

Trường Đại học CNTT – Khoa Hệ thống thông tin

BÀI 6: KỸ THUẬT OLAP (ONLINE ANALYTICAL PROCESSING)

Giảng viên: ThS. Nguyễn Thị Kim Phụng
ThS. Hồ Trung Thành

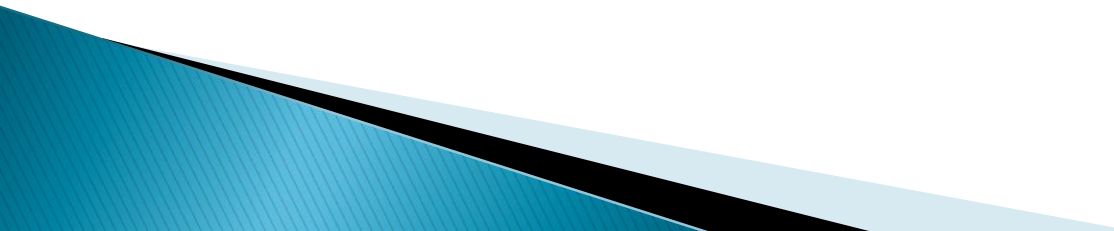
Kỹ thuật OLAP trong phân tích dữ liệu

- ▶ Giới thiệu về OLAP
- ▶ Các mô hình hỗ trợ lưu trữ OLAP
 - ROLAP
 - MOLAP
 - HOLAP
- ▶ Kiến trúc khối của OLAP
- ▶ Các kỹ thuật thao tác trên dữ liệu đa chiều
 - Cuộn lên (Roll up)
 - Truy xuống (Drill down)
 - Chọn và chiếu (Slice and dice)
 - Xoay chiều (Pivot)

Giới thiệu về OLAP

- ▶ OLAP là một kỹ thuật nhằm:
 - Đáp ứng nhu cầu phân tích dữ liệu trực tuyến theo các góc độ khác nhau dựa trên cách tổ chức cơ sở dữ liệu đa chiều
 - Mọi dữ liệu đều được hiển thị tức thời.
 - Hoạt động dựa trên cơ sở dữ liệu đa chiều rất gần gũi với cách nhìn của người dùng

Giới thiệu về OLAP (2)

- ▶ OLAP còn là một kỹ thuật sử dụng các thể hiện dữ liệu đa chiều gọi là các *khối* nhằm cung cấp khả năng truy xuất nhanh đến dữ liệu của kho dữ liệu.
 - ▶ Tạo khối cho dữ liệu trong các bảng chiều, bảng sự kiện trong kho dữ liệu
 - ▶ Cung cấp khả năng thực hiện các truy vấn phức tạp.
- 

Giới thiệu về OLAP (3)

- ▶ Kho dữ liệu và kho dữ liệu theo chủ đề lưu trữ dữ liệu cho phân tích
- ▶ OLAP là kỹ thuật cho phép các ứng dụng người dùng truy xuất hiệu quả dữ liệu trong kho.
- ▶ OLAP cung cấp nhiều lợi ích cho người phân tích, ví dụ như:

Giới thiệu về OLAP (4)

- ▶ Cung cấp mô hình dữ liệu đa chiều trực quan cho phép dễ dàng lựa chọn, định hướng và khai phá dữ liệu.
- ▶ Cung cấp một ngôn ngữ truy vấn, phân tích và cung cấp sức mạnh để khám phá các mối quan hệ trong dữ liệu kinh doanh phức tạp.
- ▶ Dữ liệu được tính toán trước đối với các truy vấn thường xuyên nhằm làm cho thời gian trả lời rất nhanh đối với các truy vấn đặc biệt.

Hệ quản trị OLAP (OLAP Server)

- ▶ Là loại phần mềm hỗ trợ kỹ thuật OLAP cho các ứng dụng và gồm có hai phần chính:
 - Hệ quản trị cơ sở dữ liệu đa chiều
 - Các công cụ phân tích dữ liệu

Các mô hình lưu trữ hỗ trợ OLAP

- ▶ Dịch vụ OLAP hỗ trợ nhiều mô hình lưu trữ dữ liệu khác nhau, mỗi mô hình có các ưu và khuyết điểm riêng, chúng được sử dụng tùy theo mục đích khai thác.
- ▶ Mức cơ bản có mô hình lưu trữ cho Hệ quản trị OLAP
 - ROLAP
 - MOLAP
 - HOLAP

Mô hình Multi – Dimensional OLAP (MOLAP)

- ▶ Mô hình MOLAP lưu trữ dữ liệu cơ sở (là dữ liệu từ các bảng của kho dữ liệu hoặc trung tâm dữ liệu theo chủ đề)
- ▶ Thông tin tổng hợp (là các độ đo được tính toán từ các bảng) trong các cấu trúc đa chiều gọi là các khối.
- ▶ Các cấu trúc này được lưu bên ngoài cơ sở dữ liệu theo chủ đề hoặc kho dữ liệu.
- ▶ Ví dụ tổng sản phẩm bán được của tất cả các vùng theo quý.

Mô hình Relational OLAP (ROLAP)

- ▶ Mô hình ROLAP lưu trữ dữ liệu cơ sở và thông tin tổng hợp trong các bảng quan hệ.
- ▶ Các bảng này được lưu trữ trong cùng cơ sở dữ liệu như là các bảng của kho dữ liệu hoặc kho dữ liệu theo chủ đề.

Mô hình hybrid OLAP (HOLAP)

- ▶ Mô hình HOLAP là sự kết hợp giữa MOLAP và ROLAP.
- ▶ Cách lưu trữ này giúp người sử dụng nhanh chóng thực hiện các thao tác phân tích đa chiều trên dữ liệu tổng hợp đồng thời có thể truy xuất cấp chi tiết có khối lượng dữ liệu lớn khi cần thiết

So sánh các mô hình

Đặc điểm	MOLAP	ROLAP	HOLAP
Lưu trữ dữ liệu cơ sở	Khối	Bảng quan hệ	Bảng quan hệ
Lưu trữ thông tin tổng hợp	Khối	Bảng quan hệ	Khối
Hiệu suất thực hiện truy vấn	Nhanh nhất	Chậm nhất	Nhanh
Tiêu thụ không gian lưu trữ	Nhiều	Thấp	Trung bình
Chi phí bảo trì	Cao	Thấp	Trung bình

Kiến trúc khối của OLAP

- ▶ Đối tượng chính của OLAP là khối, là một thể hiện đa chiều của dữ liệu chi tiết và tổng hợp.
- ▶ Một khối bao gồm:
 - Nguồn dữ liệu
 - Các chiều
 - Các độ đo và các phần dành riêng.
- ▶ Các khối được thiết kế dựa trên yêu cầu phân tích của người dùng.
- ▶ Một kho dữ liệu có thể hỗ trợ nhiều khối khác nhau như khối Bán hàng, khối Bảng kiểm kê,...

Kiến trúc khối của OLAP (2)

- ▶ **Dữ liệu nguồn** của một khối chỉ ra nơi chứa kho dữ liệu cung cấp dữ liệu cho khối.
- ▶ **Các chiều** được ánh xạ từ các thông tin của các bảng chiều trong kho dữ liệu vào các mức phân cấp, ví dụ như chiều Địa lý thì gồm các mức như Lục địa, Quốc gia, Tỉnh–Thành phố.
- ▶ **Các độ đo** xác định các giá trị số từ bảng sự kiện mà được tổng hợp cho phân tích như giá bán, chi phí hoặc số lượng bán.

Kiến trúc khối của OLAP (3)

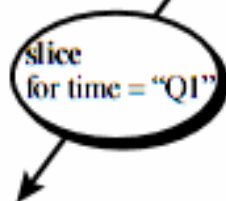
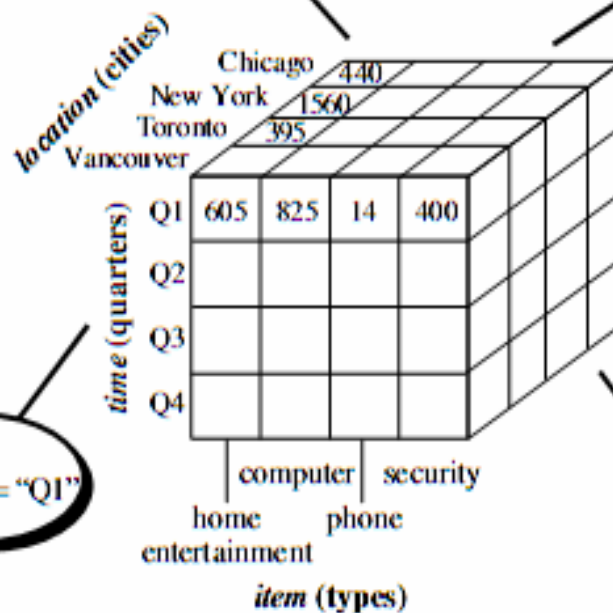
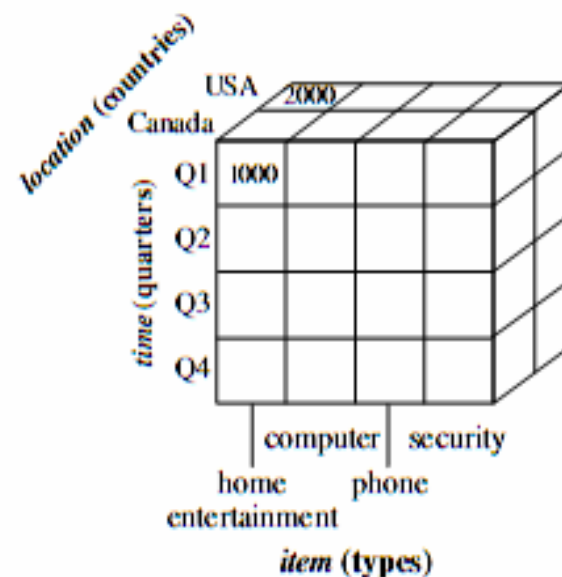
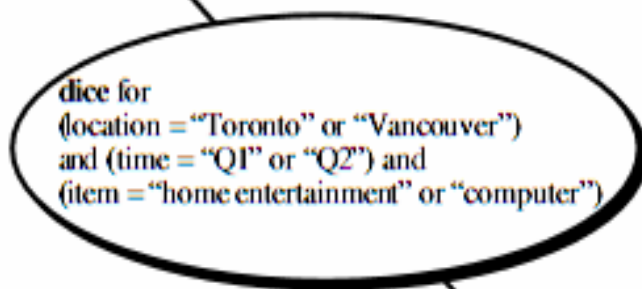
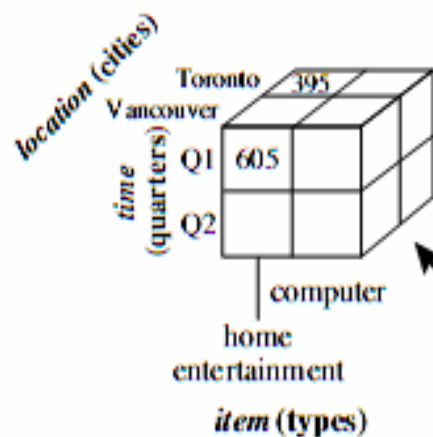
- ▶ Các partition của một khối
 - Có thể được lưu trữ độc lập trong các cách thức khác nhau với các mức độ tổng hợp khác nhau.
 - Các partition không thể hiện đối với người dùng, đối với họ một khối là một đối tượng đơn, và chúng cung cấp những tùy chọn đa dạng để quản lý dữ liệu OLAP.
- ▶ ...

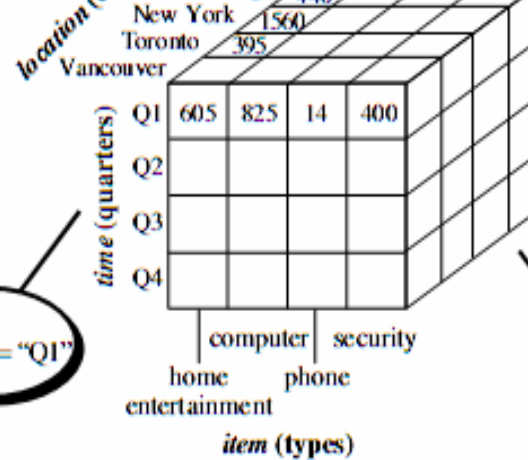
Phân tích dữ liệu

- ▶ Giới thiệu
- ▶ Công cụ SSAS trong SQL Server
- ▶ Các trình thiết kế của SSAS
- ▶ Sử dụng công cụ SSAS phân tích dữ liệu
- ▶ Ngôn ngữ truy vấn dữ liệu đa chiều MDX

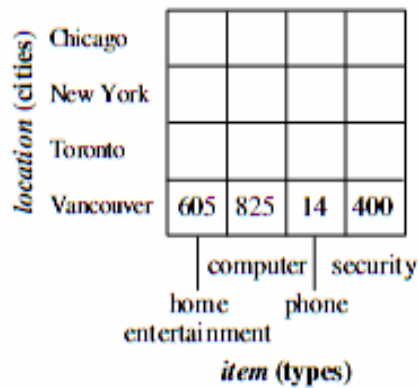
Các thao tác phân tích trên dữ liệu đa chiều

- ▶ Thao tác và phân tích trên cơ sở dữ liệu nhiều chiều được OLAP cung cấp một số công cụ phân tích từ đơn giản đến phức tạp như:
 - Cuộn lên (Roll up)
 - Truy xuống (Drill down)
 - Chọn và chiếu (Slice and Dice)
 - Xoay chiều (Pivot)

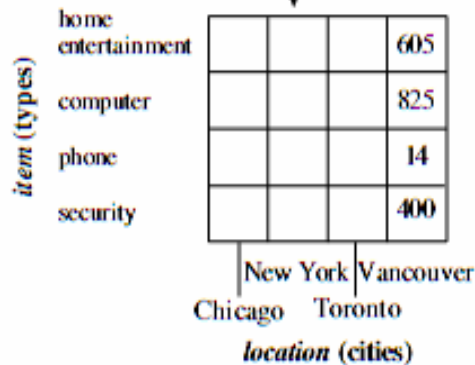




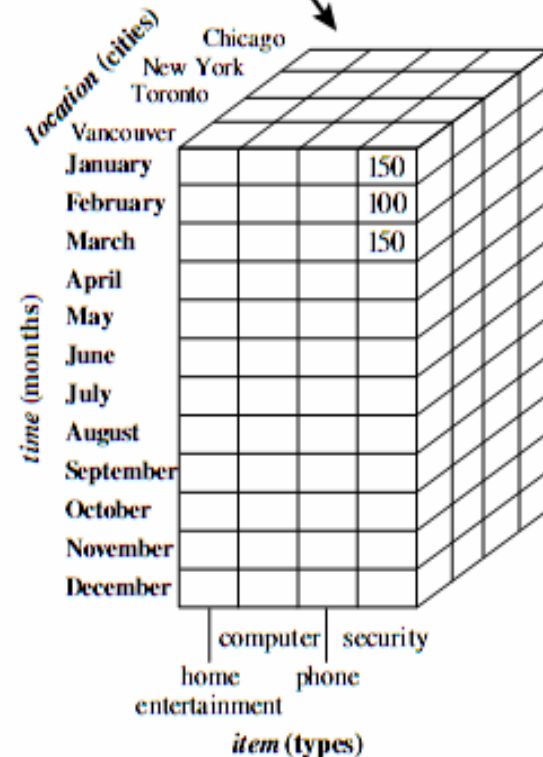
slice
for time = "Q1"



pivot



drill-down
on time
(from quarters
to months)

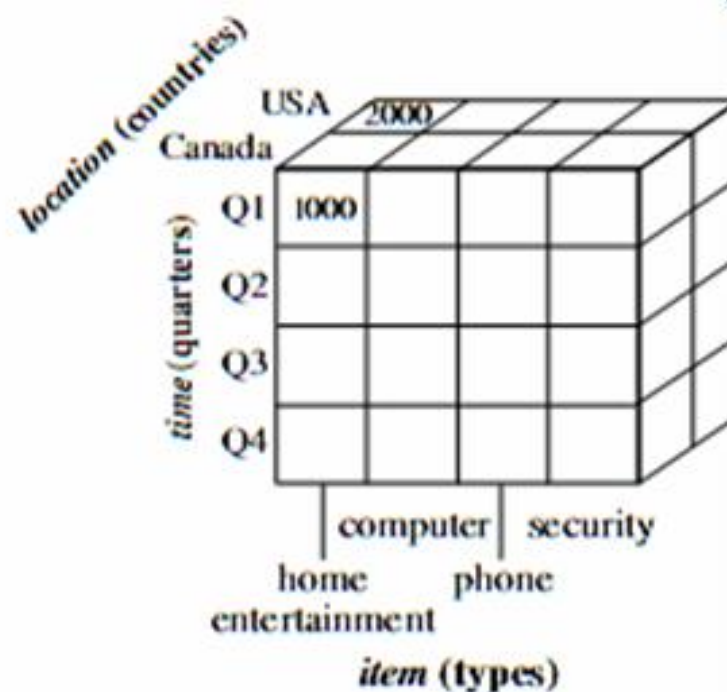
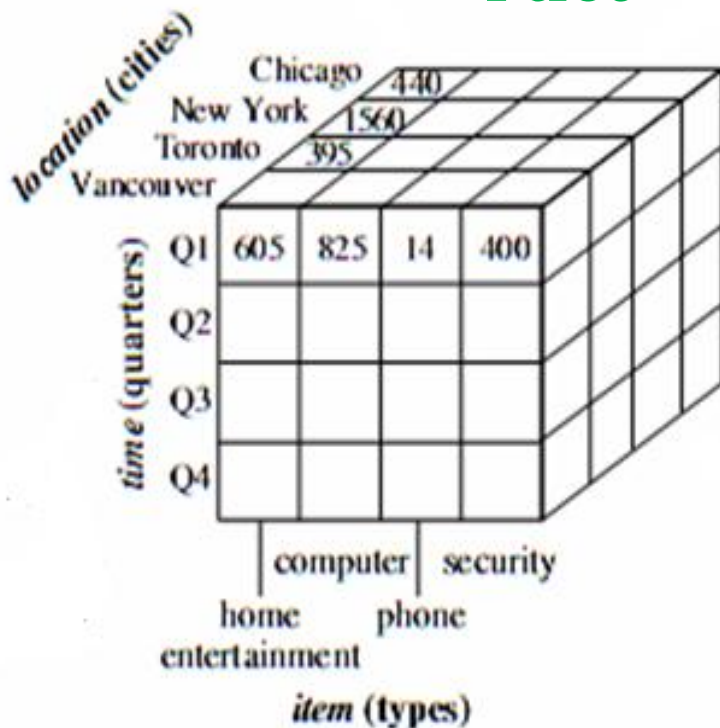


Cuộn lên (Roll – up)

- ▶ Dựa vào quan hệ cấp bậc của chiều ta có thể cuộn lên mức độ tổng quát hơn để xem dữ liệu ở mức tổng hợp.
- ▶ Ví dụ ta cuộn lên để xem dữ liệu ở mức tổng hợp theo quan hệ cấp bậc: dữ liệu ở *Thành phố (City)* → dữ liệu ở *Quốc gia (Country)* → dữ liệu ở *Vùng (Area)*...

Cuộn lên (Roll - up)

Fact



roll-up
on location
(from cities
to countries)

Dimesion

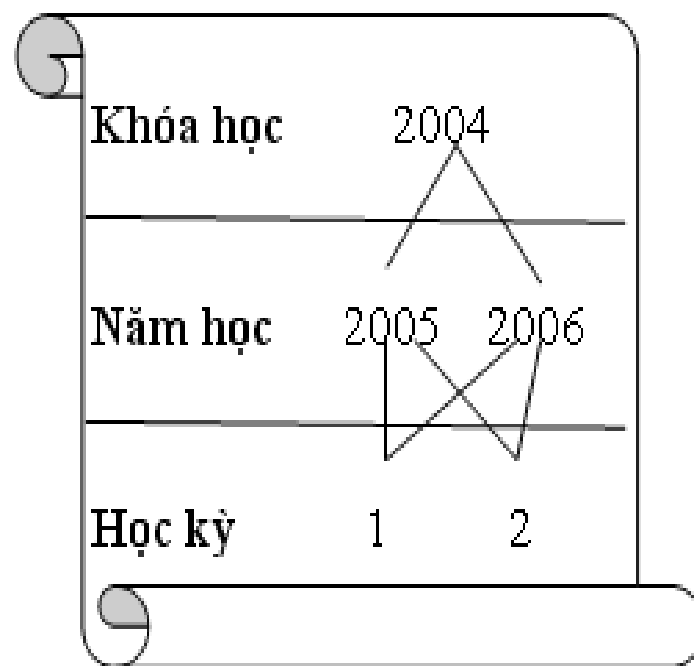
Cuộn lên theo chiều:
“Location” đi từ giá trị
mức nhỏ Cities -> tổng
hợp lên mức Countries

Location
Cities
Countries
Areas
.....

Cuộn lên (Roll – up)

- ▶ Một Ví dụ khác (Điểm tổng hợp của sinh viên): ta cuộn lên để xem dữ liệu ở mức tổng hợp theo quan hệ cấp bậc: *Học kỳ* → *Năm học* → *Khóa học*

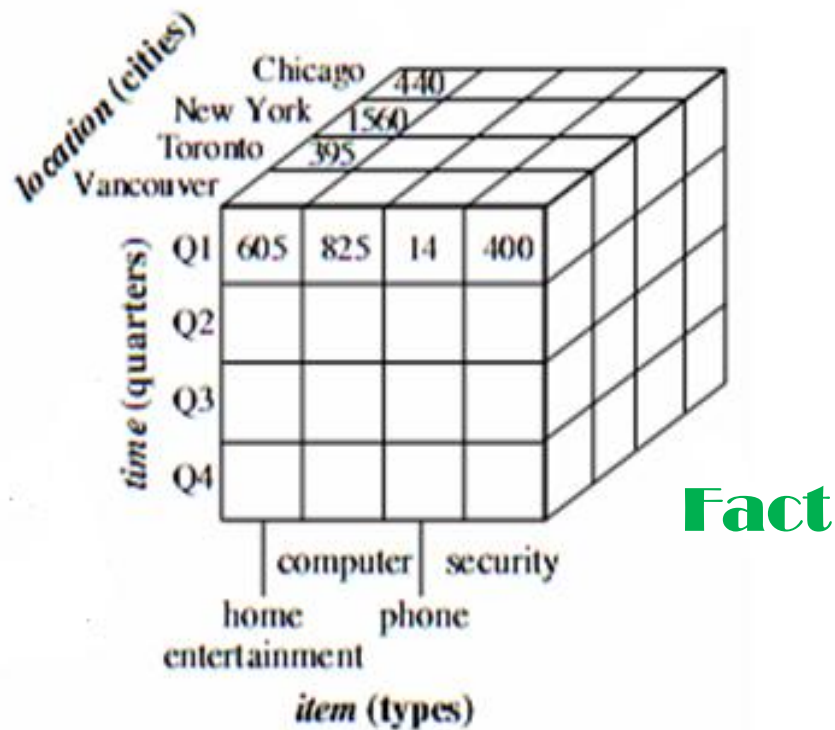
Chiều thời gian



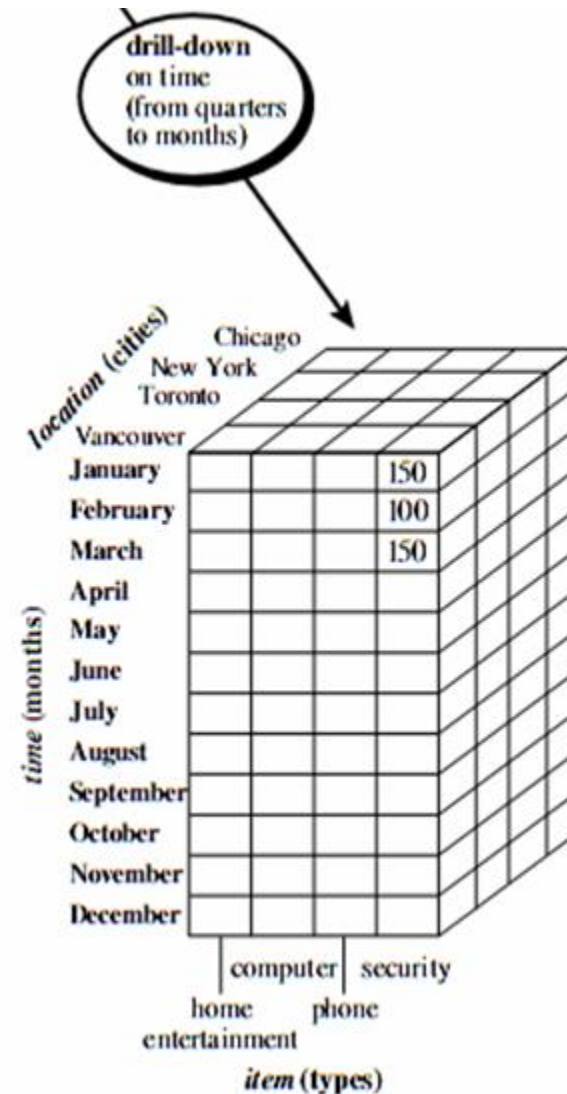
Truy xuống (Drill – down)

- ▶ Ngược lại với cuộn lên, từ mức độ tổng quát của quan hệ cấp bậc của chiều ta có thể truy xuống để xem được chi tiết của dữ liệu.
- ▶ Ví dụ ta truy xuống để xem dữ liệu ở mức chi tiết theo quan hệ cấp bậc *Quarters* → *Months* → *Weeks* → *Days*

Truy xuống (Drill - down)



Truy xuống theo chiều:
“Time” đi từ giá trị mức
lớn Quarters -> chi tiết
xuống mức Months

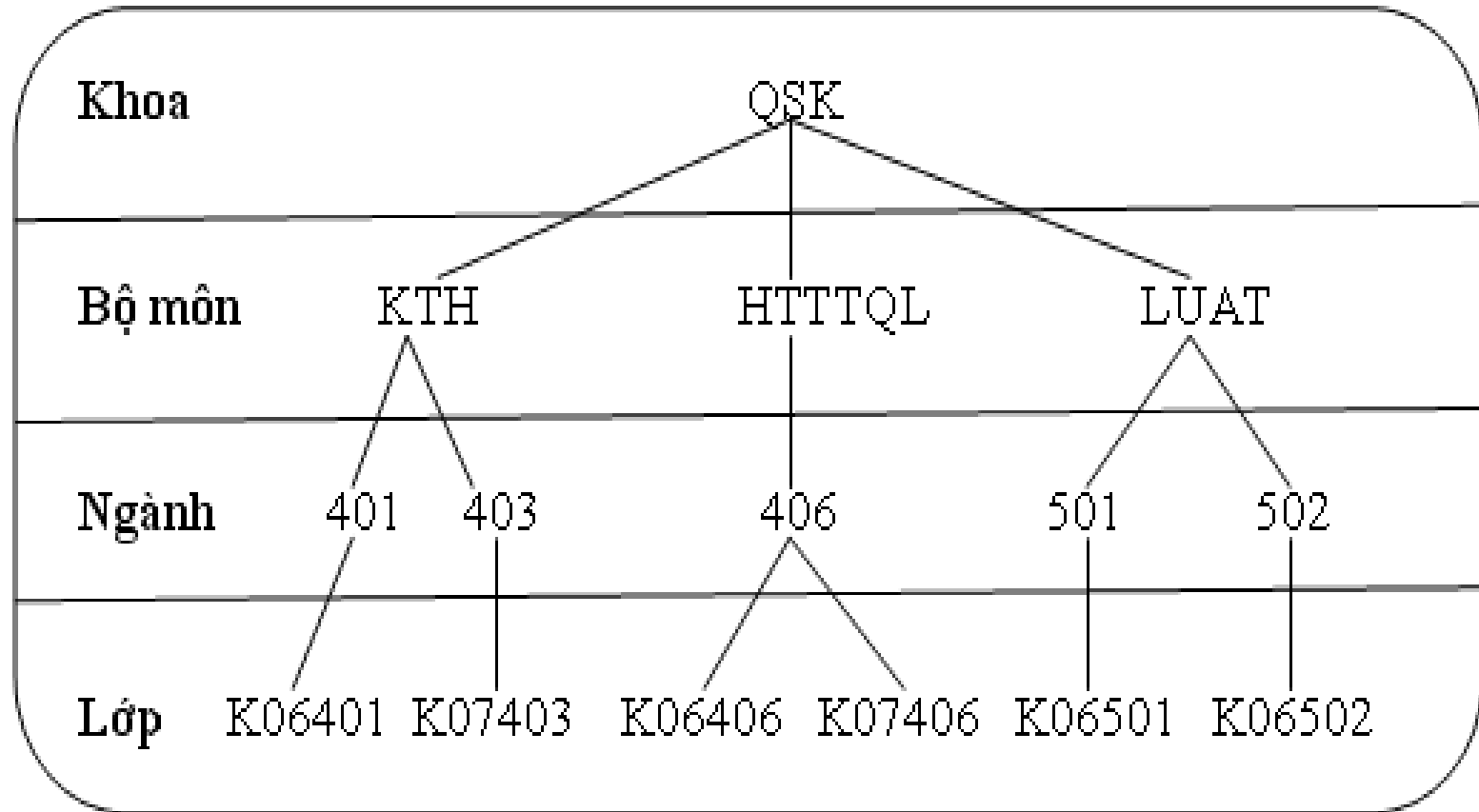


Dimension

Time
Quarters
Months
Weeks
Years
.....

Truy xuống (Ví dụ khác)

Chiều Đơn vị

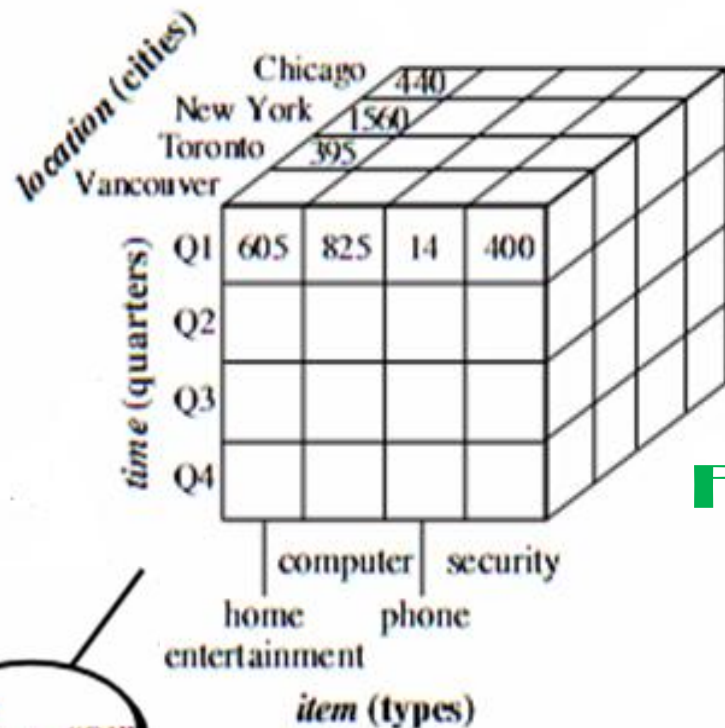


Chọn (Slice)

- ▶ **Chọn:** thực hiện một phép chọn trên một chiều của khối trung tâm để có được một khối con.
- ▶ Ví dụ ta chọn từ khối trung tâm của chiều *Time* với tiêu chí là *quarters = "Q1"*.

Chọn (Slice)

Chọn chiều: “Time” với tiêu chí là quý 1 (quarters=“Q1”) => khối con



slice for time = “Q1”

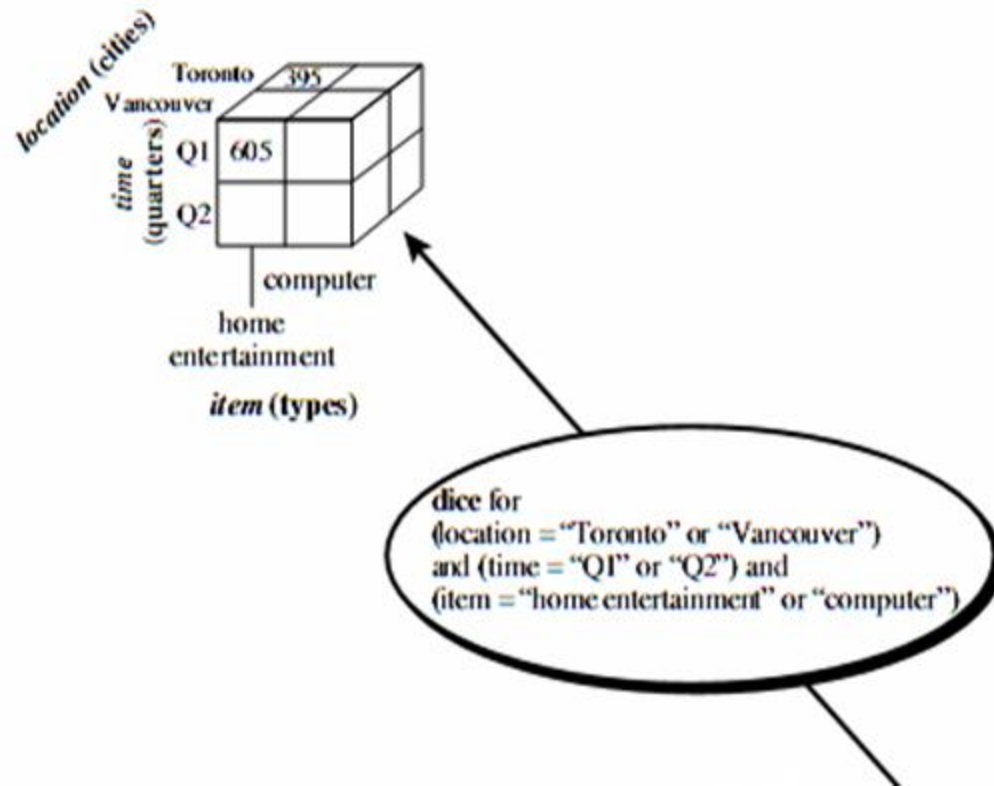
location (cities)	Chicago				
	New York				
	Toronto				
	Vancouver	605	825	14	400
		computer	security	home entertainment	phone
		item (types)			

Chọn Dimension: Time

Chiều (Dice)

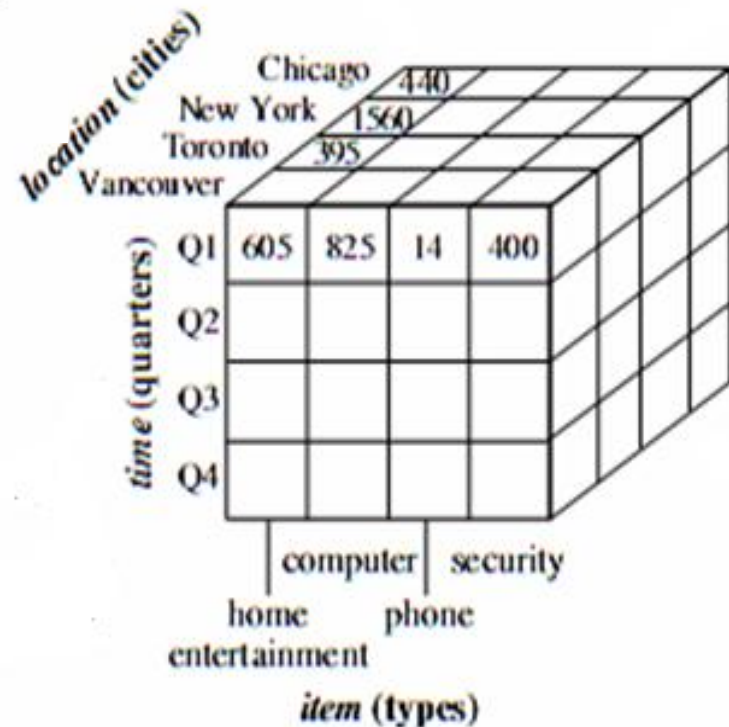
- ▶ **Chiều** là thao tác thực hiện chọn trên hai chiều hay nhiều hơn để tìm ra một khối con.
- ▶ Ví dụ chọn trên 3 chiều Location, Time và Item với các tiêu chí là: (*cities* = “Toronto” or “Vancouver”) and (*quarters* = “Q1” or “Q2”) and (*types* = “home entertainment” or “computer”).

Chiều (Dice)



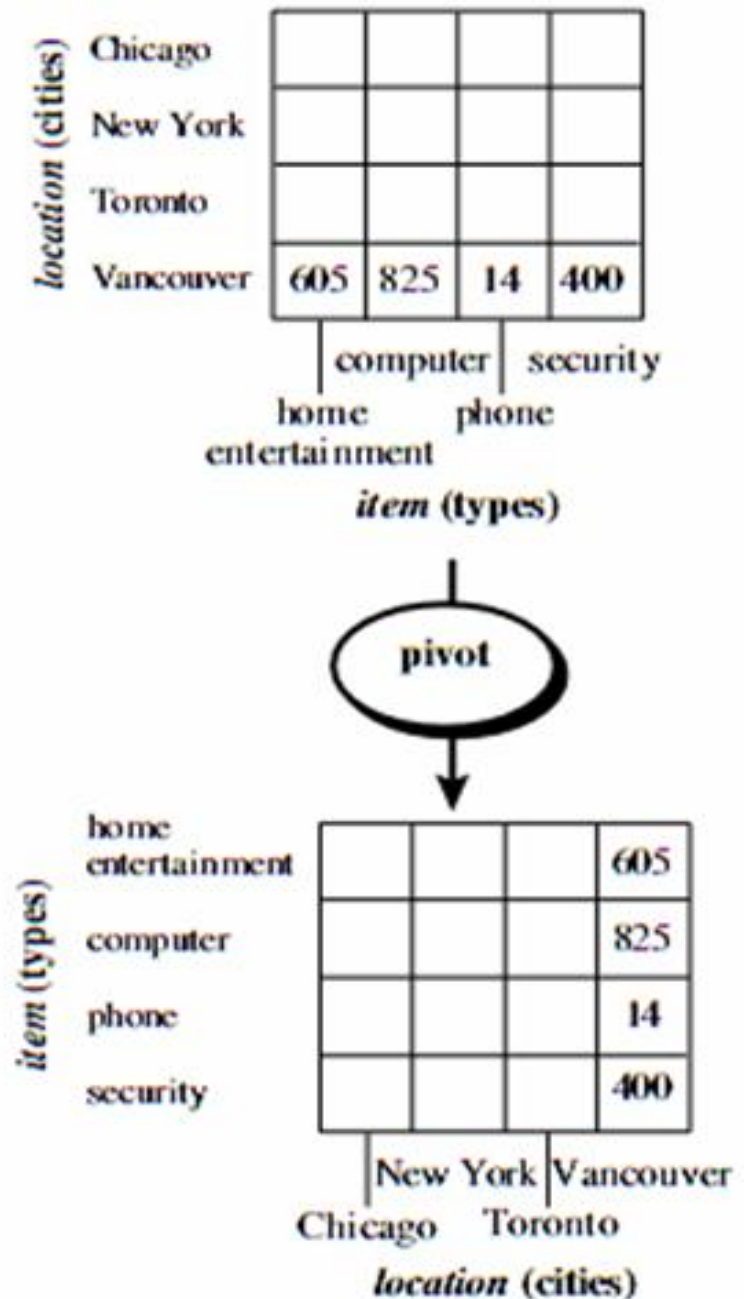
Chiều theo chiều (trục):
"Location", "Time",
"Item" với các tiêu chí
như trong hình Oval

Fact



Xoay chiều (Pivot)

- ▶ Thực hiện việc xoay khối dữ liệu để có thể xem dữ liệu ở nhiều chiều khác nhau.



Tóm tắt

- ▶ Từ cơ sở của kỹ thuật OLAP trong bài này đã giúp chúng ta hiểu hơn về kỹ thuật phân tích dữ liệu trực tuyến OLAP, những kỹ thuật phân tích trong khối dữ liệu đa chiều
- ▶ Trong bài sau chúng ta sẽ nghiên cứu phân phân tích và thiết kế kho dữ liệu ứng dụng trong thực tế