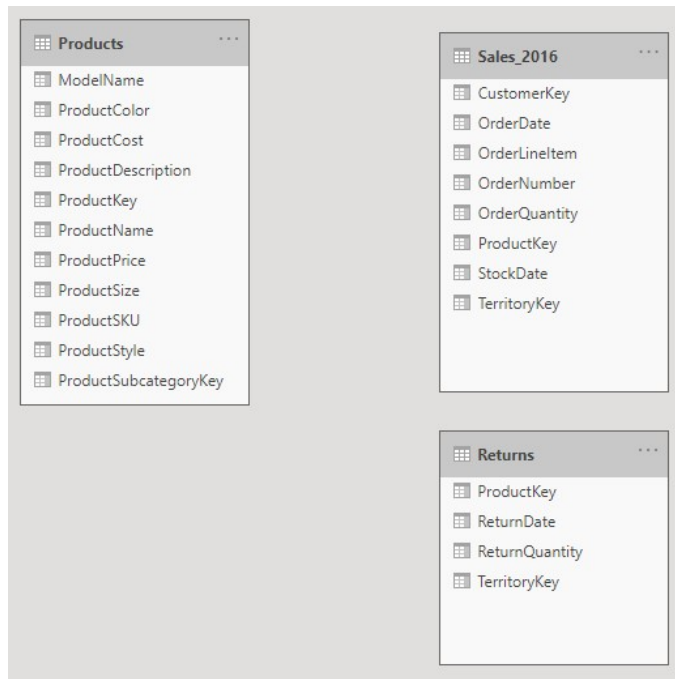


Data Model

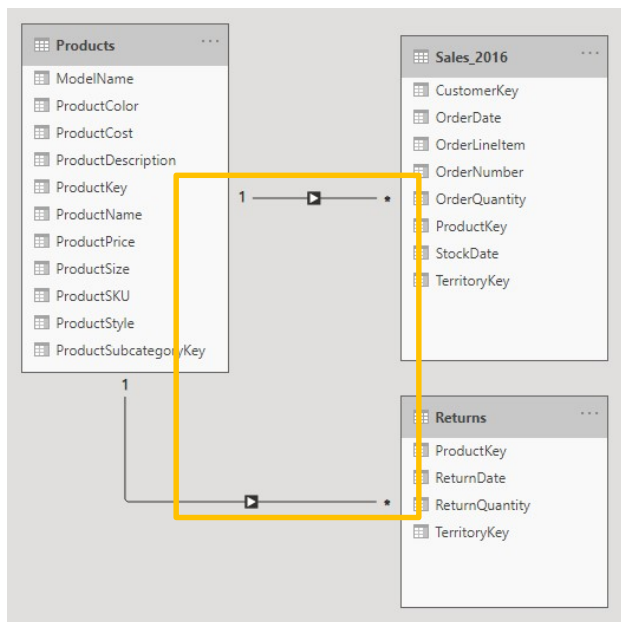


Data Model là gì?



ProductName	ReturnQuantity	OrderQuantity
All-Purpose Bike Stand	1828	36230
AWC Logo Cap	1828	36230
Bike Wash - Dissolver	1828	36230
Cable Lock	1828	36230
Chain	1828	36230
Classic Vest, L	1828	36230
Classic Vest, M	1828	36230
Classic Vest, S	1828	36230
Fender Set - Mountain	1828	36230
Front Brakes	1828	36230
Front Derailleur	1828	36230
Full-Finger Gloves, L	1828	36230
Full-Finger Gloves, M	1828	36230
Full-Finger Gloves, S	1828	36230
Half-Finger Gloves, L	1828	36230
Half-Finger Gloves, M	1828	36230
Half-Finger Gloves, S	1828	36230
Headlights - Dual-Beam	1828	36230
Headlights - Weatherproof	1828	36230
Total	1828	36230

Data Model là gì?



ProductName	ReturnQuantity	OrderQuantity
All-Purpose Bike Stand	8	118
AWC Logo Cap	46	1757
Bike Wash - Dissolver	25	757
Classic Vest, L	4	67
Classic Vest, M	7	71
Classic Vest, S	8	66
Fender Set - Mountain	54	1738
Half-Finger Gloves, L	18	325
Half-Finger Gloves, M	16	381
Half-Finger Gloves, S	15	399
Hitch Rack - 4-Bike	8	135
HL Mountain Tire	49	569
HL Road Tire	28	389
Hydration Pack - 70 oz.	25	301
LL Mountain Tire	39	704
LL Road Tire	43	835
Long-Sleeve Logo Jersey, L	15	186
Long-Sleeve Logo Jersey, M	15	165
Long-Sleeve Logo Jersey, S	12	143
Total	1828	36230

Chuẩn hóa Database

Chuẩn hóa Database là quá trình tổ chức dữ liệu tại các cột-bảng trong một mối quan hệ dữ liệu, để giảm sự dư thừa và đảm bảo tính toàn vẹn. Một số quy tắc:

- Loại bỏ dữ liệu dư thừa để giảm kích thước bảng và cải thiện tốc độ & hiệu quả xử lý
- Giảm thiểu lỗi và sự bất thường từ sửa đổi dữ liệu (chèn, cập nhật hoặc xóa...)
- Đơn giản hóa các queries và cấu trúc database có ý nghĩa phân tích

TIP: Trong việc chuẩn hóa database, các bảng nên đảm bảo mục đích *riêng biệt* và *cụ thể*

date	product_id	quantity	product_brand	product_name	product_sku	product_weight
1/1/1997	869	5	Nationeel	Nationeel Grape Fruit Roll	52382137179	17
1/7/1997	869	2	Nationeel	Nationeel Grape Fruit Roll	52382137179	17
1/3/1997	1	4	Washington	Washington Berry Juice	90748583674	8.39
1/1/1997	1472	3	Fort West	Fort West Fudge Cookies	37276054024	8.28
1/6/1997	1472	2	Fort West	Fort West Fudge Cookies	37276054024	8.28
1/5/1997	2	4	Washington	Washington Mango Drink	96516502499	7.42
1/1/1997	76	4	Red Spade	Red Spade Sliced Chicken	62054644227	18.1
1/1/1997	76	2	Red Spade	Red Spade Sliced Chicken	62054644227	18.1
1/5/1997	3	2	Washington	Washington Strawberry Drink	58427771925	13.1
1/7/1997	3	2	Washington	Washington Strawberry Drink	58427771925	13.1
1/1/1997	320	3	Excellent	Excellent Cranberry Juice	36570182442	16.4

Phân biệt Data Tables và Lookup Tables

Data Model thông thường bao gồm 2 loại bảng: **data** (hoặc “*fact*”) tables, và **lookup** (or “*dimension*”) tables

- **Data tables** thường dạng number hoặc value, thông thường ở cấp độ đầu tiên, với ID or “*key*” columns có thể sử dụng để tạo ra các **relationships**
- **Lookup tables** cung cấp thêm các diễn giải, để bổ sung thông tin, thường dạng text

The diagram illustrates a data model with three tables. The **Data Table** (Orders) is on the left, with columns OrderNumber, CustomerKey, and ProductKey. The **Customer Lookup Table** is in the top right, with columns CustomerKey, LastName, FirstName, and EmailAddress. The **Product Lookup Table** is in the bottom right, with columns ProductKey, ProductName, ProductPrice, and ModelName. Arrows indicate the relationships: a blue arrow from CustomerKey in the Data Table to the Customer Lookup Table, and a yellow arrow from ProductKey in the Data Table to the Product Lookup Table.

OrderNumber	CustomerKey	ProductKey
SO61283	26273	220
SO61283	26273	529
SO61283	26273	538
SO61284	27477	480
SO61284	27477	538
SO61285	23791	214
SO61285	23791	529
SO61285	23791	540
SO61286	11530	528
SO61286	11530	536
SO61287	15760	477
SO61287	15760	478
SO61287	15760	489
SO61288	19709	220
SO61288	19709	528
SO61288	19709	536
SO61289	23694	215
SO61289	23694	528

CustomerKey	LastName	FirstName	EmailAddress
11000	YANG	JON	jon24@adventure-works.com
11001	HUANG	EUGENE	eugene10@adventure-works.com
11002	TORRES	RUBEN	ruben35@adventure-works.com
11003	ZHU	CHRISTY	christy12@adventure-works.com
11004	JOHNSON	ELIZABETH	elizabeth5@adventure-works.com
11005	RUIZ	JULIO	julio1@adventure-works.com
11007	MEHTA	MARCO	marco14@adventure-works.com
11008	VERHOFF	ROBIN	rob4@adventure-works.com
11009	CARLSON	SHANNON	shannon38@adventure-works.com

ProductKey	ProductName	ProductPrice	ModelName
214	Sport-100 Helmet, Red	34.99	Sport-100
215	Sport-100 Helmet, Black	33.64	Sport-100
218	Mountain Bike Socks, M	9.50	Mountain Bike Socks
219	Mountain Bike Socks, L	9.50	Mountain Bike Socks
220	Sport-100 Helmet, Blue	33.64	Sport-100
223	AWC Logo Cap	8.64	Cycling Cap
226	Long-Sleeve Logo Jersey, S	48.07	Long-Sleeve Logo Jersey
229	Long-Sleeve Logo Jersey, M	48.07	Long-Sleeve Logo Jersey
232	Long-Sleeve Logo Jersey, L	48.07	Long-Sleeve Logo Jersey
235	Long-Sleeve Logo Jersey, XL	48.07	Long-Sleeve Logo Jersey
238	HL Road Frame - Red, 62	1,263.46	HL Road Frame

Đây là **Customer Lookup table** cung cấp thêm các diễn giải, thông tin bổ sung cho **customer** (last name, first name, email address, ...)

Đây là **Product Lookup table** cung cấp thông tin bổ sung cho **product** (brand, product name, price, model ...)

Đây là **Data Table** gồm cột “*quantity*”, và có thể kết nối với lookup tables qua cột “*customerkey*” và “*productkey*”

Relationships và Merged tables

Bảng gốc

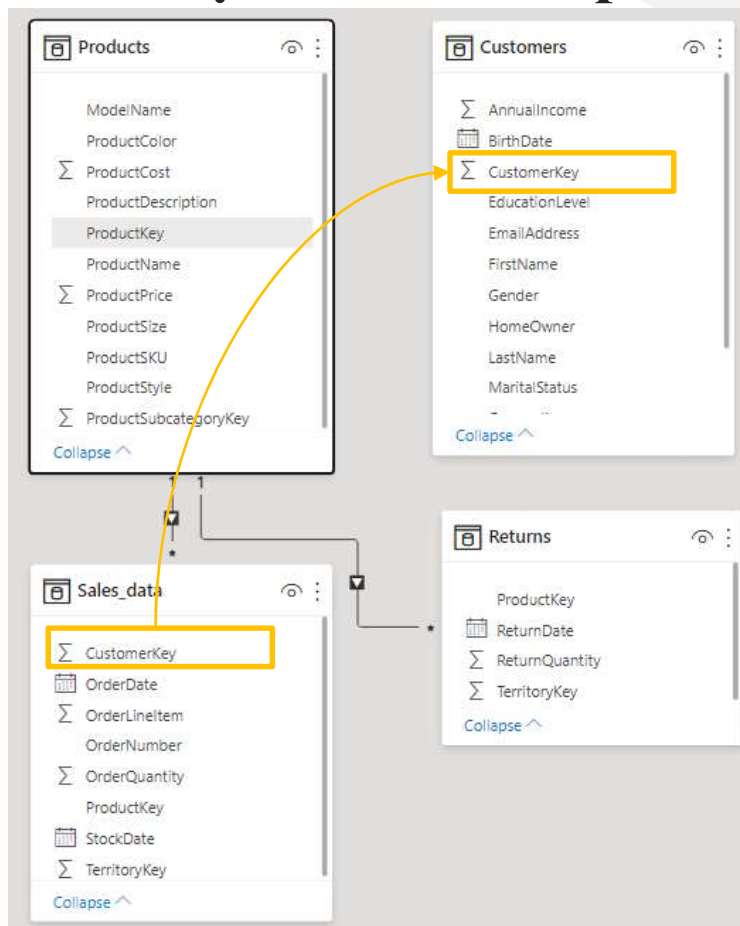
*Bổ sung thông tin về **Calendar Lookup***

*Bổ sung thông tin về **Product Lookup***

date	product_id	quantity	day_of_month	month	year	weekday	month_name	quarter	product_brand	product_name	product_sku	product_weight
1/1/1997	869	5	1	1	1997	Wednesday	January	Q1	Nationeel	Nationeel Grape Fruit Roll	52382137179	17
1/7/1997	869	2	7	1	1997	Tuesday	January	Q1	Nationeel	Nationeel Grape Fruit Roll	52382137179	17
1/3/1997	1	4	3	1	1997	Friday	January	Q1	Washington	Washington Berry Juice	90748583674	8.39
1/1/1997	1472	3	1	1	1997	Wednesday	January	Q1	Fort West	Fort West Fudge Cookies	37276054024	8.28
1/6/1997	1472	2	6	1	1997	Monday	January	Q1	Fort West	Fort West Fudge Cookies	37276054024	8.28
1/5/1997	2	4	5	1	1997	Sunday	January	Q1	Washington	Washington Mango Drink	96516502499	7.42
1/1/1997	76	4	1	1	1997	Wednesday	January	Q1	Red Spade	Red Spade Sliced Chicken	62054644227	18.1
1/1/1997	76	2	1	1	1997	Wednesday	January	Q1	Red Spade	Red Spade Sliced Chicken	62054644227	18.1
1/5/1997	3	2	5	1	1997	Sunday	January	Q1	Washington	Washington Strawberry Drink	58427771925	13.1
1/7/1997	3	2	7	1	1997	Tuesday	January	Q1	Washington	Washington Strawberry Drink	58427771925	13.1
1/1/1997	320	3	1	1	1997	Wednesday	January	Q1	Excellent	Excellent Cranberry Juice	36570182442	16.4

Merging data trong cách này tạo ra **data thừa** và sử dụng **đáng kể bộ nhớ và trình xử lý thông tin** của máy tính hơn là tạo ra các **relationship**

Cách tạo Relationships



Option 1: Click và kéo các trường dữ liệu chung giữa các bảng trong **Relationships** pane

Cách tạo Relationships

Option 2: Thêm hoặc phát hiện relationships sử dụng hộp thoại “**Manage Relationships**”

The screenshot shows the Microsoft Power BI Desktop interface. The 'Home' ribbon is active, and the 'Manage relationships' button is highlighted with a yellow box. A yellow arrow points from this button to the 'Manage relationships' dialog box. In the 'Manage relationships' dialog, the 'New...' button is highlighted with a yellow box. Another yellow arrow points from this button to the 'Create relationship' dialog box. The 'Create relationship' dialog shows a table for 'Customers' and a table for 'Sales_data' with columns like OrderDate, StockDate, OrderNumber, ProductKey, CustomerKey, and TerritoryKey. The cardinality is set to 'One to many (1:*)' and the cross filter direction is 'Single'.

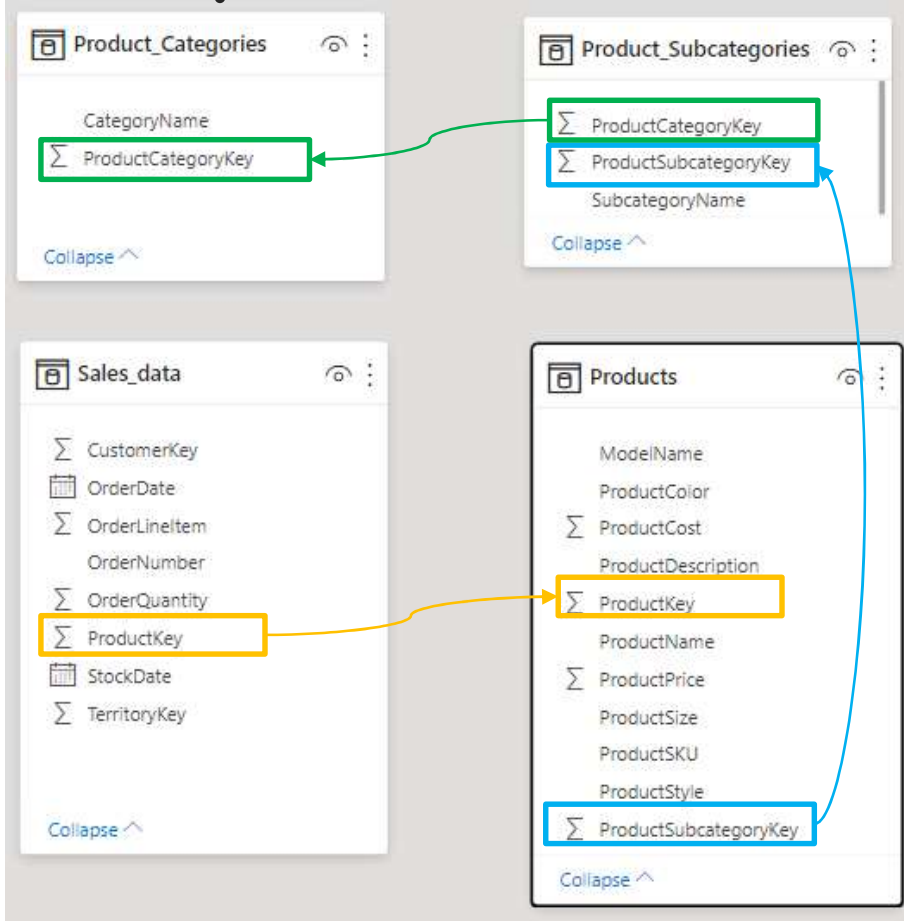
CustomerKey	Prefix	FirstName	LastName	BirthDate	MaritalStatus	Gender	
11009	MR.	SHANNON	CARLSON	Wednesday, 1 April 1964	S	M	shannoni
11106	MR.	JESSIE	LIU	Friday, 11 September 1964	S	M	jessie9@
11109	MR.	RUBEN	KAPOOR	Tuesday, 5 November 1963	S	M	ruben1@

OrderDate	StockDate	OrderNumber	ProductKey	CustomerKey	TerritoryKey	Order
Sunday, 1 January 2017	Friday, 10 January 2003	SO61305	583	20236	9	
Sunday, 1 January 2017	Friday, 31 October 2003	SO61276	488	17252	9	
Sunday, 1 January 2017	Tuesday, 10 June 2003	SO61307	582	20260	9	

Cardinality: One to many (1:*)
Cross filter direction: Single
☒ Make this relationship active
☐ Assume referential integrity
☐ Apply security filter in both directions

OK Cancel

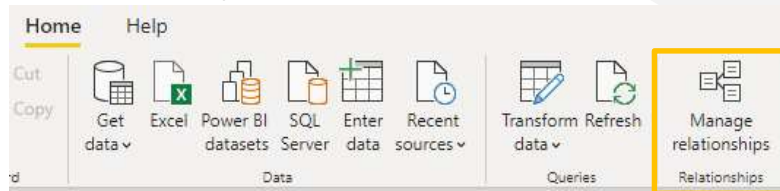
Cách tạo Snowflake



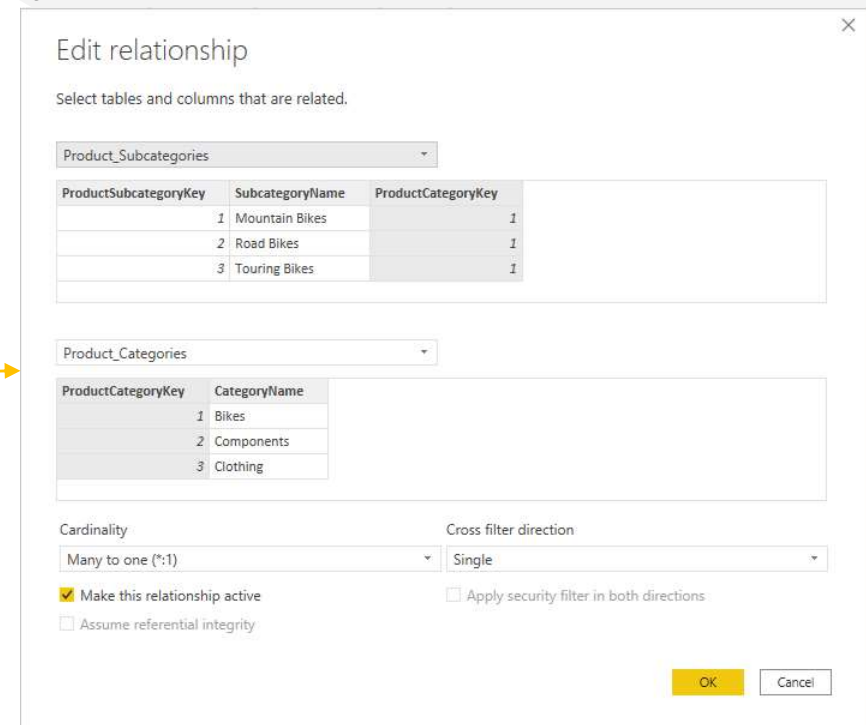
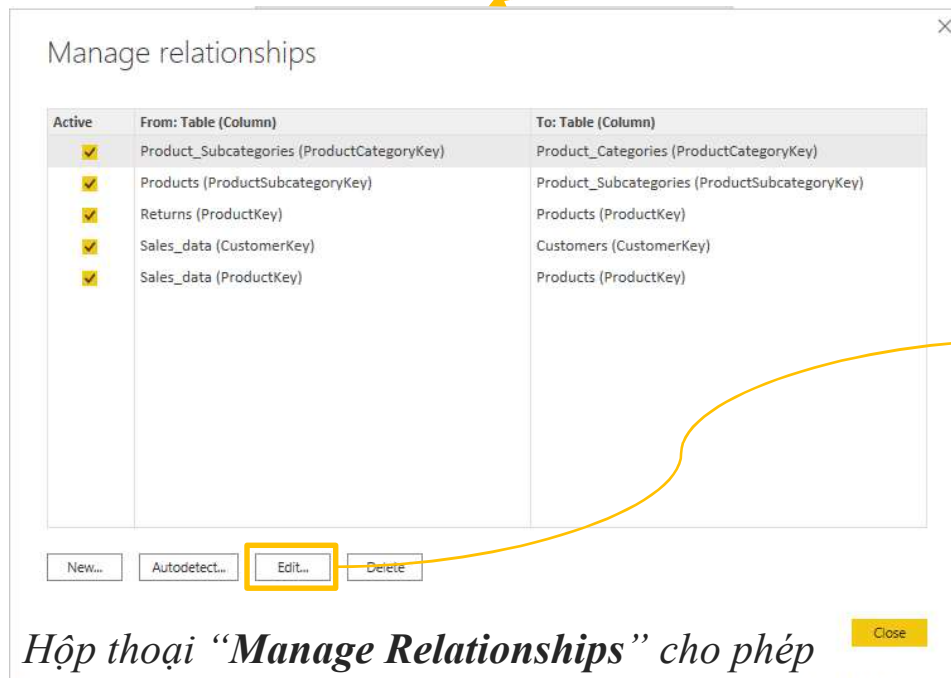
NOTE:

Models với chuỗi như hình bên được gọi là “**snowflake**” (phân biệt chuỗi “**star**” là chuỗi mà lookup tables được kết nối xung quanh một data tables trung tâm)

Cách tạo Snowflake

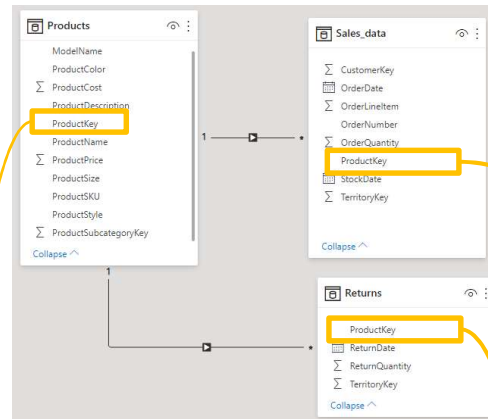


*Editing tools cho phép **activate/deactivate** relationships, xem **cardinality**, và sửa **cross filter direction***



*Hộp thoại “**Manage Relationships**” cho phép thêm, sửa hoặc xóa các relationships*

Sử dụng Fields



ReturnQuantity	OrderQuantity	ProductKey
25	949	484
54	2222	485
8	116	486
25	394	487
13	216	488
15	203	489
8	202	490
5	227	491
93	3121	528
67	2379	529
45	1572	530
39	856	535
28	1132	536
49	736	537
43	1069	538
1828	45314	

ReturnQuantity	OrderQuantity	ProductKey
1828	1218	214
1828	1074	215
1828	1152	220
1828	2394	223
1828	249	226
1828	243	229
1828	238	232
1828	208	235
1828	248	352
1828	225	354
1828	244	356
1828	251	358
1828	246	360
1828	227	362
1828	63	371
1828	45314	

ReturnQuantity	OrderQuantity	ProductKey
155	45314	477
95	45314	480
93	45314	528
77	45314	478
70	45314	214
67	45314	529
66	45314	220
56	45314	479
54	45314	485
52	45314	215
49	45314	537
46	45314	223
45	45314	530
43	45314	538
39	45314	535
1828	45314	

Trong ví dụ, chúng ta có trường **ProductKey** ở cả 3 bảng, và khi chúng ta sử dụng trường để visualization data, đã có 3 kết quả:

- TH1: Chúng ta dùng đúng trường ở Lookup Table, kết quả trả về ĐÚNG
- TH2: Chúng ta dùng trường ở Sale_Data (Data Table 1), Kết quả trả về chỉ ĐÚNG cho Sale_Data, SAI cho Return_Data
- TH3: Chúng ta dùng trường ở Return_Data (Data Table 2), Kết quả trả về chỉ ĐÚNG cho Return_Data, Sai cho Sale_Data

Cài đặt Fields 2 chiều

Edit relationship

Select tables and columns that are related.

Sales_data

OrderDate	StockDate	OrderNumber	ProductKey	CustomerKey	TerritoryKey	Order
Sunday, 1 January 2017	Friday, 10 January 2003	SO61305	583	20236	9	
Sunday, 1 January 2017	Friday, 31 October 2003	SO61276	488	17252	9	
Sunday, 1 January 2017	Tuesday, 10 June 2003	SO61307	582	20260	9	

Products

ProductKey	ProductSubcategoryKey	ProductSKU	ProductName	ModelName	Product
214	31	HL-U509-R	Sport-100 Helmet, Red	Sport-100	Universal fit, well-vent
215	31	HL-U509	Sport-100 Helmet, Black	Sport-100	Universal fit, well-vent
220	31	HL-U509-B	Sport-100 Helmet, Blue	Sport-100	Universal fit, well-vent

Cardinality

Many to one (*:1)

☒ Make this relationship active

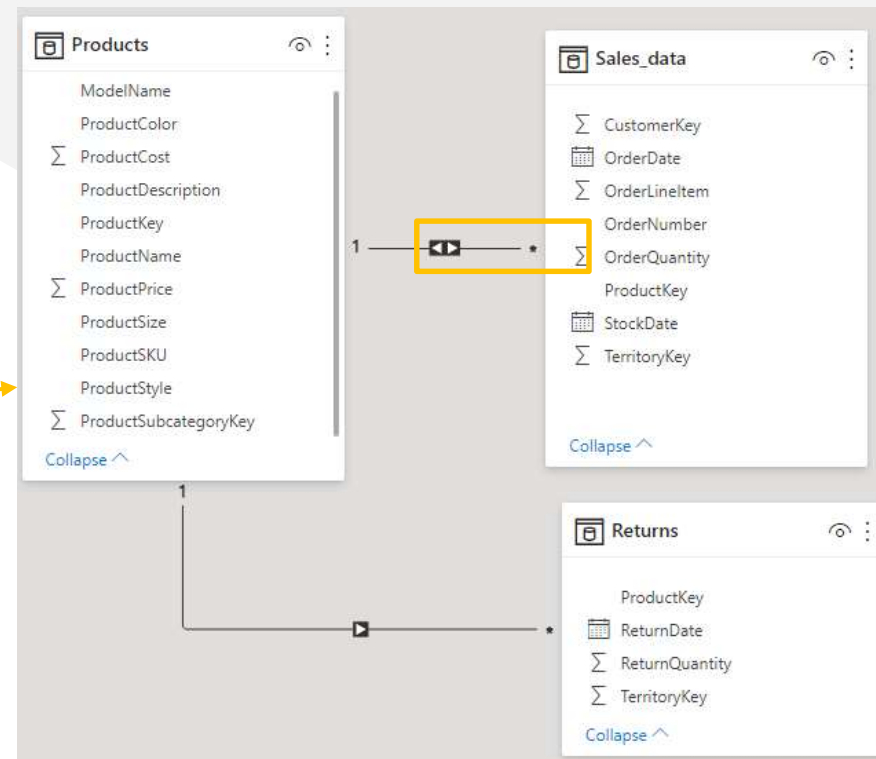
☐ Assume referential integrity

Cross filter direction

Both

☐ Apply security filter in both directions

OK Cancel



Một số lưu ý khi làm việc với Data Model

Xây dựng một Model tiêu chuẩn ngay từ khi bắt đầu

- *Xây dựng một Model tiêu chuẩn ngay từ khi bắt đầu*
- *Sử dụng Relationship thay cho Merge Table; bảng dài và hẹp luôn tốt hơn bảng ngắn và rộng*

Tổ chức lookup tables phía *trên* data tables

- *Điều này luôn nhắc cho chúng ta về flow kết nối và phân loại các loại table*

Tránh những relationship phức tạp nếu không cần thiết

- *Hạn chế sử dụng filter 2 chiều, chỉ nên sử dụng filter 1 chiều*