Knowledge Management (BCA-504)

UNIT – 1

Business Intelligence and Business Decisions:
Modeling Decision Process
Decision support systems
Group decision support system
Groupware Technologies.

Data, Information, Knowledge, Knowledge management

What do you understand by Decision support system? What are the components of DSS? Discuss the characteristics of DSS.

What do you understand by Group Decision support system? What are the components of GDSS? Discuss the characteristics of GDSS.

What is Executive information system? How does decision support system help in business?

What is groupware technology? Discuss different groupware technologies.

Operational and information system.

What is CSCW?
How is Groupware Design Different from Traditional User Interface Design?

According to Russell Ackoff, a systems theorist and professor of organizational change, the content of the human mind can be classified into five categories:

1. Data: symbols
2. Information: data that are processed to be useful; provides answers to "who", "what", "where", and "when" questions 3. Knowledge: application of data and information; answers "how" questions
4. Understanding: appreciation of "why"
5. Wisdom: evaluated understanding.

Ackoff indicates that the first four categories relate to the past; they deal with what has been or what is known. Only the fifth category, wisdom, deals with the future because it incorporates vision and design. With wisdom, people can create the future rather than just grasp the present and past. But achieving wisdom isn't easy; people must move successively through the other categories.

A further elaboration of Ackoff's definitions follows: Data...

Data is raw. It simply exists and has no significance beyond its existence (in and of itself). It can exist in any form, usable or not. It does not have meaning of itself. In computer parlance, a spreadsheet generally starts out by holding data.

Or

Data

Data is a fact that alone is not significant, as is doesn't relate to other data. Data may answer a very basic ***what*** question; such as a glossary definition, directory entry, or code listing. However, a definition or code may require knowledge, if the definition or code is complex.

Information... information is data that has been given meaning by way of relational connection. This "meaning" can be useful, but does not have to be. In computer parlance, a relational database makes information from the data stored within it.

Or

Information

Information is data that is related and is therefore in context. It can then be transformed into a Process or Procedure, making it useful. Information is data that relates ***who***, ***what***, ***where*** and ***when*** to each other, providing a baseline for a Process (i.e. control point, cycle time) or a Procedure (i.e. date, code, screen description).

Knowledge

Knowledge is the application of information. Knowledge addresses ***how*** and ***why,*** in addition to ***who***, ***what***, ***where*** and ***when***. The knowledge links all the information together to produce a comprehensive Policy, Process or Procedure.

Knowledge allows management to gain an accurate and complete picture of the enterprise Policies, Processes, and Procedures. The Policies, Processes, and Procedures become transformed into an enterprise asset.

Or

Knowledge... knowledge is the appropriate collection of information, such that it's intent is to be useful. Knowledge is a deterministic process. When someone "memorizes" information (as lessaspiring test-bound students often do), then they have amassed knowledge. This knowledge has useful meaning to them, but it does not provide for, in and of itself, an integration such as would infer further knowledge. For example, elementary school children memorize, or amass knowledge of, the "times table". They can tell you that "2 x 2 = 4" because they have amassed that knowledge (it being included in the times table). But when asked what is "1267 x 300", they can not respond correctly because that entry is not in their times table. To correctly answer such a question requires a true cognitive and analytical ability that is only encompassed in the next level... understanding. In computer parlance, most of the applications we use (modeling, simulation, etc.) exercise some type of stored knowledge.

Understanding

Understanding is what increases and supports the transition from data, to information, to knowledge, and to wisdom.

Understanding... understanding is an interpolative and probabilistic process. It is cognitive and analytical. It is the process by which I can take knowledge and synthesize new knowledge from the previously held knowledge. The difference between understanding and knowledge is the difference between "learning" and "memorizing".

Wisdom

Wisdom is complete understanding of the effects and outcomes of Knowledge. Wisdom addresses ***how*** and ***why,*** in addition to ***who***, ***what***, ***where*** and ***when*** at the Enterprise level. Enterprise Policies, Processes, and Procedures must be at this understanding level to be considered permanent, otherwise the Policy and Process may be considered Conditional.

Summary

The following summarizes the use of the Understanding Scale.

	What is represented?	What is answered?	How is it used?	Who is the Audience?
Data	Presents data with no context.	What	Simple glossary, lists, codes	Individuals Departments
Information	Information is data that is useful and has context.	Who, what, where, and when	Simple Processes and Procedures	Individual Department
			Checklists Cheatsheets	Individuals
Knowledge	Knowledge is instructions and know-how.	How and why Who, what, where, and when	Policies	Department
			Processes, and Procedures	Department Business Unit
Wisdom	Wisdom is applying Best Practices, strategic goals, functional alignment, and operational objectives.	Who, what, where, and when, why and how	Policies, Processes, and Procedures	Enterprise level Business Unit
Understanding	Understanding is the level of comprehension and application of principles and concepts.		Research and Develop Policies, Processes, and Procedures	Documentation developer Subject matter expert Management

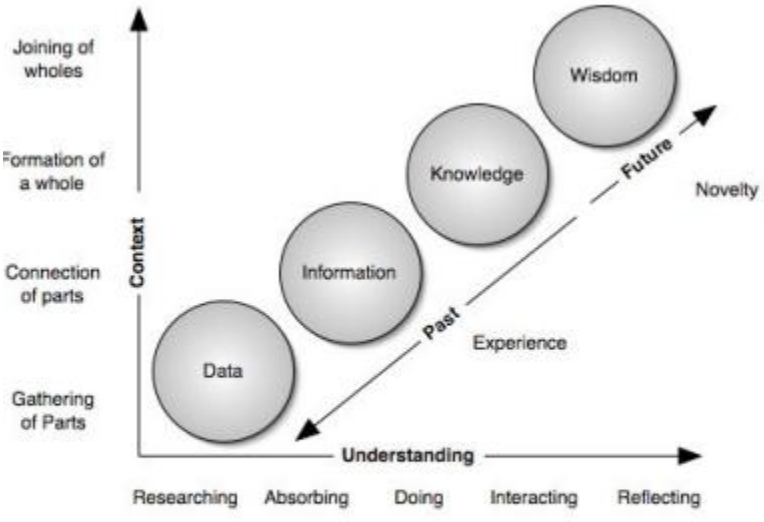
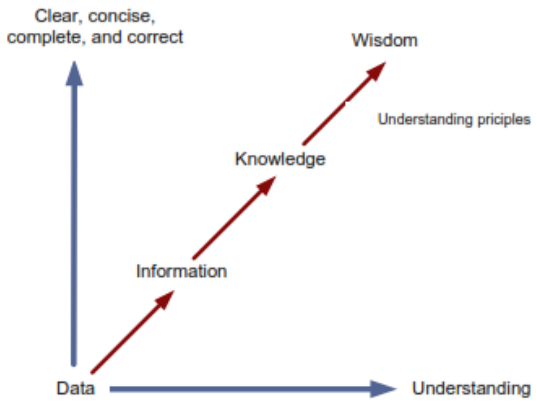


Figure 1: One view of the DIKW hierarchy (Clark, 2004)



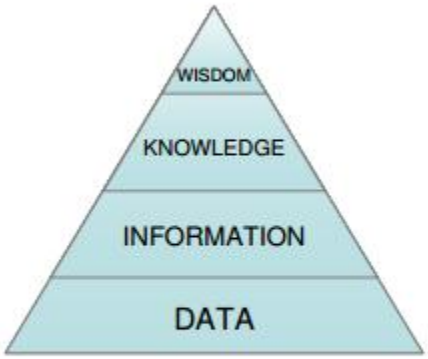
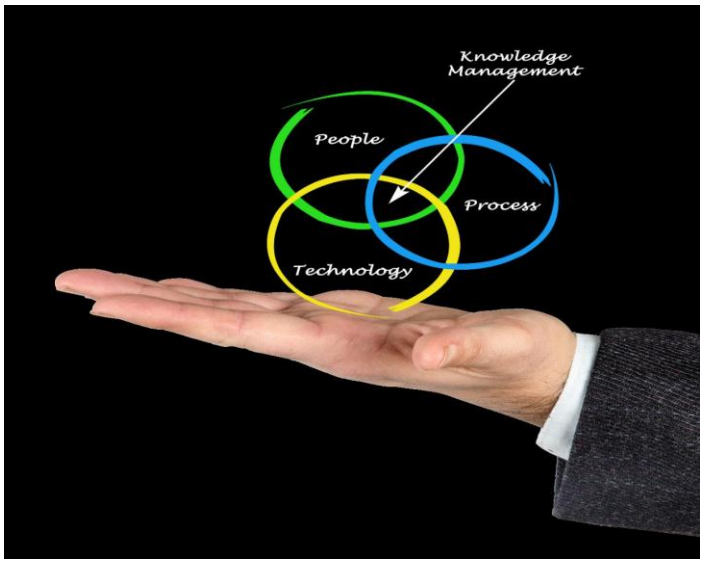


Figure 2: The Knowledge Pyramid

Knowledge management is the systematic management of an organization's knowledge assets for creating value and meeting tactical & strategic requirements. It consists of the initiatives, processes, strategies, and systems that sustain and enhance the storage, assessment, sharing, refinement, and creation of knowledge. Each enterprise should define knowledge management in terms of its own business objectives. Knowledge management is all about applying knowledge in new, previously overburdened or novel situations.



Knowledge Management is a Continuous Cycle

Knowledge management is currently seen as a continuous cycle of three processes, namely:

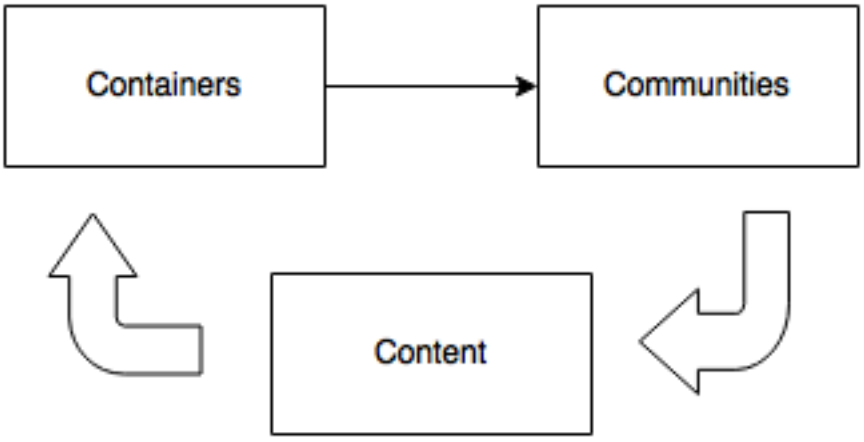
- Knowledge creation and improvement
- Knowledge distribution and circulation
- Knowledge addition and application

Knowledge management expresses a deliberate, systematic and synchronized approach to ensure the full utilization of the company's knowledge base, paired with the potential of individual skills, competencies, thoughts, innovations, and ideas to create a more efficient and effective company.

In simple words, knowledge management incorporates both **holding and storing** of the knowledge perspective, with respect to the intellectual assets.

It is the deliberate and systematic collaboration of an organization's people, technology, processes, style and structure in order to add value through reuse and innovation.

There are three distinct perspectives on Knowledge Management which leads to a different estimation and a different definition.



The Components of Knowledge Management

Knowledge management is a business activity with two primary aspects:

 Executing the knowledge component of business activities as an explicit concern of business in strategy, policy, and practice at all levels of the organization.

 Maintaining a direct link between an organization's intellectual assets both explicit (recorded) and tacit (personal know-how) and positive business results.

Application of Knowledge Management (KM) lie in the below four key areas Knowledge Management

 Globalization of Business: Organizations today are more universal i.e., they are operating in multiple sites, multilingual, and multicultural in nature.

 Leaner Organizations: Organizations are adopting to a lean strategy where they understand customer value and focus on key processes to continuously increase it. The ultimate goal is to provide perfect value to the customer through a perfect value creation process that has zero waste.

 Corporate Amnesia: We are freer as a workforce, which creates issues regarding knowledge continuity for the organization and places with continuous learning demands from knowledge worker. We no longer expect to spend our entire work life with the same organization.

 Technological Advances: The world is more connected with the advent of websites, smart phones and other latest gadgets. Advancements in technology has not only helped in better connectivity but also changed expectations. Companies are expected to have online presence around the clock providing required information as per the customer needs.

Knowledge Management serves as one of the major response to the challenge of trying to handle this complex, information overloaded work environment. As such, Knowledge management is perhaps best clustered as a science of complexity.

What do you understand by Decision support system? What are the components of DSS? Discuss the characteristics of DSS.

Decision Support System refers to a class of systems which support in the process of decision making and does not always give a decision itself. Decision Support Systems (DSS) are a specific class of computerized information system that supports business and organizational decision making activities.

A properly designed DSS is an interactive software based system intended to help decision makers compile useful information from raw data, documents, personal knowledge, and/or business models to identify and solve problems and make decisions

DSS is an application of Hebert Simon model, as discussed, the model has three phases**:**

i) Intelligence

ii) Design

iii) Choice

The DSS basically helps in the information system in the intelligence phase where the objective is to identify the problem and then go to the design phase for solution. The choice of selection criteria varies from problem to problem.

It is therefore, required to go through these phases again and again till satisfactory solution is found. In the following three phase cycle, you may use inquiry, analysis, and models and accounting system to come to rational solution.

These systems are helpful where the decision maker calls for complex manipulation of data and use of several methods to reach an acceptable solution using different analysis approach. The decision support system helps in making a decision and also in performance analysis. DSS can be built around the rule in case of programmable decision situation. The rules are not fixed or predetermined and requires every time the user to go through the decision making cycle as indicated in Herbert Simon model.

Attributes :

i) DSS should be adaptable and flexible.

ii) DSS should be interactive and provide ease of use.

iii) Effectiveness balanced with efficiency (benefit must exceed cost).

iv) Complete control by decision-makers.

v) Ease of development by (modification to suit needs and changing environment) end users.

vi) Support modeling and analysis.

vii) Data access.

viii) Standalone, integration and Web-based

DSS Characteristics :

i) Support for decision makers in semi structured and unstructured problems.

ii) Support managers at all levels.

iii) Support individuals and groups.

iv) Support for interdependent or sequential decisions.

v) Support intelligence, design, choice, and implementation.

vi) Support variety of decision processes and styles

Modern classification of DSS are

1. **Model Driven DSS** is a DSS that uses a model (quantitative) based on heuristics, optimization, simulation etc. for deriving solutions to problems. It has access to the models and has flexibility of changing the parameters of the model. Real data or transactional data from databases of TPS is then passed through the model to arrive at the solution. The system is capable of producing different scenarios.
2. **Data Driven DSS** is a DSS that gives access to time-series internal data. Data ware houses that have tools that provide facility to manipulate such data are examples of advances systems. Executive Information Systems are examples of data-driven DSS.
3. **Communications-driven DSS** is a DSS that uses network and communications technologies to support decision-relevant collaboration and communication. In such systems, communication technologies are the most important component.
4. **Document-driven DSS** is a DSS that uses computer storage and processing to provide document retrieval and analysis.
5. **Knowledge-driven** DSS is a DSS that collects and stores 'expertise' so that it can be used for decision-making when required.

Components of DSS

Even though DSS can be of several types, fundamentally each DSS will have the following components:

- **Interactive User**-System Dialog Management Subsystem-DSS requires continuous user interaction. Sometimes the system should prompt the user to give an input at other time the user should be able to control the processing. A typical user system dialog management subsystem will have the following elements:

1. **User Interface** - the user interface of a DSS has to be dynamic and GUI based. It has to be an easy to use user interface as most of the people who will be using it are not technical experts but management experts (top management) and hence the interface should be minimalist in design. also the system should be able to interact with the user in a interactive mode and hence the user interface has to be dynamic.

2. **Request Constructor** - since DSS works on an interactive dynamic mode, it needs a request constructor (incorporating aspects of Language Query Interface) which can convert the user's instructors into model understandable form, the model's data request to the database and the model's instructions/requests to the user.

- **Data Management Subsystem** - data is the most important component of a DSS. Without the data a DSS cannot function. The data management subsystem manages the data for DSS. Data is accessed in a DSS in many ways like ad hoc basis, structured query basis and heuristic search basis and hence a strong data management subsystems is required to service the varied data requests from a DSS. The subsystem has the following elements:

1. **Database Management System** - it is the data store for the DSS. It manages the data and performs all the functions that a typical DBMS package does. In fact, in most DSS a commercial DBMS or RDBMS package is used to perform this task.
2. **The Query Control** - this is a tailored element to handle the query requirements of DSS. It may connect the database, directly to the user interface or to the model base or both.
3. **Meta Data** - this contains data about the data that is stored in the database. This helps the DSS in understanding the data in the database properly and helps in creating ad hoc queries.

- **Model Management Subsystem** - this is the unique feature of a DSS. This makes the system special. However, this also makes the system very specific. There are very few examples of a generalized DSS as generalized models are not available. Those that exist work on half baked solutions. The model management subsystem may use different classes of models like,

Optimization Models, Simulation Models, Heuristic Models, Deterministic Models, Predictive Models

Each class of model is useful to solve a specific class of problems like a routing problem or a scheduling problem or a combinatorial search problem etc. Model and Model Management has several connotations in DSS literature and there have been wide ranging definitions of these terms.

What is Executive information system? How does decision support system help in business?

Executive support systems are intended to be used by the senior managers directly to provide support to non-programmed decisions in strategic management.

These information are often external, unstructured and even uncertain. Exact scope and context of such information is often not known beforehand.

This information is intelligence based:

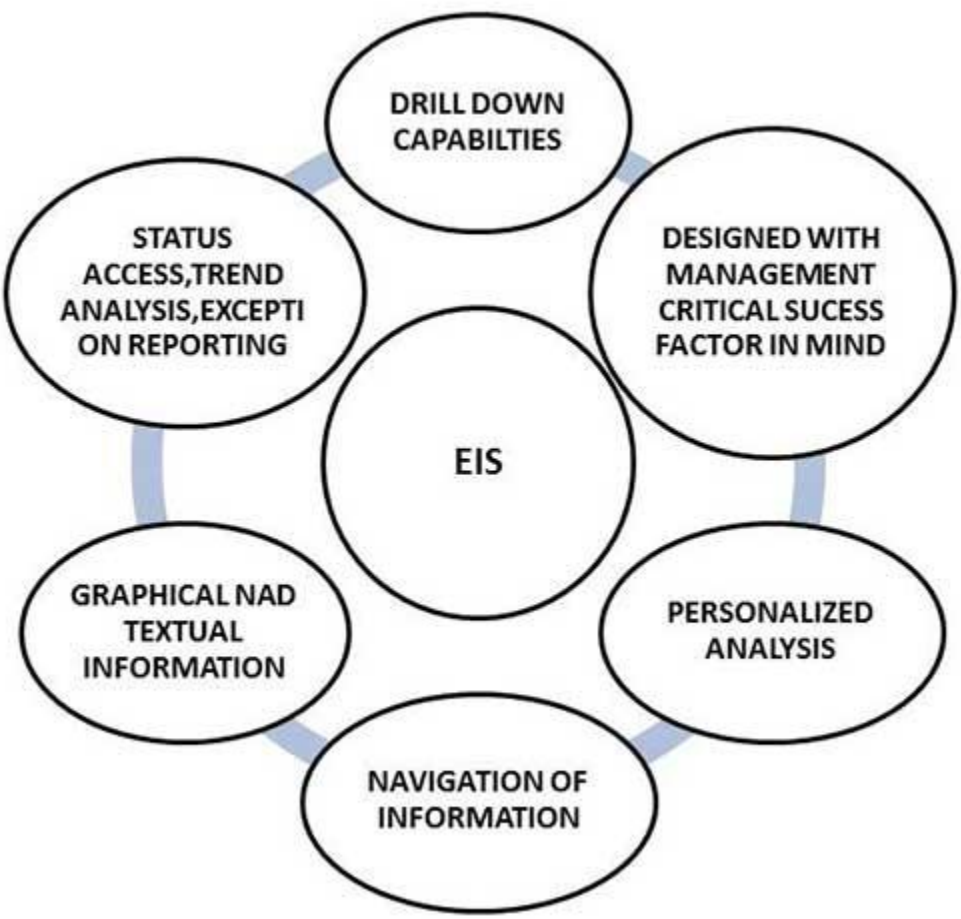
- Market intelligence
- Investment intelligence
- Technology intelligence

Examples of Intelligent Information

Following are some examples of intelligent information, which is often the source of an EIS:

- External databases
- Technology reports like patent records etc.
- Technical reports from consultants
- Market reports
- Confidential information about competitors
- Speculative information like market conditions
- Government policies
- Financial reports and information

Features of Executive Information System



Advantages of EIS

- Easy for upper level executive to use
- Ability to analyze trends
- Augmentation of managers' leadership capabilities
- Enhance personal thinking and decision-making
- Contribution to strategic control flexibility

- Enhance organizational competitiveness in the market place
- Instruments of change
- Increased executive time horizons.
- Better reporting system
- Improved mental model of business executive
- Help improve consensus building and communication
- Improve office automation
- Reduce time for finding information
- Early identification of company performance
- Detail examination of critical success factor
- Better understanding
- Time management
- Increased communication capacity and quality

Disadvantage of EIS

- Functions are limited
- Hard to quantify benefits
- Executive may encounter information overload
- System may become slow
- Difficult to keep current data
- May lead to less reliable and insecure data
- Excessive cost for small company

What do you understand by Group Decision support system? What are the components of GDSS? Discuss the characteristics of GDSS.

In general, a Group Decision Support System is a computer-based information system that is used to improve an organization's group decision making. The group decision support systems is similar to a Decision Support System in that it supports the three basic functions of data, model, and dialogue management. However, unlike an individual decision support system (DSS), a group decision support systems must interact with two or more users through a communications subsystem. Four aspects of group decision support systems that may be used to describe a particular system are listed.

A Group Decision Support System involves: · Face-to-face or non-face-to-face communication Synchronous or nonsynchronous communications (i.e. communication takes place within a single interactive session or at different times through, for example, electronic mail). ·

Close or dispersed geography (i.e., decision makers are at the same physical location or at two or more separate locations). ·

Cooperative or uncooperative (e.g., a negotiation situation) atmosphere

Group decision support systems (GDSSs), a subclass of DSSs, are defined as information technology-based support systems that provide decision-making support to groups. They refer to the systems that provide computer-based aids and communication support for decision-making meetings in organizations. The group meeting is a joint activity in which a group of people is engaged with equal or near-equal status. The activity and its outputs are intellectual in nature. Essentially, the outputs of the meeting depend on the knowledge and judgment contributed by the participants.
Differences in opinion may be settled by negotiation or arbitration.

Components of GDSS

The difference between GDSSs and DSSs is the focus on the group versus the individual decision-maker. The components of a GDSS are basically similar to those of DSS, including hardware, software, and people; but in addition, within the collaborative environment, communication and networking technologies are added for group participation from different sites. Moreover, compared with DSSs, GDSSs designers pay more attention to the user/system interface with multi-user access and system reliability because a system failure will affect a multi-user group, rather than just an individual.
There are three fundamental types of components that compose GDSSs:

1. **Software**

The software part may consist of the following components: databases and database management capabilities, user/system interface with multi-user access, specific applications to facilitate group decision-makers' activities, and modeling
capabilities.

2**. Hardware**

he hardware part may consist of the following components: I/O devices, PCs or workstations, individual monitors for each participant or a public screen for group, and a network to link participants to each other.

3. **People**

he people may include decision-making participants and /or facilitator. A facilitator is a person who directs the group through the planning process.

Benefits claimed for GDSS

There are three benefits claimed for GDSSs: increased efficiency, improved quality, and leverage that improves the way meetings run.
Due to increasing computer data processing power, communication and network performance, the speed and quality for information processing and information transmission create the opportunity for higher efficiency. Efficiency achievement
depends on the performance of hardware (e.g., PCs, LAN/WAN) and software. With regard to the software aspect of GDSSs, the software architecture with database management and an interactive interface affects system run time efficiency and performance. Improved quality of the outcomes of a group meeting implies the increased quality of alternatives examined, greater participation and contribution from people who would otherwise be silent, or decision outcomes judged to be of higher quality. In a GDSS, the outcome of a meeting or decision-making process depends on communication facilities and decision support facilities. Those facilities can help decision-making participants avoid the constraints imposed by geography. They also make information sharable and reduce effort in the decision-making process. Therefore, those facilities contribute to meeting quality improvement. Leverage implies that the system does not merely speed up the process (say efficiency), but

changes it fundamentally. In other words, leverage can be achieved through providing better ways of meeting, such as providing the ability to execute multiple tasks at the same time.

Factors that affect GDSS

Research indicates there are usually several factors affecting GDSSs,

- Anonymity
- Facility design
- Multiple public screens
- Knowledge bases and databases
- Communication network speed
- Fixed versus customized methodology
- Software design
- Group size and composition
- Satisfaction

Information needs of groups

It is fundamental and important to clearly understand what groups do and which of their activities and procedures can be and should be supported by GDSSs. Also, it is necessary to know the information needs of groups and examine how best to support these information uses with GDSSs. The information needs of groups covers a broad spectrum.

. Database access

Databases are one of the basic components of GDSSs. GDSSs offer groups the advantage of accessing databases or some on-line service for the lastest information. The databases can be internal or external databases. This is a key element in information retrieval and sharing in a group meeting. The requirements on the presentation and functions of the obtained information can be summarized as follows. information should be presented in clear and familiar ways, information presentation and all other associated management control aspects should assist the decision-maker to guide the process of judgment and choice; with an explanation facility, information containing an advice or decision suggestion enables users to know how and why results and advice are obtained; and information should be helpful to improve the precision of task situation understanding. Moreover, information needs are based on the identification of the information requirements for the particular situation.

- **Information creation**

In addition to a decision, the output of the meeting is new information. In a GDSS, all input into the computer is usually captured. In some cases, the actions of individual members of the group are stored in a database, file or some other storage format. Making a decision is not a point-event. The decision is produced based on valuable knowledge. It is worthwhile to save the valuable information in efficient ways which make it convenient for further use.

- **Dissemination of information, decisions, and responsibilities**

An often-cited advantage of GDSSs is that the participants are allowed to know what new information was created, what decision was reached, and who is responsible for follow-up or for implementing decisions.

- **On-line modeling**

On-line modeling is the next step beyond sharing existing data. For example, the participants can perform on-line analysis and send out their results or ideas to a public board.

- **Visual decision-making**

Some decisions involve visuals rather than words or numbers. Intuitively, graphics with shape, size, and color, might make it easier and faster for users to have an overall view of the information.

- **Multimedia information presentation**

The combination of visible and audible information presentation format impacts the traditional information presentation format. The benefits of multimedia presentation include better interaction, more straightforward and effective communication in the group, and decreased learning time.

- **Idea generation**

A variety of idea generation packages or methods exists for GDSSs use.

- **Voting**

This implies the ability to vote, rank, or rate.

GDSSs have an impact on the work of individuals, groups, and organizations. In general, the performance improvement and satisfaction of individuals will lead to the improvement of the group. Both hardware and software will influence GDSSs. For example, the performance of a network will directly affect data transmission. If the network slows down, it will constrain the GDSS's capability of on-line data processing. Video and audio devices are adopted to make it more straightforward for users to recognize multimedia information, which results in the improvement of efficiency and in
effectiveness, as well as in the quality of meeting outcomes. Hardware development and innovation are significant for GDSSs performance. Software is another factor that has an impact on GDSSs performance.

Characteristics of a GDSS

A GDSS has a number of unique characteristics to support a group of participants in their decision-making process:

- Special design to support creative thinking, effective communications and decision-making techniques
- Easy to use so participants from different backgrounds can all participate effectively
- Flexible so it can incorporate the different perspectives and decision-making styles of the different participants
- Automated record keeping for future review and analysis
- Parallel communication to allow multiple participants to contribute simultaneously

The most important characteristic, however, is that it provides support for a group to come to a decision. A number of different approaches can be used.

The **Delphi approach** is a structured communication technique in which experts answer questionnaires in several rounds. Participants get to see each other's input after each round and are encouraged to revise their earlier answers based on replies by others. The underlying assumption is that this type of communication will lead to a convergence toward a final answer. This approach is often used when conditions are highly uncertain and the opinion of experts is weighed heavily. A good example of this would be trying to predict the sales of a particular new product. If there are no similar products on the market, there may not be a lot of data for analysis.

The **group consensus approach** forces members to come to a unanimous decision. This is sort of like locking a team up in a room, and they can't leave before a decision is reached - but the room could be virtual, and the communications could all be electronic.

The **nominal group technique** gives each participant an equal voice, and the final decision is reached by voting. Contrary to regular voting, however, the group comes up with a number of different solutions, and these are ranked by using a voting process. Whatever the specific decision-making strategy employed, a GDSS is designed to facilitate this process.

What is CSCW?

CSCW (Computer-Supported Cooperative Work) refers to the field of study which examines the design, adoption, and use of groupware. Despite the name, this field of study is not restricted to issues of "cooperation" or "work" but also examines competition, socialization, and play. The field typically attracts those interested in software design and social and organizational behavior, including business people, computer scientists, organizational psychologists, communications researchers, and anthropologists, among other specialties.

What is Groupware?

Groupware is technology designed to facilitate the work of groups. This technology may be used to communicate, cooperate, coordinate, solve problems, compete, or negotiate. While traditional technologies like the telephone qualify as groupware, the term is ordinarily used to refer to a specific class of technologies relying on modern computer networks, such as email, newsgroups, videophones, or chat. Groupware technologies are typically categorized along two primary dimensions:

1. whether users of the groupware are working together at the same time ("realtime" or "synchronous" groupware) or different times ("asynchronous" groupware), and
2. whether users are working together in the same place ("co- located" or "face-to-face") or in different places ("non-co-located" or "distance").

Groupware offers significant advantages over single-user systems. These are some of the most common reasons people want to use groupware:

- to facilitate communication: make it faster, clearer, more persuasive
- to enable communication where it wouldn't otherwise be possible
- to enable telecommuting
- to cut down on travel costs
- to bring together multiple perspectives and expertise
- to formgroups with common interests where it wouldn't be possible to gather a sufficient number of people face-to-face
- to save time and cost in coordinating group work
- to facilitate group problem-solving
- to enable new modes of communication, such as anonymous interchanges or structured interactions

In addition to the benefits of groupware, another good reason to study usability and design issues in groupware is to avoid a failed design. Groupware is significantly more difficult to get right than traditional software. Typically, a groupware system can't succeed unless most or all of the target group is willing to adopt the system. In contrast, a single-user system can be successful even if only a fraction of the target market adopts it.

Applications of groupware technology:

There are several types of groupware applications. Comparing those design options across applications yields interesting new perspectives on well-known applications. Also, in many cases, these systems can be used together, and in fact, are intended to be used in conjunction. For example, group calendars are used to schedule videoconferencing meetings, multi-player games use live video and chat to communicate, and newsgroup discussions spawn more highly-involved interactions
in any of the other systems.

Asynchronous Groupware

Email is by far the most common groupware application (besides, of course, the traditional telephone). While the basic technology is designed to pass simple messages between 2 people, even relatively basic email systems today typically include interesting features for forwarding messages, filing messages, creating mailing groups, and attaching files with a message. Other features that have been explored include: automatic sorting and processing of messages, automatic routing, and structured communication (messages requiring certain information).

Newsgroups and mailing lists are similar in spirit to email systems except that they are intended for messages among large groups of people instead of 1-to-1 communication. In practice the main difference between newsgroups and mailing lists is that newsgroups only show messages to a user when they are explicitly requested (an "on-demand" service), while mailing lists deliver messages as they become available (an "interrupt-driven" interface). **Workflow systems** allow documents to be routed through organizations through a relatively-fixed process. A simple example of a workflow
application is an expense report in an organization: an employee enters an expense report and submits it, a copy is archived then routed to the employee's manager for approval, the manager receives the document, electronically approves it and sends it on and the expense is registered to the group's account and forwarded to the accounting department for payment. Workflow systems may provide features such as routing, development of forms, and support for differing roles and privileges.
Hypertext is a system for linking text documents to each other, with the Web being an obvious example. Whenever multiple people author and link documents, the system becomes group work, constantly evolving and responding to others' work. Some hypertext systems include capabilities for seeing who else has visited a certain page or
link, or at least seeing how often a link has been followed, thus giving users a basic awareness of what other people are doing in the system -- page counters on the Web are a crude approximation of this function. Another common multi-user feature in hypertext (that is not found on the Web) is allowing any user to create links from any page, so that others can be informed when there are relevant links that the original author was unaware of.

Group calendars allow scheduling, project management, and coordination among many people, and may provide support for scheduling equipment as well. Typical features detect when schedules conflict or find meeting times that will work for everyone. Group calendars also help to locate people. Typical concerns are privacy (users may feel that certain activities are not public matters), completeness and accuracy (users may feel that the time it takes to enter schedule information is not justified by the benefits of the calendar).

Collaborative writing systems may provide both real time support and non-real time support. Word processors may provide asynchronous support by showing authorship and by allowing users to track changes and make annotations to documents. Authors collaborating on a document may also be given tools to help plan and coordinate the authoring process, such as methods for locking parts of the document or linking separately-authored documents. Synchronous support allows authors to see each other's changes as they make them, and usually needs to provide an additional communication channel to the authors as they work (via videophones or chat).

Synchronous or Real time Groupware
Shared whiteboards allow two or more people to view and draw on a shared drawing surface even from different locations. This can be used, for instance, during a phone call, where each person can jot down notes

(e.g., a name, phone number, or map) or to work collaboratively on a visual problem. Most shared whiteboards are designed for informal
conversation, but they may also serve structured communications or more sophisticated drawing tasks, such as collaborative graphic design, publishing, or engineering applications. Shared whiteboards can indicate where each person is drawing or pointing by showing telepointers, which are color-coded or labeled to identify each person.

Video communications systems allow two-way or multi-way calling with live video, essentially a telephone system with an additional visual component. Cost and compatibility issues limited early use of video systems to scheduled videoconference meeting rooms. Video is advantageous when visual information is being discussed, but may not provide
substantial benefit in most cases where conventional audio telephones are adequate. In addition to supporting conversations, video may also be used in less direct collaborative situations, such as by providing a view of activities at a remote location.

Chat systems permit many people to write messages in real time in a public space. As each person submits a message, it appears at the bottom of a scrolling screen. Chat groups are usually formed by having a listing of chat rooms by name, location, number of people, topic of discussion, etc. Many systems allow for rooms with controlled access or with moderators to lead the discussions, but most of the topics of interest to researchers involve issues related to un-moderated real time communication including: anonymity, following the stream of conversation, scalability with number of users, and abusive users. While chat-like systems are possible using non-text media, the text version of chat has the rather interesting aspect of having a direct transcript of the conversation, which not only has long-term value, but allows for backward
reference during conversation making it easier for people to drop into a conversation and still pick up on the ongoing discussion.

Decision support systems are designed to facilitate groups in decision making. They provide tools for brainstorming, critiquing ideas, putting weights and probabilities on events and alternatives, and voting. Such systems enable presumably more rational and even-handed decisions. Primarily designed to facilitate meetings, they encourage equal participation
by, for instance, providing anonymity or enforcing turn-taking.

Multi-player games have always been reasonably common in arcades, but are becoming quite common on the internet. Many of the earliest electronic arcade games were multi-user, for example, Pong, Space Wars, and car racing games. Games are the prototypical example of multi-user situations "non-cooperative", though even competitive games require
players to cooperate in following the rules of the game. Games can be enhanced by other communication media, such as chat or video systems.

How is Groupware Design Different from Traditional User Interface Design?
Groupware design involves understanding groups and how people behave in groups. It also involves having a good understanding of networking technology and how aspects of that technology (for instance, delays in synchronizing views) affect a user's experience. All the issues related to traditional user interface design remain relevant, since the technology still involves people. However, many aspects of groups require special consideration. For instance, not only do million-person groups behave differently from 5-person groups, but the

performance parameters of the technologies to support different groups vary. Ease-of-use must be better for groupware than for single-user systems because the pace of use of an application is often driven by the pace of a conversation.

UNIT -II

Executive Information and support Systems:
Business Expert System and AI,
OLTO & OLAP
Data Warehousing
Data Marts,
Data Warehouse architecture
Tools for data warehousing

Business Expert System and Artificial Intelligence

Expert Systems

are computer programs that are derived from a branch of computer science research called *Artificial Intelligence* (AI). AI's scientific goal is to understand intelligence by building computer programs that exhibit intelligent behavior. It is concerned with the concepts and methods of symbolic inference, or reasoning, by a computer, and how the knowledge used to make those inferences will be represented inside the machine.

Of course, the term *intelligence* covers many cognitive skills, including the ability to solve problems, learn, and understand language; AI addresses all of those. But most progress to date in AI has been made in the area of problem solving -- concepts and methods for building programs that *reason* about problems rather than calculate a solution.

AI programs that achieve expert-level competence in solving problems in task areas by bringing to bear a body of knowledge about specific tasks are called *knowledge-based* or *expert systems*. Often, the term expert systems is reserved for programs whose knowledge base contains the knowledge used by human experts, in contrast to knowledge gathered from textbooks or non-experts. More often than not, the two terms, expert systems (ES) and knowledge-based systems (KBS), are used synonymously. Taken together, they represent the most widespread type of AI application. The area of human intellectual endeavor to be captured in an expert system is called the *task domain*. *Task* refers to some goal-oriented, problem-solving activity. *Domain* refers to the area within which the task is being performed. Typical tasks are diagnosis, planning, scheduling, configuration and design. An example of a task domain is aircraft crew scheduling.

Building an expert system is known as *knowledge engineering* and its practitioners are called *knowledge engineers*. The knowledge engineer must make sure that the computer has all the knowledge needed to solve a problem. The knowledge engineer must choose one or more forms in which to represent the required knowledge as symbol patterns in the memory of the computer -- that is, he (or she) must choose a *knowledge representation*. He must also ensure that the computer can use the knowledge efficiently by selecting from a handful of *reasoning methods*. The practice of knowledge engineering is described later. We first describe the components of expert systems.

The Building Blocks of Expert Systems

Every expert system consists of two principal parts: the knowledge base; and the reasoning, or inference, engine.

The *knowledge base* of expert systems contains both factual and heuristic knowledge. *Factual knowledge* is that knowledge of the task domain that is widely shared, typically found in textbooks or journals, and commonly agreed upon by those knowledgeable in the particular field.

Heuristic knowledge is the less rigorous, more experiential, more judgmental knowledge of performance. In contrast to factual knowledge, heuristic knowledge is rarely discussed, and is largely individualistic. It is the knowledge of good practice, good judgment, and plausible reasoning in the field. It is the knowledge that underlies the "art of good guessing."

Knowledge representation formalizes and organizes the knowledge. One widely used representation is the *production rule*, or simply *rule*. A rule consists of an IF part and a THEN part (also called a *condition* and an *action*). The IF part lists a set of conditions in some logical combination. The piece of knowledge represented by the production rule is relevant to the line of reasoning being developed if the IF part of the rule is satisfied; consequently, the THEN part can be concluded, or its problem-solving action taken. Expert systems whose knowledge is represented in rule form are called *rule-based systems*.

Another widely used representation, called the *unit* (also known as *frame, schema*, or *list structure*) is based upon a more passive view of knowledge. The unit is an assemblage of associated symbolic knowledge about an entity to be represented. Typically, a unit consists of a list of properties of the entity and associated values for those properties.

Since every task domain consists of many entities that stand in various relations, the properties can also be used to specify relations, and the values of these properties are the names of other units that are linked according to the relations. One unit can also represent knowledge that is a "special case" of another unit, or some units can be "parts of" another unit.

The *problem-solving model*, or *paradigm*, organizes and controls the steps taken to solve the problem. One common but powerful paradigm involves chaining of IF-THEN rules to form a line of reasoning. If the chaining starts from a set of conditions and moves toward some conclusion, the method is called *forward chaining*. If the conclusion is known (for example, a goal to be achieved) but the path to that conclusion is not known, then reasoning backwards is called for, and the method is *backward chaining*. These problem-solving methods are built into program modules called *inference engines* or *inference procedures* that manipulate and use knowledge in the knowledge base to form a line of reasoning.

The *knowledge base* an expert uses is what he learned at school, from colleagues, and from years of experience. Presumably the more experience he has, the larger his store of knowledge. Knowledge allows him to interpret the information in his databases to advantage in diagnosis, design, and analysis.

Though an expert system consists primarily of a knowledge base and an inference engine, a couple of other features are worth mentioning: reasoning with uncertainty, and explanation of the line of reasoning.

Knowledge is almost always incomplete and uncertain. To deal with uncertain knowledge, a rule may have associated with it a *confidence factor* or a weight. The set of methods for using uncertain knowledge in combination with uncertain data in the reasoning process is called *reasoning with uncertainty*. An important subclass of methods for reasoning with uncertainty is called "fuzzy logic," and the systems that use them are known as "fuzzy systems."

Because an expert system uses uncertain or heuristic knowledge (as we humans do) its credibility is often in question (as is the case with humans). When an answer to a problem is questionable, we tend to want to know the

rationale. If the rationale seems plausible, we tend to believe the answer. So it is with expert systems. Most expert systems have the ability to answer questions of the form: "Why is the answer X?" Explanations can be generated by tracing the line of reasoning used by the inference engine (Feigenbaum, McCorduck et al. 1988).

The most important ingredient in any expert system is knowledge. The power of expert systems resides in the specific, high-quality knowledge they contain about task domains. AI researchers will continue to explore and add to the current repertoire of knowledge representation and reasoning methods. But in knowledge resides the power. Because of the importance of knowledge in expert systems and because the current knowledge acquisition method is slow and tedious, much of the future of expert systems depends on breaking the knowledge acquisition bottleneck and in codifying and representing a large knowledge infrastructure.

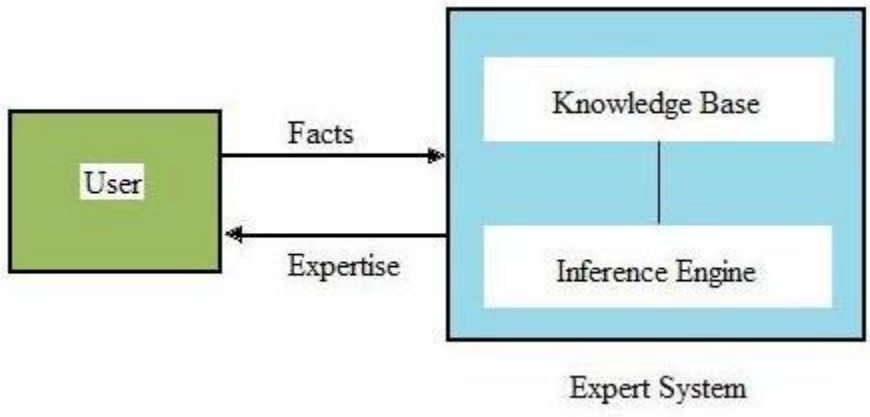
Knowledge engineering

is the art of designing and building expert systems, and knowledge engineers are its practitioners. Gerald M. Weinberg said of programming in *The Psychology of Programming*: "'Programming,' -- like 'loving,' -- is a single word that encompasses an infinitude of activities" (Weinberg 1971). Knowledge engineering is the same, perhaps more so. We stated earlier that knowledge engineering is an applied part of the science of artificial intelligence which, in turn, is a part of computer science. Theoretically, then, a knowledge engineer is a computer scientist who knows how to design and implement programs that incorporate artificial intelligence techniques. The nature of knowledge engineering is changing, however, and a new breed of knowledge engineers is emerging. We'll discuss the evolving nature of knowledge engineering later.

Today there are two ways to build an expert system. They can be built from scratch, or built using a piece of development software known as a "tool" or a "shell." Before we discuss these tools, let's briefly discuss what knowledge engineers do. Though different styles and methods of knowledge engineering exist, the basic approach is the same: a knowledge engineer interviews and observes a human expert or a group of experts and learns what the experts know, and how they reason with their knowledge. The engineer then translates the knowledge into a computer-usable language, and designs an inference engine, a reasoning structure, that uses the knowledge appropriately. He also determines how to integrate the use of uncertain knowledge in the reasoning process, and what kinds of explanation would be useful to the end user.

COMPONENTS OF AN EXPERT SYSTEM:

The expert system consists of two major components: knowledge base and inference engine.



Knowledge base contains the domain knowledge which is used by the inference engine to draw conclusions.
The inference engine is the generic control mechanism that applies the axiomatic knowledge to the task-specific

data to arrive at some conclusion. When a user supplies facts or relevant information of query to the expert system he receives advice or expertise in response. That is given the facts it uses the inference engine which in turn uses the knowledge base to infer the solution.

CHARACTERISTICS OF EXPERT SYSTEMS

High performance: They should perform at the level of a human expert.

Adequate response time: They should have the ability to respond in a reasonable amount of time. Time is crucial especially for real time systems.

Reliability: They must be reliable and should not crash.

Understandable: They should not be a black box instead it should be able explain the steps of the reasoning process. It should justify its conclusions in the same way a human expert explains why he arrived at particular conclusion.

THE APPLICATIONS OF EXPERT SYSTEMS

The spectrum of applications of expert systems technology to industrial and commercial problems is so wide as to defy easy characterization. The applications find their way into most areas of knowledge work. They are as varied as helping salespersons sell modular factory-built homes to helping NASA plan the maintenance of a space shuttle in preparation for its next flight.

Applications tend to cluster into seven major classes.

Diagnosis and Troubleshooting of Devices and Systems of All Kinds

This class comprises systems that deduce faults and suggest corrective actions for a malfunctioning device or process. Medical diagnosis was one of the first knowledge areas to which ES technology was applied (for example, see Shortliffe 1976), but diagnosis of engineered systems quickly surpassed medical diagnosis. There are probably more diagnostic applications of ES than any other type. The diagnostic problem can be stated in the abstract as: given the evidence presenting itself, what is the underlying problem/reason/cause?

Planning and Scheduling

Systems that fall into this class analyze a set of one or more potentially complex and interacting goals in order to determine a set of actions to achieve those goals, and/or provide a detailed temporal ordering of those actions, taking into account personnel, materiel, and other constraints. This class has great commercial potential, which has been recognized. Examples involve airline scheduling of flights, personnel, and gates; manufacturing job-shop scheduling; and manufacturing process planning.

Configuration of Manufactured Objects from Subassemblies

Configuration, whereby a solution to a problem is synthesized from a given set of elements related by a set of constraints, is historically one of the most important of expert system applications. Configuration applications were pioneered by computer companies as a means of facilitating the manufacture of semi-custom minicomputers

(McDermott 1981). The technique has found its way into use in many different industries, for example, modular home building, manufacturing, and other problems involving complex engineering design and manufacturing.

Financial Decision Making

The financial services industry has been a vigorous user of expert system techniques. Advisory programs have been created to assist bankers in determining whether to make loans to businesses and individuals. Insurance companies have used expert systems to assess the risk presented by the customer and to determine a price for the insurance. A typical application in the financial markets is in foreign exchange trading.

Knowledge Publishing

This is a relatively new, but also potentially explosive area. The primary function of the expert system is to deliver knowledge that is relevant to the user's problem, in the context of the user's problem. The two most widely distributed expert systems in the world are in this category. The first is an advisor which counsels a user on appropriate grammatical usage in a text. The second is a tax advisor that accompanies a tax preparation program and advises the user on tax strategy, tactics, and individual tax policy.

Process Monitoring and Control

Systems falling in this class analyze real-time data from physical devices with the goal of noticing anomalies, predicting trends, and controlling for both optimality and failure correction. Examples of real-time systems that actively monitor processes can be found in the steel making and oil refining industries.

Design and Manufacturing

These systems assist in the design of physical devices and processes, ranging from high-level conceptual design of abstract entities all the way to factory floor configuration of manufacturing processes.

BENEFITS TO END USERS

Primarily, the benefits of ESs to end users include:

- A speed-up of human professional or semi-professional work -- typically by a factor of ten and sometimes by a factor of a hundred or more.
- Within companies, major internal cost savings. For small systems, savings are sometimes in the tens or hundreds of thousands of dollars; but for large systems, often in the tens of millions of dollars and as high as hundreds of millions of dollars. These cost savings are a result of quality improvement, a major motivation for employing expert system technology.
- Improved quality of decision making. In some cases, the quality or correctness of decisions evaluated after the fact show a ten-fold improvement.
- Preservation of scarce expertise. ESs are used to preserve scarce know-how in organizations, to capture the expertise of individuals who are retiring, and to preserve corporate know-how so that it can be widely distributed to other factories, offices or plants of the company.
- Introduction of new products. A good example of a new product is a pathology advisor sold to clinical pathologists in hospitals to assist in the diagnosis of diseased tissue.

ADVANTAGES OF AN EXPERT SYSTEM

Availability: Expert systems ADVANTAGES OF EXPERT SYSTEMS

Availability: Expert systems are available easily due to mass production software.

Cheaper: The cost of providing expertise is not expensive.

Reduced danger: They can be used in any risky environments where humans cannot work with.

Permanence: The knowledge will last long indefinitely.

Multiple expertise: It can be designed to have knowledge of many experts.

Explanation: They are capable of explaining in detail the reasoning that led to a conclusion.

Fast response: They can respond at great speed due to the inherent advantages of computers over humans.

Unemotional and response at all times: Unlike humans, they do not get tense, fatigue or panic and work steadily during emergency situations.

Data Warehouse

The term "Data Warehouse" was first coined by Bill Inmon in 1990. According to Inmon, a data warehouse is a subject-oriented, integrated, time-variant, and non-volatile collection of data. This data helps analysts to take informed decisions in an organization.

An operational database undergoes frequent changes on a daily basis on account of the transactions that take place. Suppose a business executive wants to analyze previous feedback on any data such as a product, a supplier, or any consumer data, then the executive will have no data available to analyze because the previous data has been updated due to transactions.

A data warehouses provides us generalized and consolidated data in multidimensional view. Along with generalized and consolidated view of data, a data warehouses also provides us Online Analytical Processing (OLAP) tools. These tools help us in interactive and effective analysis of data in a multidimensional space. This analysis results in data generalization and data mining.

Data mining functions such as association, clustering, classification, prediction can be integrated with OLAP operations to enhance the interactive mining of knowledge at multiple level of abstraction. That's why data warehouse has now become an important platform for data analysis and online analytical processing.

Understanding a Data Warehouse

- A data warehouse is a database, which is kept separate from the organization's operational database.
- There is no frequent updating done in a data warehouse.
- It possesses consolidated historical data, which helps the organization to analyze its business.
- A data warehouse helps executives to organize, understand, and use their data to take strategic decisions.
- Data warehouse systems help in the integration of diversity of application systems.

- A data warehouse system helps in consolidated historical data analysis.

Why a Data Warehouse is Separated from Operational Databases

A data warehouses is kept separate from operational databases due to the following reasons:

- An operational database is constructed for well-known tasks and workloads such as searching particular records, indexing, etc. In contrast, data warehouse queries are often complex and they present a general form of data.
- Operational databases support concurrent processing of multiple transactions. Concurrency control and recovery mechanisms are required for operational databases to ensure robustness and consistency of the database.

- An operational database query allows to read and modify operations, while an OLAP query needs only read only access of stored data.
- An operational database maintains current data. On the other hand, a data warehouse maintains historical data.

Data Warehouse Features

The key features of a data warehouse are discussed below:

- Subject Oriented - A data warehouse is subject oriented because it provides information around a subject rather than the organization's ongoing operations. These subjects can be product, customers, suppliers, sales, revenue, etc. A data warehouse does not focus on the ongoing operations, rather it focuses on modelling and analysis of data for decision making.
- Integrated – A data warehouse is constructed by integrating data from heterogeneous sources such as relational databases, flat files, etc. This integration enhances the effective analysis of data.
- Time Variant - The data collected in a data warehouse is identified with a particular time period. The data in a data warehouse provides information from the historical point of view.
- Non-volatile - Non-volatile means the previous data is not erased when new data is added to it. A data warehouse is kept separate from the operational database and therefore frequent changes in operational database is not reflected in the data warehouse.

Note: A data warehouse does not require transaction processing, recovery, and concurrency controls, because it is physically stored and separate from the operational database.

Data Warehouse Applications

As discussed before, a data warehouse helps business executives to organize, analyze, and use their data for decision making. A data warehouse serves as a sole part of a plan-execute-assess "closed-loop" feedback system for the enterprise management. Data warehouses are widely used in the following fields:

- Financial services
- Banking services
- Consumer goods
- Retail sectors
- Controlled manufacturing

Types of Data Warehouse

Information processing, analytical processing, and data mining are the three types of data warehouse applications that are discussed below:

- **Information Processing** – A data warehouse allows to process the data stored in it. The data can be processed by means of querying, basic statistical analysis, reporting using crosstabs, tables, charts, or graphs.
- **Analytical Processing** – A data warehouse supports analytical processing of the information stored in it. The data can be analyzed by means of basic OLAP operations, including slice-and-dice, drill down, drill up, and pivoting.
- **Data Mining** - Data mining supports knowledge discovery by finding hidden patterns and associations, constructing analytical models, performing classification and prediction. These mining results can be presented using visualization tools.

Data Warehouse (OLAP)

It involves historical processing of information.
OLAP systems are used by knowledge workers such as executives, managers, and analysts
It is used to analyze the business.
It focuses on Information out.
It is based on Star Schema, Snowflake Schema, and Fact Constellation Schema.
It focuses on Information out.
It contains historical data.
It provides summarized and consolidated data.

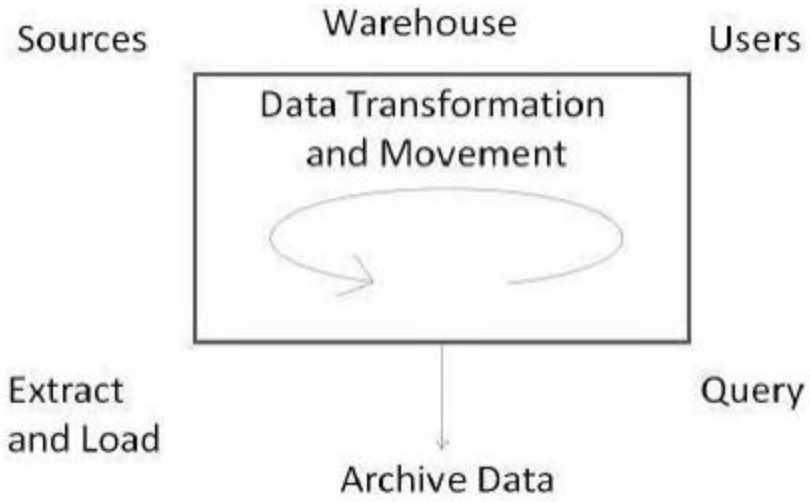
It provides summarized and multidimensional view of data.
The number of users is in hundreds.
The number of records accessed is in millions.
The database size is from 100GB to 100 TB.
These are highly flexible.

Operational Database(OLTP)
It involves day-to-day processing.
. OLTP systems are used by clerks, DBAs, or database professionals.
It is used to run the business.
It focuses on Data in.
It is based on Entity Relationship Model.
It is application oriented.
It contains current data.
It provides primitive and highly detailed data.
It provides detailed and flat relational view of data.
The number of users is in thousands.
The number of records accessed is in tens.
The database size is from 100 MB to 100 GB.
It provides high performance.

What is DataWarehousing?
Data warehousing is the process of constructing and using a data warehouse. A data warehouse is constructed by integrating data from multiple heterogeneous sources that support analytical reporting, structured and/or ad hoc queries, and decision making. Data warehousing involves data cleaning, data integration, and data consolidations.

Using Data Warehouse Information
There are decision support technologies that help utilize the data available in a data warehouse. These technologies help executives to use the warehouse quickly and effectively. They can gather data, analyze it, and take decisions based on the information present in the warehouse. The information gathered in a warehouse can be used in any of the following domains:
- **Tuning Production Strategies -** The product strategies can be well tuned by repositioning the products and managing the product portfolios by comparing the sales quarterly or yearly.
- **Customer Analysis -** Customer analysis is done by analyzing the customer's buying preferences, buying time, budget cycles, etc.
- **Operations Analysis** - Data warehousing also helps in customer relationship management, and making environmental corrections. The information also allows us to analyze business operations.

Process Flow in Data Warehouse
There are four major processes that contribute to a data warehouse:
- Extract and load the data.
- Cleaning and transforming the data.
- Backup and archive the data.
- Managing queries and directing them to the appropriate data sources



Extract and Load Process Data extraction takes data from the source systems. Data load takes the extracted data and loads it into the data warehouse**.**

Note: Before loading the data into the data warehouse, the information extracted from the external sources must be reconstructed.

Controlling the Process Controlling the process involves determining when to start data extraction and the consistency check on data. Controlling process ensures that the tools, the logic modules, and the programs are executed in correct sequence and at correct time.

When to Initiate Extract

Data needs to be in a consistent state when it is extracted, i.e., the data warehouse should represent a single, consistent version of the information to the user.

For example, in a customer profiling data warehouse in telecommunication sector, it is illogical to merge the list of customers at 8 pm on Wednesday from a customer database with the customer subscription events up to 8 pm on Tuesday. This would mean that we are finding the customers for whom there are no associated subscriptions.

Loading the Data After extracting the data, it is loaded into a temporary data store where it is cleaned up and made consistent.

Note: Consistency checks are executed only when all the data sources have been loaded into the temporary data store.

Clean and Transform Process

Once the data is extracted and loaded into the temporary data store, it is time to perform Cleaning and Transforming. Here is the list of steps involved in Cleaning and Transforming:

- Clean and transform the loaded data into a structure
- Partition the data
- Aggregation

Clean and Transform the Loaded Data into a Structure

Cleaning and transforming the loaded data helps speed up the queries. It can be done by making the data consistent:

- within itself.
- with other data within the same data source.
- with the data in other source systems.
- with the existing data present in the warehouse.

Transforming involves converting the source data into a structure. Structuring the data increases the query performance and decreases the operational cost. The data contained in a data warehouse must be transformed to support performance requirements and control the ongoing operational costs.

Partition the Data

It will optimize the hardware performance and simplify the management of data warehouse. Here we partition each fact table into multiple separate partitions.

Aggregation

Aggregation is required to speed up common queries. Aggregation relies on the fact that most common queries will analyze a subset or an aggregation of the detailed data.

Backup and Archive the Data

In order to recover the data in the event of data loss, software failure, or hardware failure, it is necessary to keep regular backups. Archiving involves removing the old data from the system in a format that allows it to be quickly restored whenever required.

For example, in a retail sales analysis data warehouse, it may be required to keep data for 3 years with the latest 6 months data being kept online. In such as scenario, there is often a requirement to be able to do month-on-month comparisons for this year and last year. In this case, we require some data to be restored from the archive.

Query Management Process

This process performs the following functions:

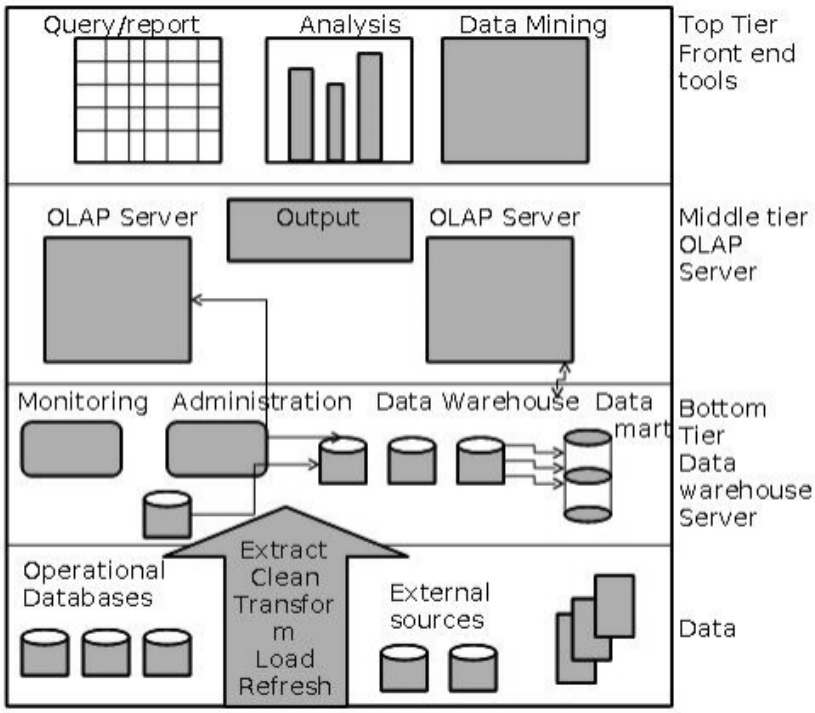
- manages the queries.
- helps speed up the execution time of queries.
- directs the queries to their most effective data sources.
- ensures that all the system sources are used in the most effective way.
- monitors actual query profiles.

The information generated in this process is used by the warehouse management process to determine which aggregations to generate. This process does not generally operate during the regular load of information into data warehouse.

Three-Tier Data Warehouse Architecture

Generally a data warehouses adopts a three-tier architecture. Following are the three tiers of the data warehouse architecture.

- **Bottom Tier** - The bottom tier of the architecture is the data warehouse database server. It is the relational database system. We use the back-end tools and utilities to feed data into the bottom tier. These backend tools and utilities perform the Extract, Clean, Load, and refresh functions.



- **Middle Tier -** In the middle tier, we have the OLAP Server that can be implemented in either of the following ways.

o By Relational OLAP (ROLAP), which is an extended relational database management system. The ROLAP maps the operations on multidimensional data to standard relational operations.
o By Multidimensional OLAP (MOLAP) model, which directly implements the multidimensional data and operations.

- **Top-Tier -** This tier is the front-end client layer. This layer holds the query tools and reporting tools, analysis tools and data mining tools. The following diagram depicts the three-tier architecture of a data warehouse:

Functions of Data Warehouse Tools and Utilities

The following are the functions of data warehouse tools and utilities:

- **Data Extraction** - Involves gathering data from multiple heterogeneous sources.
- **Data Cleaning** - Involves finding and correcting the errors in data.
- **Data Transformation** - Involves converting the data from legacy format to warehouse format.
- **Data Loading -** Involves sorting, summarizing, consolidating, checking integrity, and building indices and partitions.
- **Refreshing** - Involves updating from data sources to warehouse.

Note: Data cleaning and data transformation are important steps in improving the quality of data and data mining results.

Metadata

Metadata is simply defined as data about data. The data that are used to represent other data is known as metadata. For example, the index of a book serves as a metadata for the contents in the book. In other words, we can say that metadata is the summarized data that leads us to the detailed data.

In terms of data warehouse, we can define metadata as following:

- Metadata is a roadmap to data warehouse.
- Metadata in data warehouse defines the warehouse objects.
- Metadata acts as a directory. This directory helps the decision support system to locate the contents of a data warehouse.

Metadata Repository

Metadata repository is an integral part of a data warehouse system. It contains the following metadata:

- **Business metadata -** It contains the data ownership information, business definition, and changing policies.
- **Operational metadata** - It includes currency of data and data lineage. Currency of data refers to the data being active, archived, or purged. Lineage of data means history of data migrated and transformation applied on it.
- **Data for mapping from operational environment to data warehouse** - It metadata includes source databases and their contents, data extraction, data partition, cleaning, transformation rules, data refresh and purging rules.
- **The algorithms for summarization** - It includes dimension algorithms, data on granularity, aggregation, summarizing, etc.

Data Cube

A data cube helps us represent data in multiple dimensions. It is defined by dimensions and facts. The dimensions are the entities with respect to which an enterprise preserves the records.

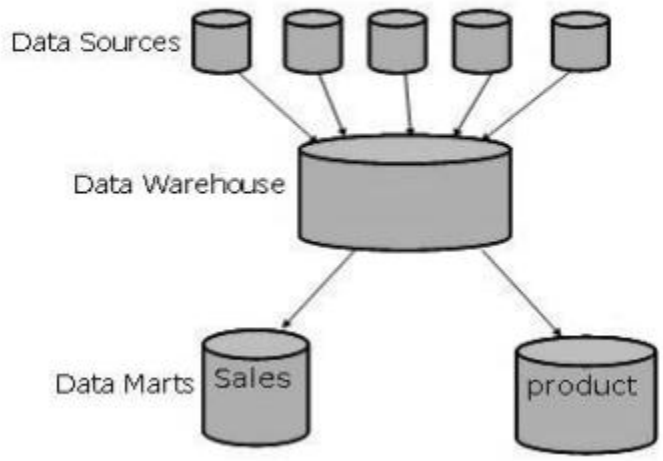
Data Mart

Data marts contain a subset of organization-wide data that is valuable to specific groups of people in an organization. In other words, a data mart contains only those data that is specific to a particular group. For example, the marketing data mart may contain only data related to items, customers, and sales. Data marts are confined to subjects.

Points to Remember About Data Marts

- Windows-based or Unix/Linux-based servers are used to implement data marts. They are implemented on low-cost servers.
- The implementation cycle of a data mart is measured in short periods of time, i.e., in weeks rather than months or years.
- The life cycle of data marts may be complex in the long run, if their planning and design are not organization-wide.
- Data marts are small in size.
- Data marts are customized by department.
- The source of a data mart is departmentally structured data warehouse.
- Data marts are flexible.

The following figure shows a graphical representation of data marts.



Online Analytical Processing Server (OLAP)

Online Analytical Processing Server (OLAP) is based on the multidimensional data model. It allows managers and analysts to get an insight of the information through fast, consistent, and interactive access to information. This chapter covers the types of OLAP, operations on OLAP, difference between OLAP, and statistical databases and OLTP.

Types of OLAP Servers

We have four types of OLAP servers:

- Relational OLAP (ROLAP)
- Multidimensional OLAP (MOLAP)
- Hybrid OLAP (HOLAP)
- Specialized SQL Servers

Relational OLAP

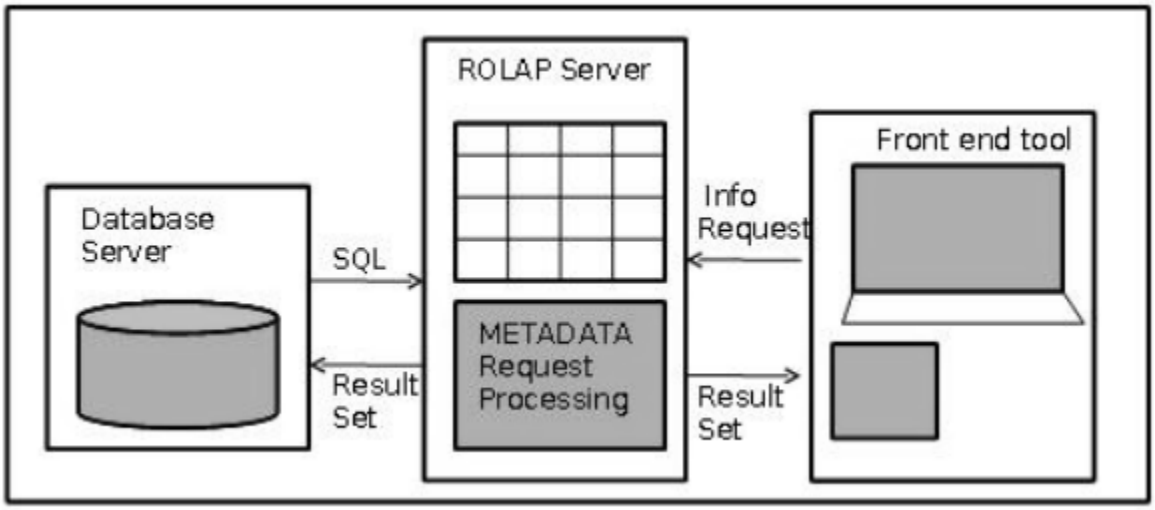
ROLAP servers are placed between relational back-end server and client frontend tools. To store and manage warehouse data, ROLAP uses relational or extended-relational DBMS.

ROLAP includes the following:

- Implementation of aggregation navigation logic.
- Optimization for each DBMS back-end.
- Additional tools and services.

Points to Remember
- ROLAP servers are highly scalable.
- ROLAP tools analyze large volumes of data across multiple dimensions.
- ROLAP tools store and analyze highly volatile and changeable data.

Relational OLAP Architecture
ROLAP includes the following components:
- Database server
- ROLAP server
- Front-end tool



Advantages
- ROLAP servers can be easily used with existing RDBMS.
- Data can be stored efficiently, since no zero facts can be stored.
- ROLAP tools do not use pre-calculated data cubes.
- DSS server of micro-strategy adopts the ROLAP approach.

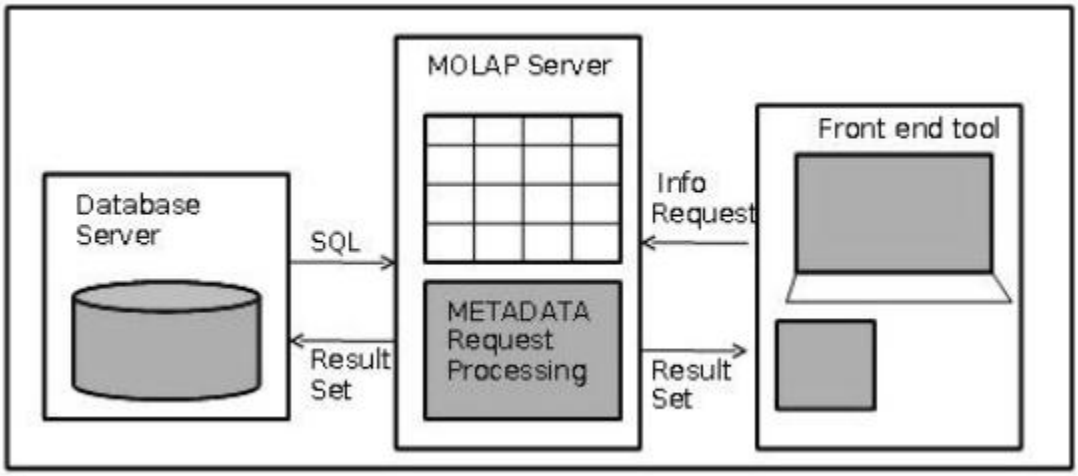
Disadvantages
- Poor query performance.
- Some limitations of scalability depending on the technology architecture that is utilized.

Multidimensional OLAP
MOLAP uses array-based multidimensional storage engines for multidimensional views of data. With multidimensional data stores, the storage utilization may be low if the dataset is sparse. Therefore, many MOLAP servers use two levels of data storage representation to handle dense and sparse datasets.

Points to Remember
- MOLAP tools process information with consistent response time regardless of level of summarizing or calculations selected.
- MOLAP tools need to avoid many of the complexities of creating a relational database to store data for analysis.
- MOLAP tools need fastest possible performance.
- MOLAP server adopts two level of storage representation to handle dense and sparse datasets.
- Denser sub-cubes are identified and stored as array structure.
- Sparse sub-cubes employ compression technology

MOLAP Architecture
MOLAP includes the following components:

- Database server
- MOLAP server
- Front-end tool



Advantages

- MOLAP allows fastest indexing to the pre-computed summarized data.
- Helps the users connected to a network who need to analyze larger, less defined data.
- Easier to use, therefore MOLAP is suitable for inexperienced users.

Disadvantages

- MOLAP are not capable of containing detailed data.
- The storage utilization may be low if the data set is sparse.

MOLAP vs ROLAP

MOLAP	ROLAP
Information retrieval is fast.	Information retrieval is comparatively slow.
Uses sparse array to store datasets.	Uses relational table.
MOLAP is best suited for inexperienced users, since it is very easy to use.	ROLAP is best suited for experienced users.
Maintains a separate database for data cubes.	It may not require space other than available in the data warehouse.
DBMS facility is weak.	DBMS facility is strong.

Hybrid OLAP Hybrid

OLAP is a combination of both ROLAP and MOLAP. It offers higher scalability of ROLAP and faster computation of MOLAP. HOLAP servers allow to store large data volumes of detailed information. The aggregations are stored separately in MOLAP store.

Specialized SQL Servers Specialized SQL servers provide advanced query language and query processing support for SQL queries over star and snowflake schemas in a read-only environment.

OLAP Operations

Since OLAP servers are based on multidimensional view of data, we will discuss OLAP operations in multidimensional data.

Here is the list of OLAP operations:

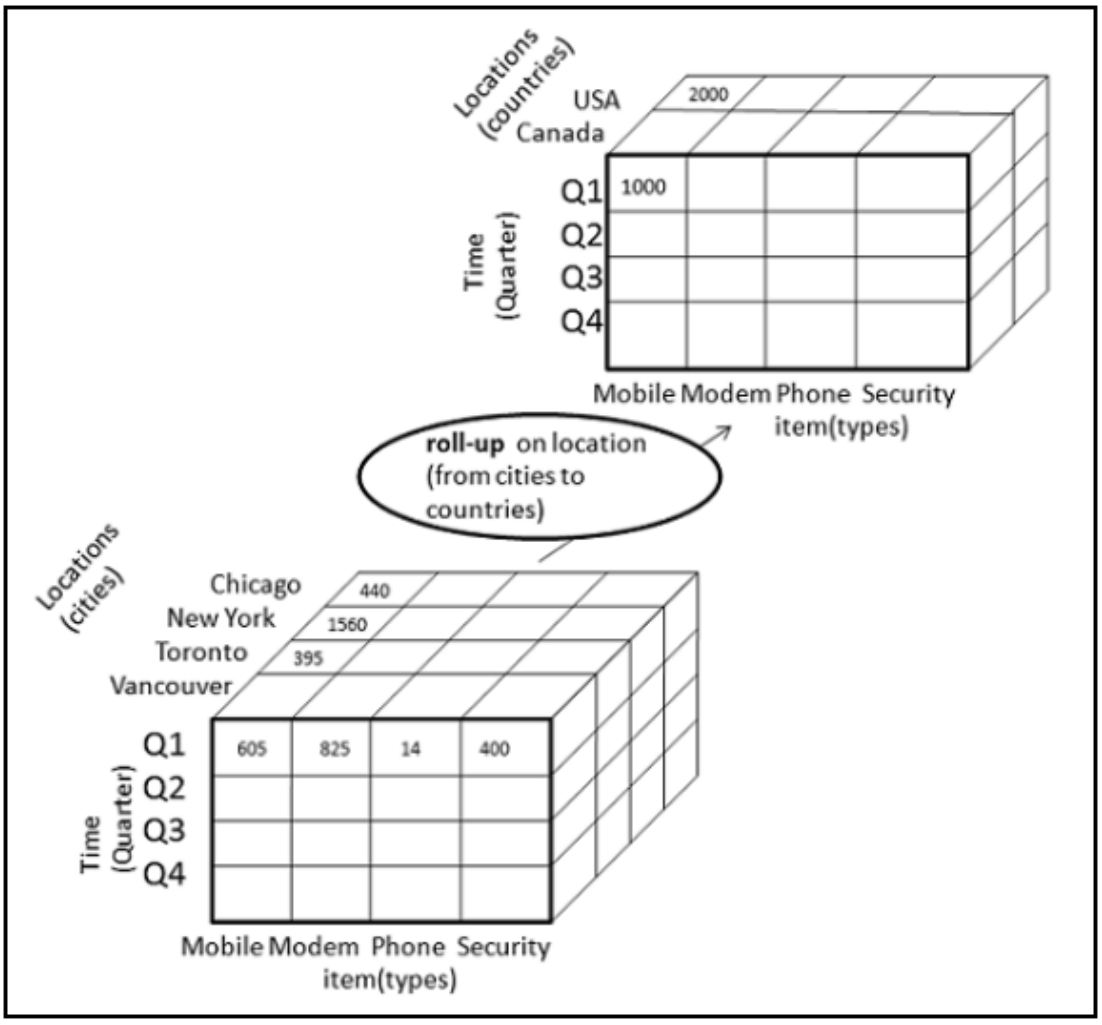
- Roll-up
- Drill-down
- Slice and dice
- Pivot (rotate)

Roll-up

Roll-up performs aggregation on a data cube in any of the following ways:

- By climbing up a concept hierarchy for a dimension
- By dimension reduction

The following diagram illustrates how roll-up works.



- Roll-up is performed by climbing up a concept hierarchy for the dimension location.
- Initially the concept hierarchy was "street < city < province < country".

- On rolling up, the data is aggregated by ascending the location hierarchy from the level of city to the level of country.
- The data is grouped into cities rather than countries.
- When roll-up is performed, one or more dimensions from the data cube are removed.

Drill-down

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

- By stepping down a concept hierarchy for a dimension
- By introducing a new dimension
- Drill-down is performed by stepping down a concept hierarchy for the dimension time.
- Initially the concept hierarchy was "day < month < quarter < year."
- On drilling down, the time dimension is descended from the level of quarter to the level of month.
- When drill-down is performed, one or more dimensions from the data cube are added.
- It navigates the data from less detailed data to highly detailed data.

Slice

The slice operation selects one particular dimension from a given cube and provides a new sub-cube. Consider the following diagram that shows how slice works.

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

Dice

Dice selects two or more dimensions from a given cube and provides a new subcube. Consider the following diagram that shows the dice operation.

The dice operation on the cube based on the following selection criteria involves three dimensions.

- (location = "Toronto" or "Vancouver")
- (time = "Q1" or "Q2")
- (item =" Mobile" or "Modem")

Pivot

The pivot operation is also known as rotation. It rotates the data axes in view in order to provide an alternative presentation of data. Consider the following diagram that shows the pivot operation.

Functions of Data Warehouse Tools and Utilities

The following are the functions of data warehouse tools and utilities:

- **Data Extraction** - Involves gathering data from multiple heterogeneous sources.
- **Data Cleaning** - Involves finding and correcting the errors in data.
- **Data Transformation** - Involves converting the data from legacy format to warehouse format.
- **Data Loading** - Involves sorting, summarizing, consolidating, checking integrity, and building indices and partitions.
- **Refreshing** - Involves updating from data sources to warehouse.

Note: Data cleaning and data transformation are important steps in improving the quality of data and data mining results.

1. What do you understand by group decision support system? What are the components of GDSS?
2. What is executive information system? How does decision support system help in business? Explain.
3. Explain the term groupware technologies in detail.
4. Define business expert system in detail along with its benefits.
5. Explain in detail data warehousing tools and utilities functions.

6. What is data warehouse? What are the goals of a data warehouse?
7. How are data marts different from data warehouse?
8. Differentiate Datamart and Warehouse.
9. Define Metadata.
10. Differentiate OLAP and OLTP.
11. Explain the three-tier architecture of a data warehouse.
12. Differentiate MOLAP and ROLTP.
13. Write all the operations of OLAP
14. Define the architecture of ROLAP and MOLAP
15. Write all the Process Flow in Data Warehousing.

UNIT-III

Multi- Dimensional analysis:

Data mining and knowledge discovery,
Data mining and Techniques,
Data mining of Advance Databases.
What is knowledge discovery?
What is datamining technologies?
What are datamining functionalities?
What do you understand by frequent pattern mining?
Classification Vs Clustering.

Section B or C

What are various steps of knowledge discovery? Discuss the role of datamining in knowledge discovery.
Explain the diagrammatic illustration for step by steps involved in the process of knowledge discovery from data base.
What is need of datamining? Explain different types of data mining techniques.
What is need of datamining? Discuss the evaluation of database system technologies.
Explain various methods for evaluating the accuracy of classification or prediction.
Describe classification and prediction. Discuss methods regarding classification.
Describe Apriori Algorithm for frequent pattern mining.

1. **Datamining terminology**

Data Mining

Data mining is defined as extracting the information from a huge set of data. In other words we can say that data mining is mining the knowledge from data. This information can be used for any of the following applications −

- Market Analysis
- Fraud Detection
- Customer Retention
- Production Control
- Science Exploration

Data Mining Engine

Data mining engine is very essential to the data mining system. It consists of a set of functional modules that perform the following functions −

- Characterization
- Association and Correlation Analysis
- Classification
- Prediction
- Cluster analysis
- Outlier analysis
- Evolution analysis

Knowledge Base

This is the domain knowledge. This knowledge is used to guide the search or evaluate the interestingness of the resulting patterns.

Knowledge Discovery

Some people treat data mining same as knowledge discovery, while others view data mining as an essential step in the process of knowledge discovery. Here is the list of steps involved in the knowledge discovery process −

- Data Cleaning
- Data Integration
- Data Selection
- Data Transformation
- Data Mining
- Pattern Evaluation
- Knowledge Presentation

User interface

User interface is the module of data mining system that helps the communication between users and the data mining system. User Interface allows the following functionalities −

- Interact with the system by specifying a data mining query task.
- Providing information to help focus the search.
- Mining based on the intermediate data mining results.
- Browse database and data warehouse schemas or data structures.
- Evaluate mined patterns.
- Visualize the patterns in different forms.

Data Integration

Data Integration is a data preprocessing technique that merges the data from multiple heterogeneous data sources into a coherent data store. Data integration may involve inconsistent data and therefore needs data cleaning.

Data Cleaning

Data cleaning is a technique that is applied to remove the noisy data and correct the inconsistencies in data. Data cleaning involves transformations to correct the wrong data. Data cleaning is performed as a data preprocessing step while preparing the data for a data warehouse.

Data Selection

Data Selection is the process where data relevant to the analysis task are retrieved from the database. Sometimes data transformation and consolidation are performed before the data selection process.

Clusters

Cluster refers to a group of similar kind of objects. Cluster analysis refers to forming group of objects that are very similar to each other but are highly different from the objects in other clusters.

Data Transformation

In this step, data is transformed or consolidated into forms appropriate for mining, by performing summary or aggregation operations.

2. **Data Mining Applications**

Here is the list of areas where data mining is widely used −

- Financial Data Analysis
- Retail Industry
- Telecommunication Industry
- Biological Data Analysis
- Other Scientific Applications
- Intrusion Detection

Financial Data Analysis

The financial data in banking and financial industry is generally reliable and of high quality which facilitates systematic data analysis and data mining. Some of the typical cases are as follows −

- Design and construction of data warehouses for multidimensional data analysis and data mining.
- Loan payment prediction and customer credit policy analysis.
- Classification and clustering of customers for targeted marketing.

- Detection of money laundering and other financial crimes.

Retail Industry

Data Mining has its great application in Retail Industry because it collects large amount of data from on sales, customer purchasing history, goods transportation, consumption and services. It is natural that the quantity of data collected will continue to expand rapidly because of the increasing ease, availability and popularity of the web.

Data mining in retail industry helps in identifying customer buying patterns and trends that lead to improved quality of customer service and good customer retention and satisfaction. Here is the list of examples of data mining in the retail industry −

- Design and Construction of data warehouses based on the benefits of data mining.
- Multidimensional analysis of sales, customers, products, time and region.
- Analysis of effectiveness of sales campaigns.
- Customer Retention.
- Product recommendation and cross-referencing of items.

Telecommunication Industry

Today the telecommunication industry is one of the most emerging industries providing various services such as fax, pager, cellular phone, internet messenger, images, e-mail, web data transmission, etc. Due to the development of new computer and communication technologies, the telecommunication industry is rapidly expanding. This is the reason why data mining is become very important to help and understand the business.

Data mining in telecommunication industry helps in identifying the telecommunication patterns, catch fraudulent activities, make better use of resource, and improve quality of service. Here is the list of examples for which data mining improves telecommunication services −

- Multidimensional Analysis of Telecommunication data.
- Fraudulent pattern analysis.
- Identification of unusual patterns.
- Multidimensional association and sequential patterns analysis.
- Mobile Telecommunication services.
- Use of visualization tools in telecommunication data analysis.

Biological Data Analysis

In recent times, we have seen a tremendous growth in the field of biology such as genomics, proteomics, functional Genomics and biomedical research. Biological data mining is a very important part of Bioinformatics. Following are the aspects in which data mining contributes for biological data analysis −

- Semantic integration of heterogeneous, distributed genomic and proteomic databases.
- Alignment, indexing, similarity search and comparative analysis multiple nucleotide sequences.
- Discovery of structural patterns and analysis of genetic networks and protein pathways.
- Association and path analysis.
- Visualization tools in genetic data analysis.

Other Scientific Applications

The applications discussed above tend to handle relatively small and homogeneous data sets for which the statistical techniques are appropriate. Huge amount of data have been collected from scientific domains such as geosciences, astronomy, etc. A large amount of data sets is being generated because of the fast numerical simulations in various fields such as climate and ecosystem modeling, chemical engineering, fluid dynamics, etc. Following are the applications of data mining in the field of Scientific Applications −

- Data Warehouses and data preprocessing.
- Graph-based mining.
- Visualization and domain specific knowledge.

Intrusion Detection

Intrusion refers to any kind of action that threatens integrity, confidentiality, or the availability of network resources. In this world of connectivity, security has become the major issue. With increased usage of internet and availability of the tools and tricks for intruding and attacking network prompted intrusion detection to become a critical component of network administration. Here is the list of areas in which data mining technology may be applied for intrusion detection −

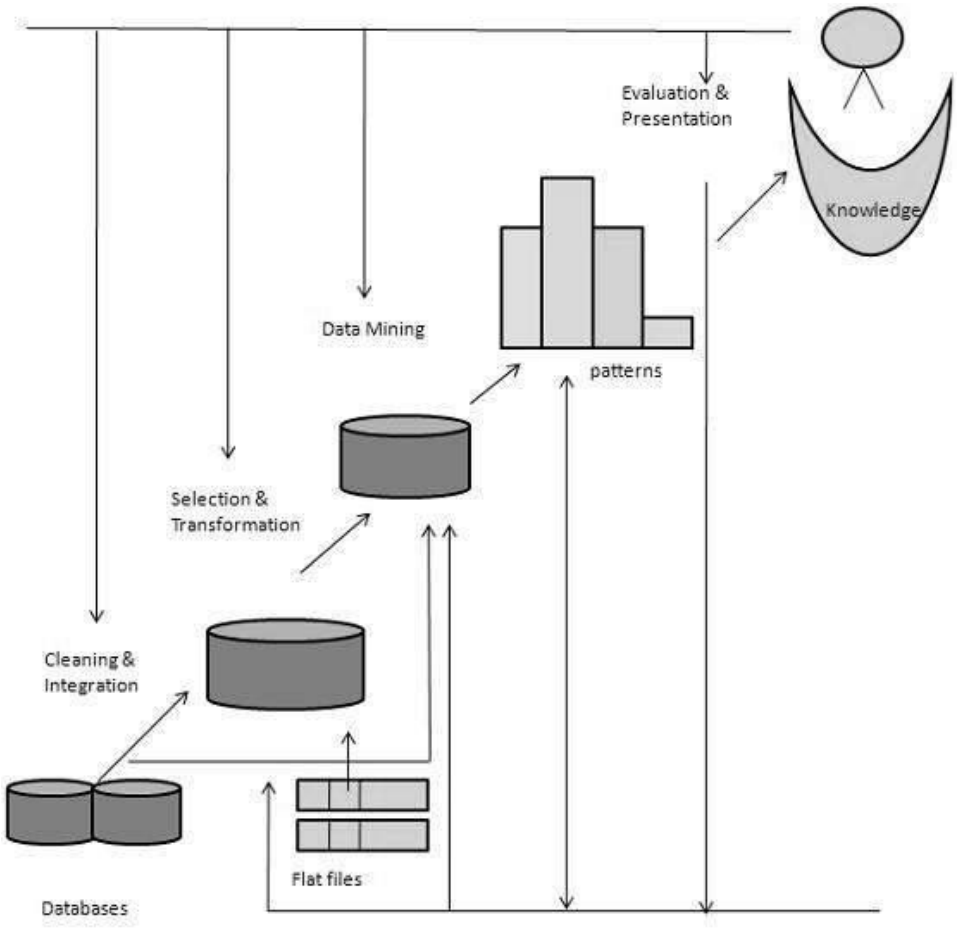
- Development of data mining algorithm for intrusion detection.
- Association and correlation analysis, aggregation to help select and build discriminating attributes.
- Analysis of Stream data.
- Distributed data mining.
- Visualization and query tools.

3. **What is Knowledge Discovery?**

Data mining as an essential step in the process of knowledge discovery. Here is the list of steps involved in the knowledge discovery process −

- **Data Cleaning** − In this step, the noise and inconsistent data is removed.
- **Data Integration** − In this step, multiple data sources are combined.
- **Data Selection** − In this step, data relevant to the analysis task are retrieved from the database.
- **Data Transformation** − In this step, data is transformed or consolidated into forms appropriate for mining by performing summary or aggregation operations.
- **Data Mining** − In this step, intelligent methods are applied in order to extract data patterns.
- **Pattern Evaluation** − In this step, data patterns are evaluated.
- **Knowledge Presentation** − In this step, knowledge is represented.

The following diagram shows the process of knowledge discovery −



4. **Datamining tasks:** Data mining deals with the kind of patterns that can be mined. On the basis of the kind of data to be mined, there are two categories of functions involved in Data Mining −
1. **Descriptive**
2. **Classification & prediction**

Descriptive Function

The descriptive function deals with the general properties of data in the database. Here is the list of descriptive functions −

- Class/Concept Description
- Mining of Frequent Patterns
- Mining of Associations
- Mining of Correlations
- Mining of Clusters

Class/Concept Description

Class/Concept refers to the data to be associated with the classes or concepts. For example, in a company, the classes of items for sales include computer and printers, and concepts of customers include big spenders and budget spenders. Such descriptions of a class or a concept are called class/concept descriptions. These descriptions can be derived by the following two ways −

- **Data Characterization** − This refers to summarizing data of class under study. This class under study is called as Target Class.
- **Data Discrimination** − It refers to the mapping or classification of a class with some predefined group or class.

Mining of Frequent Patterns

Frequent patterns are those patterns that occur frequently in transactional data. Here is the list of kind of frequent patterns −

- **Frequent Item Set** − It refers to a set of items that frequently appear together, for example, milk and bread.
- **Frequent Subsequence** − A sequence of patterns that occur frequently such as purchasing a camera is followed by memory card.
- **Frequent Sub Structure** − Substructure refers to different structural forms, such as graphs, trees, or lattices, which may be combined with item−sets or subsequences.

Mining of Association

Associations are used in retail sales to identify patterns that are frequently purchased together. This process refers to the process of uncovering the relationship among data and determining association rules.

For example, a retailer generates an association rule that shows that 70% of time milk is sold with bread and only 30% of times biscuits are sold with bread.

Mining of Correlations

It is a kind of additional analysis performed to uncover interesting statistical correlations between associated-attribute−value pairs or between two item sets to analyze that if they have positive, negative or no effect on each other.

Mining of Clusters

Cluster refers to a group of similar kind of objects. Cluster analysis refers to forming group of objects that are very similar to each other but are highly different from the objects in other clusters.

Classification & prediction: Classification is the process of finding a model (or function) that describes and distinguishes data classes or concepts, for the purpose of being able to use the model to predict the class of objects whose class label is unknown.
The derived model is based on the analysis of a set of training data (i.e. data object whose class label is known.)
The derived model may be represented in various forms, such as classification(if-then) rule, decision tree, neural network.

There are two forms of data analysis that can be used for extracting models describing important classes or to predict future data trends. These two forms are as follows −

- Classification
- Prediction

Classification models predict categorical class labels; and prediction models predict continuous valued functions. For example, we can build a classification model to categorize bank loan applications as either safe or risky, or a prediction model to predict the expenditures in dollars of potential customers on computer equipment given their income and occupation.

What is classification?

Following are the examples of cases where the data analysis task is Classification −

- A bank loan officer wants to analyze the data in order to know which customer (loan applicant) are risky or which are safe.
- A marketing manager at a company needs to analyze a customer with a given profile, who will buy a new computer.

In both of the above examples, a model or classifier is constructed to predict the categorical labels. These labels are risky or safe for loan application data and yes or no for marketing data.

What is prediction?

Following are the examples of cases where the data analysis task is Prediction −

Suppose the marketing manager needs to predict how much a given customer will spend during a sale at his company. In this example we are bothered to predict a numeric value. Therefore the data analysis task is an example of numeric prediction. In this case, a model or a predictor will be constructed that predicts a continuous-valued-function or ordered value.

Note − Regression analysis is a statistical methodology that is most often used for numeric prediction.

How Does Classification Works?

With the help of the bank loan application that we have discussed above, let us understand the working of classification. The Data Classification process includes two steps −

- Building the Classifier or Model
- Using Classifier for Classification

Building the Classifier or Model

- This step is the learning step or the learning phase.
- In this step the classification algorithms build the classifier.
- The classifier is built from the training set made up of database tuples and their associated class labels.
- Each tuple that constitutes the training set is referred to as a category or class. These tuples can also be referred to as sample, object or data points.

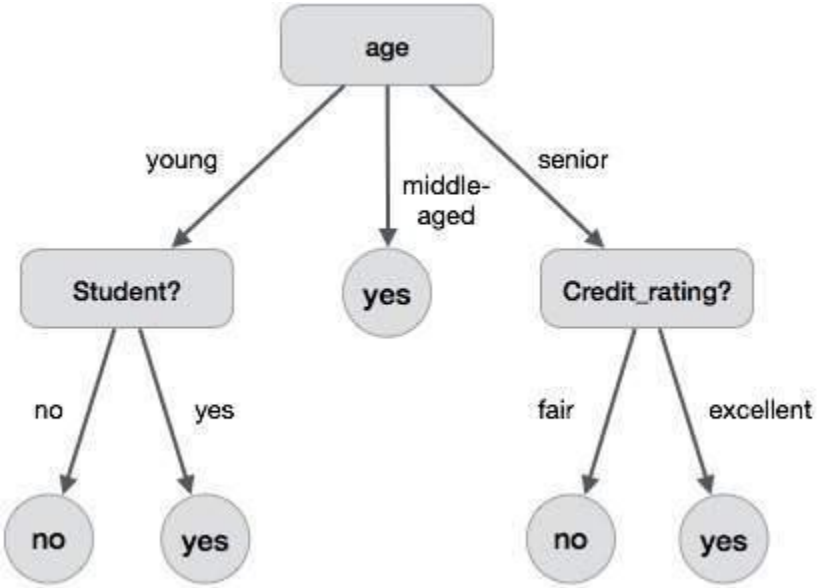
Using Classifier for Classification

In this step, the classifier is used for classification. Here the test data is used to estimate the accuracy of classification rules. The classification rules can be applied to the new data tuples if the accuracy is considered acceptable.

Decision tree:

A decision tree is a structure that includes a root node, branches, and leaf nodes. Each internal node denotes a test on an attribute, each branch denotes the outcome of a test, and each leaf node holds a class label. The topmost node in the tree is the root node.

The following decision tree is for the concept buy_computer that indicates whether a customer at a company is likely to buy a computer or not. Each internal node represents a test on an attribute. Each leaf node represents a class.



The benefits of having a decision tree are as follows −

- It does not require any domain knowledge.
- It is easy to comprehend.
- The learning and classification steps of a decision tree are simple and fast.

5. **Clustering vs Classification**

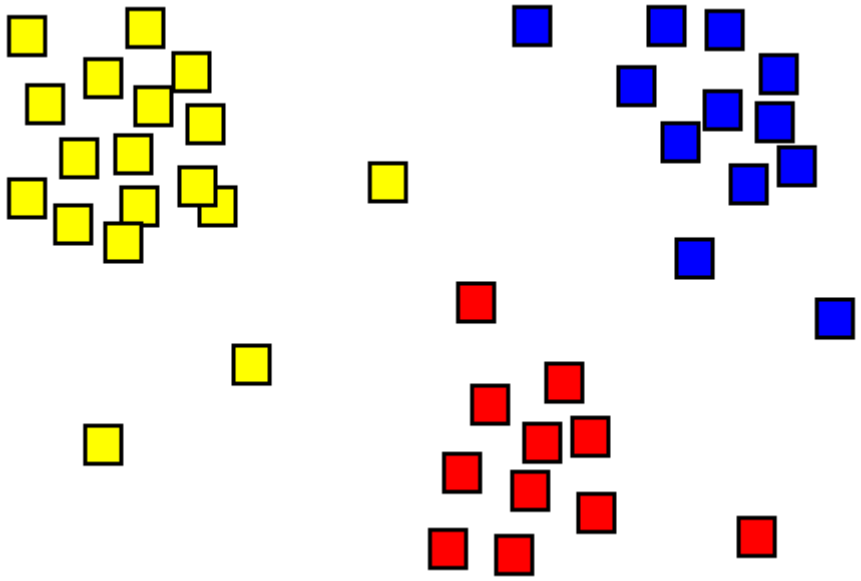
Though clustering and classification appear to be similar processes, there is a difference between them based on their meaning. In the data mining world, clustering and classification are two types of learning methods. Both these methods characterize objects into groups by one or more features. The key difference between clustering and classification is that **clustering is an unsupervised learning technique used to group similar instances on the basis of features** whereas **classification is a supervised learning technique used to assign predefined tags to instances on the basis of features.**

What is Clustering?

Clustering is a method of grouping objects in such a way that objects with similar features come together, and objects with dissimilar features go apart. It is a common technique for statistical data analysis used in machine learning and data mining. Clustering can be used for exploratory data analysis and generalization.

Clustering belongs to unsupervised data mining, and clustering is not a single specific algorithm, but a general method to solve the task. Clustering can be achieved by various algorithms. The appropriate cluster algorithm and parameter settings depend on the individual data sets. It is not an automatic task, but it is an iterative process of discovery. Therefore, it is necessary to modify data processing and parameter modeling until the result achieves

the desired properties. K-means clustering and Hierarchical clustering are two common clustering algorithms used

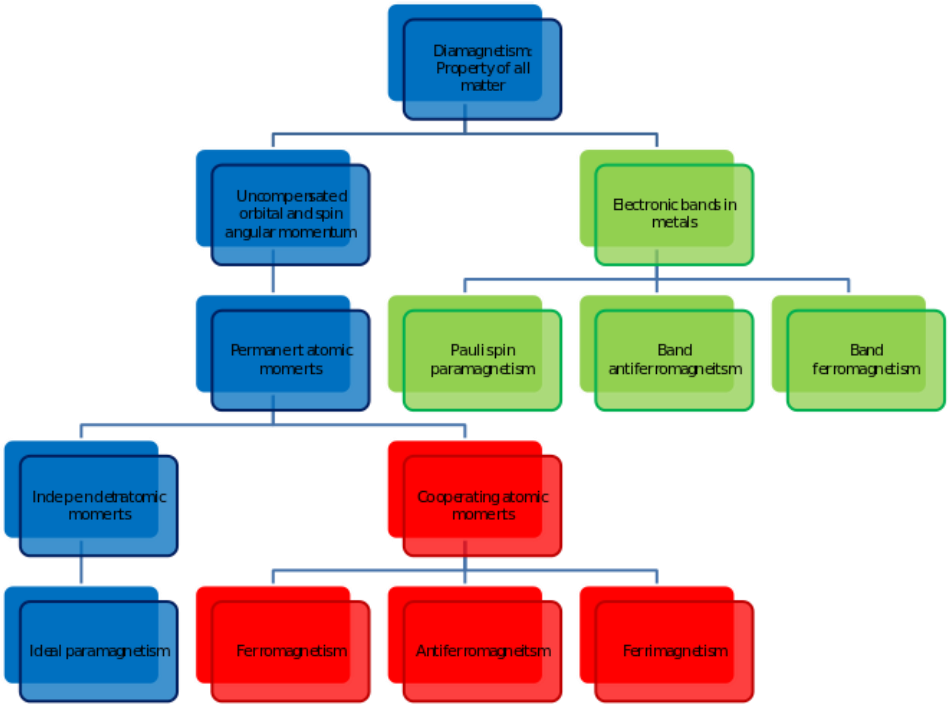


in data mining.

What is Classification?

Classification is a process of categorization where objects are recognized, differentiated and understood on the basis of the training set of data. Classification is a supervised learning technique where a training set and correctly defined observations are available.

The algorithm which implements classification is often known as the classifier, and the observations are often known as the instances. K-Nearest Neighbor algorithm and decision tree algorithms are the most famous classification algorithms used in data mining.



What is the difference between Clustering and Classification?

Clustering: Clustering is an unsupervised learning technique used to group similar instances on the basis of features.

Classification: Classification is a supervised learning technique used to assign predefined tags to instances on the basis of features.

Whether you chose supervised or unsupervised should be based on whether or not you know what the 'categories' of your data are. If you know, use supervised learning. If you do not know, then use unsupervised.

6. **Apriori algorithm** Apriori algorithm is an association rule mining algorithm used in data mining. It is used to find the frequent item set among the given number of transactions. It is a classic algorithm used in data mining for learning association rules. It is nowhere as complex as it sounds, on the contrary it is very simple; let me give you an example to explain it. Suppose you have records of large number of transactions at a shopping center as follows:

Transactions	Items bought
T1	Item1, item2, item3
T2	Item1, item2
T3	Item2, item5
T4	Item1, item2, item5

Learning association rules basically means finding the items that are purchased together more frequently than others.
For example in the above table you can see Item1 and item2 are bought together frequently.

What is the use of learning association rules?

- Shopping centers use association rules to place the items next to each other so that users buy more items. If you are familiar with data mining you would know about the famous beer-diapers-Wal-Mart story. Basically Wal-Mart studied their data and found that on Friday afternoon young American males who buy diapers also tend to buy beer. So Wal-Mart placed beer next to diapers and the beer-sales went up. This is famous because no one would have predicted such a result and that's the power of data mining.
- Also if you are familiar with Amazon, they use association mining to recommend you the items based on the current item you are browsing/buying.
- Another application is the Google auto-complete, where after you type in a word it searches frequently associated words that user type after that particular word.

So as I said Apriori is the classic and probably the most basic algorithm to do it.
Let's start with a non-simple example,

Transaction ID	Items Bought
T1	{Mango, Onion, Nintendo, Key-chain, Eggs, Yo-yo}
T2	{Doll, Onion, Nintendo, Key-chain, Eggs, Yo-yo}
T3	{Mango, Apple, Key-chain, Eggs}
T4	{Mango, Umbrella, Corn, Key-chain, Yo-yo}
T5	{Corn, Onion, Onion, Key-chain, Ice-cream, Eggs}

Now, we follow a simple golden rule: we say an item/itemset is frequently bought if it is bought at least 60% of times. So for here it should be bought at least 3 times.

For simplicity
M = Mango
O = Onion
And so on……

So the table becomes

Original table:

Transaction ID	Items Bought
T1	{M, O, N, K, E, Y }
T2	{D, O, N, K, E, Y }
T3	{M, A, K, E}
T4	{M, U, C, K, Y }
T5	{C, O, O, K, I, E}

Step 1: Count the number of transactions in which each item occurs, Note 'O=Onion' is bought 4 times in total, but, it occurs in just 3 transactions.

Item	No of transactions
M	3
O	3
N	2
K	5
E	4
Y	3
D	1
A	1
U	1
C	2
I	1

Step 2: Now remember we said the item is said frequently bought if it is bought at least 3 times. So in this step we remove all the items that are bought less than 3 times from the above table and we are left with

Item	Number of transactions
M	3
O	3
K	5
E	4
Y	3

This is the single items that are bought frequently. Now let's say we want to find a pair of items that are bought frequently. We continue from the above table (Table in step 2)

Step 3: We start making pairs from the first item, like MO,MK,ME,MY and then we start with the second item like OK,OE,OY. We did not do OM because we already did MO when we were making pairs with M and buying a Mango and Onion together is same as buying Onion and Mango together. After making all the pairs we get,

Item pairs
MO
MK
ME
MY
OK
OE
OY
KE
KY
EY

Step 4: Now we count how many times each pair is bought together. For example M and O is just bought together in {M,O,N,K,E,Y}
While M and K is bought together 3 times in {M,O,N,K,E,Y}, {M,A,K,E} AND {M,U,C, K, Y}
After doing that for all the pairs we get

Item Pairs	Number of transactions
MO	1
MK	3
ME	2
MY	2
OK	3
OE	3
OY	2
KE	4
KY	3
EY	2

Step 5: Golden rule to the rescue. Remove all the item pairs with number of transactions less than three and we are left with

Item Pairs	Number of transactions
MK	3
OK	3
OE	3
KE	4
KY	3

These are the pairs of items frequently bought together.
Now let's say we want to find a set of three items that are brought together.
We use the above table (table in step 5) and make a set of 3 items.

Step 6: To make the set of three items we need one more rule (it's termed as self-join),
It simply means, from the Item pairs in the above table, we find two pairs with the same first Alphabet, so we get

- OK and OE, this gives OKE
- KE and KY, this gives KEY

Then we find how many times O,K,E are bought together in the original table and same for K,E,Y and we get the following table

Item Set	Number of transactions
OKE	3
KEY	2

While we are on this, suppose you have sets of 3 items say ABC, ABD, ACD, ACE, BCD and you want to generate item sets of 4 items you look for two sets having the same first two alphabets.

- ABC and ABD -> ABCD
- ACD and ACE -> ACDE

And so on … In general you have to look for sets having just the last alphabet/item different.

Step 7: So we again apply the golden rule, that is, the item set must be bought together at least 3 times which leaves us with just OKE, Since KEY are bought together just two times.

Thus the set of three items that are bought together most frequently are O,K,E.

7. Data mining techniques

There are several major ***data mining techniques*** have been developing and using in data mining projects recently including *association, classification, clustering, prediction, sequential patterns* and *decision tree*. We will briefly examine those data mining techniques in the following sections.

Association

Association is one of the best-known data mining technique. In association, a pattern is discovered based on a relationship between items in the same transaction. That's is the reason why association technique is also known as *relation technique*. The association technique is used in *market basket analysis* to identify a set of products that customers frequently purchase together.

Retailers are using association technique to research customer's buying habits. Based on historical sale data, retailers might find out that customers always buy crisps when they buy beers, and, therefore, they can put beers and crisps next to each other to save time for customer and increase sales.

Classification

Classification is a classic data mining technique based on machine learning. Basically, classification is used to classify each item in a set of data into one of a predefined set of classes or groups. Classification method makes use of mathematical techniques such as decision trees, linear programming, neural network and statistics. In classification, we develop the software that can learn how to classify the data items into groups. For example, we can apply classification in the application that "given all records of employees who left the company, predict who will probably leave the company in a future period." In this case, we divide the records of employees into two groups that named "leave" and "stay". And then we can ask our data mining software to classify the employees into separate groups.

Clustering

Clustering is a data mining technique that makes a meaningful or useful cluster of objects which have similar characteristics using the automatic technique. The clustering technique defines the classes and puts objects in each class, while in the classification techniques, objects are assigned into predefined classes. To make the concept clearer, we can take book management in the library as an example. In a library, there is a wide range of books on various topics available. The challenge is how to keep those books in a way that readers can take several books on a particular topic without hassle. By using the clustering technique, we can keep books that have some kinds of similarities in one cluster or one shelf and label it with a meaningful name. If readers want to grab books in that topic, they would only have to go to that shelf instead of looking for the entire library.

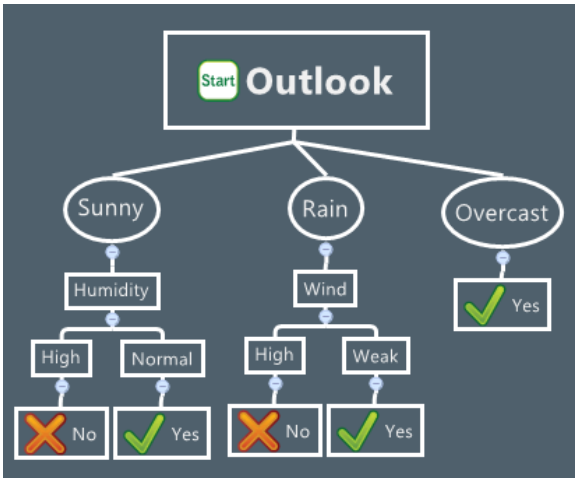
Prediction The prediction, as its name implied, is one of a data mining techniques that discovers the relationship between independent variables and relationship between dependent and independent variables. For instance, the prediction analysis technique can be used in the sale to predict profit for the future if we consider the sale is an independent variable, profit could be a dependent variable. Then based on the historical sale and profit data, we can draw a fitted regression curve that is used for profit prediction.

Sequential Patterns Sequential patterns analysis is one of data mining technique that seeks to discover or identify similar patterns, regular events or trends in transaction data over a business period.

In sales, with historical transaction data, businesses can identify a set of items that customers buy together different times in a year. Then businesses can use this information to recommend customers buy it with better deals based on their purchasing frequency in the past.

Decision trees

The A decision tree is one of the most common used data mining techniques because its model is easy to understand for users. In decision tree technique, the root of the decision tree is a simple question or condition that has multiple answers. Each answer then leads to a set of questions or conditions that help us determine the data so that we can make the final decision based on it. For example, We use the following decision tree to determine whether or not to play tennis:



Starting at the root node, if the outlook is overcast then we should definitely play tennis. If it is rainy, we should only play tennis if the wind is the week. And if it is sunny then we should play tennis in case the humidity is normal.

We often combine two or more of those data mining techniques together to form an appropriate process that meets the business needs.

8. **Data mining tools**

Data mining is not all about the tools or database software that you are using. You can perform data mining with comparatively modest database systems and simple tools, including creating and writing your own, or using off the shelf software packages. Complex data mining benefits from the past experience and algorithms defined with existing software and packages, with certain tools gaining a greater affinity or reputation with different techniques. Ex oracle data miner, data to knowledge, sas, clementine, intelligent miner etc.

9. Bayesian and Neural networks

Data mining is the process of extracting nontrivial and potentially useful information, or knowlege, from the enormous data sets available in experimental sciences (historical records, reanalysis, GCM simulations, etc.), providing explicit information that has a readable form and can be used to solve diagnosis, classification or forecasting problems. Traditionally, these problems were solved by direct hands-on data analysis using standard statistical methods, but the increasing volume of data has motivated the study of automatic data analysis using more complex and sophisticated tools which can operate directly from data. Thus, data mining identifies trends within data that go beyond simple analysis. Modern data mining techniques (association rules, decision trees, Gaussian mixture models, regression algorithms, neural networks, support vector machines, Bayesian networks, etc.) are used in many domains to solve association, classification, segmentation, diagnosis and prediction problems.
Among the different data mining algorithms, probabilistic graphical models (in particular Bayesian networks) is a sound and powerful methodology grounded on probability and statistics, which allows building tractable joint probabilistic models that represents the relevant dependencies among a set of variables (hundreds of variables in real-life applications).

Formally, Bayesian networks are directed acyclic graphs whose nodes represent variables, and whose arcs encode conditional independencies between the variables. The graph provides an intuitive description of the dependency model and defines a simple factorization of the joint probability distribution leading to a tractable model which is compatible with the encoded dependencies. Efficient algorithms exist to learn both the graphical and the probabilistic models from data, thus allowing for the automatic application of this methodogy in complex problems. Bayesian networks that model sequences of variables (such as, for example, time series of historical records) are called dynamic Bayesian networks. Generalizations of

Bayesian networks that can represent and solve decision problems under uncertainty are called influence diagrams.
On the other hand, neural networks are nonlinear models inspired in the functioning of the brain which have been designed to solve different problems. Thus, multi-layer perceptrons are regression-like algorithms to build a deterministic model y=f(x), relating a set of predictors, x, and predictands, y (figure below, left). Self-Organizing Maps (SOM) are competitive networks designed for clustering and visualization purposes (right).

UNIT-IV
KNOWLEDGE MANAGEMENT

Define knowledge management and knowledge management process.
Explain the types of knowledge
Write brief about failure of knowledge management
Characteristics of knowledge
What are different issues of challenges for knowledge management?
Difference between tacit and explicit knowledge
Define expert knowledge
Advantages and disadvantages of knowledge management
Limitations of knowledge management
Define knowledge architecture in detail
What are different benefits of knowledge management.
WRITE DOWN THE PHASES OF KNOWLEDGE MANAGEMENT

Define

1. RER
2. CASE STUDY
3. KNOWLEDGE BANK
4. KNOWLEDGE CAFÉ
5. KNOWLEDGE MARKETPLACE
6. COTS
7. BRAINSTORMING
8. ROI

"Knowledge management is really about recognizing that regardless of what business you are in, you are competing based on the knowledge of your employees"

Introduction of Knowledge Management

Knowledge management is essentially about getting the right knowledge to the right person at the right time. This in itself may not seem so complex, but it implies a strong tie to corporate strategy, understanding of where and in what forms knowledge exists, creating processes that span organizational functions, and ensuring that initiatives are accepted and supported by organizational members. Knowledge management may also include new knowledge creation, or it may solely focus on knowledge sharing, storage, and refinement. For a more comprehensive discussion and definition, see my knowledge management definition.
It is important to remember that knowledge management is not about managing knowledge for knowledge's sake. The overall objective is to create value and leverage and refine the firm's knowledge assets to meet organizational goals.

Implementing knowledge management thus has several dimensions including:

- **Strategy:** Knowledge management strategy must be dependent on corporate strategy. The objective is to manage, share, and create *relevant* knowledge assets that will help meet tactical and strategic requirements.

- **Organizational Culture:** The organizational culture influences the way people interact, the context within which knowledge is created, the resistance they will have towards certain changes, and ultimately the way they share (or the way they do not share) knowledge.
- **Organizational Processes:** The right processes, environments, and systems that enable KM to be implemented in the organization.
- **Management & Leadership:** KM requires competent and experienced leadership at all levels. There are a wide variety of KM-related roles that an organization may or may not need to implement, including a CKO, knowledge managers, knowledge brokers and so on. More on this in the section on KM positions and roles.
- **Technology:** The systems, tools, and technologies that fit the organization's requirements - properly designed and implemented.
- **Politics:** The long-term support to implement and sustain initiatives that involve virtually all organizational functions, which may be costly to implement (both from the perspective of time and money), and which often do not have a directly visible return on investment.

why is knowledge management useful? It is useful because it places a focus on knowledge as an actual asset, rather than as something intangible. In so doing, it enables the firm to better protect and exploit what it knows, and to improve and focus its knowledge development efforts to match its needs.

In other words:

- It helps firms learn from past mistakes and successes.
- It better exploits existing knowledge assets by re-deploying them in areas where the firm stands to gain something, e.g. using knowledge from one department to improve or create a product in another department, modifying knowledge from a past process to create a new solution, etc.
- It promotes a long term focus on developing the right competencies and skills and removing obsolete knowledge.
- It enhances the firm's ability to innovate.
- It enhances the firm's ability to protect its key knowledge and competencies from being lost or copied.

Unfortunately, KM is an area in which companies are often reluctant to invest because it can be expensive to implement *properly*, and it is extremely difficult to determine a specific ROI. Moreover KM is a concept the definition of which is not universally accepted, and for example within IT one often sees a much shallower, information-oriented approach. Particularly in the early days, this has led to many "KM" failures and these have tarnished the reputation of the subject as a whole. Sadly, even today, probably about one in three blogs that I read

on this subject have absolutely nothing to do with the KM that I was taught back in business school. I will discuss this latter issue in greater detail in the future.

Types of Knowledge

Once knowledge is created, it exists within the organization. However, before it can be reused or shared it must be properly recognized and categorized. Within business and KM, two types of knowledge are usually defined, namely explicit, tacit knowledge and Embedded knowledge.

- **Explicit Knowledge:** This is largely a process of sorting through documents and other records, as well as discovering knowledge within existing data and knowledge repositories. For the latter, IT can be used to uncover hidden knowledge by looking at patterns and relationships within data and text. The main tools/practices in this case include intelligence gathering, data mining (finding patterns in large bodies of data and information), and text mining (text analysis to search for knowledge, insights, etc.). Intelligence gathering is closely linked to expert systems (Bali et al 2009) where the system tries to capture the knowledge of an expert, though the extent to which they are competent for this task is questionable (Botha et al 2008).
- **Tacit knowledge:** Discovering and detecting tacit knowledge is a lot more complex and often it is up to the management in each firm to gain an understanding of what their company's experts actually know. Since tacit knowledge is considered as the most valuable in relation to sustained competitive advantage, this is a crucial step, a step that often simply involves observation and awareness. There are several qualitative and quantitative tools/practices that can help in the process; these include knowledge surveys, questionnaires, individual interviews, group interviews, focus groups, network analysis, and observation. IT can be used to help identify experts and communities. Groupware systems and other social/professional networks as well as expert finders can point to people who are considered experts, and may also give an indication of the knowledge these people/groups possess.
- **Embedded knowledge:** This implies an examination and identification of the knowledge trapped inside organizational routines, processes, products etc, which has not already been made explicit. Management must essentially ask "why do we do something a certain way?" This type of knowledge discovery involves observation and analysis, and the use of reverse engineering and modeling tools.

ROI Illustrating the return-on-investment (ROI) for a portal solution or knowledge management (KM) system measuring the ROI on improved processes and increased economic value of employee performance. Thus, rather than employing traditional notions of value and assets as noted in standard accounting practices, KM solutions are tools managers should use to support opportunities for process improvement and redesign. ROI that measures value from this perspective creates new areas of value from an organization's existing, undervalued assets. A well-developed measurement methodology for implementing a KM system may illustrate ROI, justify expenditures for implementing the system, and provide a format to ensure that process improvement occurs. A well-thought-out KM system has the capability of becoming the “digital nervous system” of an organization, tying all areas to the strategic goals of an organization.

Write brief about failure of knowledge management?

KM Failure Factors

Based on the works of numerous researchers and authors, I arrived at two categories of factors, namely "causal" and "resultant".

Causal factors refer to fundamental problems within the organization, which lead to conditions that are not suitable for KM. They are not always easily visible and they lead to a number of symptoms, which I have termed “resultant” factors.

Causal Failure Factors:

- Lack of performance indicators and measurable benefits
- Inadequate management support
- Improper planning, design, coordination, and evaluation
- Inadequate skill of knowledge managers and workers
- Problems with organizational culture
- Improper organisational structure

Resultant Failure Factors:

- Lack of widespread contribution
- Lack of relevance, quality, and usability
- Overemphasis on formal learning, systematisation, and determinant needs
- Improper implementation of technology
- Improper budgeting and excessive costs
- Lack of responsibility and ownership
- Loss of knowledge from staff defection and retirement

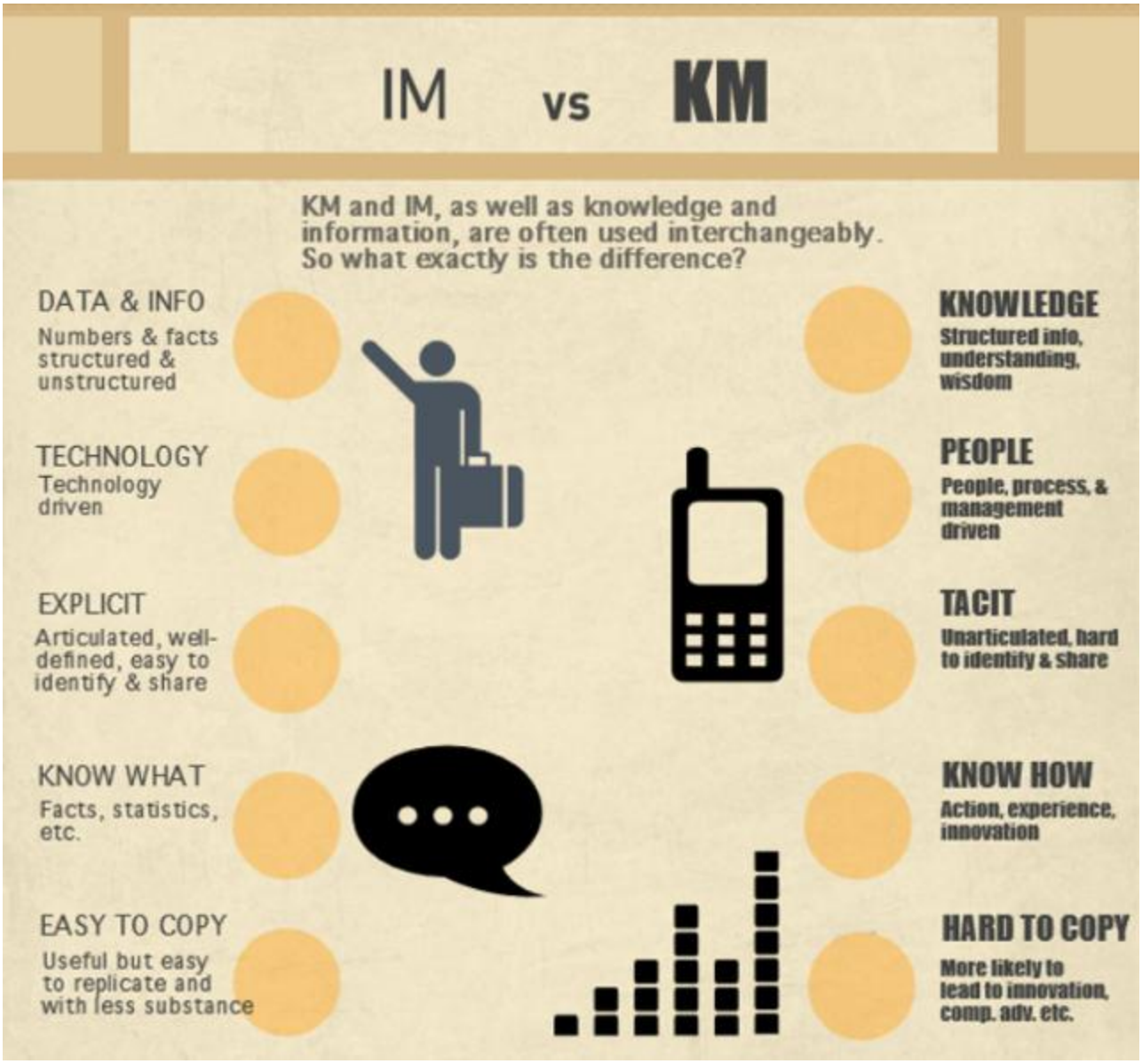
Difference between Knowledge management and information management

Information and IM:

- Focus on data and information
- Deal with unstructured and structured facts and figures.
- Benefit greatly from technology, since the information being conveyed is already codified and in an easily transferrable form.
- Focus on organizing, analyzing, and retrieving - again due to the codified nature of the information.
- Is largely about know-what, i.e. it offers a fact that you can then use to help create useful knowledge, but in itself that
- fact does not convey a course of action (e.g. sales of product x are up 25% last quarter).
- Is easy to copy - due to its codified and easily transferrable nature.

Knowledge and KM:

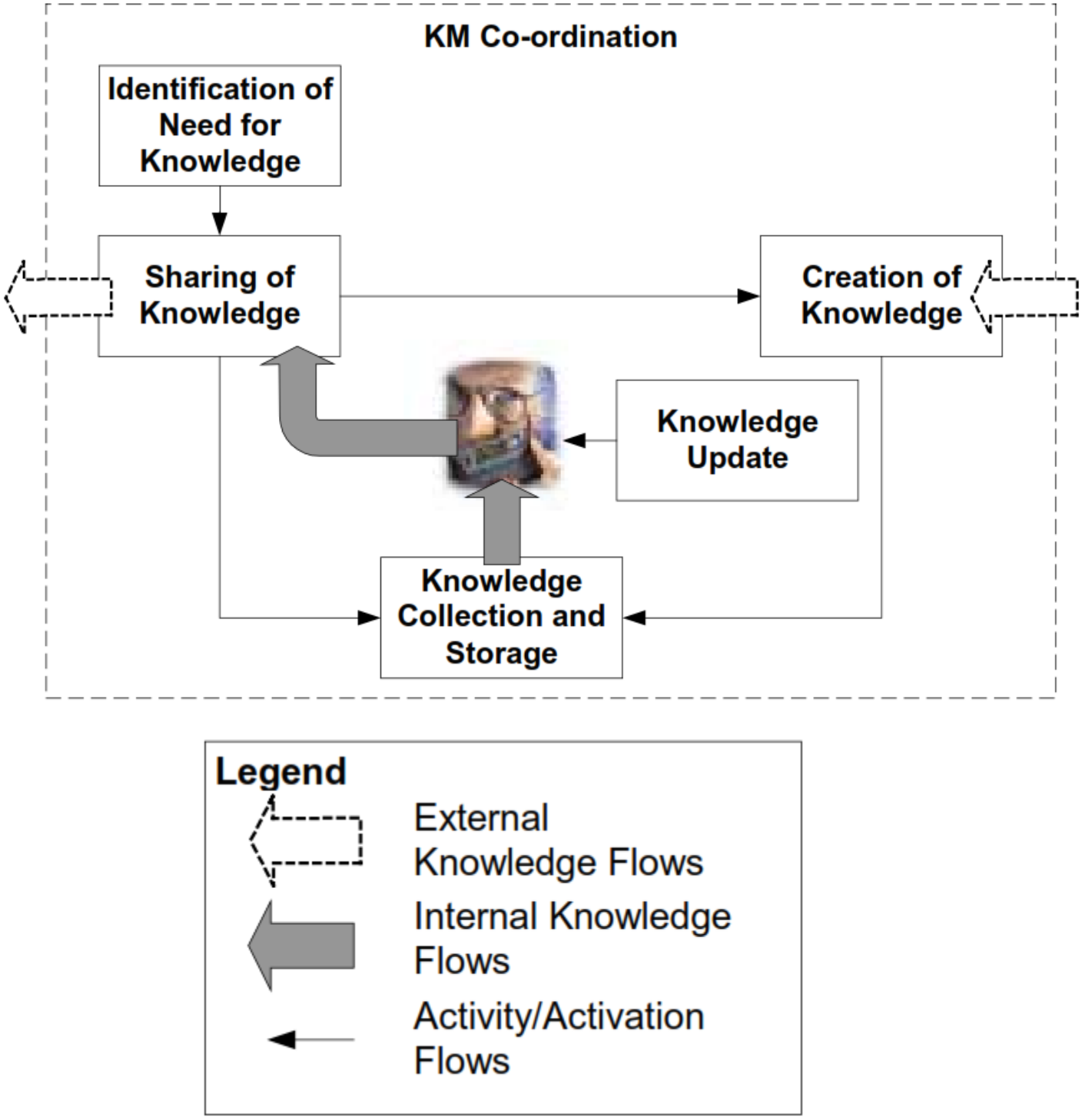
- Focus on knowledge, understanding, and wisdom
- Deal with both codified and uncodified knowledge. Uncodified knowledge - the most valuable type of knowledge - is found in the minds of practitioners and is unarticulated, context-based, and experience-based.
- Technology is useful, but KM's focus is on people and processes. The most valuable knowledge cannot effectively be (directly) transferred with technology, it must be passed on directly from person to person.
- Focus on locating, understanding, enabling, and encouraging - by creating environments, cultures, processes, etc. where knowledge is shared and created.
- Is largely about know-how, know-why, and know-who
- Is hard to copy - at least regarding the tacit elements. The connection to experience and context makes tacit knowledge extremely difficult to copy. This is why universities cannot produce seasoned practitioners - there are some things (the most important things) that you simply cannot teach from a textbook (or other codified source of information/explicit knowledge). These are learnt in the field and understood on an intuitive level. You cannot easily copy or even understand this intuition without the right experience, context, etc. - and it is this intuition that represents the most valuable organizational knowledge.



Knowledge management process

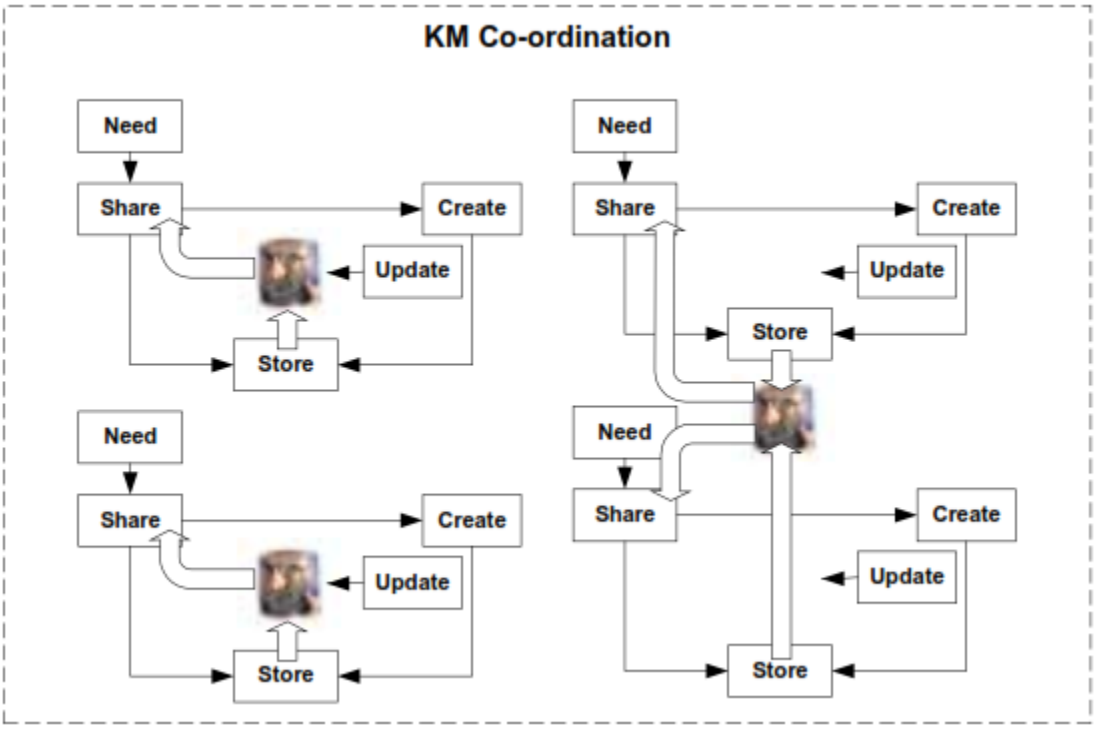
The operational processes present the processes of actually carrying out KM, i.e. knowledge collection, sharing, update, etc. Before elaborating on the processes and their sub-processes in the following sections, an overview of the model is

provided below: Figure shows the main processes of the model and their basic dependencies.



Overview of the Main Processes.

The co-ordination processes are underlying the operational processes. In Figure, this is shown by the rectangle lying behind all other processes. The operational processes are presented as the following main processes: "Identification of Need", "Sharing", "Creation", "Collection and Storage", and "Update". Please note that there are two processes that represent the main process "Sharing" in the model: "Knowledge Pull" and "Knowledge Push". The arrows connecting the processes



Knowledge management domain

provide an overview of the interaction and knowledge flows. The picture in the middle represents the place where the knowledge is stored. The purpose of this picture, showing a human and a machine, is to express the variety of possible ways of storing knowledge, including both technical (databases, documents, videos) and non-technical
(human mind) repositories.
The general concept of the process model is that within the coordinating processes the operational processes are planned and initiated. Together these make up the KM system. The main processes are described in the following. "Identification of Need for Knowledge" identifies a need for knowledge and determines it. "Sharing" is initiated in order to find out whether knowledge that already exists in the system can be used. This covers both the searching for knowledge by a person who needs the knowledge ("Knowledge Pull") and the feeding of knowledge to recipients who are known to be in need of it
("Knowledge Push"). If the needed knowledge is not available yet, "Creation of Knowledge" is initiated. Consequently, the new knowledge (the result) has to be collected – this is done in "Knowledge Collection and Storage".

Characteristics of knowledge: The most important characteristic of knowledge is non-rivalry, which means that one person's use of an idea does not preclude another person using it at the same time.

1. Knowledge is contextual and it can be re-used
2. Benefits of knowledge obtained only if it is applied
3. The values of knowledge may change over time
4. Knowledge has to be renewed or maintained
5. It can be difficult to transfer, capture and distribute knowledge
6. It is developed through learning processes
7. Depends on memory, past experience, expertise, knowledge transfer mechanisms, opportunities
8. Facilitates effectiveness and 'sense-making'
9. Knowledge enables higher learning.
10. Knowledge creation and utilization is enhanced with technology.

Difference between knowledge and information

Information	Knowledge
Static	Dynamic
Independent of the individual	Dependent on individual
Explicit	Tacit
Digital	Analogue
Easy to duplicate	Must be re-create
Easy to broadcast	Face-to-face mainly
No instrinctic meaning	Meaning has to be personally assigned

Difference between Tacit knowledge and Explicit knowledge
The distinction between tacit and explicit knowledge is perhaps the most fundamental concept of knowledge management. Such a distinction was first made by Michael Polyani in the 1960s, but it forms one of the central planks of Nonaka and Takeuchi's book The Knowledge-Creating Company (1995)

Tacit knowledge (knowing-how): knowledge embedded in the human mind through experience and jobs. Know-how and learning embedded within the minds of people. Personal wisdom and experience, context-specific, more difficult to extract and codify. Tacit knowledge Includes insights, intuitions.

Explicit knowledge (knowing-that): knowledge codified and digitized in books, documents, reports, memos, etc. Documented information that can facilitate action. Knowledge what is easily identified, articulated, shared and employed.

Thus, explicit (already codified) and tacit (embedded in the mind).

Explicit knowledge	Tacit (implicit) knowledge
Objective, rational, technical	Subjective, cognitive, experiential learning
Structured	Personal
Fixed content	Context sensitive/specific
Context independent	Dynamically created
Externalized	Internalized
Easily documented	Difficult to capture and codify
Easy to codify	Difficult to share
Easy to share	Has high value
Easily transferred/ taught/learned	Hard to document
Exists in high volumes	Hard to transfer/teach/learn
	Involves a lot of human interpretation

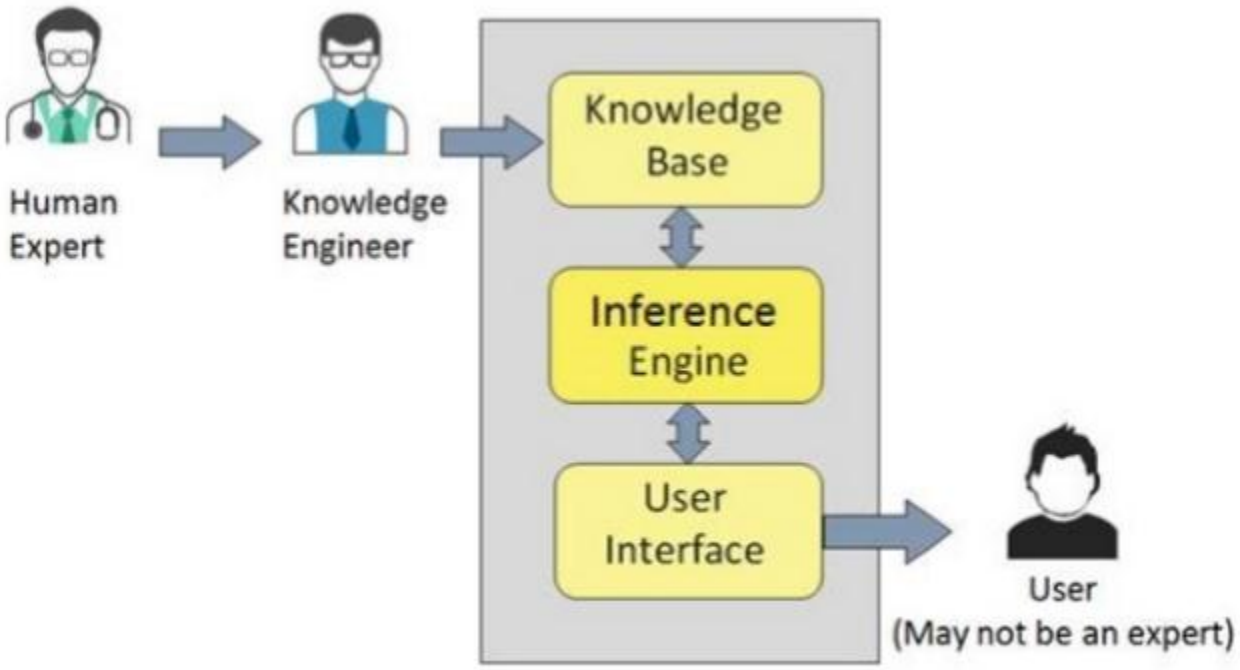
Expert system

Expert systems (ES) are one of the prominent research domains of AI. It is introduced by the researchers at Stanford University, Computer Science Department.
The expert systems are the computer applications developed to solve complex problems in a particular domain, at the level of extra-ordinary human intelligence and expertise.

Artificial intelligence based system that converts the knowledge of an expert in a specific subject into a software code. This code can be merged with other such codes (based on the knowledge of other experts) and used for answering questions (queries) submitted through a computer. Expert systems typically consist of three parts:

(1) **knowledge base** which contains the information acquired by interviewing experts, and logic rules that govern how that information is applied;
(2) **Inference engine** that interprets the submitted problem against the rules and logic of information stored in the knowledge base; and an
(3) **User Interface** that allows the user to express the problem in a human language such as English.

Let us see them one by one briefly −



Despite its earlier high hopes, expert systems technology has found application only in areas where information can be reduced to a set of computational rules, such as insurance underwriting or some aspects of securities trading. Also called rule based system

Characteristics of Expert Systems

- High performance
- Understandable
- Reliable
- Highly responsive

Capabilities of Expert Systems

The expert systems are capable of −

- Advising
- Instructing and assisting human in decision making
- Demonstrating
- Deriving a solution
- Diagnosing
- Explaining
- Interpreting input
- Predicting results
- Justifying the conclusion
- Suggesting alternative options to a problem

Expert Systems Limitations

No technology can offer easy and complete solution. Large systems are costly, require significant development time, and computer resources. ESs have their limitations which include −

- Limitations of the technology
- Difficult knowledge acquisition
- ES are difficult to maintain
- High development costs

Applications of Expert System

The following table shows where ES can be applied.

Application	Description
Design Domain	Camera lens design, automobile design.
Medical Domain	Diagnosis Systems to deduce cause of disease from observed data, conduction medical operations on humans.
Monitoring Systems	Comparing data continuously with observed system or with prescribed behavior such as leakage monitoring in long petroleum pipeline.
Process Control Systems	Controlling a physical process based on monitoring.
Knowledge Domain	Finding out faults in vehicles, computers.
Finance/Commerce	Detection of possible fraud, suspicious transactions, stock market trading, Airline scheduling, cargo scheduling.

Challenges of knowledge management

In order to maximize the benefit of knowledge management within your business you may have to overcome the following challenges:

- **Capturing and recording business knowledge -** ensure your business has processes in place to capture and record business knowledge.
- **Sharing information and knowledge** – develop a culture within your business for sharing knowledge between employees.
- **Business strategy and goals** – without clear goals or a business strategy in place for the knowledge gathered the information will be of no use to your business.
- **Knowledge management systems** – these systems can be costly and complex to understand but when utilised properly can provide huge business benefits. It is important that staff are fully trained on these systems so that they collect and record the right data.

Advantages of knowledge management

Consider the measurable benefits of capturing and using knowledge more effectively in your business. The following are all possible outcomes:

- An improvement in the **goods or services** you offer and the processes that you use to sell them. For example, identifying market trends before they happen might enable you to offer products and services to customers before your competitors.
- Increased **customer satisfaction** because you have a greater understanding of their requirements through feedback from customer communications.
- An increase in the quality of your **suppliers**, resulting from better awareness of what customers want and what your staff require.
- Improved **staff productivity**, because employees are able to benefit from colleagues' knowledge and expertise to find out the best way to get things done. They'll also feel more appreciated in a business where their ideas are listened to.
- Increased business **efficiency**, by making better use of in-house expertise.
- Better **recruitment and staffing** policies. For instance, if you have increased knowledge of what your customers are looking for, you're better able to find the right staff to serve them.
- The ability to **sell or license** your knowledge to others. You may be able to use your knowledge and expertise in an advisory or consultancy capacity. In order to do so, though, make sure that you protect your intellectual property.

Benefits of knowledge management

All organization can benefit from their people sharing, inmoving, reusing, collaborating and learning.

1. Enabling fast and better decision making.
2. Making it easy to find relevant information and resource.
3. Reusing ideas, documents and expertise.
4. Avoiding redundant efforts.
5. Avoiding making the same mistake again.
6. Taking advantage of existing expertise and experience.
7. Communicating important information widely and quickly.
8. Providing methods, tools, templates, techniques and examples.
9. Enabling the organization to leverage its size.
10. Stimulating innovation and growth.

Limitations of Knowledge management

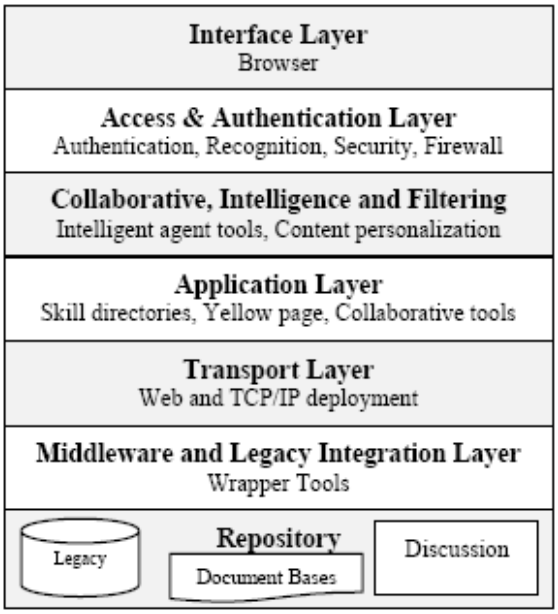
1. Failure to use company knowledge properly can lead to a great loss of time, resources, and even organization failure.
2. Knowledge sharing is crucial part of making knowledge management systems works, but most organization fail to share proper knowledge.
3. Extracting information from workers who possess valuable company knowledge can also be a difficult and lengthy process.
4. Another major disadvantages of knowledge management system is the lack of company strategy to fully utilize the information that it collects.
5. Without an implementation strategy or goal in place for the knowledge, the information useless.
6. Knowledge management systems are complex and hard to understand for the average worker, and training worker to use knowledge management system is costly.

Knowledge Management System Architecture

Developing a KMS is a complex task and requires a careful planning before selecting the tools for supporting the knowledge processes. The designed system architecture should suit the organizational culture and business needs. KMS can be as simple as a file folder until a complex business intelligence system which uses an advanced data visualization and artificial intelligence. Thus, we have studied several KMS architectures which aim to support knowledge management processes and collaboration in the organization. We found that even if there are differences between architectures in term of functions and services, the major components of architecture are comparable. The general KMS architecture is proposed by Tiwana [Tiwana 02]. He pointed out that the KMS should comprise four major components: repository, collaborative platform, network, and culture.

1. Repository holds explicated formal and informal knowledge, such as declarative knowledge, procedural knowledge, causal knowledge, and context. This component acts as the core of KMS which aims to store and retrieve knowledge for future use.
2. Collaborative platform supports distributed work and incorporates pointers, skills databases, expert locators, and informal communications channels.
3. Network means both physical and social networks that support communication and conversation. Physical network is a 'hard' network such as intranet, shared space, and back bone. Social network is a 'soft' network such as Communities of Practice (CoP), associations, and working groups.
4. Culture is the enabler to encourage sharing and use of the KMS. Research has revealed that the greatest difficulty in KM is ''changing people's behavior,'' and the current biggest impediment to knowledge transfer is ''culture''.

These four components are considered as the basis elements for each knowledge management system. However, other tools could be integrated to enhance the quality of services of the system. Tiwana also proposed seven-layer KMS architecture [Tiwana 02] which is the integration of these four components and their supportive information technologies.



Seven layers KMS architecture [Tiwana 02]

Actually, seven layer KMS architecture is just a reflection of OSI model (Open Systems Interconnection basic reference model). This model tries to represent the functions and tools of KMS in terms of layer that the

knowledge passed though. This architecture might suit with complex systems which require network and data manipulation.

Why is it helpful to view the building of a KM system as a life cycle?

It is important to have a life cycle in building knowledge management systems, because the life cycle provides structure and order to the process. Additionally, the life cycle provides a breakdown of the activities into manageable steps, good documentation for possible changes in the future, coordination of the project for a timely completion, and regular management review at each phase of the cycle.

Write down the phases of knowledge management

A winning knowledge management program increases staff productivity, product and service quality, and deliverable consistency by capitalizing on intellectual and knowledge-based assets.

Many organizations leap into a knowledge management solution (e.g. document management, data mining, blogging, and community forums) without first considering the purpose or objectives they wish to fulfill or how the organization will adopt and follow best practices for managing its knowledge assets long term.

A successful knowledge management program will consider more than just technology. An organization should also consider:

- **People**. They represent how you increase the ability of individuals within the organization to influence others with their knowledge.
- **Processes**. They involve how you establish best practices and governance for the efficient and accurate identification, management, and dissemination of knowledge.
- **Technology**. It addresses how you choose, configure, and utilize tools and automation to enable knowledge management.
- **Structure**. It directs how you transform organizational structures to facilitate and encourage cross-discipline awareness and expertise.
- **Culture**. It embodies how you establish and cultivate a knowledge-sharing, knowledge-driven culture.

8 Steps to Implementation

Implementing a knowledge management program is no easy feat. You will encounter many challenges along the way including many of the following:

- Inability to recognize or articulate knowledge; turning tacit knowledge into explicit knowledge.
- Geographical distance and/or language barriers in an international company.
- Limitations of information and communication technologies.
- Loosely defined areas of expertise.
- Internal conflicts (e.g. professional territoriality).
- Lack of incentives or performance management goals.
- Poor training or mentoring programs.

- Cultural barriers (e.g. “this is how we've always done it” mentality).

The following eight-step approach will enable you to identify these challenges so you can plan for them, thus minimizing the risks and maximizing the rewards. This approach was developed based on logical, tried-and-true activities for implementing any new organizational program. The early steps involve strategy, planning, and requirements gathering while the later steps focus on execution and continual improvement.

Define

1. **RER: Rapid Evidence Review:** A Rapid Evidence Review is a way of reviewing research and evidence on a particular issue. It looks at what has been done in particular area and records the main outcomes. Evidence review can be run in several ways. Some are more exhaustive in their execution and ambitious in their scope.
   The RER provides a quicker review but still useful way of gathering and consolidating knowledge. It’s useful building block from which to start work on a new project.

2. **CASE STUDY:** A case study is a written examination of a project, or important part of a project. It has a clear structure that brings out key qualitative and quantitative information from the project. Case studies are also published with a broad audience in mind, so it is useful to bring the most useful and transferable information to the fore.

3. **KNOWLEDGE BANK:** Knowledge banks are online services and resources which hold information, learning and support, giving you the power to improve your council. They are typically used to showcase the work of an organization and provide signposts to documents, article and toolkits.

4. **KNOWLEDGE CAFÉ:** A knowledge café people brings together to have open creative conversation on topics of mutual interest. It can be organized in a meeting or workshop format, but the emphasis should be on following dialogue that allows people to share ideas and learn from each other. It encourages people to explore issues that require discussion in order to build a consensus around an issue.

Why Use a Knowledge Café?
In an organization, especially in a hierarchical organization, people are not often given the opportunity to 'reflect' on discussions. People are normally tied to performance pressures. Therefore, much of the value that could be gained from good discussion, dialogue, and reflection is lost.

5. **KNOWLEDGE MARKETPLACE:** Knowledge Marketplace – Modelled on the E-Business Net Market concept, several knowledge-trading places have recently been established. In a Knowledge Marketplace, a third party vendor hosts a web site grouping together many suppliers of knowledge services. Suppliers may include expert advisors, vendors providing product support services, KM job placement agencies, procedures for the evaluation of KM and portal software, and research companies providing industry benchmarks and best practice case studies. Two types of Knowledge Marketplace exist.one provides common information and services to all industries while the other offers only certain services to a specific industry.

6. **COTS: Customized Off The Shelf (COTS)** – this is the traditional and most popular way of deploying application services. Based on the organizational needs, the applications will be identified and then examined against the functional needs of the organization. A short-period test may follow to identify the most suitable application. Once an application is acquired,

customization of the standard features is usually performed to integrate it into the organization's information system.

7. **BRAINSTORMING:** Brainstorming is a process where a group of people meet to focus on a problem, or idea, and explore such ideas with a view to coming up with solutions, or further developing the ideas. The participants express or contribute their ideas as they strike them and then build on the ideas raised by others. All the ideas are noted down and are not criticized. Only when the brainstorming session is over are the ideas evaluated. Brainstorming helps in problem solving and in creating new knowledge from existing knowledge.

Brainstorming is a simple way of helping a group of people to generate new and unusual ideas. The process is actually split into two phrases: divergence and convergence. During the divergent phase, everyone agrees to delay their judgment. In other words, all ideas will be treated as valid. During the convergent phrase, the participants use their judgment but do so in a 'positive' manner—that is, they look for what they like about the ideas before finding flaws.

When to use brainstorming

Used to bring different perspectives to a problem, find key areas to focus on in a project or test new methods, brainstorming usually happens in a workshop or meeting with small and large groups working together on ideas.

8. **ROI: Return on investment** measures the gain or loss generated on an investment relative to the amount of money invested. ROI is usually expressed as a percentage and is typically used for personal financial decisions, to compare a company's profitability or to compare the efficiency of different investments.

9. **COP** Communities of Practice (CoP) are also called knowledge communities, knowledge networks, learning communities, communities of interest and thematic groups. These consist of a group of people of different skill sets, development histories and experience backgrounds that work together to achieve commonly shared goals (Ruggles, 1997). These groups are different from teams and task forces. People in a CoP can perform the same job or collaborate on a shared task, e.g. software developers, or work together on a product, e.g. engineers, marketers, and manufacturing specialists.

What are Social Network Services?

A social network is a group of people who share a common area of interest. Social network services are online systems that support social networking. The core services they offer usually include

1. Finding people who have similar interests or needs;
2. Aggregating people into groups, or subgroups, and being able to communicate with those groups; and
3. Sharing content, such as documents links to relevant websites, or even streaming video.

What is Knowledge Mapping?

Knowledge Mapping is a process by which organizations can identify and categorize knowledge assets within their organization—people, processes, content, and technology. It allows an organization to leverage the existing expertise resident in the organization, as well as identify barriers and constraints to fulfilling strategic goals and objectives. It is constructing a road map to locate the information needed to make the best use of resources, independent of source or form