



# Chương 4: Mô hình dữ liệu đa chiều



## Nội dung chương

- Các khái niệm chính của mô hình dữ liệu đa chiều
  - Dữ kiện (Fact)
  - Chiều (Dimension)
  - Độ thô
  - Tổng hợp
- Các mô hình lưu trữ



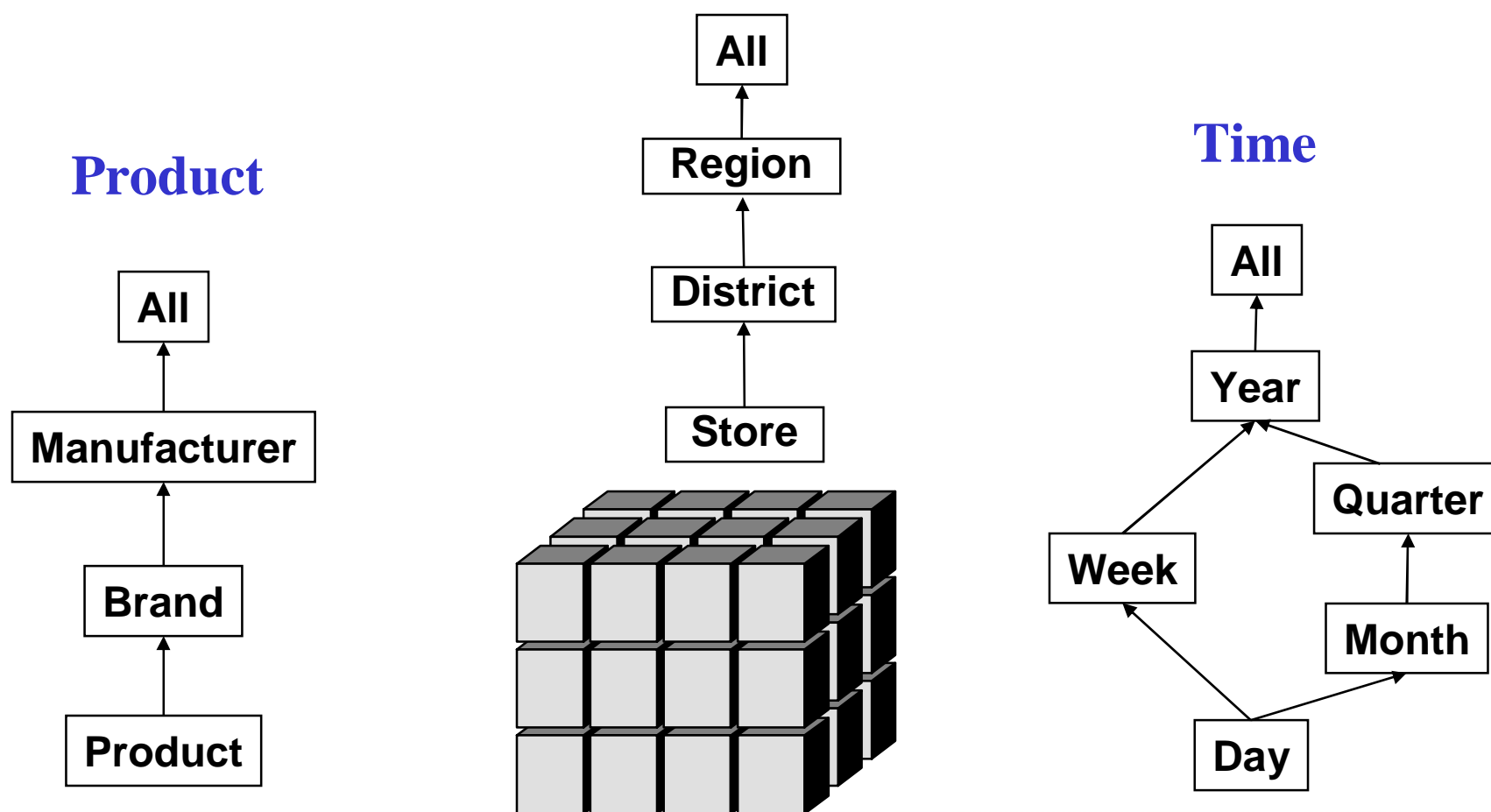
## Mô hình dữ liệu đa chiều

- Được đề xuất và thiết kế cho một mục đích phân tích dữ liệu
- Mô hình dữ liệu này không phù hợp cho hệ thống OLTP
- Mô hình dữ liệu này được thao tác bởi các công cụ OLAP
  - Các công cụ này cung cấp các phương tiện truy vấn mạnh dựa trên thiết kế mô hình dữ liệu đa chiều
  - Ví dụ như: TARGET Analysis, SQL OLAP Server



# Mô hình dữ liệu đa chiều

## Location





## Các thành phần chính

- Các dữ kiện (Facts)
  - Miêu tả các vùng kinh doanh
  - Không thay đổi khi nó đã được sinh ra
  - Được lưu tại một cấp thô nào đó
- Các chiều (Dimensions)
  - Thông tin tham chiếu qua đó các dữ kiện có thể được cấu trúc cho việc phân tích
  - Định nghĩa các phân cấp
- Và các khối đa chiều (Cubes)



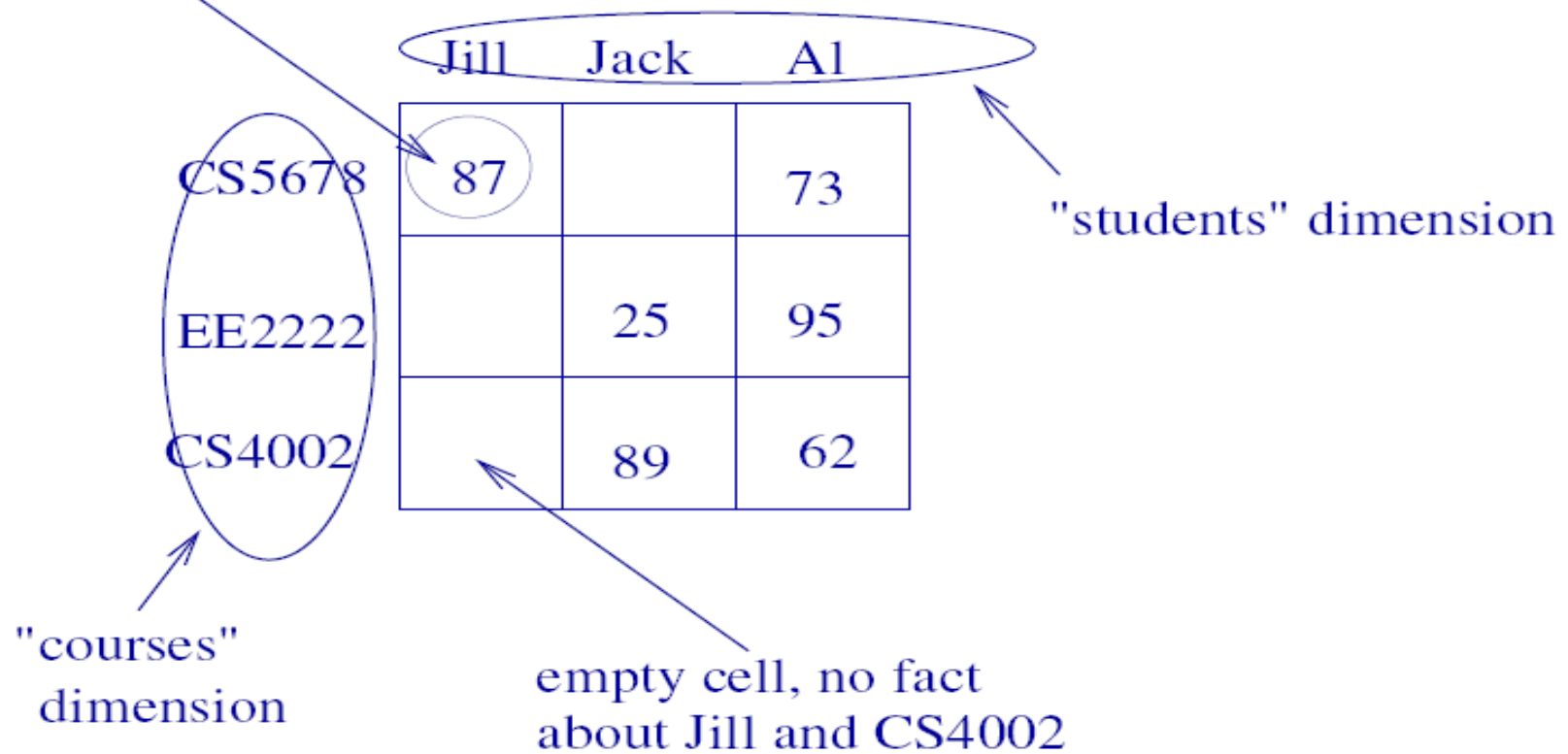
# Khối

- Một khối có thể có nhiều chiều
  - Nếu có nhiều hơn 3 chiều sẽ được gọi là ‘siêu khối’ (“hypercube”)
  - Về mặt lý thuyết thì số chiều là không hạn chế
  - Thường thì số chiều là từ 4 đến 12
- Một khối bao gồm nhiều ô dữ liệu
  - Là một liên kết giữa các giá trị của chiều
  - Một ô có thể là rỗng (không có dữ liệu cho liên kết này)
  - Khối thưa (**sparsecube**): có nhiều ô rỗng
  - Khối dày đặc (**densecube**): có ít ô rỗng



# Ví dụ về ô rỗng

measure value, for fact "Jill got 87 in CS5678"





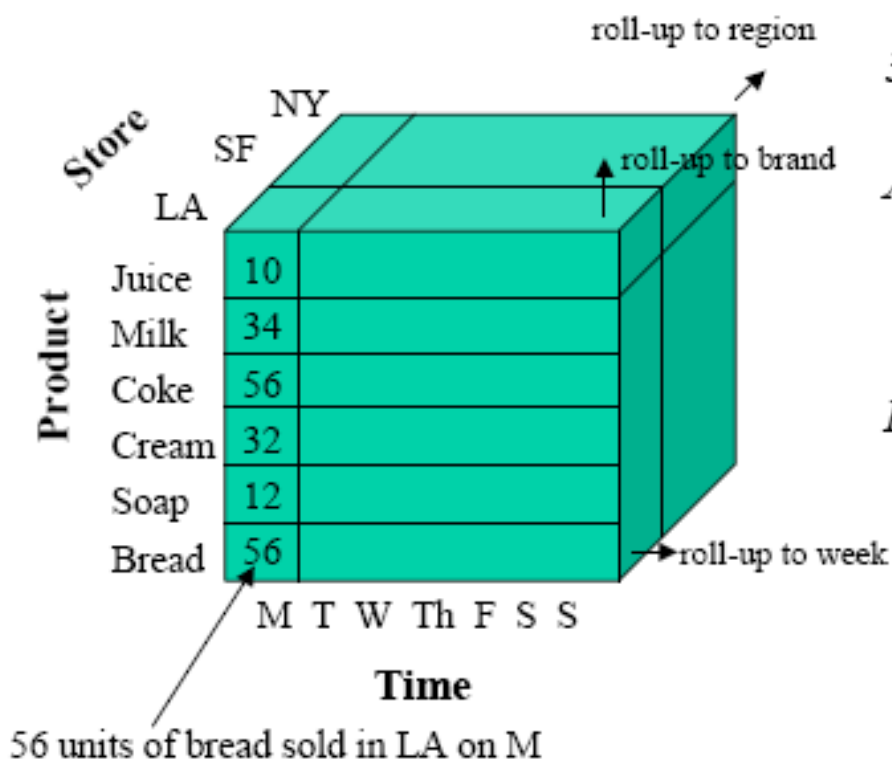
## Ví dụ về dữ kiện và chiều

- Kho Grocery
- Dữ kiện:
  - POS: point of sales sự kiện bán hàng
- Chiều:
  - Thời gian (Time)
  - Vị trí Kho (Location - Store)
  - Sản phẩm (Product)
  - Quảng cáo (Promotion)





# Khối Grocery



*3 Dimensions:*

Time, Product, Store

*Attributes:*

Product (upc, price, ...)

Store ...

...

*Hierarchies:*

Product → Brand → ...

Day → Week → Quarter

Store → Region → Country



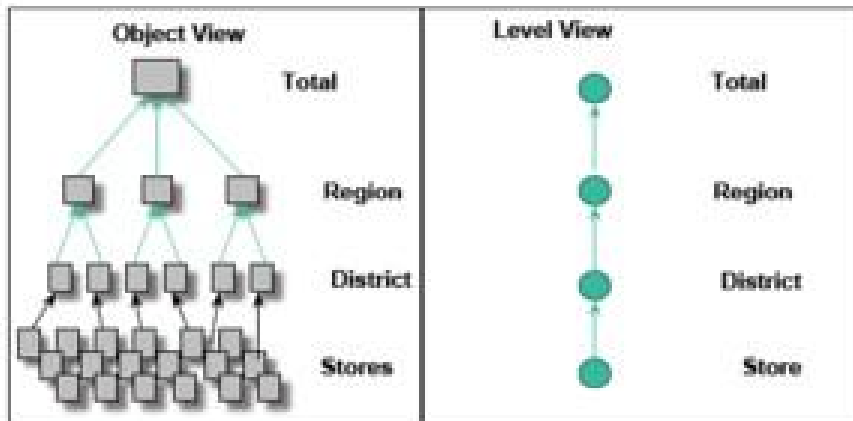
## Các khái niệm về chiều

- Miền phân cấp
- Các cấp (Levels), cấp tương ứng với các độ thô
- Lược đồ chiều Dimension schema
- Các toán tử chiều Dimension operators

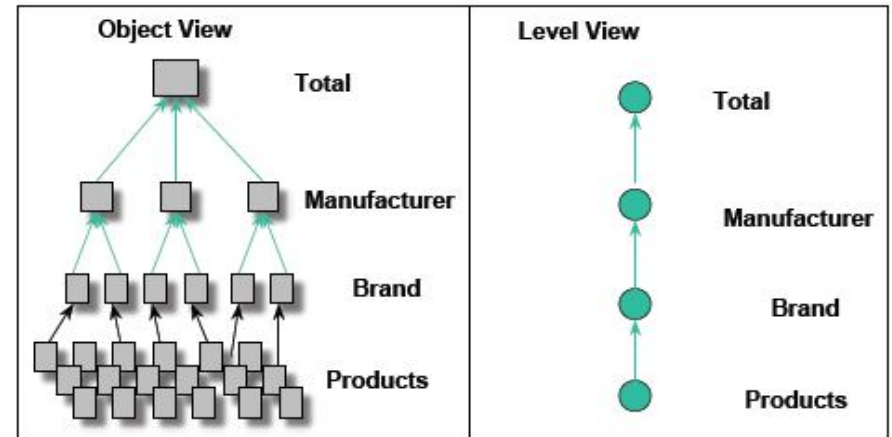


# Các ví dụ về chiều

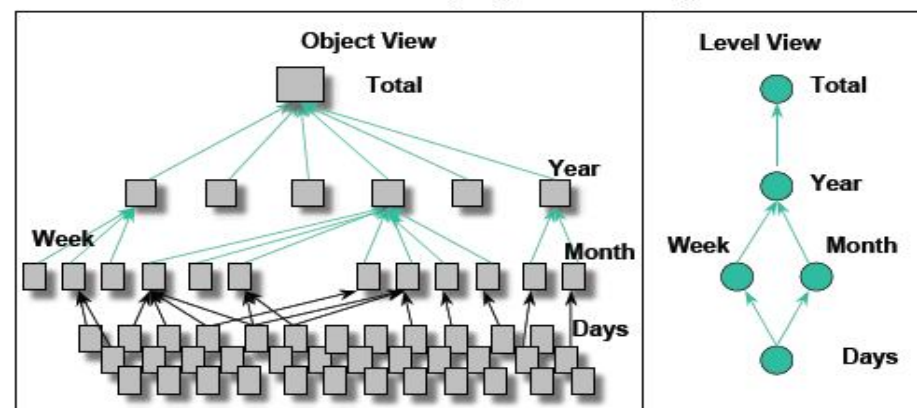
Location Dimension



Product Dimension



Time Dimension (many arrows missing)



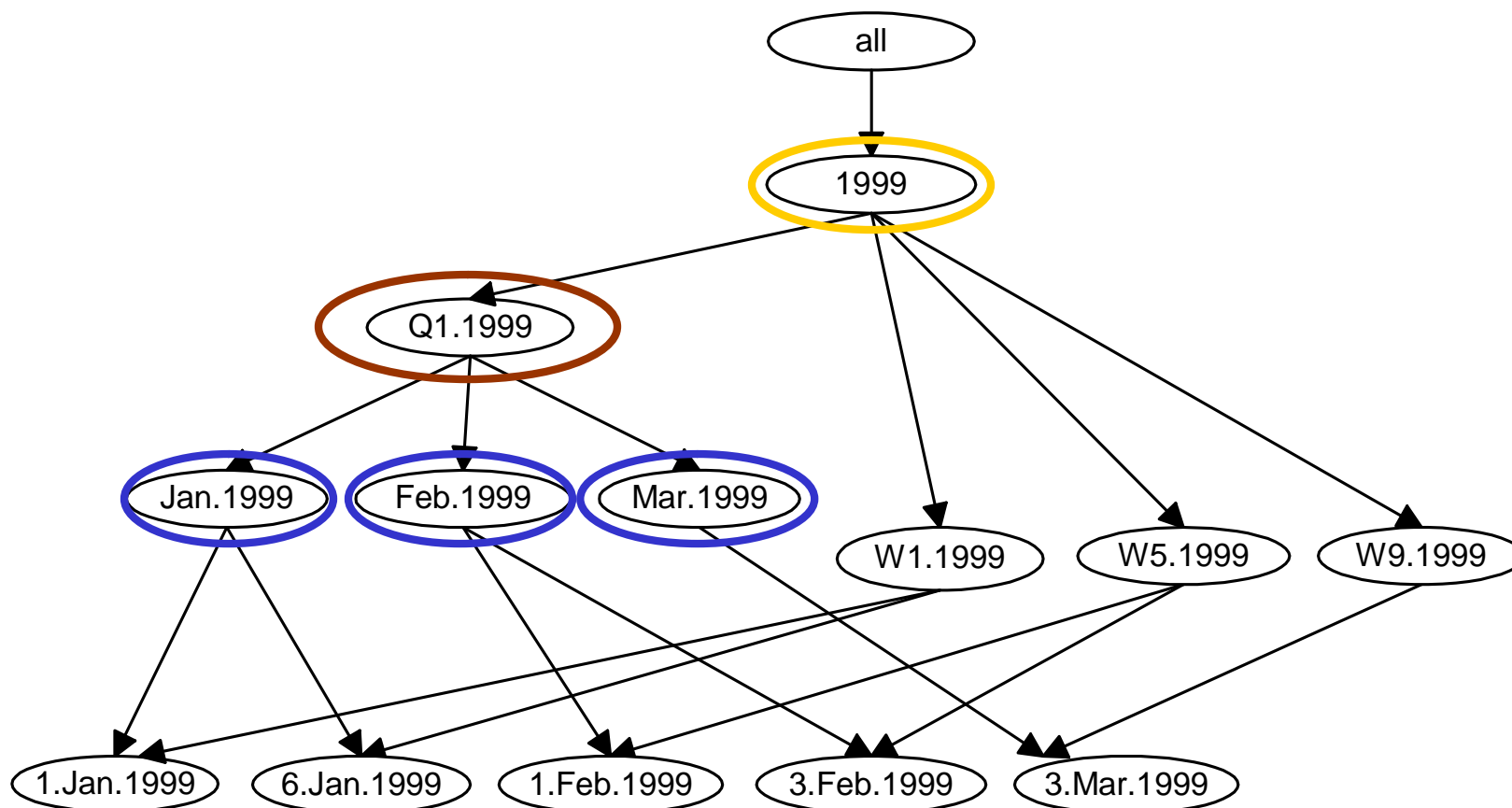


# Thuộc tính chiều

- Một chiều bao gồm nhiều thuộc tính. Ví dụ như:
  - Time dimension: day, month, year
  - Product dimension: ProductID, LineID, BrandID
- Một chiều được cấu trúc/ tổ chức ở dạng phân cấp:
  - Time dimension: day → week → quarter
  - Product dimension: product → brand → manufacturer
- Các thuộc tính xác định cấp nào mà một phần tử chiều thuộc vào
- Các phần tử time có cùng năm, tháng thuộc về cấp 'month'
- Tất cả phần tử product có cùng brand thuộc về cấp 'brand'



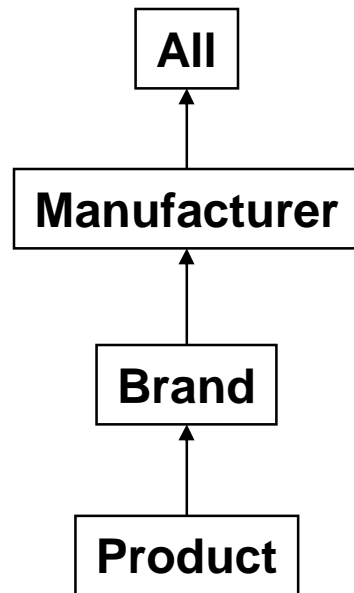
# Ví dụ về vùng phân cấp



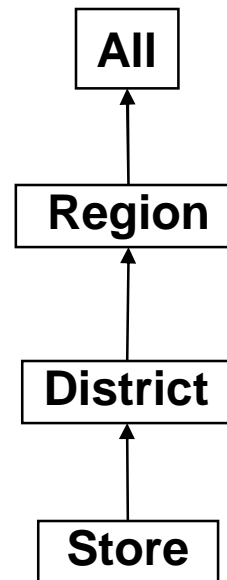


# Lược đồ chiều và phân cấp

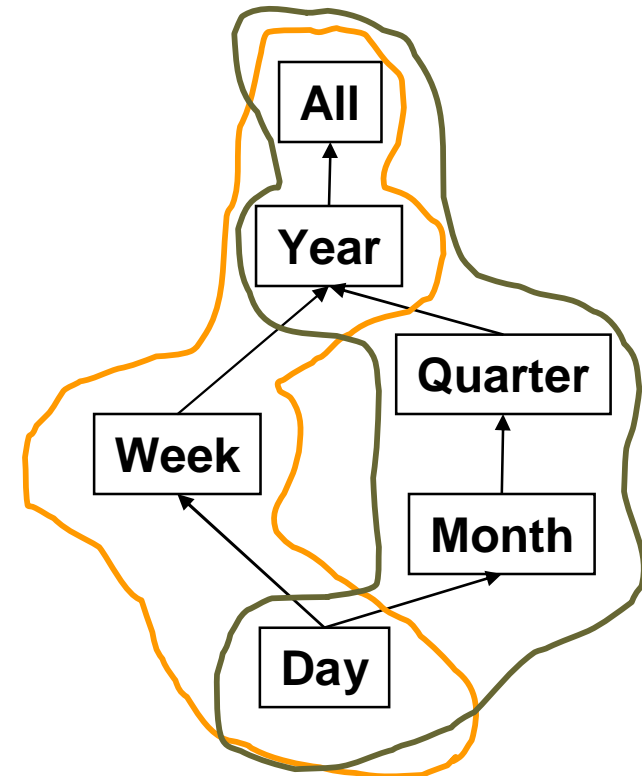
## Product



## Geography

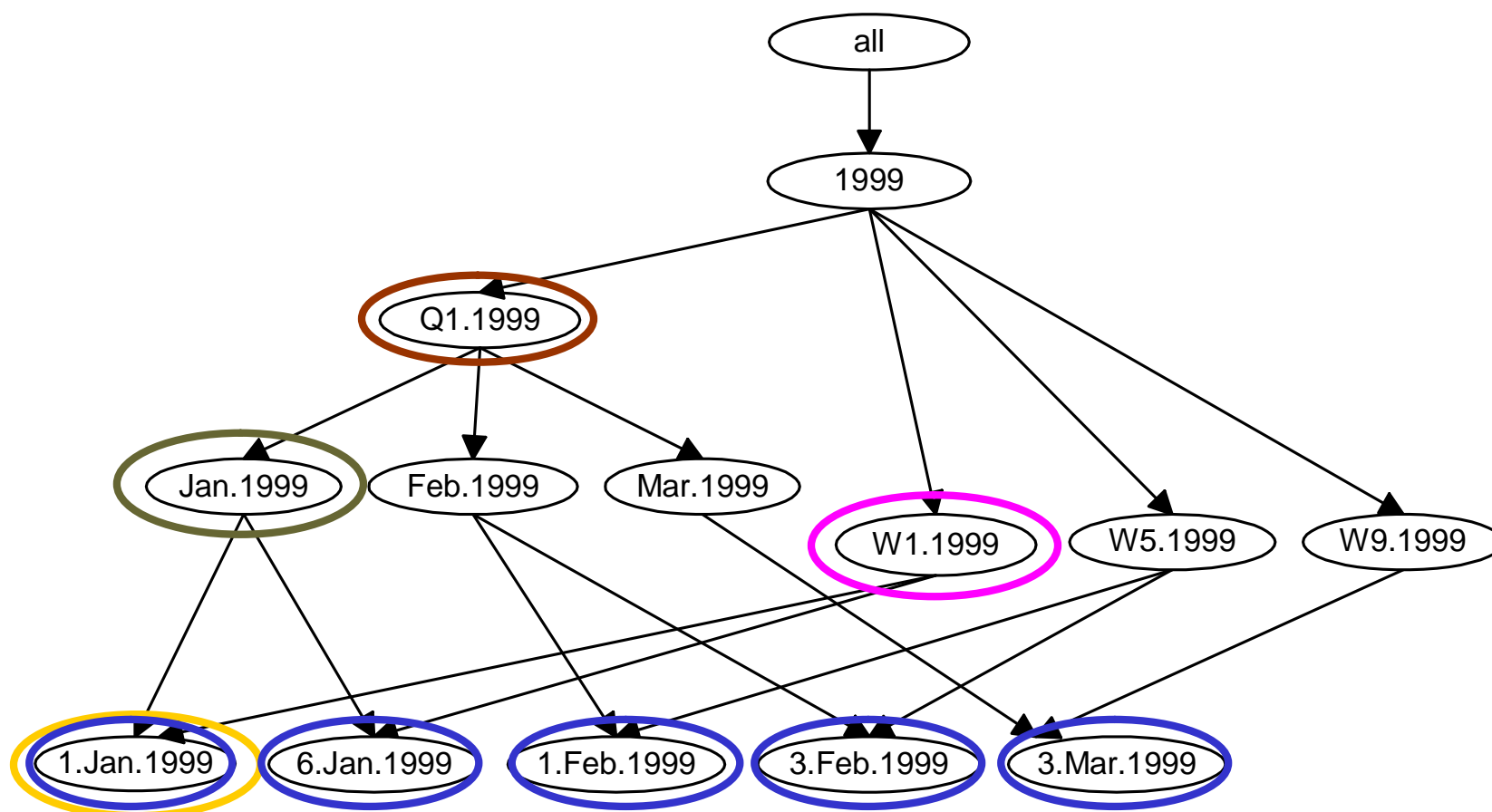


## Time



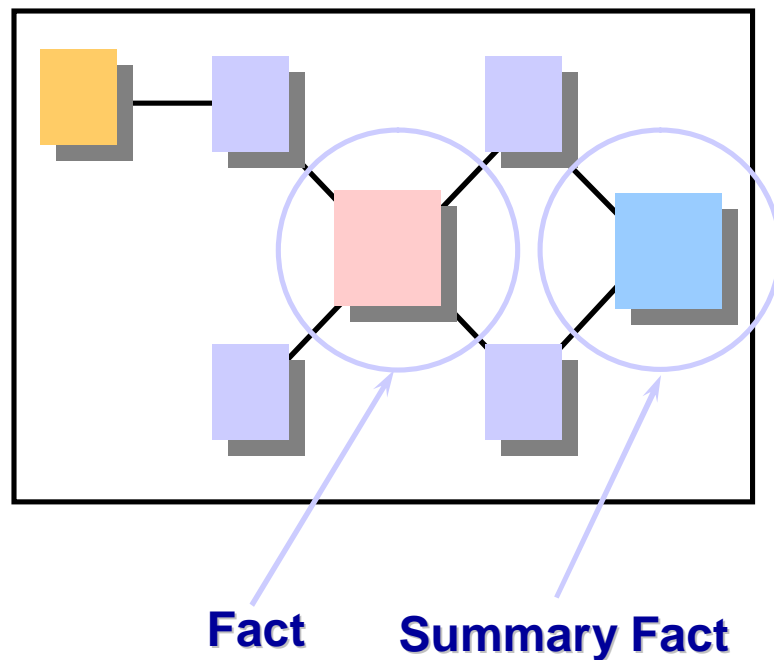


# Các toán tử chiều



## Dữ kiện (Fact Data)

- Các dữ kiện số (Numerical measures)
- Được truy cập bởi các chiều

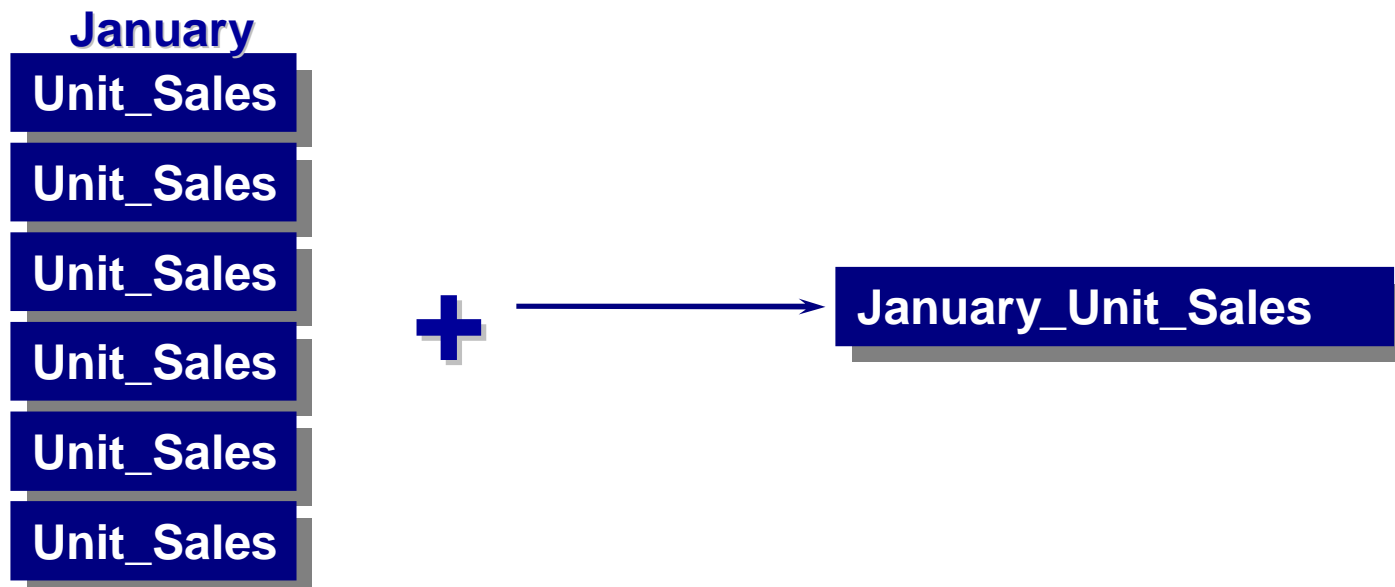
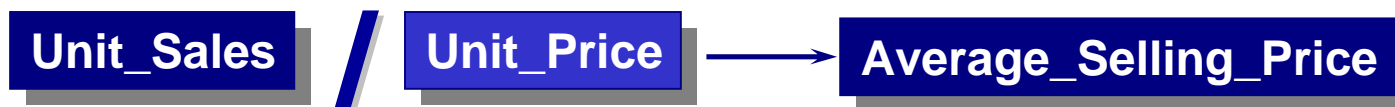






## Dữ kiện (Fact Data)

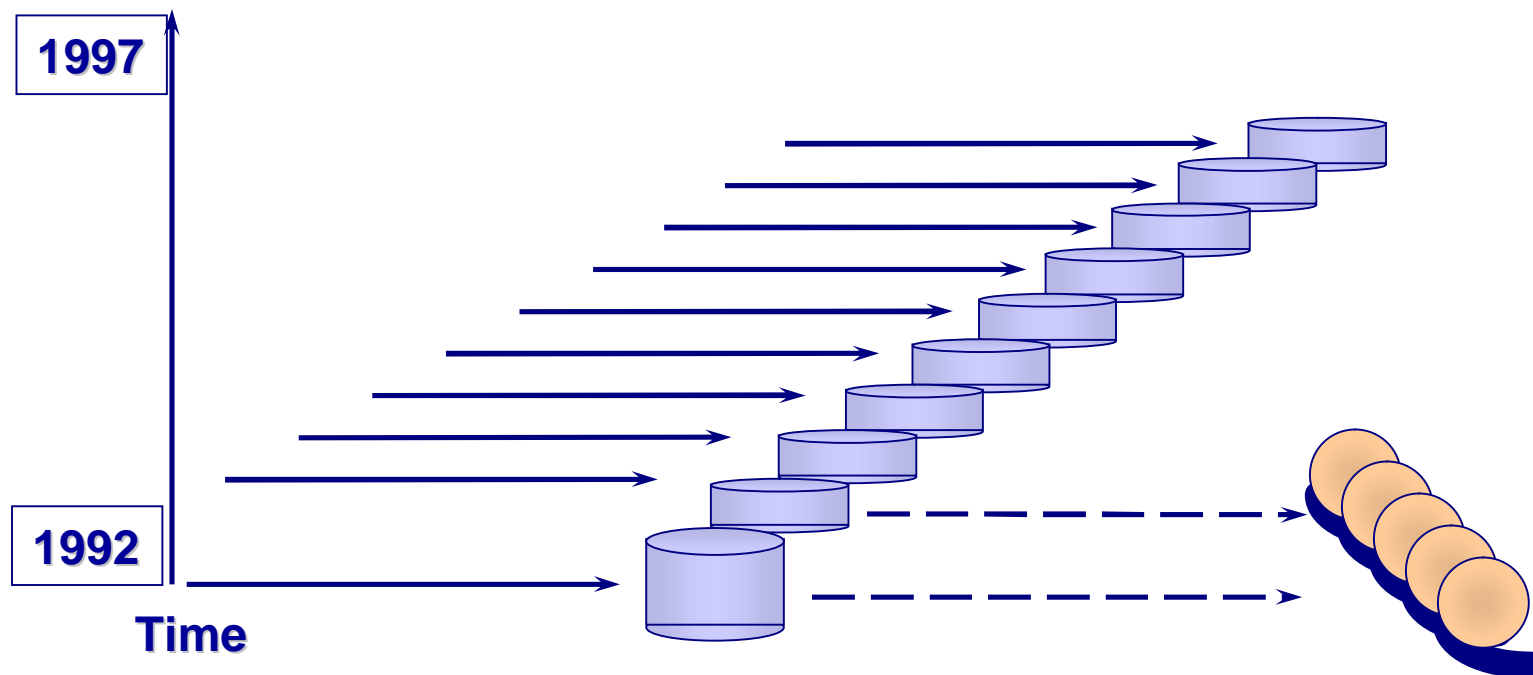
- Chứa dữ kiện suy dẫn





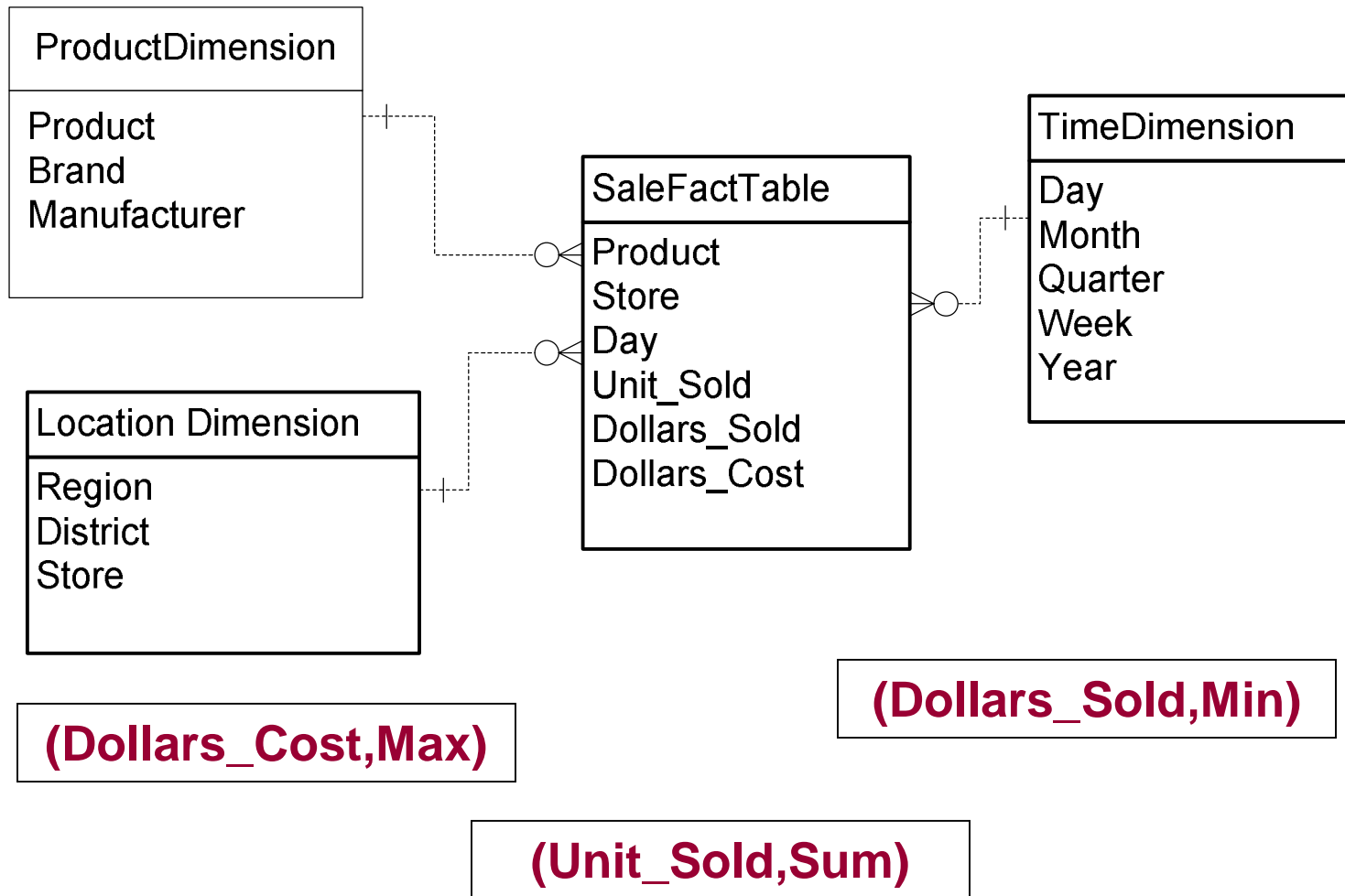
## Dữ kiện (Fact Data)

- Các bảng chứa dữ liệu lớn
- Dữ liệu được dán nhãn thời gian



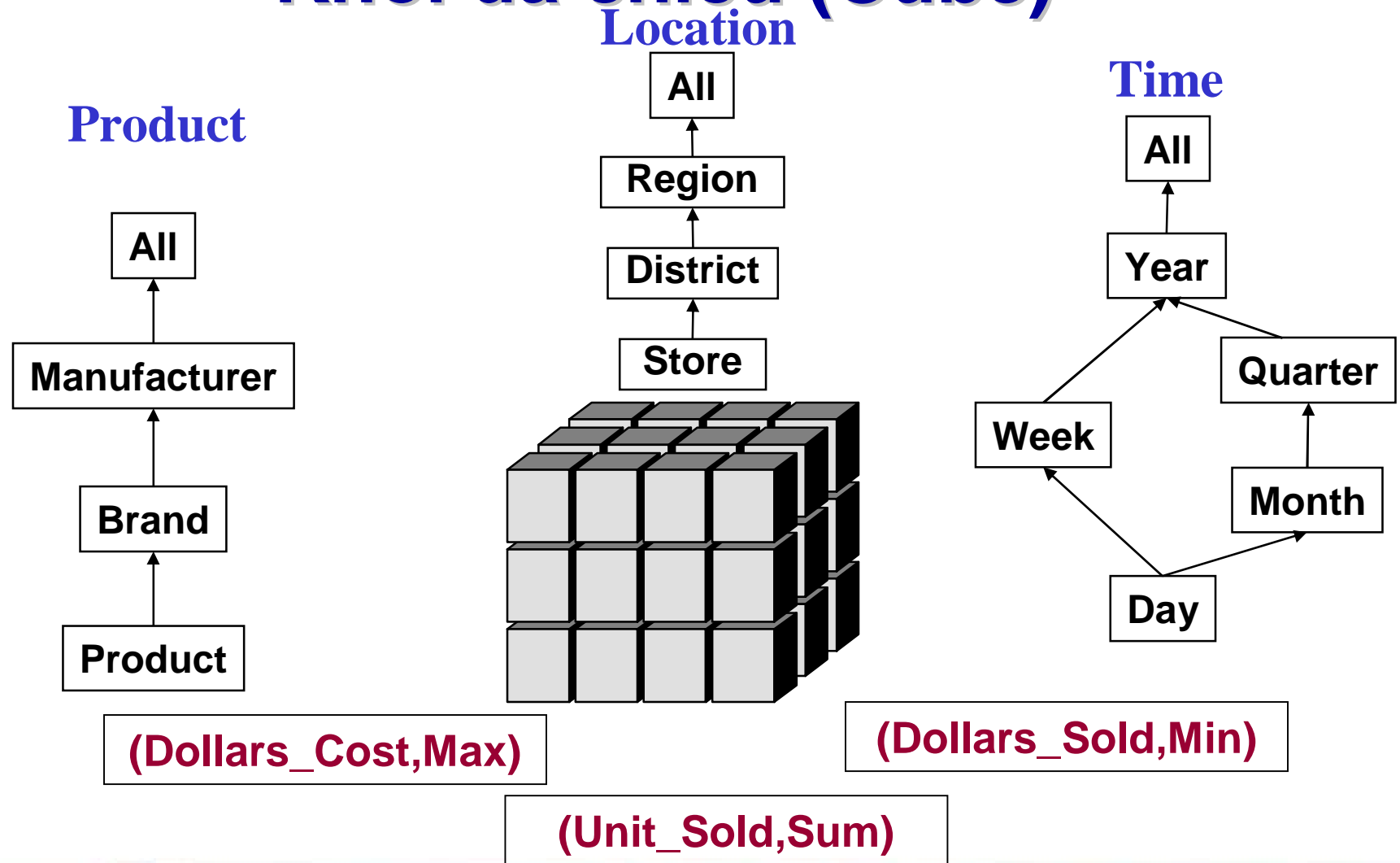


# Ví dụ về dữ kiện





# Khối đa chiều (Cube)



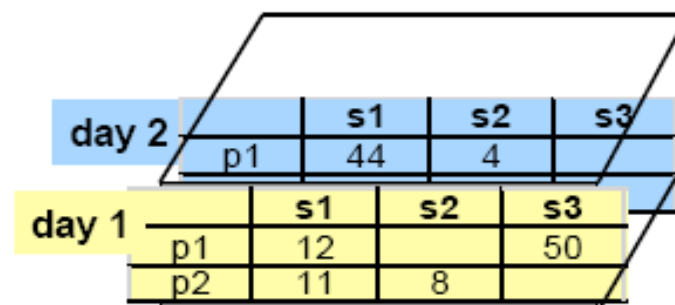


## Ví dụ về khối 3 chiều

Fact table view:

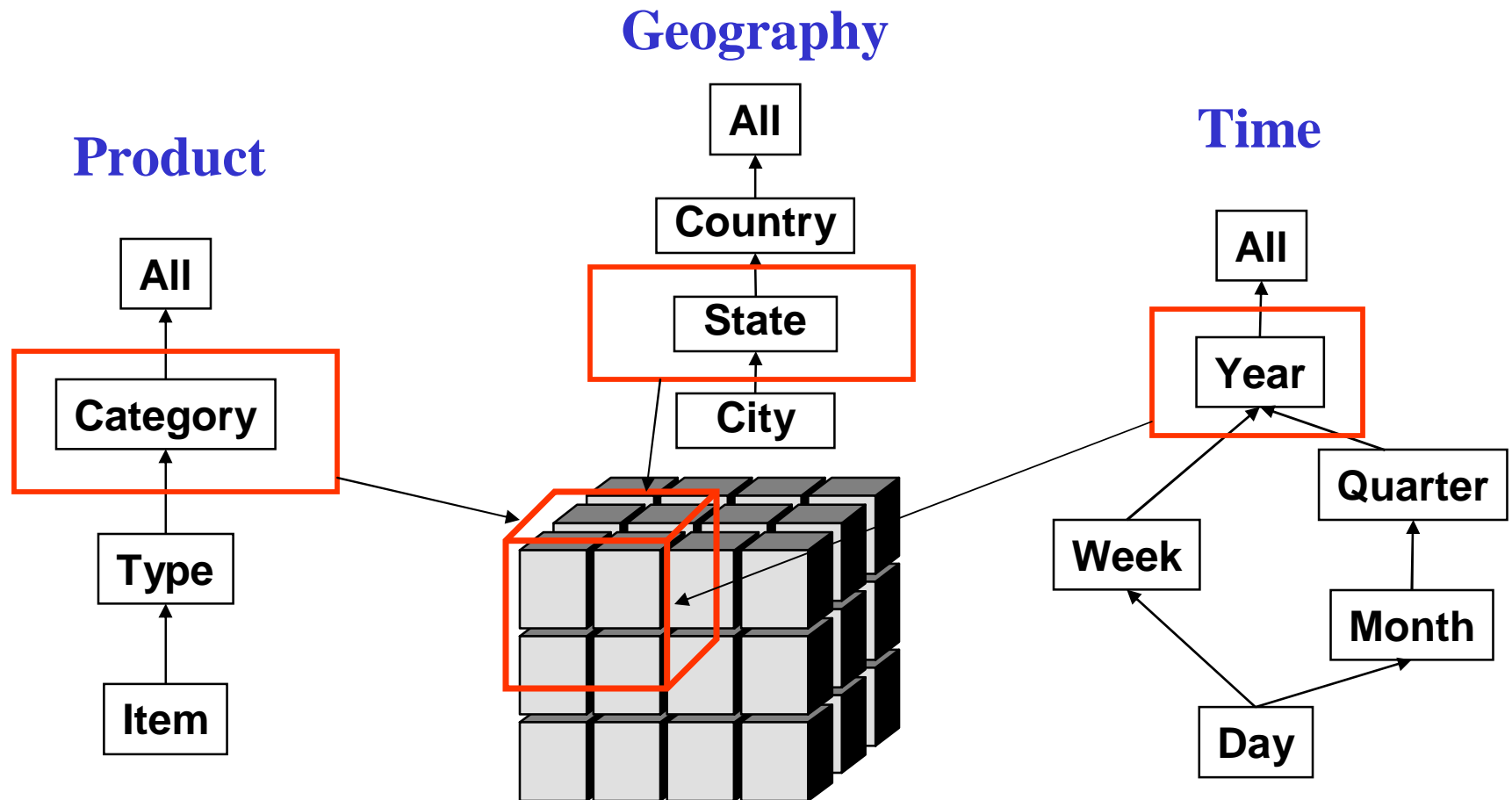
sale	prodlid	storeld	date	amt
	p1	s1	1	12
	p2	s1	1	11
	p1	s3	1	50
	p2	s2	1	8
	p1	s1	2	44
	p1	s2	2	4

Multi-dimensional cube:



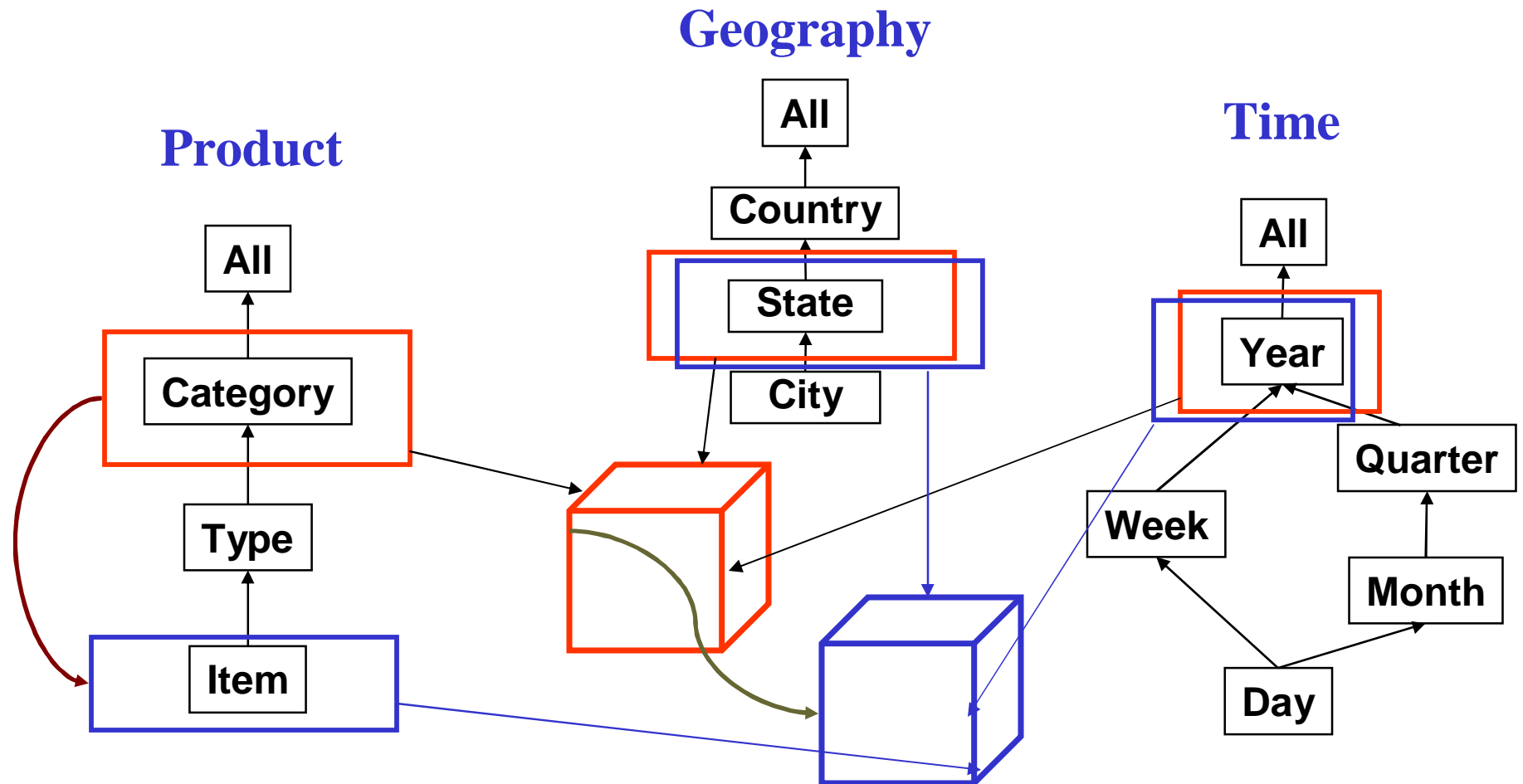


# Ví dụ về tổng hợp (Aggregation)



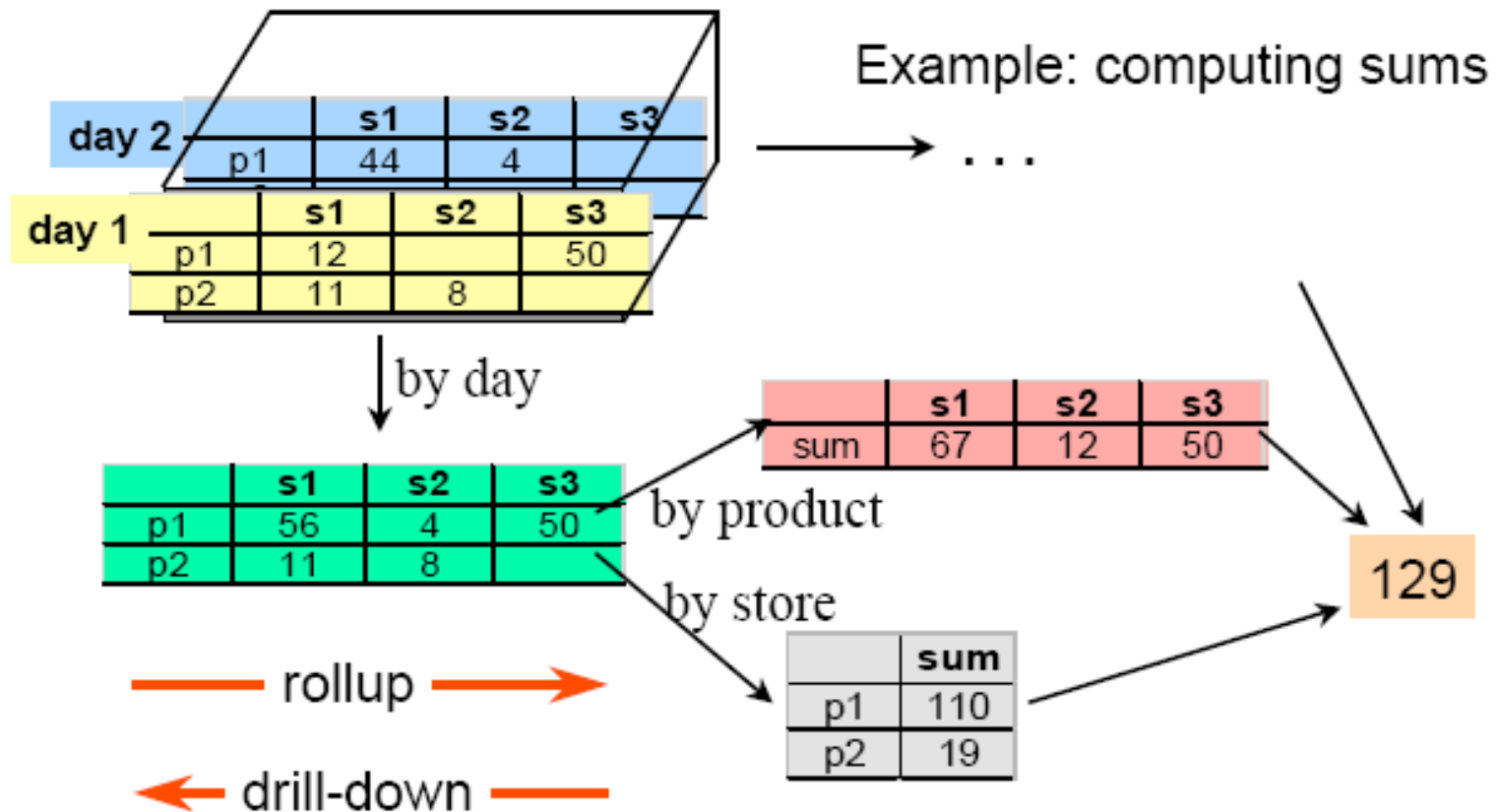


# Ví dụ của chuyển dịch





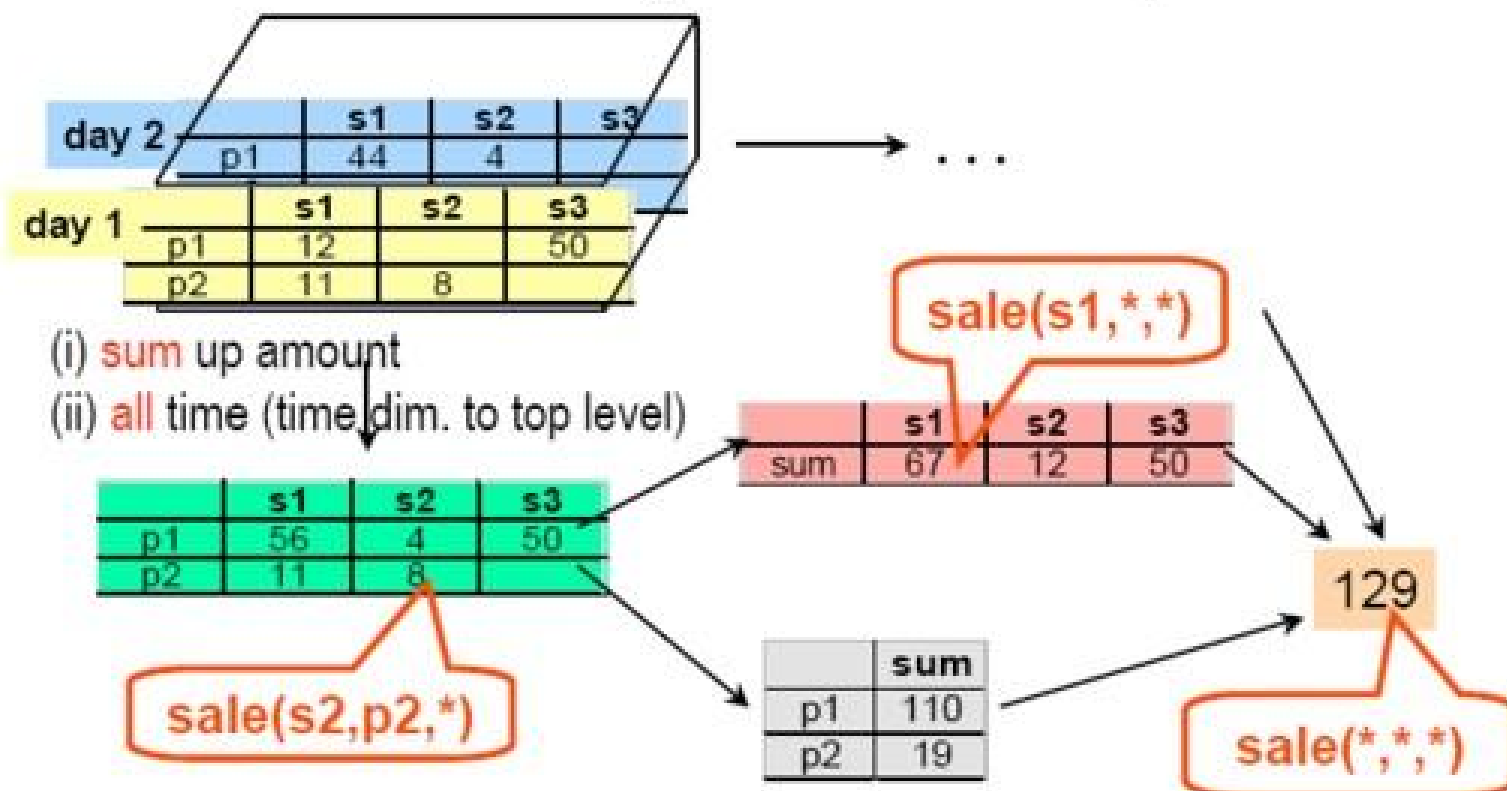
# Ví dụ: rolling up và drilling down





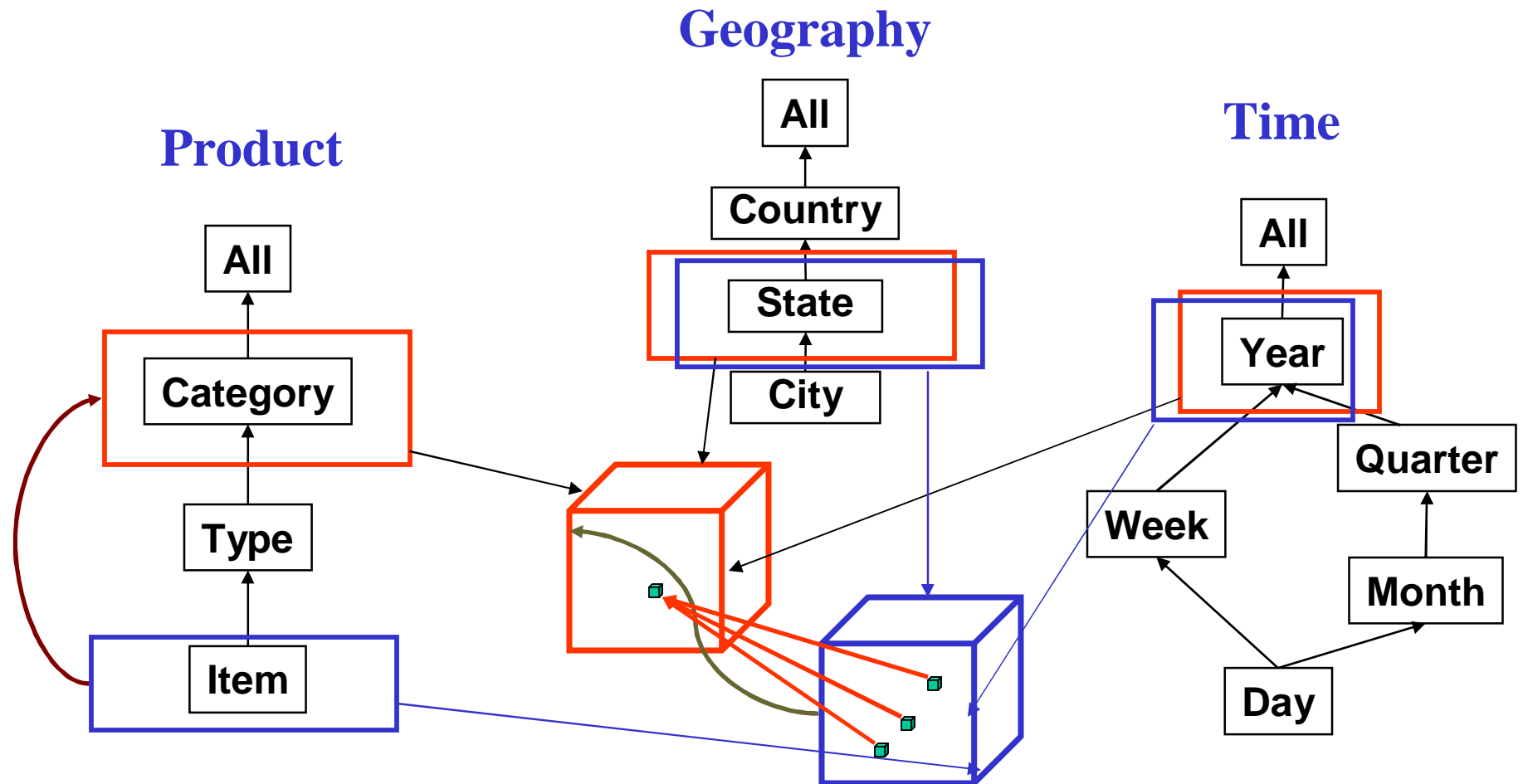
# Ví dụ của rolling up

Parameters of Roll-up: (i) aggregation function for measures  
(ii) which dimension up to which level



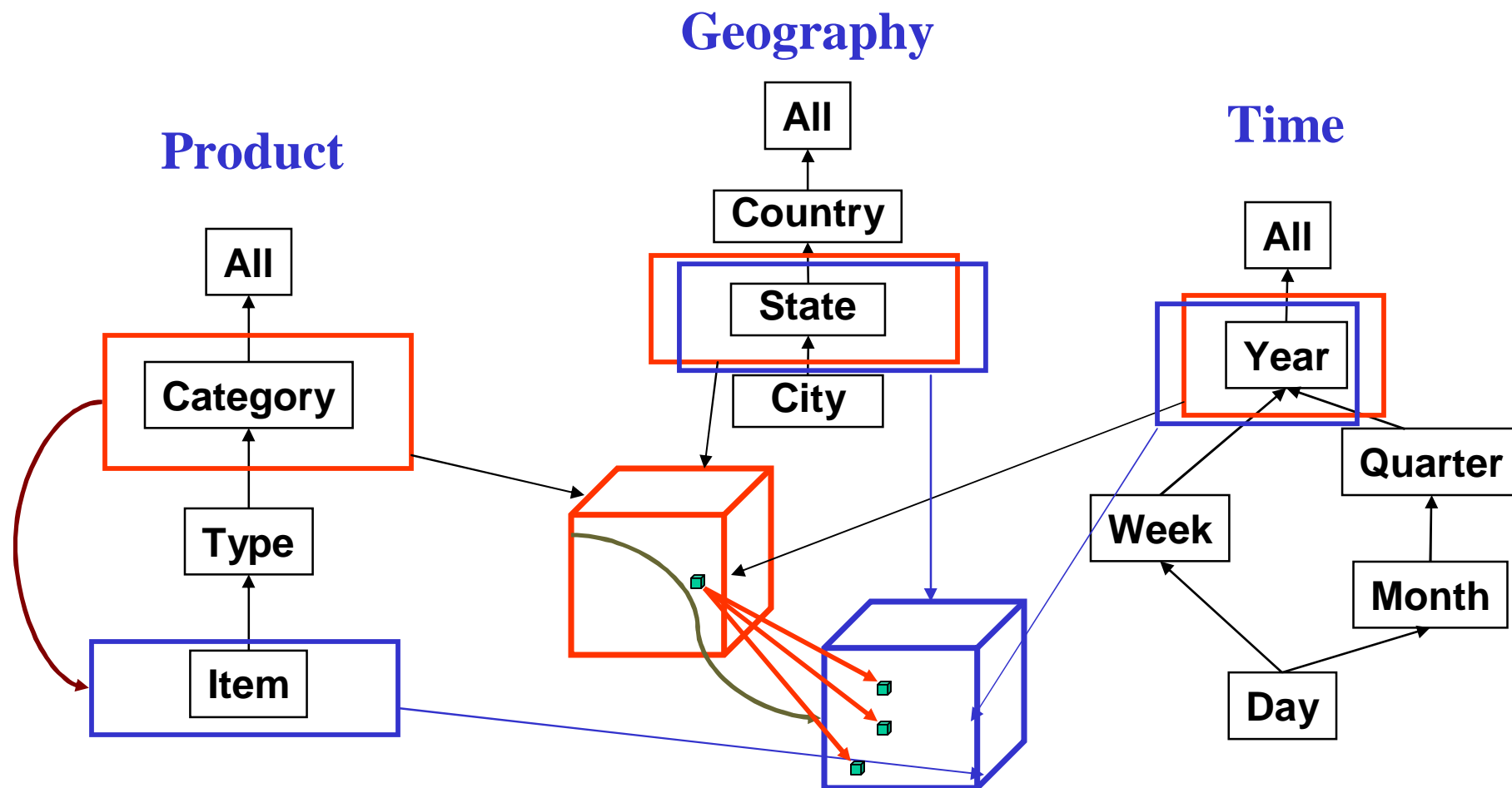


# Ví dụ của rolling up



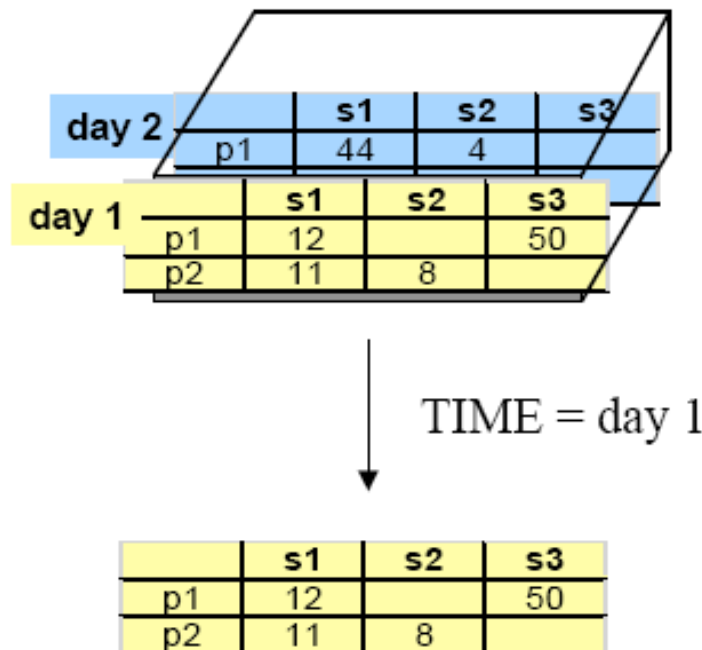


# Ví dụ của Drilling down





# Ví dụ về Slice và Pivot



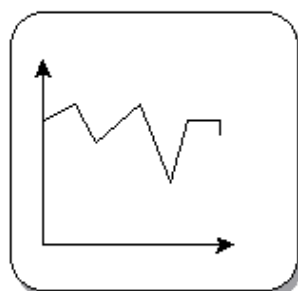
		Sales (\$ millions)		
	Products	Time		
		d1	d2	
Store s1	Electronics	\$5.2		
	Toys	\$1.9		
	Clothing	\$2.3		
	Cosmetics	\$1.1		
Store s2	Electronics	\$8.9		
	Toys	\$0.75		
	Clothing	\$4.6		
	Cosmetics	\$1.5		

		Sales (\$ millions)	
	Products	d1	
		Store s1	Store s2
Store s1	Electronics	\$5.2	\$8.9
	Toys	\$1.9	\$0.75
	Clothing	\$2.3	\$4.6
	Cosmetics	\$1.1	\$1.5
Store s2	Electronics		
	Toys		
	Clothing		



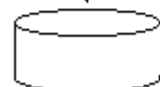
# ROLAP

Presentation Layer  
(Clientseite)



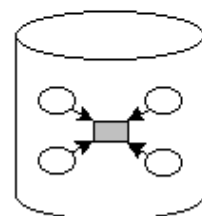
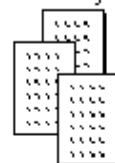
Applicationlayer  
(Serverside)

Applications-server



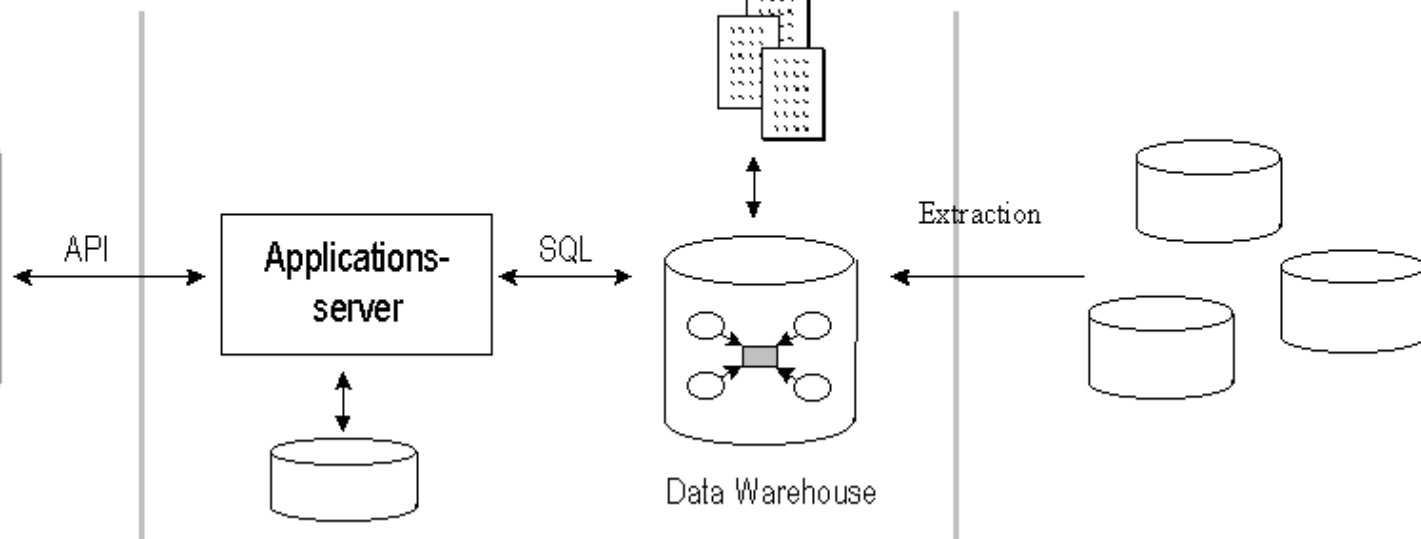
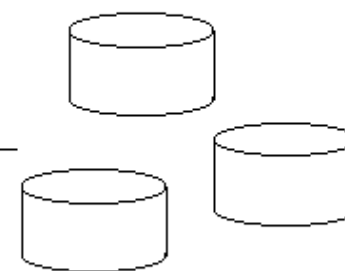
Metadata

Summary Tables



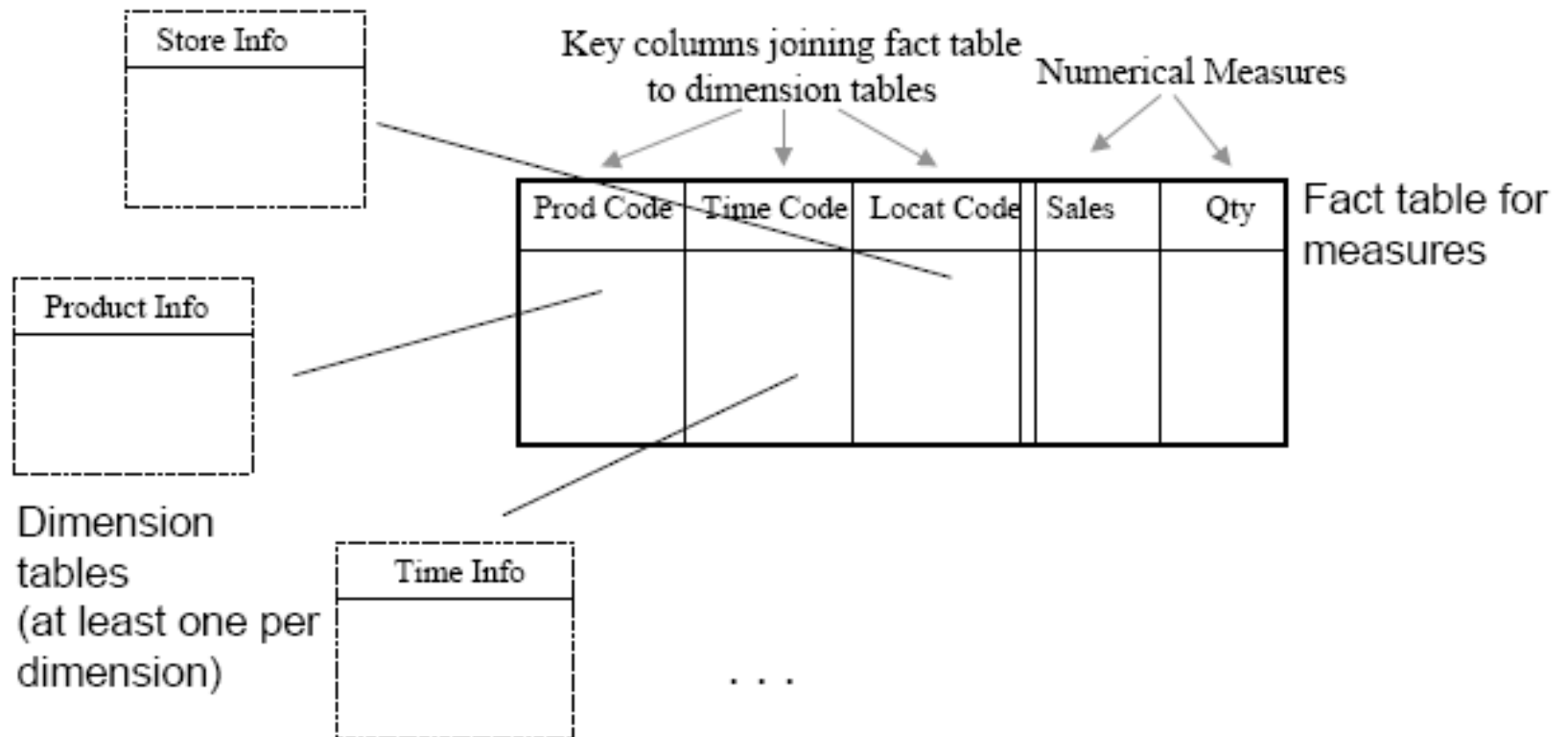
Data Warehouse

Operational Database  
Layer





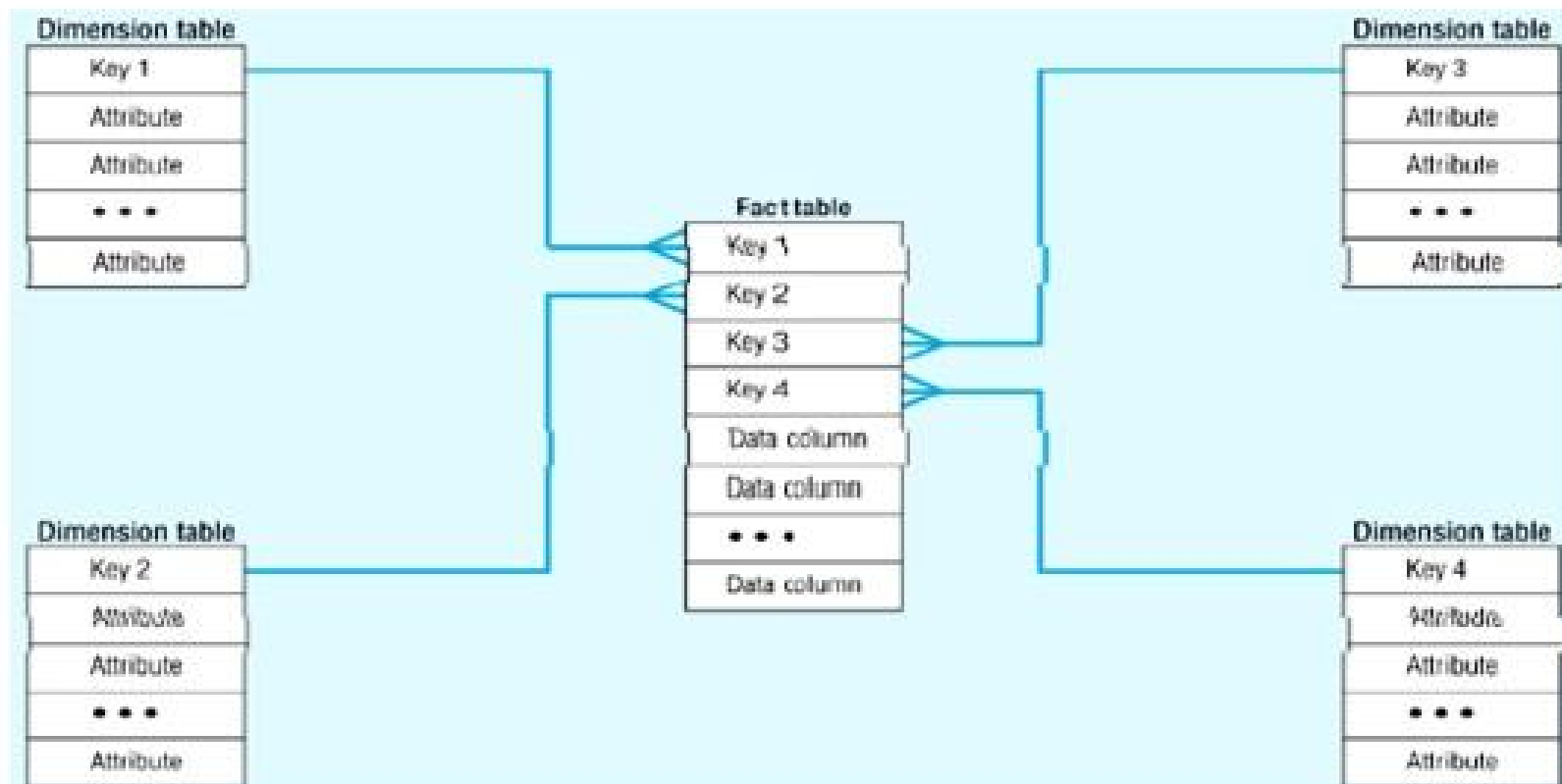
# Mô hình dữ liệu đa chiều trong csdl quan hệ





# Lược đồ hình sao tổng quát

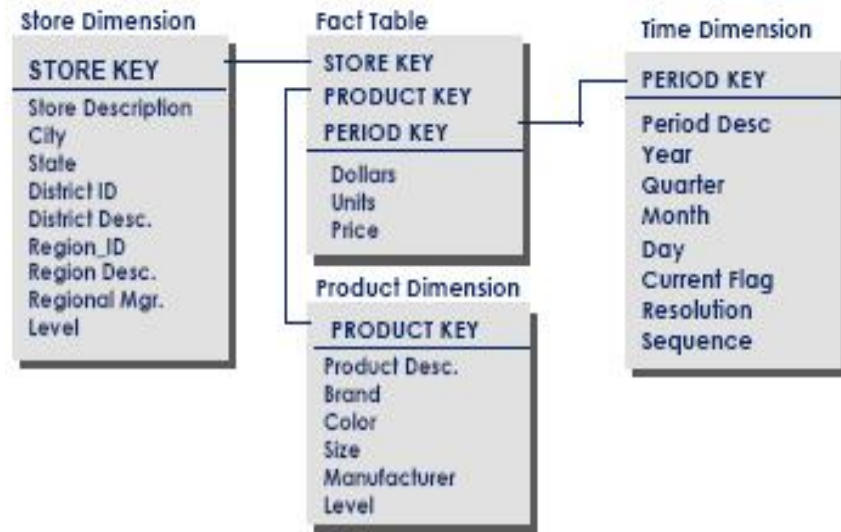
generic schema: 4 dimensions, numerous measures





# Lược đồ hình sao

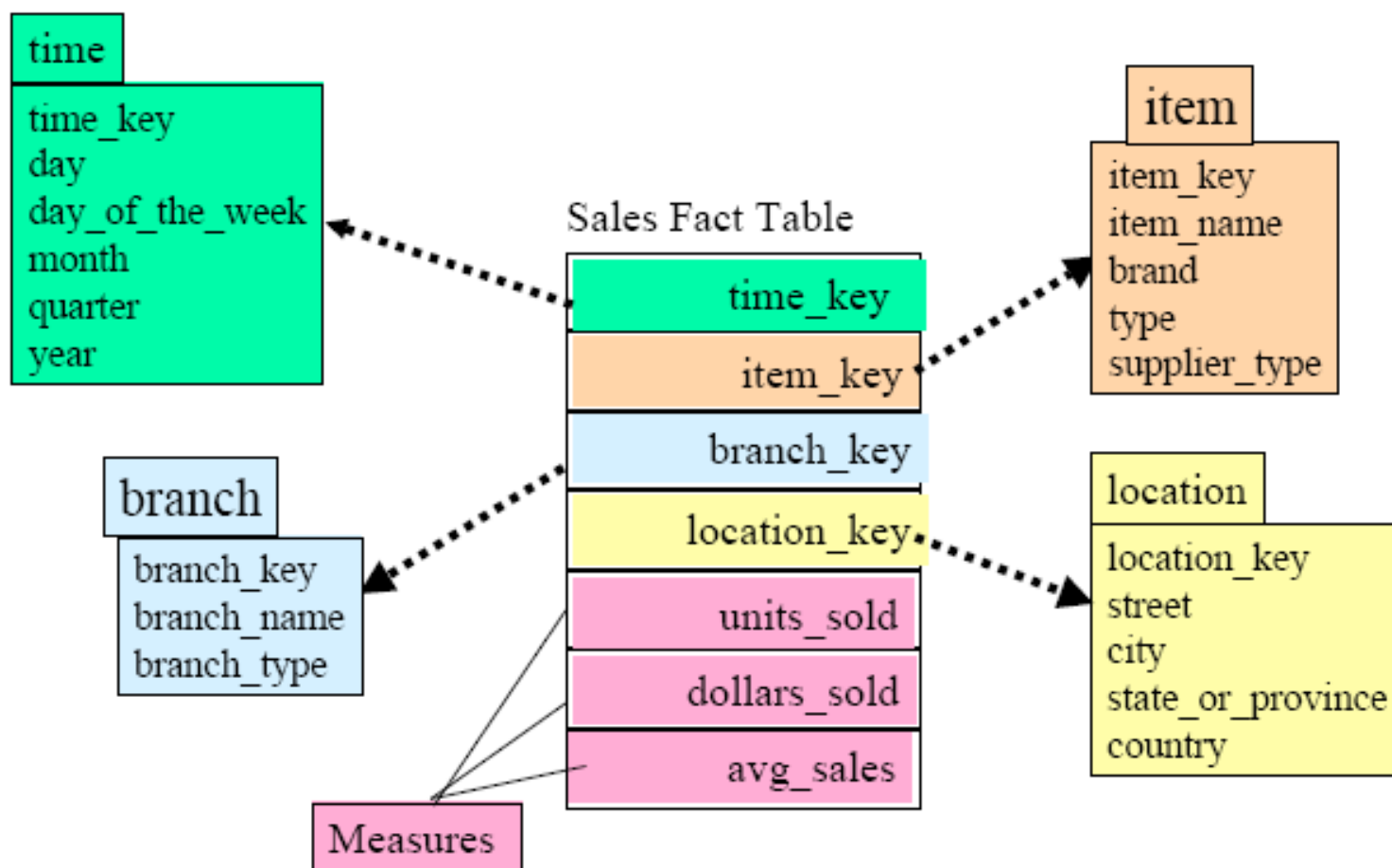
- Phân đoạn dữ liệu
- Dữ liệu không chuẩn hóa
- Một bản dữ liệu và các chiều bao quanh
- Các câu hỏi nhằm vào bản dữ liệu và được cấu trúc bởi các bản chiều
- Mô hình nổi bật của KDL





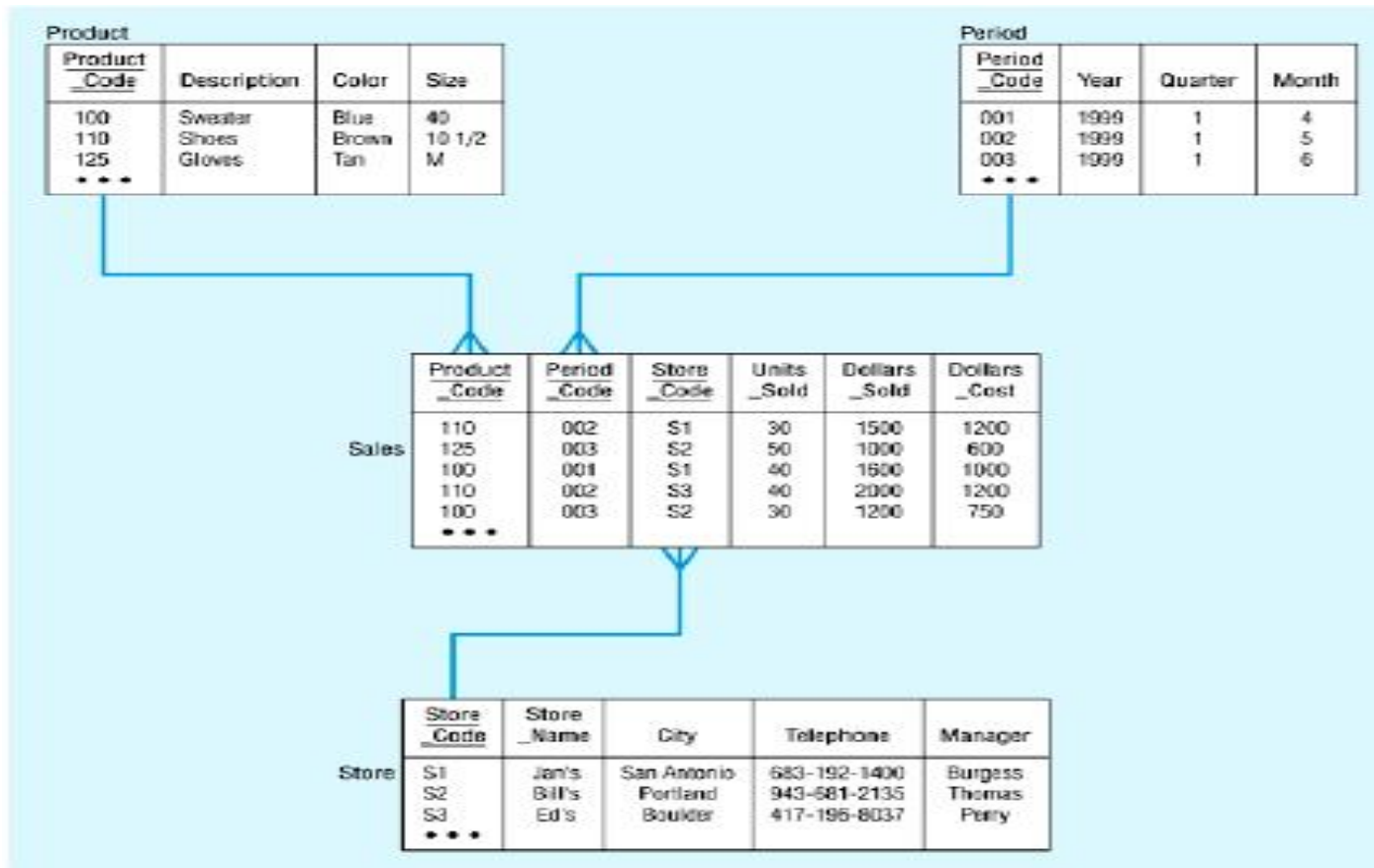


# Ví dụ lược đồ hình sao có 4 chiều



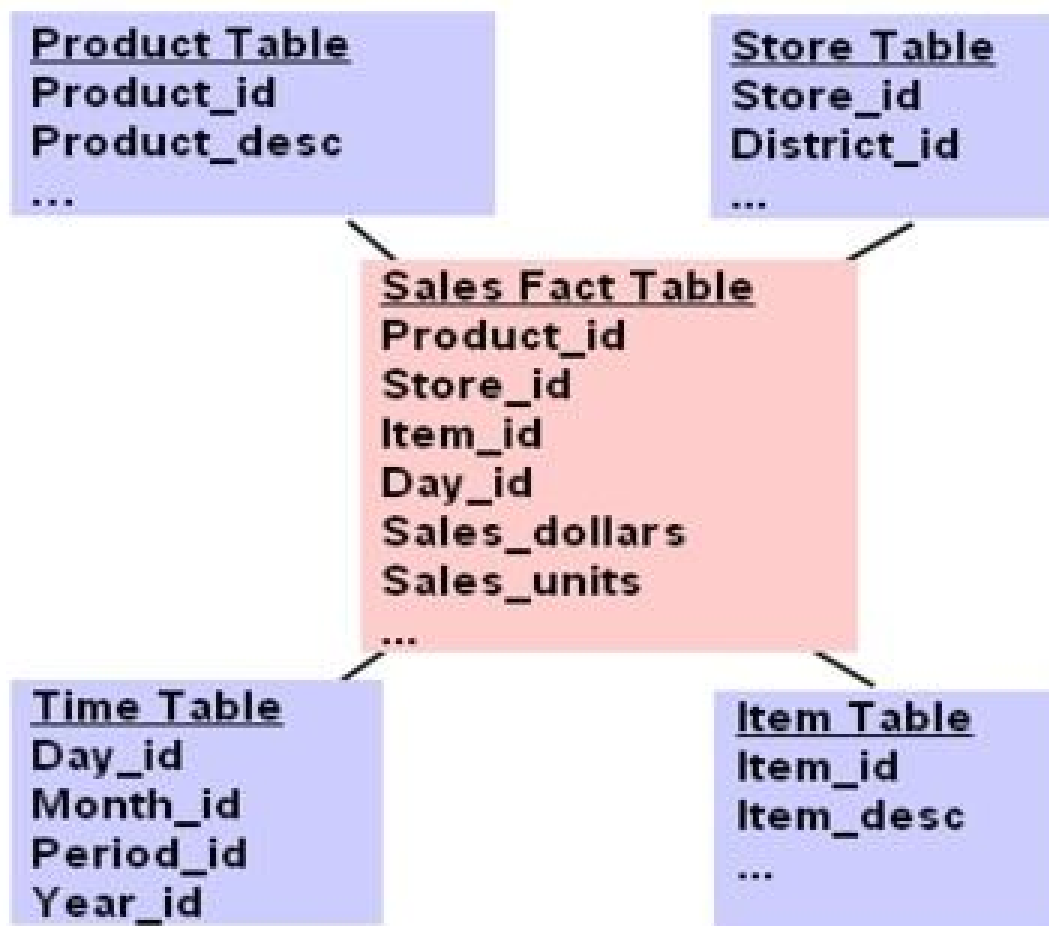


# Ví dụ lược đồ hình sao với dữ liệu





## Các lược đồ của ROLAP



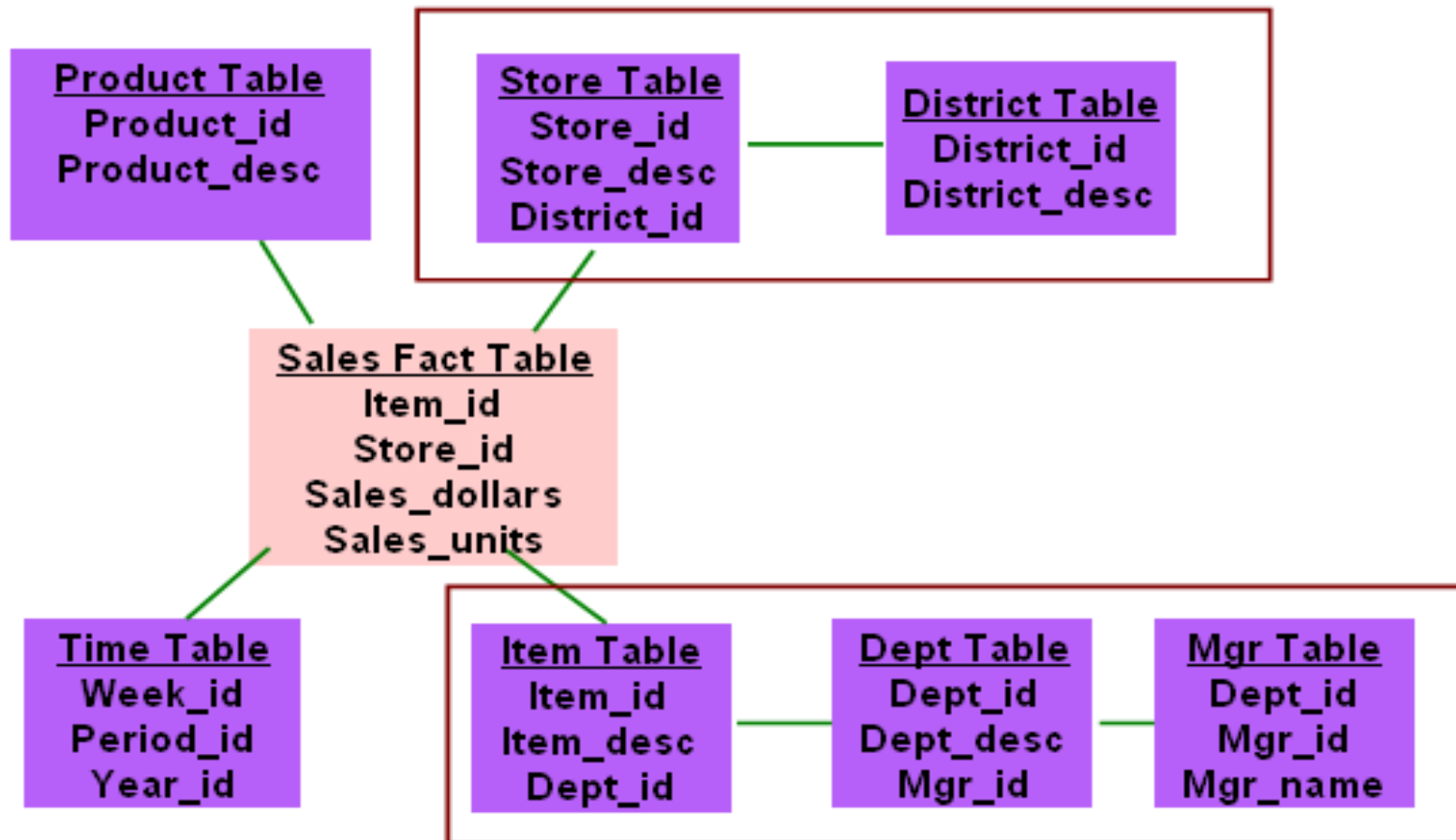


## Lược đồ hình bông tuyết

- Căn cứ vào lược đồ hình sao
- Bản dữ kiện giống như lược đồ hình sao
- Các chiều được chuẩn hoá (chuẩn 3)
- Các chiều được cấu trúc rõ ràng

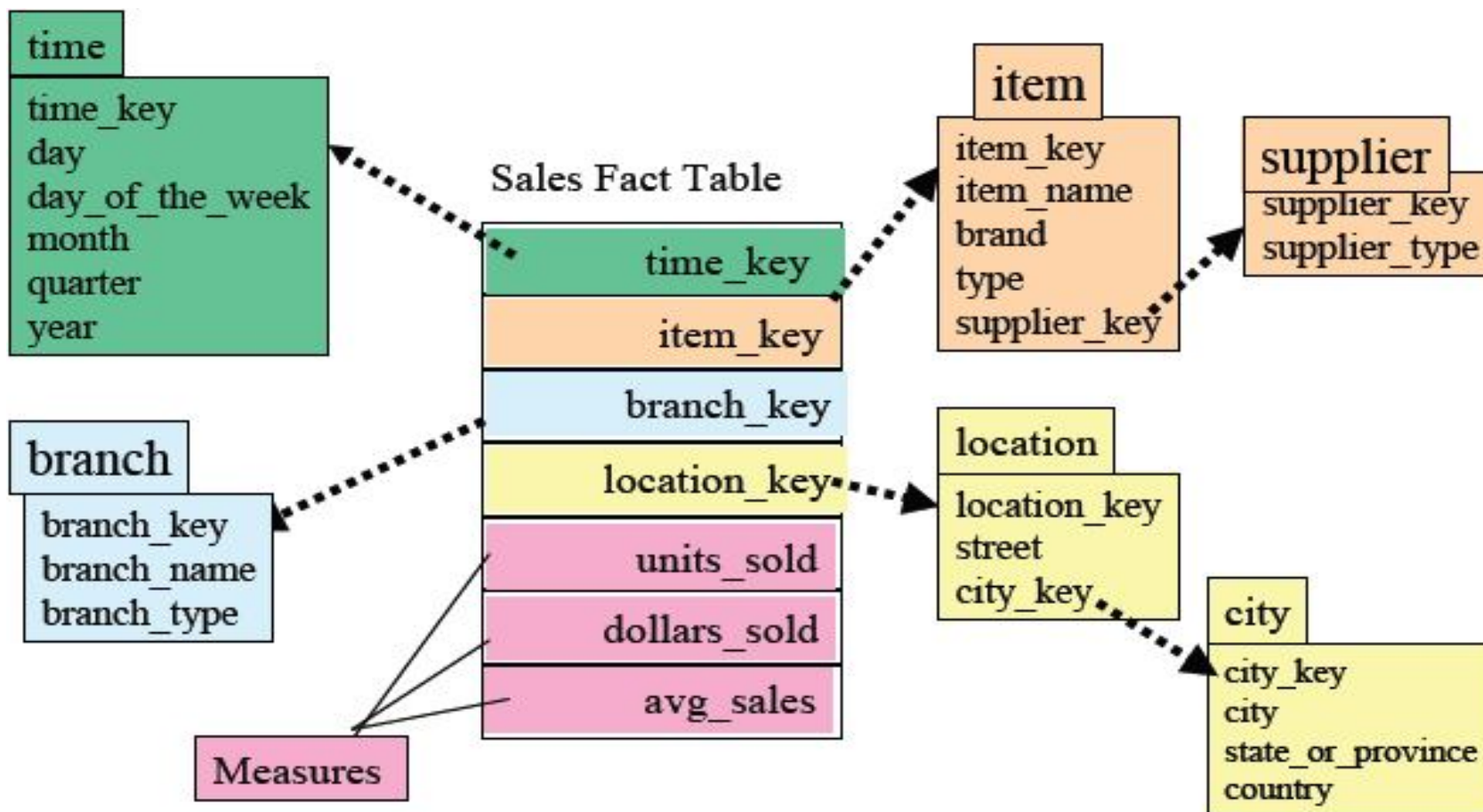


# Ví dụ về lược đồ bông tuyết



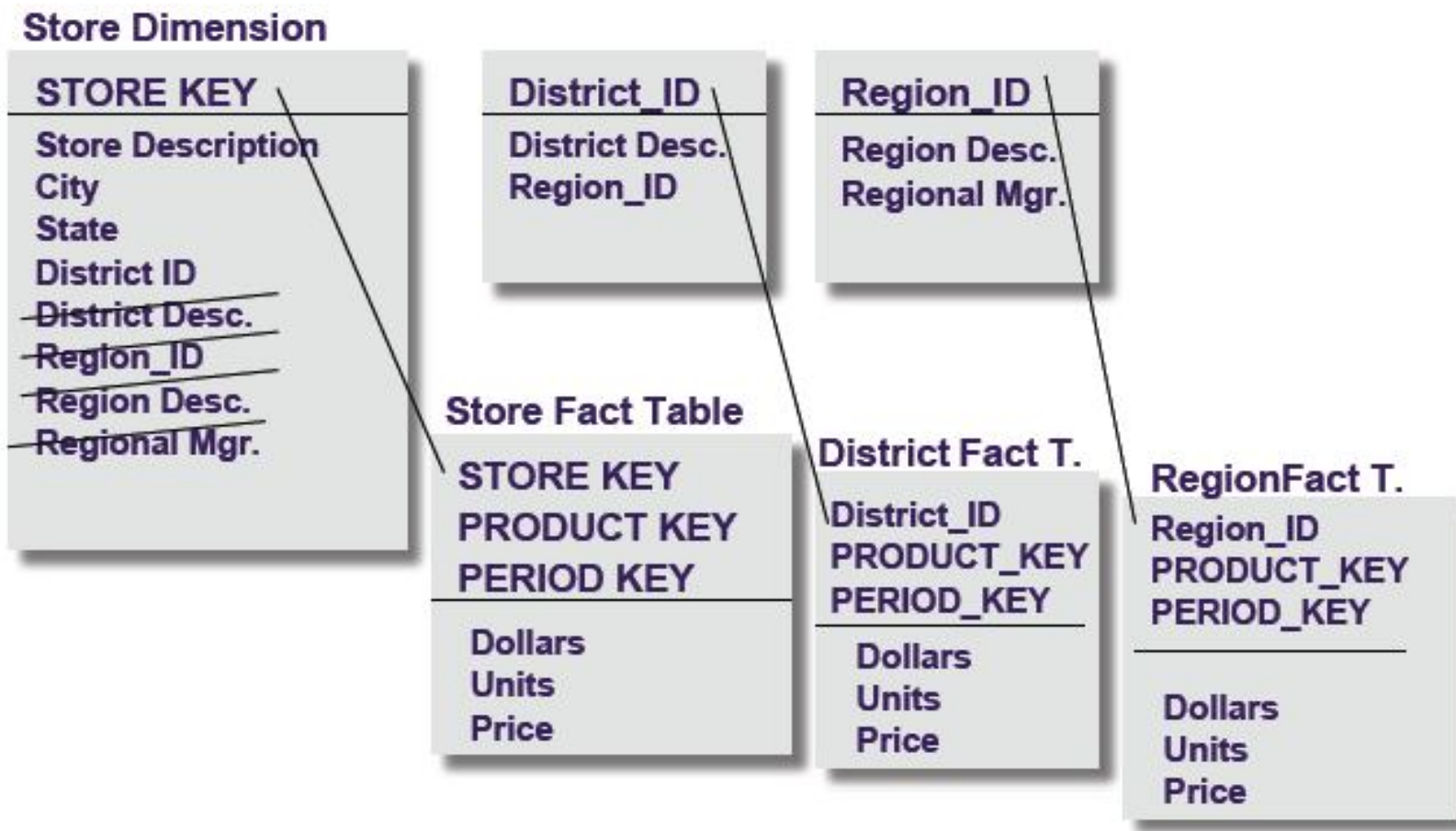


# Ví dụ về lược đồ bông tuyết



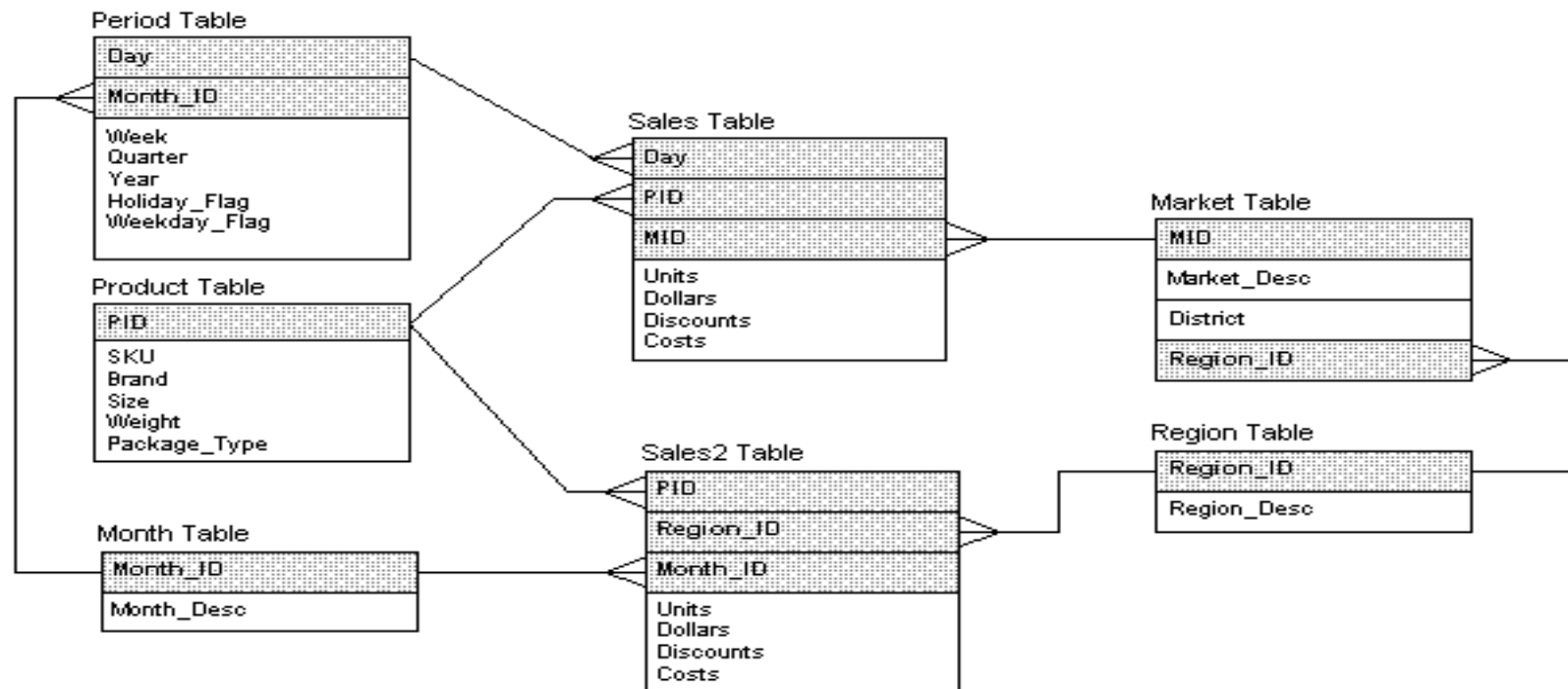


# Chiều trong lược đồ hình bông tuyết



# Các lược đồ của ROLAP (6)

## Đa fact hình sao (Multi fact star schema)

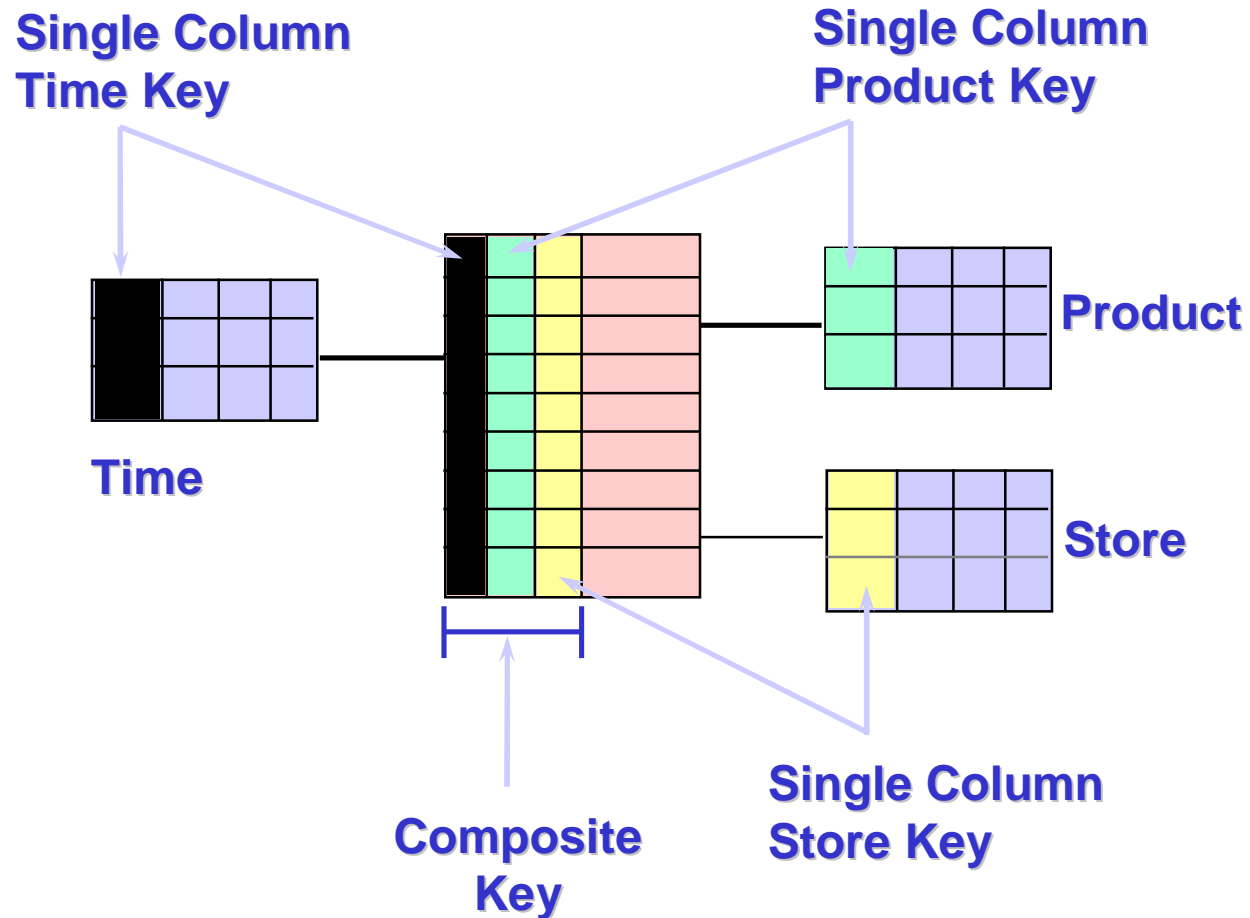


Example of Realistic Retail Star Schema  
(Multi Fact Star Schema)





# Khóa trong lược đồ ROLAP





## Các khoá

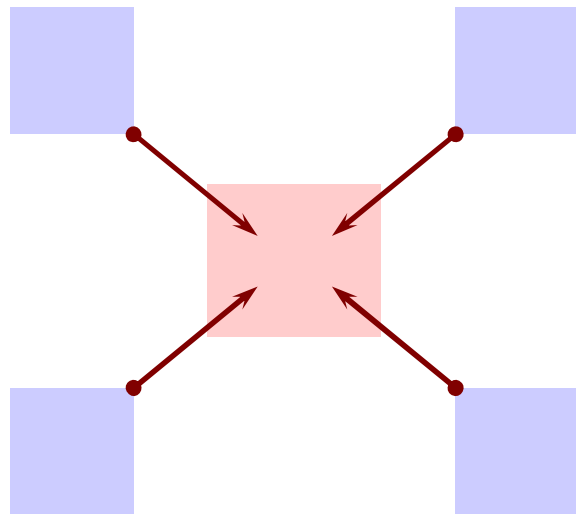
- Khóa chính nằm trên các bản chiều
- Khoá ngoài nằm trên bản dữ kiện
- Tạo lập các index để tăng tốc độ
- Khoá được xác định trong giai đoạn thiết kế
- Các khóa kết hợp có thể được áp dụng



# Các thuộc tính của bản dữ kiện

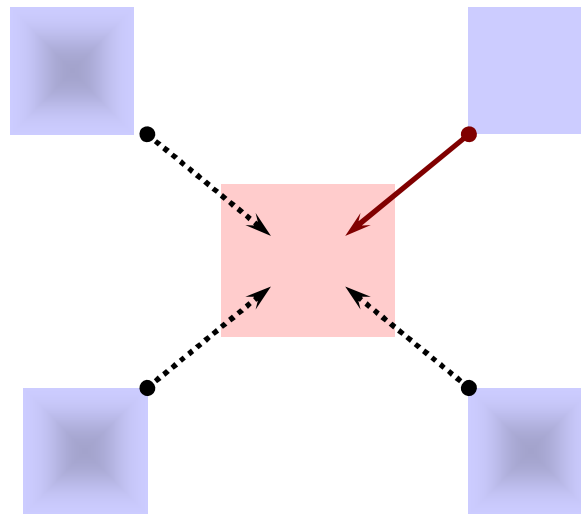
## Additive

Liên kết với tất cả các chiều



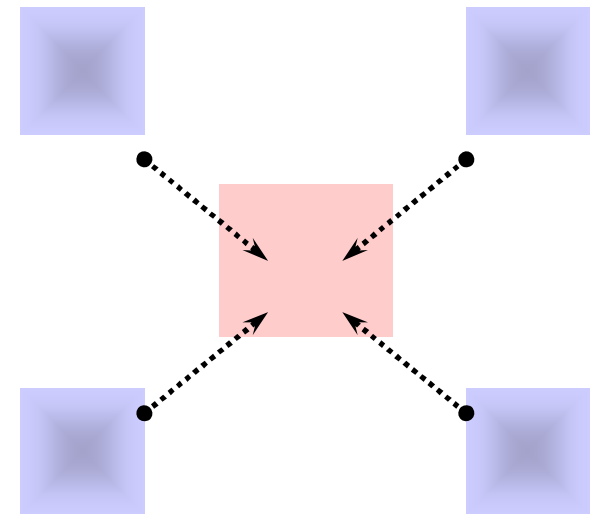
## Semi-additive

Liên kết với một số chiều



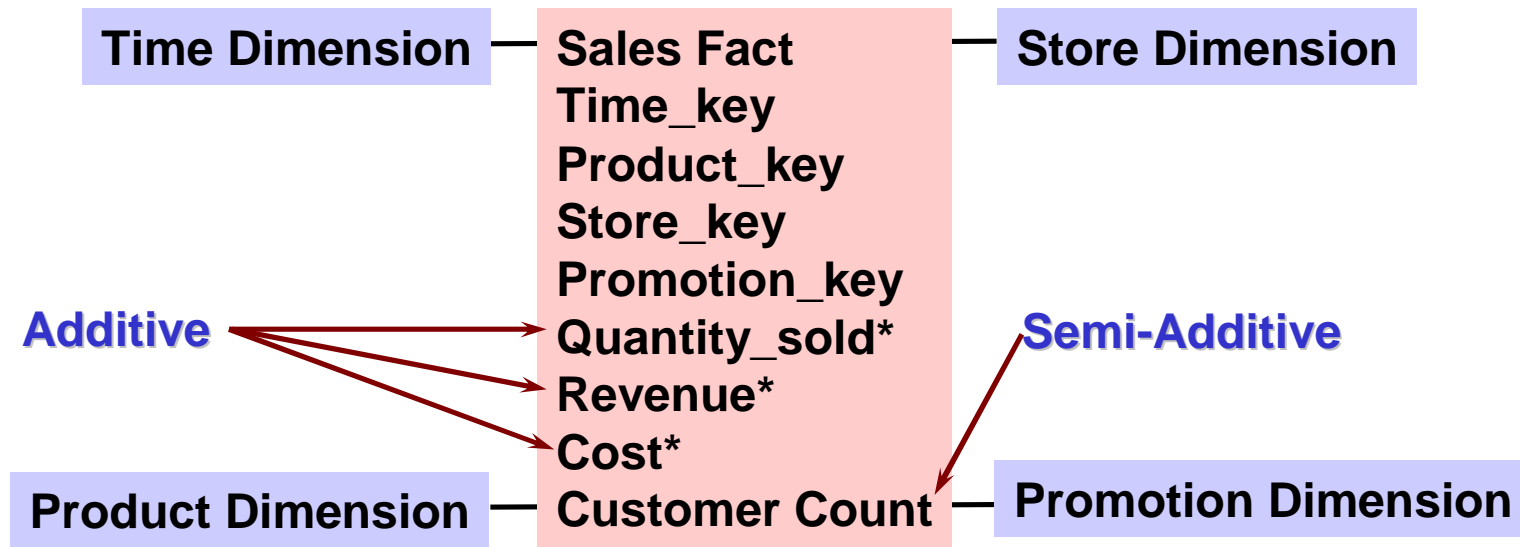
## Non-additive

Không liên kết





# Các thuộc tính của bản dữ kiện

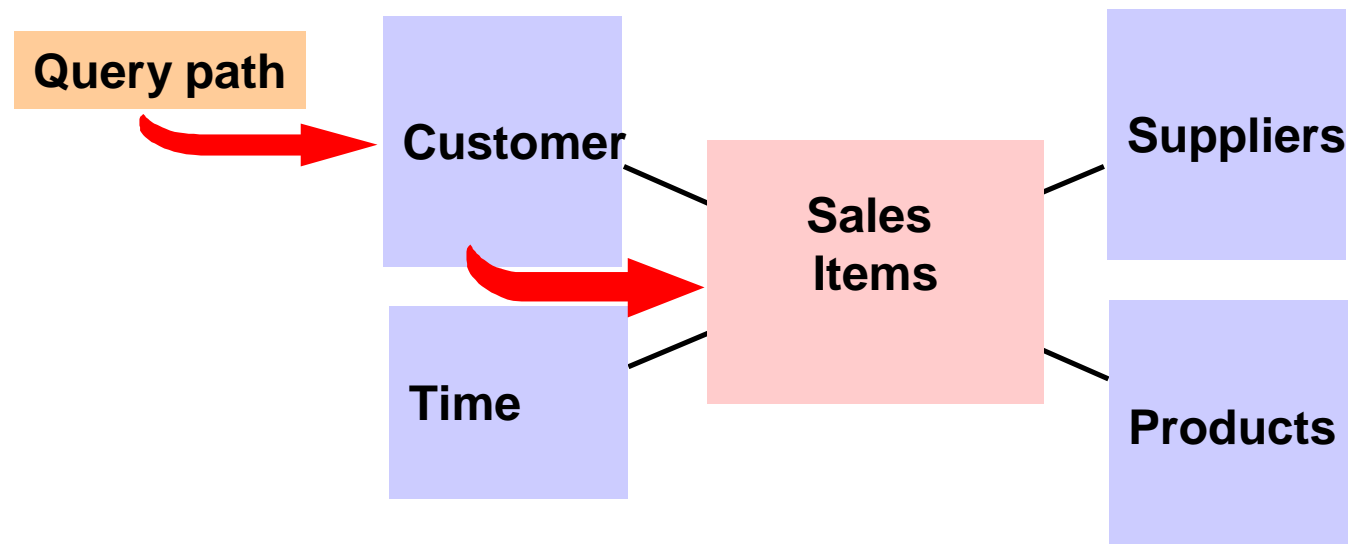


- Cost by Product + Store + Time
- Cost by Product + Promotion + Time
- Revenue by Product + Promotion + Time
- Revenue by Product + Promotion + Store + Time ...



## Chiều dữ liệu

- Dữ liệu của chiều xác định cấu trúc và miền của câu hỏi
- Dữ liệu của chiều liên kết với dữ kiện thông qua các khóa





## Chuẩn hóa chiều

- Dữ liệu chuẩn hóa không
  - Dư thừa
  - Trùng lặp
  - Các cột khóa độc lập
- Dữ liệu không chuẩn hóa
  - Tăng tốc độ trả lời câu hỏi
  - Tồn tại trong kho dữ liệu



## Độ thô dữ liệu

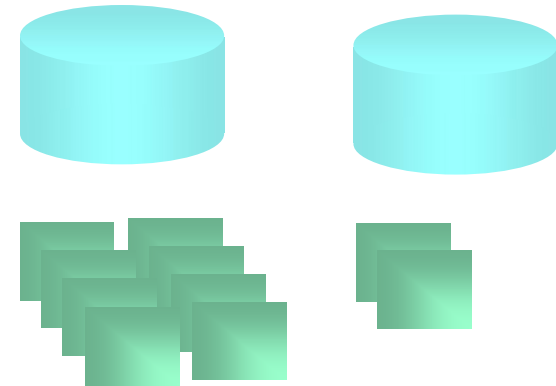
- Dữ liệu được làm cho phù hợp với KDL tại một mức độ chắc chắn (độ thô)
- Căn cứ vào độ thô đó các tổng hợp được định nghĩa
- Độ thô cao: nhiều dữ liệu hơn, độ thô thấp: ít dữ liệu hơn
- Các câu hỏi:
  - Độ thô nào đã có sẵn ?
  - Độ thô nào hợp lý và tiện dụng cho KDL

*Xu hướng lưu trữ dữ liệu có độ thô cao nhất, vì khi đã giảm độ thô thì dữ liệu chi tiết không thể phục hồi được*



## Độ thô dữ liệu

- Tác động đến kho dl
  - Dung lượng của kdl
  - Cấp độ phân tích
  - Tính mềm dẻo
- Cấp độ chi tiết của dữ liệu
  - Ngày, Tháng, Năm
  - Tỉnh, quốc gia,...
- Phân tích các yêu cầu để tránh dư thừa hay thiếu hụt dữ liệu
- Khảo sát, làm mới nếu cần thiết







## Các bảng tổng hợp

- Mục đích: trả lời nhanh các câu hỏi thường gặp
- Các bước:
  - Nhận diện các câu hỏi thường gặp
  - Nhận diện các chiều và các tổng hợp tương ứng
  - Định nghĩa phân cấp của các tổng hợp
  - Tạo một cách có chọn lựa các bản dữ kiện tổng hợp tính toán trước
  - Tạo các bảng chiều tổng hợp



## Các bảng tổng hợp

Việc dùng các tổng hợp được tính toán và lưu trữ trước là một hiệu quả lớn đối với kho dữ liệu để điều khiển và làm tăng hiệu suất câu hỏi

Kimball 1996