REPORTING AND QUERY TOOLS AND APPLICATIONS

Tool Categories:

There are **five categories** of decision support tools

- Reporting
- Managed query
- Executive information system
- OLAP
- Data Mining

Reporting Tools

Production Reporting Tools

Companies generate regular operational reports or support high volume batch jobs, such as calculating and printing pay checks

Report writers

- Crystal Reports/Accurate reporting system
- User design and run reports without having to rely on the IS department

Managed query tools

- Managed query tools shield end user from the Complexities of SQL and database structures by inserting a meta layer between user and the database
- **Metalayer :**Software that provides subject oriented views of a database and supports point and click creation of SQL

Executive information system

- First deployed on main frame system
- Predate report writer and managed query tools
- Build customized, graphical decision support apps or briefing books
- Provides high level view of the business and access to external sources eg custom, on-line news feed
- EIS Apps highlight exceptions to business activity or rules by using color-coded graphics

OLAP Tools

- Provide an intuitive way to view corporate data
- Provide navigation through the hierarchies and dimensions with the single click
- Aggregate data along common business subjects or dimensions
- Users can drill down across, or up levels

Data mining Tools

- Provide insights into corporate data that are nor easily discerned with managed query or OLAP tools
- Use variety of statistical and AI algorithm to analyze the correlation of variables in data
- Interesting patterns and relationship to investigate

a)IBM's Intelligent Miner

b) DataMind Corp's DataMind

NEED FOR APPLICATIONS

Some tools and apps can format the retrieved data into easy-to-read reports, while others concentrate on the on-screen presentation

As the complexity of questions grows this tools may rapidly become Inefficient.

Consider various access types to the data stored in a data warehouse

- Simple tabular form reporting
- Ad hoc user specified queries
- Predefined repeatable queries
- Complex queries with multi table joins ,multilevel sub queries, and sophisticated search criteria
- Ranking
- Multivariable analysis
- Time series analysis
- Data visualization, graphing, charting and pivoting
- Complex textual search
- Statistical analysis
- Interactive Drill down reporting an analysis
- AI techniques for testing of hypothesis
- Information Mapping
- Interactive drill-down reporting and analysis

The first four types of access are covered by the combine category of tools we will call query and reporting tools.

There are three types of reporting

- **Creation and viewing of standard reports**
- **Definition and creation of ad hoc reports**
- **Data exploration**

COGNOUS IMPROMPTU

What is impromptu?

Impromptu is an interactive database reporting tool. It allows Power Users to query data without programming knowledge.

It is only capable of reading the data.

Impromptu's main features includes,

· **Interactive reporting capability**

· **Enterprise-wide scalability**

· **Superior user interface**

- **Fastest time to result**

- **Lowest cost of ownership**

Catalogs

Impromptu stores metadata in subject related folders..

A catalog contains:

- **Folders**—meaningful groups of information representing columns from one or more tables

- **Columns**—individual data elements that can appear in one or more folders

- **Calculations—expressions** used to compute required values from existing data

- **Conditions**—used to filter information so that only a certain type of information is displayed

- **Prompts**—pre-defined selection criteria prompts that users can include in reports they create

- **Other components,** such as metadata, a logical database name, join information, and user classes

You can use catalogs to

- **view, run, and print reports**

- **export reports to other applications**

- **disconnect from and connect to the database**

- **create reports**

- **change the contents of the catalog**

- **add user classes**

Prompts

You can use prompts to

- **filter reports**

- **calculate data items**

- **format data**

Picklist Prompts

A picklist prompt presents you with a list of data items from which you select one or more values, so you need not be familiar with the database.

The values listed in picklist prompts can be retrieved from

- **a database via a catalog when you want to select information that often changes.**

- **a column in another saved Impromptu report, a snapshot, or a HotFile**

A report can include a prompt that asks you to select a product type from a list of those available in the database. Only the products belonging to the product type you select are retrieved and displayed in your report.

Reports

Reports are created by choosing fields from the catalog folders. This process will build a SQL (Structured Query Language) statement behind the scene. No SQL knowledge is required to use Impromptu. The data in the report may be formatted, sorted and/or grouped as needed. Titles, dates, headers and footers and other standard text formatting features (italics, bolding, and font size) are also available.

Once the desired layout is obtained, the report can be saved to a report file.

This report can then be run at a different time, and the query will be sent to the database. It is also possible to save a report as a snapshot. This will provide a local copy of the data. This data will not be updated when the report is opened.

Cross tab reports, similar to Excel Pivot tables, are also easily created in Impromptu.

Frame-Based Reporting

Frames are the building blocks of all Impromptu reports and templates. They may contain report objects, such as data, text, pictures, and charts.

There are no limits to the number of frames that you can place within an individual report or template. You can nest frames within other frames to group report objects within a report.

Different types of frames and its purpose for creating frame based reporting

- **Form frame:** An empty form frame appears.

- **List frame:** An empty list frame appears.

- **Text frame:** The flashing I-beam appears where you can begin inserting text.

- **Picture frame:** The Source tab (Picture Properties dialog box) appears. You can use this tab to select the image to include in the frame.

- **Chart frame:** The Data tab (Chart Properties dialog box) appears. You can use this tab to select the data item to include in the chart.

- **OLE Object:** The Insert Object dialog box appears where you can locate and select the file you want to insert, or you can create a new object using the software listed in the Object Type box.

Impromptu Features

- **Unified query and reporting interface:** It unifies both query and reporting interface in a single user interface

- **Object oriented architecture:** It enables an inheritance based administration so that more than 1000 users can be accommodated as easily as single user

- **Complete integration with Power Play:** It provides an integrated solution for exploring trends and patterns

- **Scalability:** Its scalability ranges from single user to 1000 user

· **Security and Control:** Security is based on user profiles and their classes.

· **Data presented in a business context:** It presents information using the terminology of the business.

· **Over 70 pre defined report templates:** It allows users can simply supply the data to create an interactive report

· **Frame based reporting:** It offers number of objects to create a user designed report

· **Business relevant reporting:** It can be used to generate a business relevant report through filters, pre conditions and calculations

· **Database independent catalogs:** Since catalogs are in independent nature they require minimum maintenance

OLAP

OLAP stands for Online Analytical Processing.

It uses database tables (fact and dimension tables) to enable multidimensional viewing, analysis and querying of large amounts of data.

E.g. OLAP technology could provide management with fast answers to complex queries on their operational data or enable them to analyze their company's historical data for trends and patterns.

Online Analytical Processing (OLAP) applications and tools are those that are designed to ask —complex queries of large multidimensional collections of data.‖ Due to that OLAP is accompanied with data warehousing.

NEED

- The key driver of OLAP is the multidimensional nature of the business problem.

- These problems are characterized by retrieving a very large number of records that can reach gigabytes and terabytes and summarizing this data into a form information that can by used by business analysts.

- One of the limitations that SQL has, it cannot represent these complex problems.

- A query will be translated in to several SQL statements. These SQL statements will involve multiple joins, intermediate tables, sorting, aggregations and a huge temporary memory to store these tables.

- These procedures required a lot of computation which will require a long time in computing.

- The second limitation of SQL is its inability to use mathematical models in these SQL statements. If an analyst, could create these complex statements using SQL statements, still there will be a large number of computation and huge memory needed.

- Therefore the use of OLAP is preferable to solve this kind of problem.

MULTIDIMENSIONAL DATA MODEL

The multidimensional data model is an integral part of On-Line Analytical Processing, or OLAP

Multidimensional data model is to view it as a cube. The cable at the left contains detailed sales data by product, market and time. The cube on the right associates sales number (unit sold) with dimensions-product type, market and time with the unit variables organized as cell in an array.

This cube can be expended to include another array-price-which can be associates with all or only some dimensions. As number of dimensions increases number of cubes cell increase exponentially.

Dimensions are hierarchical in nature i.e. time dimension may contain hierarchies for years, quarters, months, weak and day. GEOGRAPHY may contain country, state, city etc.
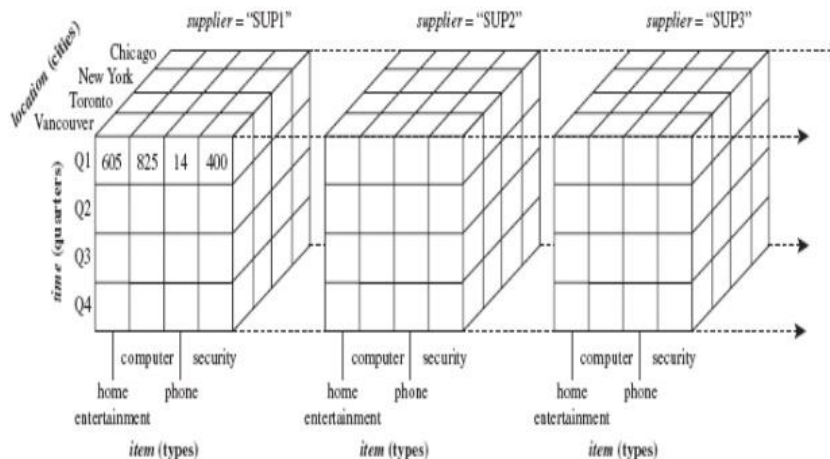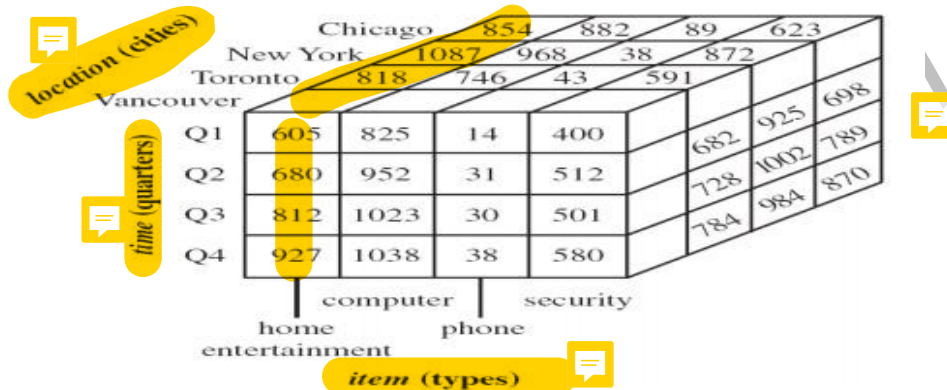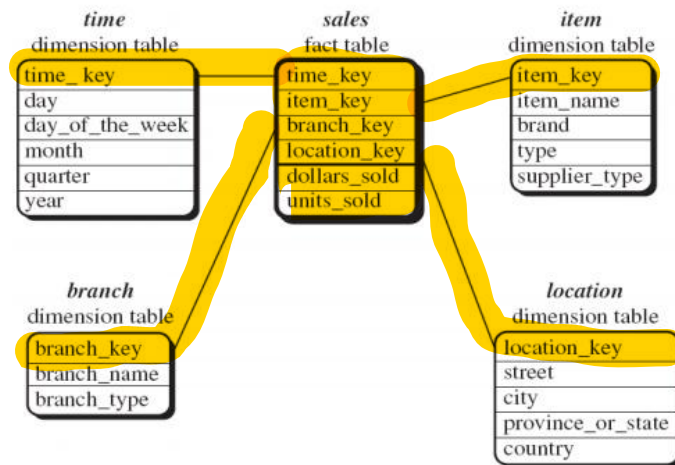

Star schema: A fact table in the middle connected to a set of dimension tables

It contains:

- A large central table (fact table)

- A set of smaller attendant tables (dimension table), one for each dimension

Snowflake schema: A refinement of star schema where some dimensional hierarchy is further splitting (normalized) into a set of smaller dimension tables, forming a shape similar to snowflake

However, the snowflake structure can reduce the effectiveness of browsing, since more joins will be needed
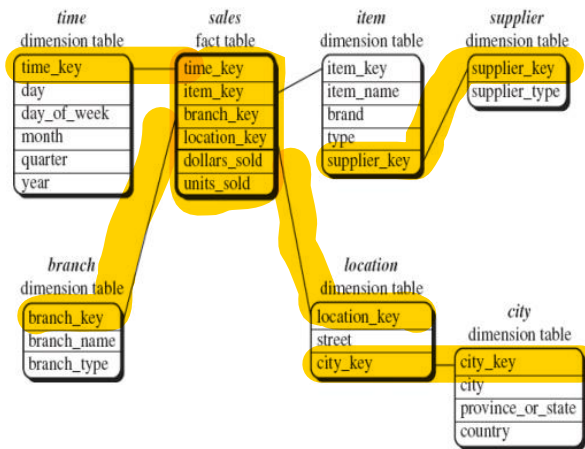
Fact constellations: Multiple fact tables share dimension tables, viewed as a collection of stars, therefore called galaxy schema or fact constellation.
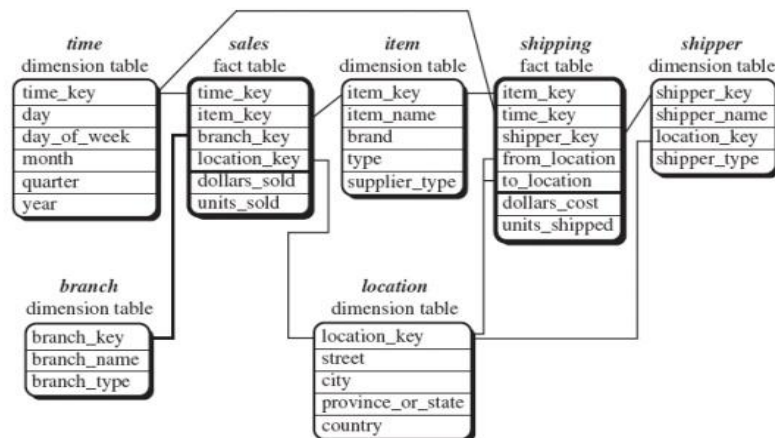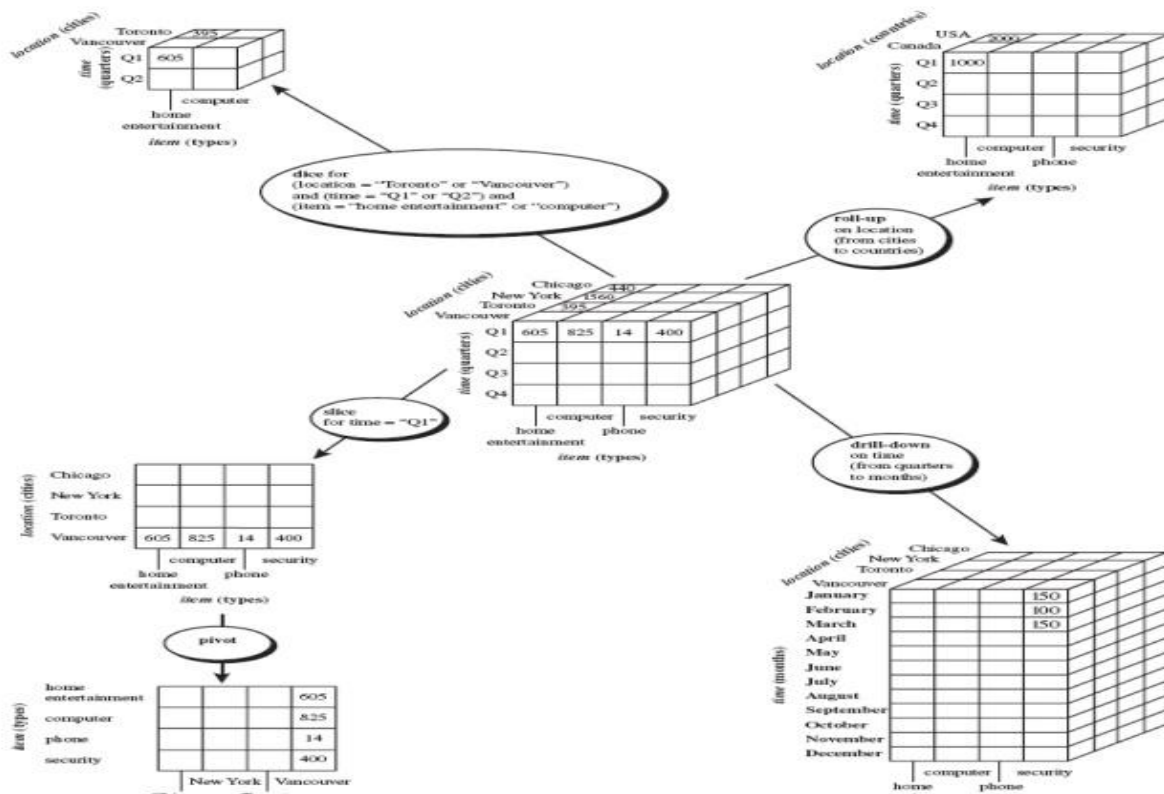
In this cube we can observe, that each side of the cube represents one of the elements of the question. The x-axis represents the time, the y-axis represents the products and the z-axis represents different centers. The cells of in the cube represents the number of product sold or can represent the price of the items.

This Figure also gives a different understanding to the drilling down operations. The relations defined must not be directly related, they related directly.

The size of the dimension increase, the size of the cube will also increase exponentially. The time response of the cube depends on the size of the cube.

OPERATIONS IN MULTIDIMENSIONAL DATA MODEL:

Aggregation (*roll-up*)

Dimension reduction: e.g., total sales by city

Summarization over aggregate hierarchy: e.g., total sales by city and year -> total sales by region and by year

Selection (*slice*) defines a subcube

e.g., sales where city = Palo Alto and date = 1/15/96

Navigation to detailed data (*drill-down*)

e.g., (sales – expense) by city, top 3% of cities by average income

Visualization Operations (e.g., Pivot or dice)

OLAP GUIDELINES

Dr. E.F. Codd the —father of the relational model, created a list of rules to deal with the OLAP systems. Users should priorities these rules according to their needs to match their business requirements.

These rules are:

1. Multidimensional conceptual view: The OLAP should provide an appropriate multidimensional Business model that suits the Business problems and Requirements.

2. Transparency: The OLAP tool should provide transparency to the input data for the users.

3. Accessibility: The OLAP tool should only access the data required only to the analysis needed.

4. Consistent reporting performance: The Size of the database should not affect in any way the performance.

5. Client/server architecture: The OLAP tool should use the client server architecture to ensure better performance and flexibility.

6. Generic dimensionality: Data entered should be equivalent to the structure and operation requirements.

7. Dynamic sparse matrix handling: The OLAP too should be able to manage the sparse matrix and so maintain the level of performance.

8. Multi-user support: The OLAP should allow several users working concurrently to work together.

9. Unrestricted cross-dimensional operations: The OLAP tool should be able to perform operations across the dimensions of the cube.

10. Intuitive data manipulation: Consolidation path re-orientation, drilling down across columns or rows, zooming out, and other manipulation inherent in the consolidation path outlines should be accomplished via direct action upon the cells of the analytical model, and should neither require the use of a menu nor multiple trips across the user interface.‖(Reference 4)

11. Flexible reporting: It is the ability of the tool to present the rows and column in a manner suitable to be analyzed.

12. Unlimited dimensions and aggregation levels: This depends on the kind of Business, where multiple dimensions and defining hierarchies can be made.

In addition to these guidelines an OLAP system should also support:

Comprehensive database management tools: This gives the database management to control distributed Businesses

The ability to drill down to detail source record level: Which requires that The OLAP tool should allow smooth transitions in the multidimensional database.

Incremental database refresh: The OLAP tool should provide partial refresh.

Structured Query Language (SQL interface): the OLAP system should be able to integrate effectively in the surrounding enterprise environment.

OLTP vs OLAP

OLTP stands for On Line Transaction Processing and is a data modeling approach typically used to facilitate and manage usual business applications. Most of applications you see and use are OLTP based. OLTP technology used to perform updates on operational or transactional systems (e.g., point of sale systems) 💬

OLAP stands for On Line Analytic Processing and is an approach to answer multi-dimensional queries. OLAP was conceived for Management Information Systems and Decision Support Systems. OLAP technology used to perform complex analysis of the data in a data warehouse. 💬

CATEGORIES OF OLAP TOOLS

MOLAP 💬

This is the more traditional way of OLAP analysis. In MOLAP, data is stored in a multidimensional cube. The storage is not in the relational database, but in proprietary formats. That is, data stored in array-based structures.

Advantages:

· **Excellent performance**: MOLAP cubes are built for fast data retrieval, and are optimal for slicing and dicing operations.

· **Can perform complex calculations**: All calculations have been pre-generated when the cube is created. Hence, complex calculations are not only doable, but they return quickly.

Disadvantages:

· **Limited in the amount of data it can handle:** Because all calculations are performed when the cube is built, it is not possible to include a large amount of data in the cube itself. This is not to say that the data in the cube cannot be derived from a large amount of data. Indeed, this is possible. But in this case, only summary-level information will be included in the cube itself.

· **Requires additional investment**: Cube technology are often proprietary and do not already exist in the organization. Therefore, to adopt MOLAP technology, chances are additional investments in human and capital resources are needed.
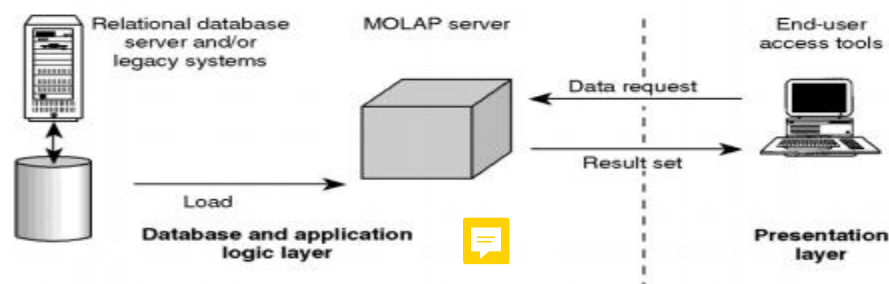

Figure Describes The Relation of the MOLAP with the server and end user.

Examples: Hyperion Essbase, Fusion (Information Builders)

ROLAP

This methodology relies on manipulating the data stored in the relational database to give the appearance of traditional OLAP's slicing and dicing functionality. In essence, each action of slicing and dicing is equivalent to adding a —WHERE‖ clause in the SQL statement. Data stored in relational tables

---

Advantages:

·   **Can handle large amounts of data:** The data size limitation of ROLAP technology is the limitation on data size of the underlying relational database. In other words, ROLAP itself places no limitation on data amount.

·   **Can leverage functionalities inherent in the relational database**: Often, relational database already comes with a host of functionalities. ROLAP technologies, since they sit on top of the relational database, can therefore leverage these functionalities.

Disadvantages:

·   **Performance can be slow:** Because each ROLAP report is essentially a SQL query (or multiple SQL queries) in the relational database, the query time can be long if the underlying data size is large.

·   **Limited by SQL functionalities**: Because ROLAP technology mainly relies on generating SQL statements to query the relational database, and SQL statements do not fit all needs (for example, it is difficult to perform complex calculations using SQL), ROLAP technologies are therefore traditionally limited by what SQL can do. ROLAP vendors have mitigated this risk by building into the tool out-of-the-box complex functions as well as the ability to allow users to define their own functions.
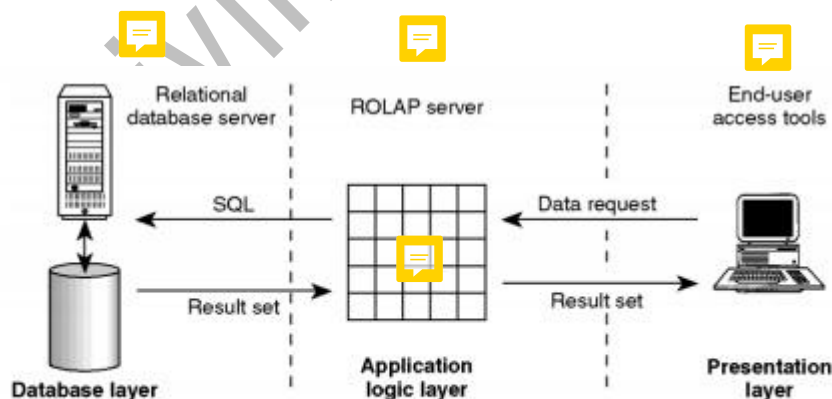

Figure describes The Relation of the ROLAP with the server and end user.

Examples: Microstrategy Intelligence Server, MetaCube (Informix/IBM)

HOLAP (MQE: Managed Query Environment) 💬

HOLAP technologies attempt to combine the advantages of MOLAP and ROLAP. For summary-type information, HOLAP leverages cube technology for faster performance. It stores only the indexes and aggregations in the multidimensional form while the rest of the data is stored in the relational database.

OLAP TOOLS AND THE INTERNET

The mainly comprehensive premises in computing have been the internet and data warehousing thus the integration of these two giant technologies is a necessity. The advantages of using the Web for access are inevitable. 💬

These advantages are:

· The internet provides connectivity between countries acting as a free resource.

· The web eases administrative tasks of managing scattered locations.

· The Web allows users to store and manage data and applications on servers that can be managed, maintained and updated centrally.

These reasons indicate the importance of the Web in data storage and manipulation.

The Web-enabled data access has many significant features, such as:

· The first

· The second

· The emerging third

· HTML publishing

· Helper applications

· Plug-ins

· Server-centric components

· Java and active-x applications

The primary key in the decision making process is the amount of data collected and how well this data is interpreted. Nowadays, Managers aren't satisfied by getting direct answers to their direct questions, Instead due to the market growth and increase of clients their questions became more complicated. Questions are likeǁ How much profit from selling our products at our different centers per monthǁ. A complicated question like this isn't as simple to be answered directly; it needs analysis to three fields in order to obtain an answer.

The Decision making processes

1. **Identify information about the problem**

2. **Analyze the problem**

3. **Collect and evaluate the information, and define alternative solutions.**

The decision making process exists in the different levels of an organization. The speed and the simplicity of gathering data and the ability to convert this data into information is the main element in the decision process. That's why the term Business Intelligence has evolved.

Business Intelligence

As mentioned Earlier, business Intelligence is concerned with gathering the data and converting this data into information, so as to use a better decision. The better the data is gathered and how well it is interpreted as information is one of the most important elements in a successful business.

Elements of Business Intelligence

There are three main Components in Business Intelligence

Data Warehouse: it is a collection of data to support the management decisions making. It revolves around the major subjects of the business to support the management.

OLAP: is used to generate complex queries of multidimensional collection of data from the data warehouse.

Data Mining: consists of various techniques that explore and bring complex relationships in very large sets.

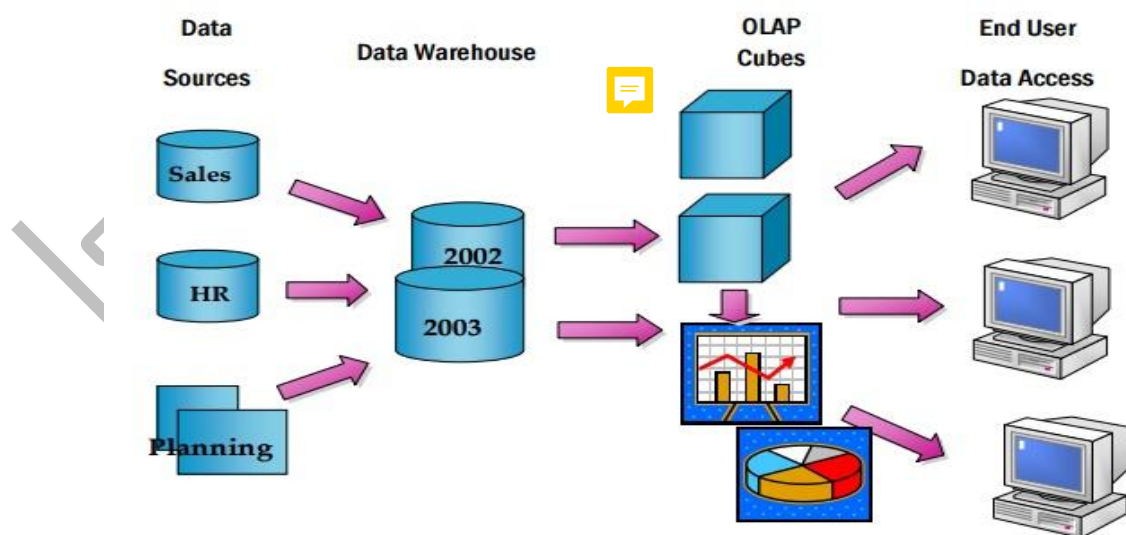In the next figure the relation between the three components are represented.

Figure (1) Shows the Connection between Elements of business Intelligence.