UNIT 3

OLAP in Data Warehouse

Mr.T.Somasundaram

Assistant Professor

Department Of Management

Kristu Jayanti College (Autonomous), Bengaluru

UNIT 3

OLAP in Data Warehouse

Concept of OLAP, Multidimensional analysis, OLAP functions and applications, OLAP models – MOLAP, ROLAP and DOLAP

Online Analytical Processing (OLAP)

Meaning:

- The term "OLAP" was coined by E.F. Cold in 1993 to refer to a type of application that allows a user to interactively analyse data.
- OLAP systems enable managers and analysts to rapidly and easily examine key performance data and perform powerful comparison and trend analyses, even on large data volumes.
- OLAP is integral part of a data warehousing solution, which is in different shades, depending on database structure and location of majority of analytical processing.
- It allows managers and analysts to get an insight of the information through fast, consistent and interactive access to information.

Online Analytical Processing (OLAP)

Meaning:

- OLAP is a category of software that allows users to analyze information from multiple database systems at the same time.
- It is a technology that enables analysts to extract and view business data from different points of view.

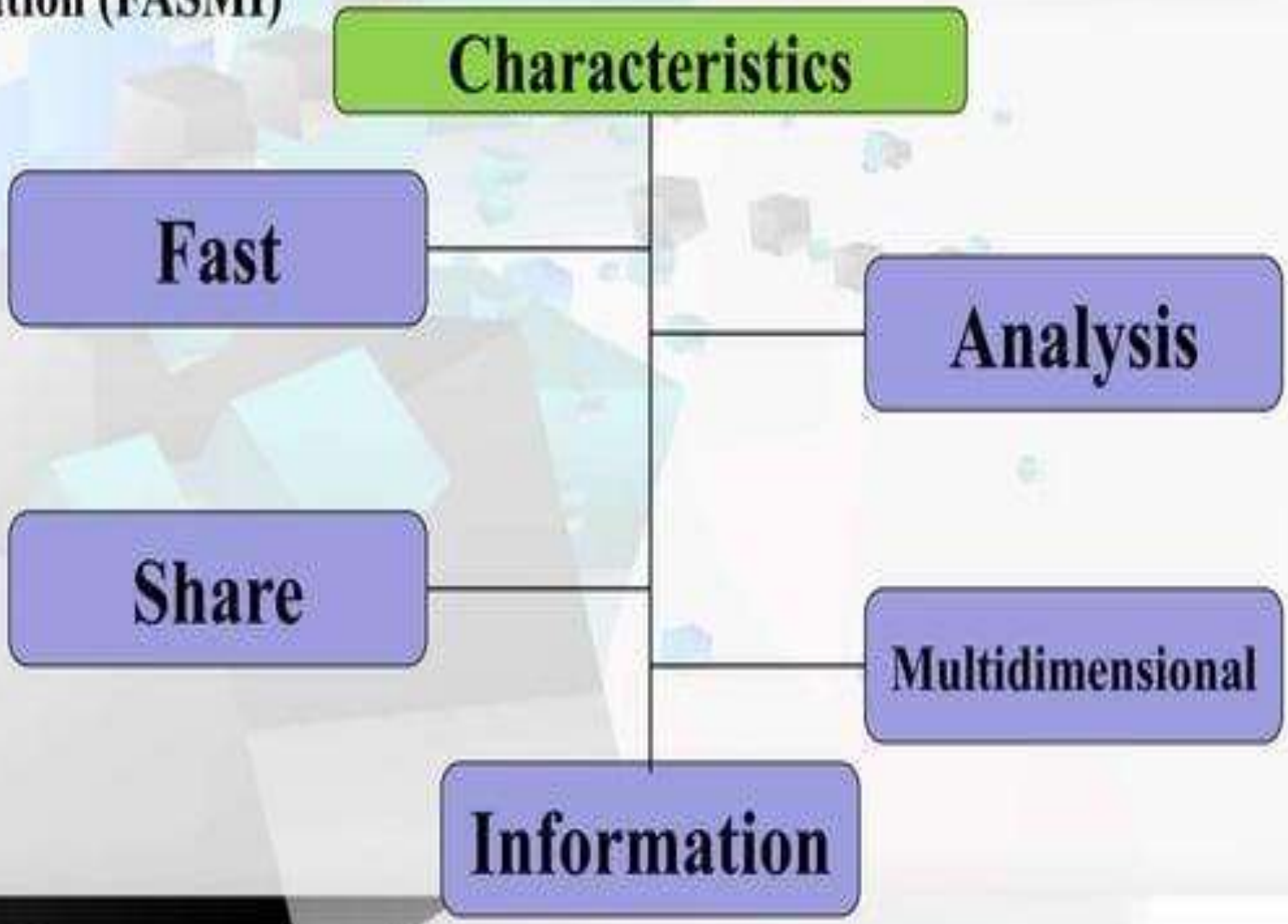
Definition:

OLAP is a method of analysing the data in a multi-dimensional format, often across multiple time periods, with the aim of uncovering the business information concealed within the data.

OLAP is the dynamic synthesis, analysis and consolidation of large volumes of multidimensional data.

Characteristics of OLAP

OLAP defined in five words – **Fast Analysis of Shred Multi-Dimensional Information (FASMI)**



Characteristics of OLAP

1. Fact – it means that the system targeted to deliver most responses to user within about 5 second and very few taking more than 20 seconds.

2. Analysis – it means that the system can cope with any business logic and statistical analysis that it relevant for the application and the user.

3. Share – it means that the system implements all the security requirements for confidentiality, if multiple write access is needed, concurrent update location at an appropriate level not all applications need users to write data back.

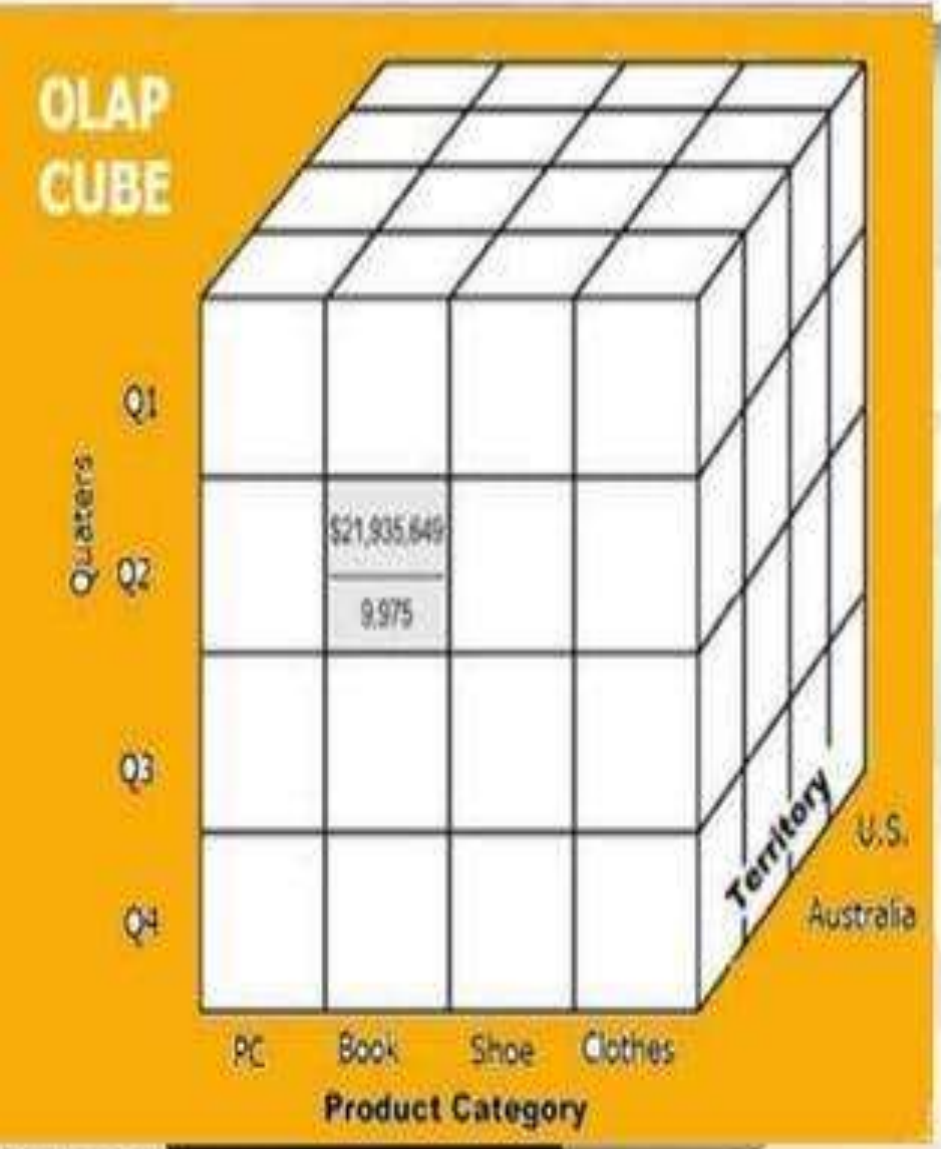
4. Multidimensional – it is the key requirement and OLAP system provide a multidimensional conceptual view of the data.

5.Information – information are all of the data and derived information needed, relevant for application.

OLAP Concept

- At the core of the OLAP concept, is an OLAP Cube.
- OLAP Cube is a data structure optimized for very quick data analysis.
- It consists of numeric facts called measures which are categorized by dimensions.
- OLAP cube is also called the hypercube.
- Data operations and analysis are performed using simple spreadsheet, where data values are arranges in row and column.
- OLAP contains multidimensional data, data usually obtained from a different and unrelated source.

OLAP Concept



- OLAP Cube consists of numeric facts called measures which are categorized by dimensions.
- OLAP Cube is also called Hypercube.
- Data operations & analysis are performed using simple spreadsheet.
- OLAP contains multidimensional data & cube can store and analyze multidimensional data in a logical and orderly manner.

Multidimensional Analysis of OLAP

- ✓ In order to understand OLAP architectures that can deliver this value, it is first necessary to understand the nature of multidimensional data.
- ✓ Multidimensional server is to provide fast, linear access to the data regardless of the way the data is being requested.
- ✓ The objective is to retrieve the data equally fast, regardless of the requested dimensions.
- ✓ Most OLAP servers in fact, achieve fast response to computed results by computing them in advance.
- ✓ One dimension is selected as the fact dimension and this dimension forms the columns of the fact table.

Multidimensional Analysis of OLAP

- ✓ The data is retrieved from the relational database into the client tool by SQL queries.
- ✓ One of the primary problems was maintaining database integrity and ensuring consistent data updates.
- ✓ Database is consistently maintained and that transaction updates can be performed in a fast and efficient manner.
- ✓ In order to provide the multidimensional views of the data, it requires relational database that the data be organized in the star or snowflake schema described previously.
- ✓ There are two major OLAP architectures, MOLAP and RAP provide their own physical multidimensional databases.

OLAP Operations (Functions)

- Data Warehouse uses Online Analytical Processing (OLAP) to formulate and execute user queries.
- OLAP provides aggregate data (measurements) along a set of dimensions, in which –

a) Each dimension is described by a set of attributes (i.e.) each dimension table includes a set of attributes.

b) Each measure depend on a set of dimensions that provides context for the measure (E.g.) for the reseller company data warehouse, the measure is number of sold units, which are described by corresponding location, time and item type.

c) All dimensions are assumed to uniquely determine the measure. (E.g.) for reseller company the location, time, producer and item type provide all necessary information to determine context of a particular no. of sold units.

OLAP Operations

There are five basic OLAP Operations:

1. ROLL UP:

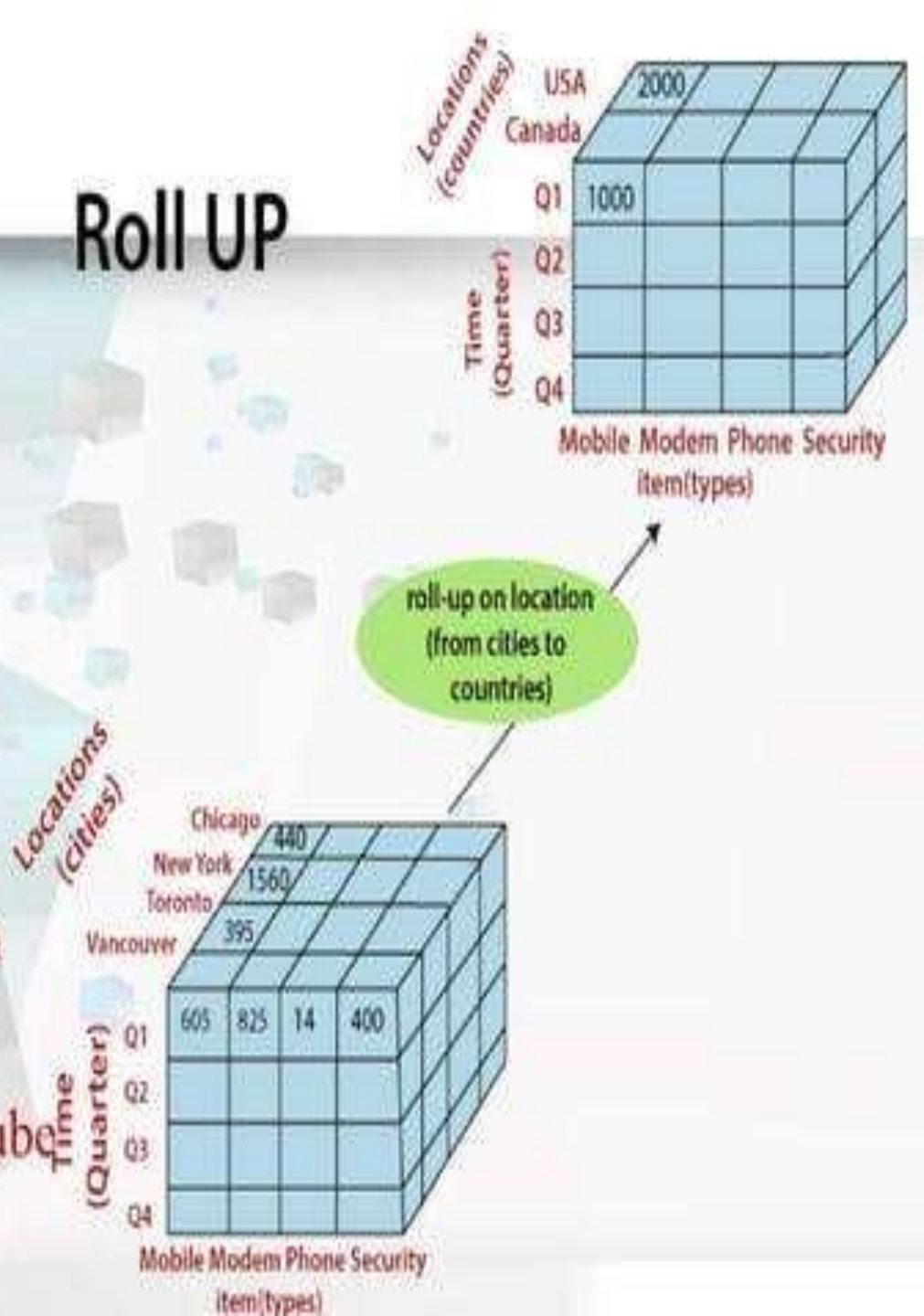
This operation performs aggregation on a data cube in any of the following way:

i) By Climbing up a concept hierarchy for a dimension.

ii) By dimension reduction.

(E.g.) In this figure,

- Roll up operation is performed by climbing up a concept hierarchy for the dimension location.
- Initially the concept hierarchy was "street < city < province < country"
- Data is aggregated by ascending the location hierarchy from level of city to level of country.
- Data is grouped into cities rather than countries.
- One or more dimensions from data cube are removed.



2. DRILL DOWN:

Drill-down operation is reverse of the roll-up. This operation is performed by either of the following way:

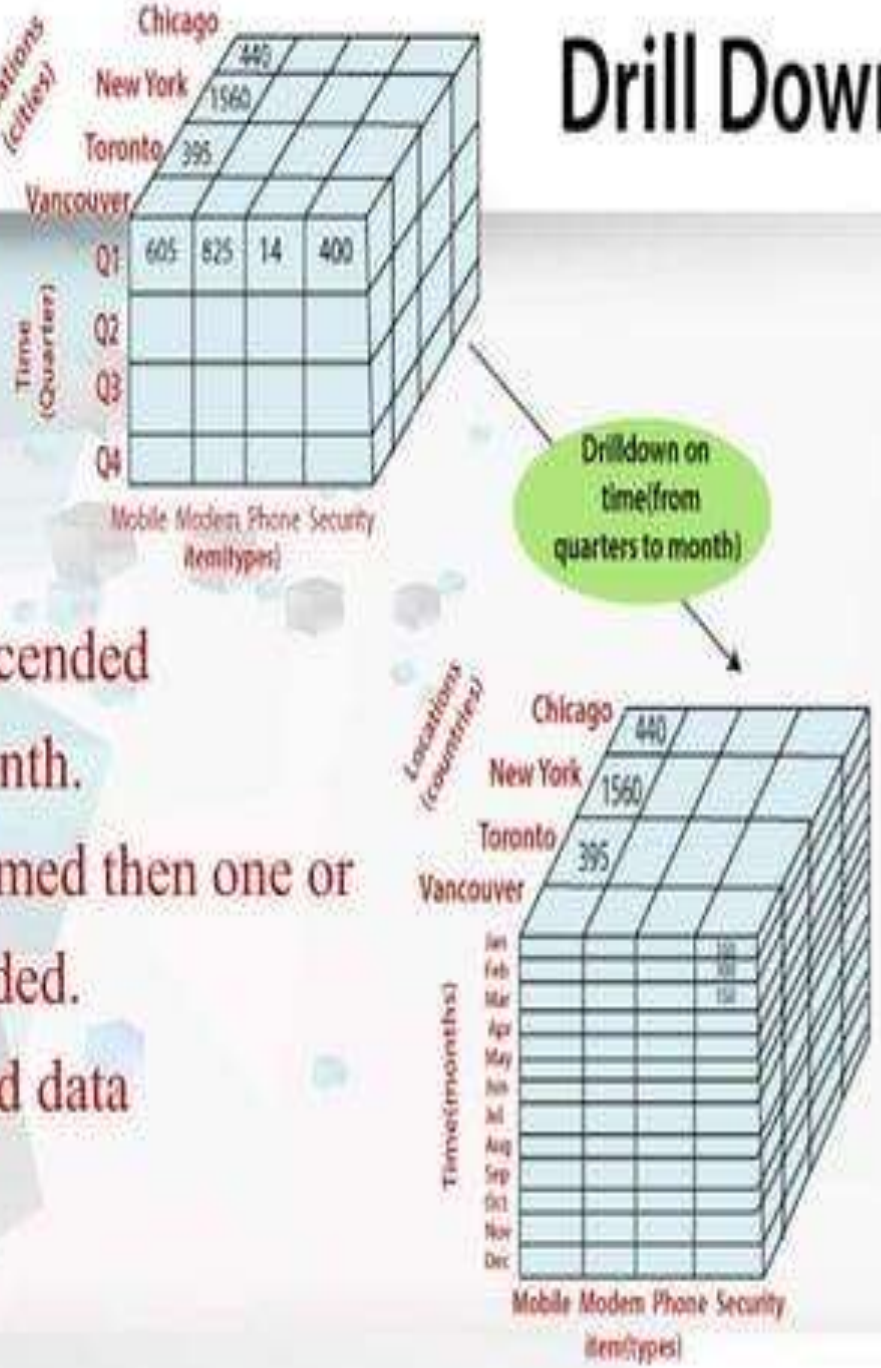
i) By stepping down a concept hierarchy for a dimension.

ii) By introducing new dimension.

Drill Down

(E.g.) In this figure,

- Drill down operation is performed by stepping down a concept of hierarchy for dimension table.
- Initially the concept hierarchy was "day < month < quarter < year"
- On drill-up the time dimension is descended from the level quarter to the level of month.
- When drill-down operations is performed then one or more dimensions from data cube are added.
- If navigates the data from less detailed data to highly detailed data..

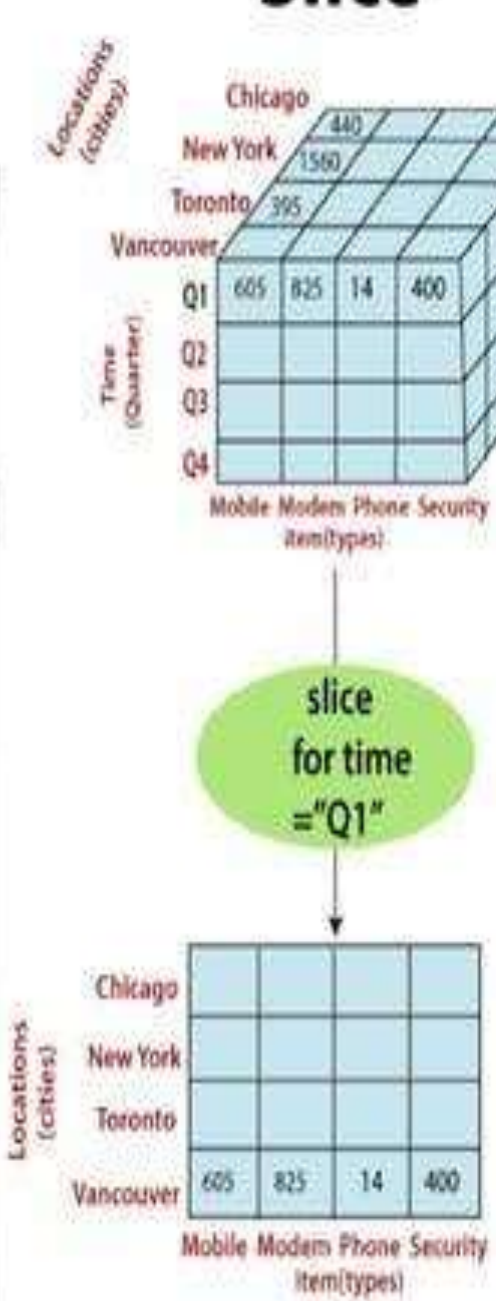


3. SLICE:

The Slice operation performs selection of one dimension on a given cube and gives us a new sub cube:

(E.g.) In this figure,

- Slice operation is performed for the dimension time using the criterion time = "Q1".
- It will form a new sub cube by selecting one or more dimensions.



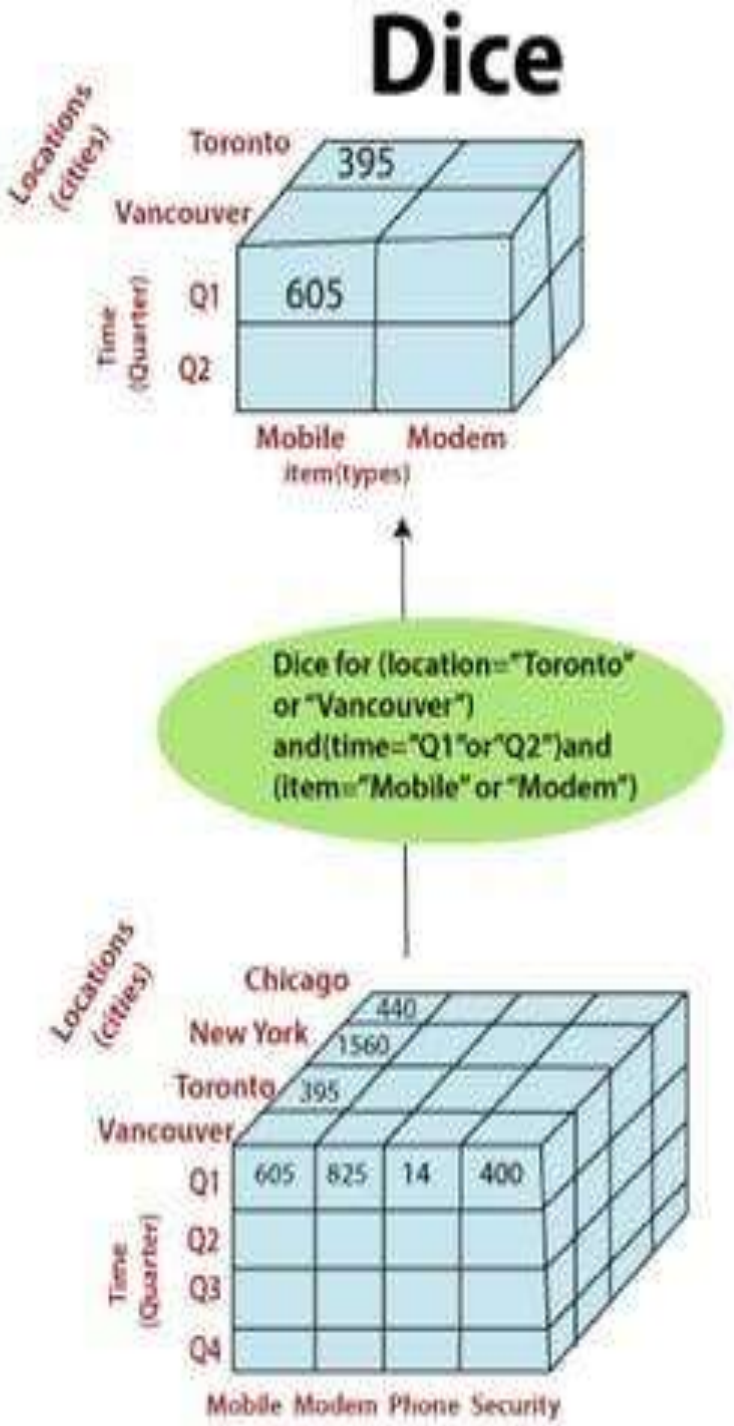
4. DICE:

The dice operation performs selection of two or more dimension on a given cube and gives us a new sub cube:

(E.g.) In this figure,

- The dice operation on the cube based on the following selection criteria that involve three dimensions:

1. (location = "Toronto" or "Vancouver")
2. (time = "Q1" of "Q2")
3. (item = "Mobile" or "Modem")



5. PIVOT:

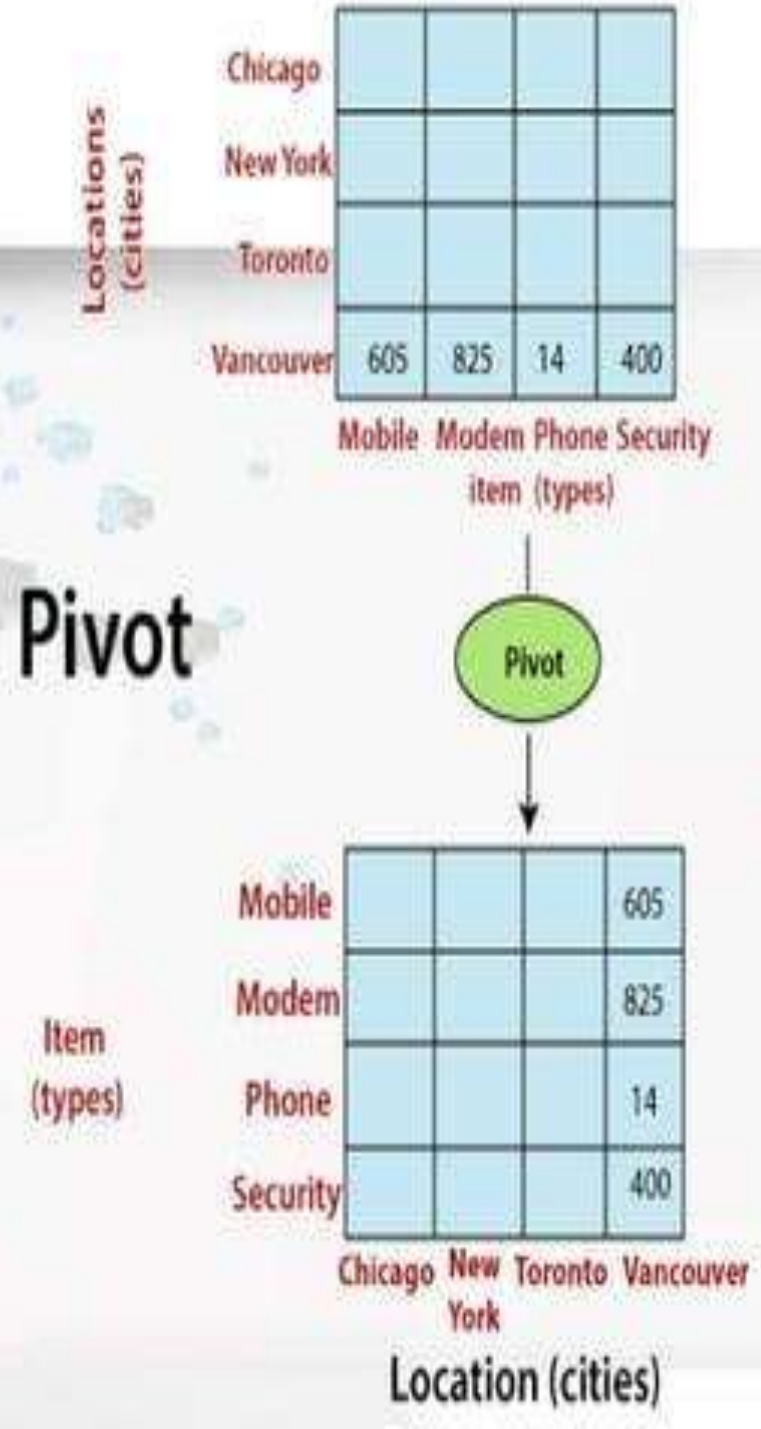
This is a visualization operation that rotates the data axes in an alternative presentation of data.

Pivot operation is also known as rotation.

It rotates the data axes in view in order to provide an alternative presentation of data.

(E.g.) In this figure,

- In this the item and location axes in 2-D slice are rotated.



Implementation Steps of OLAP

The major steps or activities for implementation of OLAP are –

- ✓ Dimensional modelling
- ✓ Design and building of the Multidimensional database (MDDB)
- ✓ Selection of the data to be moved into the OLAP system
- ✓ Data acquisition or extraction for the OLAP system
- ✓ Data loading into the OLAP server
- ✓ Computation of data aggregation and derived data
- ✓ Implementation of application on the desktop
- ✓ Provision of User training

Few Companies OLAP Implementation

- ***The World Bank*** – performs complex statistical analysis on a mass of world wide econometric data.
- ***HP*** – provides speedy operational reports using a desktop OLAP over the corporate intranet to numerous users.
- ***British Airways*** – reduces processing costs through better analysis using OLAP databases tied to new general ledger.
- ***Barclays Bank*** – manages risk on loans for maximum profit.
- ***IBM Finance*** – provides finance portal for cost-effective analysis, reporting and performance management.
- ***Deluxe Corp*** – performs more accurate forecasting through planning and analysis applications.

Advantages & Disadvantages of OLAP

Advantages:

1. Flexibility:

- ✓ It means business users of OLAP applications can become more self-sufficient.
- ✓ It access to strategic information equals more effective decision making.

2. Faster Service:

- ✓ It is possible to build an OLAP system using software designed for transaction processing or data collection.
- ✓ Developers can deliver applications to business users faster, providing better service and it reduces the applications backlog.

3. Reduction of application backlog:

- ✓ It reduces the applications backlog by making business users self-sufficient enough to build their own models.
- ✓ OLAP applications are dependent on data warehouses and transaction processing systems to refresh their source level data.
- ✓ It gains more self-sufficient users without relinquishing control over integrity of data.

4. Reduction of Network Traffic:

- ✓ IT also realizes more efficient operations through OLAP.
- ✓ IT reduces the query drag and network traffic on transaction systems or data warehouse.

5. Efficient Use of Resources:

- ✓ It provide the ability to model real business problems and a more efficient use of people resources, OLAP enables the organization to respond more quickly to market demands.

Disadvantages:

1. Complexity to Administer:

- ✓ OLAP systems provide a comfortable environment for the end user.
- ✓ Complexity increases the level of knowledge required to create business intelligence systems.

2. Data Mart required:

- ✓ OLAP system requires a data mart with a star or snowflake layout.
- ✓ Data must be copied from the OLTP systems into the data mart.

3. Latency:

- ✓ Data mart is required by OLAP, there is automatically some latency in the business intelligence.

4. Read-Only:

- ✓ OLAP data is read-only and if doesn't support, then changes will reflect in OLAP data.

Applications of OLAP

OLAP applications are used by a variety of the functions of an organization -

1. Finance and Accounting:

- ✓ Budgeting.
- ✓ Financial performance analysis.
- ✓ Activity-based costing.
- ✓ Financial modelling.

2. Sales and Marketing:

- ✓ Sales analysis and forecasting.
- ✓ Promotion analysis.
- ✓ Market and Customer segmentation.
- ✓ Market research analysis.
- ✓ Customer analysis.

3. Production:

- ✓ Production Planning.
- ✓ Defect analysis.

OLTP Vs OLAP

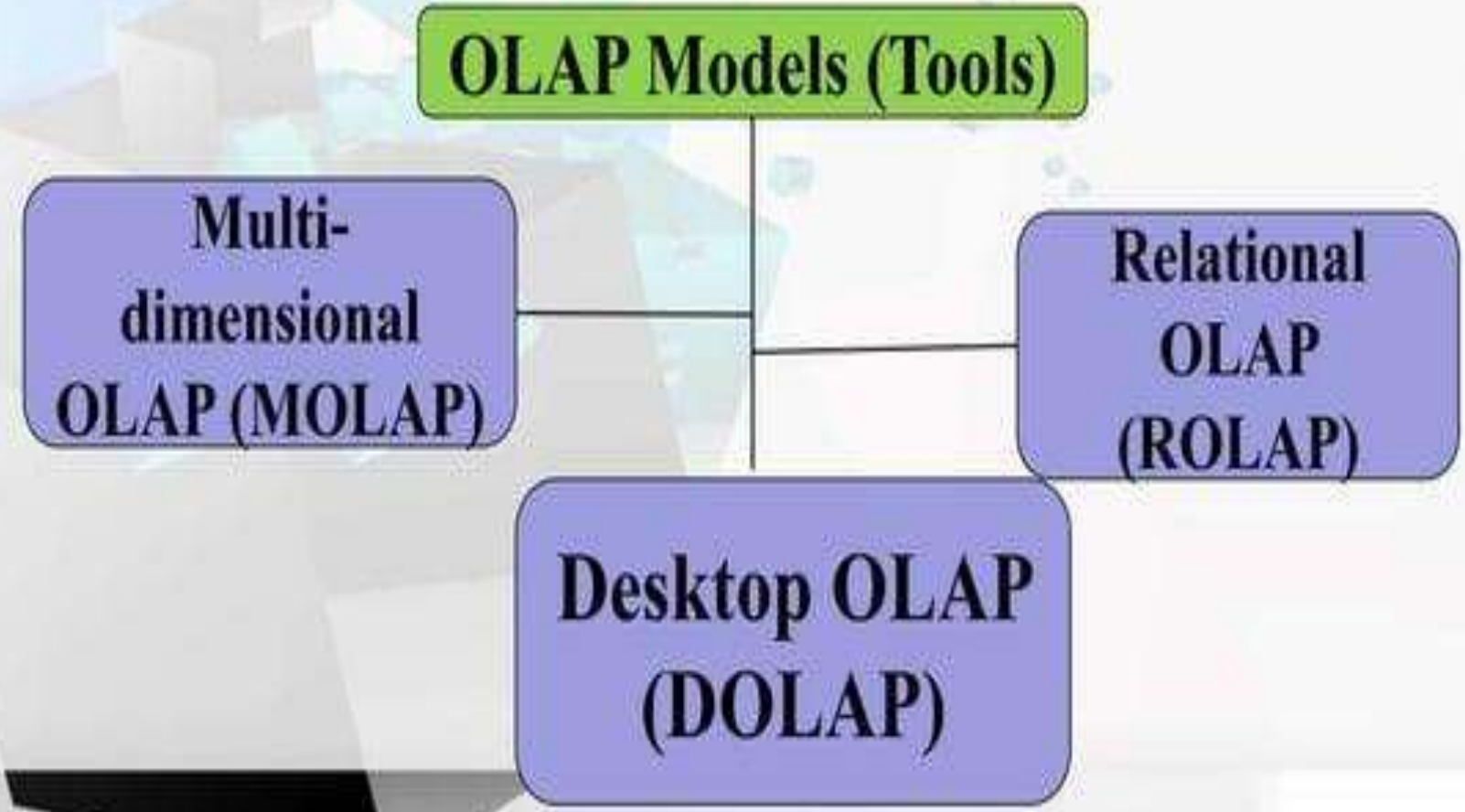
Basis	OLTP	OLAP
Source of Data	Operational data, OLTPs are the original source of data	Consolidation data, OLAP data comes from the various OLTP databases
Purpose of Data	To control and run fundamental business tasks	To help with planning, problem solving and decision support
What the Data Reveals	A snapshot of ongoing businesses processes	Multi-dimensional views of various kinds of business activities
Inserts and Updates	Short and fast inserts and updates initiated by end users	Periodic long-running batch jobs refresh the data
Queries	Relatively standardized and simple queries returning relatively few records	Often complex queries involving aggregations
Processing Speed	Typically very fast	Depends on the amount of data involved, batch data refreshes and complex queries may take many hours, query speed can be improved by creating indexes

OLTP Vs OLAP

Basis	OLTP	OLAP
Space requirements	Can be relatively small if historical data is achieved	Larger due to the existence of aggregation structures and history data, requires more indexes than OLTP
Data base Design	Highly normalized with many tables	Typically de-normalized with fewer tables, use of star or snowflake schemas
Backup and Recovery	Backup religiously, operational data is critical to run the business, data loss is likely to entail significant monetary loss and legal liability	Instead of regular backups, some environments may consider simply reloading the OLTP data as a recovery method

OLAP Models

- The OLAP server understands how data is organized in the database and has special functions for analysing the data.
- The three major alternatives for implementing OLAP applications are:



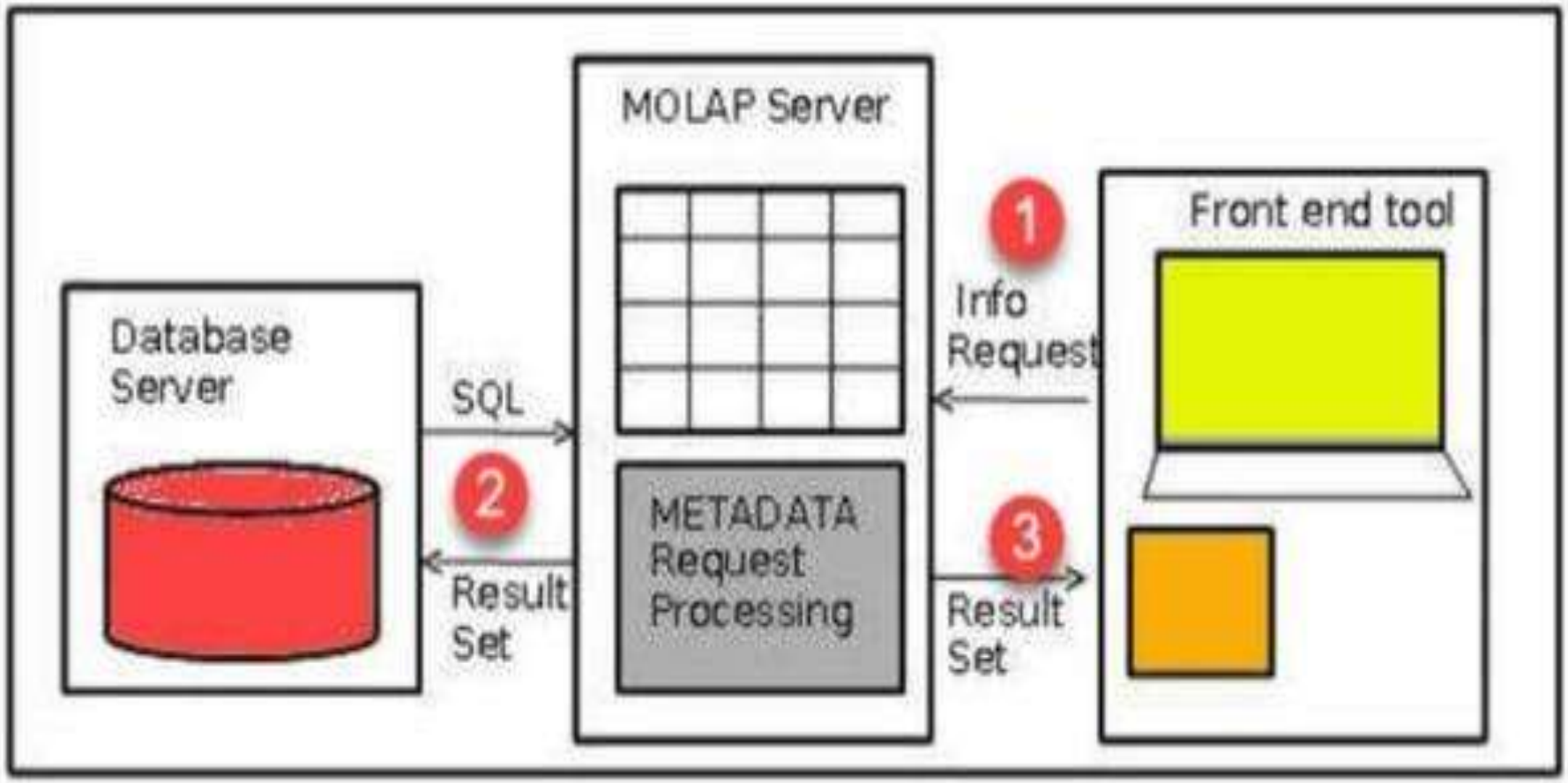
1. Multidimensional OLAP (MOLAP):

- ✓ In MOLAP model, data for analysis is stored in multidimensional databases.
- ✓ Specialized data structures (multidimensional) need to organize, navigate and analyze the data in aggregated form.
- ✓ Large multidimensional arrays form the storage structures.

(E.g.) To store no. of 500 units for Product A in Jan month 2009, in store S1 will be stored in an array represented by the values (Product A, 2009/01, Store S1).

- ✓ The array value indicates the location of the cells and it is intersection of the values of the dimension attributes.
- ✓ It provide capability to consolidate and fabricate summarized cubes during the process that loads data into MDDBs from main data warehouse.

MOLAP Architecture



MOLAP Architecture

Advantages:

- ✓ It provides excellent performance when data is accessed as intended at design time.
- ✓ It supports pre-defined analysis of trends over time (E.g.) sales by customer type by geographic area).
- ✓ All calculations have been pre-generated when cube is created.
- ✓ Complex calculations are not only feasible, but they return quickly.
- ✓ The data for analysis is stored and maintained in a persistent structure, which reduces the overhead of performing calculations and building aggregations during query processing.

Disadvantages:

- ✓ The inflexible nature of the multi dimension hierarchy limits the ability to support multiple business areas from the one database (i.e.) different views of same data.
- ✓ When dimension requirements change, the data cube may need to be reorganized.
- ✓ The aggregated nature of data makes it difficult to 'drill down' to data required by many analysis applications.
- ✓ It require new skills and toolkits to build and maintain multidimensional databases, thus increasing the development and maintenance cost.
- ✓ As the user base is much lower than for RDBMS there are fewer tools and utilities to support multi-dimensional DBMS.
- ✓ There is a heavy processing requirement to re-calculate all of the aggregations every time that new data is added to the cube.
- ✓ They are not suited to large databases (i.e.) > 10 GB.

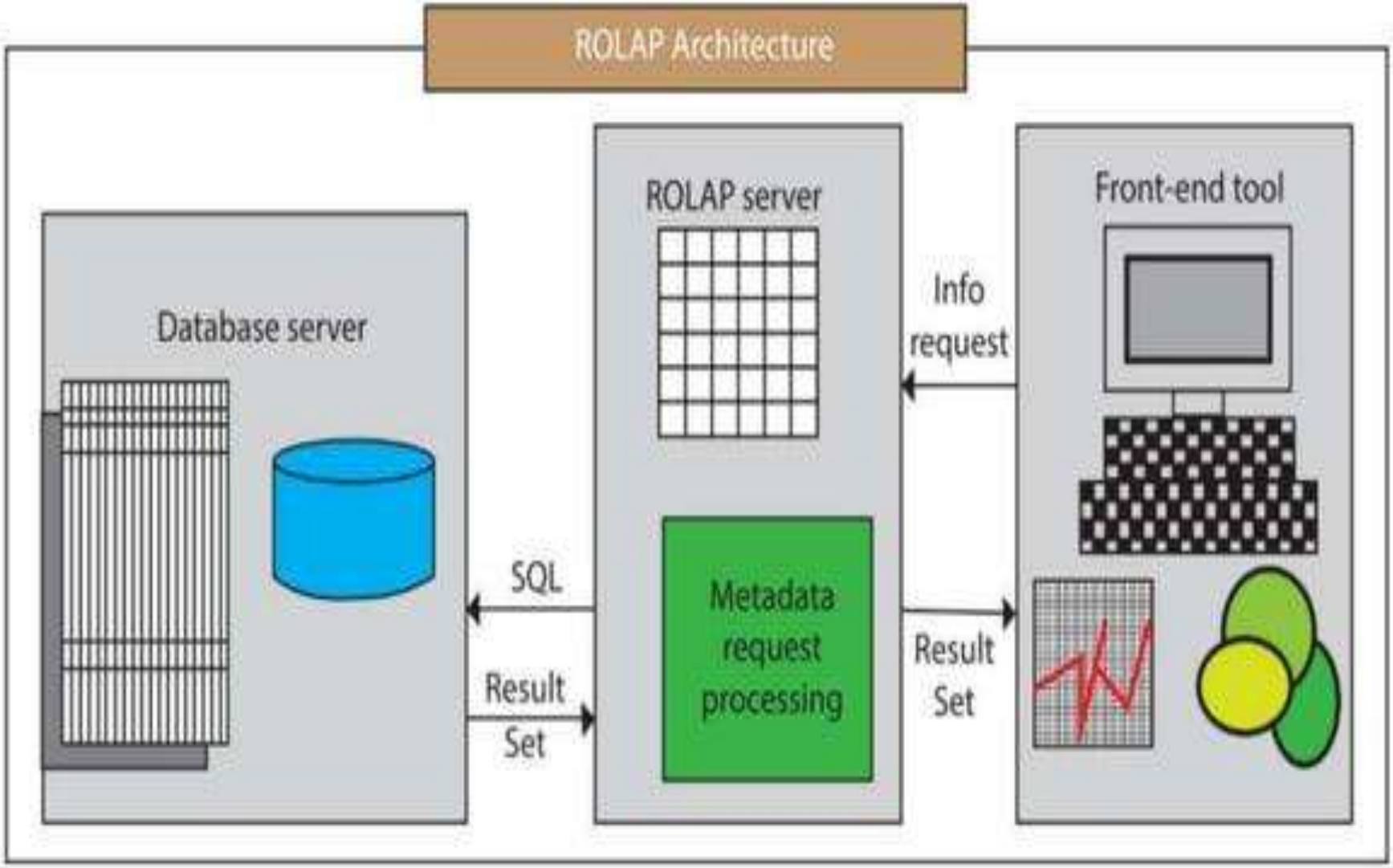
2. Relational OLAP (ROLAP):

- ✓ ROLAP supports large databases while enabling good performance, platform portability, exploitation of hardware advances like parallel processing, robust security, multi-user concurrent access and openness to multiple vendor's tools.
- ✓ ROLAP is based on familiar, proven and already selected technologies.
- ✓ The metadata layer supports the mapping of dimensions to the relational tables. (i.e.) supports summarizations and aggregations.

ROLAP has three distinct characteristics -

- Supports all the basic OLAP features and functions.
- Stores data in a relational form.
- Supports some form of aggregation.

ROLAP Architecture



Advantages:

- ✓ It perform more robust types of analysis than multi-dimensional OLAP, incorporating a wider range of data views.
- ✓ It supports flexible analysis (i.e.) analysis can explore the data.
- ✓ It allows multi-dimensional views over the same data tables, to support different business areas.
- ✓ It can support large volumes of detailed data for 'drill down analysis.
- ✓ The processing load can be shifted onto a server platform, allowing to use with network computers.
- ✓ There is no requirement to return aggregations when additional data is added to the relational databases.

Disadvantages:

- ✓ Using relational databases for multi-dimensional analysis can lead to performance issues when data is extracted from the databases, if database is not well designed.
- ✓ The denormalized database design gives rise to substantial data redundancy.
- ✓ Many relational OLAP products don't provide same degree of data visualization found in multi-dimensional OLAP tools.
- ✓ Relational data structures require more time for simple OLAP analysis than multi-dimensional OLAP tools because the data must be extracted prior to analysis.

3. Desktop OLAP (DOLAP):

- ✓ DOLAP model allows a user to download a piece of data from a database or source and work with it locally or on their desktop.
- ✓ DOLAP is single-tier, desktop-based OLAP technology.
- ✓ It is able to download a relatively small hypercube from a central point, usually from data mart or data warehouse and perform multidimensional analyses.
- ✓ Data sets are limited to boundaries defined by user with no access to granular data.
- ✓ Cubes contain summarized data, organized in a fixed structure of dimensions.
- ✓ It is ideal for well-understood, recurring analytic questions and reporting.

Advantages:

- ✓ User friendly – user can pivot and manipulate data locally from the returned result set stored on desktop.
- ✓ Excellent query performance – it collects, aggregates and calculates data in advance of analysis.
- ✓ Low cost per seat and maintenance.
- ✓ Useful for mobile users who can't always connect to the data warehouse.
- ✓ Easiest to deploy among all OLAP approaches.

Disadvantages:

- ✓ Limited functionality and data capacity.

Other types of OLAP:

4. HOLAP – Hybrid OLAP Model is an application that combines MOLAP and ROLAP's greatest characteristics into a single architecture. HOLAP systems store a larger amount of detailed data in relational tables, while aggregations are saved in pre-calculated cubes. A hybrid OLAP model is provided by Microsoft SQL Server 2000.

5. Web-Enabled OLAP Models (WOLAP) – it refers to an OLAP program that may be accessed through a web browser. WOLAP is a three-tiered architecture that consists of three components: a client, middleware, and database server, as opposed to standard client/server OLAP systems.

6. Spatial OLAP Models (SOLAP) - combines the capabilities of both GIS and OLAP into a single user interface. It helps with both spatial and non-spatial data management. Spatial OLAP, for example, can be used to analyze regional weather trends.