

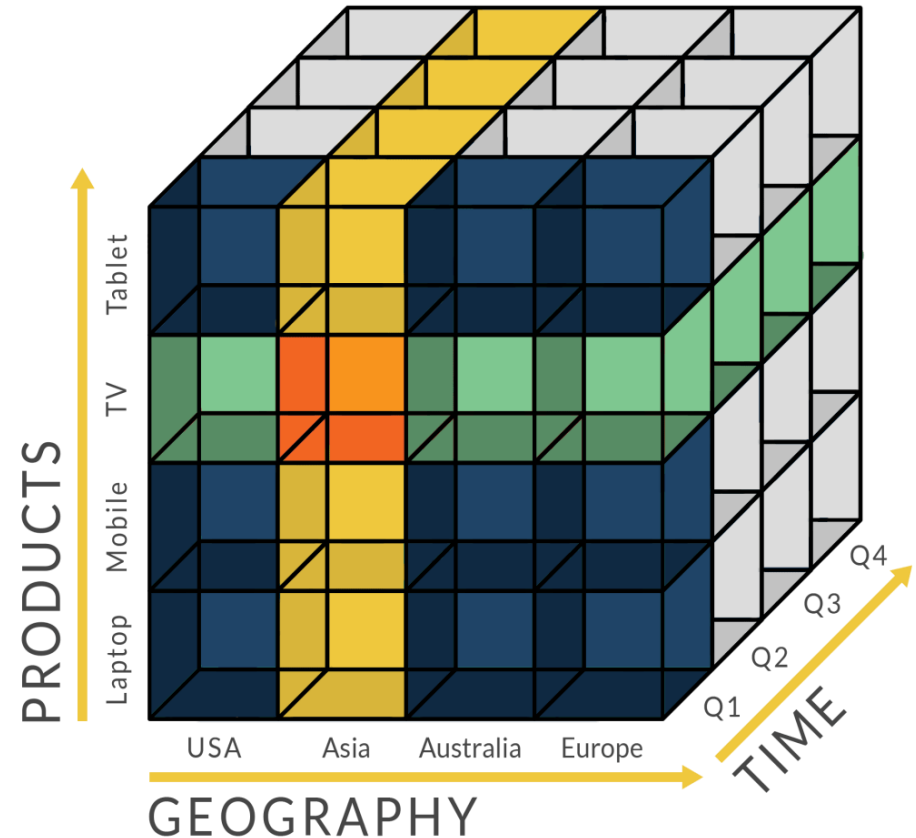
Unit 3

OLAP Technology for Data Mining

OLAP

Online Analytical Processing

- OLAP **performs multidimensional analysis** of business data and provides the capability for complex calculations, trend analysis, and sophisticated data modeling.
- It is the **foundation for many kinds of business applications** for Business Performance Management, Planning, Budgeting, Forecasting, Financial Reporting, Analysis, Simulation Models, Knowledge Discovery, and Data Warehouse Reporting.
- OLAP enables **end-users to perform analysis of data in multiple dimensions**, thereby providing the insight and understanding they need for better decision making.

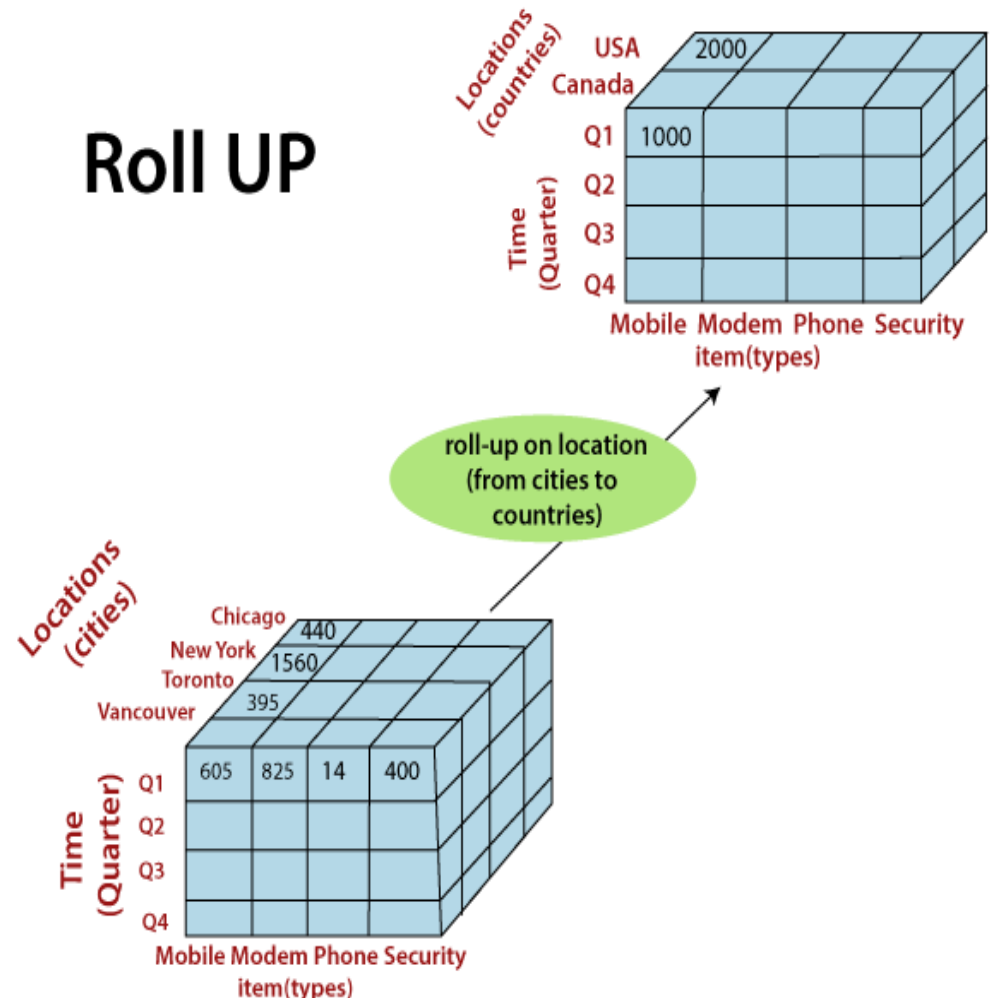


OLAP Operations in the Multidimensional Data Model

- Roll-up (drill-up)
- Drill-down (roll-down)
- Slice
- Dice
- Pivot (rotate)

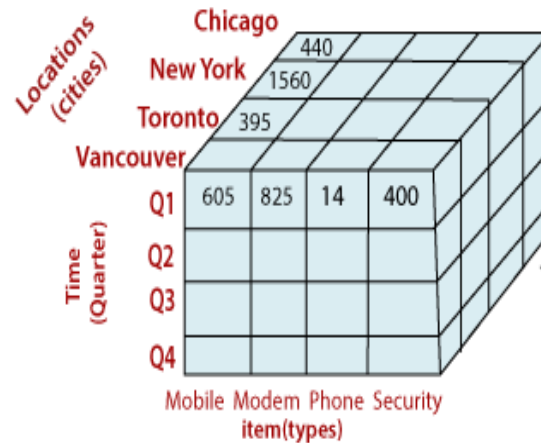
Roll-up (drill-up)

- **Performs aggregation on a data cube**, by climbing down concept hierarchies, i.e., dimension reduction
- **Zooming-out** on the data cubes.
- One or more dimensions are removed from the cube.



Drill-Down

- Reverse operation of **roll-up**
- **Zooming-in** on the data cube
- Navigates from less detailed record to more detailed data
- Performed by either **stepping down** a concept hierarchy for a dimension or adding additional dimensions.
- Drill-down adds more details to the given data



Drill Down

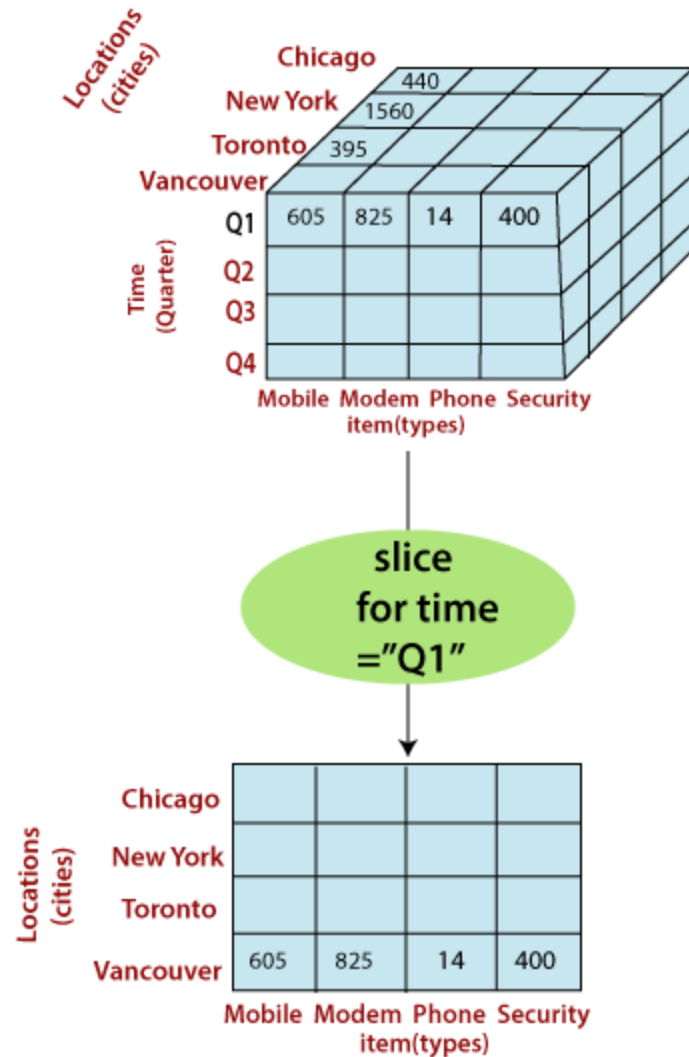
Drilldown on time(from quarters to month)



Slice

Slice

- A **slice** is a subset of the cubes corresponding to a single value for one or more members of the dimension.
- For example, a slice operation is executed when the customer wants a selection on one dimension of a three-dimensional cube resulting in a two-dimensional site.
- The **Slice operations perform a selection on one dimension of the given cube, thus resulting in a sub-cube.**



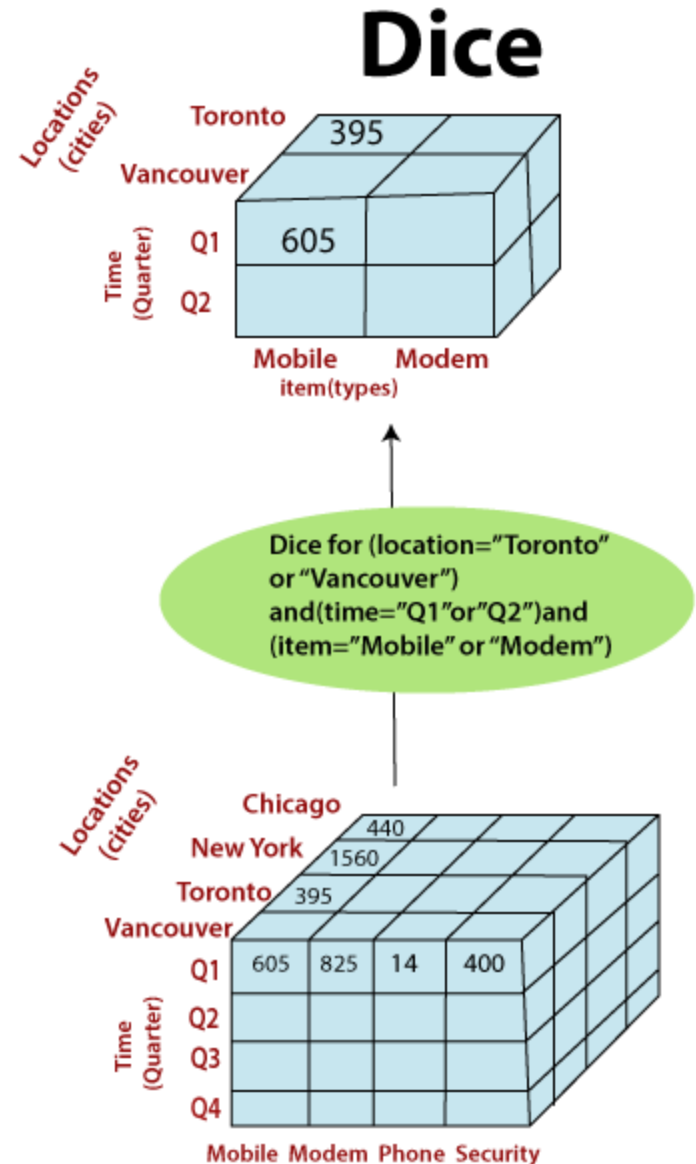
Slice is functioning for the dimensions "time" using the criterion time = "Q1".

Dice

- The dice operation describes a sub-cube by operating a **selection on two or more dimension.**

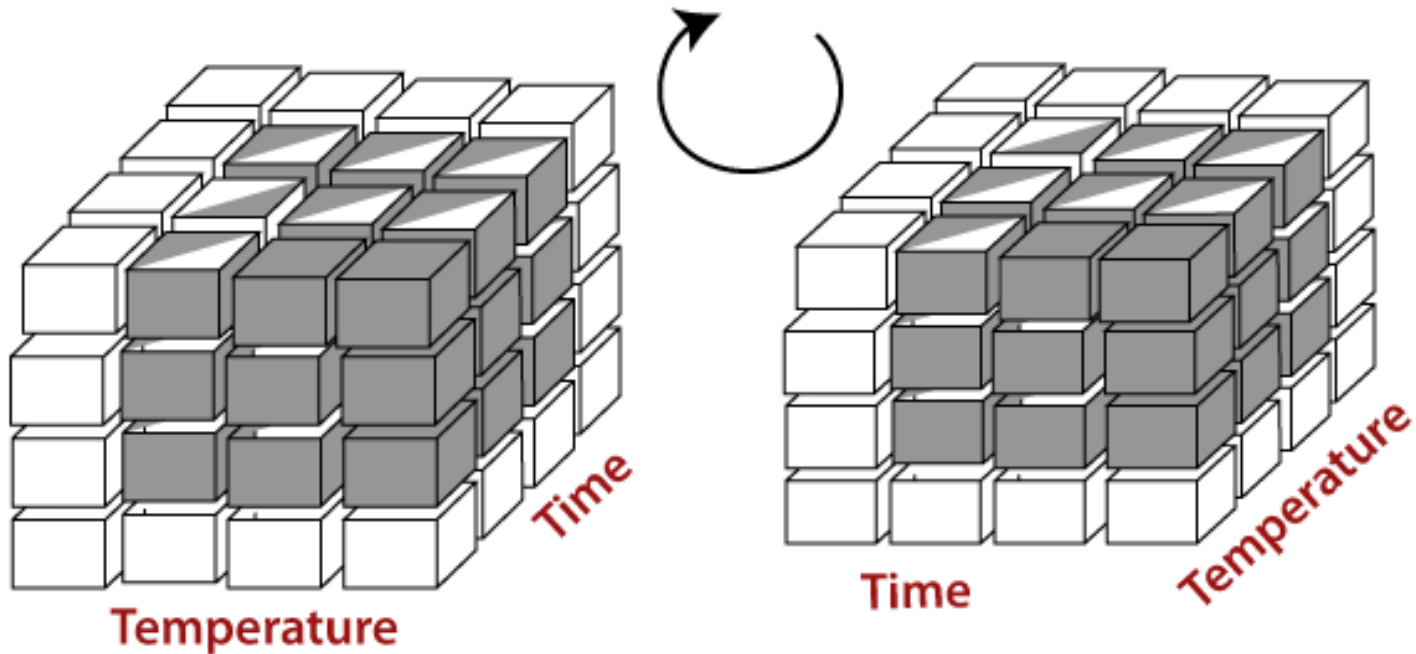
The dice operation on the cubes based on the following selection criteria involves three dimensions.

- (location = "Toronto" or "Vancouver")
- (time = "Q1" or "Q2")
- (item = "Mobile" or "Modem")

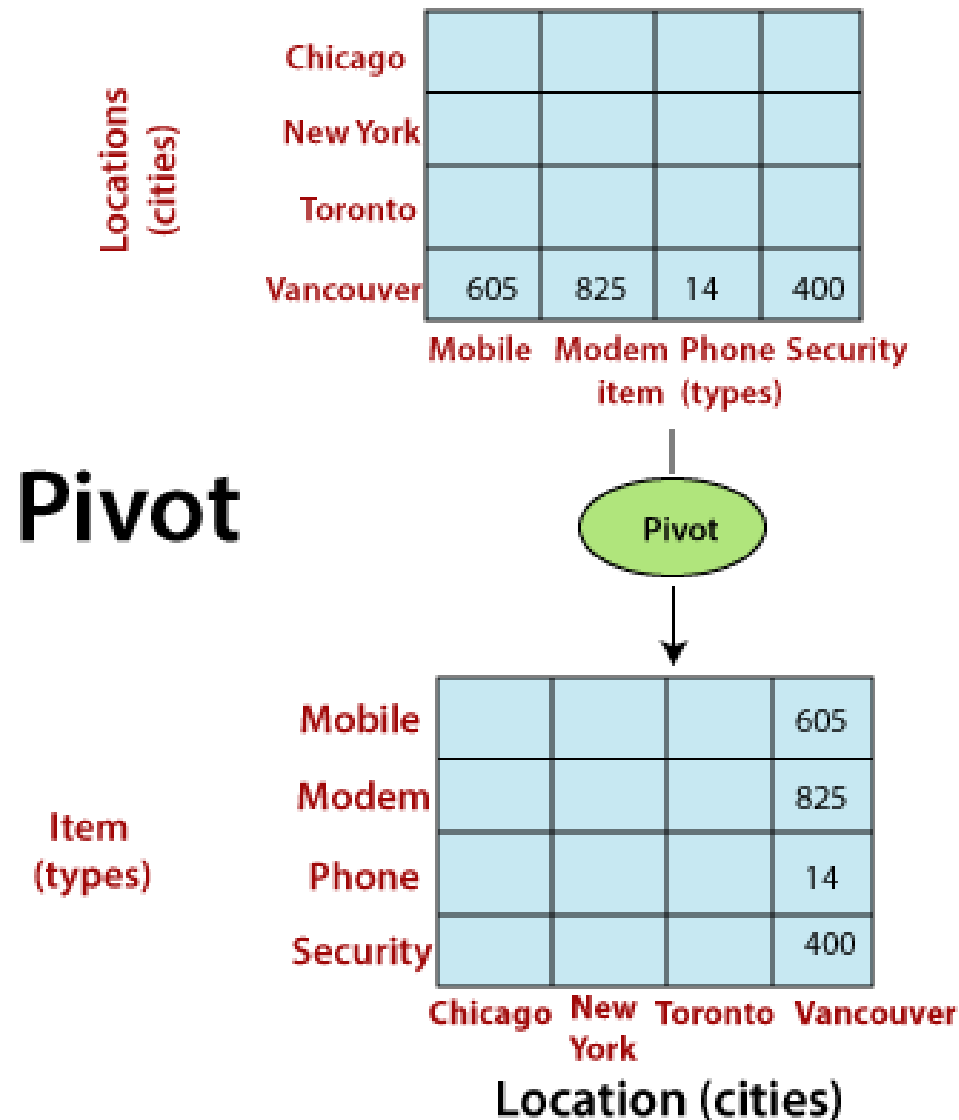


Pivot

- The pivot operation is also called a **rotation**.
- Pivot is a **visualization operations** which rotates the **data axes** in view to provide an alternative presentation of the data.



Pivot may contain **swapping the rows and columns** or moving one of the row-dimensions into the column dimensions.



OLAP Operations → Other Operations

- **Drill-across** executes queries involving (i.e., across) more than one fact table.
- The **drill-through** operation uses relational SQL facilities to drill through the bottom level of a data cube down to its back-end relational tables.
- Other OLAP operations may include **ranking the top N or bottom N items in lists**, as well as computing **moving averages, growth rates, interests, internal return rates, depreciation, currency conversions, and statistical functions**.

OLAP server

- OLAP server is **what does all the work and where the actively accessed data is stored.**
- Operate the processed multidimensional information to **provide users with consistent and fast response.**
- The server can also **populate its data structures in real time** from different databases.

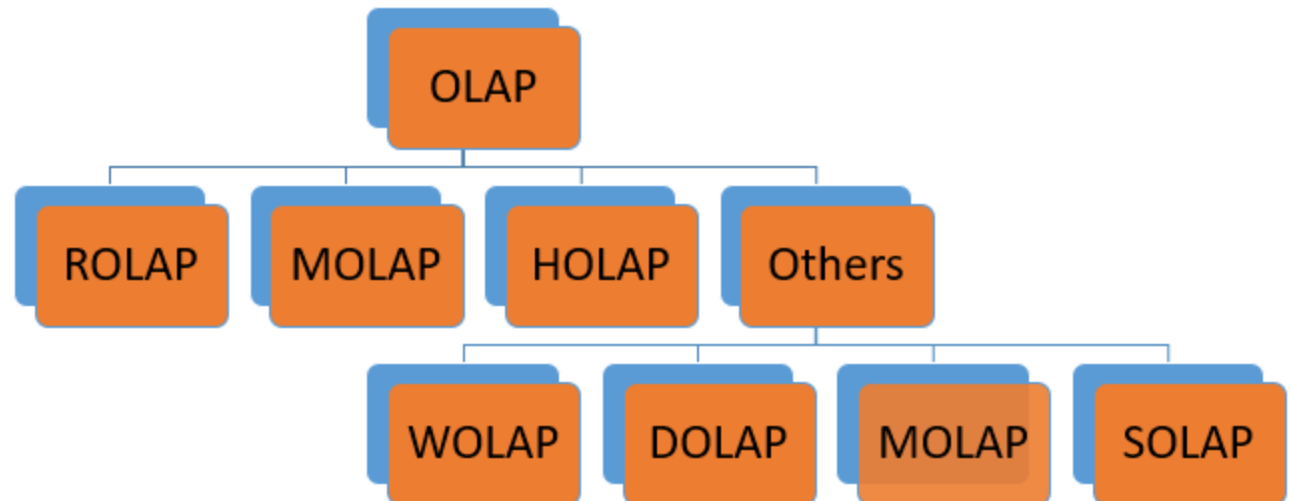
Functions of OLAP server

- Data transformation
- Storage
- Analysis
- Data presentation and access

OLAP Server Architectures

- Logically, OLAP servers present business users with multidimensional data from data warehouses or data marts, without concerns regarding how or where the data are stored.
- However, the physical architecture and implementation of OLAP servers must consider data storage issues.
- Implementations of a warehouse server for OLAP processing include the following
 - ROLAP
 - MOLAP
 - HOLAP

OLAP Hierarchical Structure

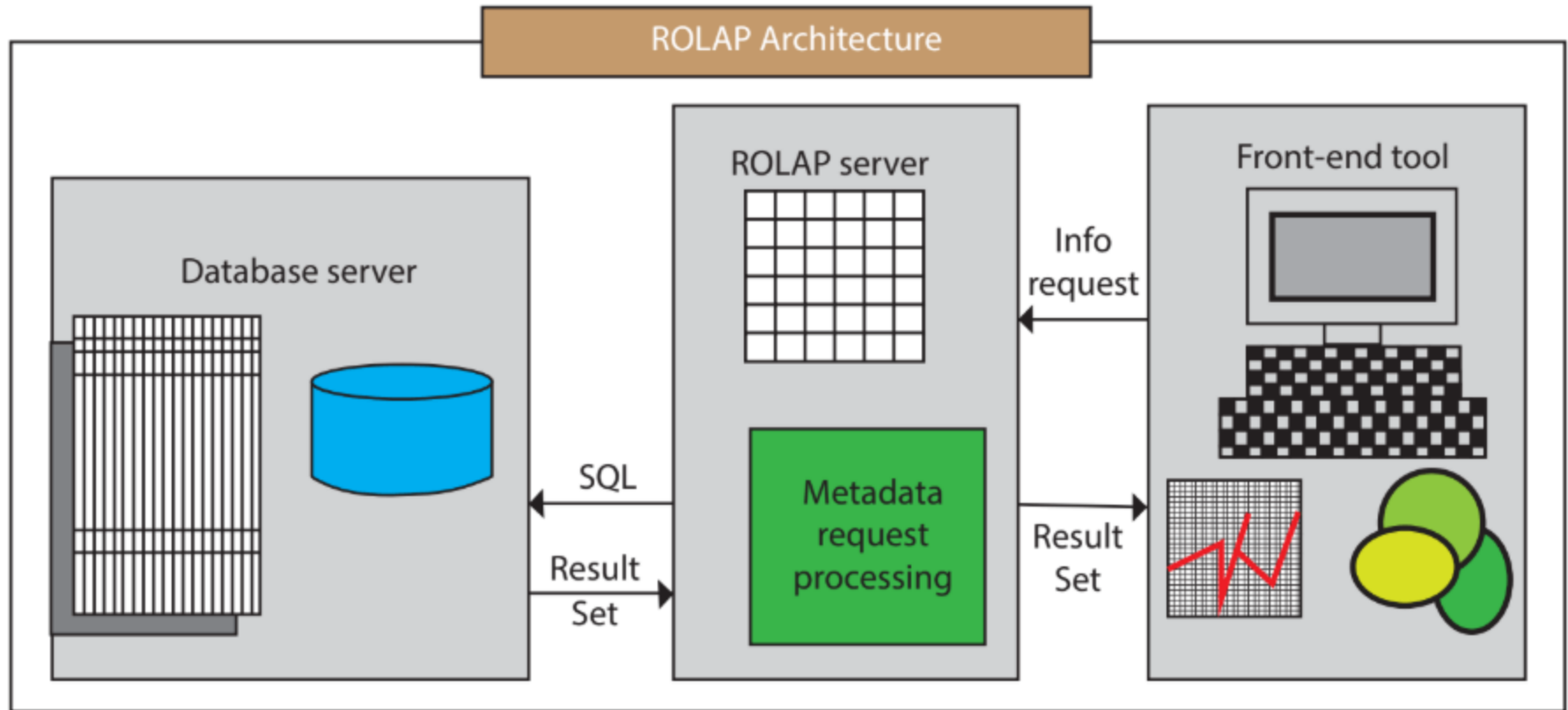


OLAP Server Architectures: ROLAP

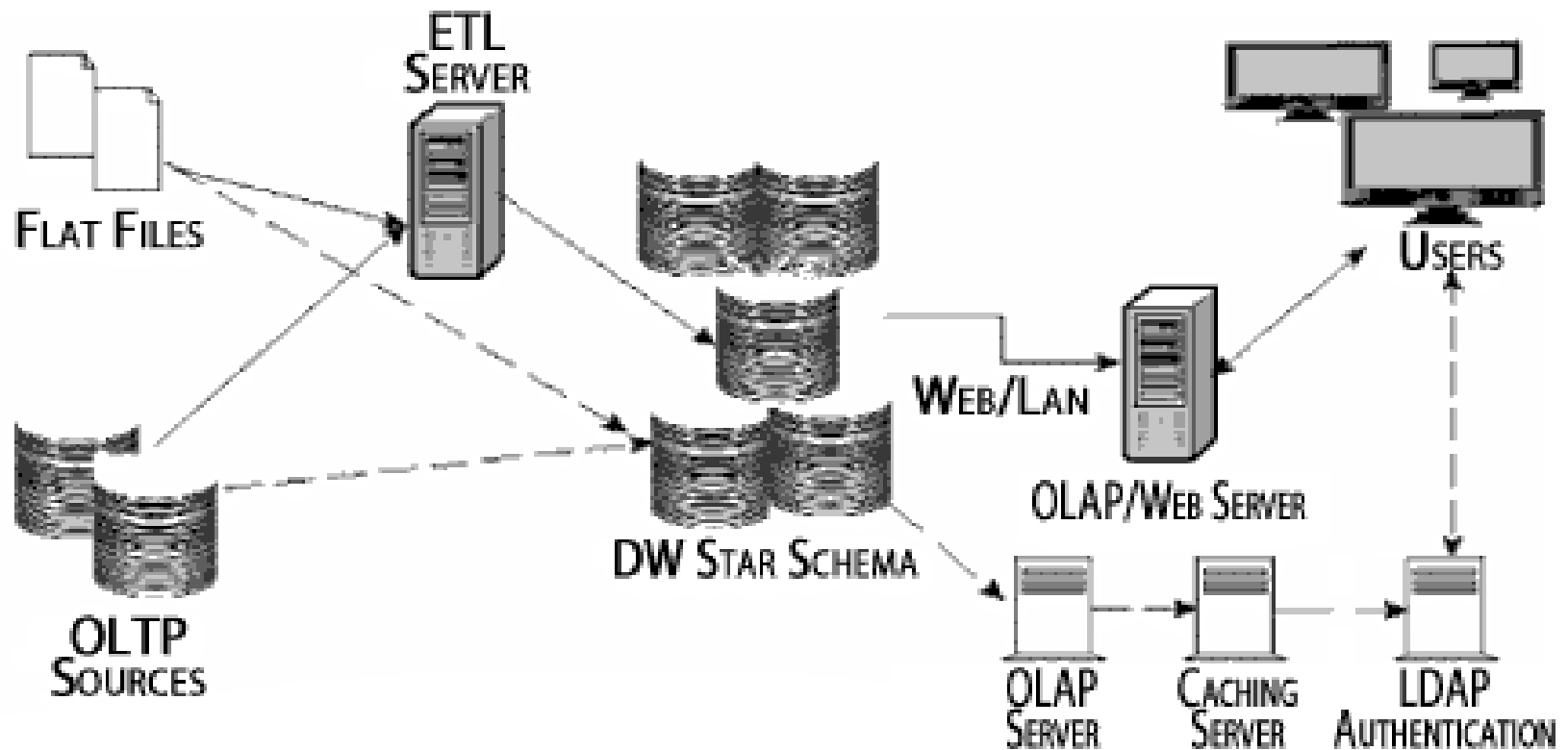
- Relational OLAP
- Fact and dimension tables are stored in relations
- These are intermediate servers which stand in between a relational back-end server and user frontend tools.
- They use a relational or extended-relational DBMS to save and handle warehouse data, and OLAP middleware to provide missing pieces.
- ROLAP technology tends to have higher scalability than MOLAP technology.
- ROLAP systems work primarily from the data that resides in a relational database, where the base data and dimension tables are stored as relational tables. This model permits the multidimensional analysis of data.

ROLAP Architecture
includes the following
components

- Database server.
- ROLAP server.
- Front-end tool.



- Relational On-Line Analytical Processing (ROLAP) performs dynamic multidimensional analysis of data stored in a relational database, rather than in a multidimensional database.
- The traditional OLAP's slice and dice functionality is equivalent to adding a WHERE clause in the SQL statement.
- The design may be structured in the form of a star or its variations.
- A typical use of ROLAP is for large data size that is infrequently queried, such as historical data.



OLAP Server Architectures: ROLAP

Advantages

- **Can handle large amounts of information:** The data size limitation of ROLAP technology is depends on the data size of the underlying RDBMS. So, ROLAP itself **does not restrict the data amount**.
- RDBMS already comes with a lot of features. So ROLAP technologies, (works on top of the RDBMS) can control these functionalities.

Disadvantages

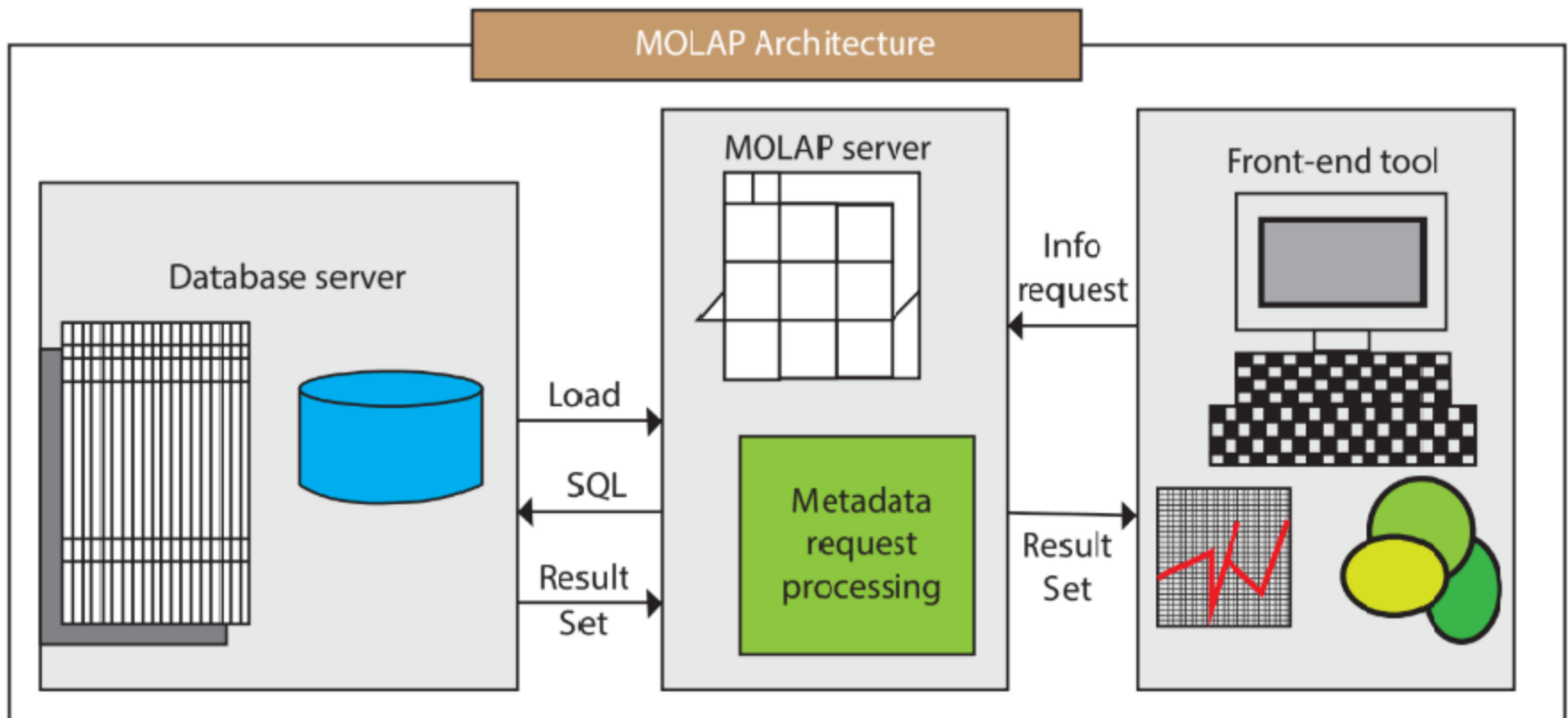
- **Performance can be slow:** Each ROLAP report is a SQL query (or multiple SQL queries) in the relational database, the **query time can be prolonged if the underlying data size is large**.
- **Limited by SQL functionalities:** ROLAP technology relies on upon developing SQL statements to query the relational database, and **SQL statements do not suit all needs**.

OLAP Server Architectures: MOLAP

- Multidimensional OLAP
- A MOLAP system is based on a native logical model that directly supports multidimensional data and operations.
- Data are stored physically into multidimensional arrays, and positional techniques are used to access them.
- One of the significant distinctions of MOLAP against a ROLAP is that data are summarized and are stored in an optimized format in a multidimensional cube, instead of in a relational database.

MOLAP Architecture includes the following components

- Database server.
 - MOLAP server.
 - Front-end tool.
- **MOLAP structure primarily reads the precompiled data.**
 - MOLAP structure has **limited capabilities to dynamically create aggregations** or to evaluate results which have not been pre-calculated and stored.
 - Applications requiring iterative and comprehensive time-series analysis of trends are well suited for MOLAP technology (e.g., financial analysis and budgeting).



OLAP Server Architectures: MOLAP(pros and cons)

Advantages

- **Excellent Performance:** A MOLAP cube is built for fast information retrieval, and is optimal for slicing and dicing operations.
- **Can perform complex calculations:** All evaluation have been pre-generated when the cube is created. Hence, complex calculations are not only possible, but they return quickly.

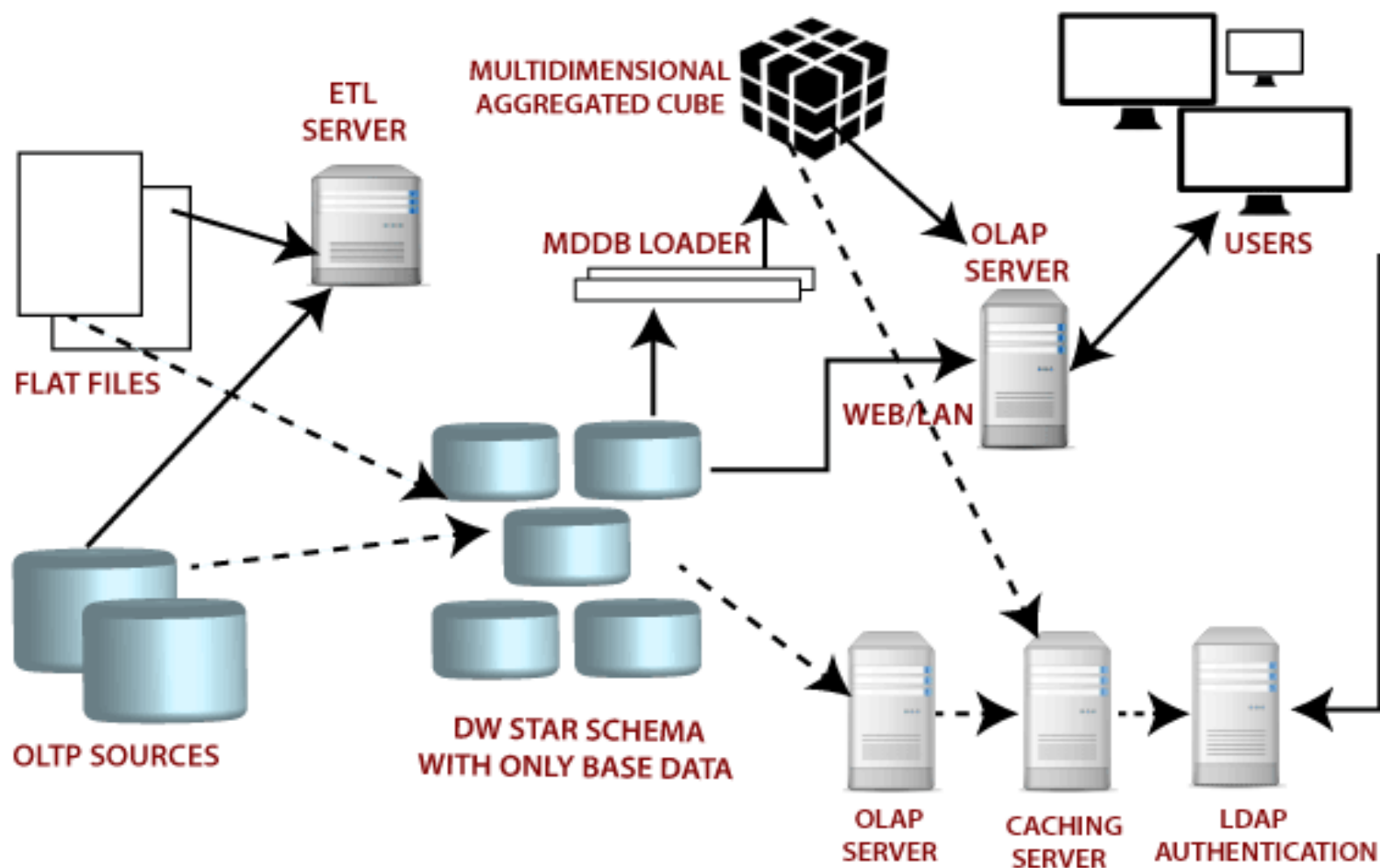
Disadvantages

- **Limited in the amount of information it can handle:** Because all calculations are performed when the cube is built, it is not possible to contain a large amount of data in the cube itself.
- **Requires additional investment:** Cube technology is generally proprietary and does not already exist in the organization. Therefore, to adopt MOLAP technology, chances are other investments in human and capital resources are needed.

OLAP Server Architectures: HOLAP

- Hybrid OLAP
- HOLAP incorporates the best features of MOLAP and ROLAP into a single architecture.
- HOLAP systems save more substantial quantities of detailed data in the relational tables while the aggregations are stored in the pre-calculated cubes.
- HOLAP also can drill through from the cube down to the relational tables for delineated data.
- The Microsoft SQL Server 2000 provides a hybrid OLAP server

HOLAP Architecture



OLAP Server Architectures: HOLAP(pros and cons)

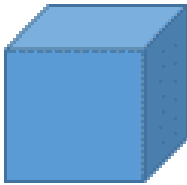
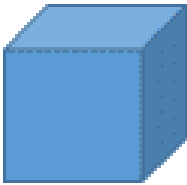
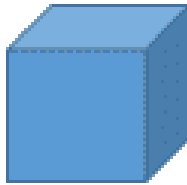
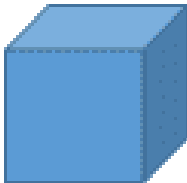
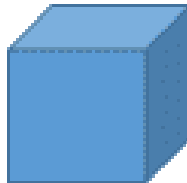
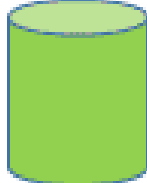
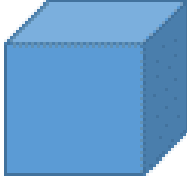
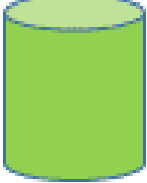
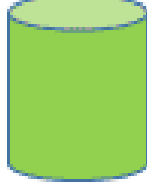
Advantages of HOLAP



- HOLAP provide benefits of both MOLAP and ROLAP.
- It provides fast access at all levels of aggregation.
- HOLAP balances the disk space requirement, as it only stores the aggregate information on the OLAP server and the detail record remains in the relational database.
- No duplicate copy of the detail record is maintained.

Disadvantages of HOLAP

- HOLAP architecture is very complicated because it supports both MOLAP and ROLAP servers.

OLAP Server Architectures: ROLAP Vs MOLAP Vs HOLAP

	MOLAP	HOLAP	ROLAP
Cube Structure			
Preprocessed Aggregates			
Detail-Level Values			

 Multidimensional Storage  Relational Storage

OLAP Server Architectures: Other Types

Web-Enabled OLAP (WOLAP) Server

WOLAP pertains to OLAP application which is **accessible via the web browser**. Unlike traditional client/server OLAP applications, WOLAP is considered to have a **three tiered architecture which consists of three components: a client, a middleware, and a database server**.

Desktop OLAP (DOLAP) Server

DOLAP permits a user to download a section of the data from the database or source, and **work with that dataset locally, or on their desktop**.

Mobile OLAP (MOLAP) Server

Mobile OLAP enables users to access and work on OLAP data and applications remotely through the use of their mobile devices.

Spatial OLAP (SOLAP) Server

SOLAP includes the capabilities of both Geographic Information Systems (GIS) and OLAP into a single user interface. It facilitates the management of both spatial and non-spatial data.

OLTP(Online Transaction Processing)

- OLTP is an operational system that supports **transaction-oriented applications** in a 3-tier architecture(presentation tier, a business logic tier, and a data store tier).
- It administers the **day to day transaction** of an organization.
- OLTP is basically **focused on query processing, maintaining data integrity in multi-access environments as well as effectiveness** that is measured by the total number of transactions per second.

OLTP → Characteristics

- OLTP uses transactions that include small amounts of data.
- Indexed data in the database can be accessed easily.
- OLTP has a large number of users.
- It has fast response times.
- It strictly performs only the predefined operations on a small number of records.
- OLTP stores the records of the last few days or a week.
- It supports complex data models and tables