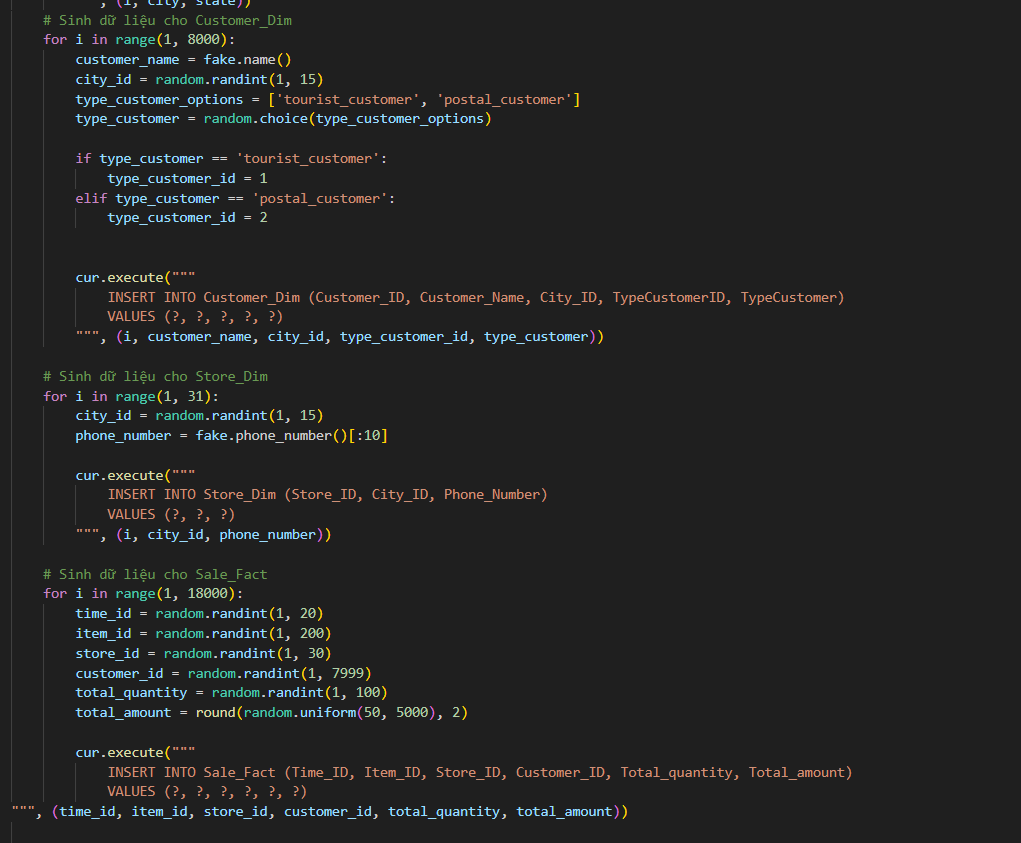
#### **3.2. Phân cấp cho chiều Location**



### 4. Sinh dữ liệu cho IDB

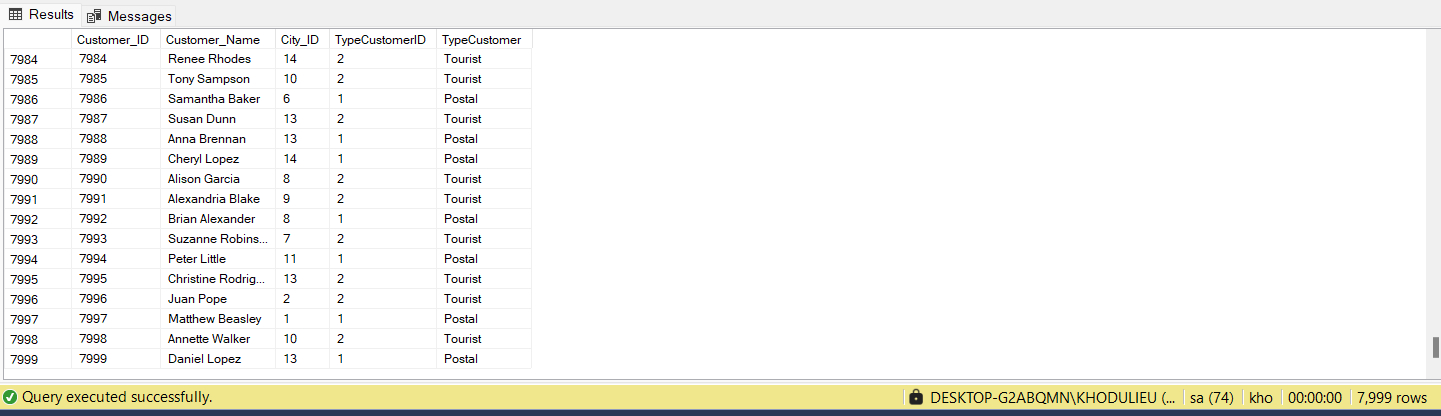
#### **4.1. Code sinh dữ liệu**



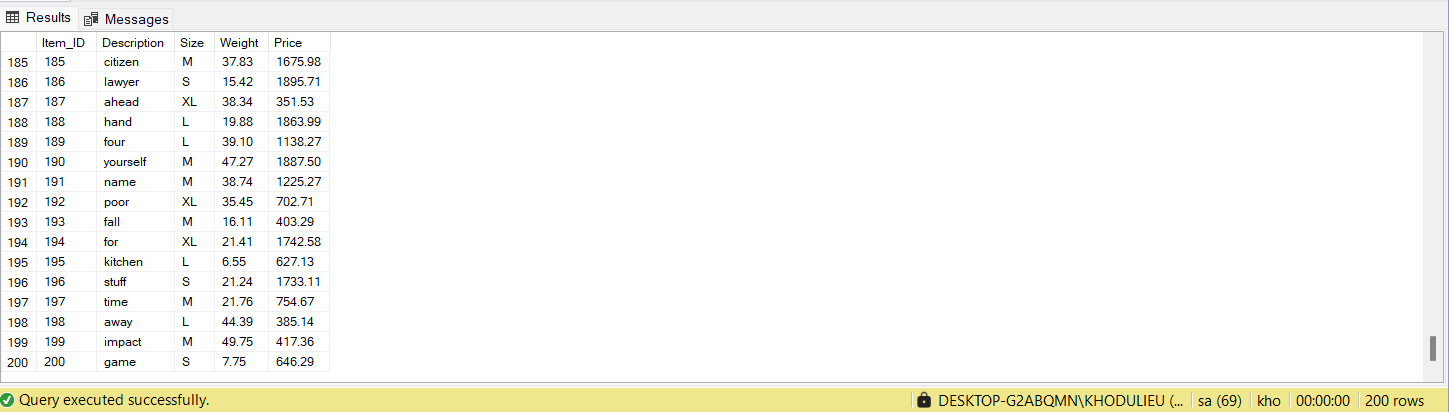


#### **4.2. Kết quả sau khi sinh dữ liệu**

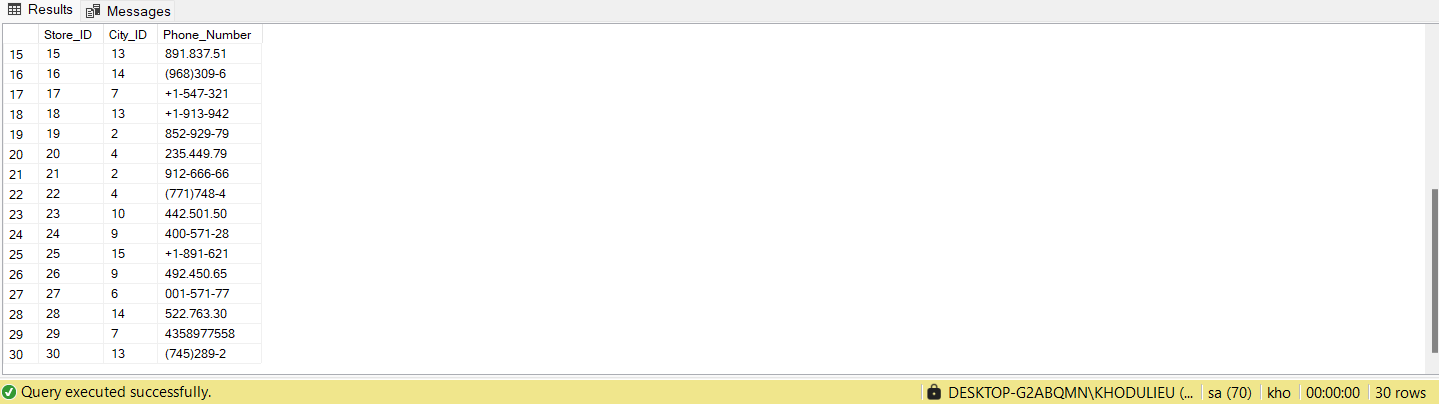
* Bảng Customer\_Dim: gồm 7999 bản ghi



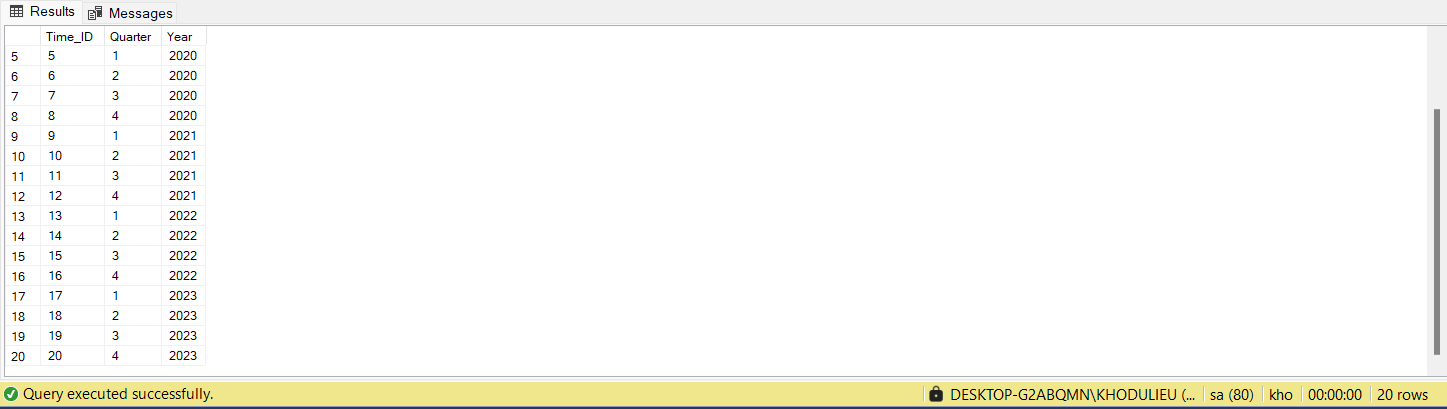
* Bảng Item\_Dim: Gồm 200 bản ghi



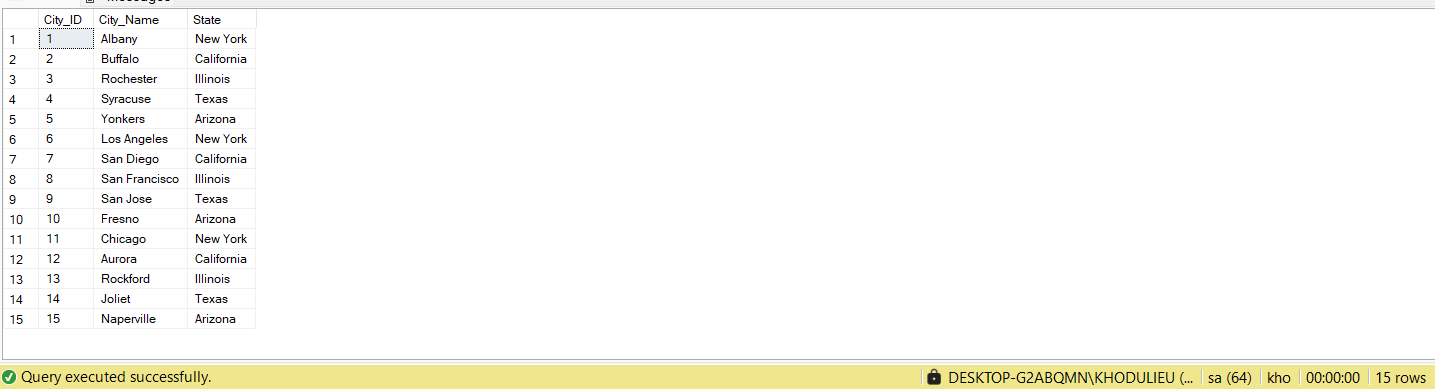
* Bảng Store\_Dim: Gồm 30 bản ghi



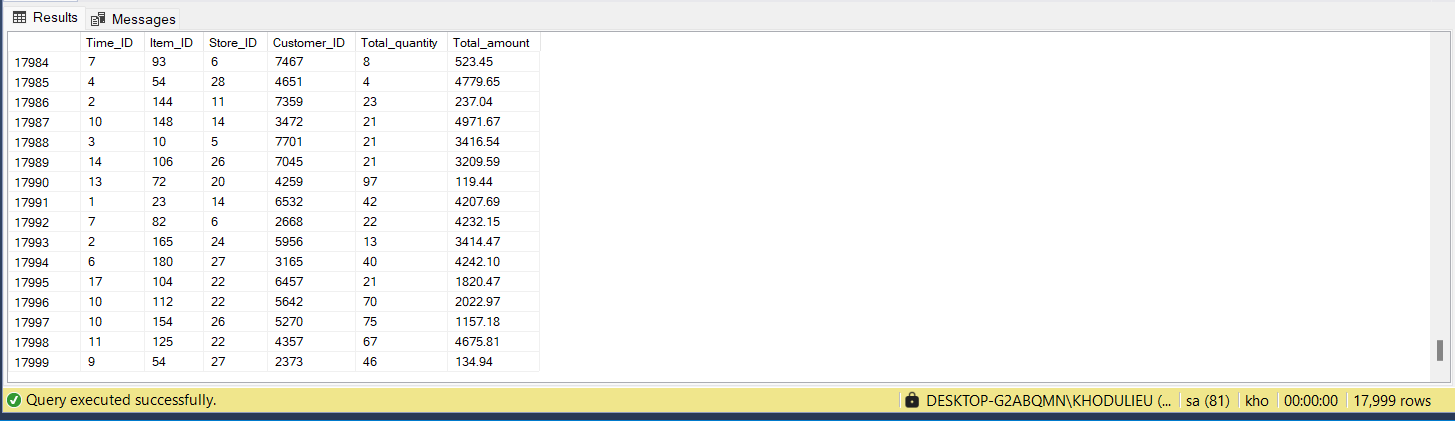
* Bảng Time\_Dim: Gồm 20 bản ghi



* Bảng Location\_Dim: Gồm 15 bản ghi



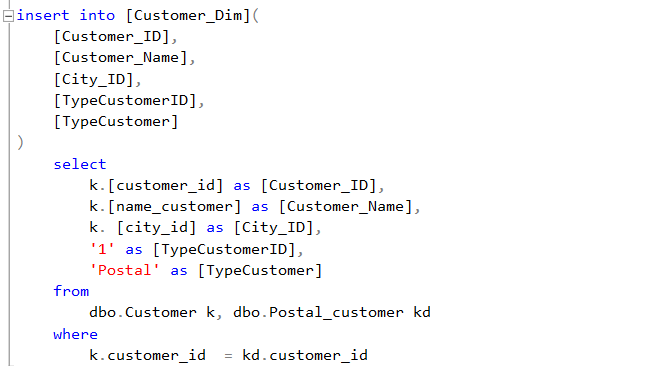
* Bảng Sale\_Fact: Gồm 17999 bản ghi

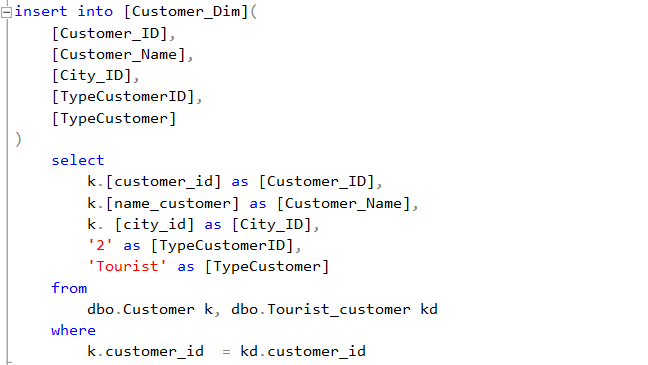


### 5. Cách ánh xạ dữ liệu từ IDB sang kho

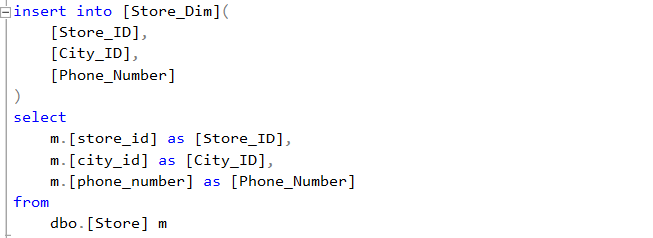
#### **5.1. Ánh xạ dữ liệu vào các bảng Dimension**

* Customer\_Dim: Bảng Customer\_Dim được đổ dữ liệu từ bảng Customer trong CSDL tích hợp. Bước đầu sẽ đổ dữ liệu những khách hàng bưu điện trước, sau đó sẽ đổ dữ liệu những khách hàng du lịch.

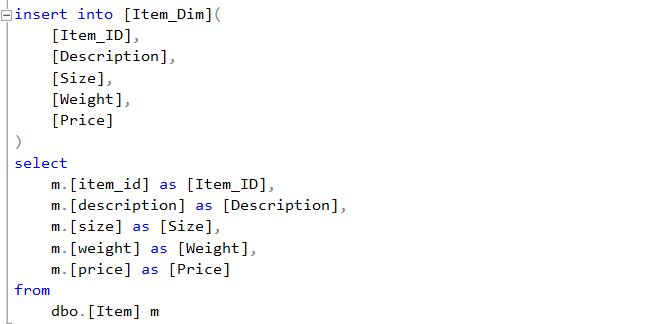




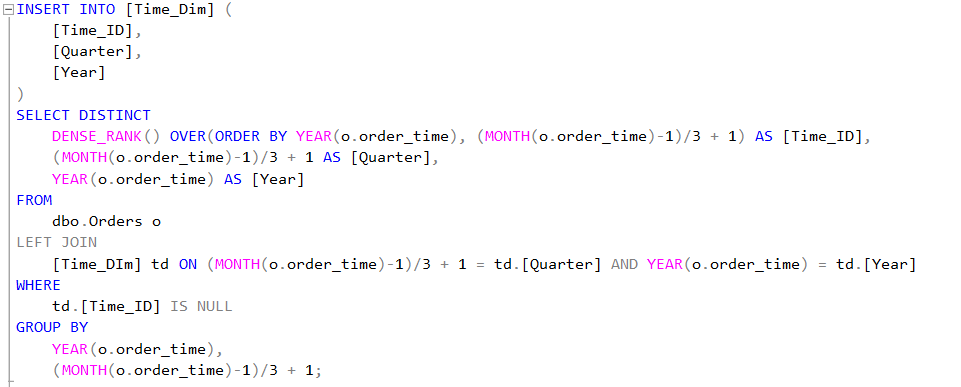
* Store\_Dim: Bảng Store\_Dim được đổ dữ liệu từ bảng Store trong CSDL tích hợp.



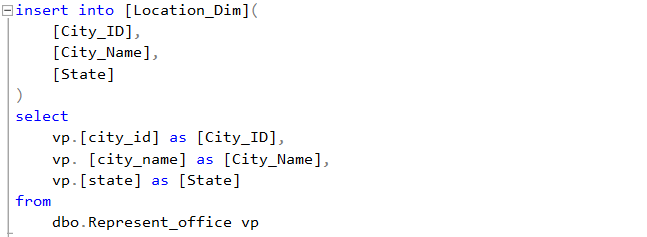
* Item\_Dim: Bảng Item\_Dim được đổ dữ liệu từ bảng Item trong CSDL tích hợp.



* Time\_Dim: Bảng Time\_Dim được đổ dữ liệu bằng cách trích xuất ra quý/năm của thuộc tính order\_time trong Orders trong CSDL tích hợp.

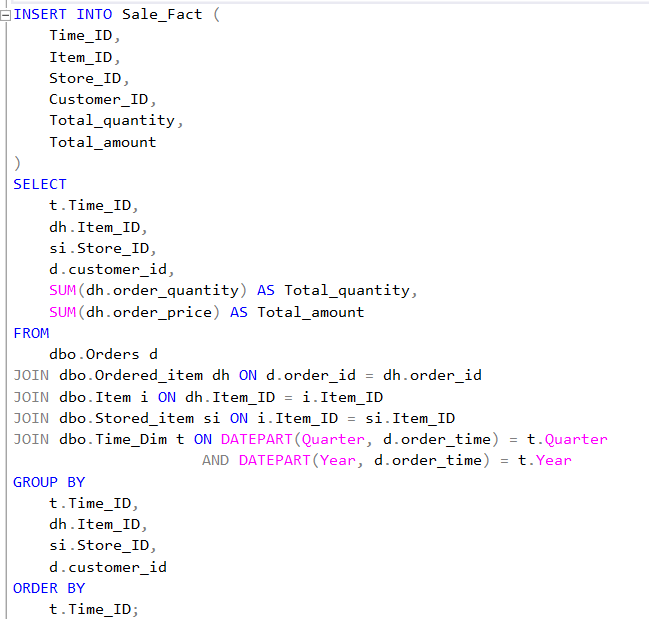


* Location\_Dim: Bảng Location\_Dim được đổ dữ liệu từ bảng Represent\_office trong CSDL tích hợp



#### **5.2. Ánh xạ dữ liệu vào bảng Sale\_Fact:**

* Bảng Sale\_Fact được đổ dữ liệu từ các bảng Orders trong CSDL tích hợp và bảng Location\_Dim, Time\_Dim, Item\_Dim, Customer\_Dim.
* Fact\_Sale: Các bản ghi chứa thông tin về tổng số lượng (Total\_quantity), tổng số tiền (Total\_amount) mà một khách hàng đã chi cho một sản phẩm ở 1 cửa hàng, thành phố trong một quý.
* Total\_quantity được tính bằng tổng số lượng mặt hàng (order\_quantity) được lấy từ bảng Ordered\_item.
* Total\_amount được tính bằng tổng giá tiền mặt hàng được (order\_price) lấy từ bảng Ordered\_item.

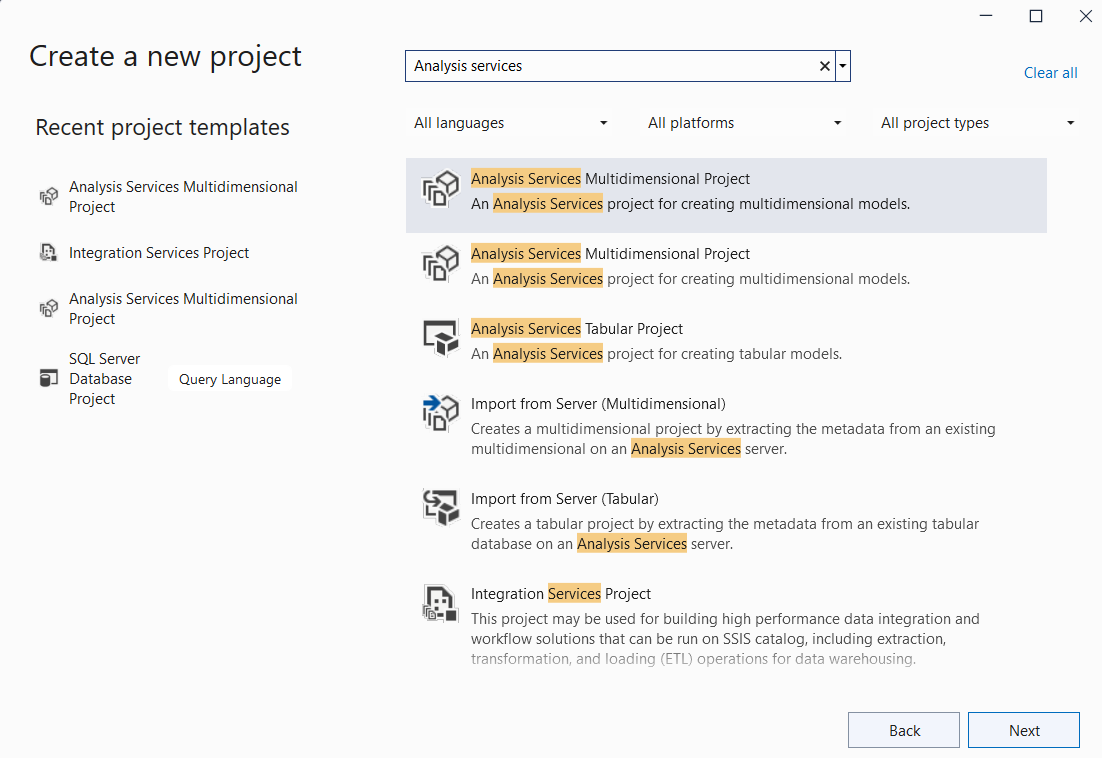


## IV. CÀI ĐẶT , TÍNH TOÁN VÀ LƯU TRỮ CÁC KHỐI DỮ LIỆU

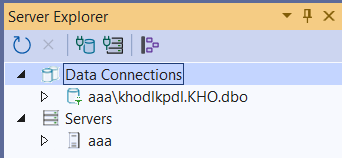
### Cài đặt

Công cụ sử dụng : SQL Server Analysis Services, Sql Server Management studio, Sql Server Integration Services

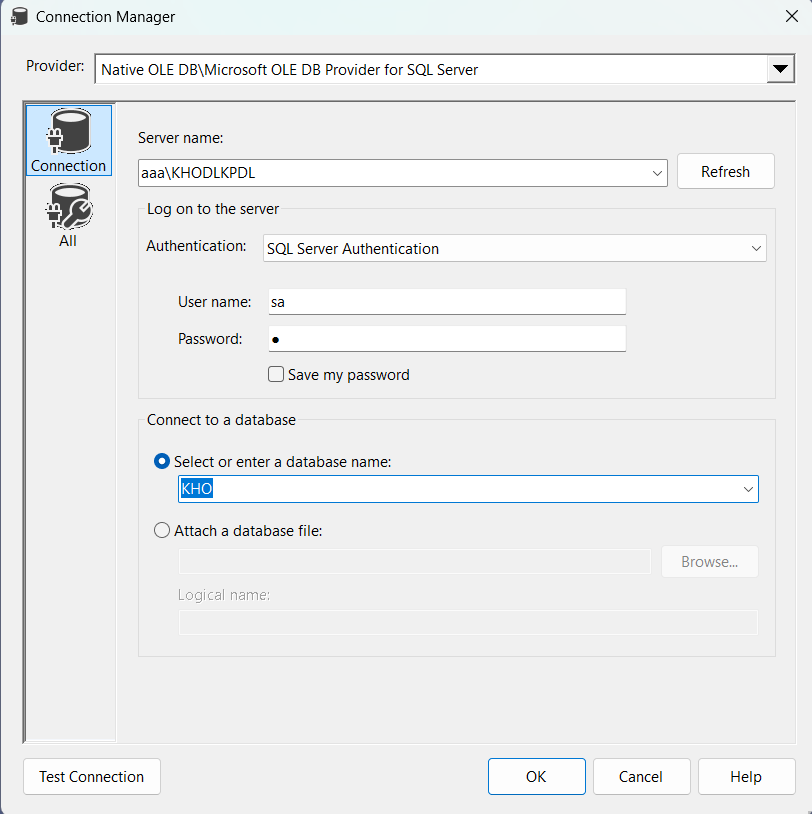
* Tạo một Analysis Services Project:
* Mở Microsoft Visual Studio -> Creat a new project -> Analysis Services Multidimensional Project, đặt tên cho project -> Creat.



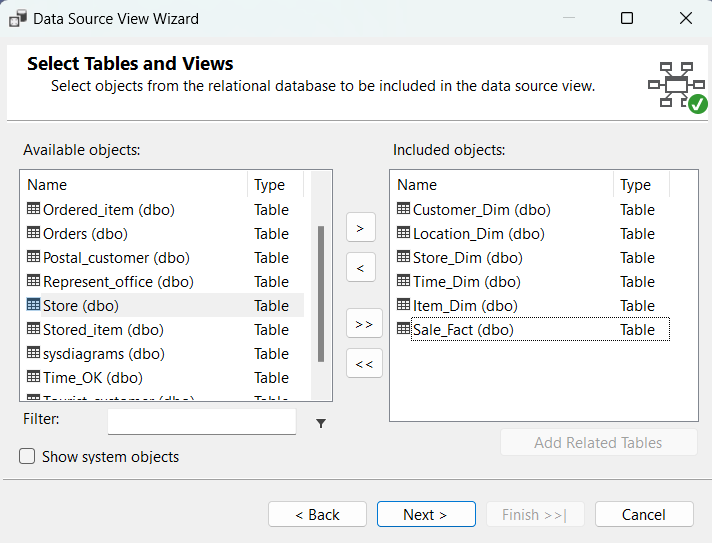
* Kết nối đến SQL Server



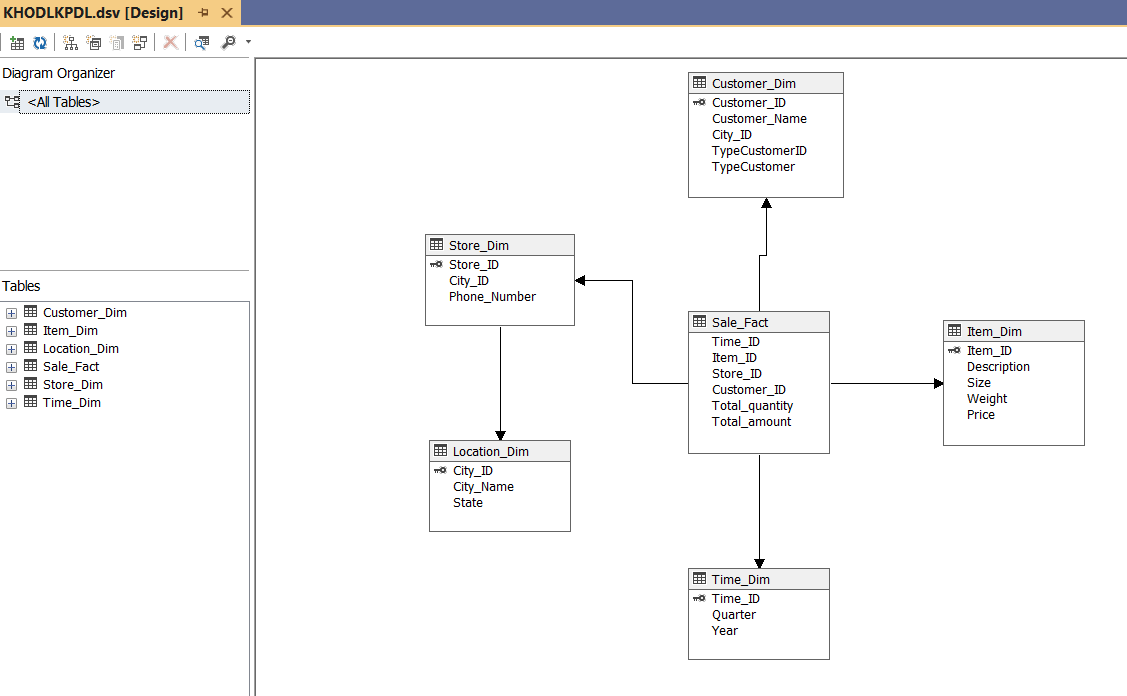
* Tạo Data Source (Kết nối giữa SASS và CSDL)
* Trong Solution Explorer -> New Data Source -> Creat a data source based on an existing or new connection -> New.
* Cửa sổ Connection manager hiên ra -> Chọn Server -> Chọn Database -> Ok.
* Chọn Next -> Inherit -> Đặt tên cho Data Source -> Finish.



* Tạo Data Source View
* Trong Solution Explorer -> New Data Source View -> Chọn Data Source vừa tạo -> Thêm các bảng cần thiết -> Next -> Đặt tên cho Data Source View -> Finish.

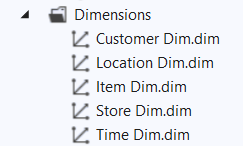


* Data source view chứa logical model của CSDL (tables, keys, columns, và các constraints) sẽ được sử dụng bởi OLAP database để tạo các data cube.

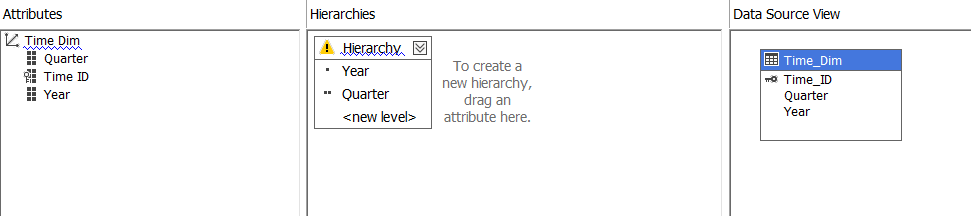


* Tạo các Dimension, thực hiện phân cấp
* Tạo các Dimension:

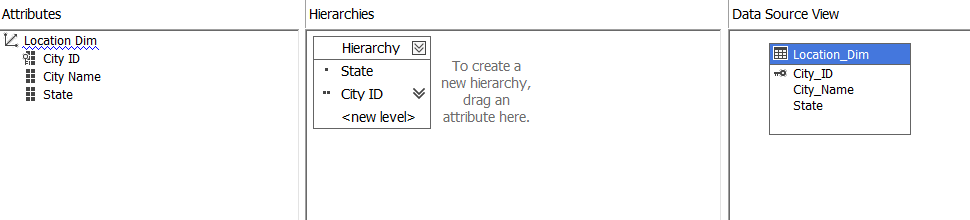
Trong Solution Explorer -> New Dimension -> Use an existing table -> Next -> Chọn các Main table cho bảng cần tạo -> Next -> Tích chọn các Available attributes cần thiết cho bảng -> Next -> Đặt tên cho Dimension -> Finish.



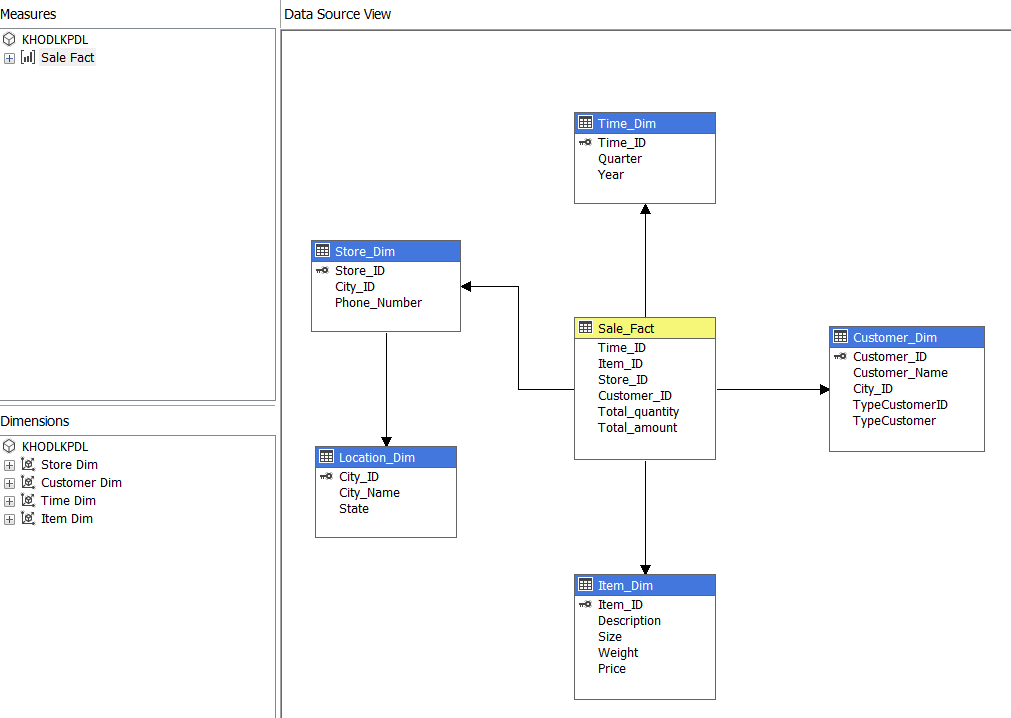
* Chúng ta có thể thay đổi các thuộc tính (attributes), các cây phân cấp (hierarchies), và các quan hệ (relationships).
* Thực hiện phân cấp
* Chiều Time:



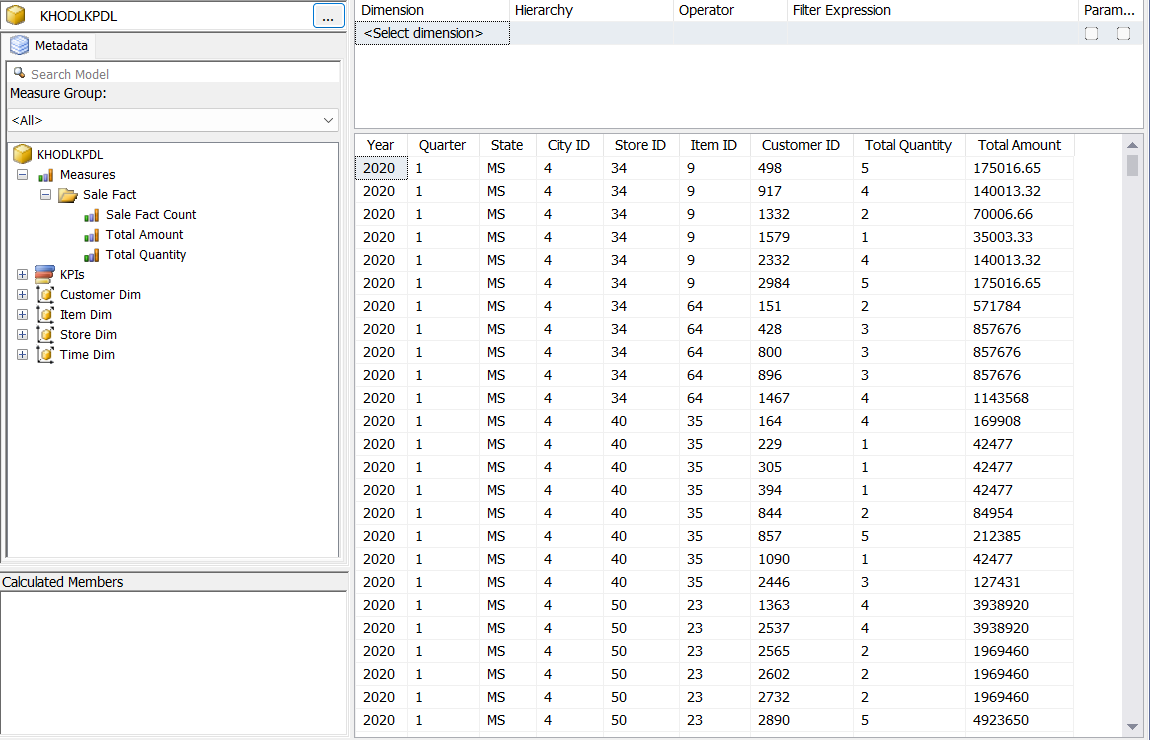
* Chiều Location:



* Tạo Data Cube
* Trong Solution Explorer -> New Cube -> Use existing tables -> Next -> Tích chọn Measure Group Tables -> Next -> Tích chọn các Measure -> Next -> Tích chọn các Dimension -> Next -> Đặt tên cho Cube -> Finish.
* Data Cube:
* Tên khối: Sale\_Fact
* Các chiều: Customer, Time, Store, Item, Location
* Các độ đo: Total\_quantity, Total\_amount



* Sau khi thực hiện xong các bước tiến hành chạy chương trình để xem báo cáo cho một khối bất kỳ.

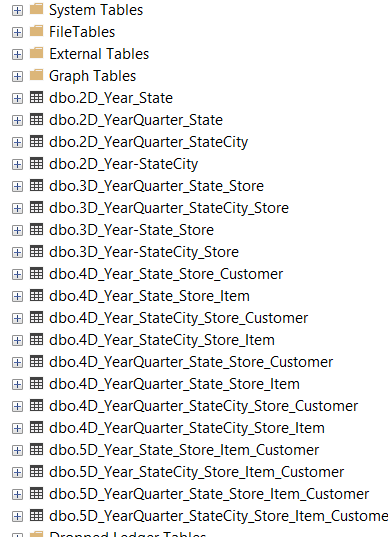


### Tính toán và lưu trữ khối dữ liệu

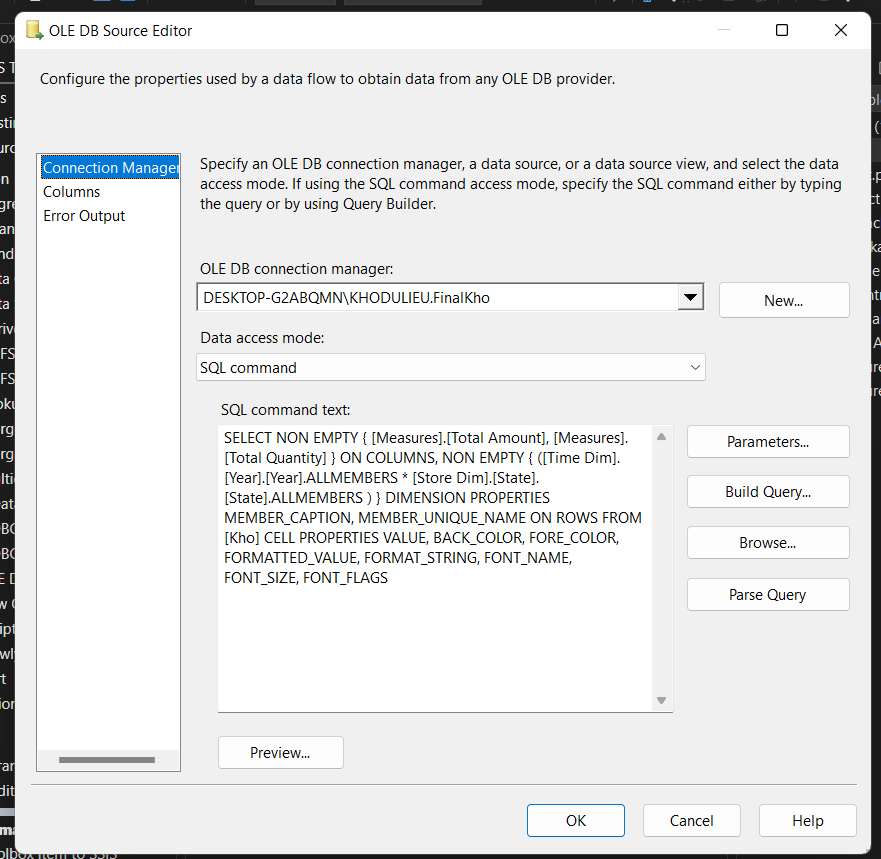
* Thiết kế khối dữ liệu phục vụ các phép toán drill down , roll up,....

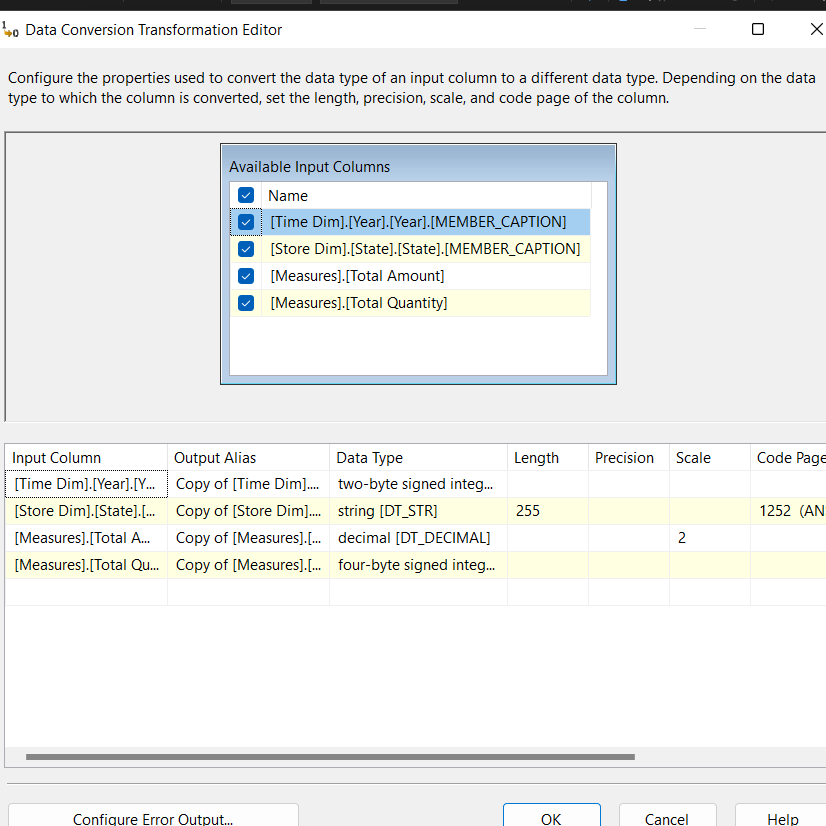
|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | Chiều | | | | | Độ đo | |
|  | STT | Tên Khối | Time | Location | Store | Item | Customer | Amount | Quantity |
| Khối 1 chiều | 1 | TimeQuarter | x |  |  |  |  | x | x |
| 2 | TimeYear | x |  |  |  |  | x | x |
| 3 | Item |  |  |  | x |  | x | x |
| 4 | State |  | x |  |  |  | x | x |
| 5 | StateCity |  | x |  |  |  | x | x |
| Khối 2 chiều | 6 | StateCity-Store |  | x | x |  |  | x | x |
| 7 | State-Store |  | x | x |  |  | x | x |
| 8 | StateCity-Customer |  | x |  |  | x | x | x |
| 9 | State-Customer |  | x |  |  | x | x | x |
| 10 | StateCity-Item |  | x |  | x |  | x | x |
| 11 | State-Item |  | x |  | x |  | x | x |
| 12 | YearQuarter- StateCity | x | x |  |  |  | x | x |
| 13 | Year-StateCity | x | x |  |  |  | x | x |
| 14 | YearQuarter-State | x | x |  |  |  | x | x |
| 15 | Year-State | x | x |  |  |  | x | x |
| 16 | YearQuarter-Item | x |  |  | x |  | x | x |
| 17 | Year-Item | x |  |  | x |  | x | x |
| Khối 3 chiều | 18 | Year-StateCity-Store | x | x | x |  |  | x | x |
| 19 | YearQuarter -StateCity-Store | x | x | x |  |  | x | x |
| 20 | Year -State-Store | x | x | x |  |  | x | x |
| 21 | YearQuarter -State-Store | x | x | x |  |  | x | x |
| 22 | YearQuarter -StateCity-Customer | x | x |  |  | x | x | x |
| 23 | StateCity-Customer-Year | x | x |  |  | x | x | x |
| 24 | State-Customer-Year | x | x |  |  | x | x | x |
| 25 | State-Customer-YearQuarter | x | x |  |  | x | x | x |
| 26 | Year-StateCity-Item | x | x |  | x |  | x | x |
| 27 | YearQuarter-StateCity-Item | x | x |  | x |  | x | x |
| 28 | Year-State-Item | x | x |  | x |  | x | x |
| 29 | YearQuarter-State-Item | x | x |  | x |  | x | x |
| 30 | StateCity-Store-Item |  | x | x | x |  | x | x |
| 31 | State-Store-Item |  | x | x | x |  | x | x |
| 31 | StateCity-Item-Customer |  | x |  | x | x | x | x |
| 32 | State-Customer-Item |  | x |  | x | x | x | x |
| 33 | StateCity-Store-Customer |  | x | x |  | x | x | x |
| 34 | State-Store-Customer |  | x | x |  | x | x | x |
| Khối 4 chiều | 35 | Year-State-Store-Customer | x | x | x |  | x | x | x |
| 36 | YearQuarter-State-Store-Customer | x | x | x |  | x | x | x |
| 37 | Year-StateCity-Store-Customer | x | x | x |  | x | x | x |
| 38 | YearQuarter-StateCity-Store-Customer | x | x | x |  | x | x | x |
| 39 | Year-State-Store-Item | x | x | x | x |  | x | x |
| 40 | YearQuarter-State-Store- Item | x | x | x | x |  | x | x |
| 41 | Year-StateCity-Store- Item | x | x | x | x |  | x | x |
| 42 | YearQuarter-StateCity-Store- Item | x | x | x | x |  | x | x |
| Khối 5 chiều  (Base) | 43 | Year-State-Store-Item-Customer | x | x | x | x | x | x | x |
| 44 | YearQuarter-State-Store- Item-Customer | x | x | x | x | x | x | x |
| 45 | Year-StateCity-Store- Item-Customer | x | x | x | x | x | x | x |
| 46 | YearQuarter-StateCity-Store- Item -Customer | x | x | x | x | x | x | x |

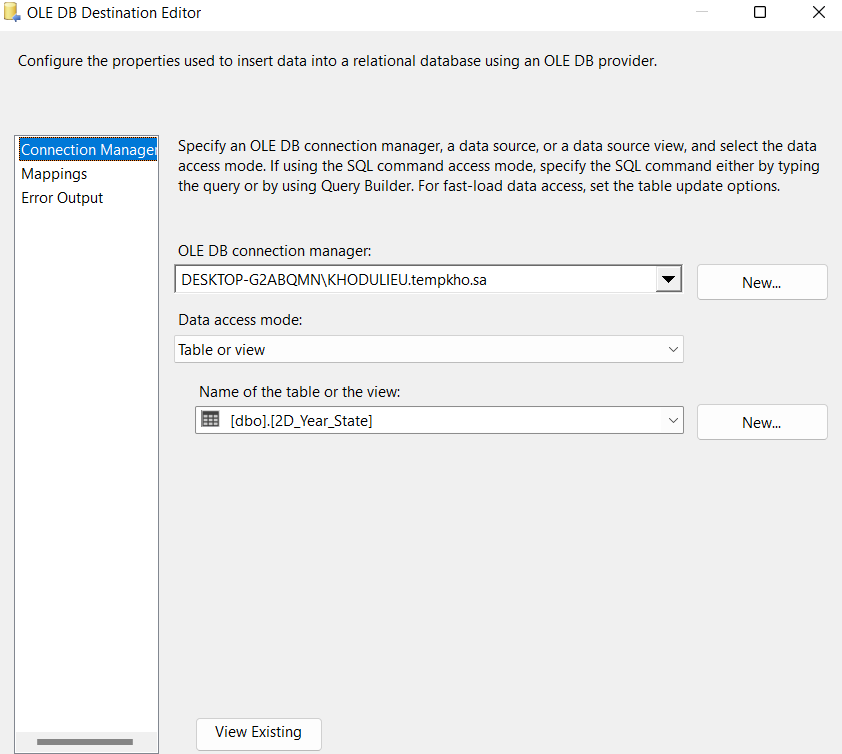
* Tải dữ liệu vào các khối

Bước 1: Tạo bảng tạm để lưu các khối 

Bước 2: Thực hiện truy vấn dữ liệu khối và thực hiện lưu vào bảng







## V .METADATA

1.Mô tả cấu trúc của Data warehouse

\*Schema(snowflake) và Dims

* 4 Chiều chính : Customer\_Dim, Item\_Dim, Store\_Dim, Time\_Dim

·

* Chiều thứ cấp: Location\_Dim

\*Data mart(Sale data mart)

- Vị trí : ‘Kho’

- Content :

+ Bảng Sale\_Fact : chứa dữ liệu bán hàng đã tổng hợp

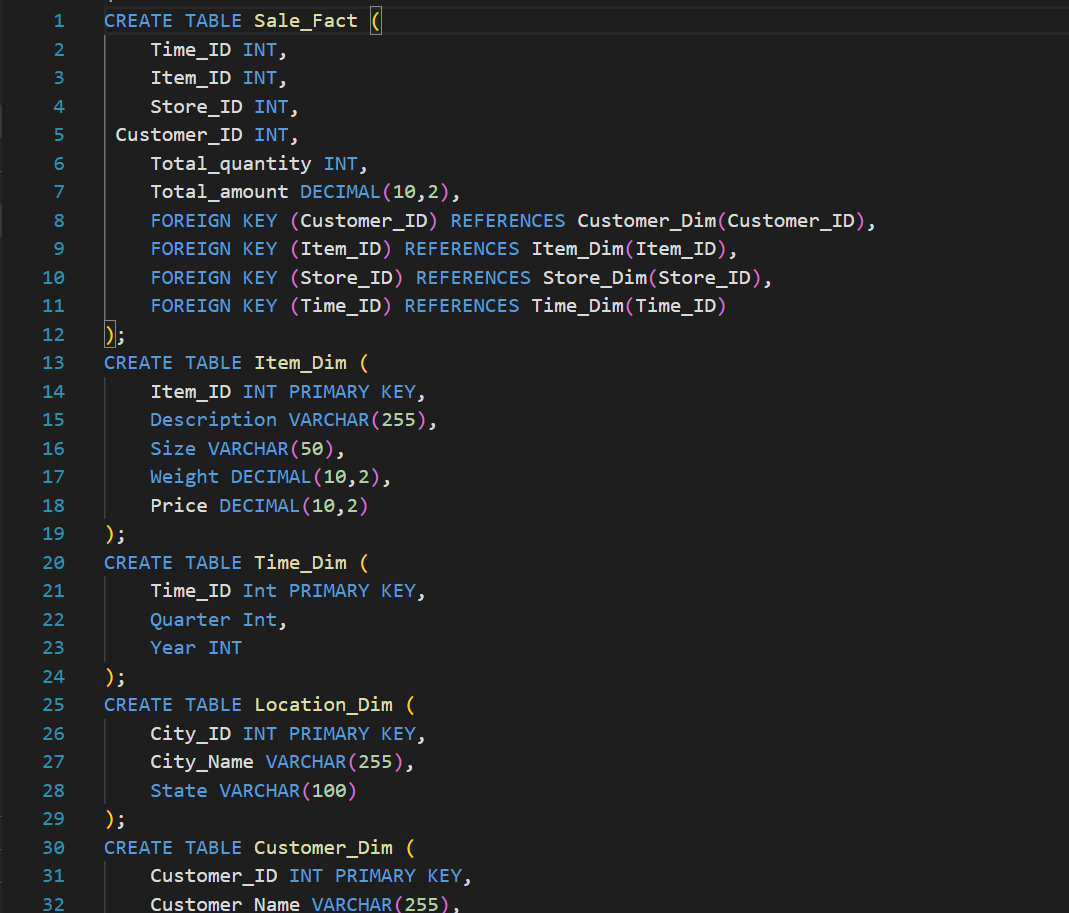
+ Bảng Item\_Dim : Chứa thông tin chi tiết về sản phẩm

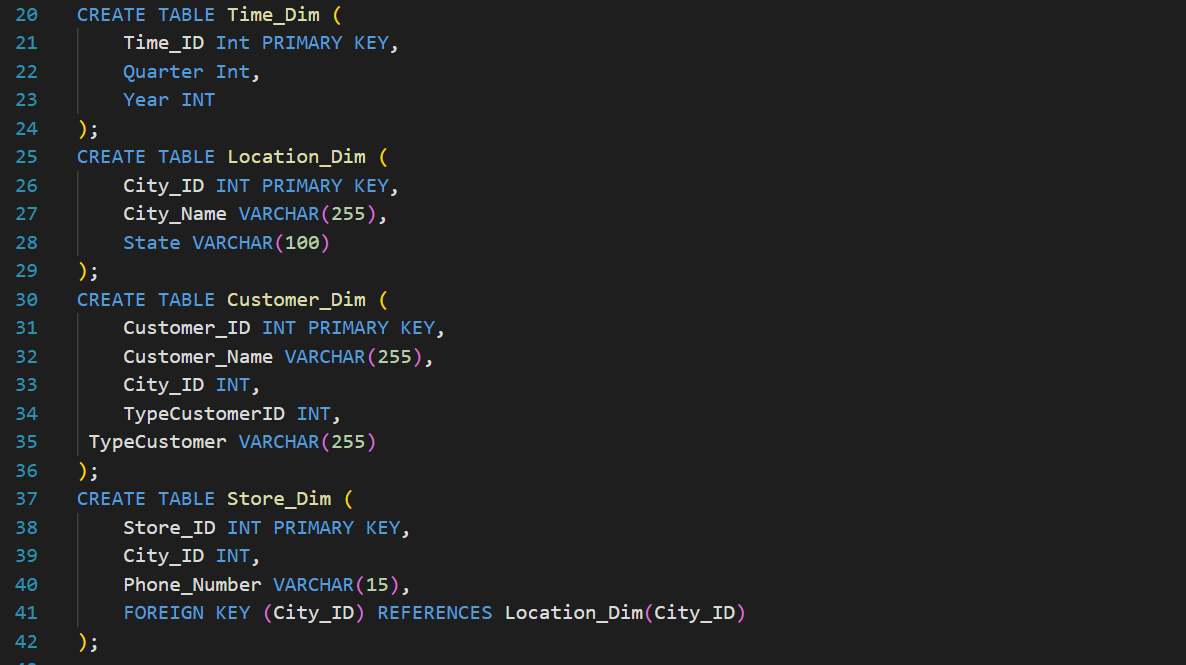
+ Bảng Time\_Dim : Chứa thông tin về thời gian

+ Bảng Location\_Dim : Chứa thông tin về địa chỉ của cửa hàng

+ Bảng Customer\_Dim : Chứa thông tin về khách hàng

+ Bảng Store\_Dim : Chứa thông tin về cửa hàng



  
 2.Operational metadata

\*currency (Tình trạng dữ liệu)

Bảng Sale\_Fact : Trạng thái: actived

+ Bảng Item\_Dim : Trạng thái: actived

+ Bảng Time\_Dim : Trạng thái: actived

+ Bảng Location\_Dim : Trạng thái: actived

+ Bảng Customer\_Dim : Trạng thái: actived

+ Bảng Store\_Dim : Trạng thái: actived

\*Số lượng truy vấn và hiệu suất

-- Lấy tổng số lượng truy vấn và thời gian phản hồi trung bình

SELECT

COUNT(\*) AS total\_queries,

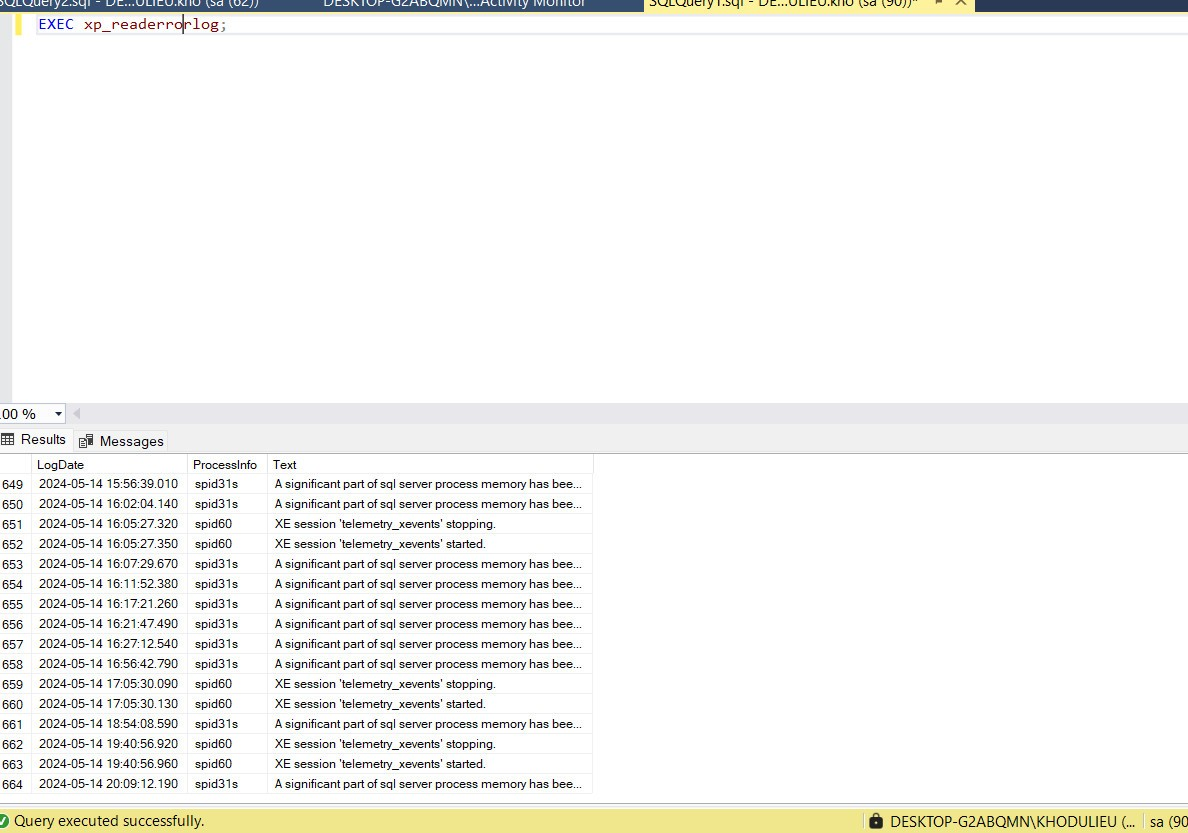
AVG(total\_worker\_time) AS avg\_response\_time\_ms

FROM sys.dm\_exec\_query\_stats;

\*Báo cáo lỗi

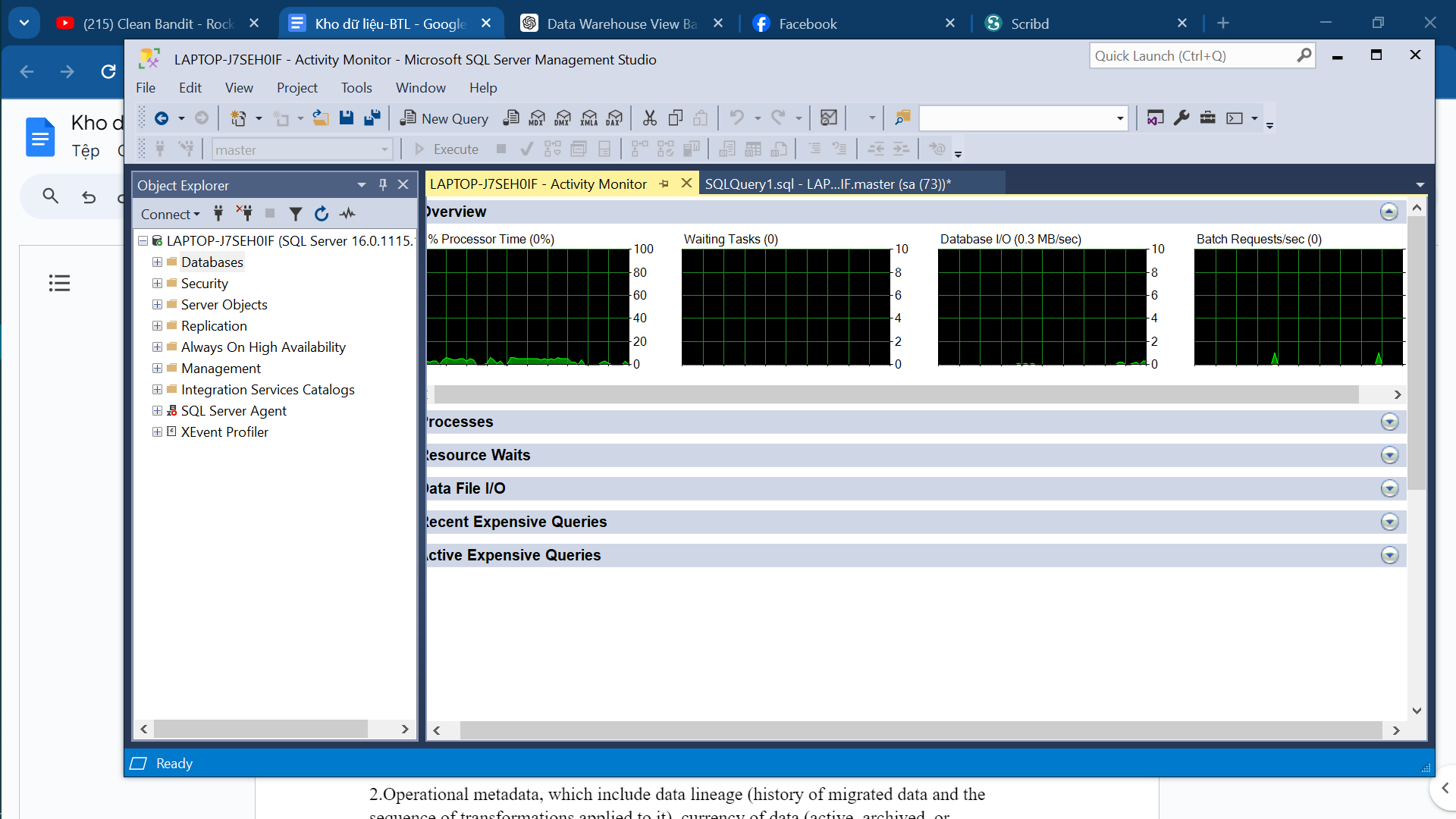
Truy cập error log bằng T-SQL trong sql server

EXEC xp\_readerrorlog;

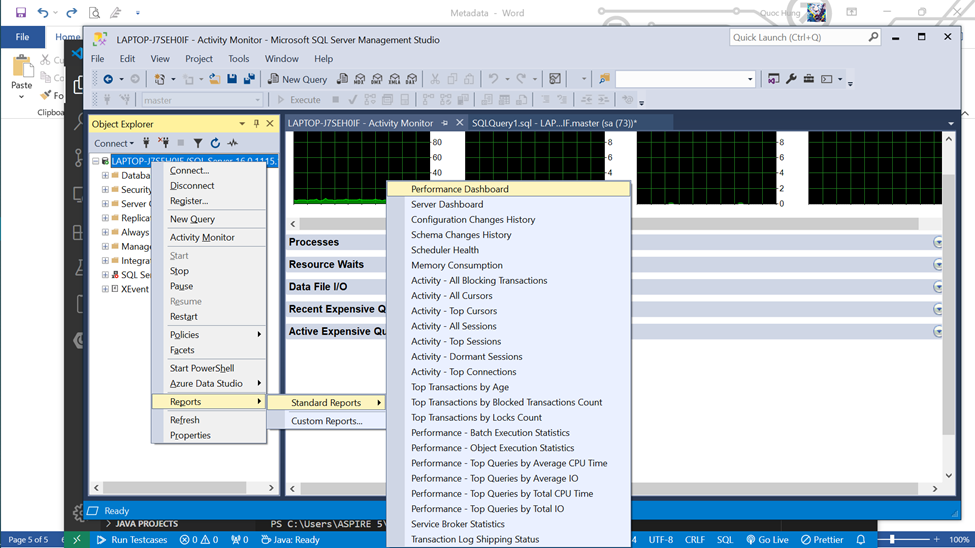


Giám sát hoạt động của Sql Server qua Reports và Activity Monitor

- Activity Monitor: giám sát các phiên kết nối,truy vấn đang chạy và tài nguyên hệ thống



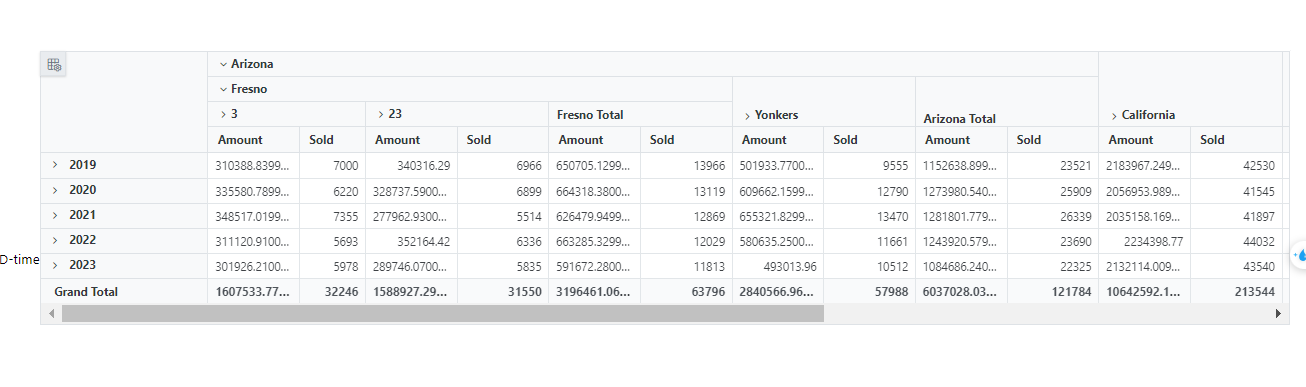
- Reports: thống kê thời gian truy vấn như thời gian thực thi,tần suất



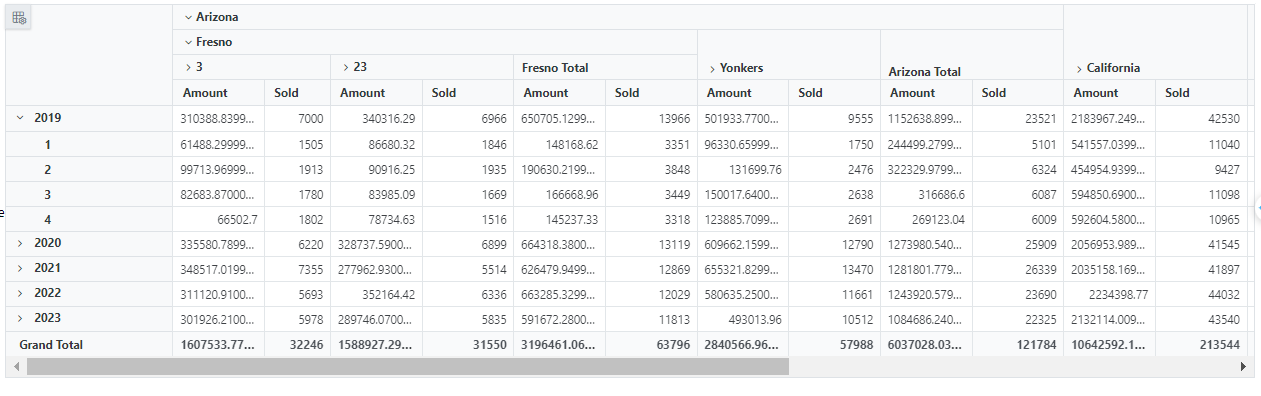
## VI. XÂY DỰNG ỨNG DỤNG BÁO CÁO PHÂN TÍCH TRỰC TUYẾN

Bảng báo cáo phân tích trực tuyến có tối đa 5 chiều

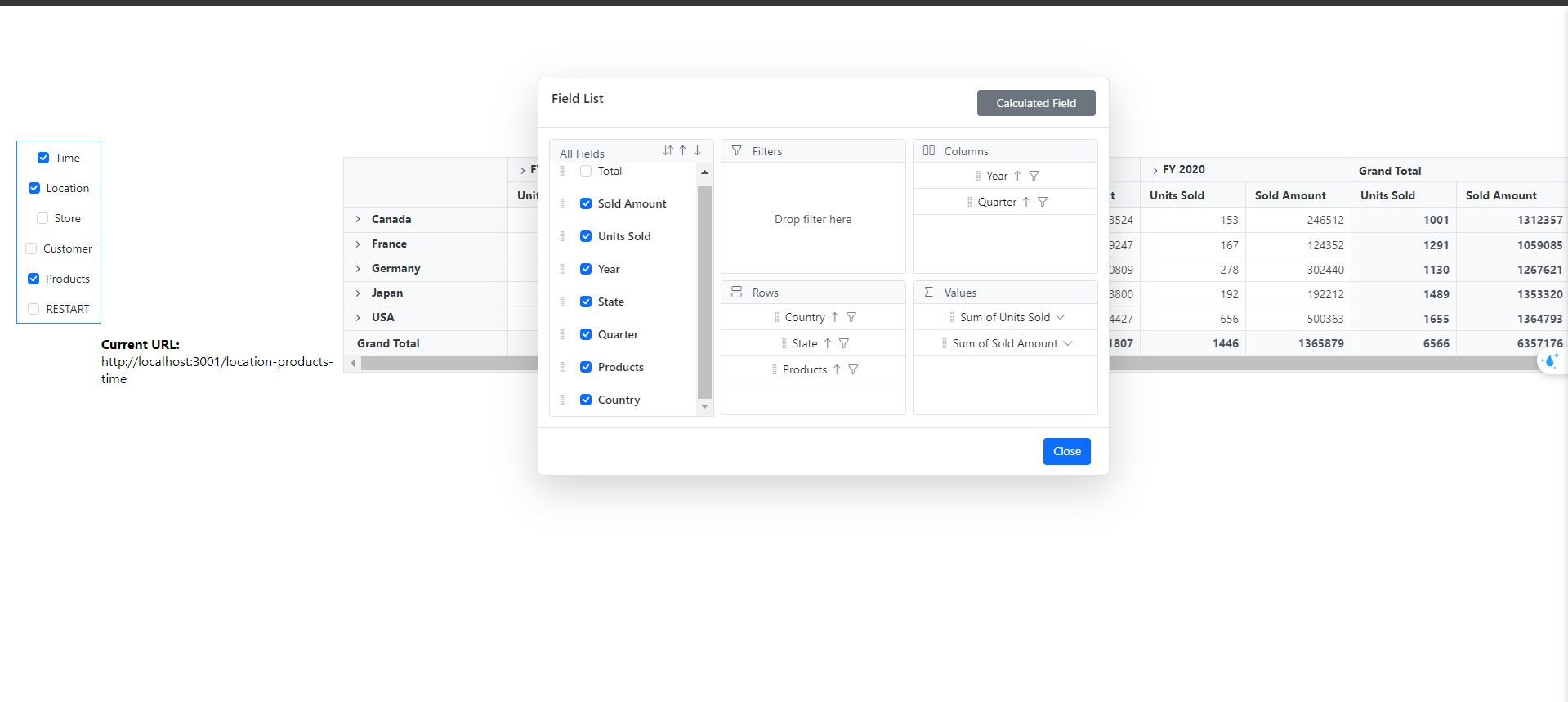
* Hình sau demo bảng 5 chiều



* Demo drill down:

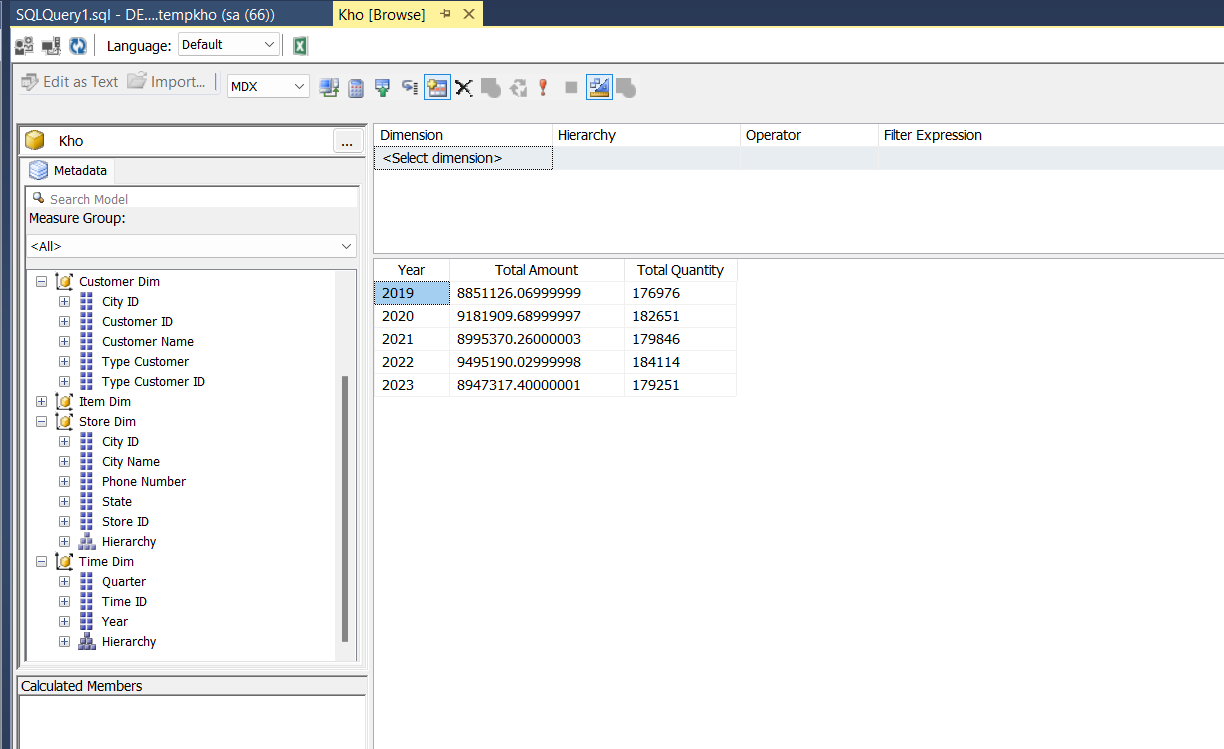


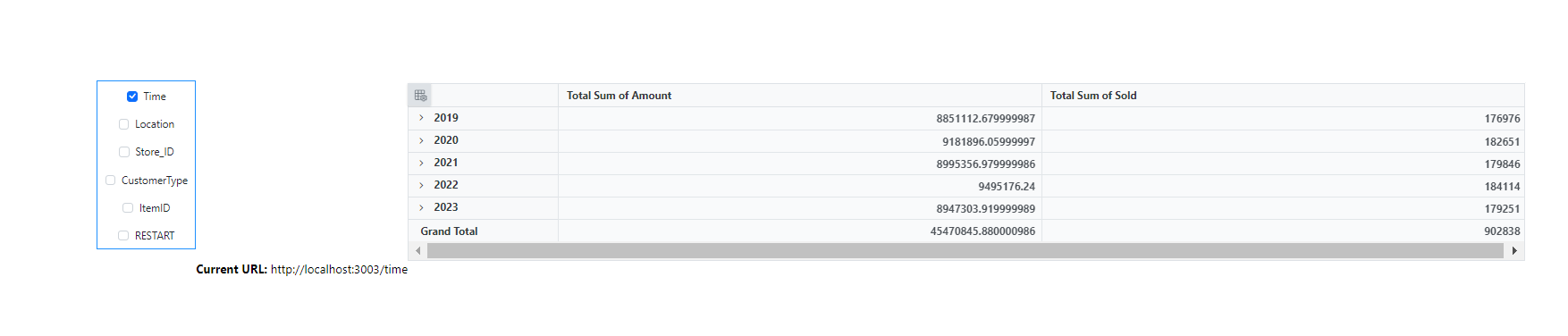
* Filter phục vụ cho việc thực hiện phép pivot, Biểu diễn các chiều theo hàng hoặc cột



## VII. KIỂM TRA TÍNH ĐÚNG ĐẮN

* Dữ liệu trong xuất lên báo cáo và dữ liệu được truy vấn trực tiếp từ kho là như nhau





## VIII. KẾT LUẬN

Nhóm đã hoàn thành được những yêu cầu của cô:

* Tích hợp hai cơ sở dữ liệu thành một cơ sở dữ liệu tích hợp
* Xây dựng lược đồ hình bông tuyết và tạo kho dữ liệu từ các lược đồ đó.
* Sinh dữ liệu cho các nguồn dữ liệu
* Ánh xạ dữ liệu từ cơ sở dữ liệu tích hợp sang kho dữ liệu.
* Tạo phân cấp để tính khối cho OLAP
* Thiết kế, xây dựng và tính sẵn các khối dữ liệu theo yêu cầu bài toán
* Chạy thử nghiệm công nghệ kho dữ liệu và xử lý phân tích trực tuyến
* Xây dựng giao diện với người dùng để demo các phép toán cơ bản của OLAP bao gồm khoan sâu xuống (drill down), cuộn lên (roll up), chiếu chọn (slide and dice), xoay (pivot).