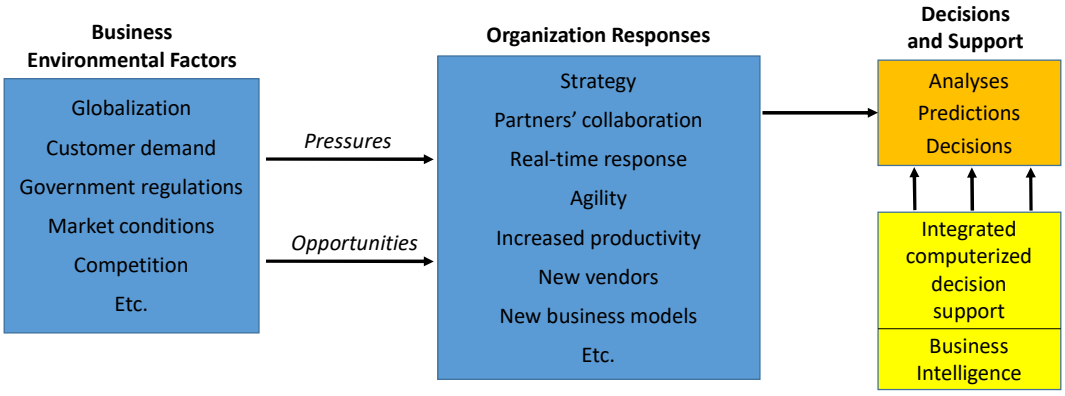
Chapter Name

Making Sense of Data for Better Decisions

The Business *Pressures-Responses-Support* Model

Companies are moving aggressively to computerized support of their operations not only on operational but also on decisional level. Many managerial actions require computerized decision support (DSS).



Business pressures results from today's business climate, responses are actions taken by companies to counter the pressures (or to take advantages of the opportunities) and computerized support facilitates the monitoring of the environment and enhances the response actions taken by organizations

Managerial Decision Making

- Managers usually make decisions by following a four-step process:
  1. Define the problem (i.e., a decision situation that may deal with some difficulty or with an opportunity).
  2. Construct a model that describes a real-world problem.
  3. Identify possible solutions to the modelled problem and evaluate the solutions.
  4. Compare, choose and recommend a potential solution to the problem.
- To follow this process one must make sure that:
  - sufficient alternatives are being considered
  - The consequences of using these alternatives can be reasonably predicted, and
  - comparisons are done properly.
- However a number of environmental factors make such an evaluation process difficult.

The Need for Decision Support Systems

- Environments are groving more complex everyday, therefore making decisions today is indeed a complex task.
- It is nearly impossible to rely on a trial-and-error approach to management for the following reasons:
  - Technology, information systems, advanced search engines, and globalization result in more and more alternative from which to choose.
  - Government regulations and the need for compliance political instability and terrorism, competition, and changing consumer demands produce more uncertainty, making it more difficult to predict consequences and the future.
  - There is a need to make rapid decisions.
  - Making mistakes may be very expensive.
- Managers must be more sophisticated; they must use the new techniques and tools to support decision making.

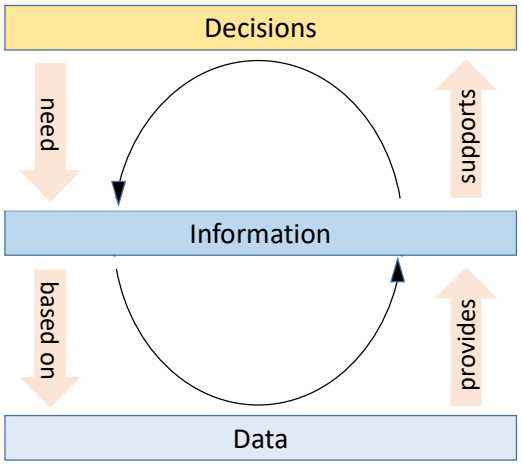
The Concept of Decision Support Systems

- In the early 1970s, Scott-Morton first articulated the major concepts of DSS.
  - He defined decision support systems (DSS) as „interactive computer-based systems, which help decision makers utilize data and models to solve unstructured problems" (Gory and Scott-Morton, 1971).
- The following is another classic DSS definition, provided by Keen and Scott-Morton (1978):
  - Decision support systems couple the intellectual resources of individuals with the capabilities of the computer to improve the quality of decisions. It is a computer-based support system for management decision makers who deal with semistructured problems.
- Note, that the term decision support system, like management information system (MIS) and other terms in the field of IT is a content free expression (i.e., it means different things to different people).
  - Therefore there is no universally accepted definition of DSS.
- Actually, DSS can be viewed as a conceptual methodology - that is a broad umbrella term.
  - However, some view DSS as a narrower, specific decision support application.

DSS as an Umbrella Term

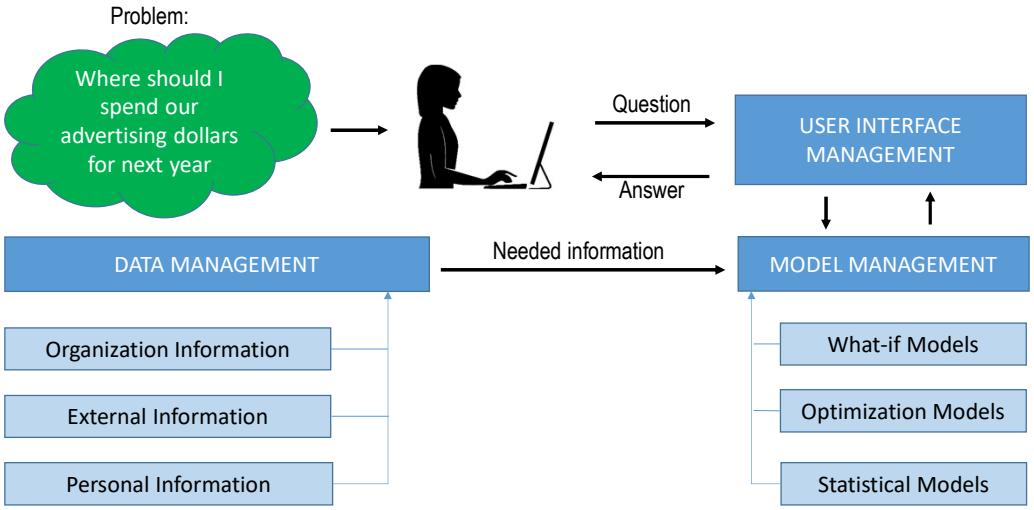
- The term DSS can be used as an umbrella term to describe any computerized system that supports decision making in an organization.
  - An organization may have a knowledge management system to guide all its personnel in their problem solving.
  - Another organization may have separate support systems for marketing, finance, and accounting; a supply chain management (SCM) system for production; and several rule-based systems for product repair diagnostics and help desks.
- DSS encompasses them all.

From Data to Decisions

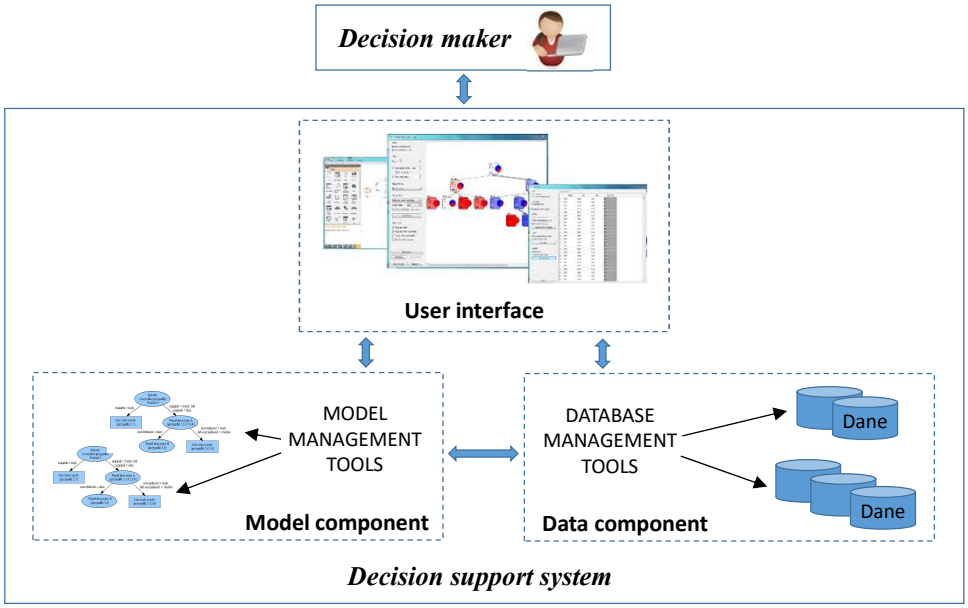


- The new techniques and tools to support decision making are focused on gaining profit from the available data.
- Making decisions requires information that is based on data. Data provides the information to support decisions, and so on.
  - Data in itself provides no judgment or interpretation and no basis of action.
  - The context and use of data turns it into information.
  - Connecting pieces of available information leads to knowledge that can then support decisions.

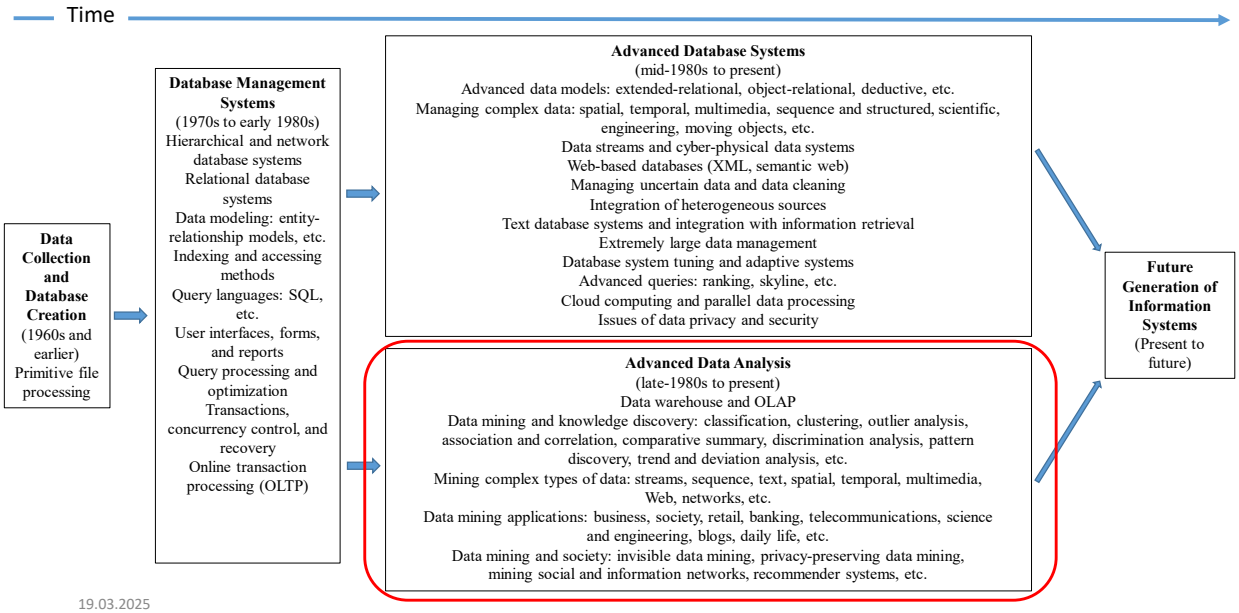
Using a Decision Support System



Components of a Decision Support System

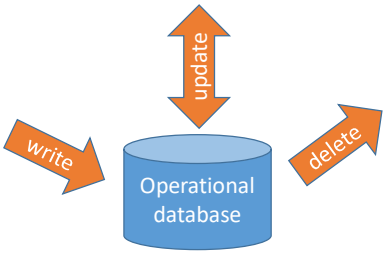


The evolution of database system technology



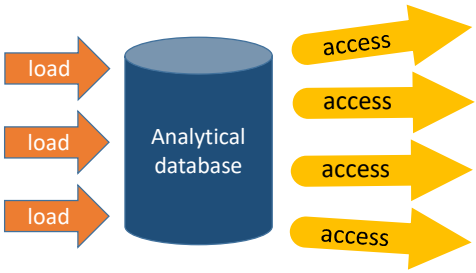
Operational vs. Analytical Databases

- Operational (e. g. relational) databases support everyday activity of a company.



Data is regularly updated on a record by record basis.

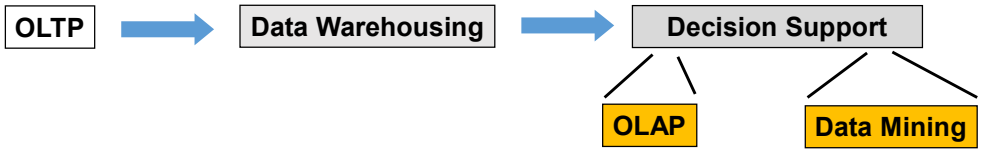
- Analytical or informational databases (e. g. data warehouses) provide an information which can be used to analyze the problems and situations.



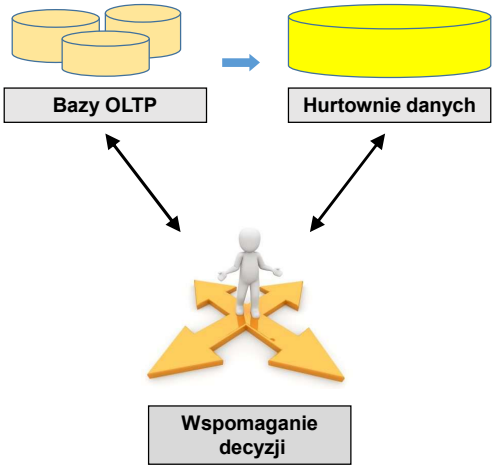
Data is loaded into the data warehouse and is accessed there but is NOT updated.

Data Driven Decision Support

- Historically, Relational Database Management Systems (RDBMSs) have been focused on the field of Online Transaction Processing (OLTP).
- Data Warehousing technology provided new possibilities to build data marts and data warehouses for Decision Support Systems.
- With OLAP and data mining features, we will be able to use the data accumulated in the OLTP databases to analyze, explain and even "predict" future data, e.g. consumer preferences, "what-if" analysis, etc.



Wspomaganie decyzji na podstawie analizy danych



- Tradycyjne, relacyjne systemy zarządzania bazami danych koncentrowały się na przetwarzaniu transakcji (*OnLine Transaction Processing* - OLTP)
- Wraz z rozwojem systemów informacyjnych pojawiła się technologia hurtowni danych (*data warehousing*) otwierając nowe możliwości w zakresie zaawansowanej analizy danych.
- Nowoczesne metody analizowania danych gromadzonych w bazach OLTP i innych repozytoriach umożliwiają m.in. uzyskiwanie podsumowań, dokonywanie predykcji oraz odkrywanie wzorców z przeznaczeniem dla celów wspomagania decyzji.

19.03.2025

Data Analysis Techniques

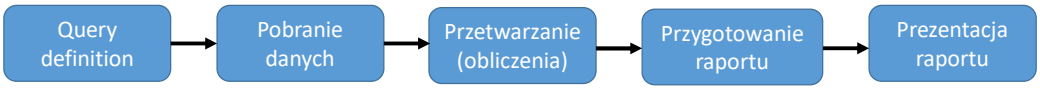
- A data warehouse is built to provide an easy to access source of high quality data.
  - It is a means to an end, not the end itself. That end is typically the need to perform analysis and decision making through the use of that source of data.
- There are several techniques for data analysis that are in common use.
  - query and reporting (formulate and display query results),
  - multidimensional analysis (analyze data content by viewing it from different perspectives), and
  - data mining (discover patterns and clustering attributes in the data that will provide further insight into the data content).

Query and Reporting

☞ Query and reporting analysis is the process of

- posing a question to be answered,
- retrieving relevant data from the data warehouse,
- transforming it into the appropriate context, and
- displaying it in a readable format.

The Process of Query and Reporting

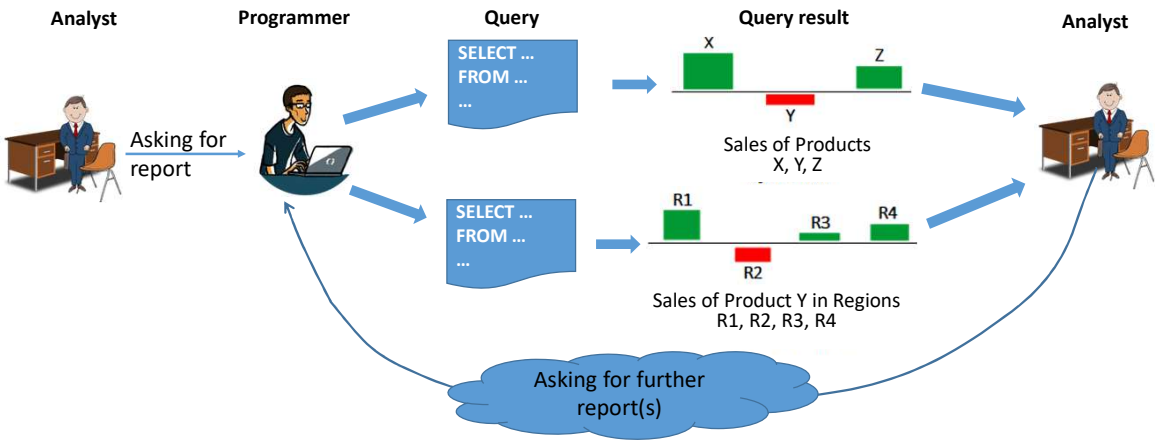


☞ It is driven by analysts who must pose those questions to receive an answer.

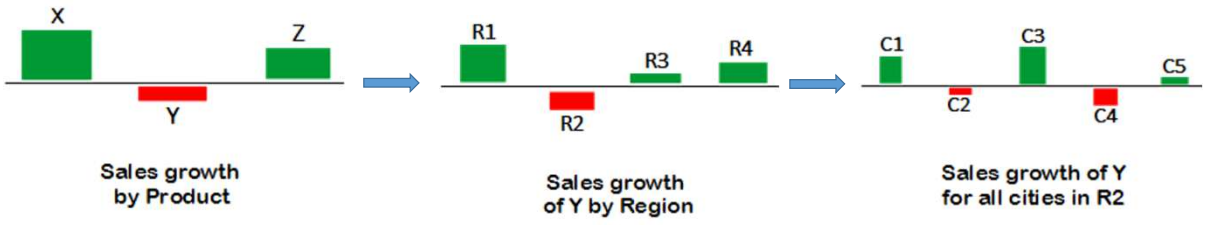
- You will find that this is quite different, for example, from data mining, which is data driven.

Query and Reporting for Data Analysis

End users are primarily interested in processing numeric values, which they use to analyze the behavior of business processes, such as sales revenue and shipment quantities.

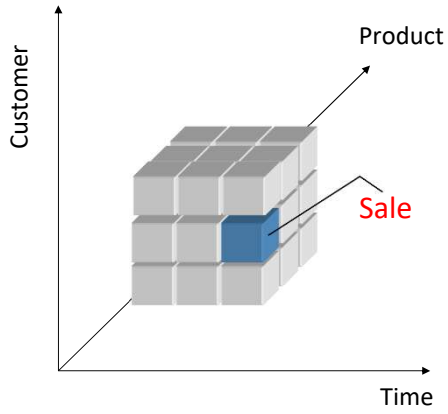


OLAP – Intuitively Explained



Using an OLAP tool the analyst can browse his data on multiple levels and dimensions in order to explain system's behavior.

Multidimensional Data Organization – Data Cubes



- Modeling data multi-dimensionally is a way to facilitate on-line business analysis and query performance.
- Using OLAP tool you can browse your data on multiple levels and dimensions in order to explain system's behavior

Note: Cubes are not limited to three dimensions.

Data Mining (1/2)

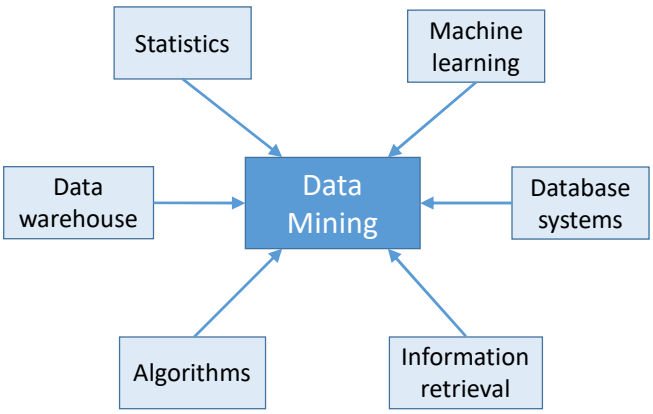
- Data mining is a relatively new data analysis technique which is very different from query and reporting and multidimensional analysis.
  - You do not ask a particular question of the data but rather use specific algorithms that analyze the data and report what they have discovered.
- Unlike query and reporting and multidimensional analysis where the user has to create and execute queries based on hypotheses, data mining searches for answers to questions that may have not been previously asked.
- The discovery could take the form of
  - finding significance in relationships between certain data elements,
  - clustering together of specific data elements, or
  - other patterns in the usage of specific sets of data elements.
- After finding these patterns, the algorithms can infer rules. These rules can then be used to generate a model that can predict a desired behavior, identify relationships among the data, discover patterns, and group clusters of records with similar attributes.

Data Mining (2/2)

- In industry, in media, and in the research milieu, the term data mining is often used to refer to the entire knowledge discovery process (perhaps because the term is shorter than knowledge discovery from data).
  - Therefore, we adopt a broad view of data mining functionality: Data mining is the process of discovering interesting patterns and knowledge from large amounts of data.
  - As a knowledge discovery process, it typically involves data cleaning, data integration, data selection, data transformation, pattern discovery, pattern evaluation, and knowledge presentation.
  - The data sources can include databases, data warehouses, the Web, other information repositories, or data that are streamed into the system dynamically.
- Data mining has many successful applications, such as business intelligence, Web search, bioinformatics, health informatics, finance, digital libraries, and digital governments.

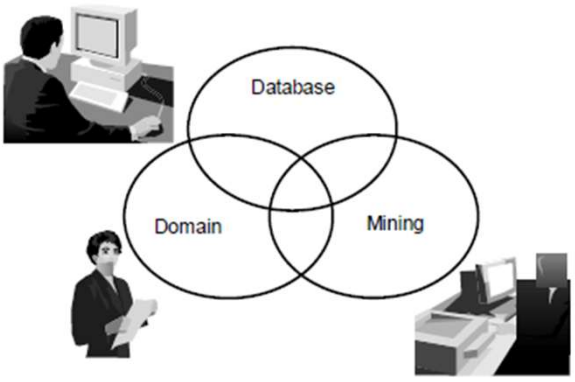
Which Technologies Are Used?

- Data mining, as a highly application-driven domain, has incorporated technologies from many other domains.
  - These include statistics, machine learning, database and data warehouse systems, information retrieval, and others.
- The interdisciplinary nature of data mining research and development contributes significantly to the success of data mining and its extensive applications.



19.03.2025

Roles in Data Mining



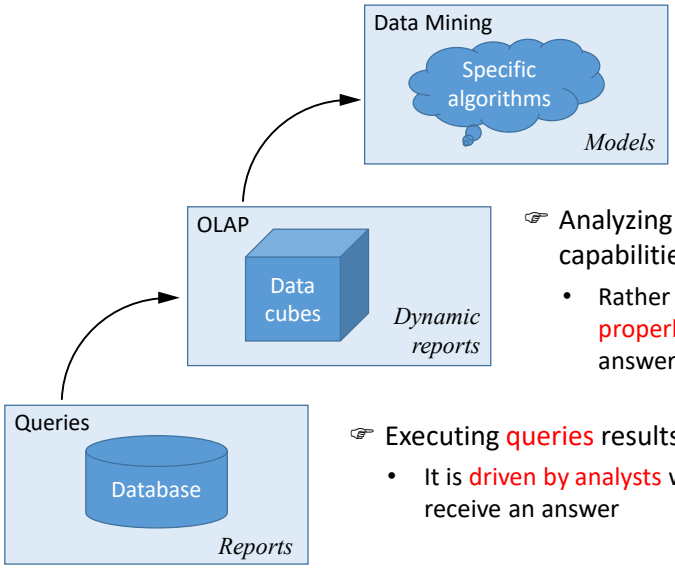
- Database administrators know where and how the company's data is stored, how to access it, and how to relate it to other data stores.
- Domain experts know the business environment, the processes, the customers, and the competitors.
- Mining specialists are the people with a background in data analysis who have at least basic statistical knowledge. They are able to apply data mining techniques and interpret the results in a technical way.

19.03.2025

Data Mining Functionalities

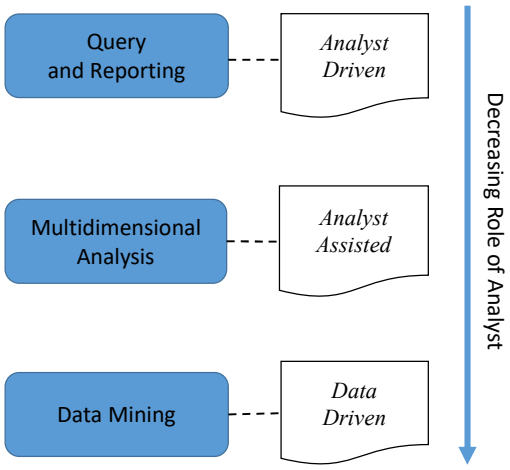
- Data mining functionalities are used to specify the kinds of patterns or knowledge to be found in data mining tasks. The functionalities include
  - characterization and discrimination;
  - the mining of frequent patterns, associations, and correlations;
  - classification and regression;
  - cluster analysis; and
  - outlier detection.
- As new types of data, new applications, and new analysis demands continue to emerge, there is no doubt we will see more and more novel data mining tasks in the future.

Evolution from Queries to Data Mining



- Data mining can help discover new insights about the business.
  - It is data driven and searches for answers to questions that may have not been previously asked.
- Analyzing data with OLAP tools, extends the capabilities of query and reporting.
  - Rather than submitting multiple queries, data is properly structured to enable fast and easy access to answers to the questions that are typically asked.
- Executing queries results in reports or charts.
  - It is driven by analysts who must pose questions to receive an answer

Role of Analyst

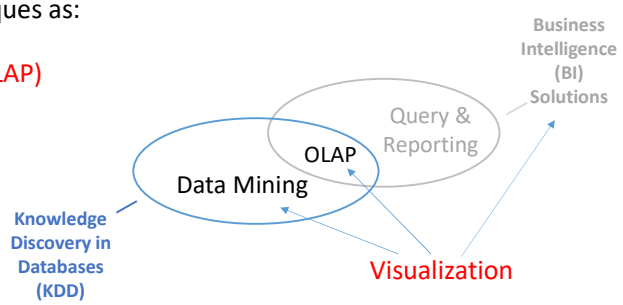


- Query and reporting analysis is fully driven by analyst who must pose questions.
- For multidimensional analysis data is structured to enable fast and easy access to answers to the questions that are typically asked.
- Data mining focuses on analyzing data rather then responding to questions.

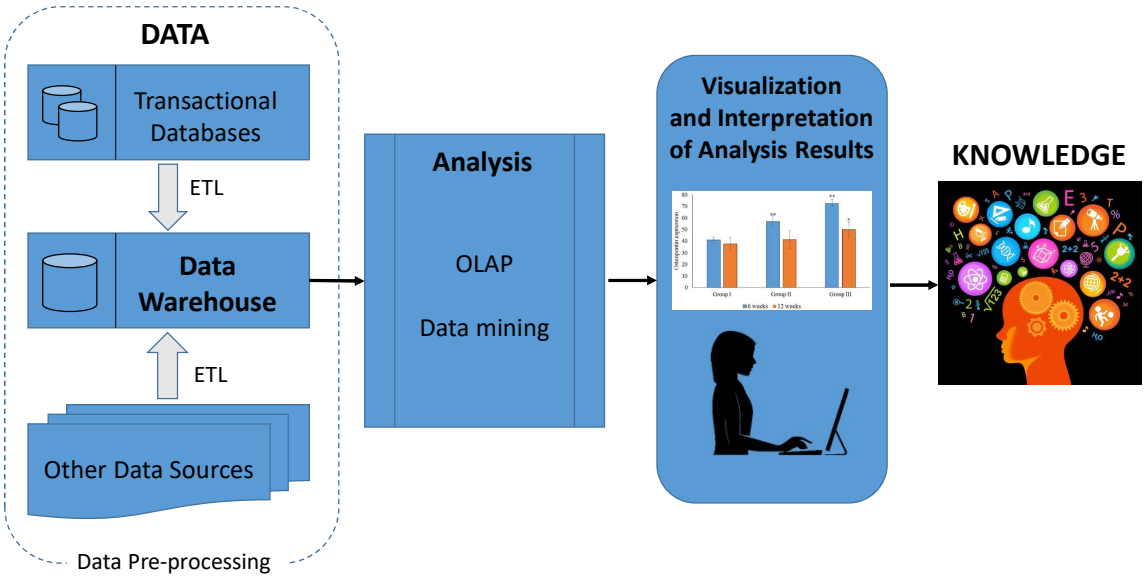
Knowledge Discovery in Databases

- The term *Knowledge Discovery in Databases* (KDD) refers to the broad process of finding knowledge in data, and emphasizes the "high-level" application of particular data mining methods.
  - It is of interest to researchers in machine learning, pattern recognition, databases, statistics, artificial intelligence, knowledge acquisition for expert systems, and data visualization.
  - ‚Knowledge' means here the relationships and patterns between data elements which can be extracted using a number of techniques as:

1) On-Line Analytical Processing (OLAP)
2) Data mining techniques



Knowledge Discovery from Data – A Schematic View



Business Intelligence

- Business Intelligence (BI) is actually a content-free expression, so it possibly means different things to different people.
- It is an umbrella-term that combines: architectures, tools, methodologies and data repositories and the related groups of users
- The major objective is to enable interactive access to data to enable manipulation of data and to give business managers and analysts the ability to conduct appropriate analysis.
- The process of BI is based on the transformation of data to information, then to decisions, and finally to actions.

Business Intelligence and Decision Support Systems

- Business intelligence comprises a collection of methodologies, processes, architectures, and technologies that transform raw data into meaningful and useful information for decision making.
- Business intelligence and decision-support systems provide assistance to managers at various organizational levels for analyzing strategic information.
  - These systems collect vast amounts of data and reduce them to a form that can be used to analyze organizational behavior.
  - This data transformation comprises a set of tasks that take the data from the sources and, through extraction, transformation, integration, and cleansing processes, store the data in a common repository called a data warehouse.
- Data warehouses have been developed and deployed as an integral part of decision support systems to provide an infrastructure that enables users to obtain efficient and accurate responses to complex queries.

Analyzing Data - Terminology