Chapter-4

Online Analytical Processing (OLAP)

Topics

- **Business Intelligence (BI) Technologies OLAP definitions**
- **Data cube & hypercube**
- **OLAP operations**
- **Types of OLAP tools OLAP Demo**

Business Intelligence (BI) Technologies

- **With the growth in data warehousing, users demand for more powerful access tools that provide advanced analytical capabilities**
- **Two main types of these access tools are**
  - **Online Analytical Processing (OLAP)**
  - **Data mining**

Business Intelligence (BI) Technologies (2)

- **OLAP and Data Mining differ in what they offer the user**
- **complementary technologies**
- **Data warehouse (or data marts) together with tools such as OLAP and /or data mining are referred to as Business Intelligence (BI) technologies**

What is OLAP?

- Online Analytical Processing (OLAP) is a system that further transforms the data into a more structured (summarized) form than tables
- OLAP is a form of Executive Information System (EIS) and Decision Support System (DSS)
- OLAP looks at data in multi-dimensional form (data cube)
- OLAP can be used by multiple users to access data in a data warehouse, e.g. via Internet
- OLAP provides managers with a quick and flexible access to large volume of data

OLAP Definitions

- **Codd (1993) – OLAP is "the dynamic synthesis, analysis, and consolidation of large volumes of multi-dimensional data."**
- **OLAP technology uses a multi-dimensional view of aggregate data to provide quick access to strategic information**

Why OLAP?

- **Users need powerful tools for the analysis of large-volume of data,**
  - **i.e. data in data warehouse**
- **Two main types of analysis tools for data warehouse are:**
  - **Online Analytical Processing (OLAP)**
    - **–"top-down" analysis**
  - **Data Mining**
    - **–"bottom-up" analysis**

Why OLAP? (2)

- **OLAP vs. general-purpose query tools**
  - **OLAP has ability to answer 'what if?' and 'why?' questions (not only 'what', 'when', 'where' and 'how much' questions)**
  - **OLAP has more advanced and interactive functionalities**
    - **Browsing**
    - **Calculations**
    - **Complex analyses**

OLAP Applications

- **OLAP applications usually have the following common features:**
  - **Multi-dimensional views of data**
    - **Data can be viewed from various perspectives, e.g.**
      - **product, location, time, etc.**
  - **Support for complex calculations**
    - **e.g. sales forecasting, moving averages, percentage growth, etc.**
  - **Time intelligence**
    - **e.g. comparisons of sales performance between**
      - **different time periods**

Multi-dimensional views of data

- **Example of 2 dimensional views of data**

City	Time	Total Revenue
Glasgow	Q1	29726
Glasgow	Q2	30443
Glasgow	Q3	30582
Glasgow	Q4	31390
London	Q1	43555
London	Q2	48244
London	Q3	56222
London	Q4	45632
Aberdeen	Q1	53210
Aberdeen	Q2	34567
Aberdeen	Q3	45677
Aberdeen	Q4	50056
.....	.....	.....
.....	.....	.....

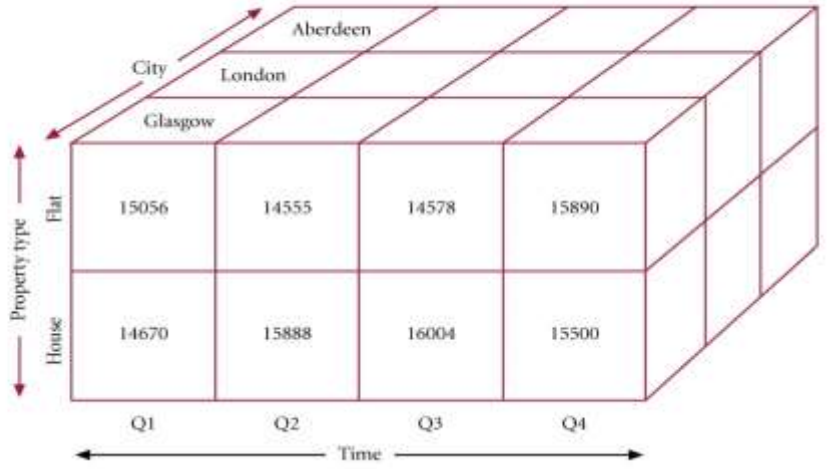
City

Time / Quarter \ City	Glasgow	London	Aberdeen	.....
Q1	29726	43555	53210	.....
Q2	30443	48244	34567	.....
Q3	30582	56222	45677	.....
Q4	31390	45632	50056	.....

Multi-dimensional views of data (2)

- Example of 3 dimensional views of data

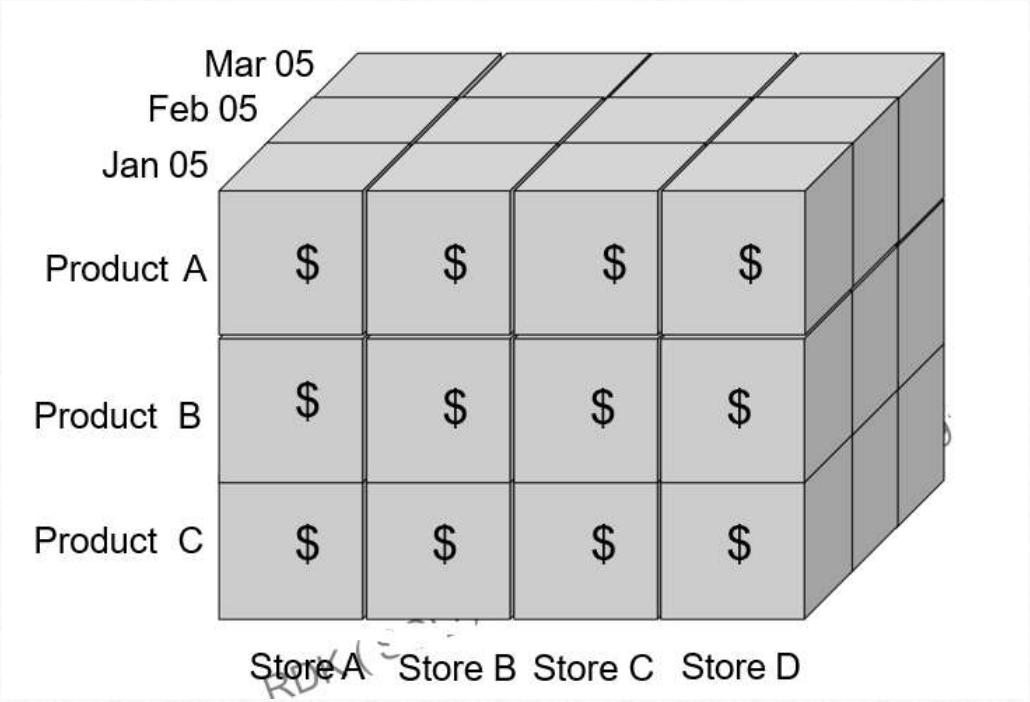
Property Type	City	Time	Total Revenue
Flat	Glasgow	Q1	15056
House	Glasgow	Q1	14670
Flat	Glasgow	Q2	14555
House	Glasgow	Q2	15888
Flat	Glasgow	Q3	14578
House	Glasgow	Q3	16004
Flat	Glasgow	Q4	15890
House	Glasgow	Q4	15500
Flat	London	Q1	19678
House	London	Q1	23877
Flat	London	Q2	19567
House	London	Q2	28677
.....	.....	.....	.....
.....	.....	.....	.....



Data Cube

- Multi-dimensional structures are best visualized as cubes of data
- Cube represents data as cells in an array Each side of a cube is a dimension
- A cube supports matrix arithmetic Hypercube is a form of data cube that has
- more than 3 dimensions
  - Hypercube can be represented as cube that contains cubes for other dimensions (cubes within cubes)
  - As number of dimensions increases, number of the
- cube’s cells increases exponentially

Data Cube Example



OLAP Operations

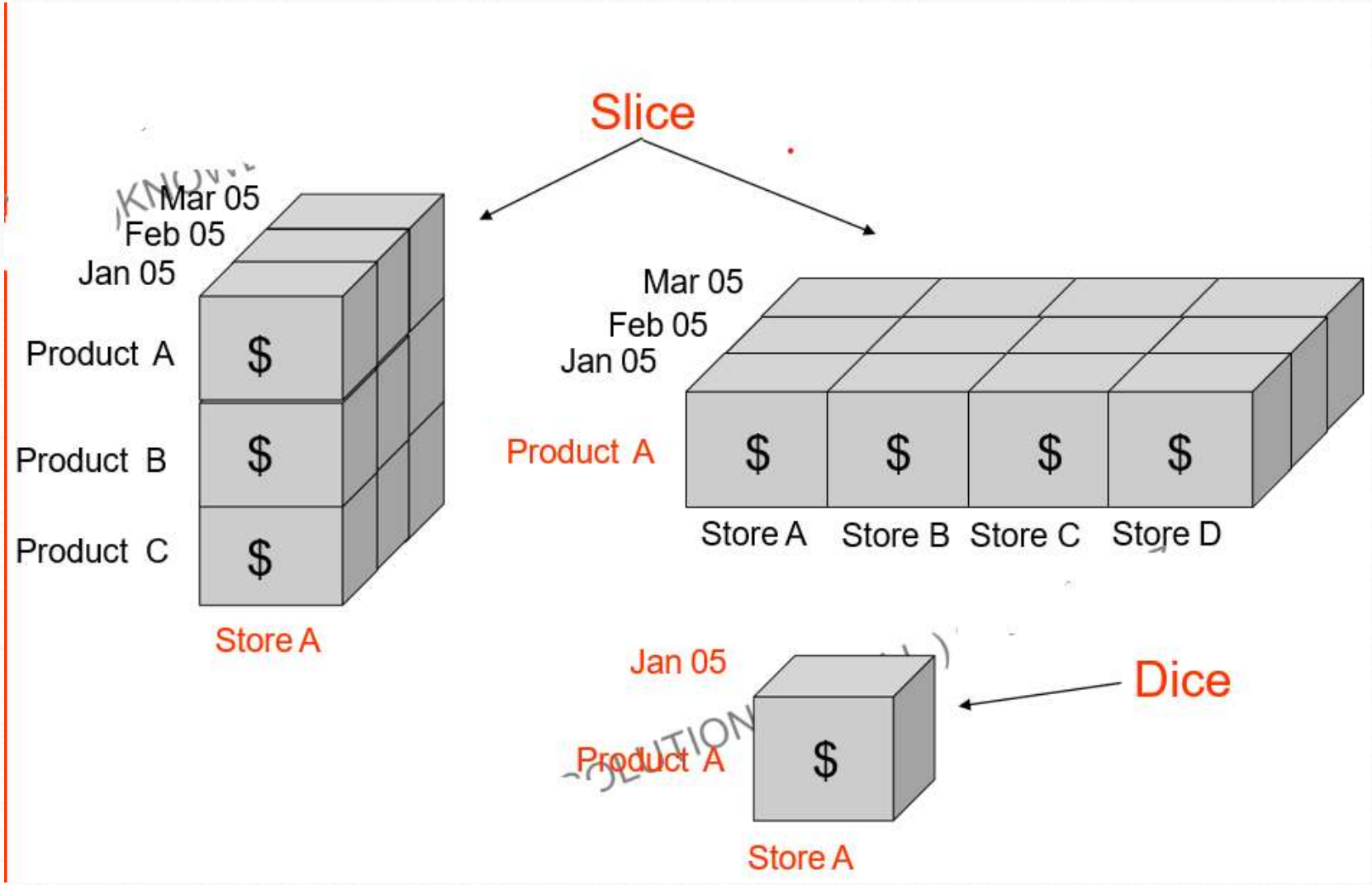
- **Slice**
  - **Select data on a single dimension of a data cube**
- **Dice**
  - **Extracts a sub-cube from the original cube**
- **Roll-up (aggregation)**
  - **Combing of cells for one dimension**
  - **Generalization, e.g. Jan, Feb, Mar = Quarter 1**
  - **May be used with “concept hierarchy”**

Drill-down

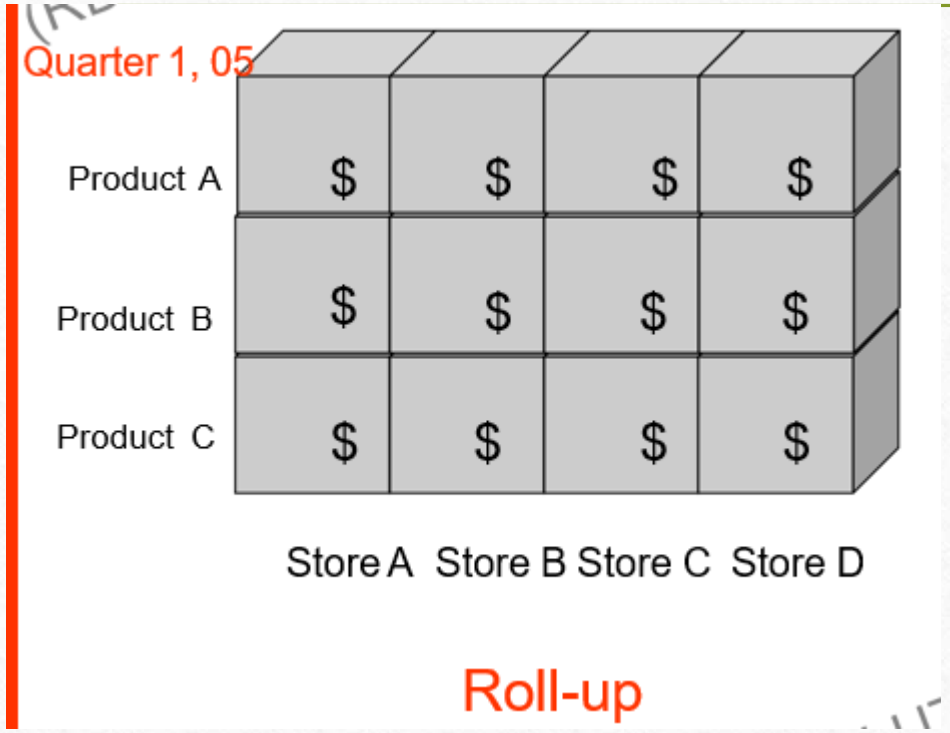
- **Reverse of “Roll-up” operation**
- **Examine data at level of greater detail, e.g. Northern Region = Chiang Mai, Chiang Rai, …**

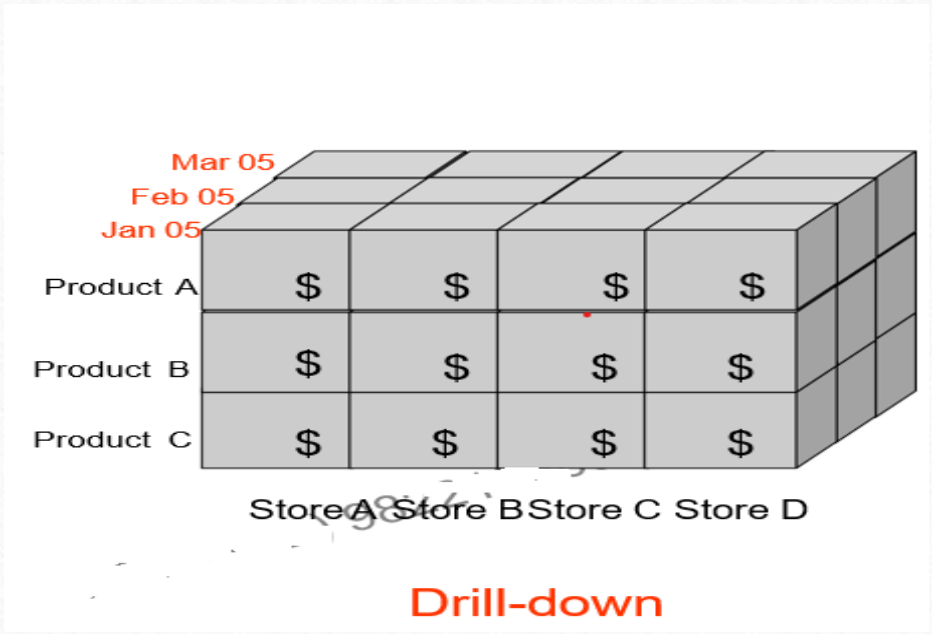
Rotation (pivot)

- **Allow user to view data from a new perspective**
- **Axis rotation**

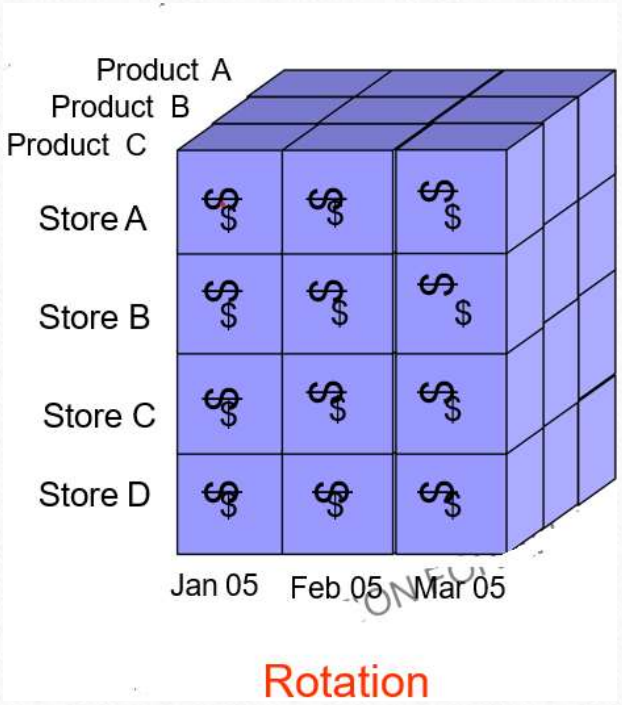

Slice
Mar 05
Feb 05
Jan 05
Product A
Product B
Product C
$
Store A
Mar 05
Feb 05
Jan 05
Product A
$
$
$
$
Store A
Store B
Store C
Store D
Jan 05
Product A
$
Store A
Dice

OLAP Operations (3)



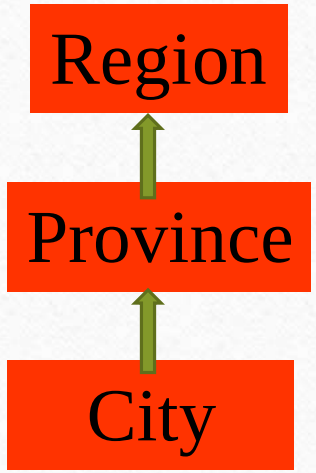


OLAP Operations (4)



Concept Hierarchy

- **Attribute may have concept hierarchies associated with**
- **Examples**



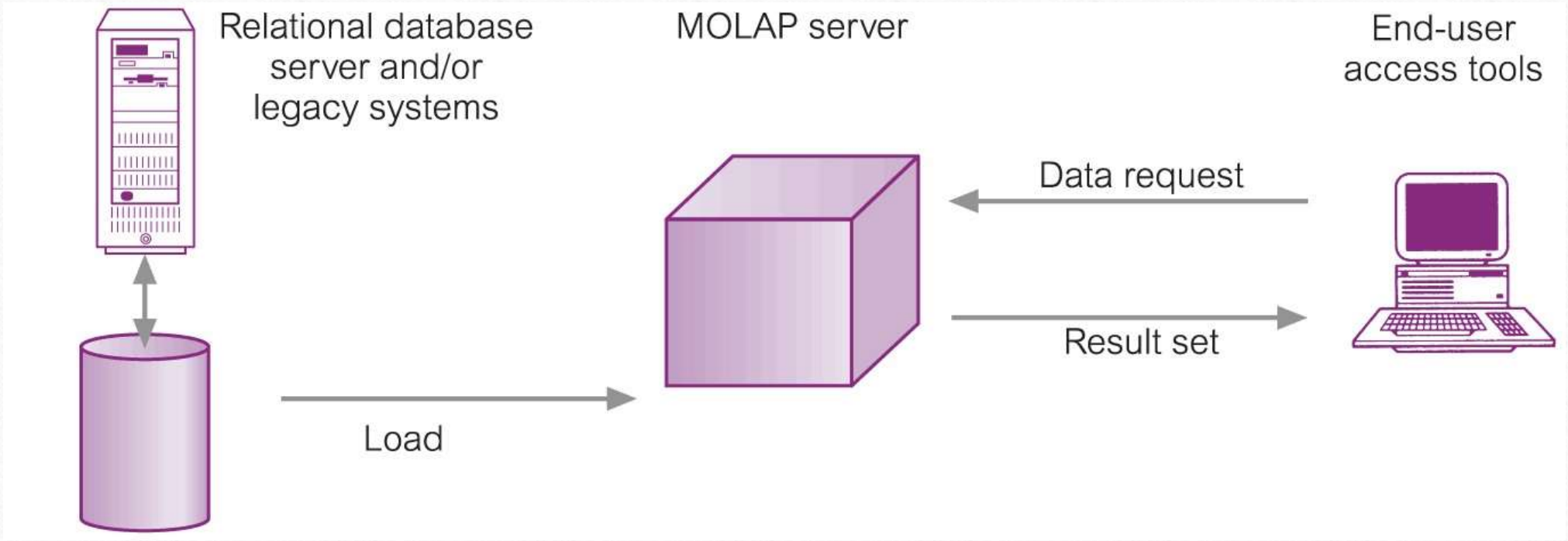
Types of OLAP tools

- **OLAP tools are categorized based on how they store and process multi-dimensional data**
- **4 main types of OLAP tools:**
  - **Multi-dimensional OLAP (MOLAP)**
  - **Relational OLAP (ROLAP)**
  - **Hybrid OLAP (HOLAP)**
  - **Desktop OLAP (DOLAP)**

Multi-dimensional OLAP (MOLAP)

- Use Multi-dimensional Database Management System (MDDBMS) to organize and analyze data
- Use some efficient storage techniques to
- minimize disk space requirement
- Provides good performance when data is used as designed
- Provide a tight coupling between data structure
- and presentation layer
- Access to data structure may be provided via application programming interfaces (APIs)

MOLAP Architecture



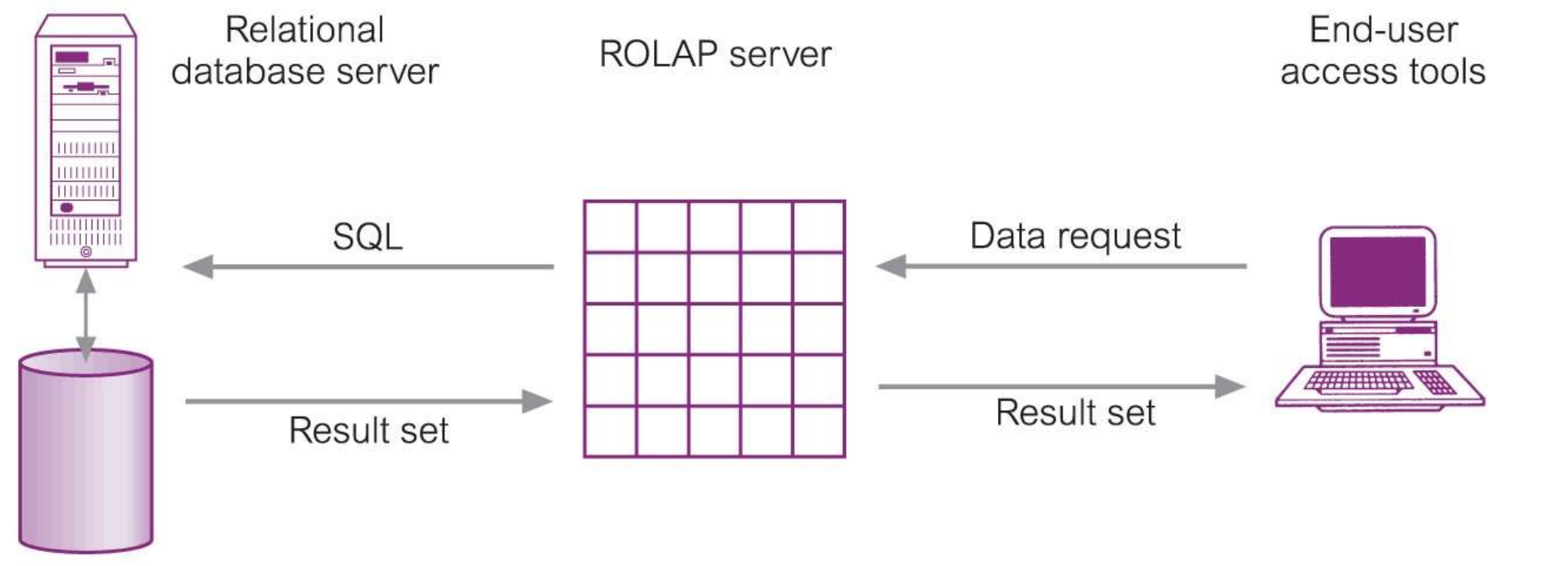
MOLAP Issues

- **MOLAP products require different skills and tools to build and maintain the database, thus increasing the cost and complexity of support**
- **MDDBMS is a new and immature technology**
- **(compared to RDBMS)**

Relational OLAP (ROLAP)

- **Fastest-growing type of OLAP technology MOLAP databases has some limitations**
  - **Not all data can be efficiently stored in MOLAP databases**
- **Uses supports from RDBMS**
  - **avoids need to create multi-dimensional database**
  - **creates multi-dimensional views from relational database**
- **May use SQL to support multi-dimensional data analysis**

ROLAP Architecture



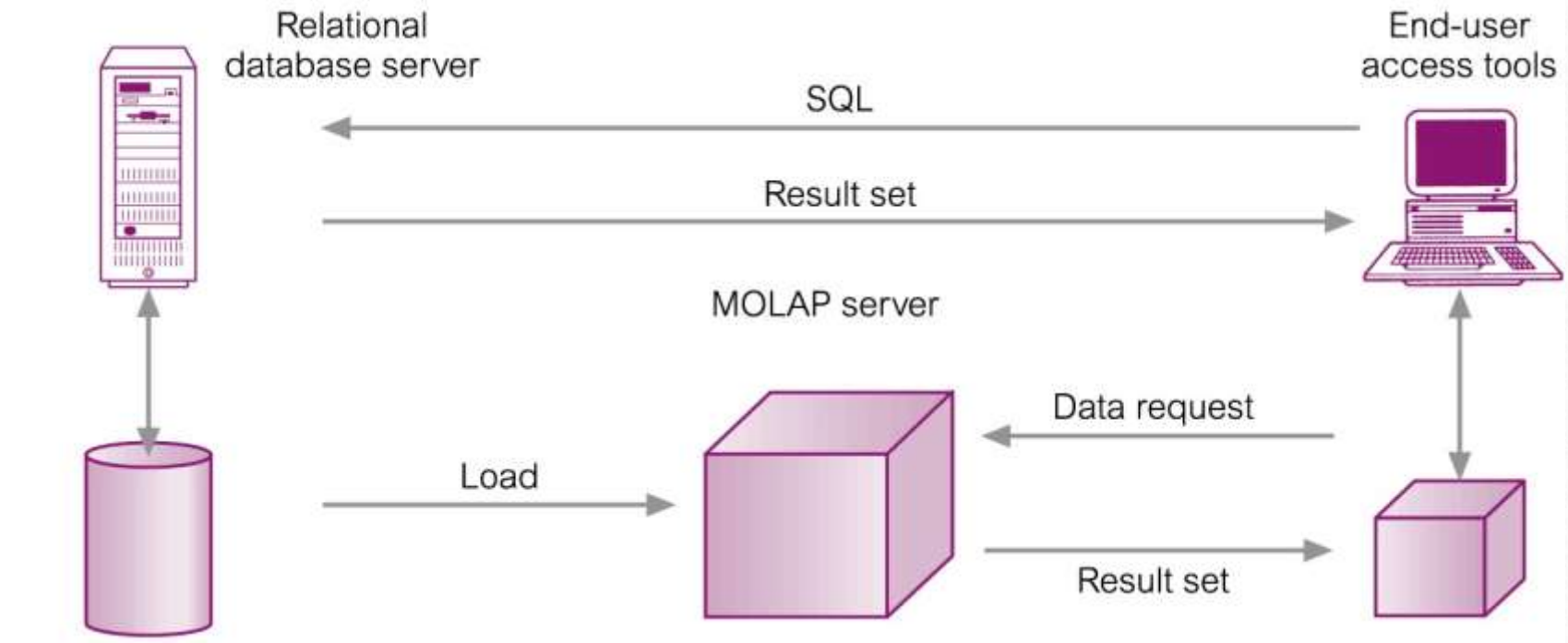
ROLAP Issues

- **Need to create a middleware to work with multi-dimensional applications**
- **The middleware must convert relational data**
- **structure to multi-dimensional data structure**
- **Performance problems for complex queries that require complex transformations from relational data**

Hybrid OLAP (HOLAP)

- **Provide query support for both RDBMS and MDDBMS**
  - **Query data directly from the RDBMS using SQL or via a MOLAP server in the form of a data cube**
- **May cause data redundancy and inefficient network**
  - **usage**

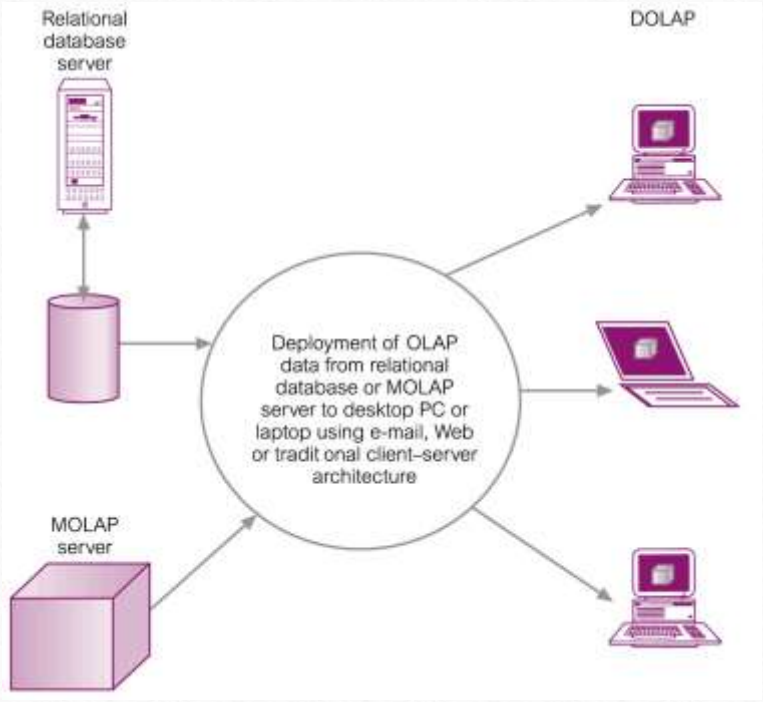
HOLAP Architecture



Desktop OLAP (DOLAP)

- **Store and process the OLAP data on client side**
- **Data are held on client machines**
  - **Database may be distributed in advance, or created on demand (e.g. through the Web)**
  - **The maintenance of database is usually done by a central server**
- **DOLAP uses the power of desktop PC to perform multi-dimensional calculations**

DOLAP Architecture



DOLAP Issues

- **Security (access control) can be difficult**
- **Can not utilize access control feature of DBMS Current trends are towards thin client machines**
  - **Complex calculations are increasingly moved to**
  - **server machine rather than client machine**

OLAP Benchmark

- **APB-1 (OLAP Council, 1998) is a standard for OLAP benchmark**
  - **Measurement of OLAP server performance**
- **APB-1 evaluates OLAP server performance for the following operations:**
  - **Loading of data**
  - **Aggregation of data**
  - **Complex Calculations**
  - **Time series analysis**
  - **Complex Queries**
  - **Drill-down through hierarchies**
  - **Multiple online sessions**
  - **etc.**

OLAP Benchmark (2)

- **A benchmark metric used by APB-1 is AQM (Analytical Queries per Minute)**
- **AQM measures the number of analytical queries that**
- **an OLAP server can process per minute**
- **The time is measured from when the data is loaded until the results are returned to user**

OLAP Extensions to SQL

- **SQL has limited capability to support complex management queries**
- **ANSI adopted a set of OLAP functions as an extension to SQL**
  - **IBM and Oracle jointly proposed these extensions in 1999 as part of the current SQL standard**
- **The extensions are referred to as the 'OLAP package':**
  - **Feature T431, 'Extended Grouping capabilities'**
  - **Feature T611, 'Extended OLAP operators**