# ON-LINE ANALYTICAL PROCESSING

*Dr. Yashvardhan Sharma*

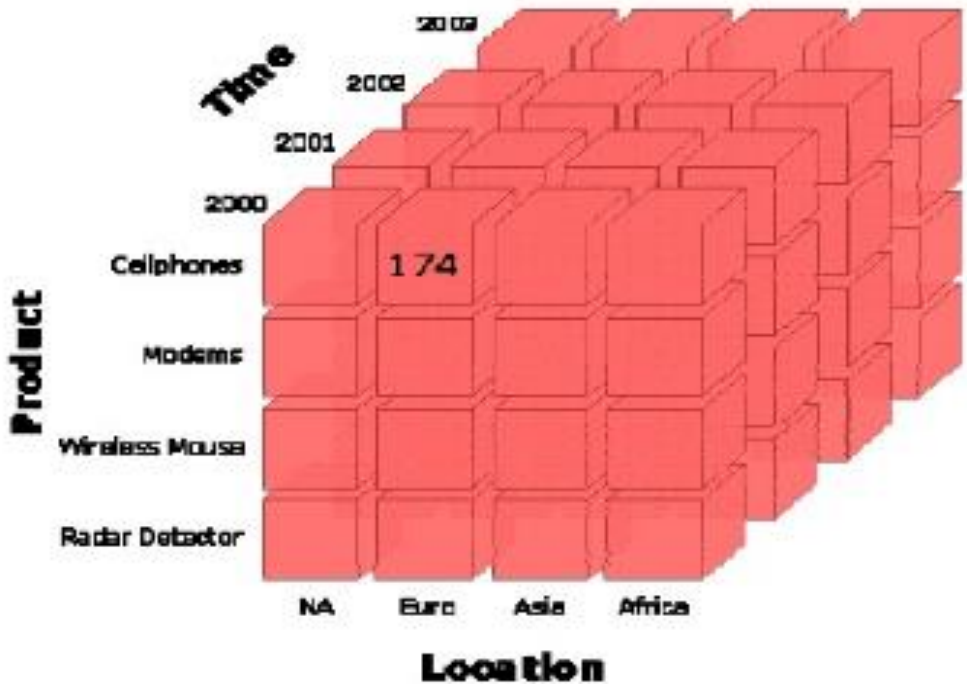
*Department of Computer Science & Information Systems*

*BITS, Pilani*

*L10 (Lec 17)*

# OLAP Definition

OLAP (online analytical processing) is computer processing that enables a user to easily and selectively extract and view data from different points of view.



# Typical calculations in the query requests

- Roll-ups to provide summaries and aggregations along the hierarchies of the dimensions
- Drill-downs from the top level to the lowest along the hierarchies of the dimensions, in combinations among the dimensions
- Simple calculations, such as computation of margins
- Share calculations to compute the percentage of parts to the whole
- Algebraic equations involving key performance indicators
- Moving averages and growth percentages
- Trend analysis using statistical methods

# OLTP vs Data warehouse

| Characteristic | OLTP Systems | Data Warehouse |
|---|---|---|
| Analytical capabilities | Very low | **Moderate** |
| Data for a single session | Very limited | **Small to medium size** |
| Size of result set | Small | **Large** |
| Response time | Very fast | **Fast to moderate** |
| Data granularity | Detail | **Detail and summary** |
| Data currency | Current | **Current and historical** |
| Access method | Predefined | **Predefined and ad hoc** |
| Basic motivation | Collect and input data | **Provide information** |
| Data model | Design for data updates | **Design for queries** |
| Optimization of database | For transactions | **For analysis** |
| Update frequency | Very frequent | **Generally read-only** |
| Scope of user interaction | Single transactions | **Throughout data content** |

# Limitations of other analysis methods

- No support for multidimensionality
- Inability to drill down to lower levels in the dimensions
- No support to rotate result by switching rows and columns
- Inability of SQL fetch results for complex calculations and handling time series data
- No alteration of the presentation of the result data sets possible

# OLAP Advantages

- It can reorganize metrics along several dimensions and allow data to be viewed from different perspectives
- Supports multidimensional analysis
- Ability to roll up , drill down within each dimension
- Fast response
- Can be implemented on web
- Highly interactive analysis

# Simple OLAP Session

| LINE | TOTAL SALES |
|---|---|
| Clothing | $12,836,450 |
| Electronics | $16,068,300 |
| Video | $21,262,190 |
| Kitchen | $17,704,400 |
| Appliances | $19,600,800 |
| Total | $87,472,140 |

1 High level summary by product line

2 Drill down by year

| LINE | 1998 | 1999 | 2000 | TOTAL |
|---|---|---|---|---|
| Clothing | $3,457,000 | $3,590,050 | $5,789,400 | $12,836,450 |
| Electronics | $5,894,800 | $4,078,900 | $6,094,600 | $16,068,300 |
| Video | $7,198,700 | $6,057,890 | $8,005,600 | $21,262,190 |
| Kitchen | $4,875,400 | $5,894,500 | $6,934,500 | $17,704,400 |
| Appliances | $5,947,300 | $6,104,500 | $7,549,000 | $19,600,800 |
| Total | $27,373,200 | $25,725,840 | $34,373,100 | $87,472,140 |

3 Rotate columns to rows

| YEAR | Clothing | Electronics | Video | Kitchen | Appliances | TOTAL |
|---|---|---|---|---|---|---|
| 1998 | $3,457,000 | $5,894,800 | $7,198,700 | $4,875,400 | $5,947,300 | $27,373,200 |
| 1999 | $3,590,050 | $4,078,900 | $6,057,890 | $5,894,500 | $6,104,500 | $25,725,840 |
| 2000 | $5,789,400 | $6,094,600 | $8,005,600 | $6,934,500 | $7,549,000 | $34,373,100 |
| Total | $12,836,450 | $16,068,300 | $21,262,190 | $17,704,400 | $19,600,800 | $87,472,140 |

# Demand for OLAP

- Traditional tools of report writers, query products, spreadsheets, & language interfaces do not match the user expectations as far as performing multidimensional analysis with complex calculations is concerned.
- Tools used with OLTP and basic DW environments do not match up to the task

# OLAP is the Answer!

***OLAP is a category of software technology that enables analysts, managers, and executives to gain insight into the data through fast, consistent, interactive, access in a wide variety of possible views of information that has been transformed from raw data to reflect the real dimensionality of the enterprise as understood by the user.***

# What is OLAP?

***OLAP software provides the ability to analyze large volumes of information to improve decision making at all levels of an organization.***

# What is OLAP?

***A wide spectrum of multidimensional analysis involving intricate calculations and requiring fast response times.***

# What is OLAP?

***OLAP has two immediate consequences: online part requires the answers of queries to be fast, the analytical part is a hint that the queries itself are complex***

***i.e., Complex questions with Fast Answers!***

# Why a separate OLAP tool?

- **Empowers end users to do own analysis**
- **Frees up IS backlog of report requests**
- **Ease of use**
- **No knowledge of tables or SQL required**

# OLAP Characteristics

- **Multi-user environment**
- **Client-server architecture**
- **Rapid response to queries, regardless of DB size and complexity**

# Guidelines for an OLAP System

- **Multidimensional Conceptual View:** It conforms to how the users perceive business problems
- **Transparency:** It helps to enhance the efficiency and productivity of the users through transparent technology, underlying data repository, computing architecture, and the diverse nature of source data
- **Accessibility:** Provide access only to the data that is actually needed to perform the specific analysis, presenting a single, coherent, and consistent view to the users
- **Consistent Reporting Performance:** Provide consistent run time, response time, or machine utilization
- **Client/Server Architecture:** It provides optimum performance, flexibility, adaptability, and interoperability
- **Generic dimensionality:** one logical structure for all dimensions

# Guidelines for an OLAP System

- **Dynamic Sparse Matrix Handling:** The system must be able to dynamically deduce the distribution of the data and adjust the storage and access to achieve and maintain consistent level of performance
- **Multiuser Support**
- **Unrestricted Cross Dimensional Operations:** Provide ability for the system to recognize dimensional hierarchies and automatically perform roll-up and drill-down operations within a dimension or across dimensions
- **Intuitive Data Manipulation**
- **Flexible reporting:** Every dimension, including any subsets, must be able to be displayed with equal ease
- **Unlimited Dimensions and Aggregation Levels**

# Data Warehouse & OLAP

- **OLAP is a software system that works on top of a DW**
- **A front-end tool for a DW**
- **Information delivery system for the DW**
- **Compliments the information delivery capacities of a DW**

# Why is OLAP useful?

- **Facilitates multidimensional data analysis by pre-computing aggregates across many sets of dimensions**
- **Provides for:**
  - **Greater speed and responsiveness**
  - **Improved user interactivity**

# OLAP Functionalities

**OLAP software provides some or all of the following functionalities:**

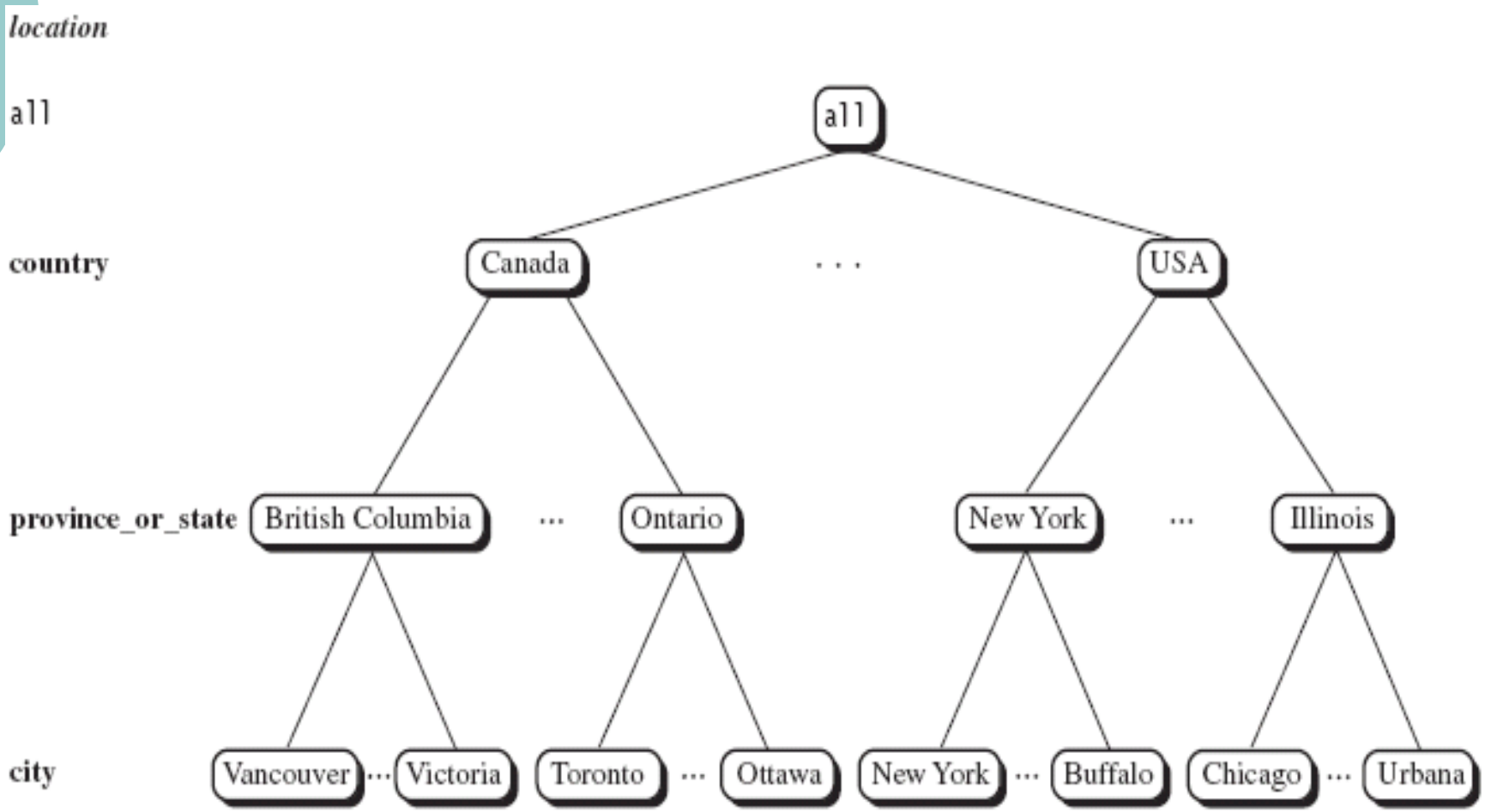
- **Drilling down from high-level summaries to better understand data relationships.**
- **Viewing data from different perspectives, called pivoting or rotation**
- **Comparisons among different elements**
- **Exception reports to highlight unusual situations· Time-series analysis to identify trends**
- **Forecasting with a variety of quantitative techniques**
- **Running model-based what-if simulations to understand the interactions among the different parts of the business**

# Concept Hierarchies

- A **Concept Hierarchy** defines a sequence of mappings from a set of low-level concepts to high-level

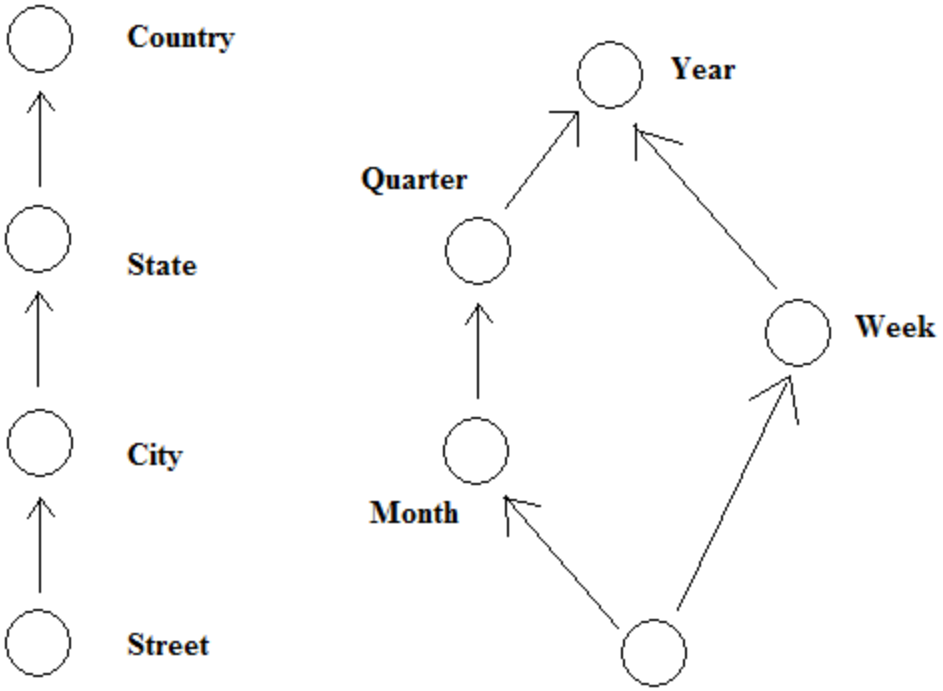
- Consider a concept hierarchy for the dimension **"Location"**

# Concept Hierarchies



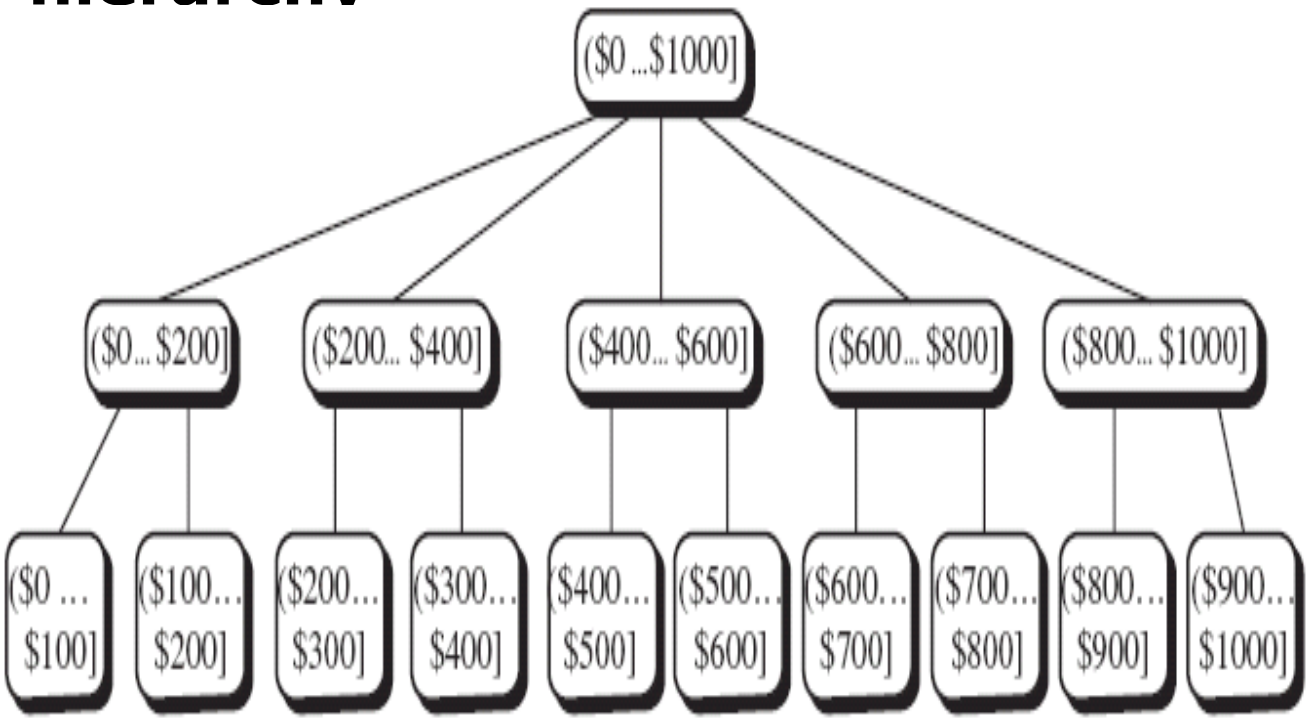
# Concept Hierarchies

- Many concept hierarchies are implicit within the database system



# Concept Hierarchies

- Concept hierarchies may also be defined by grouping values for a given dimension or attribute, resulting in a **set-grouping hierarchv**



# Dimensional Analysis

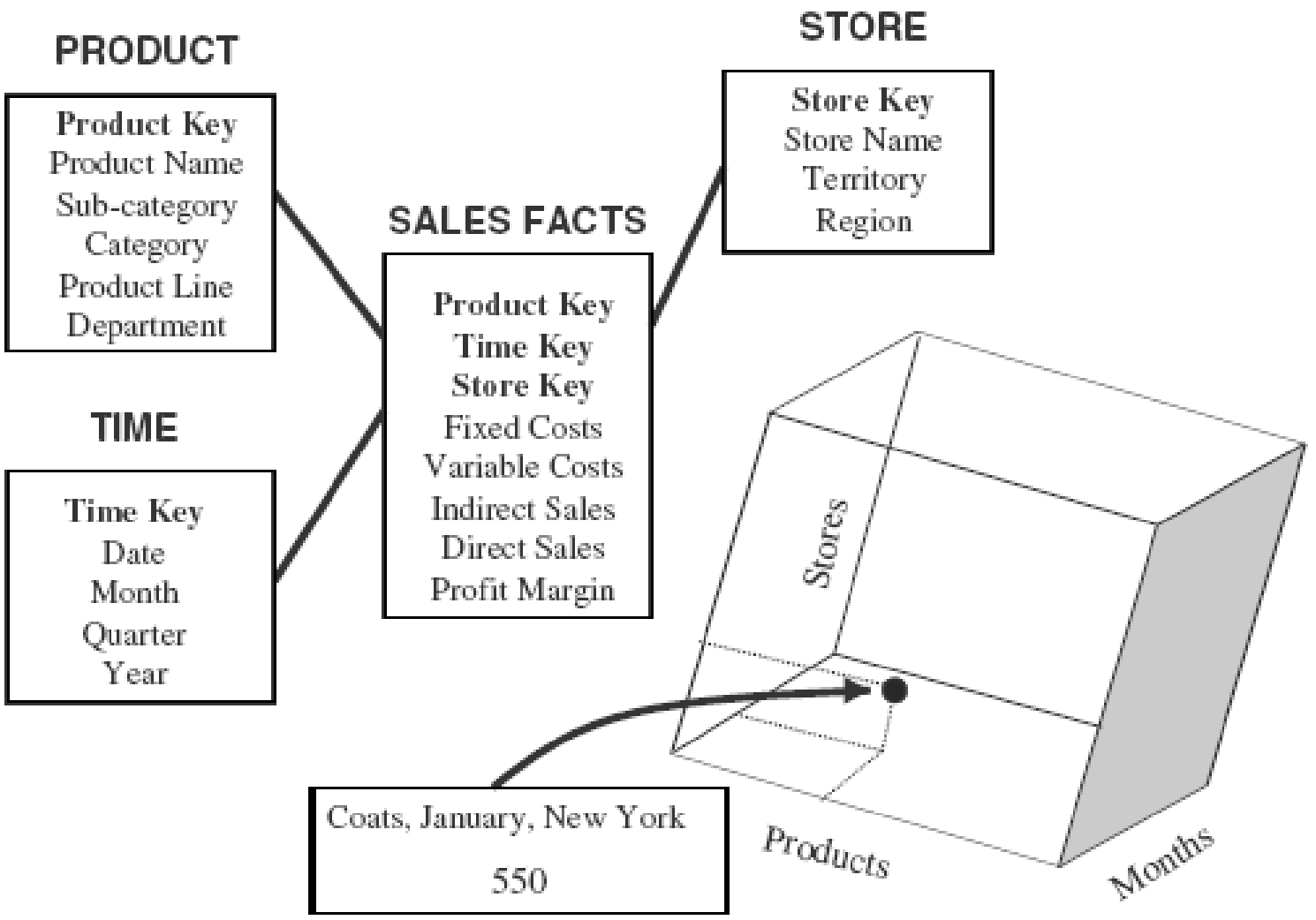



**Figure 15-5** Simple STAR schema.

Store: New York

PAGES: STORE dimension

Products

COLUMNS: PRODUCT dimension

ROWS: TIME dimension

Months

| | Hats | Coats | Jackets | Dresses | Shirts | Slacks |
|---|---|---|---|---|---|---|
| Jan | 200 | 550 | 350 | 500 | 520 | 490 |
| Feb | 210 | 480 | 390 | 510 | 530 | 500 |
| Mar | 190 | 480 | 380 | 480 | 500 | 470 |
| Apr | 190 | 430 | 350 | 490 | 510 | 480 |
| May | 160 | 530 | 320 | 530 | 550 | 520 |
| Jun | 150 | 450 | 310 | 540 | 560 | 330 |
| Jul | 130 | 480 | 270 | 550 | 570 | 250 |
| Aug | 140 | 570 | 250 | 650 | 670 | 230 |
| Sep | 160 | 470 | 240 | 630 | 650 | 210 |
| Oct | 170 | 480 | 260 | 610 | 630 | 250 |
| Nov | 180 | 520 | 280 | 680 | 700 | 260 |
| Dec | 200 | 560 | 320 | 750 | 770 | 310 |

**Figure 15-6** A Three-dimensional display.

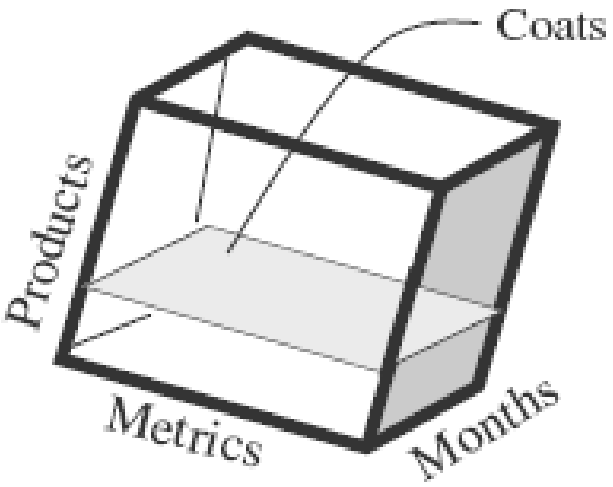
# Hypercubes

PRODUCT: Coats

PAGES: PRODUCT dimension  COLUMNS: Metrics

ROWS: TIME dimension

| | Fixed Cost | Variable Cost | Indirect Sales | Direct Sales | Profit Margin |
|---|---|---|---|---|---|
| Jan | 340 | 110 | 230 | 320 | 100 |
| Feb | 270 | 90 | 200 | 260 | 100 |
| Mar | 310 | 100 | 210 | 270 | 70 |
| Apr | 340 | 110 | 210 | 320 | 80 |
| May | 330 | 110 | 230 | 300 | 90 |
| Jun | 260 | 90 | 150 | 300 | 100 |
| Jul | 310 | 100 | 180 | 300 | 70 |
| Aug | 380 | 130 | 210 | 360 | 60 |
| Sep | 300 | 100 | 180 | 290 | 70 |
| Oct | 310 | 100 | 170 | 310 | 70 |
| Nov | 330 | 110 | 210 | 310 | 80 |
| Dec | 350 | 120 | 200 | 360 | 90 |



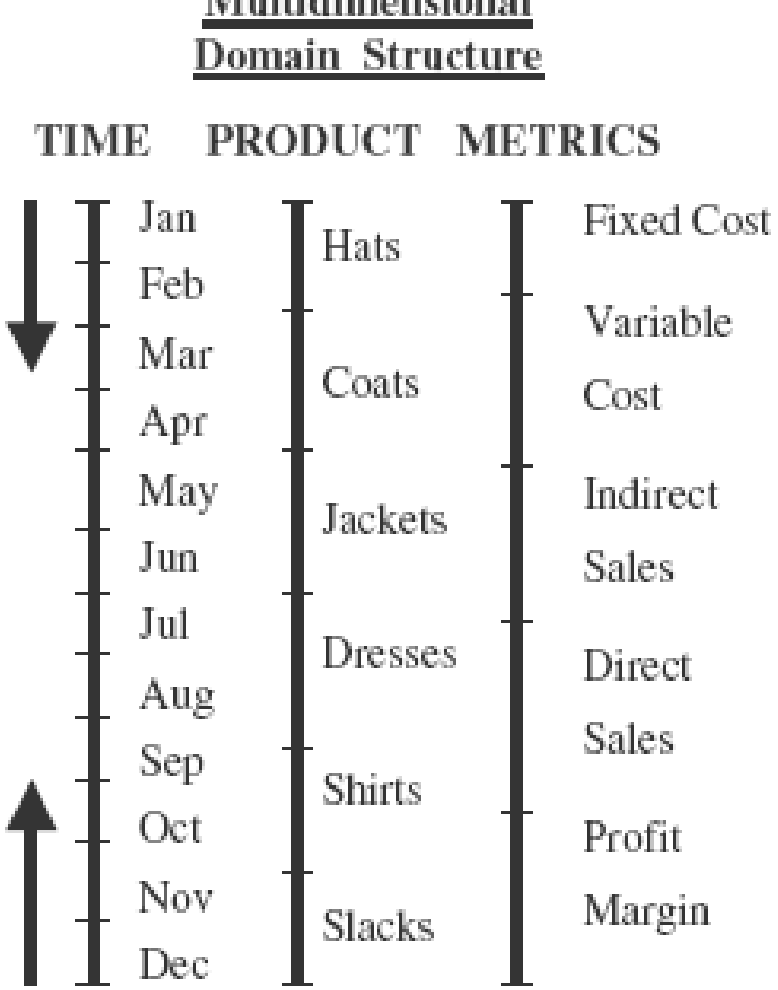




**Figure 15-7** Display of columns, rows, and pages.

- In the figure, note the three straight lines, two of which represent the two business dimensions and the third, the metrics. You can independently move up or down along the straight lines.
- Some experts refer to this representation of a multidimension as a **multidimensional domain structure (MDS)**.

## Multidimensional Domain Structure

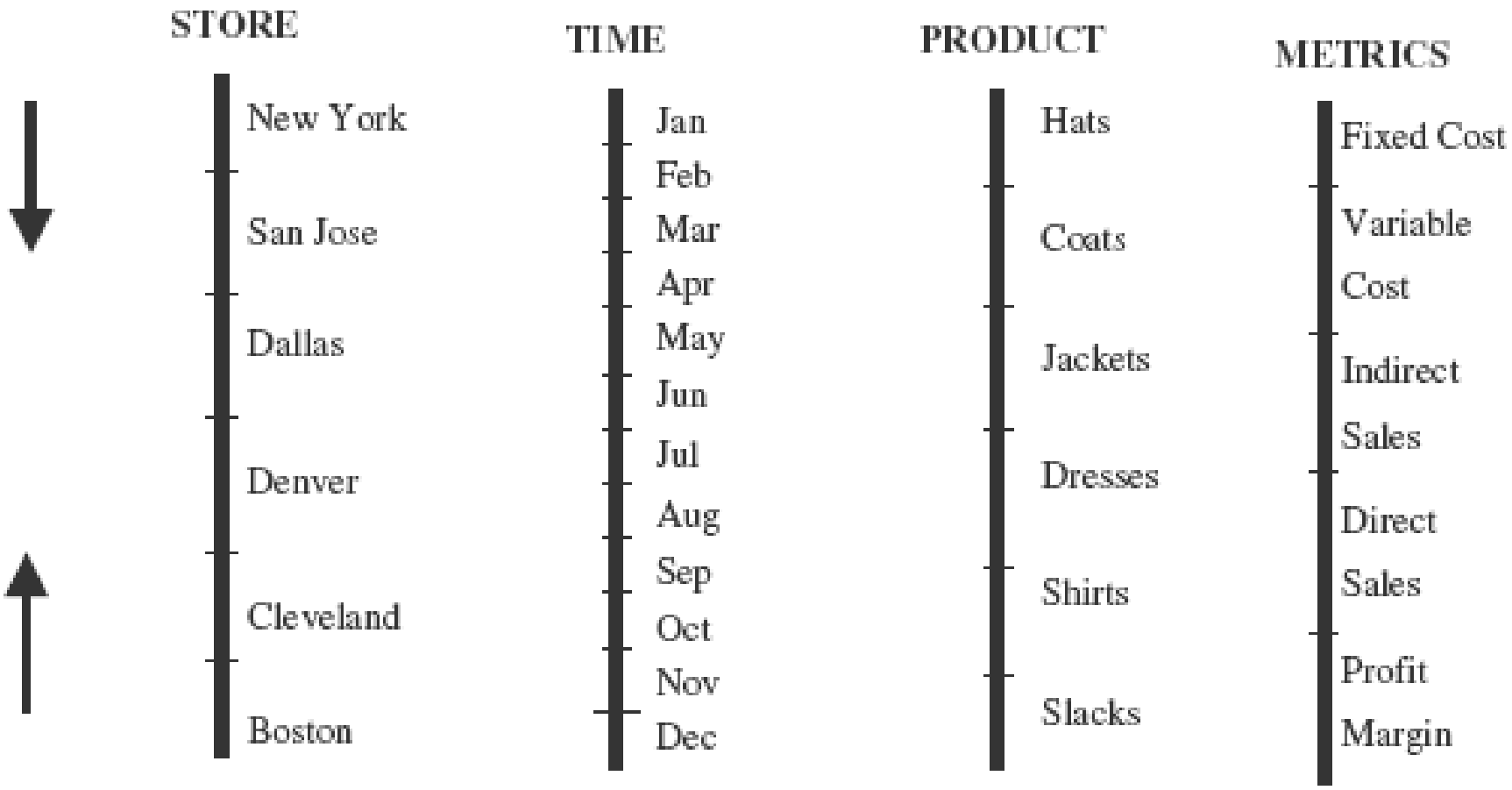
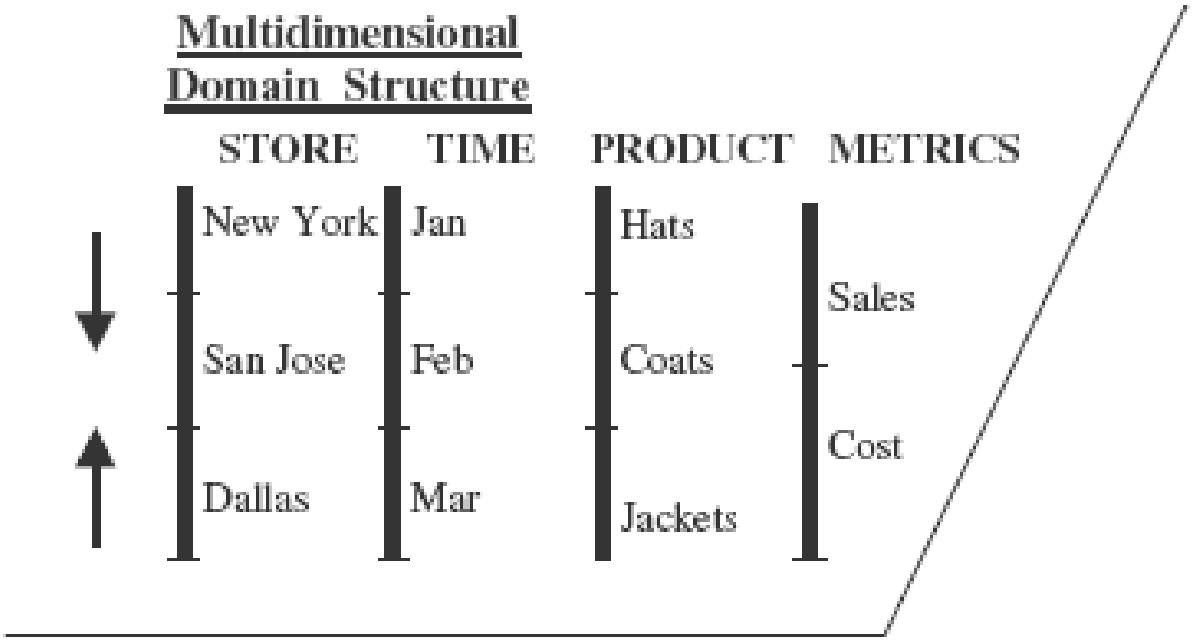



**Figure 15-8** MDS for four dimensions.




New York Store

| | Hats:Sales | Hats:Cost | Coats:Sales | Coats:Cost | Jackets:Sales | Jackets:Cost |
|---|---|---|---|---|---|---|
| Jan | 450 | 350 | 550 | 450 | 500 | 400 |
| Feb | 380 | 280 | 460 | 360 | 400 | 320 |
| Mar | 400 | 310 | 480 | 410 | 450 | 400 |

**Figure 15-9** Page displays for four-dimensional data.

## Multidimensional Domain Structure

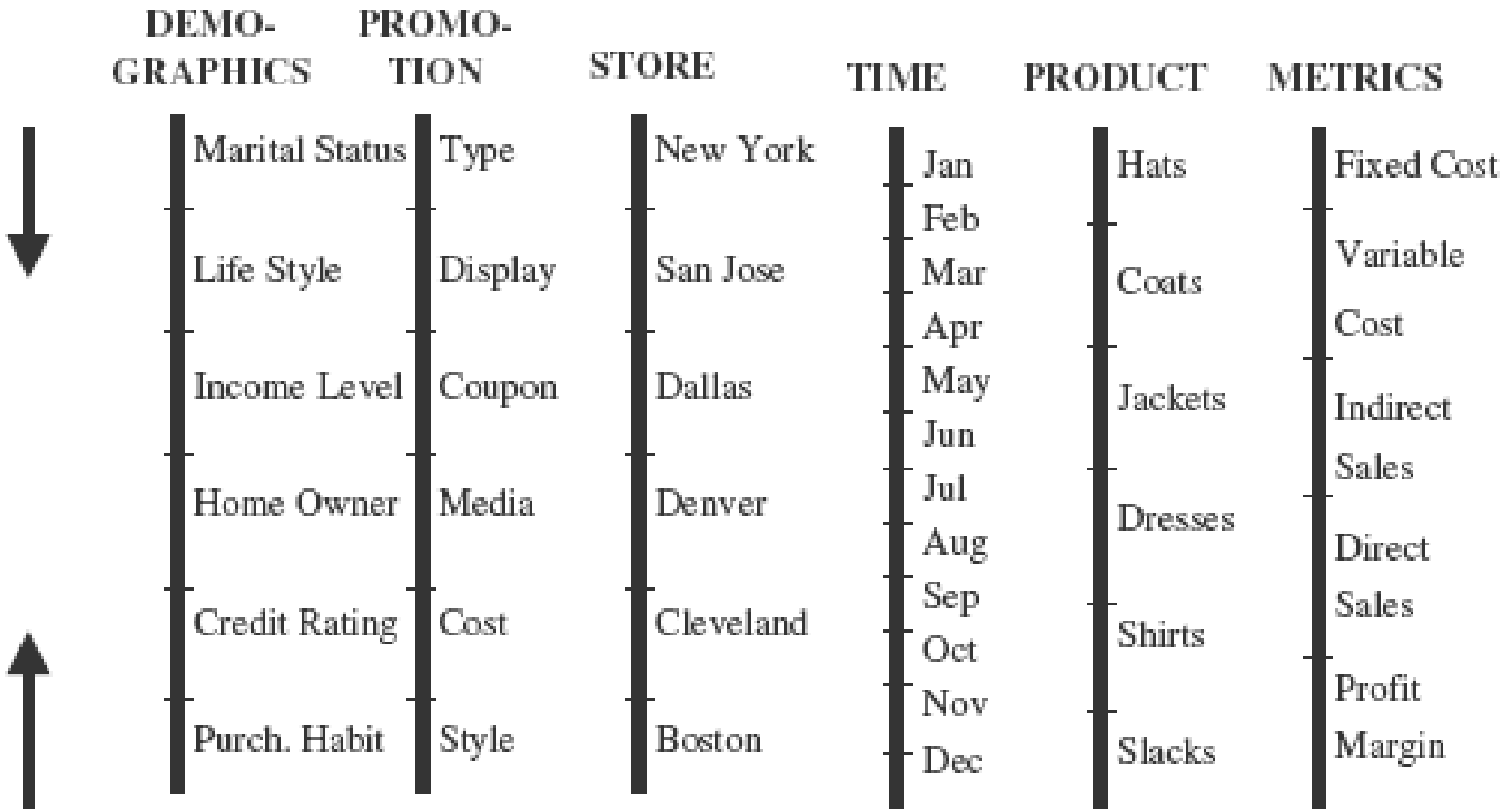
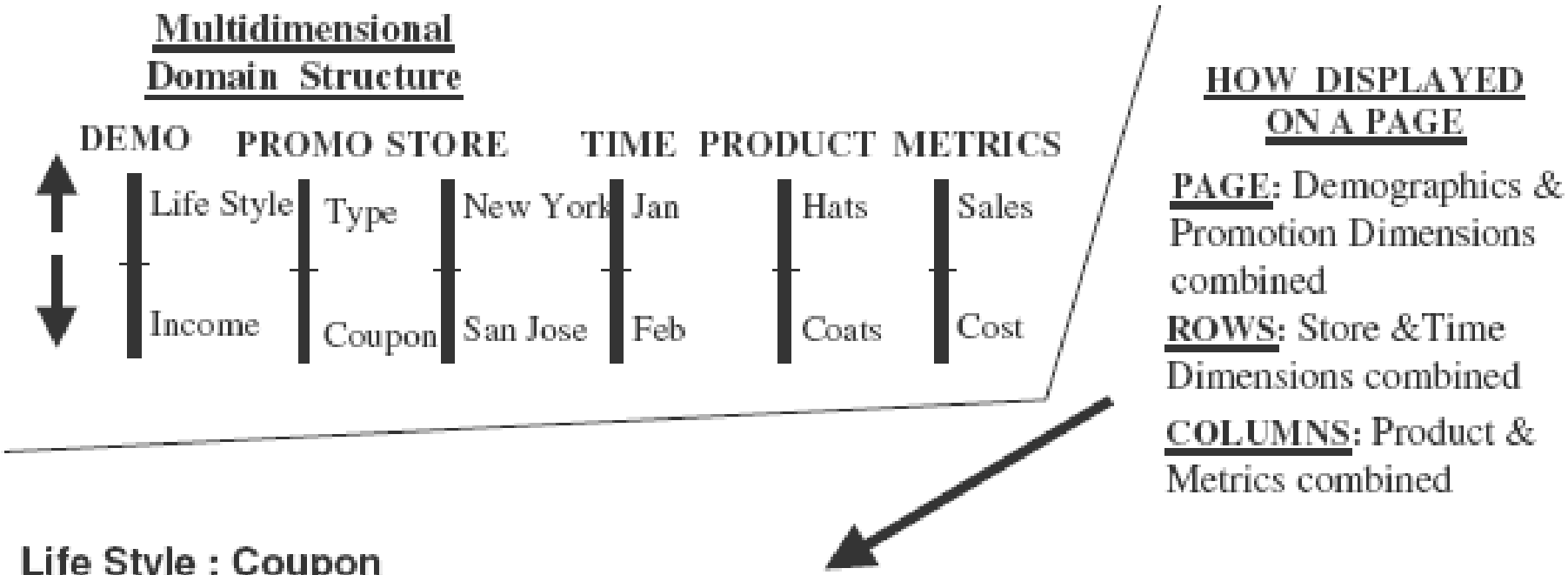



**Figure 15-10** Six-dimensional MDS.




**Life Style : Coupon**

| | | Hats | Hats | Coats | Coats |
|---|---|---|---|---|---|
| | | Sales | Cost | Sales | Cost |
| New York | Jan | 220 | 170 | 270 | 220 |
| | Feb | 190 | 140 | 230 | 180 |
| Boston | Jan | 200 | 160 | 240 | 200 |
| | Feb | 180 | 130 | 220 | 170 |

**Figure 15-11** Page displays for six-dimensional data.

**Drill down: It refers to the process of viewing data at a level of increased detail**

**Roll up: It refers to the process of viewing data with decreasing detail**

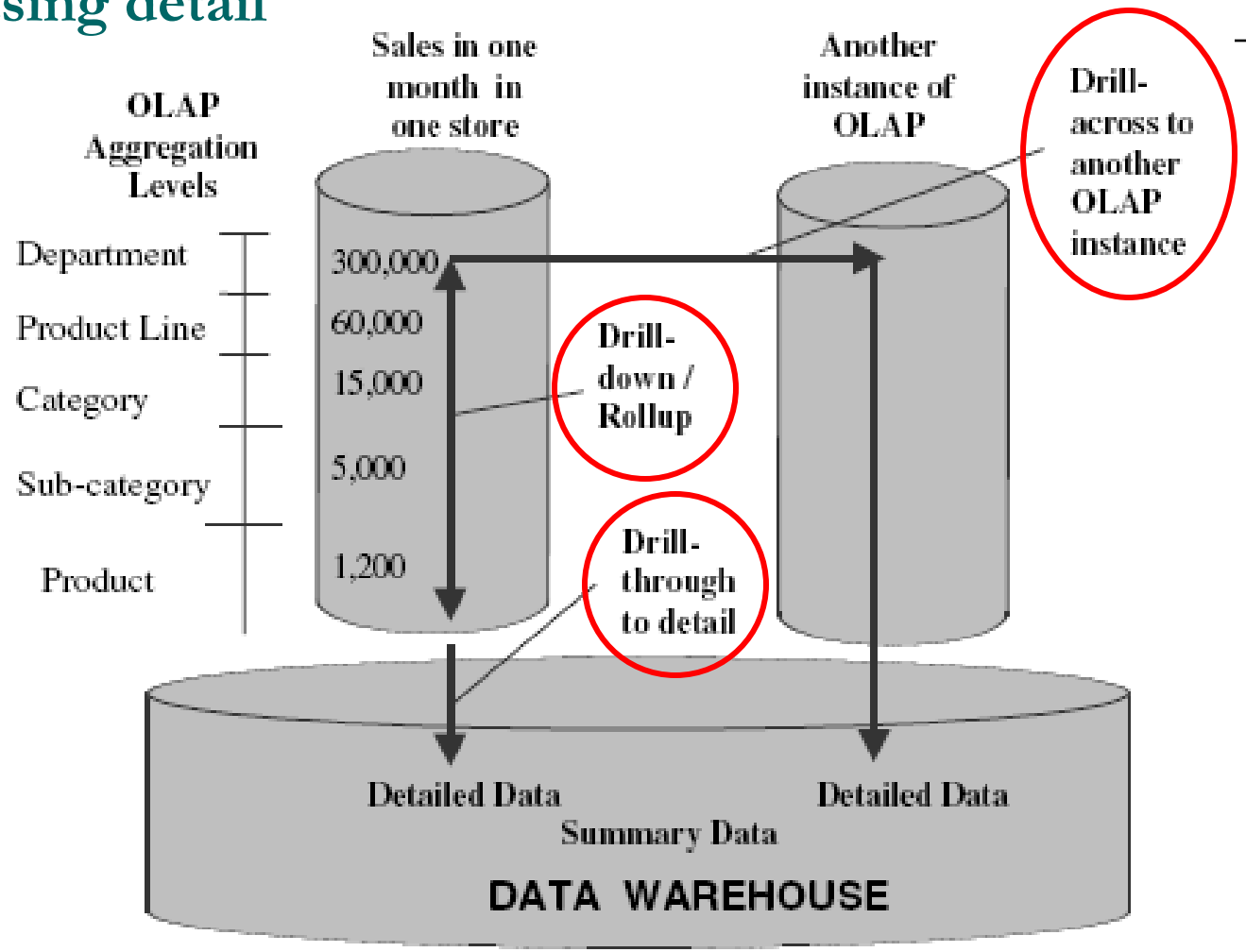



**Figure 15-12** Roll-up and drill-down features of OLAP.

# Example of roll-up

Store: New York

PAGES: STORE dimension

Products

COLUMNS: PRODUCT dimension

ROWS: TIME dimension

Months

| | Hats | Coats | Jackets | Dresses | Shirts | Slacks |
|---|---|---|---|---|---|---|
| Jan | 200 | 550 | 350 | 500 | 520 | 490 |
| Feb | 210 | 480 | 390 | 510 | 530 | 500 |
| Mar | 190 | 480 | 380 | 480 | 500 | 470 |
| Apr | 190 | 430 | 350 | 490 | 510 | 480 |
| May | 160 | 530 | 320 | 530 | 550 | 520 |
| Jun | 150 | 450 | 310 | 540 | 560 | 330 |
| Jul | 130 | 480 | 270 | 550 | 570 | 250 |
| Aug | 140 | 570 | 250 | 650 | 670 | 230 |
| Sep | 160 | 470 | 240 | 630 | 650 | 210 |
| Oct | 170 | 480 | 260 | 610 | 630 | 250 |
| Nov | 180 | 520 | 280 | 680 | 700 | 260 |
| Dec | 200 | 560 | 320 | 750 | 770 | 310 |

**Figure 15-6** A Three-dimensional display.

Store: New York

Sub-categories

PAGES: STORE dimension

COLUMNS: PRODUCT dimension

ROWS: TIME dimension

Months

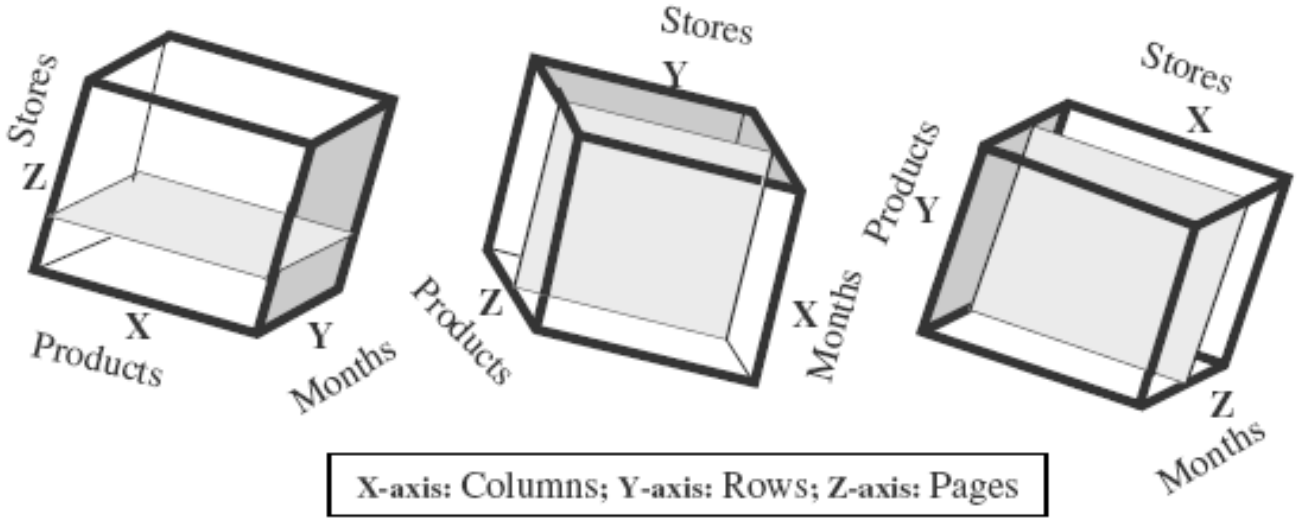
| | Outer | Dress | Casual |
|---|---|---|---|
| Jan | 1,100 | 1,020 | 490 |
| Feb | 1,080 | 1,040 | 500 |
| Mar | 1,050 | 980 | 470 |
| Apr | 970 | 1,000 | 480 |
| May | 1,010 | 1,080 | 520 |
| Jun | 910 | 1,100 | 330 |
| Jul | 880 | 1,120 | 250 |
| Aug | 960 | 1,320 | 230 |
| Sep | 870 | 1,280 | 210 |
| Oct | 910 | 1,240 | 250 |
| Nov | 980 | 1,380 | 260 |
| Dec | 1,080 | 1,520 | 310 |

**Figure 15-13** Three-dimensional display with roll-up.

# Rollup & Drill-down

- OLAP permit users to view data at ay desired level of granularity.
- Rollup: moving from finer-granularity data to coarser granularity
- Drill-down: opposite to Rollup

## Slice n dice: It is an ability to move between different combinations of dimensions when viewing data with an OLAP browser




**Slice-and-Dice or Rotation**

**Store:** New York

| | Hats | Coats | Jackets |
|---|---|---|---|
| Jan | 200 | 550 | 350 |
| Feb | 210 | 480 | 390 |
| Mar | 190 | 480 | 380 |

**Product:** Hats

| | Jan | Feb | Mar |
|---|---|---|---|
| New York | 200 | 210 | 190 |
| Boston | 210 | 250 | 240 |
| San Jose | 130 | 90 | 70 |

**Month:** January

| | New York | Boston | San Jose |
|---|---|---|---|
| Hats | 200 | 210 | 130 |
| Coats | 550 | 500 | 200 |
| Jackets | 350 | 400 | 100 |

**Figure 15-14** Slicing and dicing.

# Slicing & Dicing

- Additional Functionality that can be thought of as viewing a slice of the data cube, particularly when values for multiple dimensions are fixed.
- Slicing/Dicing simply consists of selecting specific values for these attributes, which are then displayed on top of the cross-tab

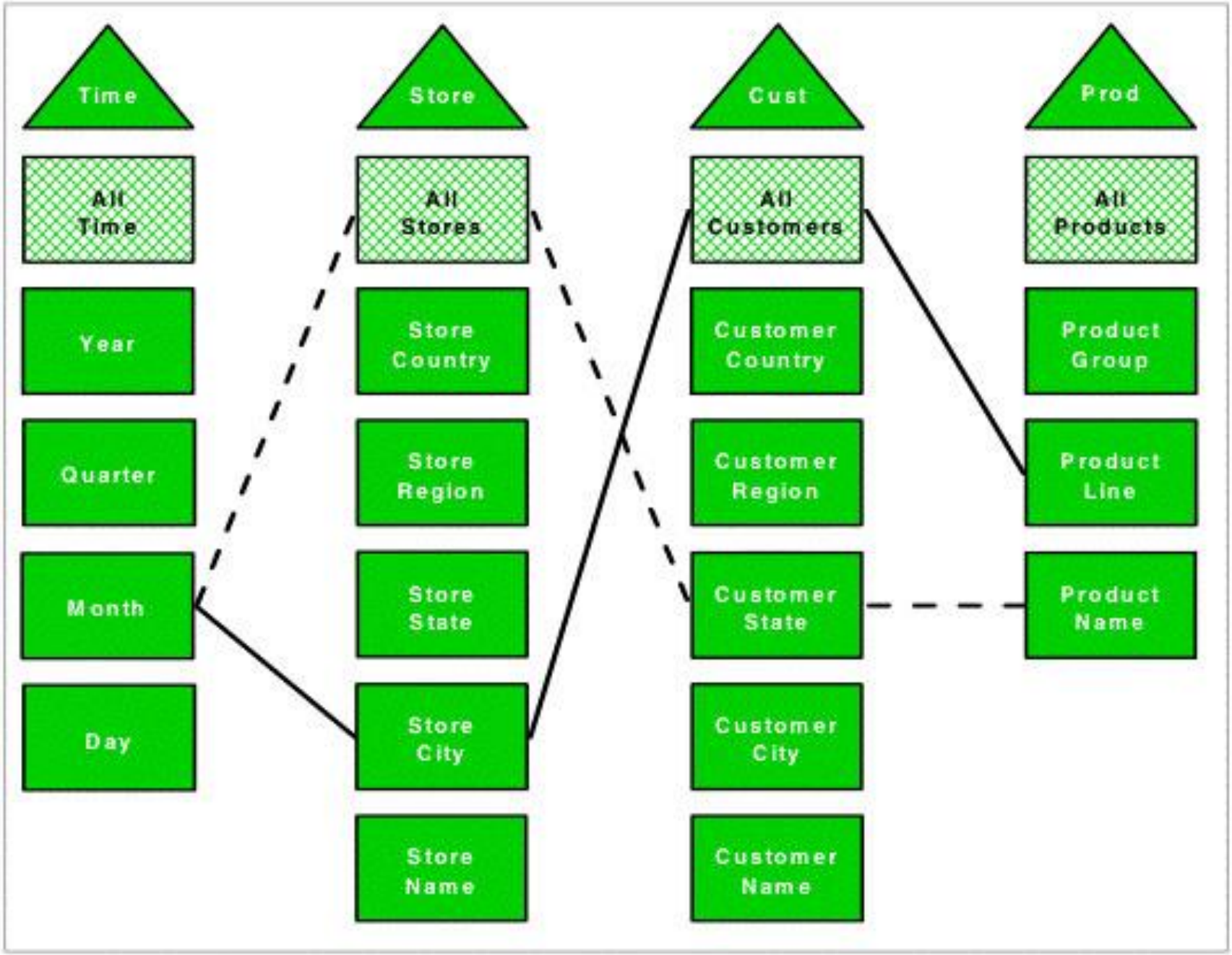
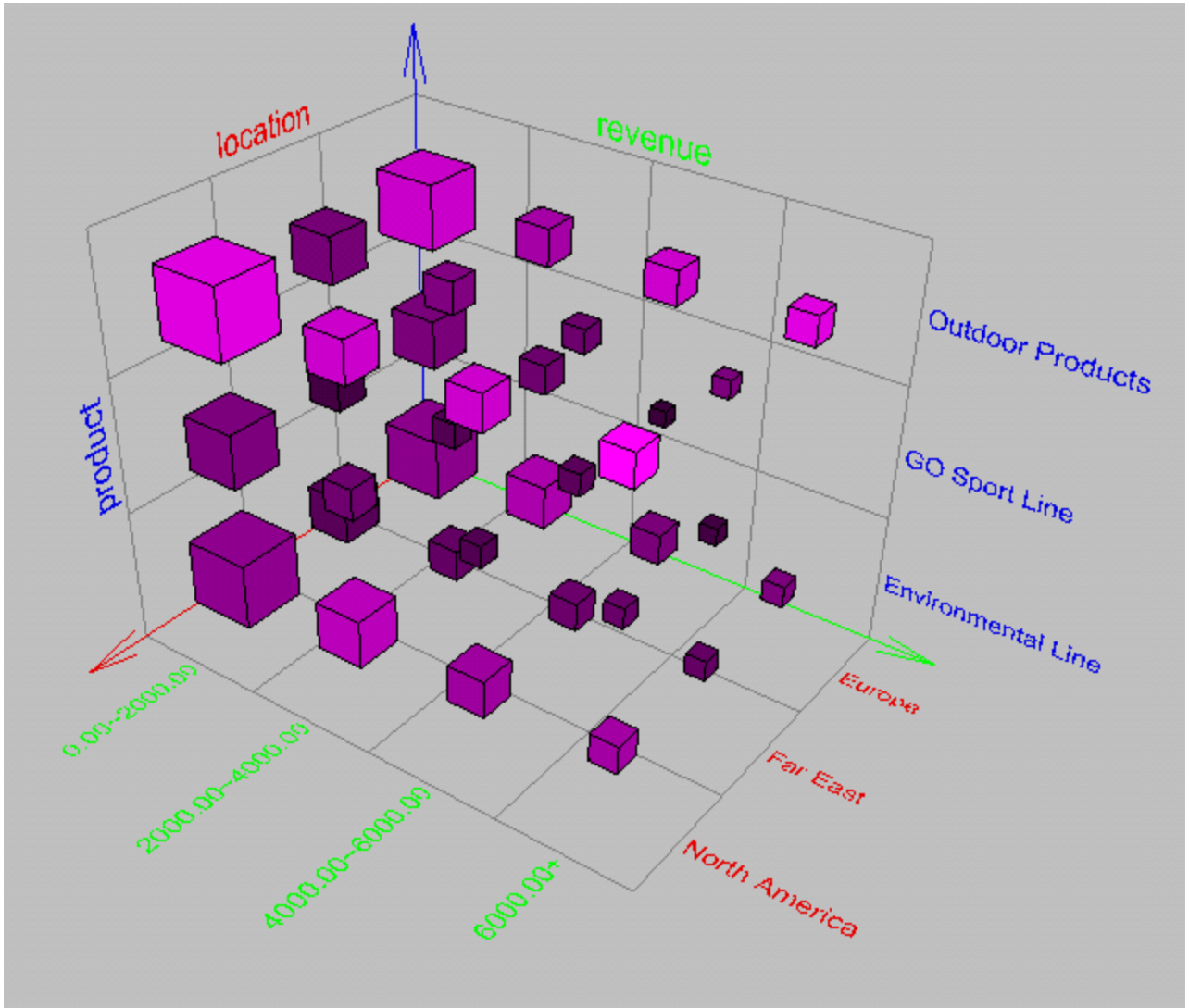


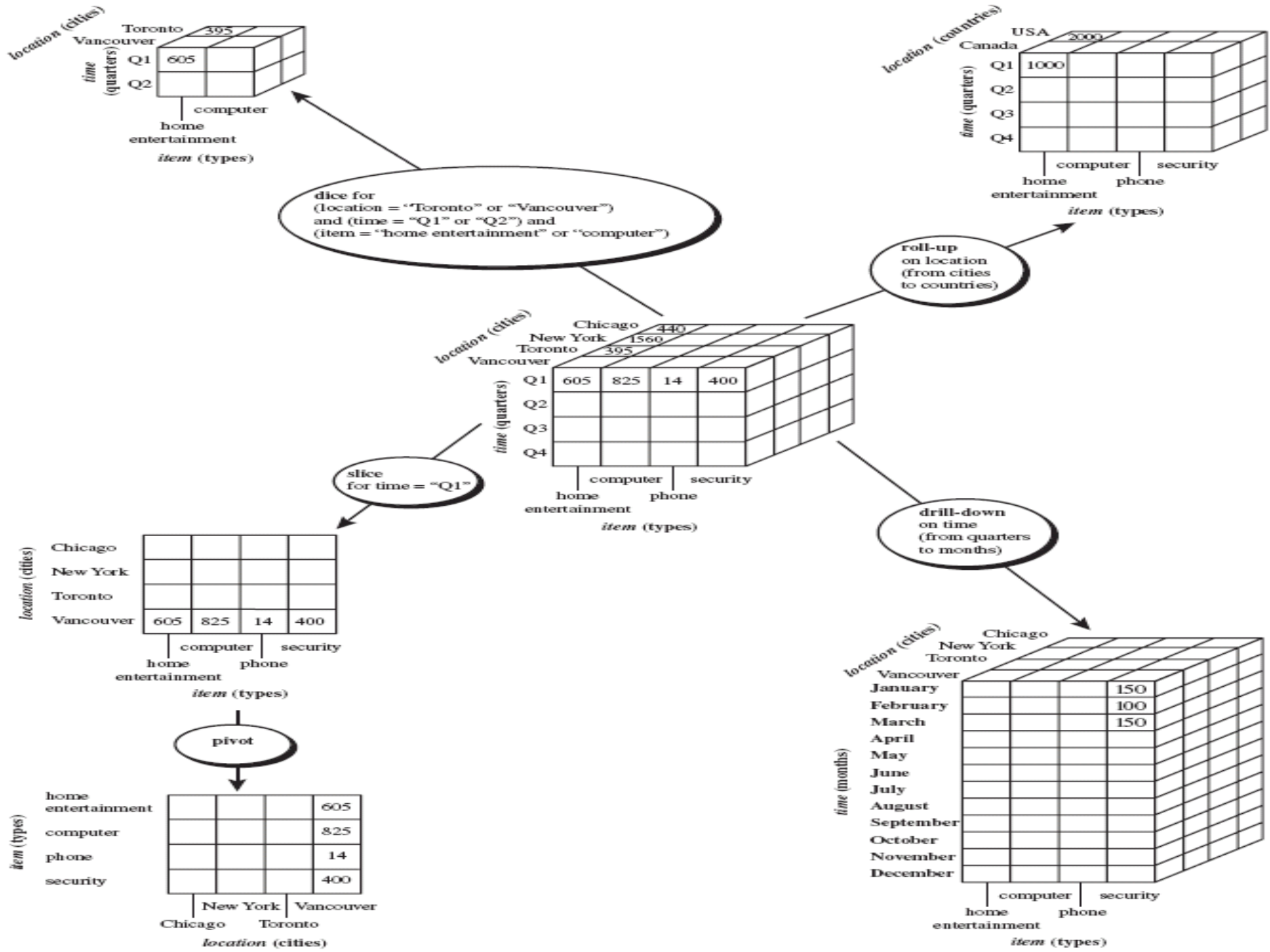

*Figure 1-2 Database slices*

# Browsing a Data Cube



# OLAP Operation

- So, how are *concept hierarchies* useful in OLAP?
- In the multidimensional model, data are organized into multiple dimensions,
- And each dimension contains multiple levels of abstraction defined by concept hierarchies


location (cities)
Toronto
Vancouver
395
time (quarters)
Q1
Q2
605
computer
home entertainment
item (types)
location (countries)
USA
Canada
2000
Q1
Q2
Q3
Q4
1000
time (quarters)
computer
security
home entertainment
phone
item (types)
dice for
(location = "Toronto" or "Vancouver")
and (time = "Q1" or "Q2") and
(item = "home entertainment" or "computer")
roll-up
on location
(from cities
to countries)
location (cities)
Chicago
New York
Toronto
Vancouver
440
1560
395
Q1
Q2
Q3
Q4
605
825
14
400
time (quarters)
computer
security
home entertainment
phone
item (types)
slice
for time = "Q1"
drill-down
on time
(from quarters
to months)
location (cities)
Chicago
New York
Toronto
Vancouver
605
825
14
400
computer
security
home entertainment
phone
item (types)
pivot
item (types)
home entertainment
computer
phone
security
605
825
14
400
New York
Vancouver
Chicago
Toronto
location (cities)
location (cities)
Chicago
New York
Toronto
Vancouver
January
February
March
April
May
June
July
August
September
October
November
December
150
100
150
time (months)
computer
security
home entertainment
phone
item (types)

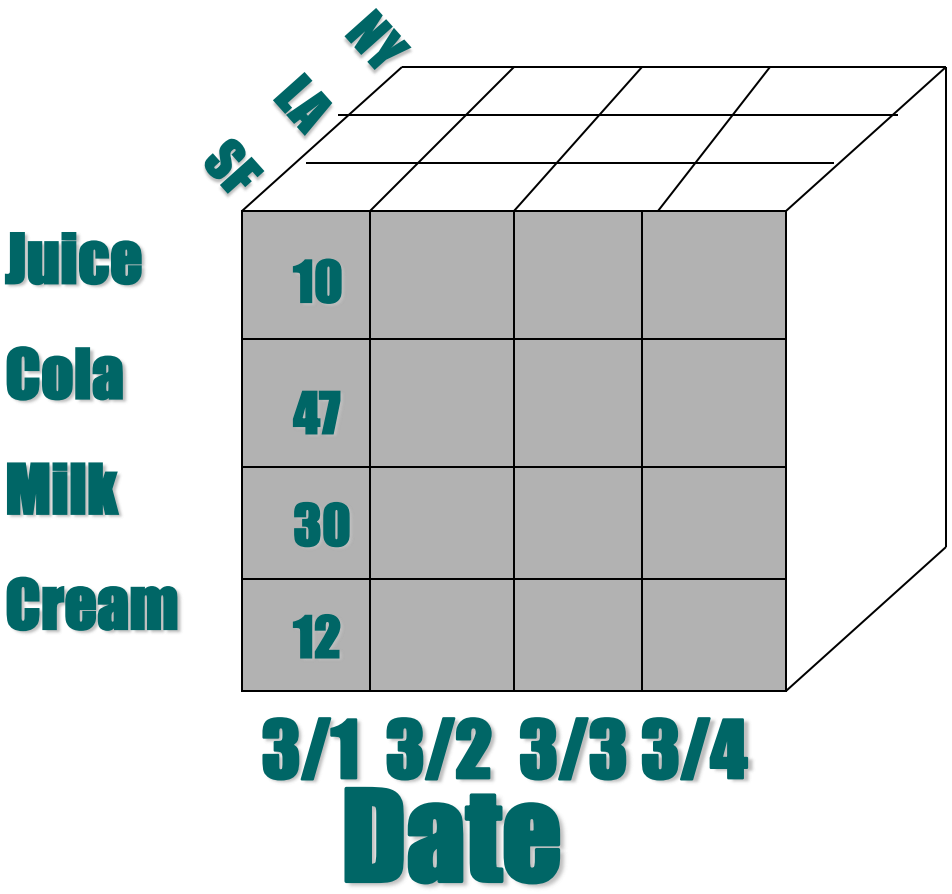
# Typical OLAP Operations

- Roll up (drill-up): summarize data
  - *by climbing up hierarchy or by dimension reduction*
- Drill down (roll down): reverse of roll-up
  - *from higher level summary to lower level summary or detailed data, or introducing new dimensions*

# Typical OLAP Operations

- Slice and dice: *project and select*
- Pivot (rotate):
  - *reorient the cube, visualization, 3D to series of 2D planes*

# Multidimensional Data

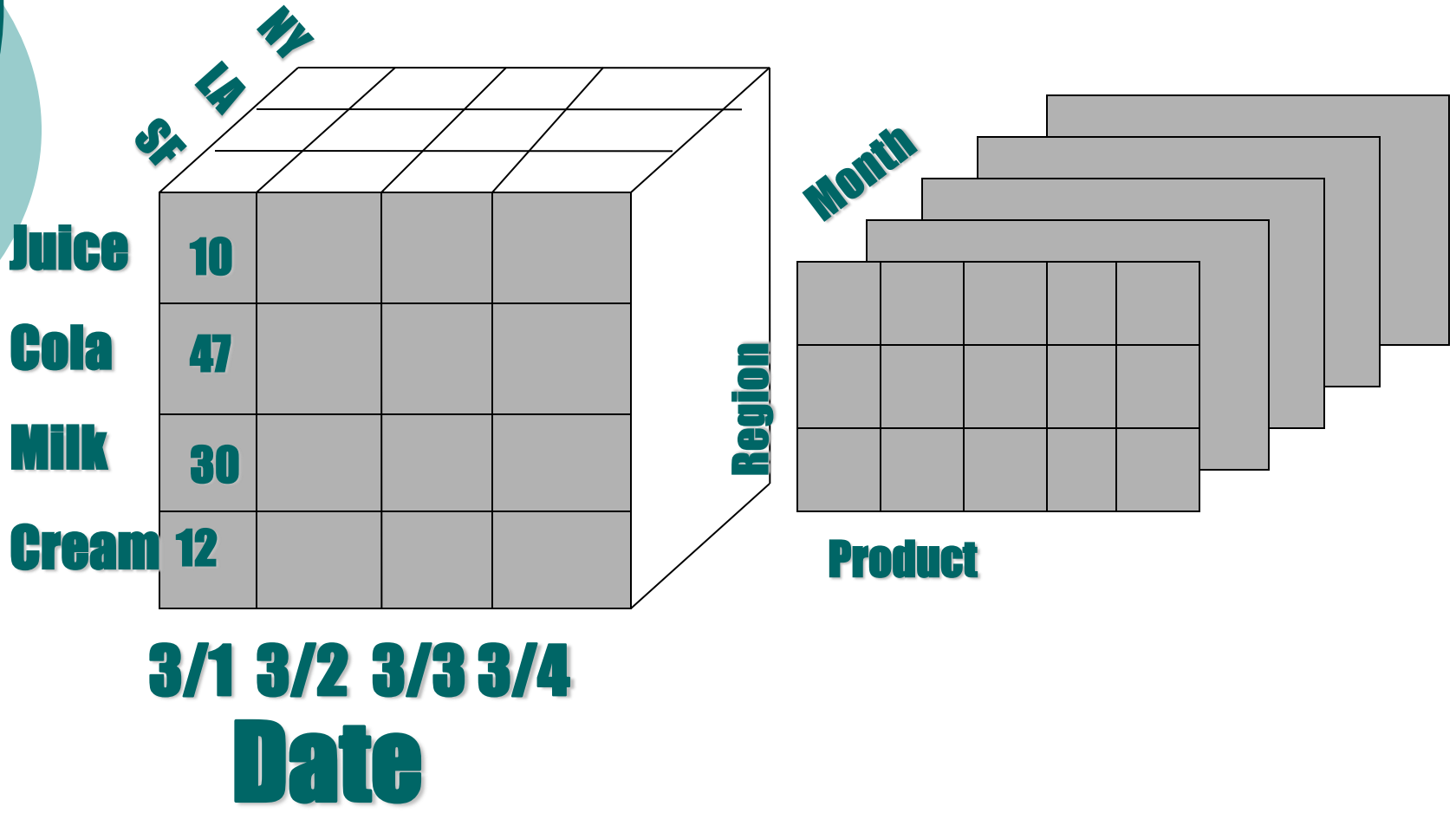



**Sales Volume as a function of time, city and product**

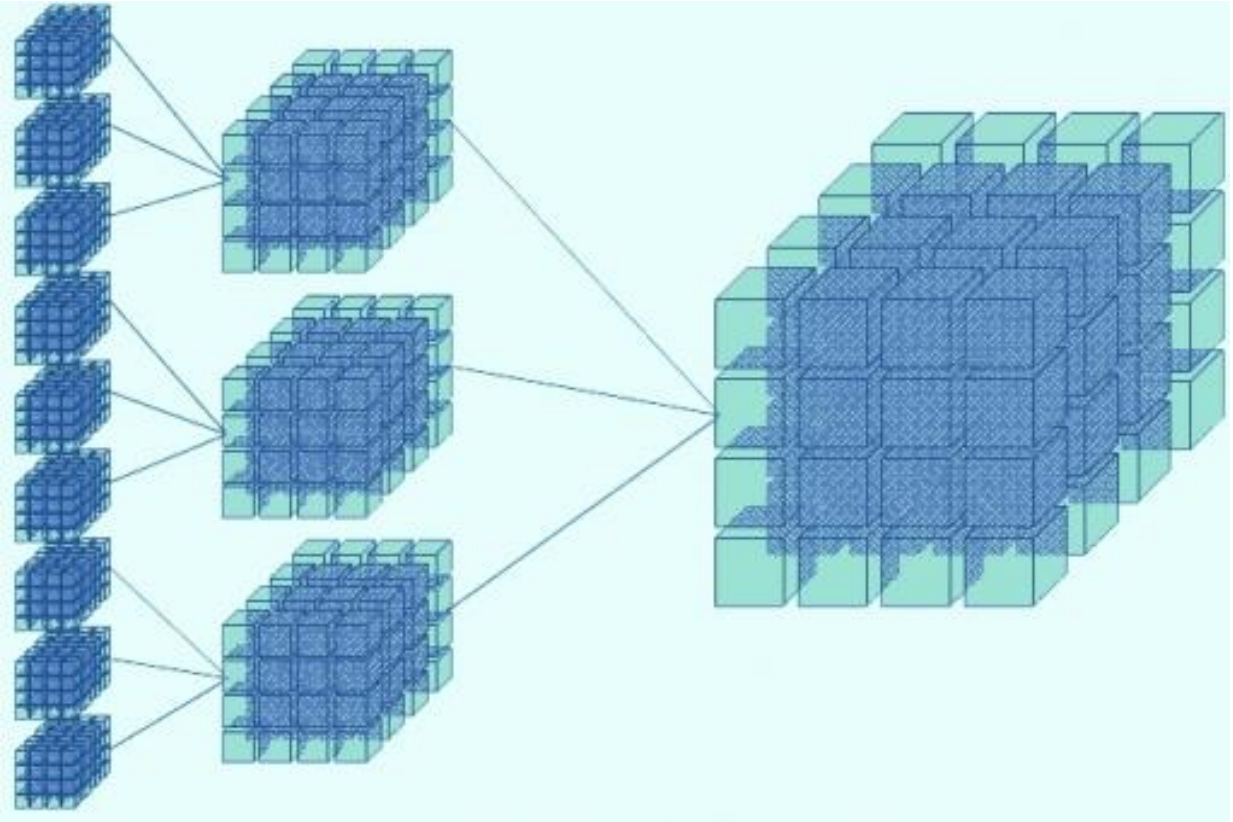
# Operations in Multidimensional Data Model

- Aggregation (*roll-up*)
  - dimension reduction:  e.g., total sales by city
  - summarization over aggregate hierarchy: e.g., total sales by city and year -> total sales by region and by year
- Selection (*slice*) defines a subcube
  - e.g., sales where city = Palo Alto and date = 1/15/96
- Navigation to detailed data (*drill-down*)
  - e.g., (sales - expense) by city, top 3% of cities by average income
- Visualization Operations (e.g., Pivot)

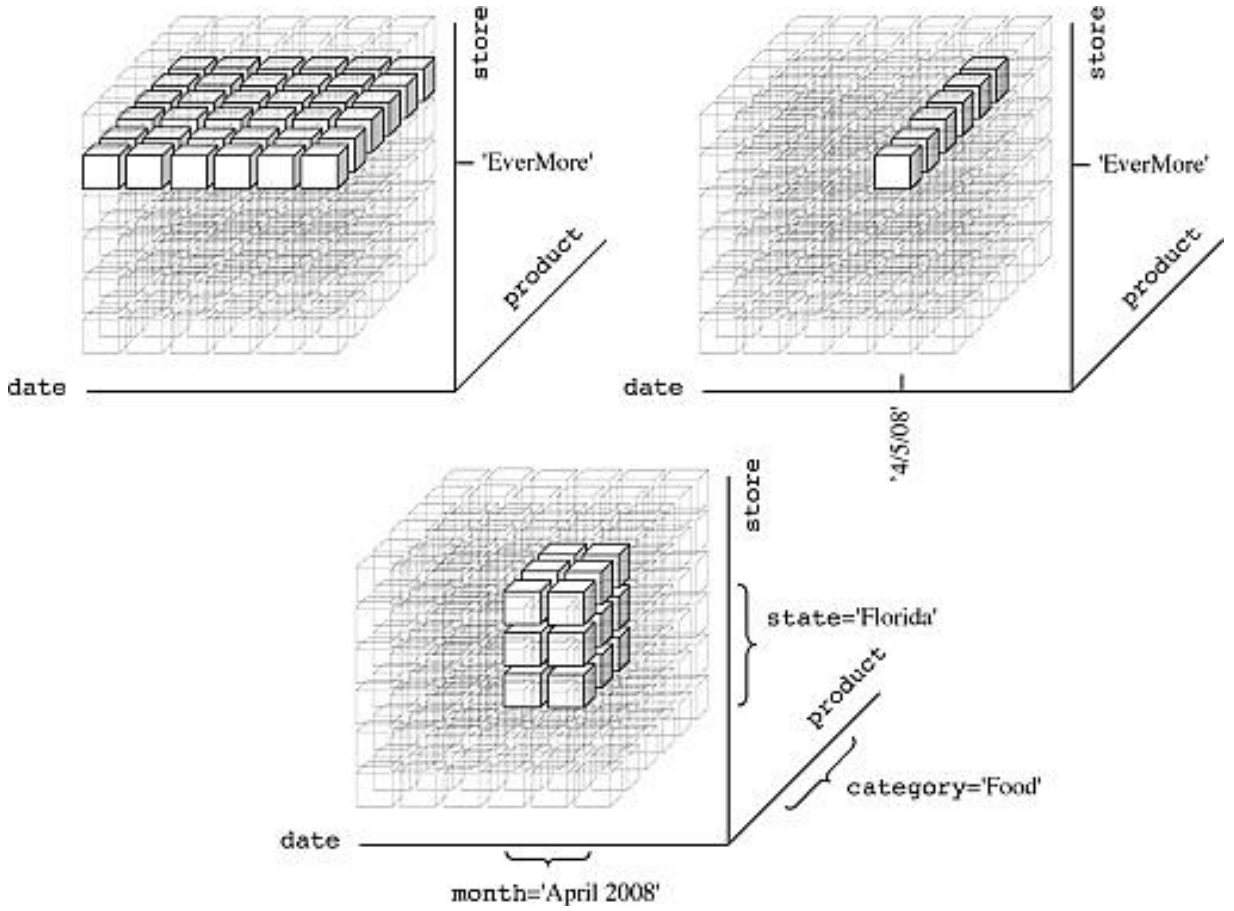
# A Visual Operation: Pivot (Rotate)



***Roll-up*:** (Aggregate, Consolidate) A roll-up involves computing all of the data relationships for one or more dimensions. To do this, a computational relationship or formula might be defined.

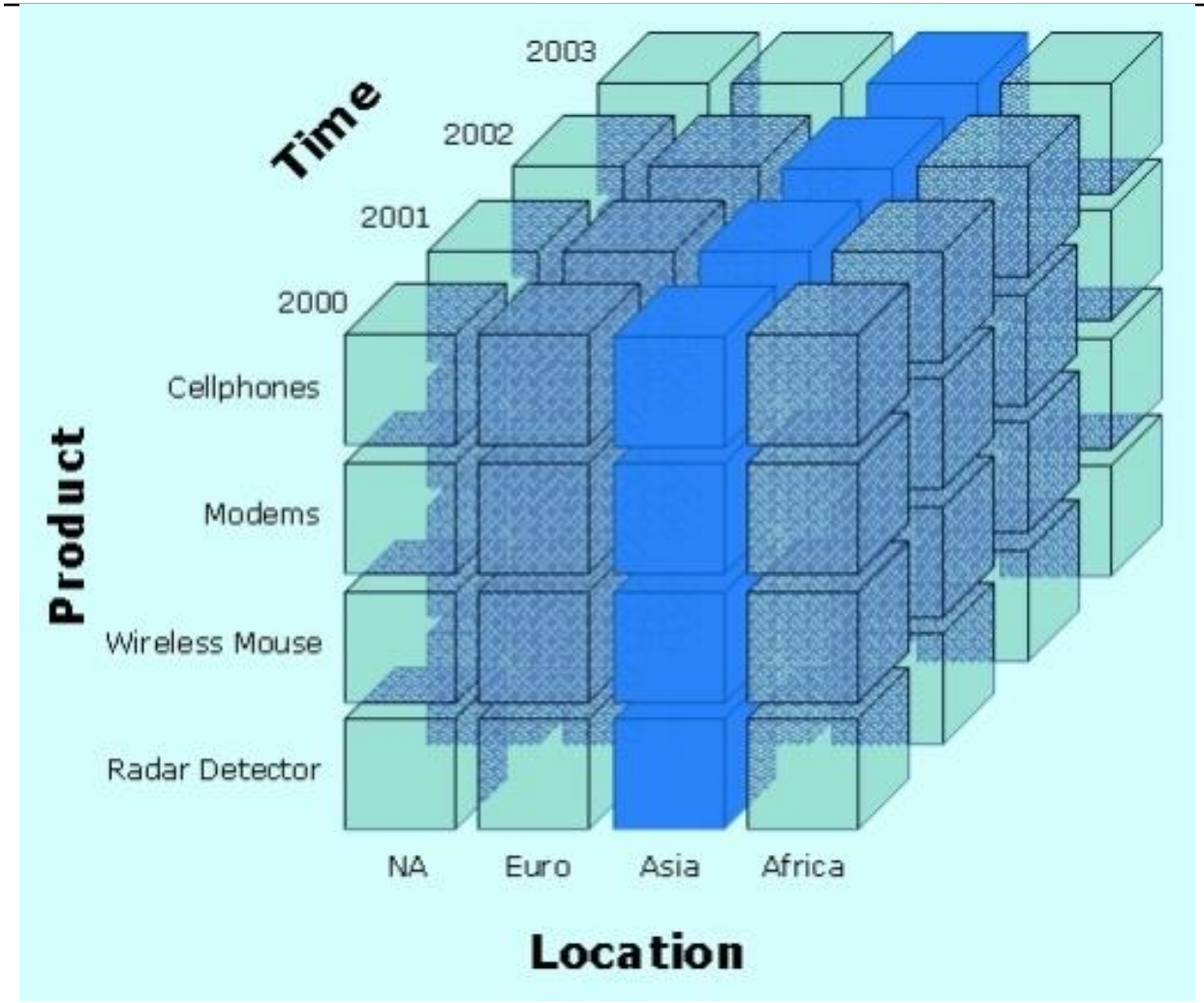


***Drill Down*:** Drilling down is a specific analytical technique whereby the user navigates among levels of data ranging from the most summarized (up) to the most detailed (down).



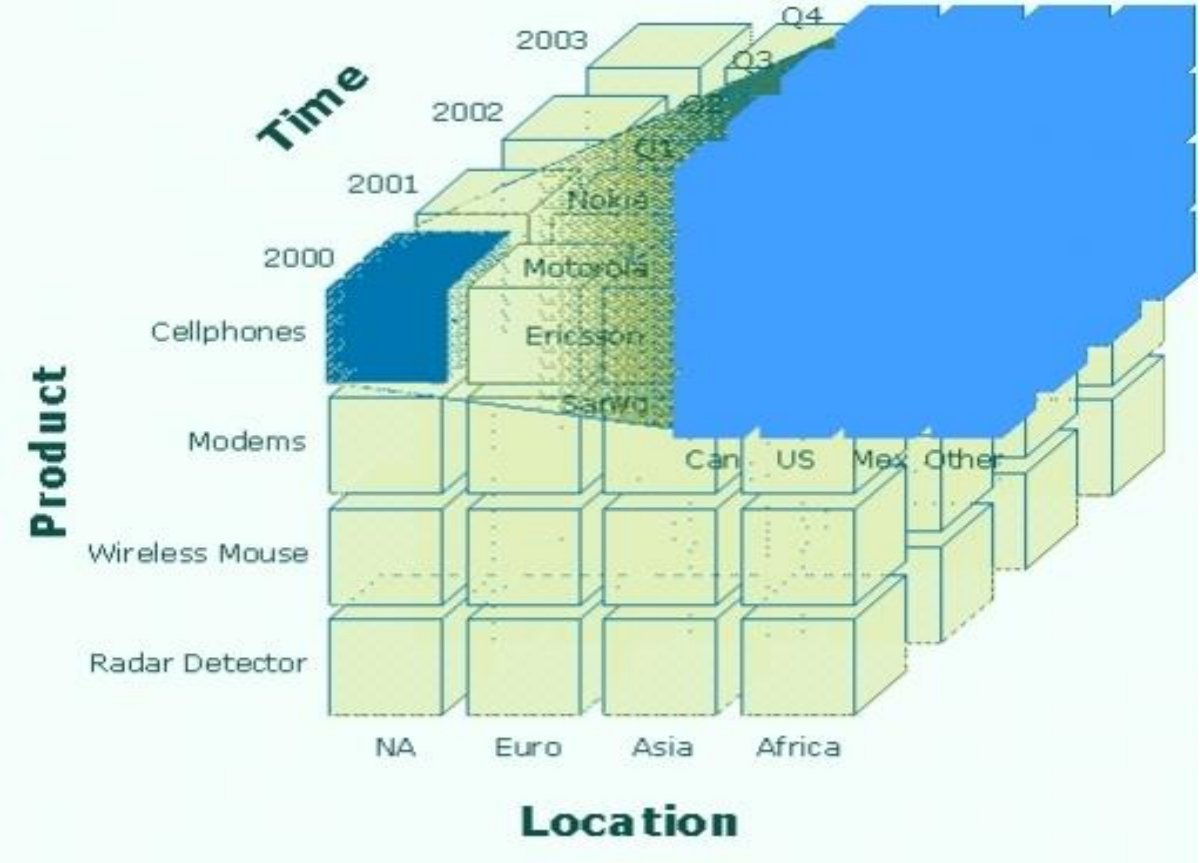
## *Slice*:

A slice is a subset of a multi-dimensional array corresponding to a single value for one or more members of the dimensions not in the subset.

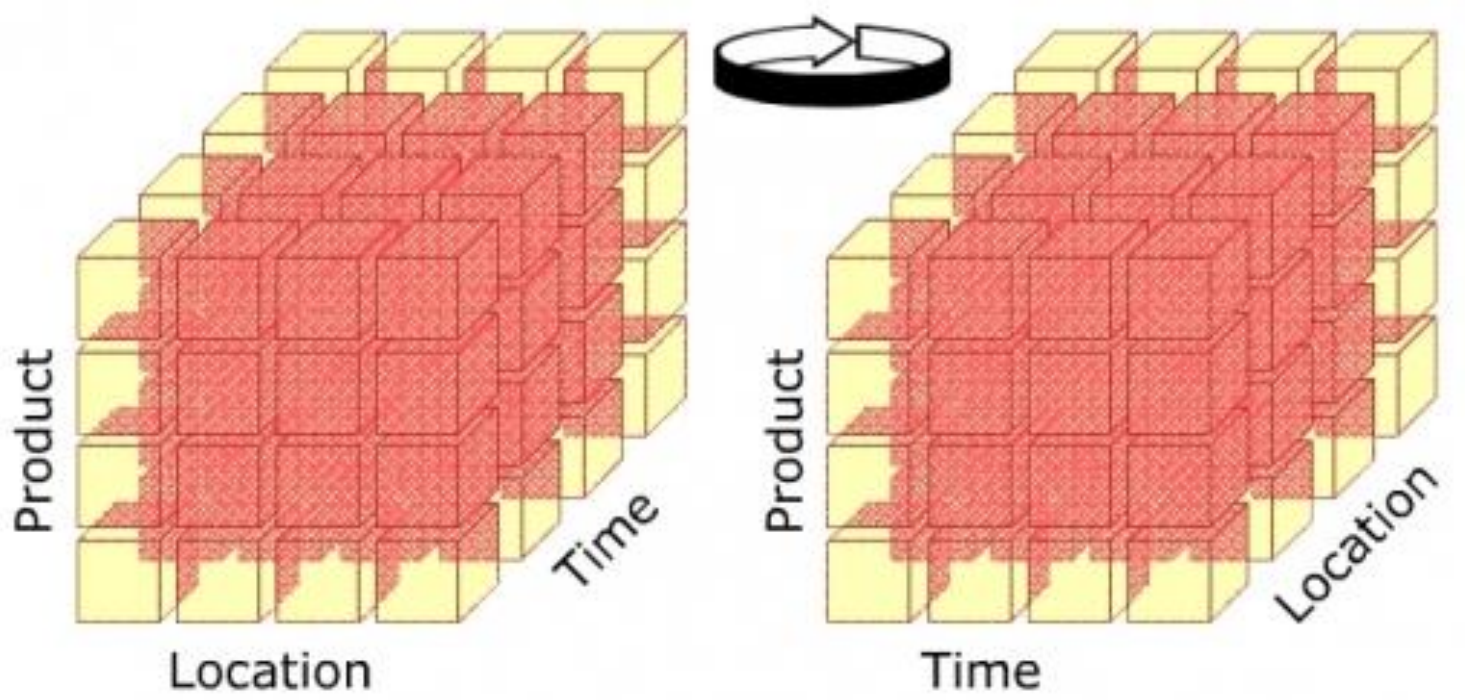


# *Dice*:

The dice operation is a slice on more than two dimensions of a data cube (or more than two consecutive slices).



***Pivot***:This operation is also called rotate operation. It rotates the data in order to provide an alternative presentation of data – the report or page display takes a different dimensional orientation.



# Uses and Benefits

- Increased productivity of business managers, analysts, and executives
- Faster delivery of applications by IT people
- Self sufficiency of users, resulting in reduction in backlog
- More efficient operation through reducing time on query executions and in net work traffic
- Ability to model real world challenges with business metrics and dimensions