SNS COLLEGE OF TECHNOLOGY

Coimbatore-35
An Autonomous Institution

Accredited by NBA – AICTE and Accredited by NAAC – UGC with 'A+' Grade
Approved by AICTE, New Delhi & Affiliated to Anna University, Chennai

DEPARTMENT OF INFORMATION TECHNOLOGY

Data Mining and Warehousing

COURSE NAME: Data Mining and Warehousing

COURSE CODE: 19ITT301

SEMESTER: 5

CONTENTS:

- Data Cube
- Multidimensional Data Model
- Schemas
- Dimensions
- Concept Hierarchies
- Measures
- OLAP Operations

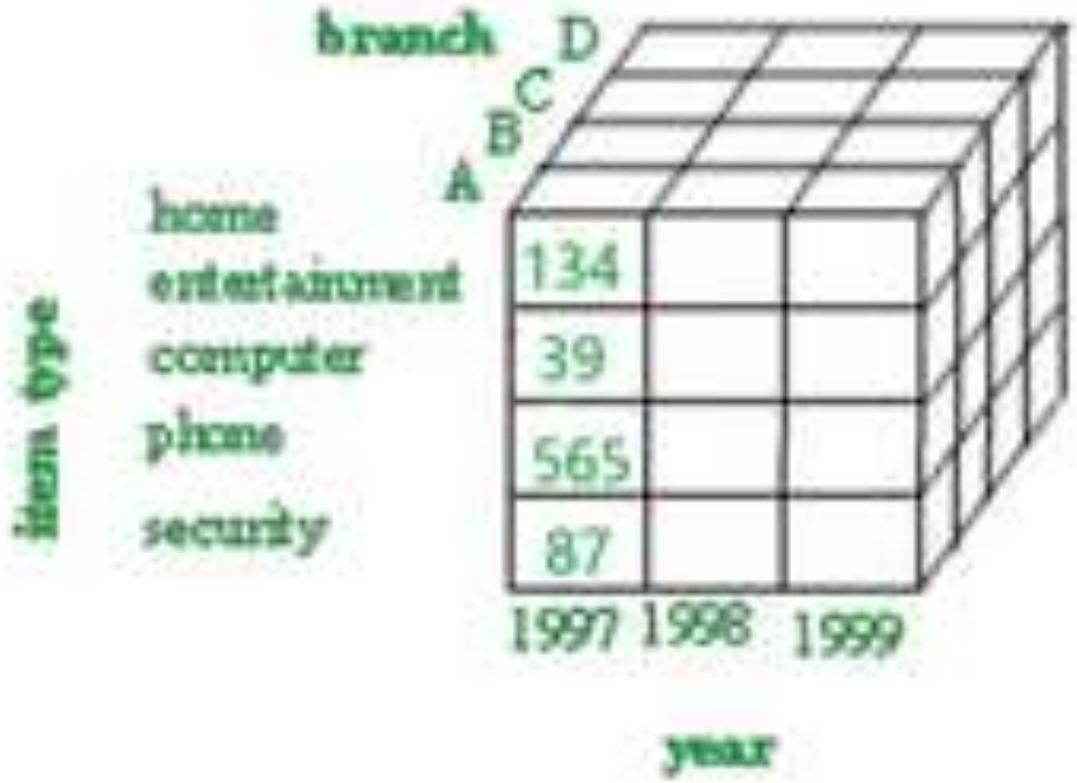
Data Cube :

A **data cube** is a multi-dimensional array of values used to represent data in a structured manner for analysis and reporting. It is particularly useful in the realm of **Online Analytical Processing (OLAP)**, enabling businesses to perform complex queries and analysis efficiently. By organizing data into multiple dimensions, data cubes facilitate a more intuitive understanding of data relationships and trends.

Real-time Applications Of Data Cubes:

1. Business Intelligence and Analytics
2. Fraud Detection and Prevention
3. Customer Behavior and Personalization
4. IoT and Smart Cities
5. Healthcare Analytics

DATA CUBE


branch
D
C
B
A
item type
home
entertainment
computer
phone
security
134
39
565
87
1997
1998
1999
year

Components

1. Dimensions

Dimensions are the axes of the cube that define the structure and context of the data. Each dimension can have different levels of granularity.

- **Time**: This dimension allows analysis over various periods, such as years, quarters, months, weeks, or days. Users can track sales trends, seasonality, and performance over time.
- **Location**: Represents geographical data, such as countries, states, cities, or specific stores. Analyzing sales by location helps businesses identify market opportunities and understand regional performance.

- **Products:** This dimension categorizes items sold, which can include product lines, brands, or specific SKUs (Stock Keeping Units). Understanding product performance helps in inventory management and marketing strategies.

- **Customer:** A dimension that captures customer data, including demographics, preferences, and purchase history. Analyzing customer behavior can drive targeted marketing efforts.

2. Measures

Measures are the quantitative data points stored in the cube that can be aggregated across dimensions. Examples include:

- **Sales Revenue:** Total income generated from sales transactions, often a primary measure for retail businesses.
- **Units Sold:** The total number of products sold during a specific period, providing insight into sales volume and product popularity.
- **Profit:** The amount earned after subtracting costs from revenues, crucial for assessing business profitability.
- **Customer Count:** The number of unique customers who made purchases, helping businesses gauge their customer base growth.

OLAP Operations

OLAP operations provide the ability to interact with and manipulate data cubes, enabling users to perform complex analysis

Slice:

The slice operation extracts a specific sub-cube by fixing one or more dimensions. For example, slicing the time dimension to view sales data for just the year 2023 provides a focused analysis without the noise of other years.

Dice:

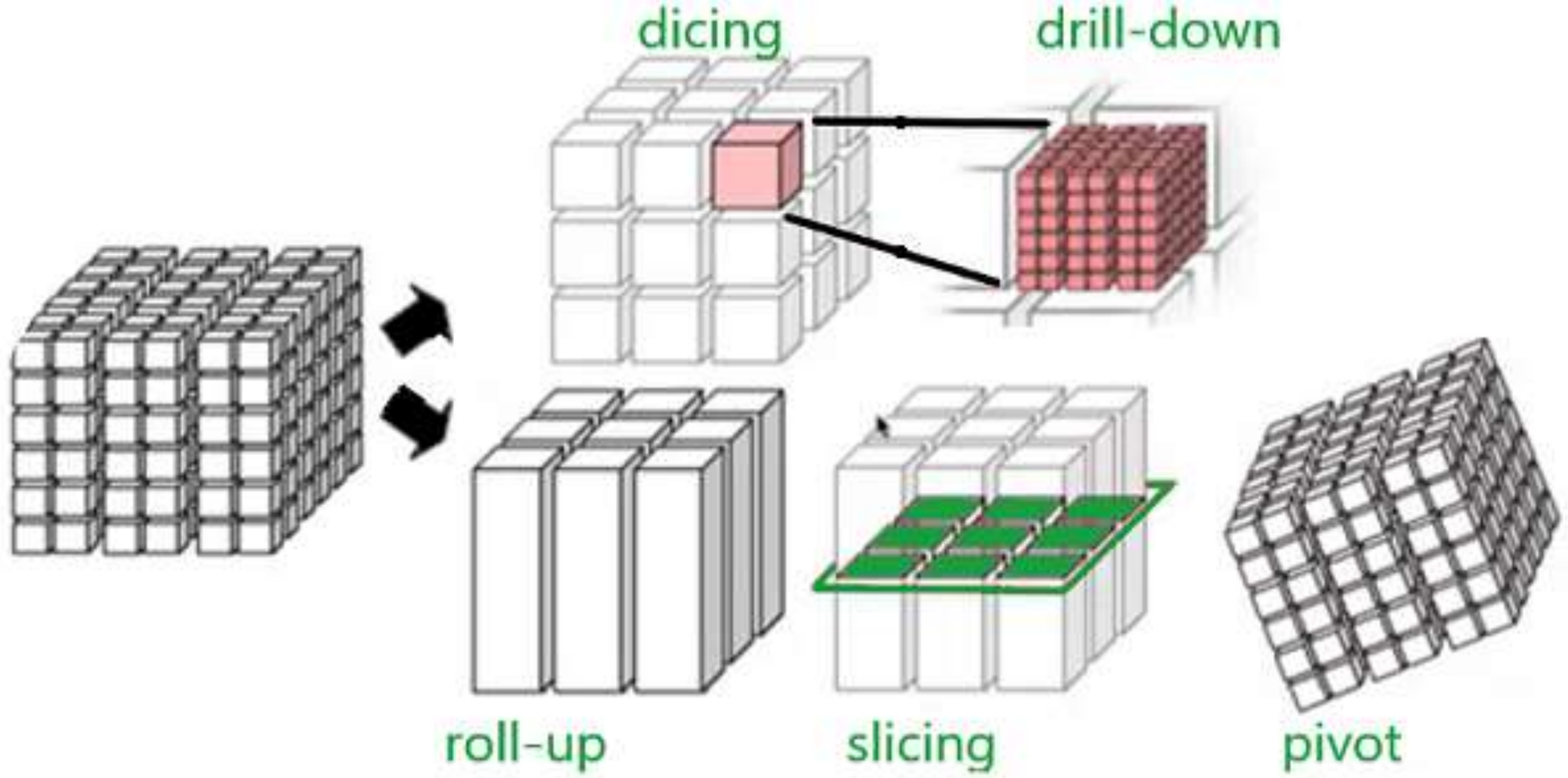
Dicing creates a sub-cube by selecting specific values from multiple dimensions. For example, if an analyst wants to see sales for the “Electronics” category in “California” during “Q1 2023,” this operation would produce a more targeted dataset.

Roll-up:

The roll-up operation aggregates data, reducing detail by summarizing it along a dimension. For instance, summarizing daily sales into monthly totals allows for higher-level insights into sales performance trends.

Drill-down:

The drill-down operation increases data granularity, allowing users to break down aggregated data into more detailed components. For example, drilling down from monthly sales to daily figures enables a deeper understanding of sales fluctuations.


dicing
drill-down
roll-up
slicing
pivot

Visualization of Data Cubes

Data cubes can be visualized in several ways to enhance understanding:

- **3D Cube Visualizations:**

  A graphical representation that illustrates the relationships between multiple dimensions, often helpful in presentations or dashboard views.

- **Pivot Tables:**

  These tables allow users to rearrange data dynamically, facilitating quick comparisons and insights across various dimensions and measures.

Advantages of data cubes:

- **Multi-dimensional analysis**:
  Data cubes enable multi-dimensional analysis of business data, allowing users to view data from different perspectives and levels of detail.

- **Interactivity:**
  Data cubes provide interactive access to large amounts of data, allowing users to easily navigate and manipulate the data to support their analysis.

- **Speed and efficiency**:
  Data cubes are optimized for OLAP analysis, enabling fast and efficient querying and aggregation of data.

- **Data aggregation:**
  Data cubes support complex calculations and data aggregation, enabling users to quickly and easily summarize large amounts of data.

•**Improved decision-making:**

Data cubes provide a clear and comprehensive view of business data, enabling improved decision-making and business intelligence.

•**Accessibility**: Data cubes can be accessed from a variety of devices and platforms, making it easy for users to access and analyze business data from anywhere.

•Helps in giving a summarised view of data.
•Data cubes store large data in a simple way.
•Data cube operation provides quick and better analysis,
•Improve performance of data.

Disadvantages of data cube:

• **Complexity**: OLAP systems can be complex to set up and maintain, requiring specialized technical expertise.

•**Data size limitations**: OLAP systems can struggle with very large data sets and may require extensive data aggregation or summarization.

•**Performance issues**: OLAP systems can be slow when dealing with large amounts of data, especially when running complex queries or calculations.

•**Data integrity:** Inconsistent data definitions and data quality issues can affect the accuracy of OLAP analysis.

•**Cost:** OLAP technology can be expensive, especially for enterprise-level solutions, due to the need for specialized hardware and software.

•**Inflexibility:** OLAP systems may not easily accommodate changing business needs and may require significant effort to modify or extend.