Introduction

1. Syllabus topic - Business intelligence:

– Business intelligence may be defined as a set of mathematical models and analysis methodologies that exploit the available data to generate information and knowledge useful for complex decision-making processes.
– The advent of low-cost data storage technologies and the wide availability of Internet connections have made it easier for individuals and organizations to access large amounts of data. Such data are often heterogeneous in origin, content and representation, as they include commercial, financial and administrative transactions, web navigation paths, emails, texts and hypertexts, and the results of clinical tests, to name just a few examples.
– Their accessibility opens up promising scenarios and opportunities, and raises an enticing question: is it possible to convert such data into information and knowledge that can then be used by decision makers to aid and improve the governance of enterprises and of public administration

1.2Effective and Timely decision:

– In complex organizations, public or private, decisions are made on a continual basis. Such decisions may be more or less critical, have long- or short-term effects and involve people and roles at various hierarchical levels.
– The ability of these knowledge workers to make decisions, both as individuals and as a community, is one of the primary factors that influence the performance and competitive strength of a given organization.

Example 1.1 –

Retention in the mobile phone industry, The marketing manager of a mobile phone company realizes that a large number of customers are discontinuing their service, leaving her company in favour of some competing provider. As can be imagined, low customer loyalty, also known as customer attrition or churn, is a critical factor for many companies operating in service industries. Suppose that the marketing manager can rely on a budget adequate to pursue a customer retention campaign aimed at 2000 individuals out of a total customer base of 2 million people. Hence, the

question naturally arises of how she should go about choosing those customers to be contacted so as to optimize the effectiveness of the campaign.

Effective decisions:

1. The application of rigorous analytical methods allows decision makers to rely on information and knowledge which are more dependable. As a result, they are able to make better decisions and devise action plans that allow their objectives to be reached in a more effective way.
2. Indeed, turning to formal analytical methods forces decision makers to explicitly describe both the criteria for evaluating alternative choices and the mechanisms regulating the problem under investigation.
3. Furthermore, the ensuing in-depth examination and thought lead to a deeper awareness and comprehension of the underlying logic of the decision-making process.

Timely decisions:

1. Enterprises operate in economic environments characterized by growing levels of competition and high dynamism.
2. As a consequence, the ability to rapidly react to the actions of competitors and to new market conditions is a critical factor in the success or even the survival of a company

Business intelligence:

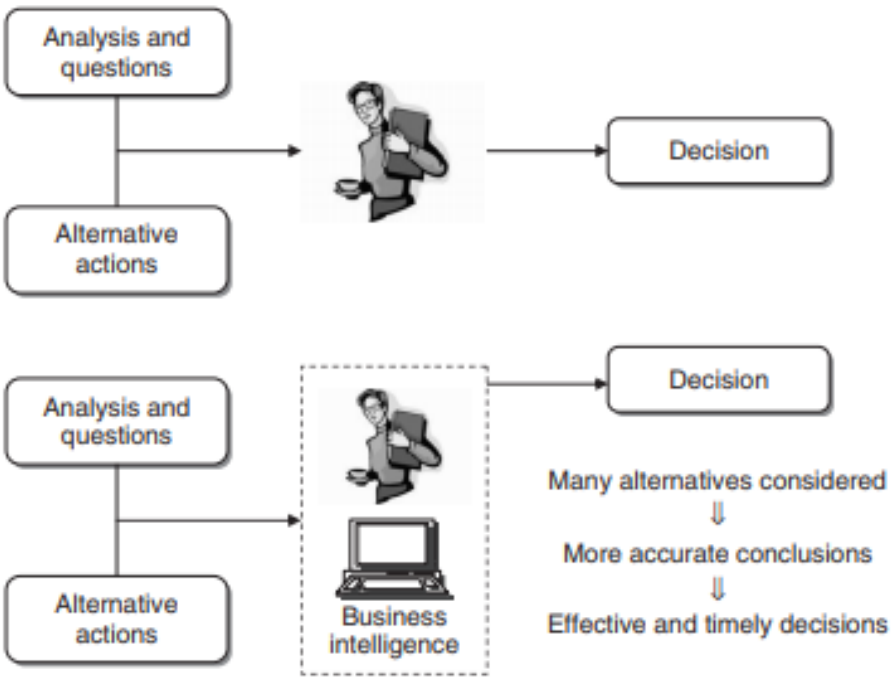


Figure: Benefits of business intelligence system

- If decision makers can rely on a business intelligence system facilitating their activity, we can expect that the overall quality of the decision-making process will be greatly improved with the help of mathematical models and algorithms; it is actually possible to analyze a larger number of alternative actions, achieve more accurate conclusions and reach effective and timely decisions.
- We may therefore conclude that the major advantage deriving from the adoption of a business intelligence system is found in the increased effectiveness of the decision-making process.

1.3 Data, Information and Knowledge

1. Data

Definition of Data:

- Data is nothing but representation of facts about objects which are distinguishable from other objects.

Examples: student data

- Consider example of student which has name, roll numbers, percentage and mobile numbers which are some facts of students. We can differentiate and identify students with above facts.

Usually data is static in nature.

1. It can represent a set of discrete facts about events.
2. Data is a prerequisite to information.
3. An organization sometimes has to decide on the nature and volume of data that is required for creating the necessary information.

2. Information:

- Information has usually got some meaning and purpose.

Definition of Information:

- Information can be considered as an aggregation of data (processed data) which makes decision making easier.

Example1:

To find Students with first class average in Engineering

- Consider above example of student which has data about all students. How to find information from above data, why we need to get information? How to process data to get information?
- Consider we need students those having average percentage greater than 60 percentages so here data is processed as per requirements.

Example2:

– To find Students with first class average in Engineering with any live backlog of any year. So we can process data and get list of student's as information per need

3. Knowledge:

– Knowledge is usually based on learning, thinking, and proper understanding of the problem area.
  1. Knowledge is not information and information is not data.
  2. Knowledge is derived from information in the same way information is derived from data.
  3. We can view it as an understanding of information based on its perceived importance or relevance to a problem area.

– By knowledge we mean human understanding of a subject matter that has been acquired through proper study and experience. It can be considered as the integration of human perceptive processes that helps them to draw meaningful conclusions

Definition of knowledge:

– By knowledge we mean human understanding of a subject matter that has been acquired through proper study and experience.

Example1:

– To find status of last two years of student's with first class and what is ratio of it.

– Knowledge is derived from information and systematic analysis of information

Example1: to find fail student's percentage of

Data, Information, Knowledge and Events

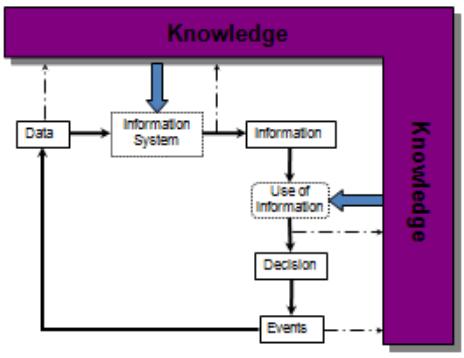


Figure: Relation between data, information and knowledge

1.4 The role of Mathematical models:

- A business intelligence system provides decision makers with information and knowledge extracted from data, through the application of mathematical models and algorithms.
- This activity may reduce to calculations of totals and percentages, graphically represented by simple histograms, whereas more elaborate analyses require the development of advanced optimization and learning models.
- In general terms, the adoption of a business intelligence system tends to promote a scientific and rational approach to the management of enterprises and complex organizations.
- Even the use of a spread sheet to estimate the effects on the budget of fluctuations in interest rates, despite its simplicity, forces decision makers to generate a mental representation of the financial flows process
- Classical scientific disciplines, such as physics, have always resorted to mathematical models for the abstract representation of real systems.
- Other disciplines, such as operations research, have instead exploited the application of scientific methods and mathematical models to the study of artificial systems, for example public and private organizations.
- The main mathematical models used in business intelligence architectures and decision support systems, as well as the corresponding solution methods
- Illustrate several related applications

The rational approach typical of a business intelligence analysis can be summarized schematically in the following main characteristics.

1. First, the objectives of the analysis are identified and the performance indicators that will be used to evaluate alternative options are defined.
2. Mathematical models are then developed by exploiting the relationships among system control variables, parameters and evaluation metrics.
3. Finally, what-if analyses are carried out to evaluate the effects on the performance determined by variations in the control variables and changes in the parameters.

1.5 The business intelligence architecture:

1. Data sources:

In a first stage, it is necessary to gather and integrate the data stored in the various primary and secondary sources, which are heterogeneous in origin and type. The sources consist for the most part of data belonging to operational systems, but may also include unstructured documents, such

as emails and data received from external providers. Generally speaking, a major effort is required to unify and integrate the different data sources.

2. Data warehouses and data marts:

– Using extraction and transformation tools known as extract, transform, load (ETL), the data originating from the different sources are stored in databases intended to support business intelligence analyses. These databases are usually referred to as data warehouses and data marts.

3. Business intelligence methodologies:

– Data are finally extracted and used to feed mathematical models and analysis methodologies intended to support decision makers. In a business intelligence system, several decision support applications may be implemented.

- Multidimensional cube analysis;
- exploratory data analysis;
- time series analysis;
- inductive learning models for data mining;
- optimization models

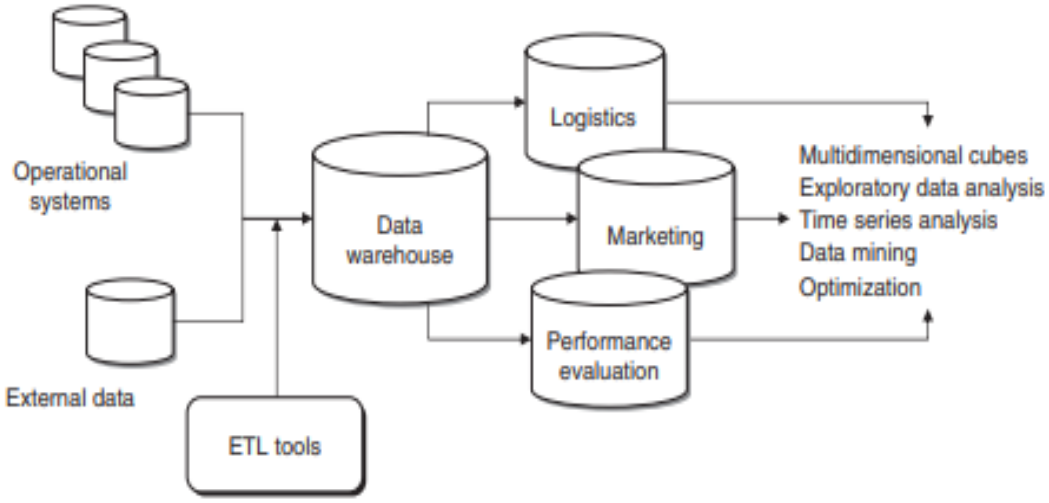


Figure: typical business intelligence architecture

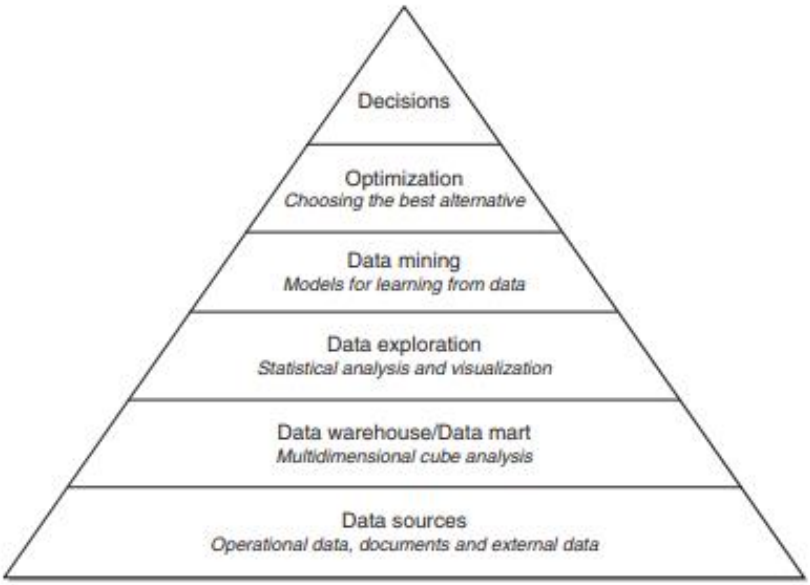


Figure: The main component of business intelligence architecture.

4. Data exploration:

- At the third level of the pyramid we find the tools for performing a passive business intelligence analysis, which consist of query and reporting systems, as well as statistical methods.
- These are referred to as passive methodologies because decision makers are requested to generate prior hypotheses or define data extraction criteria, and then use the analysis tools to find answers and confirm their original insight.
- For instance, consider the sales manager of a company who notices that revenues in a given geographic area have dropped for a specific group of customers.
- Hence, she might want to bear out her hypothesis by using extraction and visualization tools, and then apply a statistical test to verify that her conclusions are adequately supported by data.

5. Data mining:

- The fourth level includes active business intelligence methodologies, whose purpose is the extraction of information and knowledge from data.
- These include mathematical models for pattern recognition, machine learning and data mining techniques, which will be dealt with in Part II of this book.
- Unlike the tools described at the previous level of the pyramid, the models of an active kind do not require decision makers to formulate any prior hypothesis to be later verified.
- Their purpose is instead to expand the decision makers' knowledge.

6. Optimization:

– By moving up one level in the pyramid we find optimization models that allow us to determine the best solution out of a set of alternative actions, which is usually fairly extensive and sometimes even infinite. Other optimization models applied in marketing and logistics

7. Decisions:

– Finally, the top of the pyramid corresponds to the choice and the actual adoption of a specific decision and in some way represents the natural conclusion of the decision-making process.

– Even when business intelligence methodologies are available and successfully adopted, the choice of a decision pertains to the decision makers, who may also take advantage of informal and unstructured information available to adapt and modify the recommendations and the conclusions achieved through the use of mathematical models.

1.5.1 Cycle of business intelligence analysis:

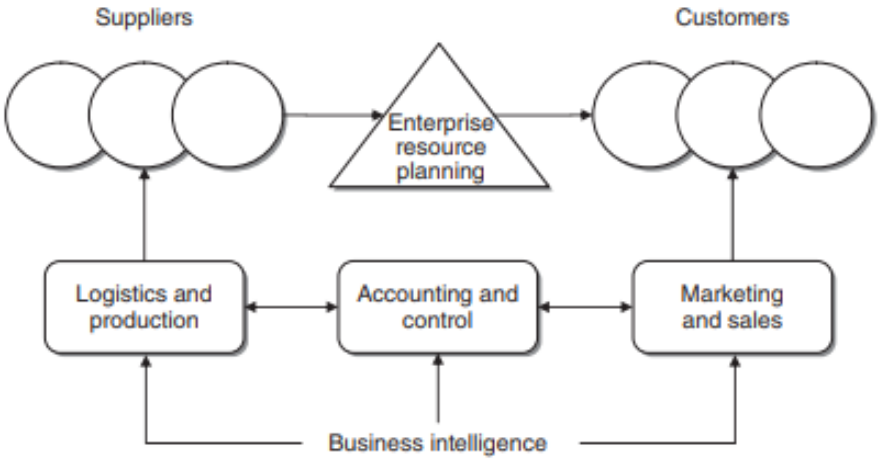


Figure: Department of an enterprise concerned with business intelligence

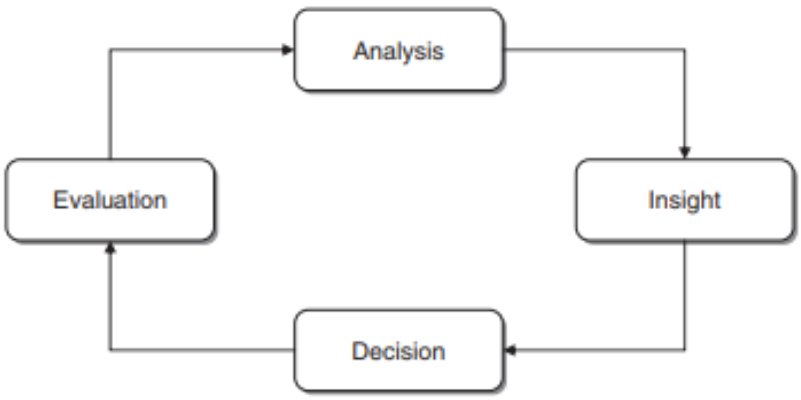


Figure: Cycle of business intelligence analysis

Analysis:

- During the analysis phase, it is necessary to recognize and accurately spell out the problem at hand. Decision makers must then create a mental representation of the phenomenon being analyzed, by identifying the critical factors that are perceived as the most relevant.
- The availability of business intelligence methodologies may help already in this stage, by permitting decision makers to rapidly develop various paths of investigation.
- For instance, the exploration of data cubes in a multidimensional analysis, according to different logical views.

Insight:

- The second phase allows decision makers to better and more deeply understand the problem at hand, often at a causal level. For instance, if the analysis carried out in the first phase shows that a large number of customers are discontinuing an insurance policy upon yearly expiration.
- It will be necessary to identify the profile and characteristics shared by such customers. The information obtained through the analysis phase is then transformed into knowledge during the insight phase.

Decision:

- During the third phase, knowledge obtained as a result of the insight phase is converted into decisions and subsequently into actions.
- The availability of business intelligence methodologies allows the analysis and insight phases to be executed more rapidly so that more effective and timely decisions can be made that better suit the strategic priorities of a given organization.
- This leads to an overall reduction in the execution time of the analysis–decision–action– revision cycle, and thus to a decision-making process of better quality.

Evaluation:

Finally, the fourth phase of the business intelligence cycle involves performance measurement and evaluation. Extensive metrics should then be devised that are not exclusively limited to the financial aspects but also take into account the major performance indicators defined for the different company departments.

1.5.2 Development of a business intelligence system:

- The development of a business intelligence system can be assimilated to a project, with a specific final objective, expected development times and costs, and the usage and coordination of the resources needed to perform planned activities.

1. Analysis:

- During the first phase, the needs of the organization relative to the development of a business intelligence system should be carefully identified.
- This preliminary phase is generally conducted through a series of interviews of knowledge workers performing different roles and activities within the organization.
- It is necessary to clearly describe the general objectives and priorities of the project, as well as to set out the costs and benefits deriving from the development of the business intelligence system.

2. Design:

- The second phase includes two sub-phases and is aimed at deriving a provisional plan of the overall architecture, taking into account any development in the near future and the evolution of the system in the mid-term.
- First, it is necessary to make an assessment of the existing information infrastructures. Moreover, the main decision-making processes that are to be supported by the business intelligence system should be examined, in order to adequately determine the information requirements.

Planning:

- The planning stage includes a sub-phase where the functions of the business intelligence system are defined and described in greater detail. Subsequently, existing data as well as other data that might be retrieved externally are assessed.
- This allows the information structures of the business intelligence architecture, which consist of a central data warehouse and possibly some satellite data marts, to be designed.
- Simultaneously with the recognition of the available data, the mathematical models to be adopted should be defined, ensuring the availability of the data required to feed each model and verifying that the efficiency of the algorithms to be utilized will be adequate for the magnitude of the resulting problems.

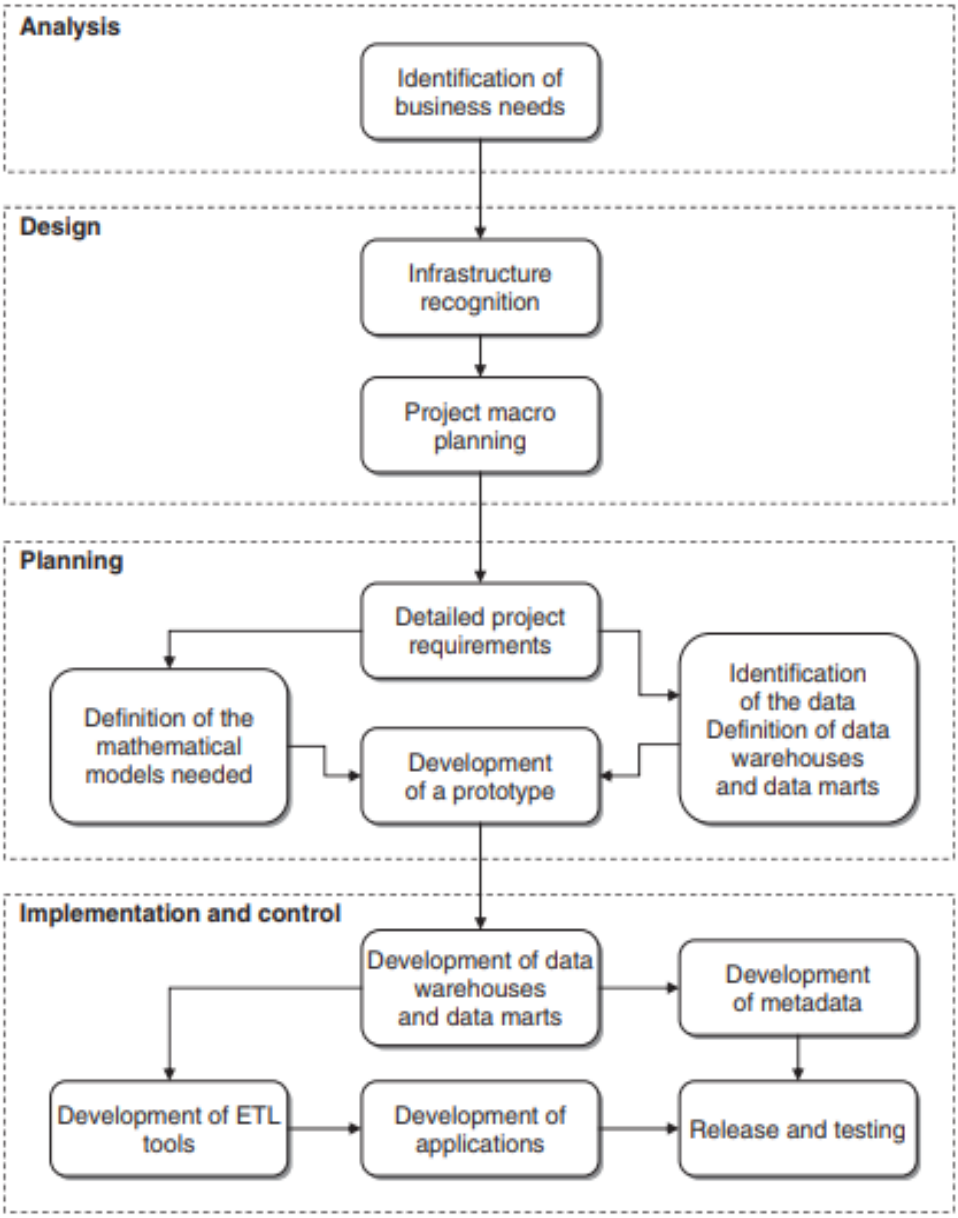


Figure: Phase in the development of a business intelligence system

Implementation and control:

- The last phase consists of five main sub-phases. First, the data warehouse and each specific data mart are developed.
- These represent the information infrastructures that will feed the business intelligence system. In order to explain the meaning of the data contained in the data warehouse and the transformations applied in advance to the primary data, a metadata archive should be created.

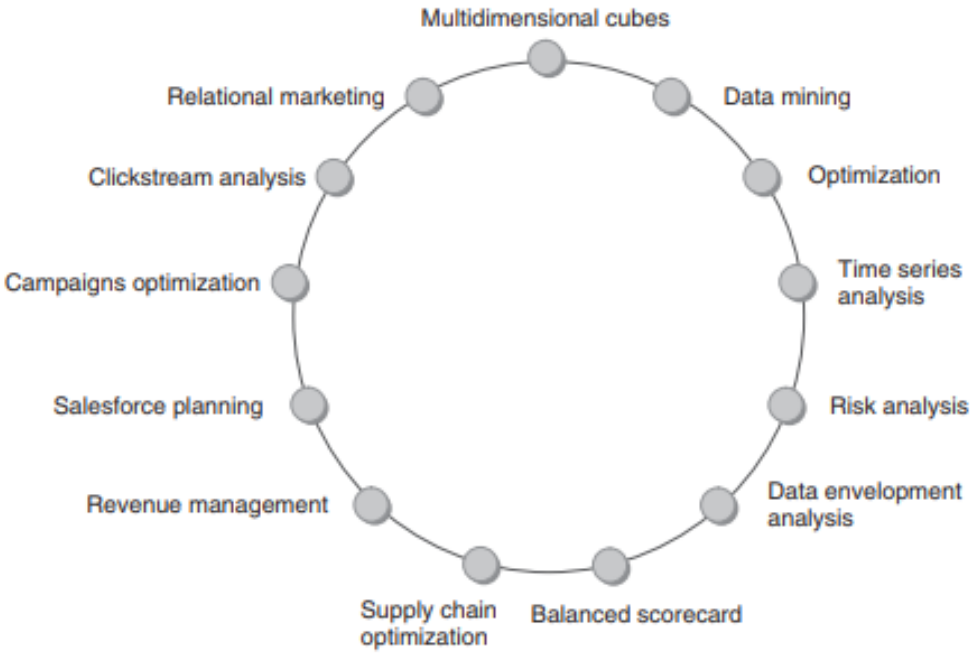


Figure: portfolio of available methodologies in a business intelligence system

2. Decision Support System:

- Decision Support System is a general term for any computer application that enhances a person or group's ability to make decisions.
- Decision Support Systems refers to an academic field of research that involves designing and studying Decision Support Systems in their context of use.
- Decision-support systems are used to make business decisions, often based on data collected by online transaction-processing systems.

2.1 Definition: "A Decision Support System (DSS)

- "A Decision Support System (DSS) is an interactive computer-based system or subsystem intended to help decision makers use communications technologies, data, documents, knowledge and/or models to identify and solve problems, complete decision process tasks, and make decisions".
- "DSS is basic component in the development of the BI Architecture."

Structure of DSS:

- "DSS is an interactive computer system helping decision makers to combine data and models to solve semi-structured and unstructured problems".

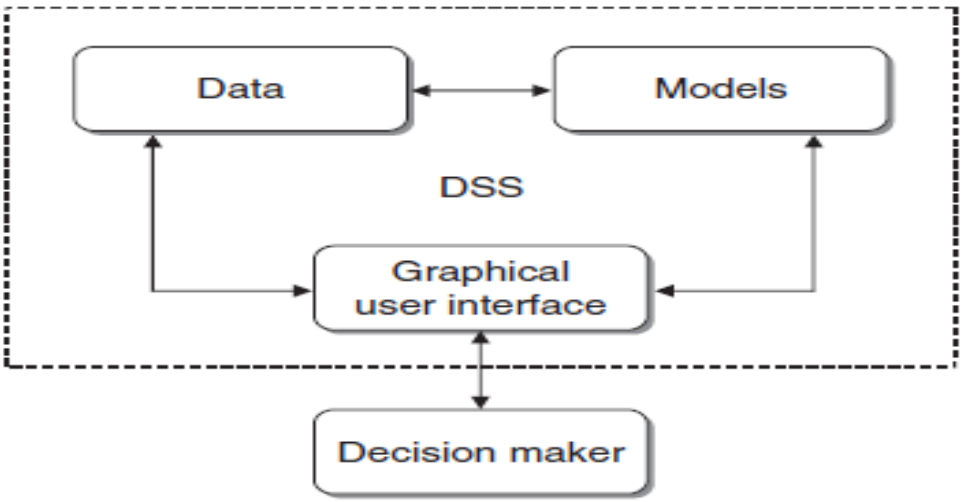


Figure: Structure of a DSS

- Decision-making is the process of identifying and choosing alternatives based on the values and preferences of the decision-maker.
- A decision is usually made with a fair degree of rationality. These decisions may concern the development of a strategic plan.
- The decision-making process is part of a broader subject usually referred to as problem solving, which refers to the process through which individuals try to bridge the gap between the current operating conditions of a system (as is) and the supposedly better conditions to be achieved in the future

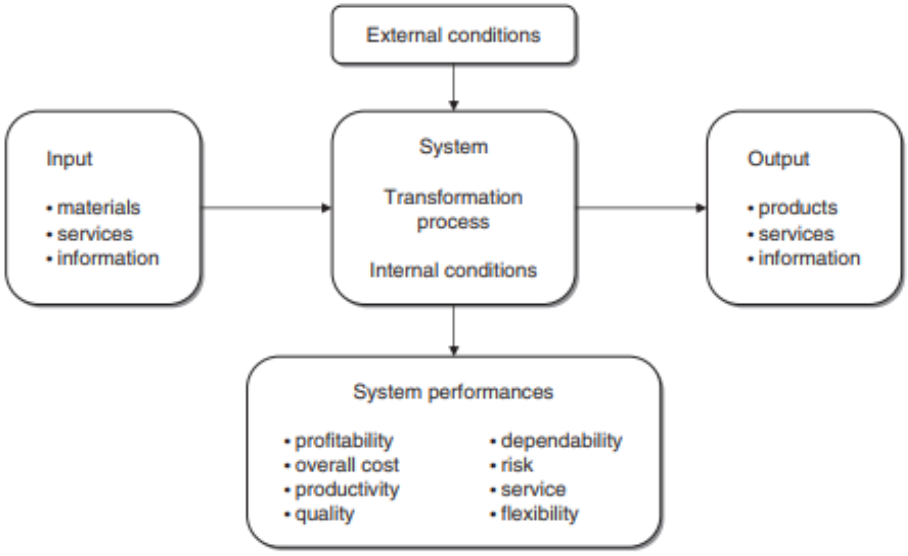


Figure: Abstract representation of System

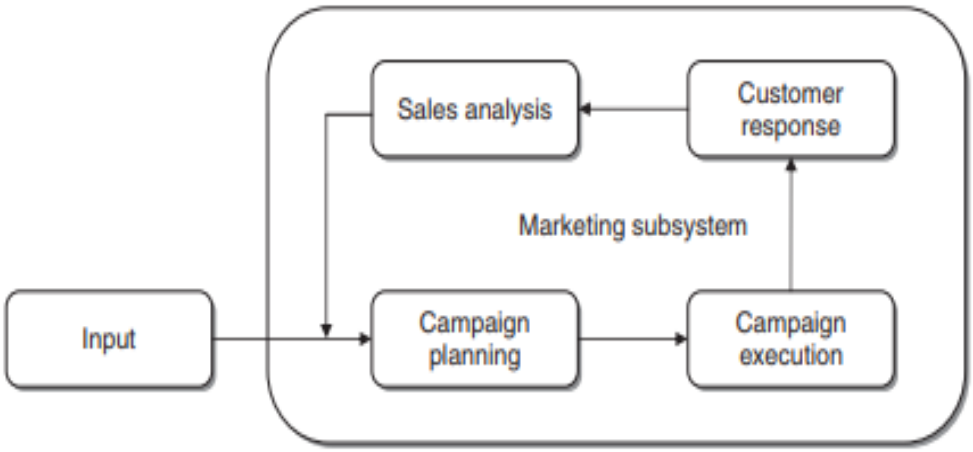


Figure: A closed cycled marketing system with feedback effect

Effectiveness:

- Effectiveness measurements express the level of conformity of a given system to the objectives for which it was designed. The associated performance indicators are therefore linked to the system output flows, such as production volumes, weekly sales and yield per share.

Efficiency:

- Efficiency measurements highlight the relationship between input flows used by the system and the corresponding output flows.
- Efficiency measurements are therefore associated with the quality of the transformation process. For example, they might express the amount of resources needed to achieve a given sales volume.

2.1.1 Integration in the decision-making process:

- A DSS should provide help for different kinds of knowledge workers, within the same application domain, particularly in respect of semi-structured and unstructured decision processes, both of an individual and a collective nature. Further, a DSS is intended for

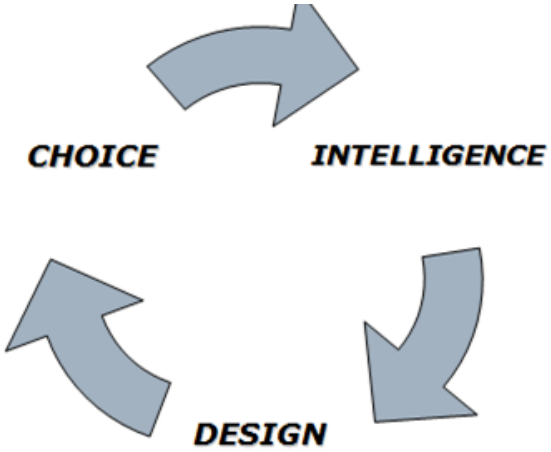


Figure: decision making process

The Early Framework of Decision Support System consists of four phases:

- **Intelligence** – Searching for conditions that call for decision;
- **Design** – Developing and analysing possible alternative actions of solution;
- **Choice** – Selecting a course of action among those;
- **Implementation** – Adopting the selected course of action in decision situation.

2.2 Representation of the decision-making process:

– In order to build effective DSSs, we first need to describe in general terms how a decision-making process is articulated.

– In particular, we wish to understand the steps that lead individuals to make decisions and the extent of the influence exerted on them by the subjective attitudes of the decision makers and the specific context within which decisions are taken.

2.2.1 Rationality and Problem Solving:

– A decision is a choice from multiple alternatives, usually made with a fair degree of rationality. Each individual faces on a continual basis decisions that can be more or less important, both in their personal and professional life.

– We will focus on decisions made by knowledge workers in public and private enterprises and organizations.

– These decisions may concern the development of a strategic plan and imply therefore substantial investment choices, the definition of marketing initiatives and related sales predictions, and the design of a production plan that allows the available human and technological resources to be employed in an effective and efficient way.

- The decision-making process is part of a broader subject usually referred to as problem solving, which refers to the process through which individuals try to bridge the gap between the current operating conditions of a system (as is) and the supposedly better conditions to be achieved in the future (to be).
- The transition of a system toward the desired state implies overcoming certain obstacles and is not easy to attain. This forces decision makers to devise a set of alternative feasible options to achieve the desired goal, and then choose a decision based on a comparison between the advantages and disadvantages of each option.
- Hence, the decision selected must be put into practice and then verified to determine if it has enabled the planned objectives to be achieved. When this fails to happen, the problem is reconsidered, according to recursive logic.
- The alternatives represent the possible actions aimed at solving the given problem and helping to achieve the planned objective.
- In some instances, the number of alternatives being considered may be small. In the case of a credit agency that has to decide whether or not to grant a loan to an applicant, only two options exist, namely acceptance and rejection of the request.
- In other instances, the number of alternatives can be very large or even infinite. For example, the development of the annual logistic plan of a manufacturing company requires a choice to be made from an infinite number of alternative options.

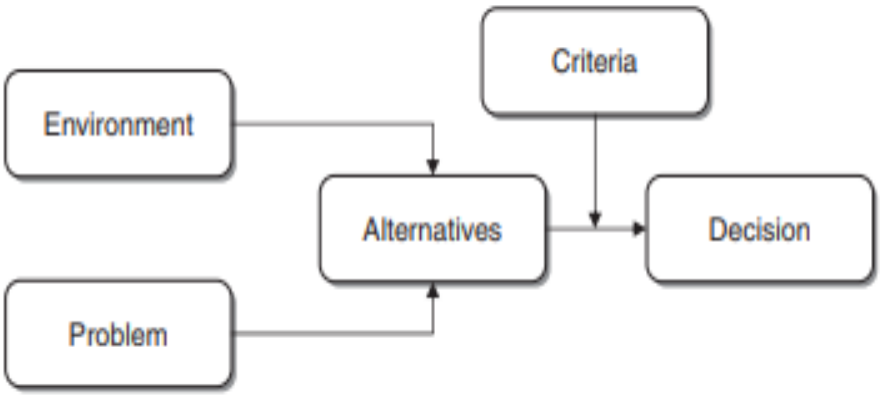


Figure: Logical flow of problem solving process

Economic:

- Economic factors are the most influential in decision-making processes, and are often aimed at the minimization of costs or the maximization of profits. For example, an annual logistic plan may be preferred over alternative plans if it achieves a reduction in total costs.

Technical:

- Options that are not technically feasible must be discarded. For instance, a production plan that exceeds the maximum capacity of a plant cannot be regarded as a feasible option.

Legal:

- Legal rationality implies that before adopting any choice the decision makers should verify whether it is compatible with the legislation in force within the application domain.

Ethical:

- Besides being compliant with the law, a decision should abide by the ethical principles and social rules of the community to which the system belongs.

Procedural:

- A decision may be considered ideal from an economic, legal and social standpoint, but it may be unworkable due to cultural limitations of the organization in terms of prevailing procedures and common practice.

Political:

- The decision maker must also assess the political consequences of a specific decision among individuals, departments and organizations

2.2.2 The Decision making Process:

- A compelling representation of the decision-making process was proposed in the early 1960s, and still remains today a major methodological reference.
- The model includes three phases, termed intelligence, design and choice. Figure shows extended version of the original scheme, which results from the inclusion of two additional phases, namely implementation and control.

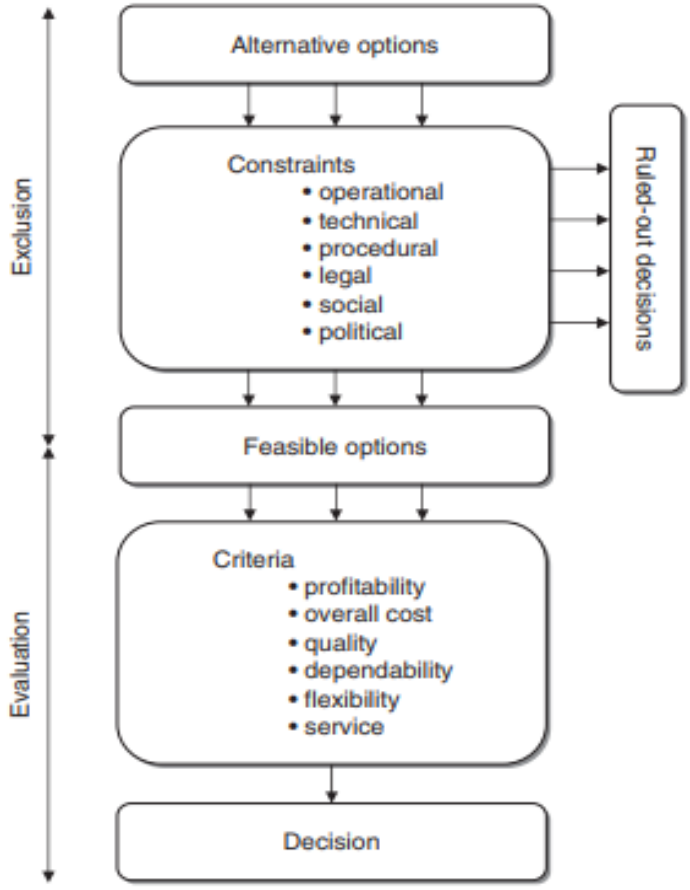


Figure: Logical structure of the decision making process

Intelligence:

- In the intelligence phase the task of the decision maker is to identify, circumscribe and explicitly define the problem that emerges in the system under study.
- The analysis of the context and all the available information may allow decision makers to quickly grasp the signals and symptoms pointing to a corrective action to improve the system performance.
- For example, during the execution of a project the intelligence phase may consist of a comparison between the current progress of the activities and the original development plan.
- In general, it is important not to confuse the problem with the symptoms. For example, suppose that an e-commerce bookseller receives a complaint concerning late delivery of a book order placed on-line. Such inconvenience may be interpreted as the problem and be tackled by arranging a second delivery by priority shipping to circumvent the dissatisfaction of the customer.
- On the other hand, this may be the symptom of a broader problem, due to an understaffed shipping department where human errors are likely to arise under pressure.

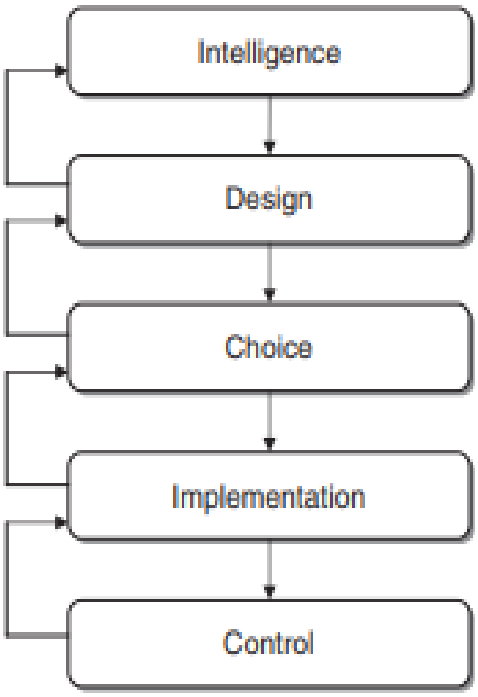


Figure: Phases of Decision making process

Design:

- In the design phase actions aimed at solving the identified problem should be developed and planned.
- At this level, the experience and creativity of the decision makers play a critical role, as they are asked to devise viable solutions that ultimately allow the intended purpose to be achieved.
- Where the number of available actions is small, decision makers can make an explicit enumeration of the alternatives to identify the best solution.
- If, on the other hand, the number of alternatives is very large, or even unlimited, their identification occurs in an implicit way, usually through a description of the rules that feasible actions should satisfy.
- For example, these rules may directly translate into the constraints of an optimization model.

Choice:

- Once the alternative actions have been identified, it is necessary to evaluate them on the basis of the performance criteria deemed significant.
- Mathematical models and the corresponding solution methods usually play a valuable role during the choice phase.

- For example, optimization models and methods allow the best solution to be found in very complex situations involving countless or even infinite feasible solutions.
- On the other hand, decision trees can be used to handle decision-making processes influenced by stochastic events.

Implementation:

- When the best alternative has been selected by the decision maker, it is transformed into actions by means of an implementation plan. This involves assigning responsibilities and roles to all those involved into the action plan.
- Control. Once the action has been implemented, it is finally necessary to verify and check that the original expectations have been satisfied and the effects of the action match the original intentions.
- In particular, the differences between the values of the performance indicators identified in the choice phase and the values actually observed at the end of the implementation plan should be measured.
- In an adequately planned DSS, the results of these evaluations translate into experience and information, which are then transferred into the data warehouse to be used during subsequent decision-making processes.

The most relevant aspects characterizing a decision-making process can be briefly summarized as follows.

- Decisions are often devised by a group of individuals instead of a single decision maker. • The number of alternative actions may be very high, and sometimes unlimited. • The effects of a given decision usually appear later, not immediately.
- The decisions made within a public or private enterprise or organization are often interconnected and determine broad effects. Each decision has consequences for many individuals and several parts of the organization.
- During the decision-making process knowledge workers are asked to access data and information, and work on them based on a conceptual and analytical framework.
- Feedback plays an important role in providing information and knowledge for future decision-making processes within a given organization.
- In most instances, the decision-making process has multiple goals, with different performance indicators, that might also be in conflict with one another.

- Many decisions are made in a fuzzy context and entail risk factors. The level of propensity or aversion to risk varies significantly among different individuals.
- Experiments carried out in a real-world system, according to a trial-and error scheme, are too costly and risky to be of practical use for decision making.
- The dynamics in which an enterprise operates, strongly affected by the pressure of a competitive environment imply that knowledge workers need to address situations and make decisions quickly and in a timely fashion.

2.2.3. Types of Decision Support System:

- Defining taxonomy of decisions may prove useful during the design of a DSS, since it is likely that decision-making processes with similar characteristics may be supported by the same set of methodologies.
- Decisions can be classified in terms of two main dimensions, according to their nature and scope. Each dimension will be subdivided into three classes, giving a total of nine possible combinations.
- According to their nature Decision can be classified. There are different types of decision support system, are as follows:
    1. Structured Decisions
    2. Semi-structured Decisions.
    3. Unstructured Decisions

1. Structured Decision:

- A decision is structured if it is based on a well-defined and recurring decision-making procedure. In most cases structured decisions can be traced back to an algorithm, which may be more or less explicit for decision makers, and are therefore better suited for automation.
- More specifically, we have a structured decision if input flows, output flows and the transformations performed by the system can be clearly described in the three phases of intelligence, design and choice.

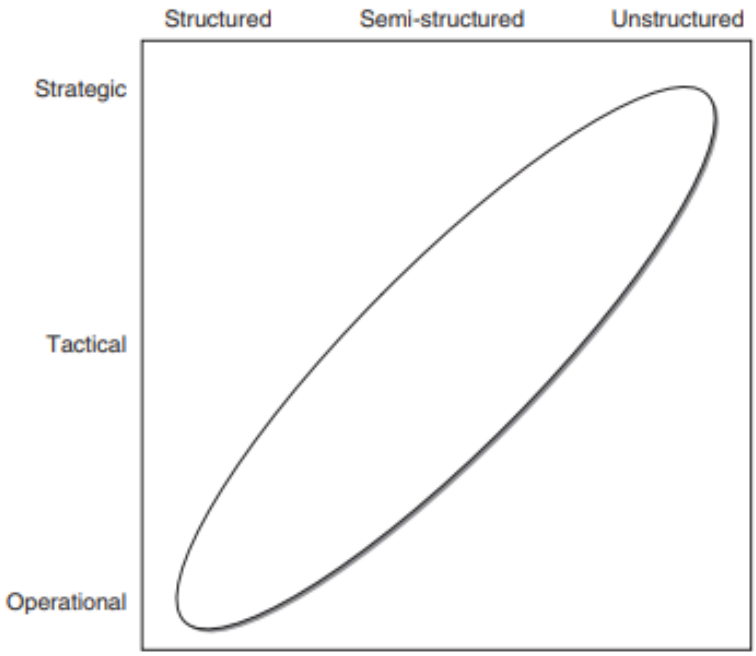


Figure: Taxonomy of Decision

Example:

- A paper mill produces for the company warehouse paper sheets in different standard sizes that are subsequently cut to size for customers. Specifically, customers submit orders in terms of type of paper, quantity and size.
- The sizes specified in the orders are usually smaller than standard sizes and must be cut out of these The paper mill is therefore forced to consider how the sizes required to fulfil orders should best be combined and cut from standard sizes so as to minimize paper waste.
- This decision is common to many industries (paper, aluminium, wood, steel, glass, fabric) and can be very well supported by optimization models.
- However, even in connection with such structured decisions, particular circumstances and specific input values may require intervention by the decision maker to modify the plans obtained by means of optimization models.
- For example, the company may wish to favor a specific request of a customer considered strategic, introducing a fast-processing lane in the cutting plan, even if this may involve more wasted material during the cutting stage.

2. Semi-structured Decisions:

- A decision is semi-structured when some phases are structured and others are not. Most decisions faced by knowledge workers in managing public or private enterprises or organizations are semi-structured.

- Hence, they can take advantage of DSSs and a business intelligence environment primarily in two ways. For the unstructured phases of the decision-making process, business intelligence tools may offer a passive type of support which translates into timely and versatile access to information.
- For the structured phases it is possible to provide an active form of support through mathematical models and algorithms that allow significant parts of the decision-making process to be automated.
- Sometimes situations may arise where the nature of a decision cannot be easily identified unambiguously. When facing the same problem, such as establishing the sale price of a product, different decision makers operating in different organizations may come up with dissimilar choices.
- For example, a first decision maker may believe that the best sale price can be obtained by comparing cost and price–demand elasticity curves. As a consequence, such decision maker may consider the choice phase of the decision-making process as structured.
- By contrast, a second decision maker may believe that the elasticity curve does not reflect all the factors influencing the response of the market to price variations since some of these elements cannot be quantified. For this individual the choice phase turns out to be unstructured or at most semi-structured.

Example:

The logistics manager of a manufacturing company needs to develop an annual plan. The logistic plan determines the allocation to each plant of the production volumes forecasted for the different market areas, the purchase of materials from each supplier with the related volumes and delivery times, the production lots for each manufacturing stage, the stock levels of sub-assemblies and end items, and the distribution of end items to the market areas. These decisions have a great economic and organizational impact that might greatly benefit from the adoption of a DSS based on large-scale optimization models. However, it is likely that in a real situation some elements are left to discretion of the decision makers, who may prefer a given logistic plan over another, even if it implies moderately higher costs compared to the optimal plan proposed by the model. For example, it might be appropriate to maintain unaltered the supply of parts purchased from a given supplier who is considered strategic for the future even though this supplier is less competitive than others that are instead preferred by the optimization model in terms of minimum cost.

3. Unstructured Decisions:

- A decision is said to be unstructured if the three phases of intelligence, design and choice are also unstructured. This means that for each phase there is at least one element in the system (input flows, output flows and the transformation processes) that cannot be described in detail and reduced to a predefined sequence of steps.
- Such an event may occur when a decision-making process is faced for the first time or if it happens very seldom. In this type of decisions the role of knowledge workers is fundamental, and business intelligence systems may provide support to decision makers through timely and versatile access to information.

Example:

Consider an enterprise that is the target of a hostile takeover by a public offer made by a direct competitor. There are various possible defensive decisions and actions that are strongly dependent on the context in which the enterprise operates and the offer is made. It is difficult to envisage a systematic description of the decision process that might be later reproduced in other similar cases. From the above examples it emerges that the nature of a decision process depends on many factors, including:

- The characteristics of the organization within which the system is placed;
- The subjective attitudes of the decision makers;
- The availability of appropriate problem-solving methodologies;
- The availability of effective decision support tools.

Depending on their scope, decisions can be classified as strategic, tactical and operational.

1. Strategic decisions:

- Decisions are strategic when they affect the entire organization or at least a substantial part of it for a long period of time.
- Strategic decisions strongly influence the general objectives and policies of an enterprise. As a consequence, strategic decisions are taken at a higher organizational level, usually by the company top management.

2. Tactical decisions:

- Tactical decisions affect only parts of an enterprise and are usually restricted to a single department. The time span is limited to a medium-term horizon, typically up to a year.
- Tactical decisions place themselves within the context determined by strategic decisions. In a company hierarchy, tactical decisions are made by middle managers, such as the heads of the company departments.

3. Operational decisions:

- Operational decisions refer to specific activities carried out within an organization and have a modest impact on the future.
- Operational decisions are framed within the elements and conditions determined by strategic and tactical decisions.

Nevertheless structure of BI is not stable; producers of business intelligence solutions may cover only some components into their products or expand utility function according to customer wish.

- **Model driven DSS** - includes systems that use accounting and financial models, representational models, and optimization models.
- **Data driven DSS** – includes file drawer and management reporting systems, data warehousing and analysis systems, Executive Information Systems (EIS) and Geographic Information Systems (GIS).
- **Knowledge driven DSS** - can suggest or recommend actions to managers. These DSS are person-computer systems with specialized problem-solving expertise. The "expertise" consists of knowledge about a particular domain, understanding of problems within that domain, and "skill" at solving some of these problems.

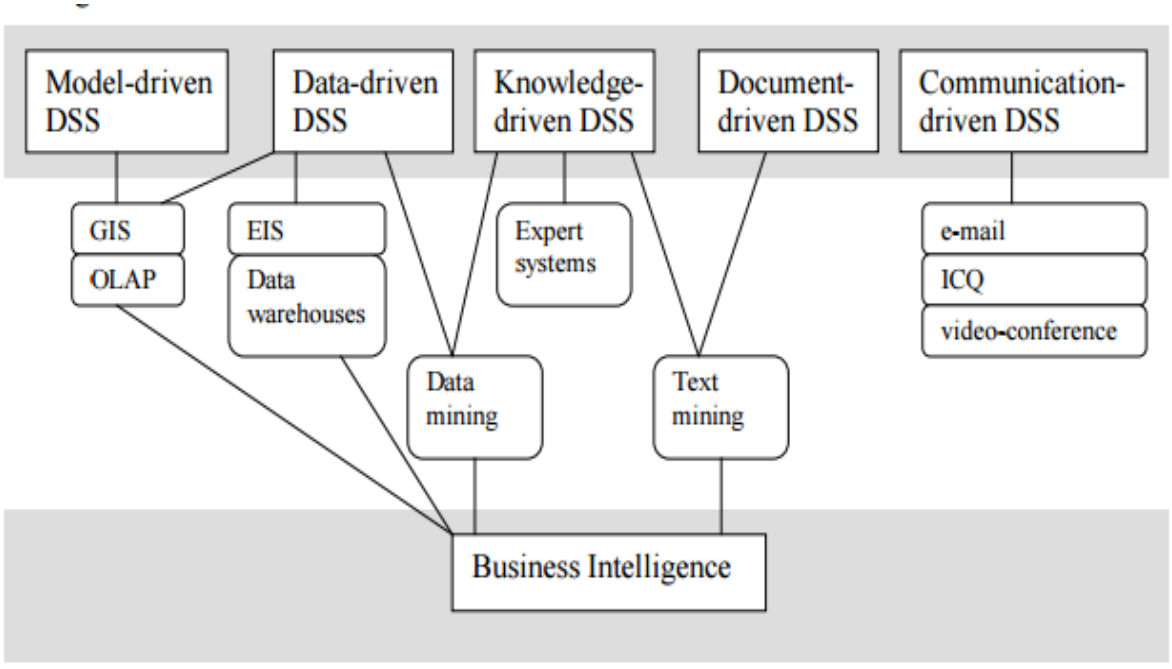


Figure: Business Intelligence in DSS

- **Document driven DSS** - integrates a variety of storage and processing technologies to provide complete document retrieval and analysis. The Web provides access to large document databases including databases of hypertext documents, images, sounds and video. A search engine is a powerful decision-aiding tool associated with this type of DSS.

- **Communication driven and group DSS** – where communication driven DSS includes communication, collaboration and coordination and GDSS focus on supporting groups of decision makers to analyze problem situations and performing group decision making tasks.

Use of Data mining & Text mining in DSS:

- **Data Mining** - It is the computational process of discovering patterns in large **data** sets ("big **data**") involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems.
- Data mining is the process of sifting through large amounts of data to produce data content relationships. Data mining tools can be used to create hybrid Data-Driven and Knowledge-Driven DSS.
- **Text mining -** also referred to as **text** data **mining**, roughly equivalent to **text** analytics, refers to the process of deriving high-quality information from **text**. High-quality information is typically derived through the devising of patterns and trends through means such as statistical pattern learning.

The characteristics of the information required in a decision-making process will change depending on the scope of the decisions to be supported, and consequently also the orientation of a DSS will vary accordingly.

	Operational	Tactical	Strategic
Accuracy	High	⟷	Low
Level of detail	Detailed	⟷	Aggregate
Time horizon	Present	⟷	Future
Frequency of use	High	⟷	Low
Source	Internal	⟷	External
Scope of information	Quantitative	⟷	Qualitative
Nature of information	Narrow	⟷	Wide
Age of information	Present	⟷	Past

Figure: Characteristics of the information in terms of the scope of decision

2.2.4 Approaches to the Decision making Process:

- The subjective orientation of decision makers across an organization strongly influences the structure of the decision-making process.
- Review the major approaches that prevail in the management of complex organizations, and examine the implications when designing a DSS. A preliminary distinction should be made between a rational approach and a political-organizational approach.

1. Rational approach.

- When a rational approach is followed, a decision maker considers major factors, such as economic, technical, legal, ethical, procedural and political, also establishing the criteria of evaluation so as to assess different options and then select the best decision.
- DSS may help both in a passive way, through timely and versatile access to information, and in an active way, through the use of mathematical models for decision making.

2. Political-organizational approach:

- When a political-organizational approach is pursued, a decision maker proceeds in a more instinctual and less systematic way.
- Decisions are not based on clearly defined alternatives and selection criteria. As a consequence, a DSS can only help in a passive way, providing timely and versatile access to information.
- It might also be useful during discussions and negotiations in those decision-making processes that involve multiple actors, such as managers operating in different departments. Within the rational approach we can further distinguish between two alternative ways in which the actual decision-making process influences decisions: absolute rationality and bounded rationality.

3. Absolute rationality:

- The term 'absolute rationality' refers to a decision-making process for which multiple performance indicators can be reduced to a single criterion, which therefore naturally lends itself to an optimization model.
- For example, a production manager who has to put together a medium-term logistic plan may be able to convert all performance indicators into monetary units, and therefore subsequently derive the solution with the minimum cost.
- This implies that non-monetary indicators, such as stock volumes or the number of days of delay in handling a given order, should be transformed into monetary measurement units.

- From a methodological perspective, this implies that a multi-objective optimization problem is transformed into a single-objective problem by expressing all the relevant factors in a common measurement unit that allows the heterogeneous objectives to be added together.

4. Bounded rationality:

- Bounded rationality occurs whenever it is not possible to meaningfully reduce multiple criteria into a single objective, so that the decision maker considers an option to be satisfactory when the corresponding performance indicators fall above or below prefixed threshold values.
- For instance, a production plan is acceptable if its cost is sufficiently low, the stock quantities are within a given threshold, and the service time is below customers' expectations. Therefore, the concept of bounded rationality captures the rational choices that are constrained by the limits of knowledge and cognitive capability.
- Most decision-making processes occurring within the enterprises and the public administration are aimed at making a decision that appears acceptable with respect to multiple evaluation criteria, and therefore decision processes based on bounded rationality are more likely to occur in practice.

– A logical and ordered process can help you to do this by making sure that you address all of the critical elements needed for a successful outcome. Working through this process systematically will reduce the likelihood of overlooking important factors. Our seven-step approach takes this into account:

Step 1: Identify the decision to be made.

– You realize that a decision must be made. You then go through an internal process of trying to define clearly the nature of the decision you must make. This first step is a very important one.

Step 2: Gather relevant information.

– Most decisions require collecting pertinent information. The real trick in this step is to know what information is needed the best sources of this information, and how to go about getting it.
– Some information must be sought from within you through a process of self-assessment; other information must be sought from outside yourself-from books, people, and a variety of other sources. This step, therefore, involves both internal and external "work".

Step 3: Identify alternatives.

– Through the process of collecting information you will probably identify several possible paths of action, or alternatives. You may also use your imagination and information to construct new

alternatives. In this step of the decision-making process, you will list all possible and desirable alternatives.

Step 4: Weigh evidence. –

- In this step, you draw on your information and emotions to imagine what it would be like if you carried out each of the alternatives to the end.
- You must evaluate whether the need identified in Step 1 would be helped or solved through the use of each alternative.
- In going through this difficult internal process, you begin to favor certain alternatives which appear to have higher potential for reaching your goal. Eventually you are able to place the alternatives in priority order, based upon your own value system.

Step 5: Choose among alternatives. –

- Once you have weighed all the evidence, you are ready to select the alternative which seems to be best suited to you. You may even choose a combination of alternatives.
- Your choice in Step 5 may very likely be the same or similar to the alternative you placed at the top of your list.

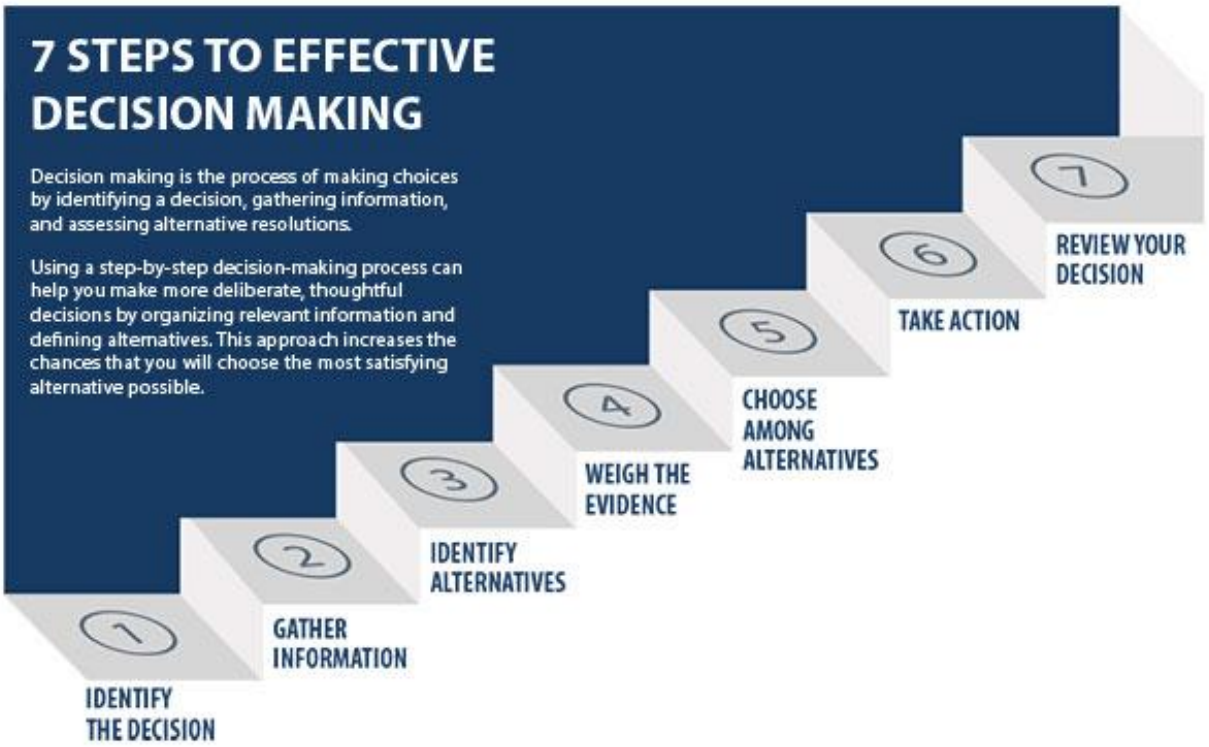


Figure: Steps of Decision making model

Step 6: Take action.

- You now take some positive action which begins to implement the alternative you chose in Step 5.

Step 7: Review.

- Decision and consequences in the last step you experience the results of your decision and evaluate whether or not it has "solved" the need you identified in Step 1.
- If it has, you may stay with this decision for some period of time. If the decision has not resolved the identified need, you may repeat certain steps of the process in order to make a new decision.
- For example, gather more detailed or somewhat different information or discover additional alternatives on which to base your decision.

Examples of business decisions:

- What items to stock?
- What insurance premium to change?
- To whom to send advertisements
- Examples of data used for making decisions
- Retail sales transaction details
- Customer profiles (income, age, gender, etc.)

2.2.5 Characteristics and Capabilities of DSS.

The Characteristics and Capabilities of DSS are:

1. Support for decision makers (mainly in semi- and un-structured situation) by bringing together human judgment and computerized information.
2. Support for all managerial levels, ranging from top executives to line managers.
3. Support for individuals (from different departments, organizational levels or different organizations) as well as groups of decision makers working somewhat independently – virtual teams through collaborative Web tools.
4. Support for independent or sequential decisions that may be made once, several times or repeatedly.
5. Support in all phases of decision-making process (*intelligence, design, choice, and implementation*).
6. Support for a variety of decision-making process and style.

7. The decision maker should be reactive, able to confront changing conditions quickly and able to adapt the DSS to meet these changes. DSS are flexible, so users can *add, delete, combine, change or rearrange basic elements*.
8. User-friendliness, strong graphical capabilities and natural language interactive human-machine interface can greatly increase the effectiveness of DSS, Most new DSS application use Web-based interfaces
9. Improvement the effectiveness of decision making rather than its efficiency. When DSS are deployed, decision making often takes longer but the decisions are better.
10. The decision maker has complete control over all steps of the decision-making process in solving a problem – a DSS aims to support not to replace the decision maker.
11. End users are able to develop and modify simple systems by themselves. Larger systems can be built with assistance from information system specialist. Online analytical process (OLAP) and data mining software, with data warehouses, allow users to build very large and complex DSS.
12. Models are generally utilized to analyze decision-making situations. The modeling capability enable experimentation with different strategies under different configurations
13. Access is provides to a variety of data sources, formats and types, including GIS, multimedia and object oriented.
14. Can be employed as a standalone tool used by an individual decision maker in one location or distributed throughout an organization and in several organizations along the supply chain. It can be integrated with other DSS or applications and it can be distributed internally and externally using networking and Web technologies.

These key DSS Characteristics and Capabilities allow decision makers to make better, more consistent decision in a timely manner and they are provided by the major DSS components.

1. Components of Decision Support System (DSS) schematic view:

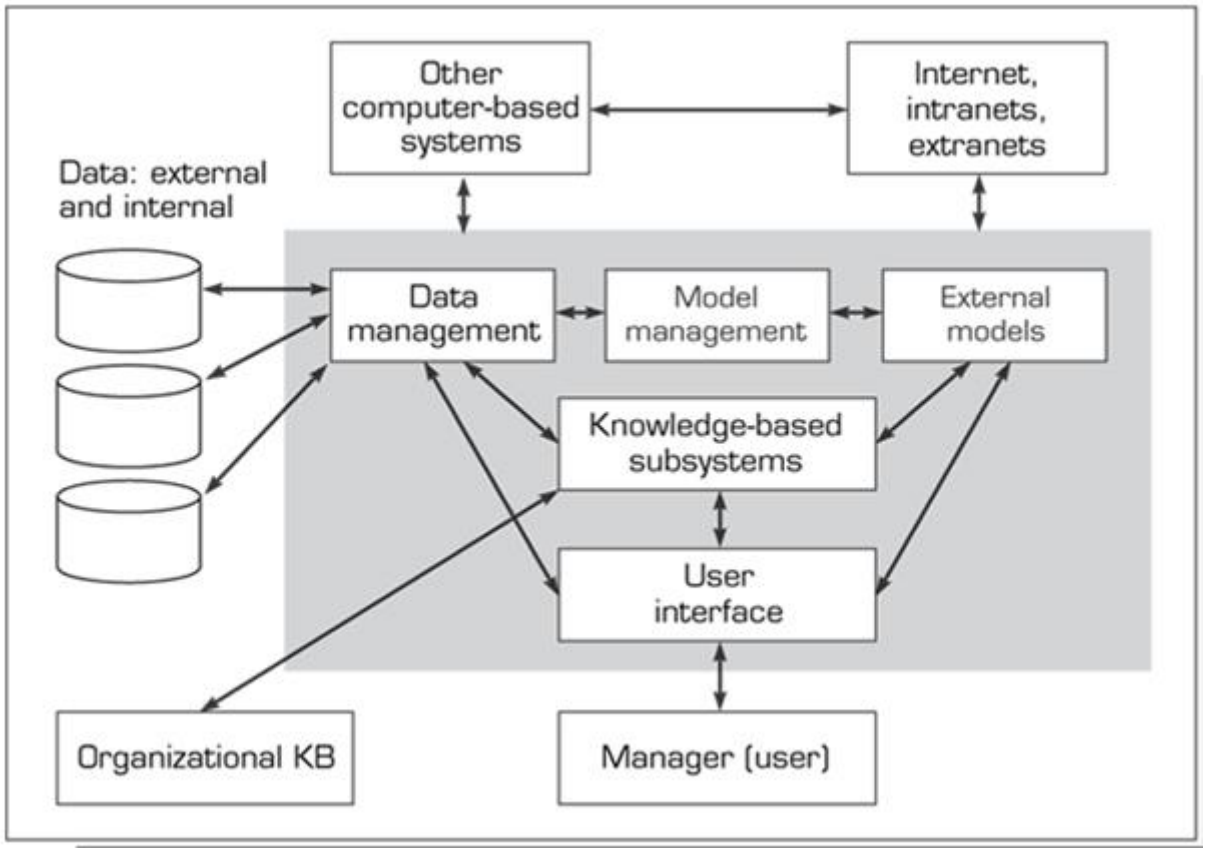


Figure: components of DSS schematic view

Subsystems:

Data management Managed by DBMS:

- Extracts data Manages data and their relationships
- Updates (add, delete, edit, change)
- Retrieves data (accesses it)
- Queries and manipulates data
- Employs data dictionary

Model management Managed by MBMS

Model management subsystem of a DSS consists of the components:

- Model base
- Model base management system
- Modeling language
- Model directory
- Model execution, integration, and command processor

User interface

- GUI
- Natural language processor
- Interacts with model management and data management subsystems
- Examples: Speech recognition, Display panel, tactile interfaces, Gesture interface

Knowledge Management and organizational knowledge base

- Many unstructured/semi structured problems need expertise (knowledge) for their solutions.
- Such expertise can be provided by some knowledge engineers who interview the domain experts and gather the information necessary for the knowledge-base.
- More advanced DSSs are equipped with a component called knowledge base management subsystem.
- This subsystem can achieve complex problem solving and it can enhance operations of other components.
- The knowledge base is where the "knowledge" of the DSS is stored. By knowledge, we mean the rules, heuristics, constraints, previous outcomes and any other "knowledge" that may have been programmed into the DSS.

Applications of Decision Support System:

1. Clinical decision support system for medical diagnosis.
2. A bank loan officer verifying the credit of a loan applicant.
3. An engineering firm that has bids on several projects and wants to know if they can be competitive with their costs.
4. DSS is extensively used in business and management. Executive dashboards and other business performance software allow faster decision making, identification of negative trends, and better allocation of business resources.
5. A growing area of DSS application, concepts, principles, and techniques is in agricultural production, marketing for sustainable development.
6. A specific example concerns the Canadian National Railway system, which tests its equipment on a regular basis using a decision support system.
7. A DSS can be designed to help make decisions on the stock market, or deciding which area or segment to market a product towards.

Attributes of a DSS:

1. Adaptability and flexibility
2. High level of Interactivity
3. Ease of use
4. Efficiency and effectiveness
5. Complete control by decision-makers
6. Ease of development
7. Extendibility
8. Support for modelling and analysis
9. Support for data access
10. Standalone, integrated, and Web-based

2.2.6 Development of Decision support system:

The development phases of a DSS

In order to develop most DSSs a specific project shows the major steps required in the development of a DSS. The logical flow of the activities is shown by the solid arrows. The dotted arrows in the opposite direction indicate revisions of one or more phases that might become necessary during the development of the system, through a feedback mechanism.

a) Planning:

- The main purpose of the planning phase is to understand the needs and opportunities, sometimes characterized by weak signals, and to translate them into a project and later into a successful DSS.

b) Analysis:

- In the analysis phase, it is necessary to define in detail the functions of the DSS to be developed.
- A response should therefore be given to the following question: What should the DSS accomplish, and who will use it, when and how?
- The analysis also involves mapping out the actual decision processes and imagining what the new processes will look like once the DSS is in place.
- Finally, it is necessary to explore the data in order to understand how much and what type of information already exists and what information can be retrieved from external sources.

c) Design:

- During the design phase the main question is: How will the DSS work?
- It is also necessary to define in detail the interactions with the users, by means of input masks, graphic visualizations on the screen and printed reports. A further aspect is the make-or-buy choice – whether to subcontract the implementation of the DSS to third parties, in whole or in part.
- Implementation. Once the specifications have been laid down, it is time for implementation, testing and the actual installation, when the DSS is rolled out and put to work.
- A further aspect of the implementation phase, which is often overlooked, relates to the overall impact on the organization determined by the new system.

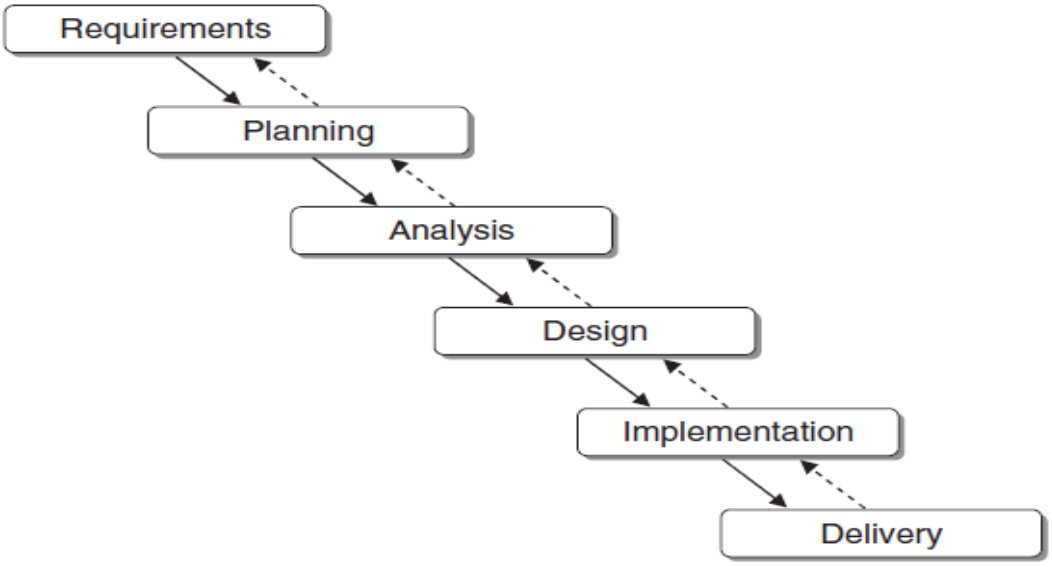


Figure: Development phases of decision support system.

d) Implementation

- Once the specifications have been laid down, it is time for implementation, testing and the actual installation, when the DSS is rolled out and put to work. Any problems faced in this last phase can be traced back to project management methods.
- A further aspect of the implementation phase, which is often overlooked, relates to the overall impact on the organization determined by the new system.
- Such effects should be monitored using change management techniques, making sure that no one feels excluded from the organizational innovation process and rejects the DSS.
- Sometimes a project may not come to a successful conclusion, may not succeed in

fulfilling expectations, or may even turn out to be a complete failure.

- However, there are ways to reduce the risk of failure. The most significant of these is based on the use of rapid prototyping development where, instead of implementing the system as a whole, the approach is to identify a sequence of autonomous subsystems, of limited capabilities, and develop these subsystems step by step until the final stage is reached corresponding to the fully developed DSS.

3. Data Warehousing:

- A data warehousing is a technique for collecting and managing data from varied sources to provide meaningful business insights. It is a blend of technologies and components which allows the strategic use of data.
- It is electronic storage of a large amount of information by a business which is designed for query and analysis instead of transaction processing.
- It is a process of transforming data into information and making it available to users in a timely manner to make a difference.
- A warehouse is a subject oriented, integrated, time-variant, and non-volatile collection of data in support management's decision making process.
- He defined the terms is sentence as follows.

1. **Subject oriented**

   Data that gives information about a particular subject instead of about a company's on-going operations

2. **Integrated**

   Data that is gathered into the data warehouse from a variety of sources and merged into a coherent whole

3. **Time variant**

   All data in the data warehouse is identified with a particular time period

4. **Non-volatile**

   Data is stable in a data warehouse. More data is added but data is never removed. This enables management to gain a consistent picture of the business.

3.1 Benefits of Data Warehousing:

1. Potential high returns on investment and delivers enhanced business intelligence:

Implementation of data warehouse requires a huge investment in lakhs or rupees. But it helps the organizations to take strategic decision based on past historical data organization

can improves the results of various result of various processes like marketing segmentation, inventory and management and step.

2. Competitive advantages:

As previously unknown and unavailable data is available in data warehouse decision makers can access that data to take decisions to gain the competitive advantages.

3. Save time:

As the data from multiple sources is available is integrated from business users can access data from one place. There is no retrieve the data multiple sources.

4. Better enterprise intelligence:

It improves the customer service and productivity.

5. High Quality Data:

Data in data warehouse is cleaned and transfer into desired format. So data quality is High.

3.1.1. Benefits of Data Warehouse:

Characteristics / Features of a Data Warehouse

A common way of introducing data warehousing is to refer to the characteristics of a data warehouse.

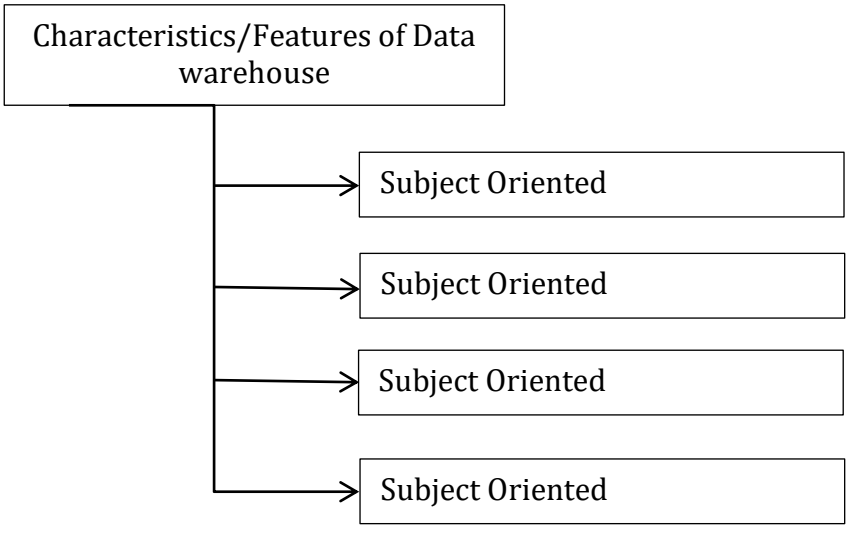


Figure: Characteristics / Features of Data Warehouse

1. Subject Oriented:

- Data warehouse are designed to help analyze data for example, to learn more about banking data, a warehouse can to build that concentrates on transaction loans etc.
- This warehouse can be used to answer questions like "which customer has taken maximum loan amount for last year?". This ability to define a data warehouse by subject matter, loan in this case, makes the data warehouse subject oriented.

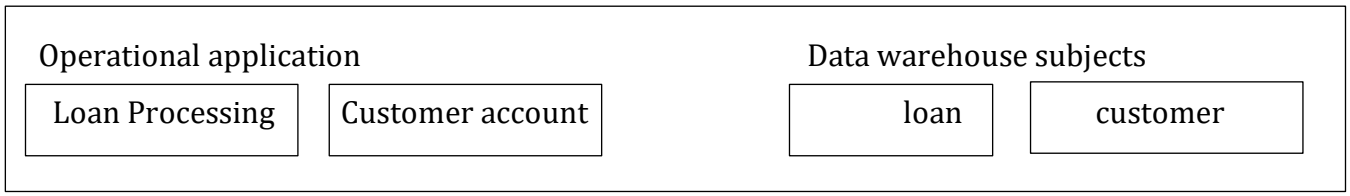


Figure: Data warehouse in subject oriented

2. Integrated:

- A data warehouse is constructed by integrating multiple, heterogeneous data sources like, relational databases, flat, files, on-line transaction records.
- The data collected is cleaned and then data integration techniques an applied, which ensure consistency in mining conventions, encoding structure, attribute measures etc., among different data sources

Example:

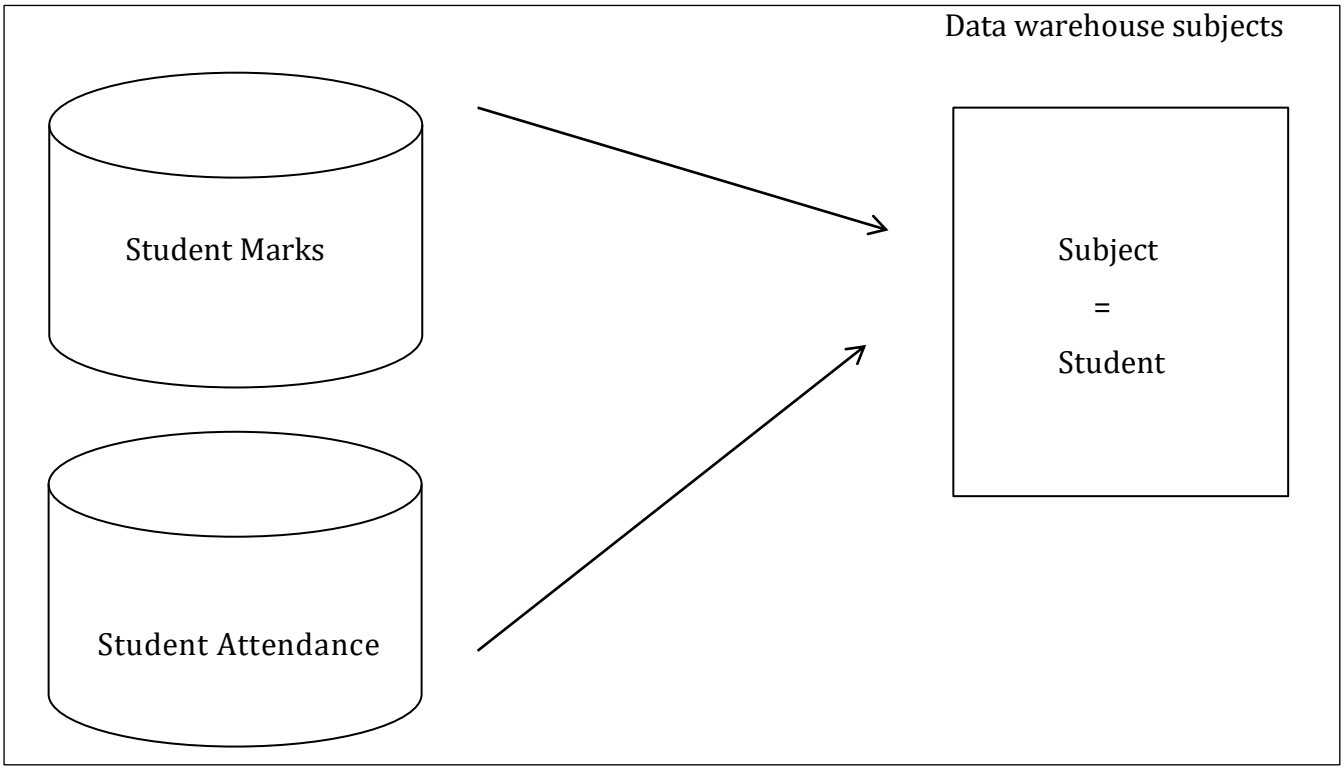


Figure: Integrated Data Warehouse

3. Non-volatile:

- Non-volatiles means that once data entered into the warehouse, it cannot be removed or changed because the purpose of warehouse is to analyze the data.

4. Time Variant:

- A data warehouse maintains historical data. For e.g. A customer record has details of his job. A Data warehouse would maintain all his previous job (historical information) when compared to a transactional system which only maintains current job due to which its not possible to retrieve older records.

3.2. Types of Data Warehouse:

Three main types of Data warehouse:

1. Enterprise Data Warehouse
2. Operational Data Store
3. Data Mart

1. Enterprise Data Warehouse:

- Enterprise Data Warehouse is a centralized warehouse. It provides decision support service across the enterprise.
- It offers a unified approach for organizing and representing data.
- It also provides the ability to classify data according to the subject and give access according to those divisions.

2. Operational Data Store:

- Operational Data Store, which is also called ODS, are nothing but data store required when neither Data warehouse nor OLTP systems support organizations reporting needs.
- In ODS, Data warehouse is refreshed in real time.
- It is widely preferred for routine activities like storing records of the Employees.

3. Data Mart:

- A data mart is a subset of the data warehouse.
- It specially designed for a particular line of business, such as sales, finance, sales or finance. In an independent data mart, data can collect directly from sources.

3.2.1. General Stages of Data Warehouse:

The following are the general use of the data warehouse.

1. Offline operational database

2. Offline Data Warehouse
3. Real time Data Warehouse
4. Integrated Data Warehouse

1. Offline operational database:

- Data is just copied from an operational system to another server. In this way, loading, processing, and reporting of the copied data do not impact the operational system's performance.

2. Offline Data Warehouse:

- Data in the Data warehouse is regularly updated from the Operational Database. The data in Data warehouse is mapped and transformed to meet the Data warehouse objectives.

3. Real time Data Warehouse:

- Data warehouses are updated whenever any transaction takes place in operational database. For example, Airline or railway booking system.

4. Integrated Data Warehouse:

- Data Warehouses are updated continuously when the operational system performs a transaction. The Data warehouse then generates transactions which are passed back to the operational system.

3.2.2 Component of Data Warehouse:

There are four component of Data Warehouse:

1. Load Manager
2. Warehouse Manager
3. Query Manager
4. End-User Access tools.

1. Load Manager:

- Load manager is also called the front component.
- It performs with all the operations associated with the extraction and load of data into the warehouse. These operations include transformations to prepare the data for entering into the Data warehouse.

2. Warehouse Manager:

- Warehouse manager performs operations associated with the management of the data in the warehouse.
- It performs operations like analysis of data to ensure consistency, creation of indexes and views, generation of demoralization and aggregations, transformation and merging of source data and archiving and baking-up data.

3. Query Manager:

- Query manager is also known as backend component.
- It performs all the operation operations related to the management of user queries. The operations of this Data warehouse component are direct queries to the appropriate tables for scheduling the execution of queries.

4. End-User Access Tools:

- This is categorized into five different groups like
  1. Data Reporting
  2. Query Tools
  3. Application development tools
  4. EIS tools,
  5. OLAP tools and data mining tools.

3.3 Difference between OLTP and OLAP system:

- OLAP (Online Analytical Processing) supports the multidimensional view of data.
- OLAP provides fast, steady, and proficient access to the various view of information
- The complex query can be processed
- It's easy to analyze information by processing complex queries on multidimensional view of data
- Data warehouse is generally used to analyze the information where huge amount of historical data is stored.
- Information in data warehouse is related to more than one dimension like sales, market trends, buying patterns, suppliers etc.

1. Application Differences:

SR.NO	OLTP(Online Transaction Processing)	OLAP(On-Line Analytical Processing)
01	Transactional Oriented	Subject Oriented
02	High Create / Read / Update / Delete (CRUD) activity.	High Read activity
03	Many Users	Few Users
04	Continuous updates – many sources	Batch updates – single sources
05	Real-time information	Historical information
06	Tactical decision – making	Strategic planning
07	Controlled, Customized delivery	"Uncontrolled", generalized delivery
08	Operational database	Informational database

2. Modelling Objective Differences:

SR.NO	OLTP(Online Transaction Processing)	OLAP(On-Line Analytical Processing)
01	High transaction volumes using few records at a time	Low transaction volumes using many records at a time
02	Balancing needs of online v/s scheduled batch processing	Design for on-demand online processing
03	Highly volatile data	Non-volatile data
04	Data redundancy –BAD	Data redundancy –GOOD
05	Few levels of granularity	Multiple levels of granularity
06	Complex database designs used by IT personnel	Simpler database designs with business-friendly construct.

3. Model Differences:

SR.NO	OLTP(Online Transaction Processing)	OLAP(On-Line Analytical Processing)
01	Single purpose model supports Operational system.	Multiple Models – supports International systems
02	Full set of Enterprise data	Subnet of Enterprise data
03	Eliminate redundancy	Plan of redundancy
04	Natural or surrogate keys	Surrogate keys
05	Validate model against business function analysis	Validate model against reporting requirements
06	Technical metadata depends on business requirements	Technical metadata depends on data mapping result
07	This moment is time is important	Many moments in time are essential elements.

4. Big data:

- Big data is broad term for data sets so large or complex that traditional data Processing applications are inadequate.
- Now days the amount of data created by various advanced technologies like social networking sites, E-commerce etc. is very large. It is really difficult to store such huge data by using the traditional data storage facilities.
- Big data means huge amount of data, it is a collection of large datasets that cannot be processed using traditional computing techniques. Big Data is complex and difficult to store, Maintain or access in regular file system, big data becomes a complete subject, which involves different techniques, tools, and frameworks.
- Big Data is described as volumes of data available is changing level of complexity, Produced at different velocities and changing level of ambiguity, that cannot be processed using conventional technologies, processing methods, algorithms, or any commercial off the shelf solution

Sources of Big Data:

- There are various sources of Big Data. Now days in number of fields such huge data get created. Following are the some of the fields.

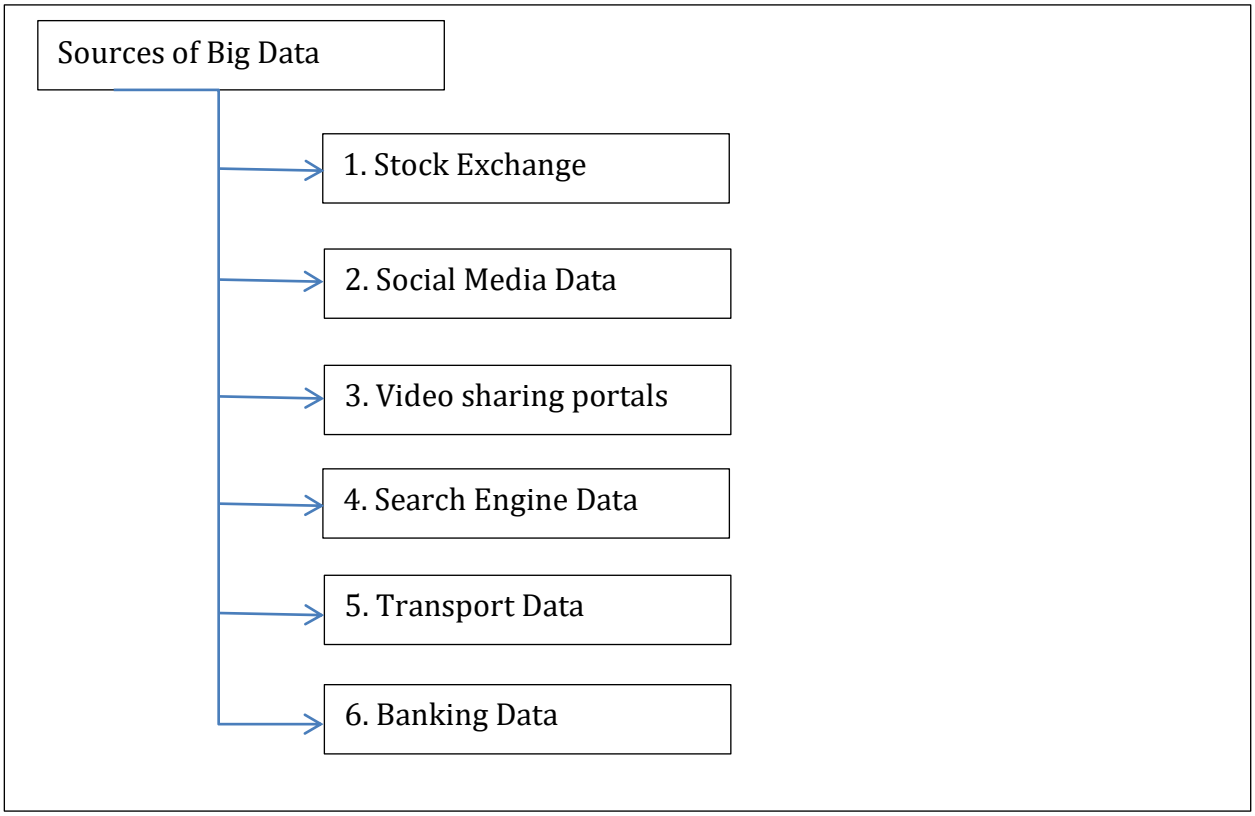


Figure: Sources of Big Data

1. Stock Exchange:

- The data in the share market regarding information about prices and status details of shares of thousands of companies is very huge.

2. Social Media Data:

- The data of social networking sites contains information about all the account holders, their posts, chat history, advertisements etc., on topmost sites like Facebook and WhatsApp, there are literally billons of users.

3. Video sharing portals:

- Video sharing portals like YouTube, Vimeo etc. contains millions of videos each of which requires lots of memory to store.

4. Search Engine Data:

- The search engines like Google and Yahoo holds lot much of metadata regarding various sites.

5. Transport Data:

- Transport data contains information about model, capacity, distance and availability of various vehicles.

6. Banking Data:

- The big giants in banking domains like SBI of ICICI hold large amount of data regarding huge transactions of account holders.

4.2 Characteristics of Big data and consideration:

1. Volume:

- The name Big Data itself is related to a size which is enormous. Size of data plays a very crucial role in determining value out of data. Also, whether a particular data can actually be considered as a Big Data or not, is dependent upon the volume of data.
- **'Volume'** is one characteristic which needs to be considered while dealing with Big Data.

2. Velocity:

- The term **'velocity'** refers to the speed of generation of data. How fast the data is generated and processed to meet the demands, determines real potential in the data.

- Big Data Velocity deals with the speed at which data flows in from sources like business processes, application logs, networks, and social media sites, sensors, Mobile devices, etc. The flow of data is massive and continuous.

3. Variability:

- Hampering the process of being able to handle and manage the data effectively.

4. Variety:

- Variety refers to heterogeneous sources and the nature of data, both structured and unstructured. During earlier days, spread sheets and databases were the only sources of data considered by most of the applications.
- Nowadays, data in the form of emails, photos, videos, monitoring devices, PDFs, audio, etc. are also being considered in the analysis applications.
- This variety of unstructured data poses certain issues for storage, mining and analysing data.

4.3 Benefits of Big data Processing:

- Ability to process Big Data brings in multiple benefits, such as-
  - Businesses can utilize outside intelligence while taking decisions
- Access to social data from search engines and sites like Facebook, twitter are enabling organizations to fine tune their business strategies.
  - Improved customer service
- Traditional customer feedback systems are getting replaced by new systems designed with Big Data technologies. In these new systems, Big Data and natural language processing technologies are being used to read and evaluate consumer responses.
  - Early identification of risk to the product/services, if any
  - Better operational efficiency
- Big Data technologies can be used for creating a staging area or landing zone for new data before identifying what data should be moved to the data warehouse. In addition, such integration of Big Data technologies and data warehouse helps an organization to offload infrequently accessed data.

5. Introduction to Hadoop:

– **Apache Hadoop** is an open-source software framework for storage and large-scale processing of data-sets on clusters of commodity hardware. There are mainly five building blocks inside this runtime environment (from bottom to top):

- Hadoop is open source, java based programming framework which supports the processing and storage of extremely large sets of data in a distributed computing environment using simple programming models.
- Hadoop has very strong processing power and the ability to handle virtually unlimited parallel tasks.
- The cluster is the set of host machines (nodes). Nodes may be partitioned in racks. This is the hardware part of the infrastructure.

5.1. Hadoop Architecture:

Hadoop Has two major layers namely:

1. Processing/Computation layer (Map Reduce), and
2. Storage layer (Hadoop Distributed File System).

- The YARN Infrastructure (Yet another Resource Negotiator) is the framework responsible for providing the computational resources (e.g., CPUs, memory, etc.) needed for application executions. Two important elements are:
-

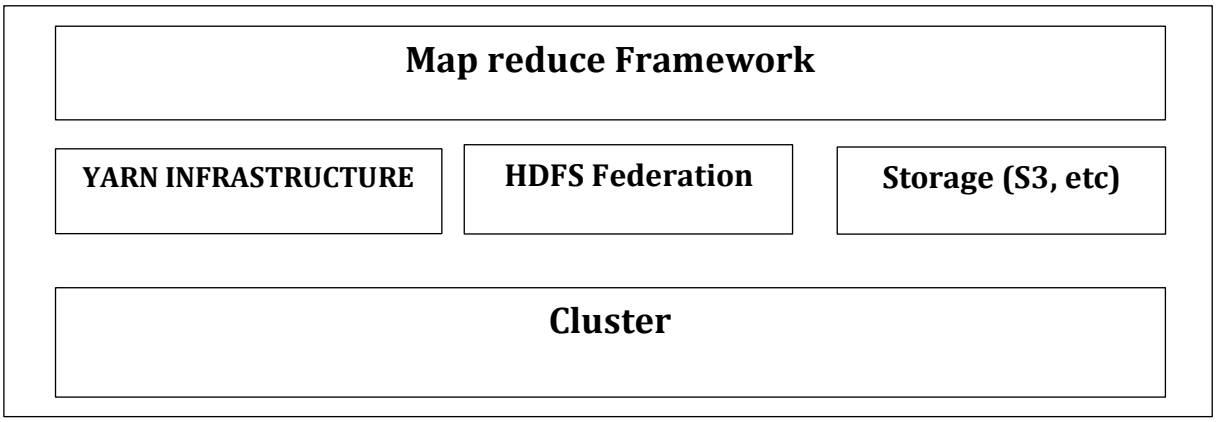


Figure: Architecture of Hadoop

- The **Resource Manager** (one per cluster) is the master. It knows where the slaves are located (Rack Awareness) and how many resources they have.
- It runs several services; the most important is the **Resource Scheduler** which decides how to assign the resources.

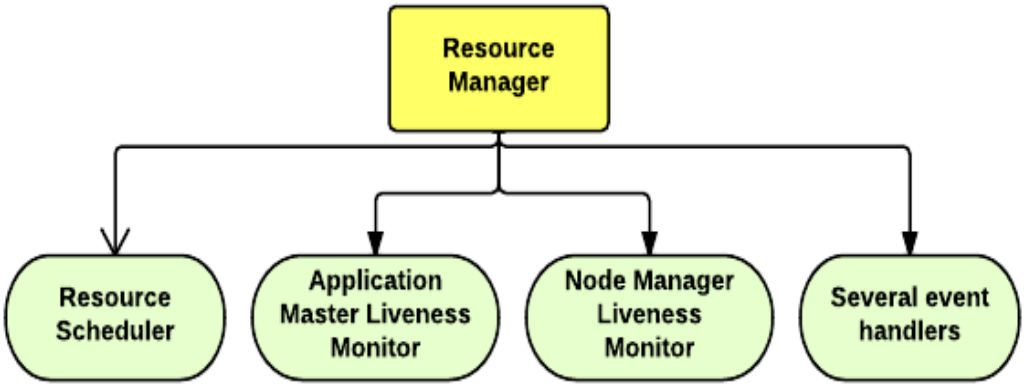
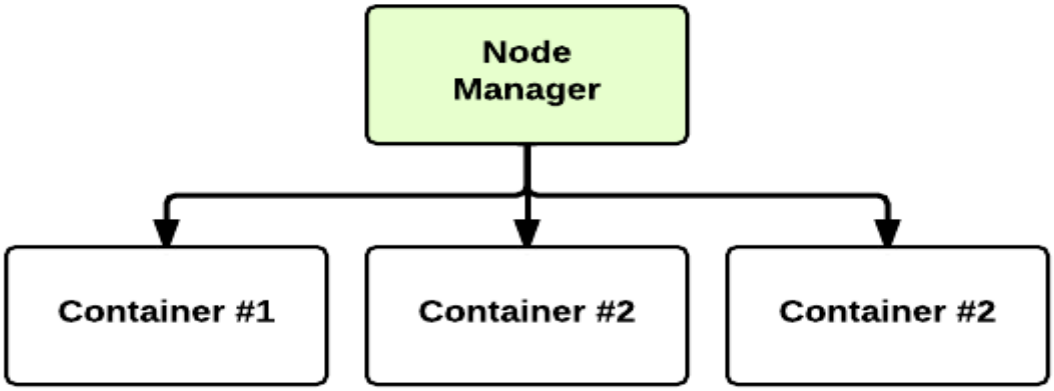


Figure: Resource Manager

- The **Node Manager** (many per cluster) is the slave of the infrastructure. When it starts, it announces himself to the Resource Manager.
- Periodically, it sends an heartbeat to the Resource Manager. Each Node Manager offers some resources to the cluster.
- Its resource capacity is the amount of memory and the number of vcores. At run-time, the Resource Scheduler will decide how to use this capacity: a **Container** is a fraction of the NM capacity and it is used by the client for running a program.



–

- The HDFS Federation is the framework responsible for providing permanent, reliable and distributed storage. This is typically used for storing inputs and output (but not intermediate ones).
- Other alternative storage solutions. For instance, Amazon uses the Simple Storage Service (S3).
- The Map Reduce Framework is the software layer implementing the Map Reduce
- The YARN infrastructure and the HDFS federation are completely decoupled and independent: the first one provides resources for running an application while the second one provides

storage. The Map Reduce framework is only one of many possible frameworks which run on top of YARN (although currently is the only one implemented).

1. Map Reduce:

- MapReduce is a parallel programming model for writing distributed applications devised at Google for efficient processing of large amounts of data (multi-terabyte data-sets),
- on large clusters (thousands of nodes) of commodity hardware in a reliable, fault-tolerant manner.
- The MapReduce program runs on Hadoop which is an Apache open-source framework.

2. Hadoop Distributed File System [HDFS]:

- The Hadoop Distributed File System (HDFS) is based on the Google File System (GFS) and provides a distributed file system that is designed to run on commodity hardware.
- It has many similarities with existing distributed file systems.
- However, the differences from other distributed file systems are significant.
- It is highly fault-tolerant and is designed to be deployed on low-cost hardware.
- It provides high throughput access to application data and is suitable for applications having large datasets. Apart from the above-mentioned two core components.
- Hadoop framework also includes the following two modules:
    - Hadoop Common: These are Java libraries and utilities required by other Hadoop modules.
    - Hadoop YARN: This is a framework for job scheduling and cluster resource management.