Business Intelligence (BI) and Data Warehousing (DW)

OLTP AND OLAP Architectures

Jackline Ssanyu

January 29, 2020

OLTP and OLAP

- Todays **IT systems** are **divided** into **Online Transaction processing (OLTP)** and **Online Analytical Processing (OLAP)** systems.
  - **OLTP** systems **provide source data to data warehouses**
  - **OLAP** systems **help to analyze it**.

OLTP

- OLTP supports transaction-oriented applications in a 3-tier architecture.
- OLTP administers day to day short online transactions of an organization, such as INSERT, DELETE, UPDATE, using SQL.
- The primary objective is data processing and not data analysis
- The main emphasis for OLTP systems is put on very fast query processing-maintaining data integrity in multi-access environments, and an effectiveness -measured by number of transactions per second.
- In OLTP database there is detailed and current data, and schema used to store transactional databases is the entity model (usually 3NF).

Example of an OLTP System

- **ATM center**: Assume that a couple has a joint account with a bank. One day both simultaneously reach different ATM centers at precisely the same time and want to withdraw total amount present in their bank account.
  - However, the person that completes authentication process first will be able to get money.
  - In this case, OLTP system makes sure that withdrawn amount will never be more than the amount present in the bank.
  - The key to note here is that **OLTP systems are optimized for transactional superiority instead data analysis**.

Other Examples OLTP Systems

- Online banking
- Online airline ticket booking
- Sending a text message
- Order entry
- Add a book to shopping cart

OLTP-OLAP Architectures.

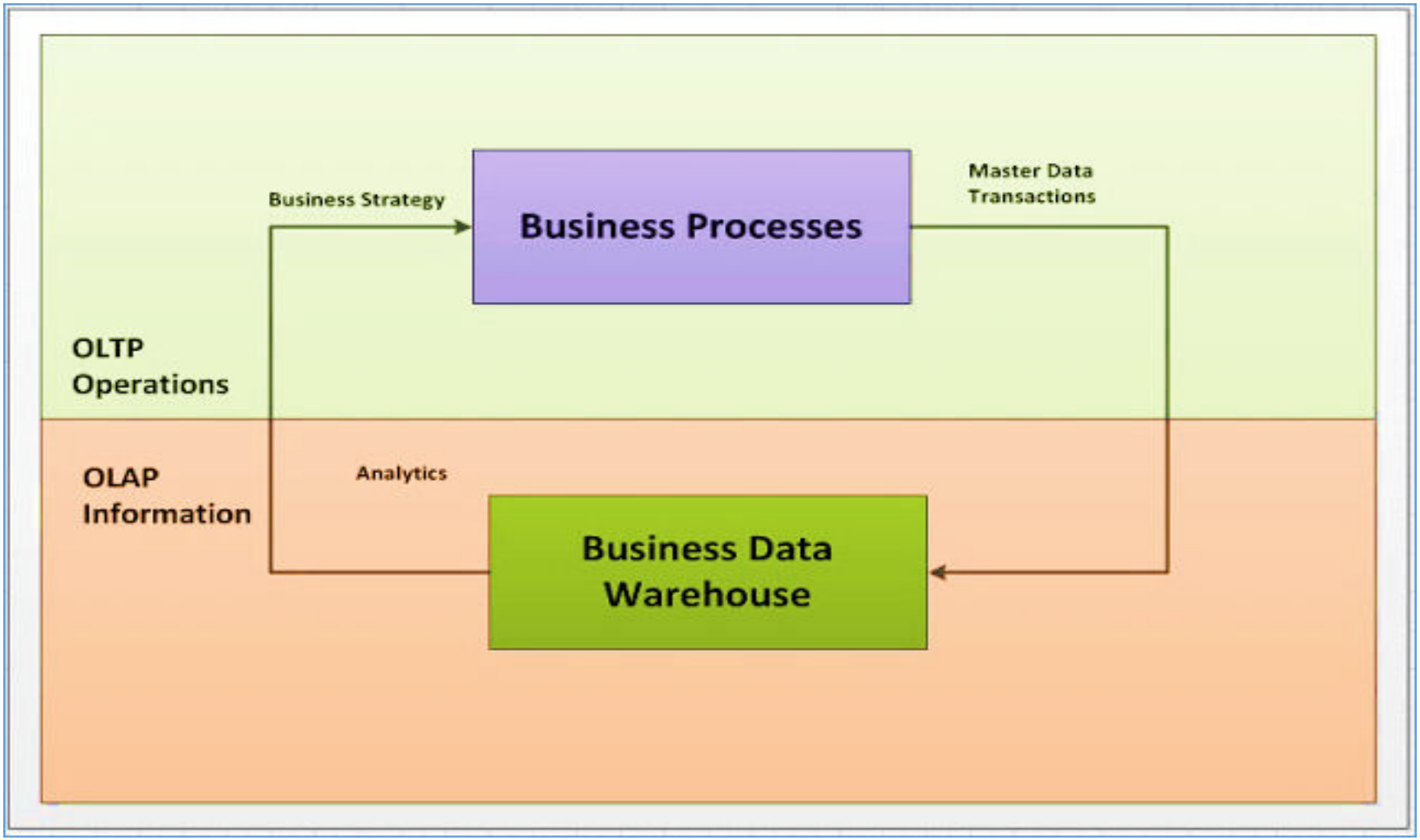


Figure: OLTP-OLAP Architectures.

OLAP

- OLAP is a category of software tools which provide analysis of data for business decisions.
- OLAP systems allow users to analyze database information from multiple database systems at one time.
- The primary **objective is data analysis and not data processing**.
- OLAP systems are characterized by relatively **low volume of transactions**.
- Queries are often very complex and involve aggregations.
- For OLAP systems a response time is an effectiveness measure.
- OLAP applications are widely used by Data Mining techniques.
- In OLAP database there is aggregated, historical data, stored in multi-dimensional schemas (usually star schema).

Example of an OLAP System

- Any Data warehouse system is an OLAP system. Some Sample uses of OLAP are as follows
  - A company might compare their mobile phone sales in September with sales in October, then compare those results with the results in another location which may be stored in a sperate database.
  - Amazon analyzes purchases by its customers to come up with a personalized homepage with products which likely interest to their customer.

OLAP Architecture.

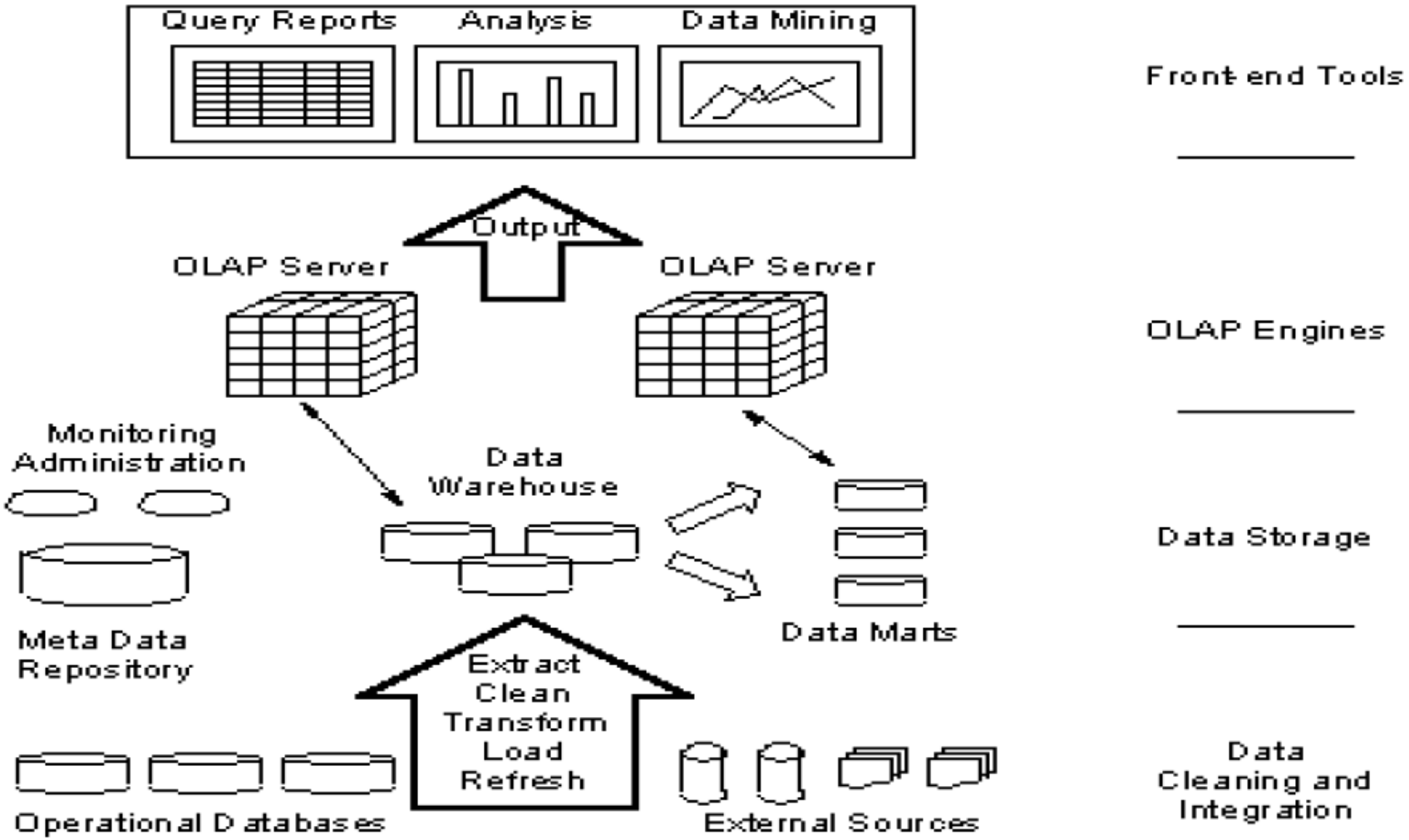


Figure: OLAP Architecture.

OLTP vs. OLAP

- In general we can assume that OLTP systems provide source data to data warehouses, whereas OLAP systems help to analyze it.

OLTP vs. OLAP Summary

- OLAP is a category of software tools that analyze data stored in a database.
- OLTP supports transaction-oriented applications in a 3-tier architecture
- OLAP creates a single platform for all types of business analysis needs which includes planning, budgeting, forecasting, and analysis.
- OLTP is useful to administer day to day transactions of an organization.
- OLAP is characterized by a large volume of data.
- OLTP is characterized by large numbers of short online transactions.

OLAP Characteristics

OLAP systems share four main characteristics:

- They use multidimensional data analysis techniques.
- They provide advanced database support.
- They provide easy-to-use end-user interfaces.
- They support the client/server architecture.

Multidimensional data analysis techniques

- Multidimensional data analysis is the process of analyzing and categorizing data based on multiple dimensions (such as, customers, products and time) and measures.
- It helps in the mathematical modelling of business processes, where large and complex data is involved across multiple regions and products.
- The analysis generated from these models can help in decision-making and planning activities of business operations.
- **Multidimensional data** is presented making use of **data cubes**.

Advanced database support

To deliver efficient decision support, OLAP tools must have advanced data access features. Such features include:

- Access to many different kinds of DBMSs, flat files, and internal and external data sources.
- Access to aggregated data warehouse data as well as to the detail data found in operational databases.
- Advanced data navigation features such as drill-down and roll-up.
- Rapid and consistent query response times.
- Support for very large databases. The data warehouse can easily and quickly grow to multiple gigabytes and even terabytes.

OLAP Operational architecture

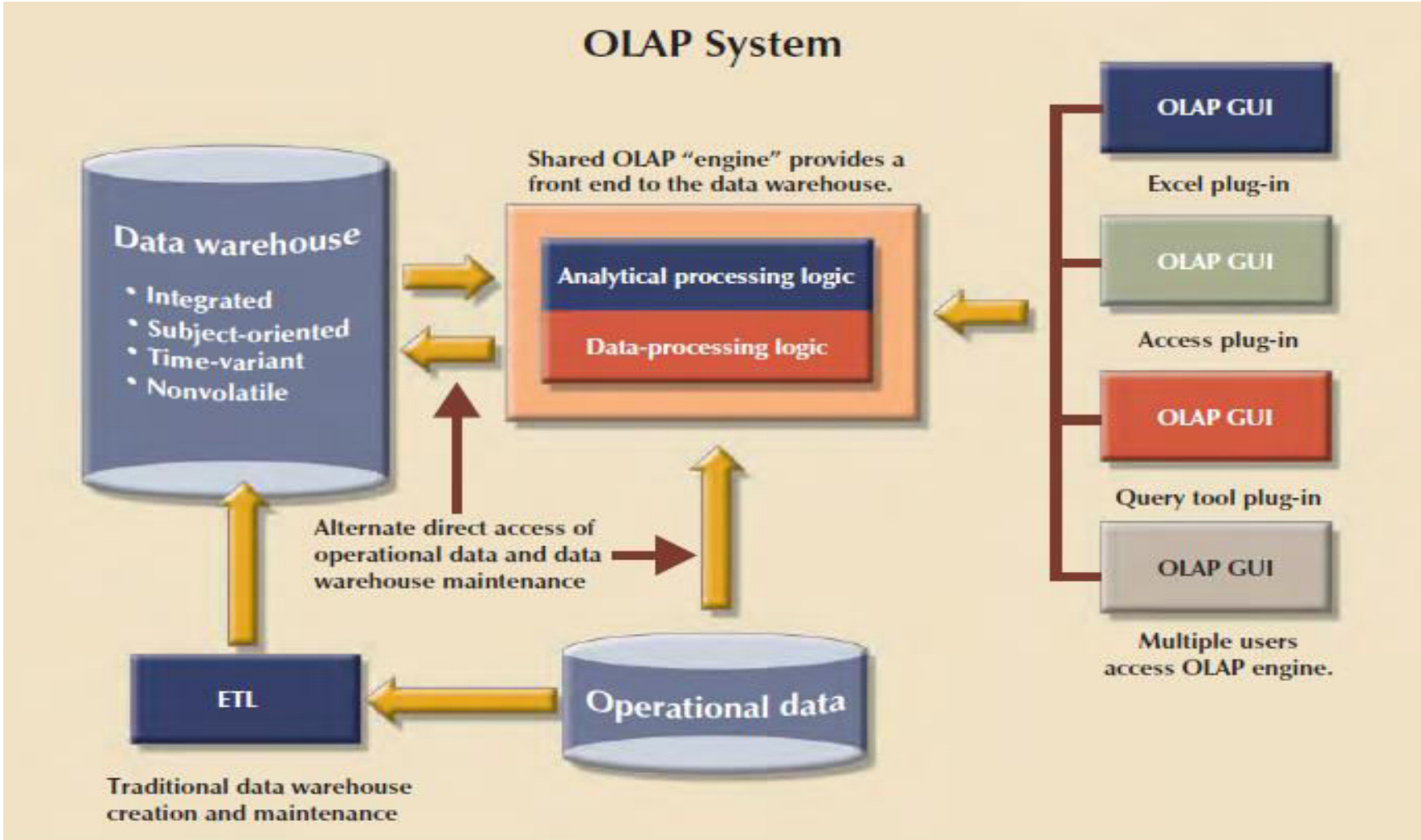


Figure: OLAP client/server architecture

OLAP client/server architecture...

As shown in the previous figure, OLAP client/server architecture is composed of three components: **OLAP GUI**, **OLAP engine** (server) and **Data warehouse's decision support database**.

- **OLAP GUI** runs on client work stations
- **OLAP engine** (server) is composed of OLAP analytical processing logic and data processing logic runs on a shared computer.
- OLAP server is the front-end to the Data warehouse's decision support database. It is the middle layer between data warehouse and end-user GUI.
- OLAP server accepts and processes the data processing requests generated by the end-user analytical tools.

OLAP client/server architecture...

- The data warehouse is created and maintained by a process/software tool (ETL) that is independent of the OLAP system.
- OLAP system might access both data storage types (operational database or data warehouse) or only one. This depends on the vendors implementation of the product selected.
- An OLAP system can also directly access operational data, transforming it and storing it in a multidimensional structure. In other words, the OLAP system can provide an alternative ETL and multidimensional data store component.
- The data warehouse handles the data component more efficiently than OLAP does, therefore it is beneficial to have a central data warehouse to serve as the large enterprise decision support database.

OLAP server with local mini-data-marts

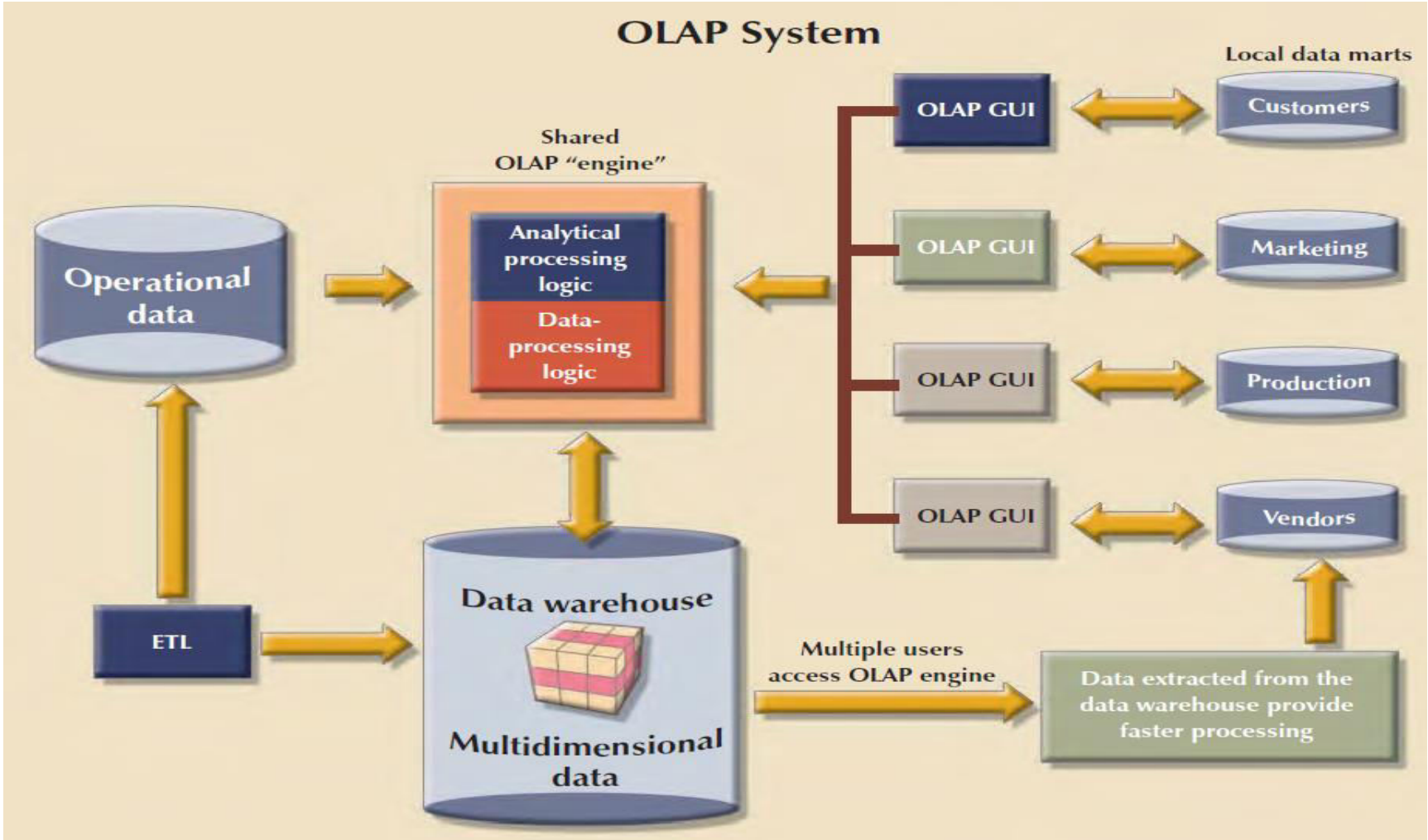


Figure: OLAP server with local mini-data-marts

OLAP server with local mini-data-marts...

- To provide better performance some OLAP systems merge data warehouse and data mart approaches.
- They store small extracts of the data warehouse at end-user work stations as shown in the previous figure.
- The main objective is to increase speed of data access and data visualization (the graphic representations of data trends and characteristics).

OLAP Models

There are three types of widely used OLAP models:

- **Relational OLAP(ROLAP)**: ROLAP is an extended RDBMS along with multidimensional data mapping to perform the standard relational operation.
- **Multidimensional OLAP (MOLAP)**:MOLAP stores data on disks in specialized multidimensional array structure.
- **Hybrid OnlineAnalytical Processing (HOLAP)**:In HOLAP approach the aggregated totals are stored in a multidimensional database while the detailed data is stored in the relational database. This offers both data efficiency of the ROLAP model and the performance of the MOLAP model.

Relational OLAP (ROLAP)

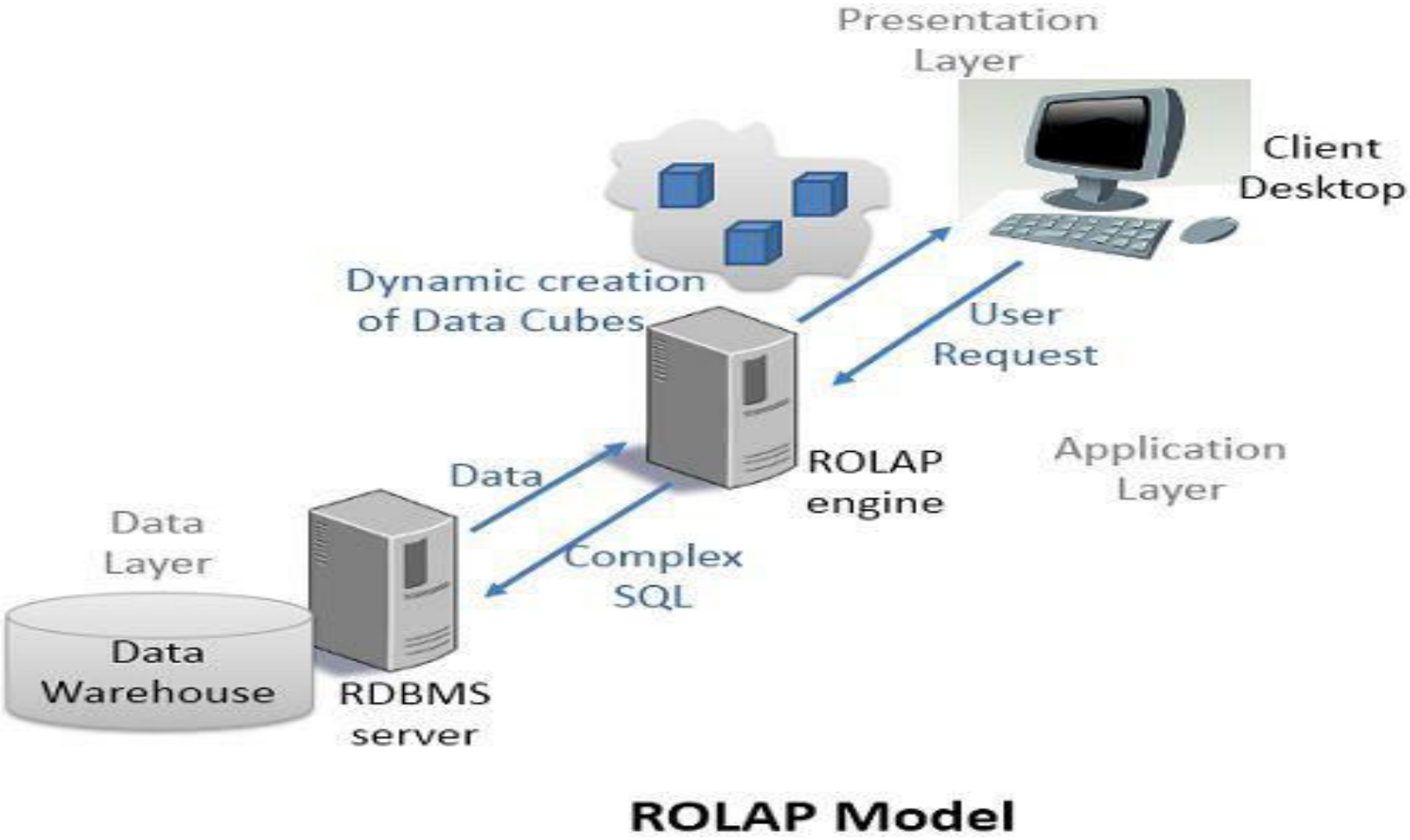


Figure: ROLAP Model

Relational OLAP (ROLAP)...

- Data is stored as in relational database i.e. rows and columns in the data warehouse.
- Data is presented in the front of the user in the multidimensional form.
- Whenever the ROLAP engine in analytical server issues a complex query, it fetches data from the main warehouse and dynamically creates a multidimensional view of data for the user.
- It is suitable for companies already using RDBMS. They can naturally extend the operational databases to Data warehouse.
- As the multidimensional view of data is created dynamically it processes slower in comparison to MOLAP.

Multidimensional Data Schema Support within the RDBMS

- Relational databases use normalization to reduce redundancies, thereby eliminating data anomalies, and to facilitate data updates.
- Unfortunately, for decision support purposes, it is easier to understand data when they are seen with respect to other data.
- Decision support data tend to be non-normalized and duplicated.
- ROLAP uses special design techniques to enable RDBMS technology to support multidimensional data representations.
- These special design techniques are: Star schema, Snow-Flake and Constellation.
- These schemas are designed to optimize data query operations rather than data update operations

Multidimensional OLAP (MOLAP)

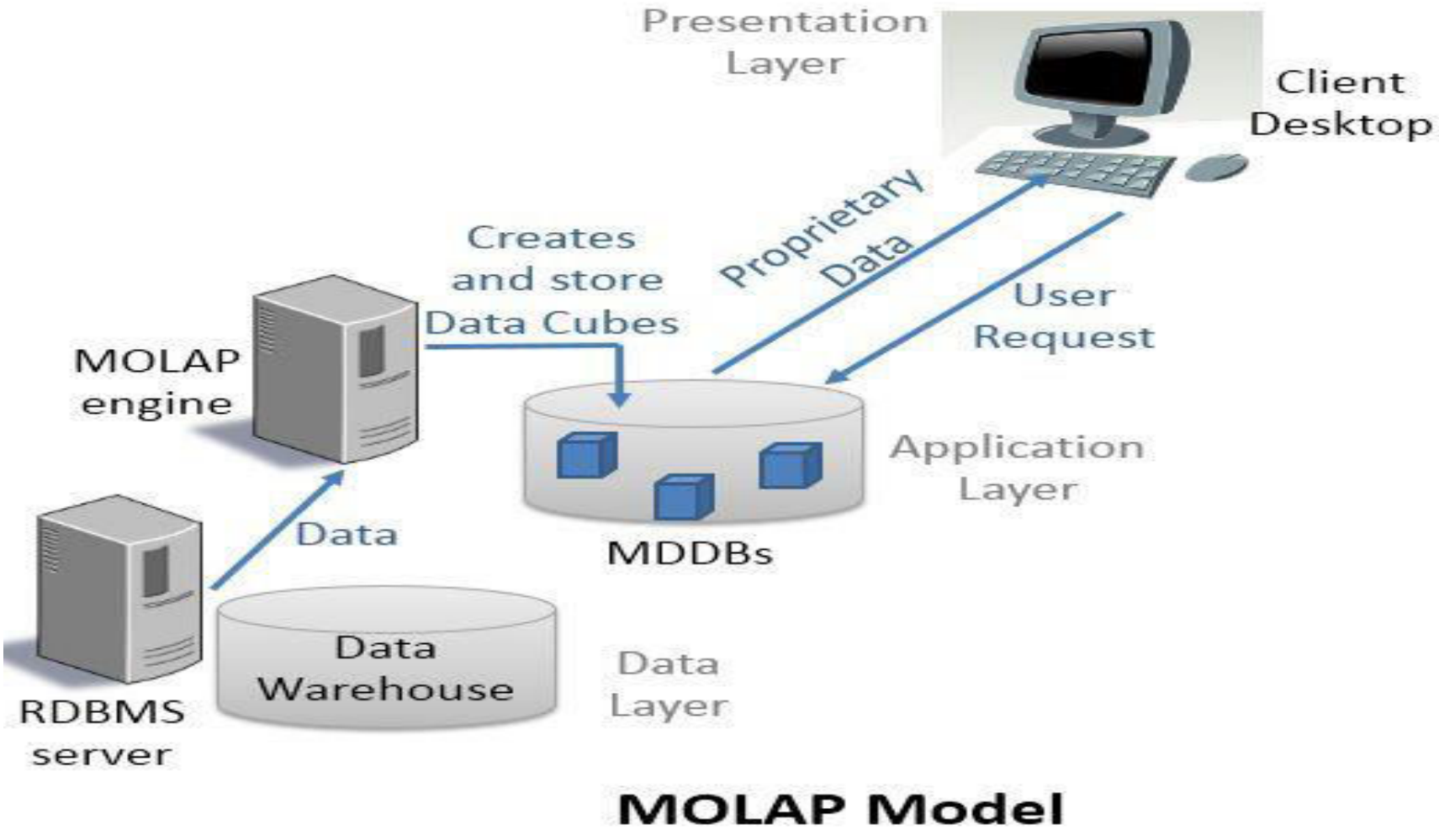


Figure: MOLAP Model

Multidimensional OLAP (MOLAP)...

- The data used for analysis is stored in specialized multidimensional databases (MDDBs).
- MDDBs are proprietary software systems which store data in a multidimensional view using **Data cubes**.
- MDBMS end users visualize the stored data as a three-dimensional cube.
- Data cubes are created by extracting data from the operational databases or from the data warehouse.
- The MOLAP engine, which resides in the application layer, provide the multidimensional view of data from MDDBs to the user.
- Thus when a user requests for the data, no time is wasted in calculating the data (as in ROLAP) and the system responses fast.

Hybrid OnlineAnalytical Processing (HOLAP)...

- Hybrid OLAP is a mixture of both ROLAP and MOLAP. It offers fast computation of MOLAP and higher scalability of ROLAP. HOLAP uses two databases.
  - Aggregated or computed data is stored in a multidimensional OLAP cube
  - Detailed information is stored in a relational database.
- ROLAP are instantly updated and HOLAP users have access to this real-time instantly updated data. MOLAP brings cleaning and conversion of data thereby improving data relevance. This brings best of both worlds.

OLAP Cubes

- OLAP Cube (also known as a Multi-dimensional Cube or data cube) is a data structure in data warehousing that allows fast analysis of data according to the multiple Dimensions that define a business problem. A multidimensional cube for reporting sales might be, for example, composed of 5 Dimensions: Salesperson, Sales, Region, Product, and time.
- Using the cube, it is easy to carry out certain activities which enable the business users to understand the trend of the business.
- Cubes are data structures that the data warehouse uses to contain the data that is imported.
- The cubes divide the data into subsets that are defined by dimensions. You use one or more cubes to create an application.

OLAP Cubes...

- A Cube can accommodate any number of arrays, or Dimensions.
- A **dimension** is a structure that categorizes **facts** and **measures** in order to enable users to answer business questions. Commonly used dimensions are people, products, place and time. For example a cube for reporting sales might be, for example, composed of 3 Dimensions: Customer, Product and Time.
- **Measures** are the core of the dimensional model and are data elements that can be summed, averaged, or mathematically manipulated. For example, sales amount, units sold are examples of measures. Measures are used to determine how a business is operating.

OLAP Cubes- Example1

- The following data Cube consists of three dimensions: Customer, Product and Time.
- It would be queried (in SQL) BY customer, BY product, BY time.
- For example, sample query could be to select the customer By time and By product.

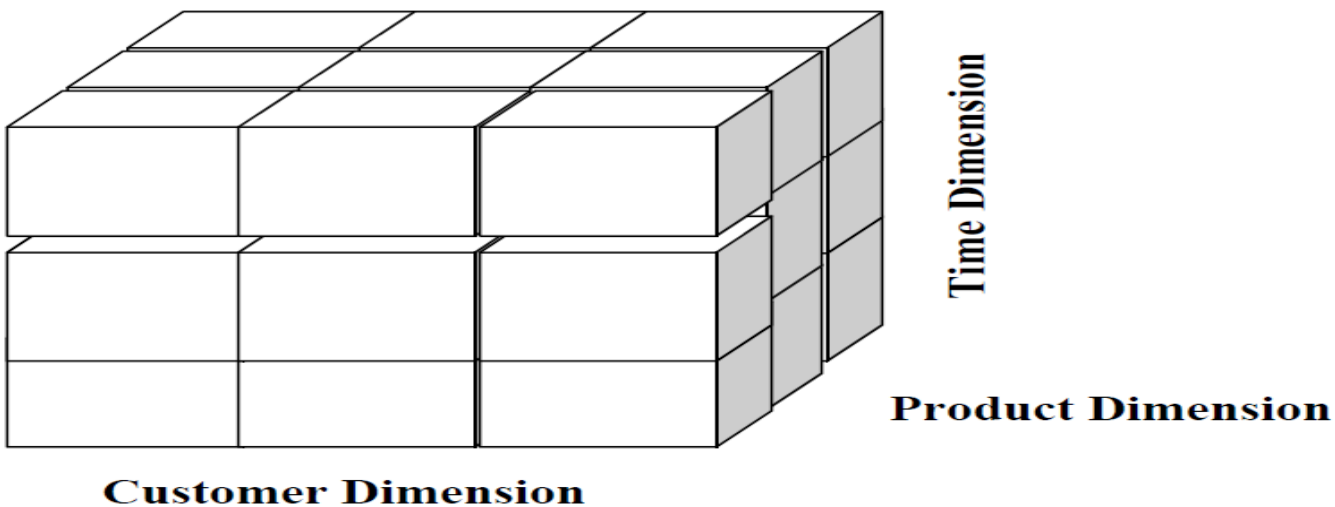


Figure: OLAP Cube

OLAP Cubes- Example2

- The following data Cube consists of three dimensions: Country, Product and Time, which would be queried.
- For example, sample query could be to select the total sales for TVs by country for the first quarter.

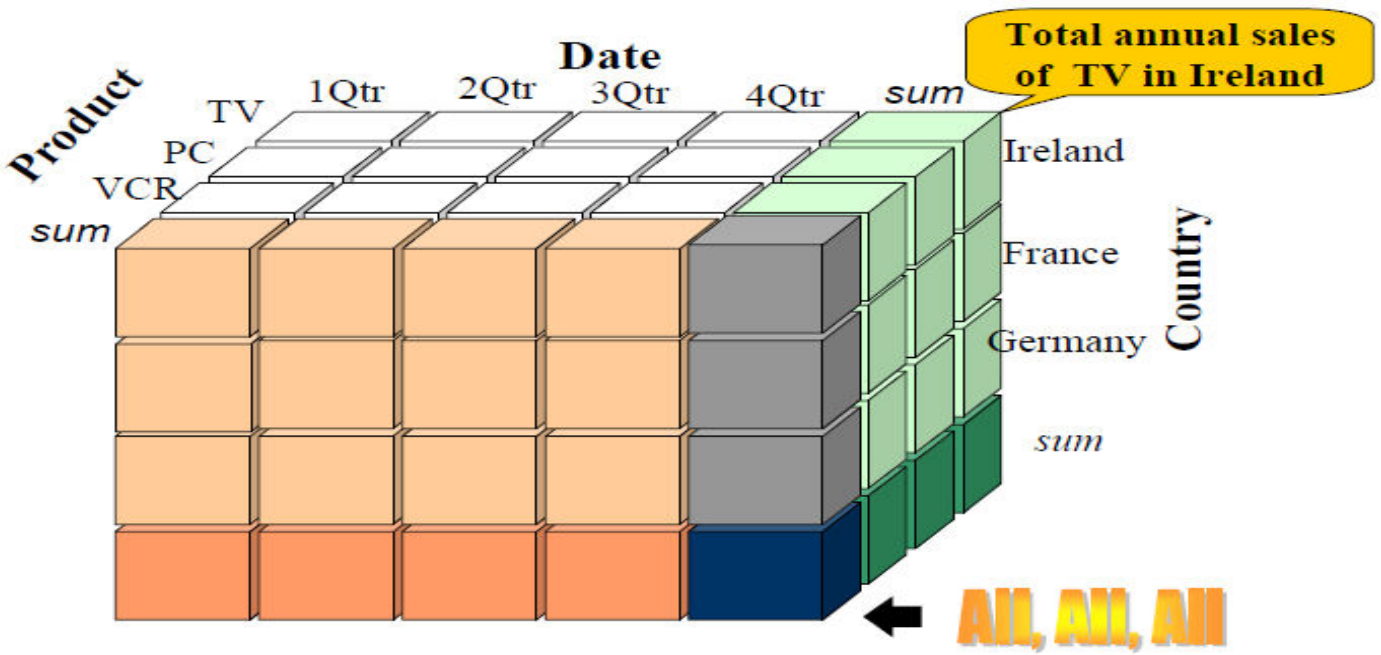


Figure: OLAP Cube 2

OLAP Operations

- In the multidimensional model, the records are organized into various dimensions, and each dimension includes multiple levels of abstraction described by concept hierarchies.
  - A concept hierarchy defines a sequence of mappings from a set of low-level concepts to higher-level, more general concepts.
  - **Example 1:** Consider a concept hierarchy for the dimension location. You can organize location by Country, county, sub-county, village and Zone.
  - **Example 2:** You can organize time by year, quarter, month, day and even hours
- This organization support users with the flexibility to view data from various perspectives.
- A number of OLAP data cube operation exist to demonstrate these different views, allowing interactive queries and search of the record at hand. Hence, OLAP supports a user-friendly environment for interactive data analysis.

Drill-Down (Cont.)

There are five basic analytical operations that can be performed on an OLAP cube:

- **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
- **Slice**: It selects a single dimension from the OLAP cube which results in a new sub-cube creation.
- **Dice** : It selects a sub-cube from the OLAP cube by selecting two or more dimensions.
- **Pivot**:It is also known as rotation operation as it rotates the current view to get a new view of the representation.

Roll-Up

It performs aggregation on the OLAP cube. It can be done by:

- Reducing dimensions
- Climbing up concept hierarchy. Concept hierarchy is a system of grouping things based on their order or level.

Roll-Up - Example1

Cities: Delhi, New York, Patiala and Los Angeles win 5, 2, 3 and 5 medals respectively. So in this example, we roll upon Location from cities to countries. i.e. More detailed data to less detailed data.

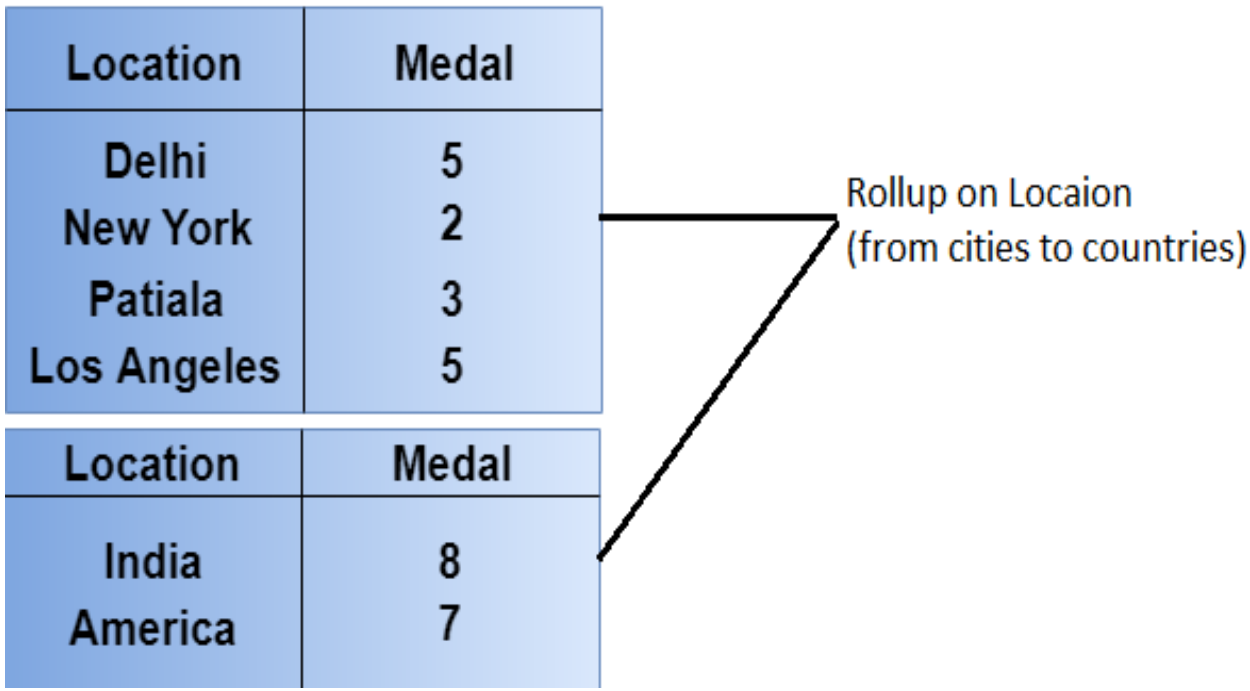
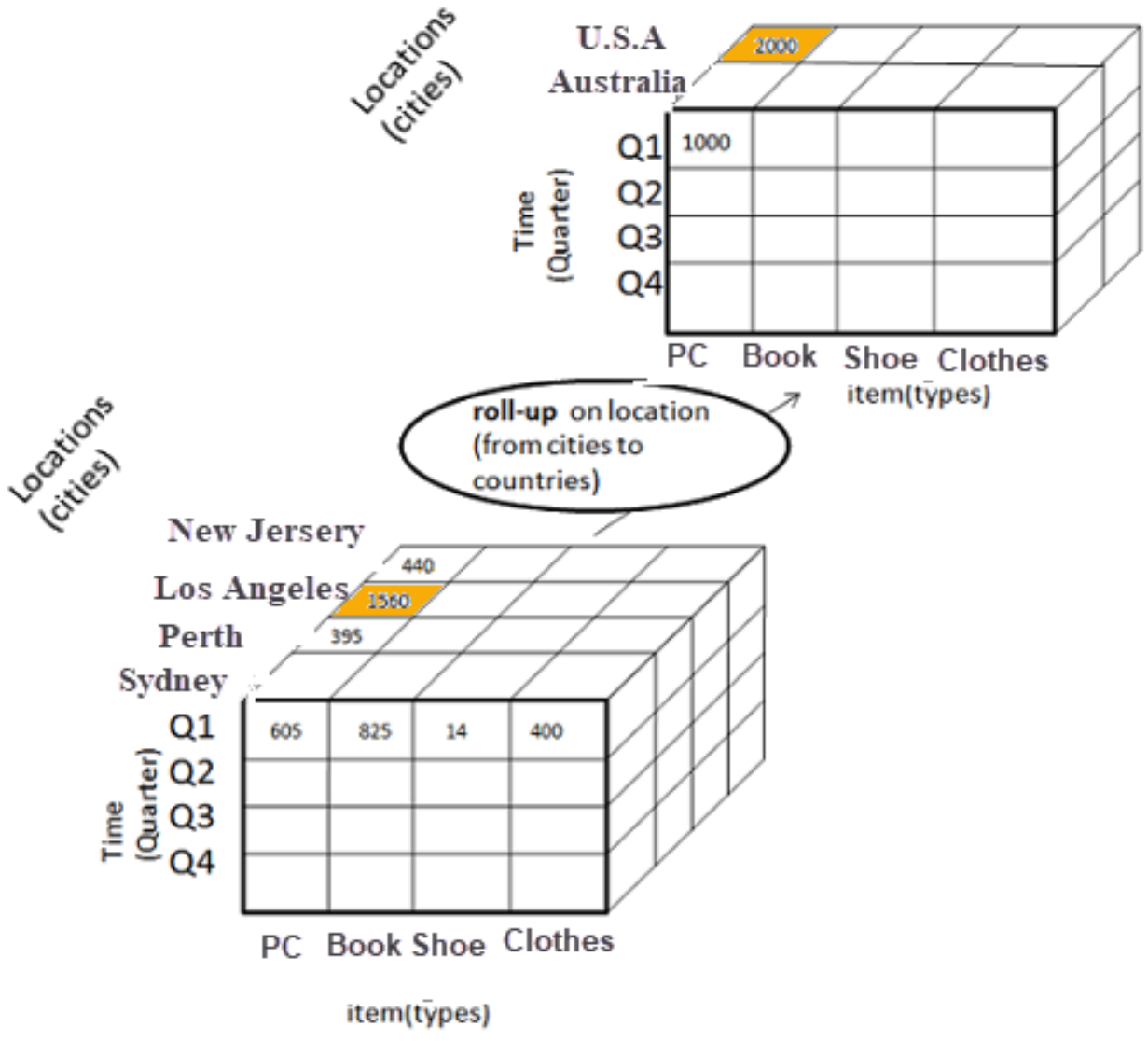


Figure: Roll up Example

Roll-Up - Example1 (Cont.)

- In this example, cities Delhi and Patiala are rolled up into country India, and cities New York and Los Angeles are rolled up into USA.
- The medals figure of Delhi and Patiala are 5 and 3 respectively. They become 8 after roll-up. The medals figure of New York and Los Angeles are 2 and 5 respectively. They become 7 after roll-up.
- In this aggregation process, data in location hierarchy moves up from city to the country.

Roll-Up - Example2



Roll-Up - Example2 (Cont.)

- In this example, the cube contains the dimensions, location, and time and item, where the location is aggregated with regard to city values, time is aggregated with respect to quarters, and an item is aggregated with respect to item types.
- cities New Jersey and Lost Angles and rolled up into country USA
- The sales figure of New Jersey and Los Angeles are 440 and 1560 respectively. They become 2000 after roll-up
- In this aggregation process, data is location hierarchy moves up from city to the country.
- In the roll-up process at least one or more dimensions may need to be removed. In this example, Quarter dimension is removed.

Drill-Down

The drill-down operation (also called roll-down) is the reverse operation of roll-up. Drill-down is like zooming-in on the data cube. It navigates from less detailed record to more detailed data. Drill-down can be performed by:

- Moving down in the concept hierarchy
- Adding a new dimension

Drill-Down Example1

Drill-down on Location form countries to cities. i.e. less detailed data to more detailed data.

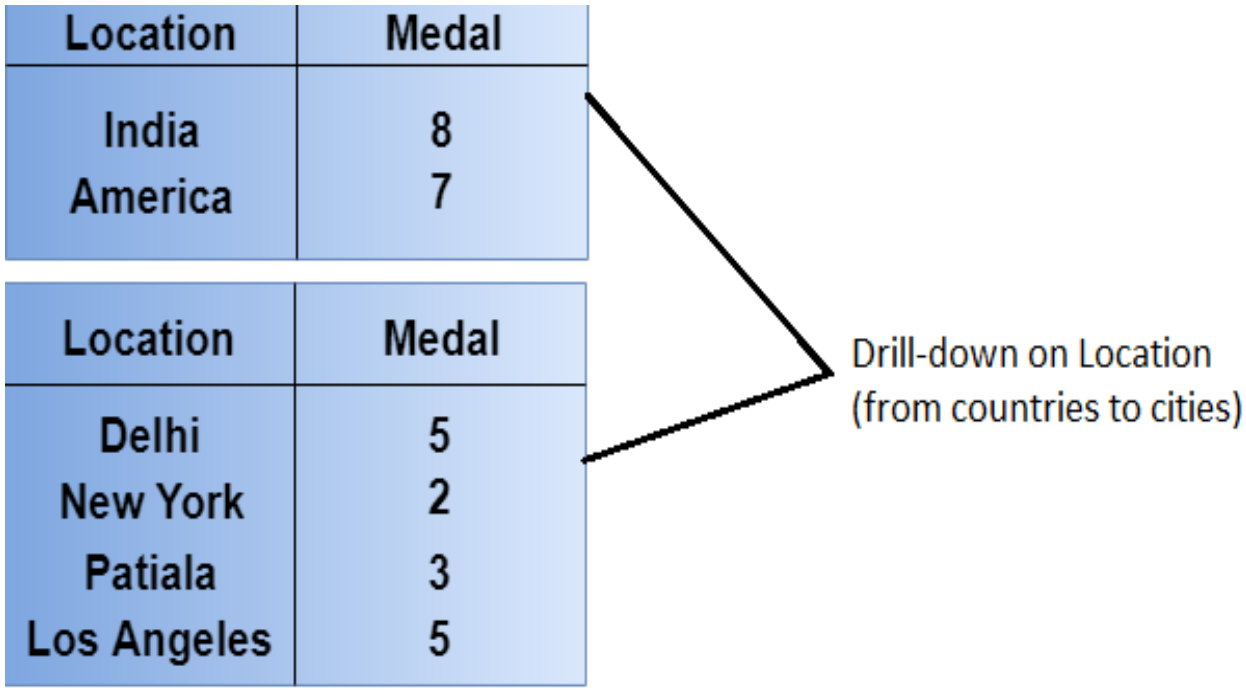
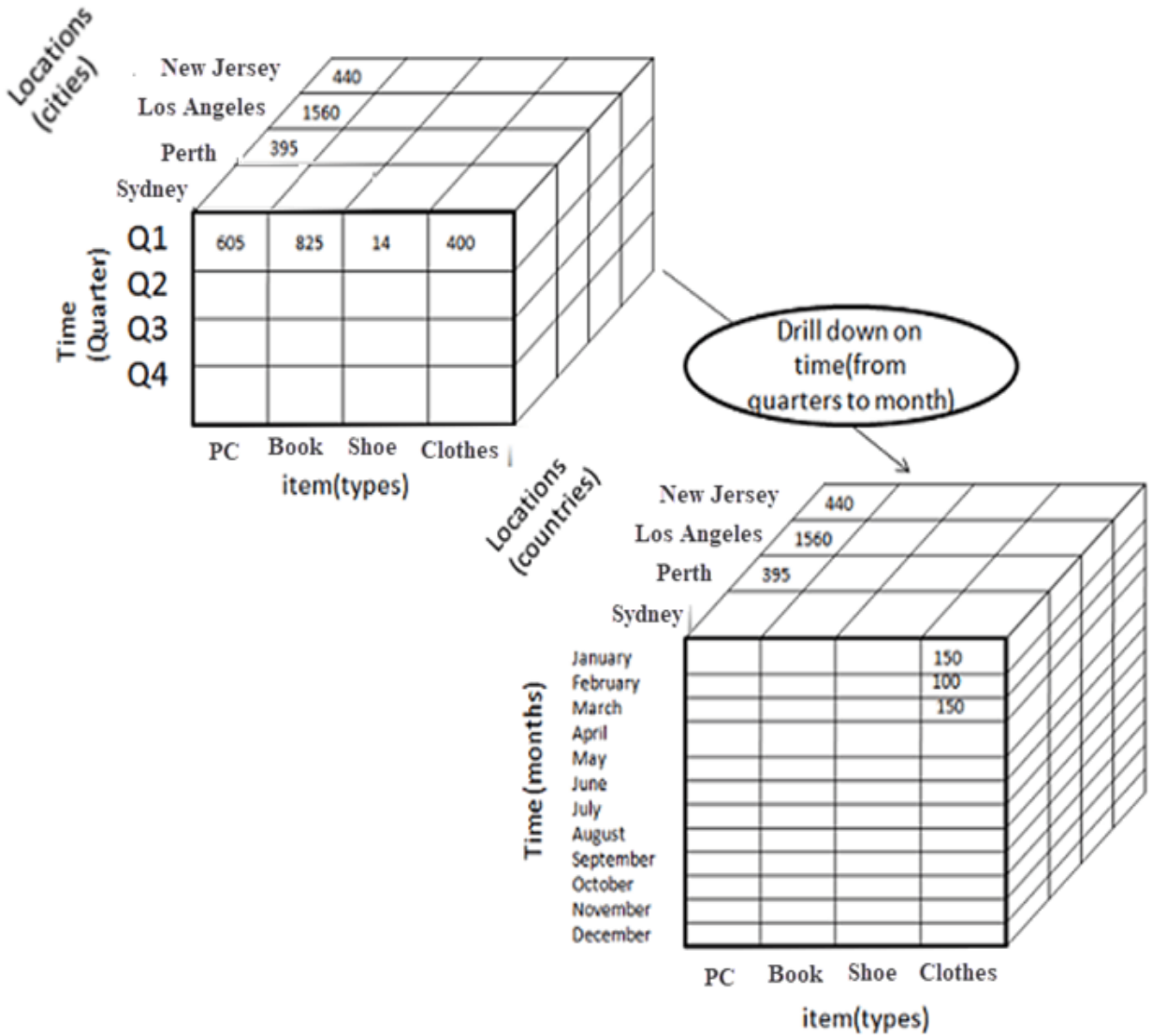


Figure: DrillDown Example

Drill-Down - Example2



Drill-Down - Example2 (Cont.)

In the previous figure:

- Quarter Q1 is drilled down to months January, February, and March. Corresponding sales are also registers.
- In this example, dimension months are added.

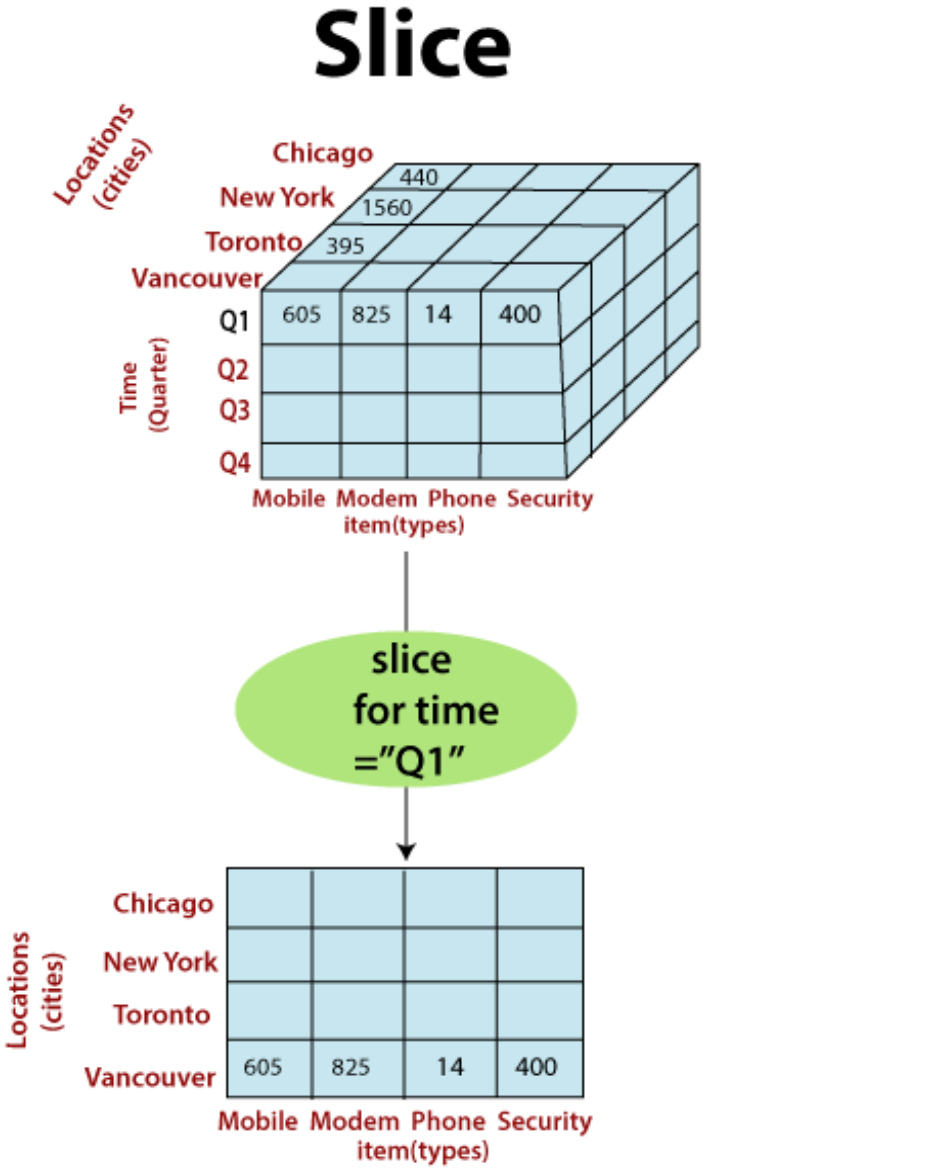
Slice

- The slice operation performs a selection on one dimension of the given cube, resulting in a subcube.Reduces the dimensionality of the cubes.
- For example, if we want to make a select where Medal = 5

Location	Medal
Delhi	5
Los Angeles	5

Figure: Example of Slice

Slice - Example2



Slice - Example2 (Cont.)

In the previous example:

- Dimension Time is Sliced with Q1 as the filter.
- A new cube is created altogether.

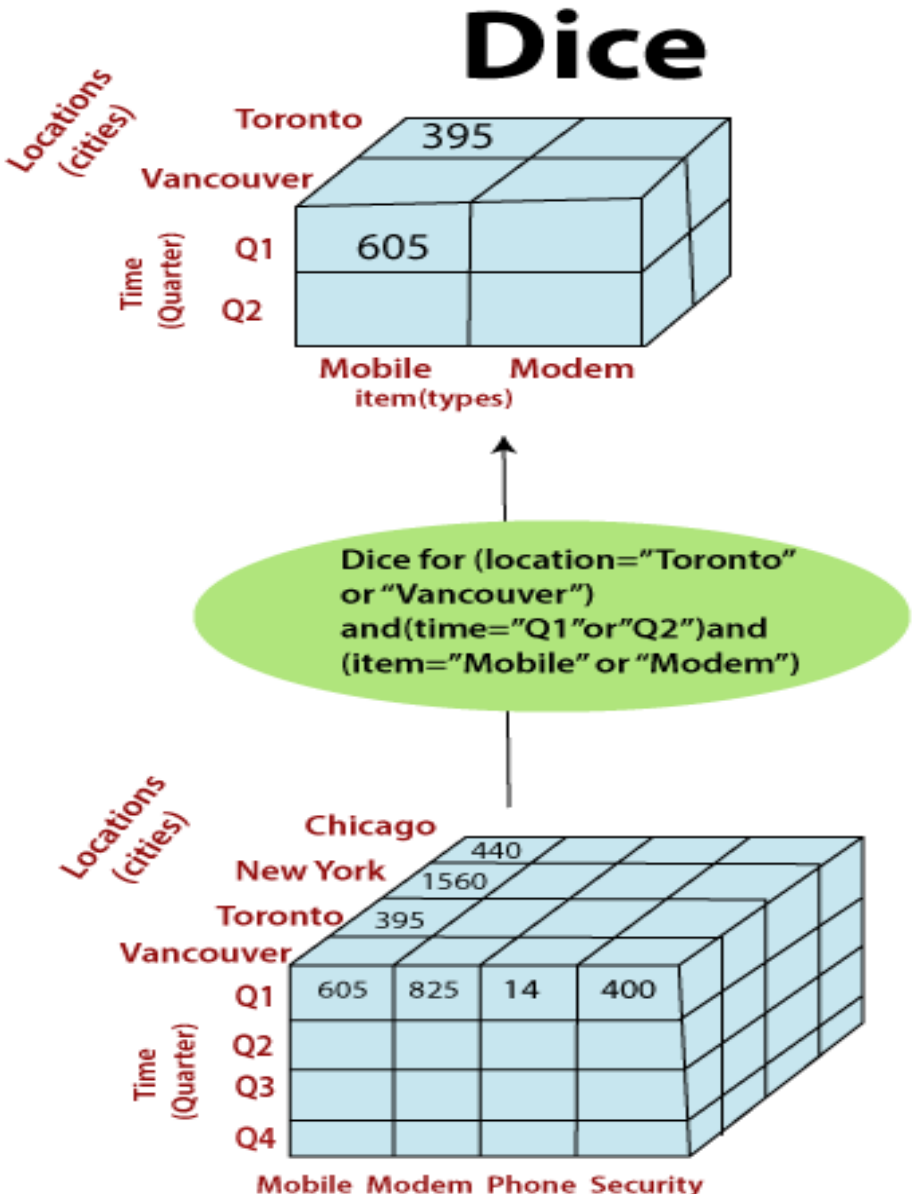
Dice

- The dice operation defines a sub-cube by performing a selection on two or more dimensions.
- For example, if we want to make a select where Medal = 3 or Location = New York

Location	Medal
Patiala	3
New York	2

Figure: Example of Dice

Dice - Example2



Dice - Example2 (Cont.)

In the previous example, 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.
- It may contain swapping the rows and columns or moving one of the row-dimensions into the column dimensions.

Pivot - General Example

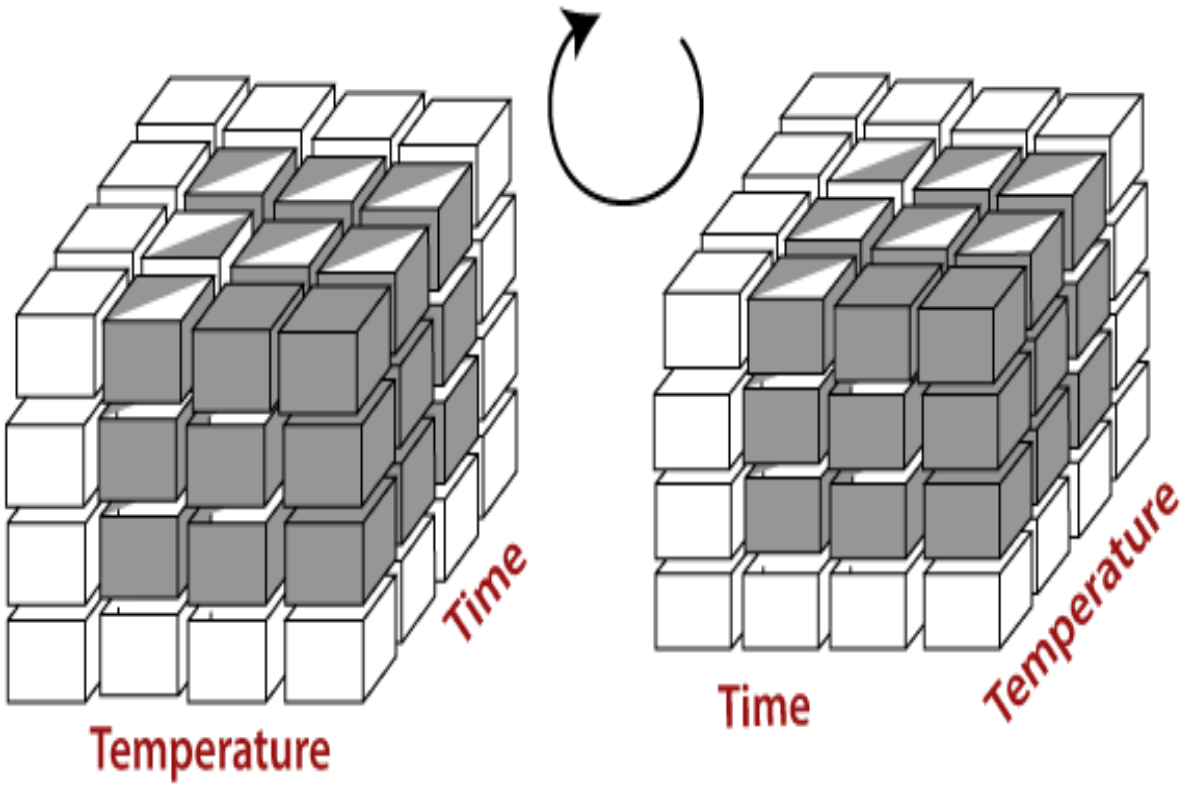
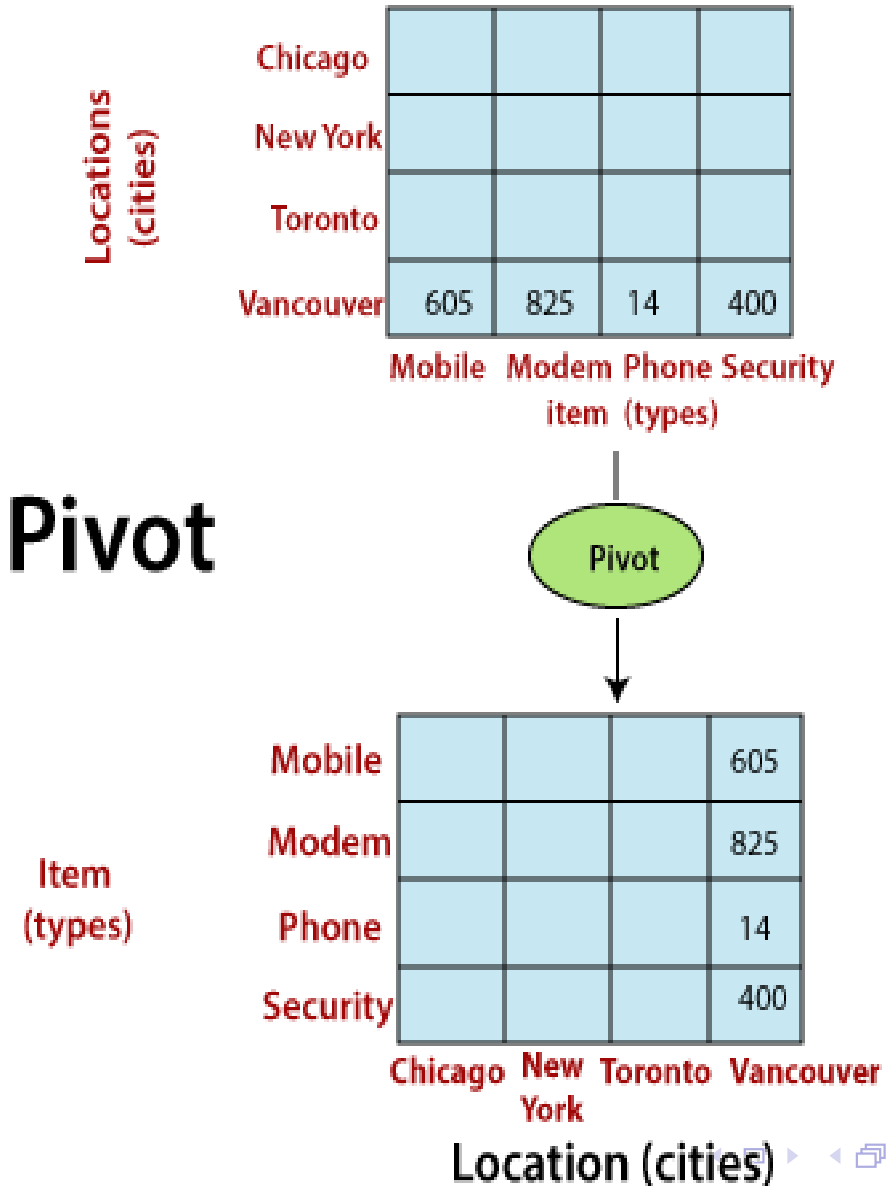
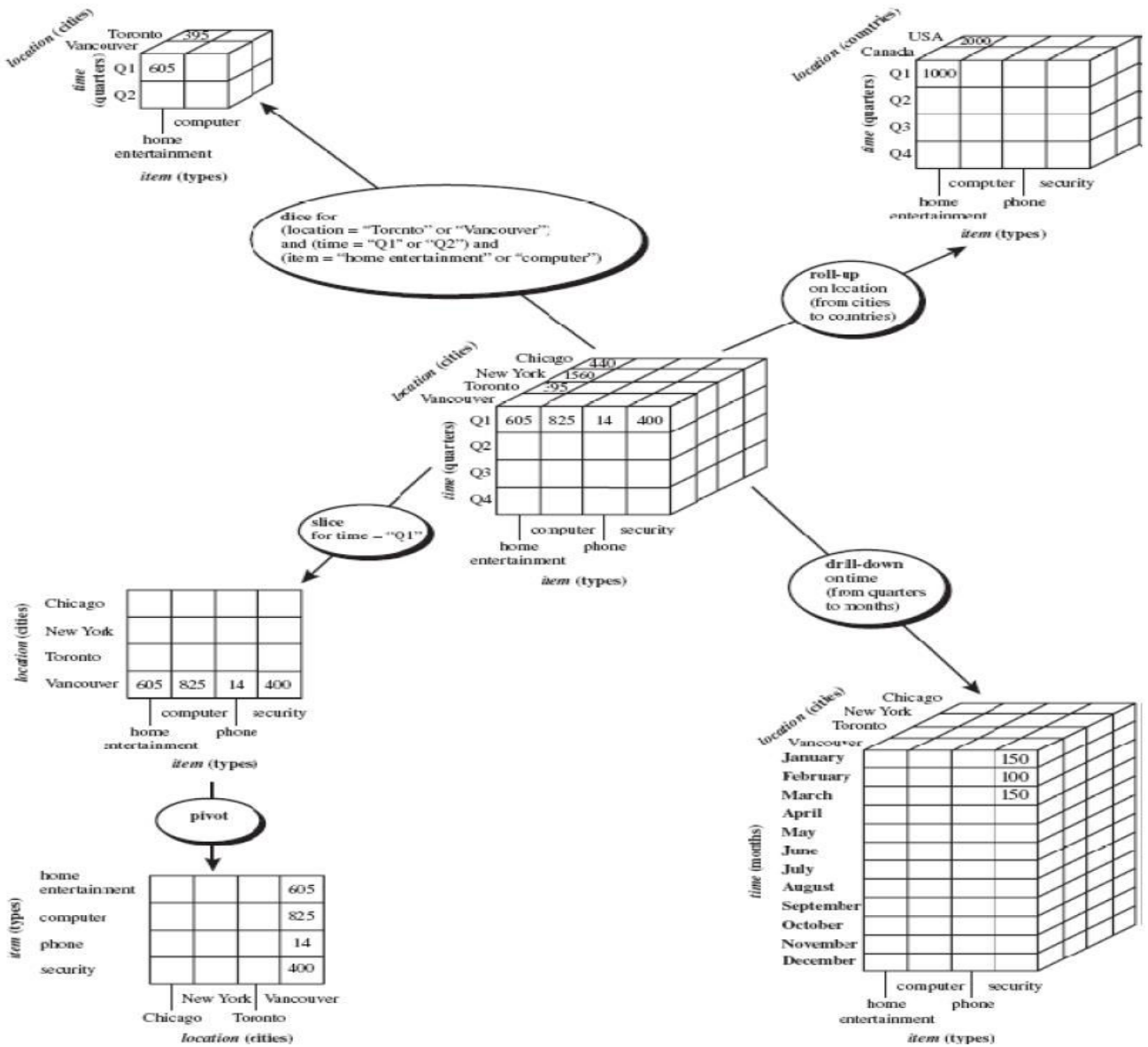


Figure: Illustration of Pivot

Pivot - Specific Example



Typical OLAP Operations - Summarized



Other OLAP Operations

- Executes queries containing more than one fact table.
- Other OLAP operations may contain ranking the top-N or bottom-N elements in lists, as well as calculate moving average, growth rates, and interests, internal rates of returns, depreciation, currency conversions, and statistical tasks.
- OLAP offers analytical modeling capabilities, containing a calculation engine for determining ratios, variance, etc. and for computing measures across various dimensions.
- It can generate summarization, aggregation, and hierarchies at each granularity level and at every dimensions intersection.
- OLAP also provide functional models for forecasting, trend analysis, and statistical analysis. In this context, the OLAP engine is a powerful data analysis tool.

Business Application of BI

We can break business applications of BI into two areas.

- Operational data use - how data helps run the business
- Strategic information use - which helps improve the business.

These both are valuable, and without the operational use of information a business could not survive.

Strategic Uses of information through BI Analytics

Customer analytics: Help a company better understand who their customers are and how to maximize the value of each customer in various aspects:

- **Customer profiling:** marketing tool that businesses use to understand their customers and helps to make better business decisions.
  - It involves a description of a customer or group of customers that includes demographic, geographic, and psychographic characteristics, as well as buying patterns, creditworthiness, and purchase history.
  - Demographic - the statistical characteristics of human populations (such as age or income) used especially to identify markets.
  - Psychographic - study of personality, values, opinions, attitudes, interests, and lifestyles.

Strategic Uses of information through BI Analytics...

- **Targeted marketing**
  - Identifies an audience likely to buy services or products and promotes those services or products to that audience.
  - Once these key groups are recognized, companies develop marketing campaigns and specific products for those preferred market segments.
- **Personalization**
  - As more business moves online, the browser acts as a proxy for the company's first interface with the customer.
  - the process of crafting a presentation to the customer based on that customer's profile, is the modern-day counterpart to the old-fashioned salesperson.
  - Web site personalization exploits customer profiles to dynamically collect content designed for an individual, and it is meant to enhance that customer's experience.

Strategic Uses of information through BI Analytics...

- **Collaborative Filtering**
  - A method of making automatic predictions about the interests of a user by collecting preferences or taste information from many users (collaborating). For example:
    - E-commerce Web sites that suggest alternate or additional purchases based on other peoples preferences. In other words, the information on a Web page may suggest that people who have purchased product X also have purchased product Y.
    - These kinds of suggestions are the result of a process called collaborative filtering, which evaluates the similarity between the preferences of groups of customers. This kind of recommendation generation creates relatively reliable cross-sell and up-sell opportunities.

Strategic Uses of information through BI Analytics...

- **Customer lifetime value:**
  - The lifetime value of a customer is a measure of a customers profitability over the lifetime of the relationship, which incorporates the costs associated with managing that relationship and the revenues expected from that customer.
  - Customer analytics incorporates metrics for measuring customer lifetime value.

Strategic Uses of information through BI Analytics...

Human capital productivity analytics:

- These are tools evaluating on-time performance, labor costs, production yield, etc., all as factors of how staff members work.
- This information can also be integrated into an information repository and analyzed for value.

Strategic Uses of information through BI Analytics...

Business productivity analytics:

- Assist to identify specific factors that affect the number of defective items produced, such as time of day, the source of raw materials used, and even the individuals who staff a production line. Helps companies to improve quality production.
- Can assist management in resource planning and staffing.

Strategic Uses of information through BI Analytics...

Supply Chain Analytics:

- Tools which contain techniques to find meaningful patterns and knowledge in order, shipment and transactional data.
- allow a company's management to track performance and reliability by supplier, evaluating and rating the quality of the products supplied.
- There are different methods by which a company delivers its products to its customers, each with its own cost schedule. Tools let management accurately determine the best way to move product.
- Helps to predict inventory requirements
- Helps a company to predict demand for specific products within certain areas, so that they can be stored in nearby warehouses.

Questions

- What are the most relevant differences between operational and decision support data?
- In a tabular form, give the differences between OLTP and OLAP.
- In a tabular form give the key differences between ROLAP and MOLAP.
- State the advantages and disadvantages of ROLAP and MOLAP.
- OLAP data is stored in multidimensional databases. Explain.
- What is data cleansing? What is its purpose?