INFORMATION TECHNOLOGY IN DAIRY INDUSTRY

A. P. Ruhil

R. R. B. Singh

INFORMATION TECHNOLOGY IN DAIRY INDUSTRY

A. P. Ruhil

Dairy Economics, Statistics and Management Division

NDRI, Karnal

R. R. B. Singh

Dairy Technology Division

NDRI, Karnal

Keep smiling.....

Index

Lesson 1

INFORMATION TECHNOLOGY: IMPORTANCE AND APPLICATIONS IN DAIRY INDUSTRY

1.1 Introduction

This lesson will introduce the concept and strength of Information technology (IT), importance of computerization in dairy industry, applications of IT in Dairy Industries and an overview of course contents. These topics will be useful to create IT awareness among the students and its importance and applications in dairy industry.

IT has emerged as a frontier know how for addressing complex problems related to any subject. Innovations made in the field of information technology have revolutionized the ways of collection, storage, retrieval, processing & communication of qualitative & value added information. Information technology has made a great impact on the society. In computer age, not even a single area whether scientific, commercial, art, social, etc., is untouched from the benefits of information technology. It has, therefore, become an integral part of academic curriculum of almost every discipline. IT has played a significant role in food and dairy processing as well, particularly in the solution of problems, data collection and analysis, online production monitoring, image processing, sensory evaluation, quality control, process control, etc., in addition to routine applications such as word processing and graphics.

The dairy sector has grown many folds during the last few decades, the changes being particularly perceivable after the liberalization of Indian economy, which has opened new opportunities in this area. Though India is the number one milk producing country in the world, most of the milk is consumed in raw form rather than milk products partly due to lack of adequate milk processing facilities. The demand for milk and milk products is increasing day by day. With the rise in economic status, the consumer is more conscious about quality of milk products. Multi-national companies are coming forward for setting up of large dairy farms and milk processing plants with huge investments. To keep high quality standards, these upcoming plants are fully automated. Most of the machines are process control and computer numerical control (CNC) based to perform routine plant operations. Activities related to personnel, finance, marketing, inventory management, supply chain management etc. are being computerized using high end software such as ERP, SAP, etc.

1.2 Information Technology

IT has become a popular word among common men. IT is a converging spectrum of technologies that includes telecommunications (such as telephony, cable, satellite, and computer networks), computing (computers, intranet, the Internet, software and mobile phones) and broadcasting (Radio and TV technologies). Internet is the latest in the range of IT spectrum. IT facilitates in collection, storage, processing, analyzing, retrieval, transfer and dissemination of information and communication services.

Information technology is associated with mankind since the development of printing press. Therefore, it is not a new item, only the method and tools have changed from old time. Developments in the field of IT have transformed the world in information and communication sector. They have made old services

more efficient and as well as added new services and new dimensions in the services. IT was the key technology of the 20th century and will also play a lead role in 21st century. Reduction in the prices of computing and communication devices has played a significant role in bringing IT to the door step of common man. IT has made a great impact on all corners of the society. Today we are living in the age of "Information Society", that is a society where the ability to access, search, use, create, and exchange information is the key for individual and collective well being. Internet along with World Wide Web (WWW) is one of the most interesting and exciting phenomena of IT for information dissemination.

Information technology includes all matters concerned with the furtherance of computer science and technology and with the design, development, installation and implementation of information systems and applications. Information technology architecture is an integrated framework for acquiring and evolving IT to achieve strategic goals. It has both logical and technical components. Logical components include objectives, functional and information requirements, system configurations, and information flows. Technical components include IT standards and rules that will be used to implement the logical structure. *The technology of computers, telecommunications, and other devices that integrate data, equipment, personnel, and problem solving methods in planning and controlling business activities is called Information Technology*.

Government of India has also recognized the importance of IT applications in all functional areas including animal husbandry and dairying. This is evidently reflected from the recommendations of the National Task Force on IT for linking all villages through computer network. Subsequently, some progressive states and organizations/institutions have made pioneering efforts in this direction. As a result, a few villages have been linked together through a computer network in some regions of the country, viz., Warna Wired Villages (in Maharashtra state) and MS Swaminathan Info Villages in Chennai, Information Village of MANAGE, Hyderabad, e-choupals developed by ITC Ltd., Secundrabad, etc. These info villages are connected thorough hybrid form of wired and wireless technologies for communications. They are using Indian regional and local languages for effective interaction among the farmers and the experts such as Tamil, Marathi, Hindi, etc.

1.2.1 Strength of IT

Major strength of IT lies in collection, storage, processing, analyzing, retrieval and dissemination of information. The application of electronic communication, information digitization and imaging technologies offers new and effective methods of developing and delivery of instructions, training and learning over the traditional methods. Following are the advantages of using IT tools:

- Speed and accuracy of data processing
- Large storage capacity of data or information
- Variety of tasks can be performed
- Automation of activities for example process control applications
- Reduction in cost in long run
- Reduced manpower requirement
- Better quality of work, improved customer image, management confidence, etc.
- Promoting informatics led resource planning and management
- Strengthening research, education, training, extension and development linkages
- Improved marketing of milk and milk products
- E-commerce for direct linkages between local producers, traders, retailers and suppliers.

1.3 Importance of Computerization in Dairy Industry

In India, dairy sector is growing very fast. At present, India is the largest milk producing country in the world. India's milk production is increasing at the rate of more than 3.5% while in the world it is increasing at rate of 1%, approximately. A number of dairy plants in organized sector are coming up to handle large quantity of milk to meet the demand of public for providing quality milk and milk products at low cost. A large dairy plant involves diversified activities like procurement of milk, processing of milk, manufacturing of milk products, marketing, inventory control, human resource management, etc. The complexity in these operations increases with scaling up of handling capacity of milk, manufacturing variety of milk products and multiple locations of manufacturing units. Therefore, it is imperative to employ state-of-the-art information technology tools to manage multifaceted operations of dairy plant and to disseminate information efficiently and effectively by minimizing losses during handling, processing, manufacturing and transportation of milk and milk products.

Applications of information technology have been proved very beneficial in management and operation of dairy plants. Computer based information systems and process controlled machines have assisted the management in speeding up data processing activities and maintaining quality control of milk products. For example various computerized operations like receiving of milk at dock yard and checking its quality instantly, timely payment to suppliers, monitoring manufacturing process, labor efficiency, PFA standard of milk products, tracking employee's record, etc., have shown tremendous improvement and impact of IT on end users.

Realizing the importance of computerization and information technology two decades back, National Dairy Research Institute (NDRI), Karnal introduced one full semester course titled "**Computer Application in Dairy Industry**" in the 4th year curriculum of B.Tech (Dairy Technology) programme in the year 1990. Subsequently, various State Agriculture Universities (SAU) adopted this course for B.Tech. (DT) programme. High level committee established by ICAR for reforming UG and PG courses offered by ICAR institutes and SAUs, has revised the contents of this course and renamed it as "**IT in Dairy Industry**" which is currently being offered during 3rd year of B.Tech. (DT) programme.

1.4 IT Applications in Dairying

Dairying is one of the prospective area in which IT can be effectively applied particularly for office automation, plant operations, food and dairy processing. Successful attempts have been made in the past to computerize various aspects of dairying to improve productivity and efficiency of dairy plants. A few important IT tools and applications which have immensely benefitted the dairy industry are discussed below:

1.4.1 Computer network

A computer network may be defined as "an interconnected collection of autonomous computers". Two computers are said to be interconnected if they are able to exchange information. The connections may be through copper wires, optical fibers, and wireless maeans. A network may cover a small geographic area connecting devices in a single building or group of buildings that is known as Local Area Network (LAN). A network that covers a large area such as a state, country or the world is called a Wide Area Network (WAN). Computer networks are used for sharing resources and communication within organizations.

1.4.2 The Internet

Internet is a powerful tool to disseminate information among masses worldwide. Internet is a global collection of computers, which are interconnected to each other through wired or wireless media and

makes communication possible among people in a common language. It provides a number of services and tools to access information, for example, e-mail, video conferencing, Usenet and newsgroups, marketing and advertisements, hosting Web sites etc. Internet is growing explosively because of its enormous potential in providing the latest information.

1.4.3 Touch screen information kiosk

This is a powerful medium to disseminate reliable information among a large group of workers and users. It is just like an ordinary computer with touch screen monitor. Information to be disseminated is prepared using visual graphics and multimedia tools. Kiosk can be used to display standard practices on dairy operations, staff position, organizational setup, products manufactured by dairy plant or other informative material to educate visitors and staff members.

1.4.4 Wireless sensor network

A wireless sensor network generally consists of spatially distributed sensor nodes and base station(s) (or "gateway") that can communicate with a number of wireless sensor nodes scattered in a region via a radio link to cooperatively monitor physical or environmental conditions, such as temperature, humidity, sound, vibration, pressure, motion, air pollutants, etc. Data are collected at the wireless sensor node, compressed, and transmitted to the gateway directly or, if required, it uses other wireless sensor nodes to forward data to the gateway. Widespread use of sensor network can improve the monitoring of dairy plant and generate valuable inputs for crisis management. The variety of potentially useful sensors can be applied in different situations to improve the productivity and efficiency of dairy plants, for example, environmental sensors to monitor temperature, video cameras/ image sensors for security, plant operations and observing behavior of staff, visitors, etc. Sensor networks are also being used for real time monitoring of product manufacturing process to diagnose defects in milk products.

1.4.5 Radio frequency identification

Radio Frequency Identification (RFID) is a dedicated short range communication system. RFID may be defined as "a system that transmits the identity (in form of a unique serial number) of an object (or person) wirelessly using radio waves". This technology is commonly implemented using RFID tags (also known as transponders) mounted or attached on objects and readers/ writers (also known as interrogators). RFID tags consist of two parts - an integrated circuit (containing unique code, memory for storing and processing of data, modulating and demodulating radio signals, etc.) and antenna for receiving and transmitting signals. These components are enclosed in a temper proof plastic cover of different sizes and shapes. RFID tags are low cost, unique and tamper proof. RFID technology is being used worldwide to trace the dairy and livestock products in stores and during transportation. RFID is also used in various sections dairy plants for tracking the products in storage as well as on production lines.

1.4.6 Global positioning system

The Global Positioning System (GPS) is a satellite-based global navigation system that provides geographical location (longitude, latitude and altitude) and time information of an object anywhere in the world for 24 hours a day in all weather conditions. GPS is owned, operated and maintained by US Government and is freely accessible by anyone with a GPS receiver for general use around the world. GPS is made up of a network of 24 well spaced satellites placed into the orbit. The location accuracy varies from 10 to 100 meters for most equipment while accuracy can be pinpointed to within one meter

with special military equipment. GPS equipment is widely used in science and has now become sufficiently low-cost so that almost anyone can own a GPS receiver.

GPS is useful to trace vehicles during milk transportation from collection centers to central dairy dock yard, and marketing of dairy products to minimize malpractices on route. GPS device is mounted on the vehicle, central computer in control room will keep track of the vehicle through satellite communication. With the advent of low cost GPS devices this technology is gaining popularity.

1.4.7 Decision support system

Decision Support System (DSS) is a specific class of computer based information system that supports technological and managerial decision-making by assisting the organization with knowledge about ill-structured or semi-structured issues. Various applications of DSS in diary industry are such as *Cow Culling Decision Support System*, *MCLONE3* - A Decision Support System for Management of Liquid Dairy and Swine Manure, *Dairypert*, *NutMan*, *Dairypro* - An individual Feed Allocation Decision Support System for the Dairy Farm, Economics for opening a dairy farm for a given number of animals. DSS developed at NDRI, Karnal - An interactive data- driven DSS on dairy production aspects, MSI-NDRI, Web-enabled information system for online searching, ordering and maintenance of dairy cultures, Multimedia information system on transferable dairy technologies.

1.4.8 Expert system

An expert system is a computer program that contains stored knowledge and solves problems in a specific field in much the same way that a human expert would. The knowledge typically comes from a series of conversations between the developer of an expert system and one or more subject matter experts. Expert systems are developed when the problem occurs frequently, and is complex, knowledge of the expert is required, uncertainty is involved, problem is dynamic and involves heuristic factors, and practical value is high. Examples: Animal disease diagnosis system, maintaining quality of dairy products, etc.

1.4.9 Artificial neural networks

Artificial Neural Network (ANN) is an emerging IT tool in the area of artificial intelligence, where the goal is to develop tools capable of performing cognitive functions such as reasoning and learning. This approach consists of creating a system to mimic the processing power of the human brain. ANN tools have been used by researchers to solve problems related to dairy and food processing worldwide and now in India also. Artificial neural networks are used for pattern recognition aimed at characterization of food properties and for predicting quality deterioration during storage life of the products. A few applications of ANN in dairy are:

- Machine perception including machine vision, electronic tongue and electronic nose for sensory evaluation
- Spectral data interpretation for identification of functional groups and quantitative analysis
- Food Microbiology and Food Fermentation
- Predicting physical, chemical, and functional properties of food products during processing and distribution.
- Prediction of shelf life of dairy products based on sensory evaluation

1.4.10 Akashganga

Shree Kamdhenu Electronics Private Ltd. (SKEPL) developed an IT enabled "**Automatic Milk Collection Station**" marketed under the brand name of AKASHGANGA. It automates the milk collection process at local Dairy Cooperative Society (DCS). This system incorporates an electronic weighing system, a milk analyzer to test milk quality, a personal computer, and accounting and management software that cover all functionalities of the Primary Co-operative Societies. Compared to manual procedures, the AKASHGANGA system reduces labor handling and increases efficiency and transparency in the system, which creates a base for improving the quality of milk produced. That means, milk can be sent on to the cooperative union for processing more quickly, reducing spoilage; farmers can see for themselves the weight and quality of their milk via a display and get printed receipt, which increases their trust in the cooperative process. In addition, farmers are paid immediately, rather than few days later as under manual procedures; and local cooperatives need fewer employees and have better records and reports for planning purposes. The purpose is to integrate entire operations at DCS, e.g., milk receipt, milk sales to unions, local sales, purchases, product sales to farmers, financial procedures such as account ledgers, balance sheet/ profit & loss, audit and inventory management.

1.4.11 ERP applications at Amul dairy

Amul dairy is the largest dairy cooperative society in India with collection of 6 million liters of milk per day from about 10,600 separate village cooperative societies which have approximately 2.1 million milk producing members. To manage such a large network of societies and members is great challenge so as efficiently to improve the system for:

- Weighing of milk and determination of fat and SNF contents
- Calculation of the purchase price
- Timely payment to milk producers without loss of much time to test the milk and work out the purchase price.

Thus the management, in consultation with Tata Consultancy Services implemented Enterprise Resource Planner (ERP) as solution to keep pace with dynamically changing business environment. The project was named as Enterprise wise Integrated Application System (EIAS). This include following IT related actions:

- Automatic Milk Collection System Units (AMCUS) at village societies have been installed in the first phase to automate milk producers logistics

- AMCSU facilitates the capture of member information, milk fat content, volume collected and amount payable to each member
- Indian Institute of Management (IIM), Ahmedabad developed and provided an application software Dairy Information System Kiosk (DISK) to facilitate data analysis and decision support in improving milk collection
- The software DISK offers an extensive knowledge and service delivery mechanism through a Dairy Portal.
- DISK contains a database on the history of cattle owned by the farmers, medical history of these cattle, reproductive cycle and history of diseases
- As a large amount of detailed history on milk production is available in the database, the system can be used to forecast milk collection and monitor the produce from individual sellers.

By implementing IT strategies, the management observed radical changes in business performance. It improved the delivery mechanism and brought transparency in business operations. Processing time for payment to milk producers reduced from a week to few minutes. Distributers are able to place their orders online using Website. ERP helped in monitoring of crucial management practices like demand versus supply and decision making process in easy way since real time data is available on click of a mouse button.

****** ☺ ******

Lesson 2

DATA, INFORMATION AND KNOWLEDGE

2.1 Introduction

This lesson introduces the concept of data, information, knowledge and wisdom. Different types of information and the quality of information are also discussed in the lesson. These topics will be useful for students to develop an understanding of basic difference among data – information – knowledge.

Man learns from its surrounding environment. This has led to accumulation of knowledge since beginning of human history. Creation of knowledge is a complex process and takes a long time to build up knowledge base about any particular subject. Knowledge is dynamic and it is stored in the brain and also embedded in the practice of livelihood of a person. Knowledge is a personal map/ model of the world. Information and data are two major sources to build knowledge base. Data is description of events in understandable coding form. Description of data through coding patterns is in practice even before the Stone Age. People were used to communicate through codes, which were understandable to other party also. For example, pebbles were being used for counting purpose during the early period. After that symbols were developed for counting and recording facts. The recorded data were used to extract meaningful information. Thus, information can be defined as data that is meaningful or useful to the recipient. Information tries to provide answers to "who", "what", "where", and "when" type of questions. Information allows us to expand our knowledge beyond the range of our senses. Knowledge is basically an application of data and information for answering "how" questions. For example, with reference to a dairy plant, data may be daily milk receipt and milk sale figures recorded in registers for many years. Information may be extracted from this data as what is the average milk procurement and milk sale month and year wise? Knowledge drawn from the recorded data may be “low procurement during summer period is due to most of the animals getting dry during this period. Therefore, some breeding policy or strategy may be drawn so that the dry animals are equally distributed throughout the year instead of summer period only” and “keep sufficient stock of dry milk and butter to reconstitute the milk to meet the demand during lean period”.

2.2 Types of Information

Information can be classified on the basis of the purpose for which it is being utilized into following three broad categories:

2.2.1 Strategic information the course of actions to be taken up by a business. For example whether to diversify the business, open a new branch etc. For such decisions one has to obtain information on trend

This is the information needed for long term planning and directing s, impact of advertising, effect of reduction or increase in prices, impact of adverting etc. This type of information is highly unstructured i.e. not well defined. Further, the volume of information in strategic planning is small and difficult to obtain.

2.2.2 Tactical information

This is the information needed to take short term decisions to run business efficiently. For example, should credit limits of customer be changed? How much of each item should be stocked in inventory? Tactical information requires specifically designed processing of data. Most of it is obtainable from day to day collection of routine data. The volume of tactical information is more than strategic information.

2.2.3 Operational information

This is the information needed for day to day operations of a business organization. For example, list of items to be reordered in inventory management system, list of defaulting customers, list of the suppliers who did not supply the material on time to send remainders to them etc. Operational information is easy to obtain by straight forward clerical processing of data. Further the volume of such information is much more than tactical information.

In addition to the above mentioned types of information, there is another type of information called as **statutory information** required by laws and to be sent to government departments. Statutory information pertains to the information relating to a statute, code, or written law, which helps to understand the particular law. The procedures for producing such information are normally clearly specified and required straight forward processing of data.

Different levels of management in an organization require different type of information as per their functional need. Though there is no specific demarcation line to distinguish between the levels or layers of management, however, the management structure of an organization can be described broadly into three levels to carry out the business operations. In a large business organization like a dairy plant it is essential to delegate responsibilities to specialists in each area and make them accountable for their efficient functioning. For example a dairy may have number of sections like procurement, processing, production, sale and marketing, human resource development (HRD), accounts and finance, etc. Each section is headed by a middle level manager and they report to General Manager who is overall in-charge of the organization. Middle level managers in turn will have many assistants who are responsible for specific day-to-day operations. These are called line managers. Further line managers are assisted by the clerical staff to perform clerical and routine tasks. Thus the management structure can be shown using a pyramid shape as (fig. 2.1) given below. Information required at different level of management is also shown in the same figure.

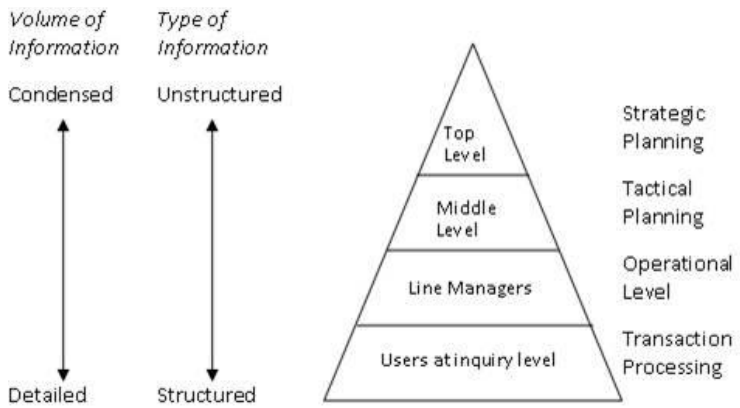


Figure 2.1 Management structure and information required at different levels

2.3 Quality of information

The quality of information refers to its suitability or its reliability. Quality is not an absolute concept; rather it is defined within a context. Some important attributes of information which may affect the quality of information are described as follows:

Quality	**How to ensure quality**
Accuracy	Ensure correct input and processing rule
Completeness	Include all data
Trustworthy	Do not hide unpleasant information
Adequacy	Information required for taking decisions must be available in sufficient quantity and in concise form
Tailor-made	It is essential to give brief summarized information to ensure quick assimilation and action as per requirement
Timely	Information should be given to the consumer when he needs it. Delayed information may sometimes be of no value.
Up-to-date	It should include all data available at the time of processing. In other words, timely information does not mean in time only rather it means timely as well as up-to-date information.
Relevant	Information should satisfy needs of the users and must be useful in taking decisions
Explicitness	Recipient should be able to extract required information from the report without

	wasting any time on further processing
Authority	Information should be given to right person with correct permission
Cost	Generation of information costs to producer/ user therefore, it must be produced and used judicially.
Brief	Summarize relevant information
Understandable	Use attractive format and graphical charts.

2.4 Information Systems

Information is a vital resource to an individual as well as to an organization for efficient operation and management. It is regarded as a resource just like other traditional 'four Ms' – men, money, material and machines. Information is not a natural resource such as water and minerals, nor is it a resource that can be purchased from the market such as man power or raw materials. It is generated within the organization from data collected over a period of time. Sometimes it is also acquired from external sources. At every stage, we need information to take decisions either at individual or group level. The decision making process becomes more complex at an organizational level. Accurate and reliable information must be available timely for taking rational decisions. Managers have the responsibility for solving the larger problems of a company and making decisions among alternative solutions. In the past, they sought information from miscellaneous haphazard sources and processed the information on a personal basis. In today's world, the complexities in the functioning of business have increased many folds due to globalization. The manager's task has become more difficult nowadays because of:

- Liberalization and globalization, organizations are required to compete not locally but globally, therefore managers have to take quick decisions to be ahead of other competitors.
- In the information age, data are being doubled in every 2-3 years. Manager has to extract meaningful information and hidden patterns from this large volume of data, failing which he may not be able to take right decisions at the right time.

In view of above mentioned constraints, it has become necessary to have a well planned and designed information system in an organization to provide timely, reliable and useful information to enable the management to take speedy and rational decisions. It is a complex entity which encompasses many areas such as information science, information technologies, foundation concepts, business application areas and non business areas also. Information systems provide following support to an organization

- Business process and operations.
- Decision making process.
- Strategic planning for competitive advantages.

The organizational information system collects, transmits, processes and stores data, and retrieves and distributes information to the various users in the organization. If properly developed, managed and used, information systems can provide the most cost-effective resource for the organizations. Information system provides an interface between human activity systems and information technology.

2.5 Applications of Information System

There is a wide range of applications of information system that are implemented in today's world. Although IS applications are countless but to name a few such as:

- Sales analysis, production performance and cost trend reporting system

- Product Pricing, profitability forecasting and risk analysis system
- Sales and inventory processing and accounting System
- Product manufacturing system
- Milk supply chain management
- Geographic Information systems for agriculture
- Agricultural Resources Information System (http://agris.nic.in) (AgRIS)
- Agricultural Marketing Information System (AGMARKNET)
- M-commerce (Wireless and mobile applications)

2.6 Computer Based Information System

A Computer Based Information System (CBIS) is an information system that uses the power of computers to store, retrieve, process, analyze, and transmit data to the users of the system in an organization. A computer based information system provides fast and centralized access of data to end users so that users can extract and analyze the data as per their requirement. A computer based information systems have number of advantages over manual system. A few characteristics of CBIS are as follows:

- **Information coverage:** Central information systems provide organizations with the advantages of having large amounts of data, covering many different fields, accessibility via central source. Information coverage is a huge advantage for any organization, because having vast amounts of useful data from different department streamlines data access and increases productivity.
- **Centralized data:** Management and access of data is centralized. Most large-scale businesses and organizations use some sort of central database to store, manage and retrieve organizational data. For example MySQL, PostgreSQL or Microsoft SQL, Oracle, Ingres, etc., provide centralized database solutions.
- **Access efficiency:** CBIS improves data access and organizing efficiency which allows workers to focus more on the task rather than managing data manually.
- **Extensibility:** Computer-based information systems can be customized and extended as per the requirement of an organization.

2.7 Trends in Information Systems

In the last few decades there has been revolutionary change in the use of information and information systems in organizations. Many large business companies are using IS and IT to get competitive edge over other companies. For example, providing services through mobile phones to attract customers, Enterprise Resource Planning (ERP), web based information systems, geographical information systems, global positioning system etc. Information was regarded as a necessary evil associated with the development, production, and marketing of products or services. Information was merely considered as by product of transactions in the organizations. Therefore, IS during 1950s were primarily designed for transaction processing and record keeping with the aim to reduce cost of routine data processing activities in accounting areas. Such transaction processing activities were given name as Electronic Data Processing (EDP).

Another role was added, as the concept of Management Information System (MIS) was conceived in data processing domain in early 1960s. By this time organization started recognizing the information as an important tool which could support the general management tasks. This new role focused on

developing business applications to provide requisite information for decision making with predefined management reports to managerial end users.

After some time, it was realized by the management that information produced in pre-specified forms were not sufficient to take decision in many decision making situations. So the concept of DSS was initiated by 1970s. The new role of IS was to provide managerial end users with ad hoc and interactive support for their decision making process. This interactive support would be tailored to the unique decision making style of managers as they face specific type of real world problems.

The evolutionary phase of IT continued in 1980s, several new roles of information systems appeared. Following significant developments were occurred during this period:

- Development of microcomputers and application software packages gave new concept of *end user computing*. The end users were able to process their daily routine work on their personal computer devices without waiting for centralized support.
- Development of new computing paradigm of Artificial Intelligence (AI) opened new avenues in business applications. A number of information systems developed based on AI techniques such as virtual reality applications, robotics, natural language processing, *expert systems* (ES), *knowledge-based systems* (KBS) etc. The application of AI tools tried to replace human interventions in manual and intellectual activities with machines and software, thus, freeing up the knowledge worker for some other complex tasks. Expert systems can work as consultants to users by providing expert advice in limited subject areas in absence of experts. A number of expert systems are available for dairy industry and agriculture for example wheat expert system developed by IASRI, New Delhi to provide expert advice on wheat related problems, touch screen kiosk for cattle health knowledge dissemination among dairy farmers developed by Rajiv Gandhi College of Veterinary and Animal Sciences, Puducherry, animals disease diagnosis system etc.
- By late 1980s, information came into a new role as strategic resource, which was capable of providing competitive advantages or a strategic weapon to fight competition over other competitors. With this concept a new type of information system emerged called as Strategic Information System (SIS). In this concept, IT becomes an integral component of business processes, products, and services that help a company to gain competitive advantage in the global market.

The period of late 1990s observed revolutionary emergence of a new kind of information system called as ERP system. This system is specific to an organization that integrates all facets of the business including planning, manufacturing, sales, inventory control, order tracking, customer service, finance, human resources and marketing etc. virtually every business functions. ERP attempts to integrate all departments and functions across a company onto a single computer system that can serve particular needs of all different departments. The primary advantages of ERP are improved efficiency, information integration for better decision making, faster response time to customer queries, flexible decision making, data sharing, etc., which resulted into better corporate image, improved customer goodwill, customer satisfaction etc.

Rapid growth of ICT during 1990s brought many improvements in communication facilities and infrastructure. The most wonderful research achievement of 21st century emerged in form of Internet. Since then it has dramatically changed the capabilities of information systems in business. The Internet based and Web enabled information systems are dominating in the operations and management of today's business for example e-commerce, online shopping, e-governance, e-agriculture, etc.

Presently (2000 onward) a new era of mobile applications in business is emerging with the advances in wireless and mobile computing technologies. Numbers of companies have started using this technology

for developing new applications such as m-commerce on the pattern of e-commerce, providing services through mobile phones as well as managing routine business activities through mobile phones. The development of **Android** operating system, launched by Google has added new dimensions in development of mobile applications. The future lies in mobile applications. Table 2.1 summarizes the changes taken place in information systems since 1950s onward.

During last few decades it has been observed that the role of information systems in the organization is shifting to support business processes rather than individual functions. The focus is outwards to customers, rather than inwards to procedures. Business functions are changing rapidly. This poses a challenge to existing information systems, which are often inappropriately structured to meet these changing needs. It also poses a challenge to people who design, work and use these systems, since they may hold outdated assumptions.

Table 2.1 Historical trend in information systems

Time period	Role of Information	Type of IS	Aim of Information System
2000s - onwards	Reaching the unreached people, Competitive Weapon	M-commerce	Wireless and Mobile phone based services
1990s-2000s	Tracking, Control, Drill down, Strategic resource, Competitive Weapon	ERP, E-business	An integrated system including all facets of the a company, the Internet based e-business and e-commerce to reach remote customers/ end users
1980s-1990s	Inference, Strategic resource, Competitive Weapon	EIS, ES, SIS	Critical information for top management, knowledge based expert advices to end users, Strategic services and products for competitive advantage
1970s-1980s	Decision, Specific management control	DSS	Interactive ad hoc support to managers for decision making
1960s-1970s	General purpose support	MIS	Information in pre-specified format to support decision making
1950s-1960s	A by-product of data transactions	EDP	Transaction processing, record keeping, Accounting and Inventory

Level of participation of end users in IS
Level of IS in business and management

****** ☺ ******

LESSON 3

TYPES OF INFORMATION SYSTEMS

3.1 Introduction

This lesson introduces different types of information systems being used in business organization for efficient operations and management. These topics will be useful for students to familiarize themselves with different kinds of information systems and their objectives and functioning.

The discipline of information system is still in evolutionary phase. Previous chapter had discussed the growth and expansion of information systems since usage of computers in business data processing. Over the years, needs of management and role of end users increased in business operations to compete at global level and accordingly more sophisticated information systems emerged to meet users' requirements. Conceptually, applications of information systems that have been developed to support business activities can be classified in many different ways. However, these information systems can be broadly categorized into two categories as follows on the basis of their roles in operations and management of a business.

- Operations Support Systems
- Management Support Systems

In addition to above categories, several other information systems provide support in both operations as well as management of the business. Some information system may be categorized under general support systems, for example, business expert systems, functional business system and strategic information systems. Fig. 3.1 depicts the categories of the various information systems based on their role in operations and managerial decision making.

3.1.1 Operations Support Systems

Data is generated as by-product of transactions or when an event takes place in an organization. For example, each transaction generate some data in events such as receipt of milk at dock yard, distribution of milk to different sections for product manufacturing, handling losses during manufacturing process, quality control of milk products etc. This data is recorded, processed and used to carry out routine business operations. Information systems which use such kind of data are called operations support systems. These systems produce a variety of information products for internal and external use. However, such systems do not emphasize on producing specific information products that can be used by managers. In order to make use of such products further processing of the output is usually required. Such processing is done by management information systems. Role of operations support systems is to efficiently process business transactions of routine nature (such systems are called as Transaction processing system (TPS)), control industrial processes, support organizational communication and update databases.

3.1.2 Management Support Systems

These systems provide information and decision support for effective decision making by managers. There are various types of information systems such as MIS, DSS and EIS that support a variety of decision making process at different managerial level in an organization.

3.1.3 General Support Systems

General support system includes various information systems such as strategic information system (SIS) which provides support to business organization to gain strategic advantage over its competitors, functional business systems (FBS) which focus on operational and managerial applications in support of basic business functions such as accounting, inventory, marketing, production etc., business expert systems (BES) that provides expert advice on specific complex problems and acts as an expert in specific problem area.

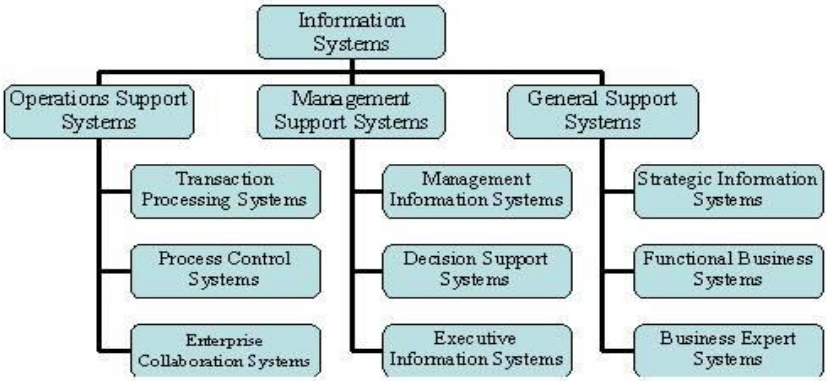


Fig. 3.1 Classification of information systems based on their role in business

3.2 Management Information Systems (MIS)

There is no consensus on definition of the term MIS. A number of terminologies are being used for MIS such as "information processing system", "information and decision system", "organizational information system", or simply "information system" to refer to the computer based information system. However MIS in general can defined as "*An integrated user-machine system for providing information to support operations, management, and decision making functions in an organization. The system utilizes computer hardware and software; manual procedures; models for analysis, planning, control and decision making; and a database*".

An integrated system does not mean that MIS is a single, monolithic structure rather it means that various components of information system fit into an overall design. MIS is a broad concept rather than a single system. MIS enhances job performance throughout an organization. At highest management levels, it provides information to help management to make strategic decisions. At other levels, MIS provides the means through which organizational activities are monitored and information is distributed to management, employees, and customers. An effective MIS ensures to meet appropriate presentation formats and time frames required by operations and senior management. A schematic diagram of MIS is shown in fig. 3.2 given below.

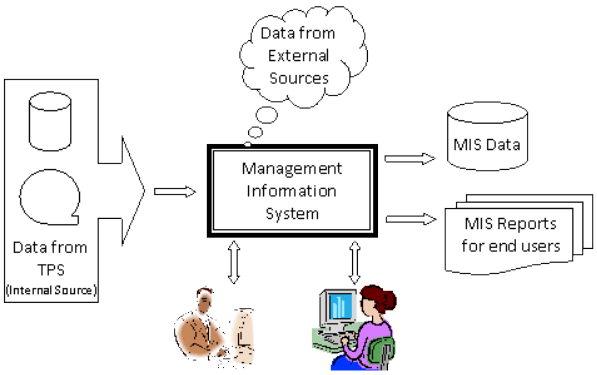


Fig.3.2 Conceptual frame work of MIS

MIS are designed to achieve the following goals:

- Enhance communication among employees.
- Deliver complex material throughout the institution.
- Provide an objective system for recording and aggregating data.
- Reduce expenses related to labor-intensive manual activities.
- Support the organization's information needs.

MIS should have a clearly defined framework of guidelines, policies, practices, standards, and procedures for the organization. These should be followed throughout the organization in development, maintenance, and use of MIS. It should be supportive of the institution's longer term strategic goals and objectives. On other extreme, it is also used for routine transaction processing to ensure basic control is maintained over day to day activities.

3.3 MIS versus Data Processing

Some significant distinguishable points between MIS and data processing as highlighted in table 3.1 as given below:

Table 3.1 MIS versus Data processing

Data Processing	MIS
This term was used prior to 1970s	This concept started in late 1960s and became popular in 1970s
Information is considered as by product of data processing activities	Information is considered as valuable resource for general purpose support to business operations
Primarily designed for transaction processing and record keeping with the aim to reduce cost of routine data processing activities in accounting areas	Focus on developing business applications that provided managerial end users with predefined management reports that would give managers the requisite information for decision making
Represents automation of fundamental operations, routine processing to support operations	MIS is more comprehensive, it encompasses processing of wider range of organizational functions and management processes

Process raw data generated through internal or external transactions	Generally process the information produced from data processing system
Output results in form of updated databases, bulky reports using simple and straight forward computations for line workers	Output reports in abstract form in pre-specified format for managers for decision making.
Independent approach for data processing for various applications	Integrated approach for processing data to meet the organizational information needs.

3.4 Decision Support System

To run an organization manager need to take decisions which require information from different sources. The more information you have, based on experiences or from internal/ external sources, the better your decisions will be. For making quality decisions managers need best available tools to help them. Before reaching to any conclusion managers should do following to the best of their abilities:

- Thoroughly check wide range of alternatives
- Weigh costs and risks of both positive and negative consequences
- Intensively search for new information for evaluating alternatives
- Take all new information into account, even when it doesn't support initial course of action
- Re-examine positive and negative consequences of all alternatives, including initially rejected ones
- Make detailed provisions for implementation, including contingency plans for known risks

A decision is a choice between alternatives based on estimates of the values of those alternatives. There are many approaches to decision-making since there exists wide range of domains in which decisions are made therefore concept of decision support system is very broad. In general, DSS is usually referred to computer based information system that performs a supporting role in decision making process. Computerized DSS became feasible during mid-1960s with the development of minicomputers, timesharing operating systems and distributed computing. DSS help managers to make better decisions by combining massive amounts of historical and current data from internal information systems and external sources with sophisticated analytical models and tools.

DSS may be defined as "*an interactive, flexible, and adaptable computer-based information system, especially developed for supporting solution of a non-structured management problem for improved decision making. It utilizes communication technologies, specialized databases, knowledge and/ or analytical models, computer based interactive interface that allows a decision maker to apply his/her own insights and judgment for making unstructured or semi structured business decisions*".

DSS provides interactive ad-hoc support to assist managers during decision making process. DSS tends to be used in planning, analyzing alternatives, and trial and error search for solutions. DSS incorporate variety of decision models and are thus capable of performing what-if analysis. DSS differ from traditional IS because each DSS is tailored to a specific managerial task or special problem, its use is limited to that task or problem. DSS tend to be designed to serve management control level and strategic planning level managers. These systems produce information by taking output of TPS and MIS as input data and other inputs may also be collected or generated from external environment. Since strategic information is less structured in comparison to tactical information, therefore, it is difficult to develop

DSS. An MIS uses internal data to supply useful information. A DSS uses internal data as well as external data to help analyze various decision making problems. Analyzing complex problems with interactive decisions is the primary reason for an organization to use a DSS. For example a production manager in a dairy plant may use DSS to perform what-if analysis to decide how much product should be manufactured based on expected demand in future, location and availability of raw material required to manufacture the product, risk analysis etc.

3.5 DSS versus MIS

Characteristics of MIS and DSS are compared as given below in table 3.2:

Table 3.2 Comparison of MIS and DSS

MIS	DSS
Structured decisions	Semi-structured, unstructured decisions
Reports based on routine flow of data	Focused on specific decisions / classes of decisions
General control of organization	End-user control of data, tools, and sessions
Structured information flows	Emphasizes change, flexibility, quick responses
Presentation in form of reports	Presentation in form of graphics, Greater emphasis on models, assumptions, and ad hoc queries
Traditional systems development	Develop through prototyping; iterative process

3.6 Types of Decision Support Systems

- **Data-driven DSS:** Data-Driven DSS extracts useful information from massive amount of data taken from company's TPS and MIS for taking decisions by executives. Data-Driven DSS emphasizes on manipulation of large databases of structured data and especially a time-series data of organization and sometimes external data. Simple file systems accessed by query and retrieval tools provide the most elementary level of functionality. Data warehouse systems that allow the manipulation of data by computerized tools tailored to a specific task setting provide additional functionality. Data-Driven DSS with Online Analytical Processing (OLAP) provide the highest level of functionality and decision support that is linked to analysis of large collections of historical data.

- **Model-driven DSS:** Model-Driven DSS includes systems that use accounting, financial, representational, optimization, statistical, or simulation models. Model-Driven DSS emphasizes on access and manipulation of data using models. Simple statistical and analytical tools provide the most elementary level of functionality. Model-Driven DSS use data and parameters provided by decision-makers to aid them in analyzing a situation. Such DSS are not data intensive therefore do not require large databases. These systems are separated from main Information Systems of an organization and primarily used for typical "what-if" analysis. That is, "What will be effect on profit, if we increase production of products and decrease the shipment time?" These systems rely heavily

on models to help executives to assess the impact of their decisions on the organization, its suppliers, and its customers.

- **Knowledge-driven DSS:** Knowledge-Driven DSS can suggest or recommend actions to managers. This concept is still evolving. These are computer systems with specialized problem-solving expertise stored as facts, rules, procedures, or in similar structures. The "expertise" consists of knowledge about a particular domain, understanding of problems within that domain, and "skill" at solving some of these problems. A related concept is Data Mining. It refers to a class of analytical applications that search for hidden patterns in a database. Data mining is a process of sifting through large amounts of data to produce data content relationships.
- **Document-driven DSS:** Document-Driven DSS is evolving to help managers to retrieve and manage unstructured documents and Web pages. A Document-Driven DSS integrates a variety of storage and processing technologies to provide document retrieval and analysis. WWW provides access to large document databases including hypertext documents, images, sounds and video. A document-driven DSS may access documents related to policies and procedures, product specifications, catalogs, and corporate historical documents, including minutes of meetings, corporate records, and important correspondence. A search engine is a powerful decision aiding tool associated with a Document-Driven DSS.
- **Communications-driven and group DSS:** Communications-driven DSS includes communication, collaboration and decision support technologies for making decisions. Group DSS is a hybrid decision support system that emphasizes on the usage of both communications and decision models. Group DSS is an interactive computer-based system intended to facilitate solution of problems by decision-makers working together as a group. Group DSS supports electronic communication, scheduling, document sharing, and other group productivity and decision support enhancing activities. A number of modern technologies that are used to facilitate communication in group DSS are two-way interactive video, white boards, bulletin boards, E-mail etc.

3.7 Components of DSS

Traditionally, a decision support system has the following major components:

- The user interface
- The database
- The models and analytical tools and
- The DSS architecture and network

Data-Driven, Document-Driven and Knowledge-Driven DSS need specialized database components. A Model-Driven DSS may use simple flat-file database with few records, but the model component is very important. Experience and some empirical evidence indicate that design and implementation issues vary for Data-Driven, Document-Driven, Model-Driven and Knowledge-Driven DSS. Mathematical and analytical models are major components of a Model-Driven DSS. Each model-driven DSS has specific purpose and hence different models are needed and used. Choosing an appropriate model is a key design issue. In model-driven DSS values of key variables or parameters are changed, often repeatedly, to reflect potential changes in supply, production, the economy, sales, the marketplace, costs, and/or other environmental and internal factors. Information from models is then analyzed and evaluated by decision-maker. Concept of DSS with various components is depicted in fig. 3.3.

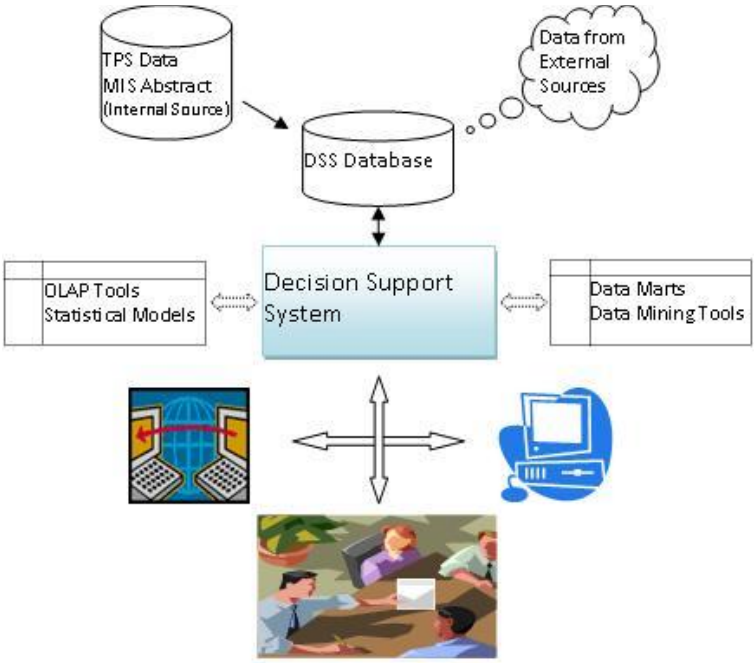


Fig. 3.3 Systematic overview of a DSS

3.8 SAP Enterprise resource planning

This term Enterprise Resource Planning (ERP) is used as a concept to integrate various activities of the organization to mange business efficiently. The basic purpose is to facilitate flow of information among various business functions within the organization. An integrated software application package known as ERP system is used to integrate all activities of an operation, product planning, development, manufacturing, finance, inventory, sales and marketing, customer relation, etc. ERP system provides all routine features of MIS and DSS to manage business efficiently as well as latest tools such as business data warehousing and business intelligence for mining novel information from historical data.

SAP is a software corporation which makes ERP systems. SAP stands for Systems Applications and Products. The original SAP idea was to provide customers with the ability to interact with a common corporate database for comprehensive range of applications which, includes managing their day to day operations, logistics, finances, month end, quarter end and yearly activities, reporting, HR, etc. SAP was initially designed to be run on the mainframe and was called R/2 (Release 2). Then quickly it shifted on to the client server model with a later release called R/3 and this was the most popular version of SAP. After R/3, later versions of their core software were launched called Enterprise Central Component (ECC). New modules to meet the growing need of organizations were added such as business intelligence (BI) to mine data from their daily operational data and extract meaningful trends that could

enable further business opportunities, customer relationship management (CRM), Process integration (PI) etc.

With the development of WWW and Internet tools, SAP also moved from standard client server architecture to a completely web-based architecture where every transaction can be run from just a browser. Web based version is called as mySAP which is a very successful product in the market. A few best known software products of the company are:

- Enterprise resource planning application (SAP ERP),
- Enterprise data warehouse solution - SAP Business Warehouse, SAP Business Objects
- Sybase mobile products and in-memory computing appliance (SAP HANA).

- Customer Relationship Management (CRM)

- Product Lifecycle Management (PLM)

  - Supply Chain Management (SCM)

  - Supplier Relationship Management (SRM)

3.9 Applications in Dairy Industry and Animal Management

A large number of applications of DSS in animal management have been reported in literature at national and international level. A few important applications are mentioned here for reference as given below:

3.9.1 HERDMAN: A new preventive herd health software to enhance dairy animal productivity
(Infovet; 302 Shariq Apt. Takoli, Kalwe, Thane, India 400 605, herdmancom@Hotmail.Com)

Vetindia Infotech Solutions, Mumbai, in collaboration with the Department of Medicine, Bombay Veterinary College, affiliated to Maharashtra Animal and Fishery Sciences University, has developed a cow / buffalo herd management software that is tailor-made to satisfy the needs of dairy industry as well as dairy farmers. It is a window-based software that provides computing and analysis strength using Visual Basic and Access/MS-SQL. Its interface is icon/menu driven which can be operated by a moderate literate farmer without any difficulty. Major characteristics of this software are:

- It maintains life-time records of animals in the farm
- Records of cows and buffaloes can be maintained simultaneously and data can be analyzed separately
- Records of all categories of animals, such as calves, heifers, adults, breeding bulls and working bullocks can be maintained
- The records of culled, sold or died animals maintained in archive files that can be retrieved easily to analyze data

It provides data integration and merging facility with other window-based software used for other purposes, such as, progeny testing, epidemiological mapping, GIS- or Info map-based data modules.

HERDMAN can easily be interfaced with milking machines, milk parlor or automatic feeding machines.

3.9.2 Empowering dairy farmers through a dairy information & services kiosk

(http://unpan1.un.org/intradoc/groups/public/documents/UNPAN/UNPAN023605.pdf)
In recent years, milk co-operative movement initiated by National Dairy Development Board (NDDB) has led to a substantial increase in milk production in India. Two main reasons for this increase are efficient procedures for collection of milk and higher profits for producers. IT has played an important role in improving milk production of the country by educating the farmers and brining efficiency and transparency in the procedures. Automation of milk procurement and billing process at 2,500 rural milk collection societies has immensely benefitted the farmers. The DISK has enabled co-operatives and farmers to manage database of milch cattle and access to information about valued services about animal health, market trends, scientific animal management practices etc. These cases demonstrate that farmers are willing to invest in technology provided the farmers are empowered through innovative systems.

****** ☺ ******

Lesson 4

KNOWLEDGE MANAGEMENT SYSTEMS

4.1 Introduction

This lesson discusses the concept of knowledge management and various types of knowledge based systems. These topics will be useful for students to have idea of current trend in the information systems for improving decision making ability of managers.

Knowledge based information systems are referred as Knowledge Management Systems (KMS). These systems manage knowledge in the organization. They support creation, organization and dissemination of knowledge to managers and other workers of the organization. Many organizations are building knowledge management systems to manage organizational learning, solution to complex problems and decisions, policy issues, and business know-how. Such systems provide quick feedback to knowledge workers and encourage behavioral changes in employees to improve business performance. For example, taking a piece of advice on complex issue such as best practices on animal health care, ethics of production management, clean milk production, transfer of technology to farmers, etc. Main objective for creation of KMS is sharing of ideas. For example, sharing the findings of "composition of cheese to increase its shelf life" by a researcher can lead to improved products that are using cheese as ingredient or it may also lead to some other innovative ideas. KMS may be stand alone or Web based systems but web based systems are more effective to disseminate knowledge to scattered and large number of stakeholders. Some of the advantages of KMS systems are:

- Sharing of valuable organizational information throughout organizational hierarchy.
- Can avoid re-inventing of wheel, reducing redundant work.
- May reduce training time for new employees
- Retention of intellectual property after an employee leaves organization

4.2 Artificial Intelligent Systems

Artificial intelligence is an outcome of synergy of many fields like computer science, mathematics, biology, philosophy, linguistics, neural science, engineering, etc. The primary goal of AI is to develop computers that can simulate the ability to think as well as see, hear, walk, talk, and feel like human beings. Computer functions are associated with human intelligence such as reasoning, learning, acquire and apply knowledge, creativity and imagination, and problem solving. AI systems learn from a given set of inputs and outputs. These store experiences in their memory, generalize them and become ready to deal with new circumstances (i.e., new inputs). AI applications may be classified under three major areas such as cognitive science, robotics, and natural interface. However, these classifications overlap each other sometimes. Figure 6.1 given below shows major application areas of AI.

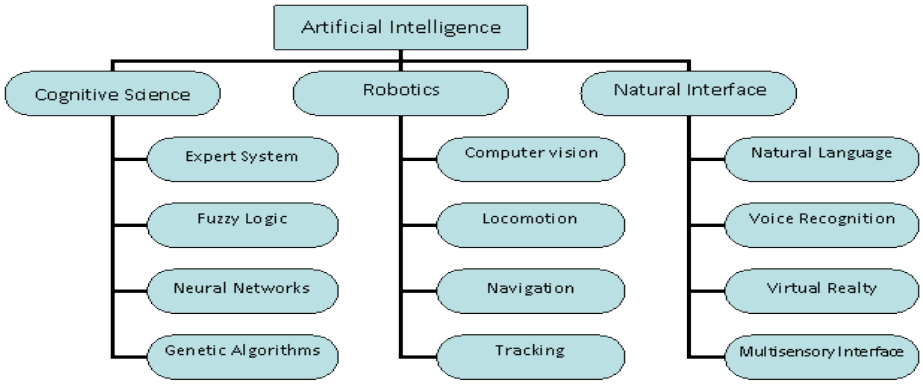


Fig. 4.1 Important application areas of artificial intelligence

Information systems that make use of artificial intelligent techniques in decision making process are known as artificial intelligent systems. In such systems, knowledge is major point of focus rather than data or information. AI technologies have been successfully explored to its full potential in the areas of problem solving i.e. concepts and methods for building programs that reason about problems rather than calculate a solution. Development of expert systems for decision making is the most widely used AI application. However, significant results have also been achieved in other areas of AI like intelligent robotics, voice recognition, computer vision, natural language processing etc. AI systems have several advantages over human intelligence as given below:

- Knowledge in AI systems is permanent
- Can be easily duplicated and disseminated
- Consistent and thorough
- Knowledge can be documented

4.2.1 Characteristics of AI system

- Symbolic processing: AI is a branch of Computer Science that deals with symbolic, non algorithmic method of problem solving.
- Heuristics: It consists of intuitive knowledge or rules of thumb, learned from experiences.
- Drawing inferences: AI systems have the ability of reasoning or drawing inferences from facts and rules using heuristics or other search approaches.
- Machine learning: AI systems have mechanical learning capabilities called machine learning in parallel with human beings.

4.3 Expert Systems

Knowledge based expert systems or simply Expert Systems (ES) are computer programs that use human knowledge to solve problems which normally would require human intelligence. Expert systems represent expertise knowledge in form of data or rules within the computer program. These rules and data can be called upon when needed to solve problems. Expert system emulates the interaction with user just like a human expert to solve a problem. Program asks a series of questions and end user provides input by selecting one or more answers from list of answers or by entering data. Program will keep on asking questions until it reached to a conclusion. The conclusion drawn by system may be a selection of single solution or a list of possible solutions arranged in order of likelihood. Expert system can explain why data is needed and how conclusions were reached. Some expert systems are complex to build while some are easy. In a very simple case, consider a tree diagram on paper describing how to solve a problem. By making a selection at each branch point, tree diagram can help to make final decision. This type of tree structured logic can easily be converted into computerized program that is easier to use.

More elaborate systems may include confidence factors allowing several possible solutions to be selected with different degree of confidence. Output of an expert system may be information, an instruction or a risk judgment. A different problem within the domain can be solved using the same program without reprogramming. Ability of these systems to explain reasoning process through back traces and to handle levels of confidence and uncertainty provides an additional feature that simple expert systems do not have.

Most of the expert systems are developed using a specialized software tool called shells. A shell contains the user Interface, script language for developing expert system, a format for declarative knowledge base, and an inference engine, different data structures, connectivity with other programs and databases etc. Some of the commonly used shells for developing expert systems are CLIPS and JESS.

- **CLIPS**: CLIPS is a public domain software tool for building expert systems. The name is an acronym for "C Language Integrated Production System." It was developed by the Software Technology Branch (STB) at the NASA/ Lyndon B. Johnson Space Center. It was released in 1986 for the first time and has undergone continual refinement and improvement ever since. CLIPS is probably the most widely used expert system tools because it is fast, efficient and free. CLIPS is written in C, extensions can be written in C, and CLIPS can be called from C. Its user interface more

closely resembles that of the programming language LISP. It supports rule-based, object-oriented and procedural programming. CLIPS inference engine provides only forward chaining.

- **JESS:** JESS is an expert system shell and it is even based on CLIPS. The name is acronym for Java Expert System Shell. JESS was originally a clone of the essential core of CLIPS, but has begun to "acquire a Java-influenced flavor of its own". JESS is completely programmed in Java by Ernest J. Friedman-Hill at the Sandia National Laboratories for the U.S. Department of Energy. It was first released in 1995. It provides both forward and backward chaining.

These shells come equipped with an inference mechanism with several levels of sophistication "forward chaining", "backward chaining" and "mixed chaining" for drawing conclusions and require knowledge to be entered according to a specified format. Forward chaining (also known as data driven reasoning) is the questioning of an expert who has no idea of the solution and investigates progressively (e.g. fault diagnosis). Inference technique uses IF - THEN rules to deduce a problem solution from initial data. In backward chaining (also known as goal driven reasoning), the engine has an idea of the target (e.g. is it okay or not?). It starts from the goal in hopes of finding the solution as soon as possible. Inference technique uses IF - THEN rules to repetitively break a goal into smaller sub-goals which are easier to prove. In mixed chaining the engine has an idea of the goal but it is not enough: it deduces in forward chaining from previous user responses all that is possible before asking the next question. So quite often he deduces the answer to the next question before asking it.

4.4 Components of Expert System

Components of an expert system include knowledge base and software modules that perform inferences (or reasoning) on knowledge in the knowledge base and communicate with users for question and answer. Figure 6.2 illustrates interrelated components of an expert system.

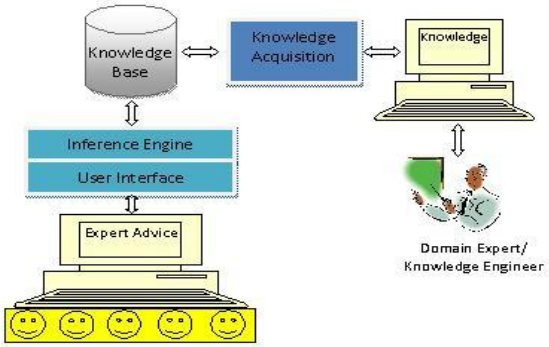


Fig. 4.2 Components of expert system

The role of individual components is described as follows:

4.4.1 Domain Expert

The individual person who is expert in solving the given problem? Domain knowledge is with him (expert). Here the system is intended to solve the problem as expert.

4.4.2 Knowledge Engineer

The individual who encodes the expert knowledge in a declarative form that can be used by expert system. Knowledge engineer acquires domain knowledge from expert and builds a knowledge base for particular problem using expert shell.

4.4.3 Knowledge base

Knowledge base of expert systems contains both factual and heuristic knowledge. Knowledge available in text book, journals and standard practices about a specific domain is known as factual knowledge. It is declarative knowledge either in true or false statements. Heuristics knowledge is informal, experiential, and judgmental knowledge of an application area that constitutes rules of good judgment in the field. Heuristics cover the knowledge of how to solve problems efficiently and effectively, how to plan steps in solving a complex problem, how to improve performance etc. In comparison to factual knowledge, heuristic knowledge is less rigorous, 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 in expert systems may be represented in many ways. Widely used method is form of production rules or simply rules. A rule consists of IF and THEN part (also called a condition and an action). Expert systems whose knowledge is represented in rule form are called *rule-based* systems. Another commonly used method of knowledge representation is *case-based*. In this method knowledge base is created in the form of cases, that is, past performance, occurrences, and experiences. Other methods of representation are *frame-based* and *object-based*.

4.4.4 The inference engine

Inference engine processes the knowledge such as rules and facts related to a specific problem. It recommends the course of action for a user. It is a computer program designed to produce reasoning based on logic. Several kinds of logics are used in expert systems such as propositional logic, predicates of order 1 or more, epistemic logic, modal logic, temporal logic, fuzzy logic, etc. With logical reasoning, the engine is able to draw conclusions from the knowledge contained in the rule base. 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. Set of methods for using uncertain knowledge in combination with uncertain data in reasoning process is called *reasoning with uncertainty*. For instance, an important subclass of methods for reasoning with uncertainty is called "fuzzy logic," and the systems that use them are known as "fuzzy systems."

The expert system process data in two ways i.e. batch or conversational. In batch mode, the expert system has all the necessary data to process from the beginning. For the user, the program works as a classical program. User provides data and receives results immediately. Reasoning and processing of data is invisible to the user. In conversational method, user interactively answers the questions during execution of the system. This method is useful when the problem is big, complex and all necessary data is not available at the starting. The software tries to solve the problem by asking the missing data from the user, gradually approaching the goal as quickly as possible.

4.4.5 User Interface

User interface is a program that controls the dialog between the user and the system. Users interact with system by putting their questions and answers and get advice from the expert system. User interface provides an easy and user friendly environment to operate the system.

4.4.6 Users

Users are the individual person who is consulting with the system to get advice on the particular problem.

4.5 Applications of expert systems

- Decision Management: Worker performance evaluation, Demographic forecast
- Diagnostic / problem shooting: Fault diagnosis, Help desk operations, debugging
- Design/ configuration: Manufacturability studies, optimum assembly plan
- Selection/ Classification: Material selection, Information classification
- Process Monitoring/ Control: Machine control, production monitoring

4.6 Benefits of expert systems

- Expert systems are faster and more consistent.
- The knowledge base may have knowledge of several experts.
- Does not tire and distract from overwork or stress.
- Preserve knowledge of experts.

4.7 Limitations of expert systems

- Can solve one type of problem in a limited domain of knowledge
- Inability to learn from experiences

- Maintenance problem
- Development cost may be too high. It includes cost of hardware, software, knowledge engineer, expert time etc.

4.8 Current Trends – Knowledge based DSS

Knowledge based DSS or intelligent DSS is the current buzz word in DSS research field. It is an extended version of DSS which derives knowledge through artificial intelligent tools like artificial neural networks, fuzzy logic, genetic algorithms, experts systems, knowledge representation methods etc. Such systems assist in decision making process where information is incomplete, imprecise, uncertain, non linear, unstructured and decisions are to be made using human judgment. Recent developments in field of information technology have further enhanced the strength of intelligent DSS for example WWW, Internet, intranet, multimedia, virtual reality, static/ mobile ad hoc networks, wired/ wireless sensor networks, RFID, image analysis, etc. have given a real momentum in quality data collection and processing.

With advances in micro-electronic equipment for example sensors, potential parameters about animal can be recorded continuously online without human intervention. The possibilities comprise electronic identification, automatic weighing, temperature and activity measurements, automatic regulation of feed and water intake, geographic positioning of individual animals and monitoring animals through video recordings using image analysis. Recent examples have shown the usage of such data for developing knowledge based DSS with the help of artificial intelligence techniques for heat detection, daily gain in body weight, water consumption, and monitoring of pregnancy rates, early diagnosis of mastitis and lameness. Purpose of these systems/ studies is to allow managers/ users to detect changes in the monitored variable for welfare of animals and increasing productivity of dairy farms.

****** ☺ ******

Lesson 5

SYSTEMS DEVELOPMENT APPROACHES

5.1 Introduction

This lesson will introduce the concept of a system, elements, characteristics, classification of systems, various approaches for system development and role of system analyst in developing a system have been discussed in this lesson. Concept of dairy business as a system has also been elaborated in this lesson. These topics will develop an understanding about systems and existence of different type of systems in real world. It will help the students in studying and analyzing the existing business systems in dairy industry.

5.2 Definition of System

The word "*system*" is possibly the most overused word in our vocabulary and is used quite often in everyday life. Generally, we speak about political system, education system, human body system, social system, transportation system, banking system, computer system etc. The common feature, which all systems share is that they are collection of elements integrated to achieve the required goals. The term system is derived from Greek word *systema*, which means an organized relationship among functioning units or components. A system exists because it is designed to achieve one or more objectives. In present context, the word *System* can be defined as a group of interdependent components in an organized way, which works together to achieve one or more objectives as per specified plan. Interdependent components may refer to physical parts or managerial steps known as subsystem of a system. For example a dairy plant can be considered as a system in which various departments and sections are joined together for a common goal.

A system may be regarded as a set of entities or elements interacting among themselves in a certain pre-specified manner to process inputs and produce outputs (objectives) in a defined time period. Output of a system may be quantitative or qualitative. The concept of a system is shown in fig. 5.1 as given below:

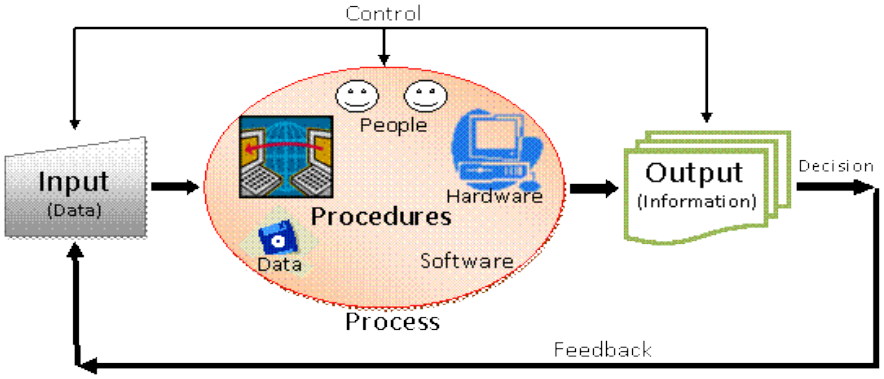


Fig. 5.1 System with input and output

5.3 Characteristics of a System

As defined, a system is an organized relationship of various subsystems or components to achieve some objectives. For example, a computer system has a number of subsystems like Keyboard, Monitor, CPU, Mouse, etc. A Computer system is just an organized relationship of these subsystems in a planned way. These subsystems are also dependent of each other to achieve set targets. Thus the definition of a system suggests the following characteristics that are present in all systems.

- **Organization:** This refers to a systematic order of components working together as per order to achieve centralized objective. It is the arrangement of subsystems in a specific structure and order that help to achieve objectives.
- **Interaction:** This refers to the manner in which each subsystem functions, interacts or communicates with other components of the system. Special interfaces are designed to share data/ information among subcomponents.
- **Interdependence:** It means that subsystems of a system are dependent on one another to achieve objectives of a system. This can be compared with a situation like a complex task which is divided into number of simple tasks (i.e. subsystems). Each simple task performs its work and produces output. Output of one task may become input of some other task. Like this a given complex problem is finally solved. Similarly in a system output of one subsystem may be an input of some other subsystem. These subsystems are coordinated and linked together according to a plan.
- **Integration:** This refers to holistic view of the system that how different subsystems are tied together to achieve central objective. A Bottom up approach is followed to integrate subsystems.
- **Planned approach:** In order to achieve centralized objectives, a system should work according to a planned approach. The planned approach is a set of pre-laid policies, procedures, rules, budget, strategies and schedules to achieve goals.
- **Central objective:** Each system is developed to achieve specified centralized objectives. All subsystems are developed and integrated to achieve centralized objectives keeping the unique identity of each subsystem. Each subsystem works independently and interacts with other subsystems. Each subsystem contributes something directly or indirectly in achieving the system objectives.

5.4 Elements of a System

In general an information system accepts input data and processes it to produce information as output for decision making. System is evaluated on the produced output and feedbacks are given to improve the system. A system has following key elements:

- **Output:** One of the major objectives of a system is to produce an output that has value to its user. Nature of output may be goods, services, information etc. Output is outcome of processing and end result of the system.
- **Inputs:** Inputs are the elements like material, manpower, data, information etc. that is entered into the system for processing.
- **Processor(s):** Processor is the element of a system that involves actual transformation of input into output. It is an operational component of a system.
- **Control:** Control element guides a system. It is a decision making subsystem that controls various activities of a system such as governing input, processing, and output. Control is achieved by correcting the deviation between output and standard.

- **Feedback:** Control in a dynamic system is achieved by feedback. Feedback compares output of a system against performance standards and accordingly information is communicated to system for necessary action. This may yield to change in input or processing and consequently the output.
- **Environment:** Environment is a super system within which an organization operates. It is the source of external elements that affect a system. It often determines how a system should work. For example vendors, competitors, Govt. policies, tax department, etc. may provide constraints and consequently influence actual performance of the system.
- **Boundaries and interface:** A system is defined by its boundaries – the limits that identify its components, processes, and interrelationships when it interfaces with another system. A system boundary indicates where one system ends and another starts. Generally boundaries between systems are not always clear cut. A complex system may have many interfaces with other systems. For example the market & sale section is concerned with sale of milk and milk products, collection of sale amount, determining demand of products in near future. This section is not concerned about how the products are being manufactured, manufacturing losses for production, etc.

5.5 Types of System

Systems have been classified in different ways. Common classifications are:

I. Physical and Abstract System
II. Deterministic and Probabilistic System
III. Open and Closed System
IV. Information System

5.5.1 Physical and abstract system

Systems can be categorized as physical and abstract systems. Physical system corresponds to concrete operational systems made up of tangible (i.e. physically available objects) entities such as people, materials, machines, and other physical things that may be static or dynamic in operation of a system. The physical systems display some activity or behavior. Elements of such system interact to achieve a common objective. For example physical parts of a dairy plant are equipment, employees, office establishment, furniture, building that facilitate operation of plant. These parts are more or less same and can be seen and counted, hence these are static. In contrast, daily receipt of milk, manufacturing of products per day, sale of products etc. keep on changing as per the need, hence these are dynamic.

Abstract systems are conceptual or non physical entities. These are an orderly arrangement of interdependent ideas or constructs which may or may not have any counterpart in real world. They may be as straight forward as formulas of relationships among sets of variables or models of an abstract conceptualization of physical situations. Model is representation of a real or planned system. For example set of instructions or procedure for manufacturing milk products (e.g. paneer, cheese, ice-cream etc.) is an example of abstract system. Steps involved in a procedure are an orderly arrangement of ideas to finally produce milk product.

5.5.2 Deterministic and probabilistic system

Deterministic system is one in which occurrence of all events is perfectly predictable. In such systems, under given description of system state at a particular time of its operation, next state can be perfectly predicted. For example numerically controlled machine are deterministic in nature.

Probabilistic system is one in which occurrence of events cannot be perfectly predicted. Though behavior of such a system can be described in terms of probability, a certain degree of error is always attached to behavior prediction of the system. An example of such system is status of items in inventory,

occurrences of a telephone call in a particular time period, average demand of milk product at a given point of time etc.

5.5.3 Open and closed system

Open systems are those systems which interact with its environment and have many interfaces with outer world. Thus these systems exchange information with the environment including random and undefined inputs. Open systems permit interaction across its boundaries; and receive inputs from and provide outputs. These systems are usually adaptive i.e. their interaction with environment is such that they change themselves according to situation which favour them to remain in existence for longer time. For example, information systems must adapt to the changing demands of users for its survival. A human being is also an example of open system. All organizations are generally open systems as they cannot work in isolation.

Closed systems are isolated from environmental influences i.e. they do not interact with its environment and do not adopt changes. Such systems are rare, but relatively closed systems are common. Systems that are relatively isolated from environment but not completely closed are termed as closed systems. A relatively closed system is one which controls its input and so is protected from environmental disturbances. For example a computer program is relatively closed system which processes predefined inputs in a predefined way. Closed or relatively closed systems are difficult to survive for a longer time since they do not interact with changing environment and will eventually die.

5.5.4 Information system

Information about a subject helps in taking decision with higher confidence since it reduces uncertainty about a state or event. An information system is the basis for interaction between users at different level. An information system can be defined as a set of devices, procedures, and operating systems designed around user based criteria to produce information and communicate it to user for planning, control, and performance.

In order to work as an effective unit, an organization has to make use of information. An information system is often regarded as another subsystem. Information system provides information for decision and control and acts as linking mechanism between functional subsystems. Major information systems are formal, informal, and computer based.

5.6 Illustration of Business of a Dairy Plant as a System

Organizations are complex systems that consist of interrelated and intertwine subsystems. Business of a dairy plant can be defined in system terms. The business set up of a dairy plant consists of various components like Milk procurement and billing, Processing of milk, Manufacturing of dairy products, Inventory of milk and milk products, Inventory of engineering parts, Marketing and Sale, Quality Control of dairy products, Manufacturing and handling losses, Accounting, HRD, R&D, etc. All these components work together to enhance profit and produce good quality of milk products based on set procedures and rules of their interactions. It can be observed that each component (i.e. subsystem) stated above is a complete system within itself and these subsystems interact with each other to pass on necessary information for smooth functioning of each subsystem as well as the whole system. For example, marketing section determines the demand of a particular dairy product for next few days and this information is passed on to Production section to fulfill demand in the market. Changes in one part of system may yield anticipated and unanticipated consequences in other parts of the system. Thus the business of a dairy plant is a system which receives resources (capital, people, plant, milk, etc), process received milk and produces the milk products. Dairy plant interact with outside environment consumers

demand, government agencies like pollution control board, income tax department, milk producers, etc., to make equilibrium in input and output.

5.7 System Development Approaches

Developing a successful information system for business application is a challenging task. System development process includes the activities like understanding, planning, designing, implementation and maintenance of systems. System development approaches have undergone reforms as per the changes in software development technology. Software technology has changed from modular to structure and now to object oriented technology. Basic goal in system development process is to produce high quality information system for solving business applications. It provides a framework that is used to structure, plan and control the process of developing a system. Various methods have been evolved and being practiced over the years. A few major methods are:

I. System Development Life Cycle method (or traditional method))
II. Structured Analysis Development method
III. Waterfall method
IV. Prototype method
V. Spiral method
VI. Object oriented Development method

Each of these methods has its own strengths and weaknesses. One method may not be suitable for all type of systems. It depends on technical nature of a system, organizational requirements, knowledge of developers etc. Basically all these methods follow either *linear* or *iterative* approach or combination of these two approaches for development of systems. In linear approach, above mentioned activities are followed in a sequence with or without overlapping of some activities to develop a system. In iterative approach, the activities may or not be repeated for improvement after getting users feedback.

5.7.1 System Development Life Cycle

The most prevalent and important method for system development is System Development Life Cycle (SDLC) method. Though it is traditional method but still it holds good in any methodology or approach as it is generic term for system development. System development process has a life cycle just like a living system. Systems are conceived, designed, developed and maintained. Over a period, numbers of changes in the existing computer system are made to accommodate new requirements of users and technological developments that give a totally different look to the system which was conceived and developed in the beginning. This is a continuous process which shows that information systems are just like living systems which came into existence and after some time they die (i.e., replaced with new one). System analysis and design are important factors of system development life cycle. For computer professionals, SDLC is a systematic method to develop a computer based system to solve a business or scientific problems in order to fulfill the needs of an organization or a customer. SDLC method helps to monitor the progress while the application development is in progress. This method monitors and controls system development beginning from fact gathering, designing, implementation and maintenance in cyclic way as a continuous process. There has been much criticism for this approach. However, other methods only supplement SDLC method rather than replace it. Different phases in SDLC are as follows:

- Study and analysis phase
- Systems design
- System development

- System implementation
- System maintenance

System development and maintenance is a continuous process as long as the system exists as the organizational requirements keep on changing throughout the year. Business processes changes rapidly in globally competitive age, and most of the systems need to be replaced or significantly updated after few years of operation. The updated system looks like a new system different from original one. This completes life cycle of a system like living entities (from origin to death). Fig. 5.2 shows all steps involved in SDLC method.

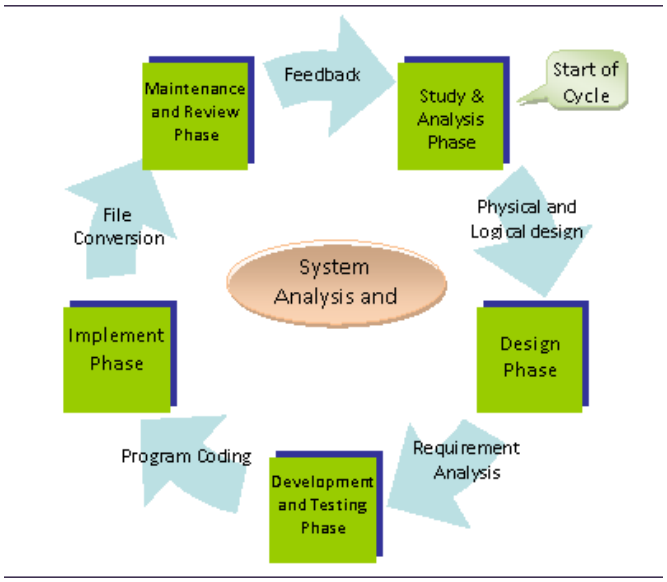


Fig. 5.2 System development life cycle method

5.7.2 Structured Analysis

Structured analysis is a traditional system development method which is time tested and easy to understand. This technique / methodology describe the system through a set of process models therefore; it is also known as *process-centric* method. System is viewed from perspective of data flowing through different processes. Function of a system is described by processes that transform data into useful information which leads to achieve the systems objectives. The system is decomposed into sub processes, sub processes into further sub processes and so on. This allows concentrating on the relevant part only by keeping other details aside. Process modeling identifies data flowing into a process, business rules that transform data to resulting output data. In addition to process modeling, structured analysis includes data organizations, database design and user interface design. Result of structured analysis is a set of related graphical diagrams, process descriptions, and data definitions. A variety of structured tools such as data flow diagram, decision table, decision tree, data dictionary, structured English are used to document the data items, data flow, logical complexities etc. All steps mentioned in SDLC are followed to develop a system in this method.

5.7.3 Waterfall method

This method follows SDLC approach for system development. System is divided into sequential phases and organized in linear order i.e. output of one phase becomes input of next phase. There may be some

overlap and splash back between phases. Emphasis is given on planning, time schedules, target dates, budgets and implementation of an entire system at one time i.e. complete the system development and implementation at once. Tight schedule is maintained while developing a system in form of time table, documentation and approval of managers/ users at the end phases and before beginning the next phase. Fig. 5.3 shows all steps involved in Waterfall method.

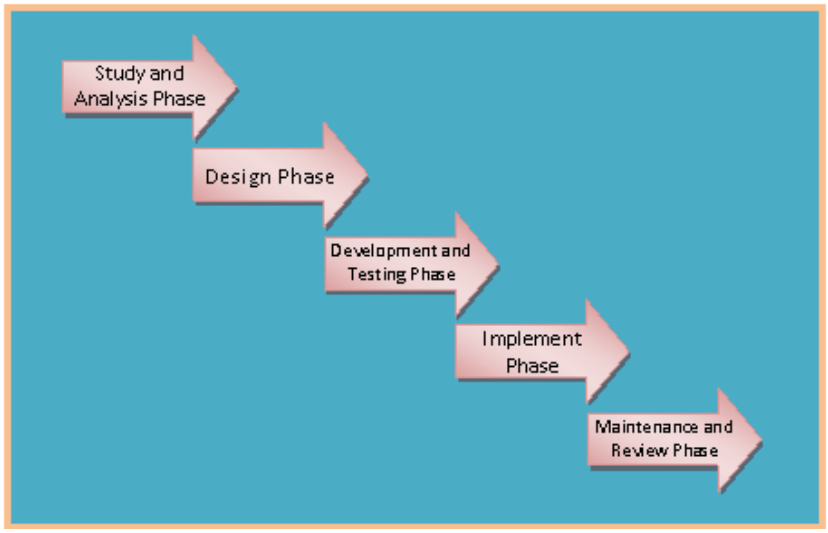


Fig. 5.3 Waterfall method for system development (linear approach)

5.7.4 Prototyping Method

This method is based on an iterative approach for system development. In this method, a prototype of the system i.e. a small working model of the system is developed instead of a complete system. Though, it is comprehensive system but does not include all requirements of users. This model is released to users at an early stage for trial and to understand the requirements before developing a complete system. This method is useful where users' requirements are difficult to identify in the beginning and requirements may also change during system development process. The system is finally developed based on the advice and feedback of the users. This approach is centered towards user involvement in analysis as well as design phase to provide satisfactory system which reduces the chances of product rejection. The activities that are involved in SDLC are also taken care of in this approach but here the emphasis is on the involvement of user. The prototype model is based on the evolutionary method of system development as shown in fig. 5.4.

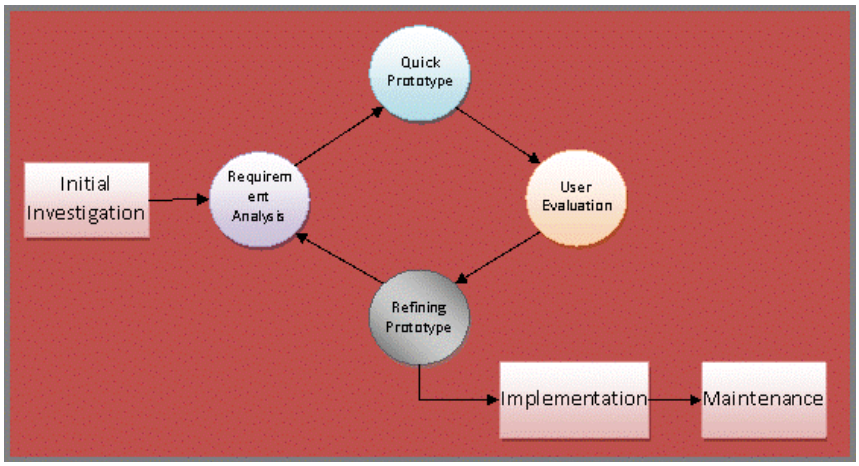


Fig. 5.4 Prototype model (iterative approach)

5.7.4.1 Reasons for developing a prototype system

- **High cost:** System, involving high cost or high risk situation, will be most suited to prototyping approach as prototyping provides an opportunity to user for directly interacting in the process of system analysis and design. Consequently reduces cost of system development.
- **Information not well defined:** Where information requirements are not well defined and the same gets identified and streamlined while working by the user in interactive mode, the prototyping provide a good opportunity to identify the system requirements in a better way.
- **Unknown information requirements:** There may be a very unique situation when one has neither information nor experience of information requirements. The prototyping approach is a good process for giving a time to gain experience and identify information requirement.

5.7.4.2 Advantages

- Users have opportunity to say about what they like the best and least features of the prototype model.
- The chances of product rejection are less due to users' involvements in system development process.
- Requirements are identified in a better way following an iterative modification process until the users' are satisfied.

5.7.5 Spiral method

This method is based on combination of linear and iterative approach for system development. A new element of risk analysis is added in this approach. The focus is on risk assessment and on minimizing system risk by breaking a system into smaller segments and providing more ease of change during development process as well as providing opportunity to evaluate risks and weigh consideration of system continuation throughout the life cycle. This method suggests that various activities involved in system development should be organized in spiral form. Each cycle of spiral consists of four stages namely planning, risk analysis, engineering and user evaluation. These stages are represented on four

quadrants as shown in fig 5.5 given below. An angular dimension represents progress in the development process.

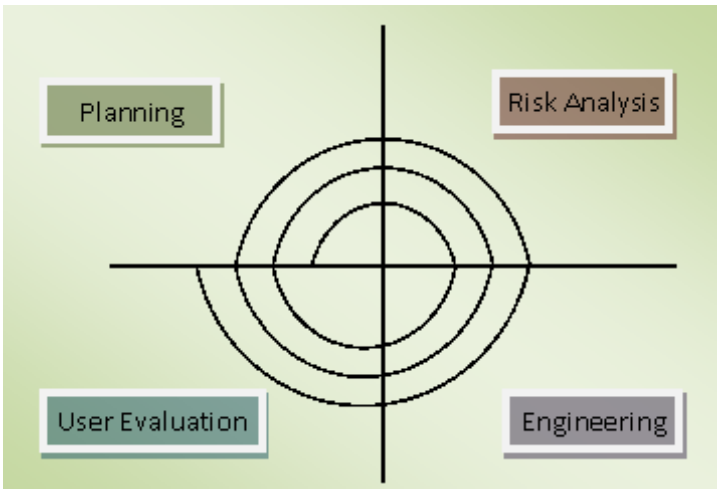


Fig. 5.5 Spiral method (combination of linear and iterative approach)

5.7.6 Object oriented analysis

Structured analysis treats processes and data as separate components, while Object Oriented Analysis (OOA) combines data and the processes that act on the data into things called objects. OOA method is used to model real world business processes and operations. In this method, the requirement of users are analyzed and understood from perspective of classes and objects found in the problem domain. Result is a set of software objects that represent people, things, transactions and events. Using object oriented programming language, a programmer transforms the objects into reusable code and components. An object is a member of a class, which is a collection of similar objects. Objects possess characteristics called properties, which it inherits from its class or possess on its own. This approach has following major phases:-

- Object Oriented Analysis
- Object Oriented Design
- Object Oriented Testing
- Implementation

5.8 Systems Analyst

A systems analyst is responsible for researching, planning and recommending software and system choices to meet an organization's business requirements. Task of a systems analyst is to identify needs and resource constraints and to translate these into a viable operation. A system analyst investigates, analyses, designs develops, installs, evaluates, and maintains an information system of an organization. To perform these tasks system analysts constantly interact with users and managers within and outside the organization. Systems analyst plays a vital role in analysis and design of a system or an application in any business/ scientific environment. Generally system analyst is supported by a team of IT department of the company.

A systems analyst studies the problems and needs of a business system to evaluate how business system and computer system can be used together to solve the problems and achieve higher profits in business. For studying and identifying the needs of an existing system or for developing a new system, an analyst has to perform the following tasks:

- Collect and study the historical background of the organization.
- Define business and system objectives.
- Identify system requirements and their specifications keeping in view of future expansion and other business requirements.
- Work out various alternative feasible solutions suiting to organizations resources;
- Identify hardware and software requirements and their selection based on design recommended.
- Design input and output procedures; data integrity rules, database, etc.
- Develop complete system and testing procedure.
- Implement, update, and maintain system till the system becomes workable solution in accordance with the system and business objectives.

To carry out the above mentioned activities, system analyst must have the following qualities and skills to play the role of investigator; change agent; system designer; IT expert; trainer and motivator etc.

5.8.1 Personal characteristics

- Communication Skill
- Quick assimilation and sharpness in understanding skills
- Far sightedness and vision
- Adaptability and Flexibility skills
- Sound temperament, patience and rationality
- Management skills
- Leadership Quality
- Training and Documentation capabilities

5.8.2 Technical skills

- Application domain and IT Knowledge
- Creativity
- Project Management
- Interdisciplinary interfaces
- Technical Capability of questioning

Lesson 6

SYSTEMS ANALYSIS AND DESIGN

6.1 Introduction

This lesson will discuss the objectives of systems analysis and design. Various activities involved in systems analysis and design like requirement analysis, feasibility study, graphical tools, input/ output design, codes design, data validation, files design, program design, system development, hardware and software requirement, implementation, testing, maintenance, etc., are elucidated in this lesson. These topics will be useful for understanding the system which will be helpful in designing the system at later stage.

In early days, system analysis and design was concerned with manmade system involving inputs, processes and outputs. But in modern times, system analysis and design deals with the process of examining and understanding the working of an existing system of an organization, identifying problems (if any) with an objective to improve the same through better methods, technology and procedures. In general, the phrase "System Analysis and Design" relates to the complete process of developing a system which involves the following steps:

1. Systems analysis
2. Systems design
3. Systems development
4. Systems implementation
5. Systems maintenance

6.2 Systems Analysis

Systems analysis is a process of collecting and interpreting facts about the present system, diagnosing problems, analyzing the business requirements and recommending improvements based on information gathered. In other words, systems analysis means identification, understanding, and examining the system, which creates base for system design in order to achieve pre-determined goals/ objectives of the system. System analysis is carried out with the following objectives

- To understand the functioning of present system.
- To analyze the users requirement for initiating computer based information system.
- To develop a logical model of feasible solution to solve problems.

Systems analysis is basically a detailed study of all important business aspects under consideration and thus the study becomes a basis for the proposed system. The proposed system may be a new system or some modification in the existing system. Systems analysis is a logical process for understanding the system. The emphasis is on investigation to know how the system is currently working, to identify requirement of users from the system and to determine what best can be done to solve the problem.

6.3 Systems Design

Logical design of an information system developed during system analysis phase defines functions and features of the system and relationship among its components. Logical design includes output that must be produced by the system; input needed by the system, and processes that must be performed by the system without regard to how tasks will be accomplished physically. During system designing phase, logical design is translated into physical design with respect to user requirements, data flow of existing system, I/O specification of the existing system. In contrast to logical design, physical design is a plan for actual implementation of the system. Physical design specification must answer the following:

I. How data will be entered?

II. How the files are organized?

III. How reports will be generated?

IV. What will be format or layout of reports?

6.4 Systems Development

During systems development phase, analyst mainly focuses on constructing a workable system which will fulfill the requirements as per the specifications laid down in system analysis and design phase. System analyst monitors the operation of development of system to ensure higher levels of satisfaction and performance. Database and tables are created for storing data using database management software packages (DBMS), programs are written for data entry with data validation checks to ensure accurate and reliable input into databases, report generation, creating GUIs, procedures, module, menus, etc. using computer programming languages or 4GL software packages. Plans are worked out to test performance of overall system.

6.5 Systems Implementation

Once the development of information system is successfully completed, it is ready for implementation. In the implementation phase, newly developed system is installed at the user's site. In other words it is the process of replacing the existing system with a new system. During this process, transition from old system to new system should be smooth and acceptable to users to avoid any unpleasant situation or chaos which may lead to rejection of the system. The new system may be a replacement of a manual system or a major modification in the existing computer based information system. Therefore, SA should carefully plan and design methods for changeover from old system to new system.

6.6 Systems Maintenance

Job of system analyst does not finish after implementation of the system at user's site rather his real acid test starts after implementation of software as most of his time will be spent on maintaining the system and meeting the users' requirement. As the organizations are in dynamic and competitive

world, evaluation and maintenance is a continuous process. Corrective actions are taken based on the users' feedback to evaluate the system. During maintenance phase, system analyst mainly focus on incorporating the changes occurring due to factors like external environmental, internal policy, users preferences, software and hardware platforms etc. to make system more efficient and effective. It should be ensured that all related components and modules of the system must be corrected whenever faults are detected and fixed. Process of monitoring, evaluating, and modifying of existing system to make the required or desirable improvements is terms as *system maintenance*. Different types of maintenance undertaken by SA are Corrective maintenance, Adaptive maintenance, Enhancement maintenance, Preventive maintenance

Detailed activities covered in above mentioned steps are summarized in table 6.1.

Table 16.1 Summary of activities involved in systems development

SN	Phases/ Activities	Key Questions	Results
1.	**Systems Analysis Phase**		
i	**Need recognition:** • Preliminary survey/ initial investigation	• What is the problem or opportunity? • Why computerized system is required?	• Statement of scope and objectives • Performance criteria
ii	**Feasibility study:** • Evaluation of existing system and procedures • Analysis of alternative candidate systems • Cost estimates	• What are the user's demonstrable needs? • Is the problem worth solving? • How can the problem be redefined? • Whether the solution to the problem will be cost effective, technical and operational viable?	• Economic Technical and Operational feasibility • Cost benefit analysis • System scope and objectives • Statement of new scope and objectives

iii	**Detailed system study:** • Study the existing system in detail to understand terminology, components, procedure, workflow and data flow, relationship between components • Data collection	• What must be done to solve the problem? • What are the facts?	• Statement showing the working of existing system • Logical model of the system – data flow diagrams, Data dictionary, system flow chart • Pertinent data • Questionnaires/interviews for data collection
iv	**System requirements:** • Identify input and output requirements and their formats • Hardware and software requirements	• How the system interact with outer environment • How different components are interacting with each others	• More suitable new input and output formats • Specifications for Hardware and software requirements
v	**Iteration for improvement:** • Above mentioned activities are repeated	• How a system can further be improved?	• Final result of system study and analysis phase
2.	**Systems Design Phase**		
i	**General design specification:** • Translation of logical design to physical design	• How must the problem be solved? • What is the new system processing flow?	• Physical design of the system • Flow of information into and out of the system • Implementation plan

ii	**Input–Output design:** • Identify inputs to the system • Identify output of the system	• How the reports will be generated • How the data will be entered • How the data will be validated • What will be the formats or layout of input and output	• New formats for output reports • New formats/ screens for data entry forms • Different types of data validation checks used during data entry • Devices and method to be used for generating output and data entry
iii	**File and Database design:** • Storage of input data and generated data	• How to design an efficient database • How the data will be stored in data files using the concepts of normalization? • What will be the structure of data files? • How the data files will be related to each other?	• Logical design of database • Schema and sub-schema of database • Classification of data files • Identification of primary key for each data file • Creation of relationship between data files
iv	**Code design:** • Designing of codes for input and output data to identify and retrieve record uniquely in files	• How to design efficient coding pattern	• List of codes for different data items for input and output
v	**Procedure and Program logic design:** • Identify procedures of the system to be designed and developed	• Does the user approve the system? • How to develop logic of	• Algorithms, flow charts and pseudo codes of programs for data entry, output reports, menu design, intermediate data

	for processing input/ output data	programs using different tools?	processing etc.
3.	**Systems Development Phase**		
i	**Database creation:** • Create the designed data base using a DBSM or some other file creation method	• Which DBMS should be selected to meet the organization requirements?	• Physically creation of database using logical design
ii	**Program writing (i.e. coding):** • Writing of programs using SQL or some other programming languages like VC++, C++, VB, Java, VB.net etc.	• Which programming language should be used for writing programs?	• Programs for solving problems, data entry, report generation, menus etc.
iii	**Testing:** • Program testing • Module testing • System testing • Complete quality and performance testing	• How well do individual programs/ modules test out? • How ready are programs for acceptance test?	• Test plans • Formal system test • Performance of software with respect to hardware
4.	**Systems Implementation Phase**		
i	**File/ system conversion (installation):** • Software is installed at	• What is the actual	• Plan of to change over to new system (At

	client site • Data is entered in files using new software or existing data is converted into file/ database structure of new system	operation? • Are there delays in loading/ converting data files?	once/ Parallel/ Location wise/ Modular) • Data is entered in new files/ database
ii	**User training:** • Train the end users about the operation of software developed and exception handling • Exposure to executive staff about the software and its capabilities	• What to teach the users? • How to train the staff?	• Plan for training the staff (Brainstorming sessions/ Seminars/ Operational/ Awareness/ On Job trainings, etc.)
iii	**Documentation:** • Prepare different kind of manuals with respect to level of usage	• Are manuals ready?	• User manual for beginners • Technical manual • System manual
5.	**Post-implementation, Maintenance and Review Phase**		
i	**Evaluation:** • The implemented system is evaluated by end users	• Is the key system running?	• Users requirements and standards are met
ii	**Maintenance:** • Collect feedback of users about the system • Focus on changes	• Should the system be modified?	• Updated the system as per the requirements/ suggestions

	associated with error correction, exceptional situation occurred, performance of system, backups and recovery of data files, etc.		• Satisfied users
iii	**Enhancement:** • Addition of new plans/ modules/ schemes due to changes in users priorities, environmental factors, organizational requirements, etc.	• How to incorporate new plans/ modules in the existing system?	• Steps involved in system analysis, design and implementations are repeated to add new modules

6.7 Requirement Analysis

Requirement analysis is a way of translating user's ideas and requirement about the problem into formal documents which laid down foundation of system design and development. This is the most difficult and error prone activity because of communication gap between users and developer. Generally user does not understand software thus, not able to explain their requirement and SA does not understand user's problem/ application. A properly conducted requirement analysis has following benefits:

- It will satisfy the business requirement which would be acceptable to users.
- It bridges the communication gap between user and system analyst.
- It reduces development cost by overcoming errors and misunderstanding at an early stage.
- It serves as benchmarks to measure overall acceptability of the developed system.

Requirement analysis is basically to learn and collect information about system under consideration. Various methods for collecting information about organization in order to study the existing system and to identify system requirements are:

- Interviewing
- Documents
- Questionnaires

- Observation

6.8 Feasibility Study

Feasibility study tries to ascertain whether the system is viable or not for the organization. System request is passed through a series of tests like what is the problem to be solved? Is the problem even worth solving? Whether it is worthwhile to proceed further or not? The series of tests is called a feasibility study or analysis which determines likelihood that the proposed system will be useful to the organization or not. Feasibility analysis is an important part for every system under consideration for development. Objective of feasibility analysis is to assess various alternative systems and to propose most suitable system for development. Depending on the system request, feasibility study may be quite simple that can be done in few hours only or it may be complex and extensive fact finding which may take number of days or months. For example, if the milk procurement section wants that reports to be produced in different order from the existing system, analyst can decide quickly whether request is feasible. On the other hand, a proposal from milk market section, for developing a new market research system to predict sales trends require more efforts. Feasibility of a proposed system can be assessed in terms of *operational*, *technical*, and *economical* feasibility.

6.8.1 Operational feasibility

An operational feasibility of system means that the proposed system has been accepted by the users and will be used effectively after it has been developed and implemented. Operational feasibility of a system ensures that the system will work after its implementation. Human resource of the organization has to be competent and enthusiastic to accept the change which is likely to take place by implementing the proposed system. In other words, management, employees, customers, users etc. must be willing and able to support the new proposed system. If users have difficulty with a new system then it may be rejected and will not produce expected benefits. Therefore motivation and involvement of users is must from beginning in development of the proposed system. Operational feasibility depends on several vital issues and analyst must answer following questions:

I. Is their sufficient support from management and users?
II. Are the current methods acceptable to user or they need change?
III. Will proposed system not result in loss of control in any area?
IV. Will the new system result in reduction of employees?
V. If yes, what will happen to the affected employees?

Analyst must get positive answers to these questions to make system operationally feasible. Generally a system developed with active involvement of users and as per rules, regulations, organization culture, union agreements, etc. would be operational feasible.

6.8.2 Technical feasibility

A technical feasibility means that the organization has resources and capability to procure the required technology and equipment (hardware, software etc) to develop, install and operate the proposed system within specified time period. Besides this, the proposed system should be capable of meeting all requirements of people working in the organization. System should be flexible and expendable to sustain for longer period. To satisfy management and users, system must ensure security, reliability, accuracy and accessibility of data in most technical way. While assessing the technical feasibility, analyst must consider following questions:

I. Does necessary technology exist for developing and implementing the proposed system and can it be acquired?

II. Does organization have requisite technical expertise? If not, can it be acquired?

III. Are the proposed devices and media capable to hold the volume of data required?

IV. Will the proposed system be compatible with the coming up technology?

V. Will the existing resources like hardware, software and employees be used?

6.8.3 Economical and financial feasibility

Economic feasibility means that the projected benefits of proposed system should outweigh the estimated cost. Proposed system will involve cost on deriving system requirement specifications, hardware, software, installation, maintenance, site preparation, training and consumable cost. Costs can be one time investment or recurring in nature. Projected benefits should justify investment in the system since finance is the main constraints for any system to be accepted or rejected. Management must be satisfied with benefits to be derived by implementing the proposed system. While assessing the economical feasibility, analyst must consider following questions and should get positive answers to these questions to make the system economical feasible:

I. Whether total cost is within the allocated budget? Does the additional cost make overall business competitive and consequently resulting in maximizing profit?

II. Do the benefits in form of reduced cost per unit in the long run, improve customer services and whether improved resource utilization justifies the investment?

Analyst must perform cost and benefit analysis of the proposed system and alternate solutions before making any recommendations to the management.

6.9 Data and Process Modeling Techniques

These techniques are used by systems analyst during system analysis to depict how various processes of a system transform data into useful information in graphical form. Output of process modeling activity is a logical model that supports business operations and meets requirements of managers and users. A logical model shows what the system must do, regardless of how it will be accomplished physically. System analyst may use one or more of the following traditional and structured analysis tools to represent system data and process graphically.

I. System flow chart
II. Data flow diagram (DFD)
III. Data dictionary
IV. Decision tree
V. Decision table

6.9.1 System flow chart

A system flowchart is a graphical/ pictorial representation of the system's discrete physical components such as programs, procedures, files, reports, screens, etc. It is a valuable presentation tool to show the interaction and interlinking of major components of a system in sequential order to achieve final meaningful results through the system. It serves as a system roadmap.

Since system flow chart emphasizes on physical implementation details about the system therefore it is not an appropriate modeling tool at an early stage of systems analysis. However, it could be a useful modeling tool at the end of the systems analysis activity, when the user implementation model is being developed.

Different symbols are used to show various components in a system flow chart. For example rectangular boxes represent operational aggregates of computer software e.g., computer programs, job steps, runs, or other units of computer software, magnetic tape and hard disk symbols are used to represent different kinds of physical files, on-line terminals and telecommunication links etc. Some of the important symbols used in drawing a system flow chart are shown in fig. 6.1 given below:

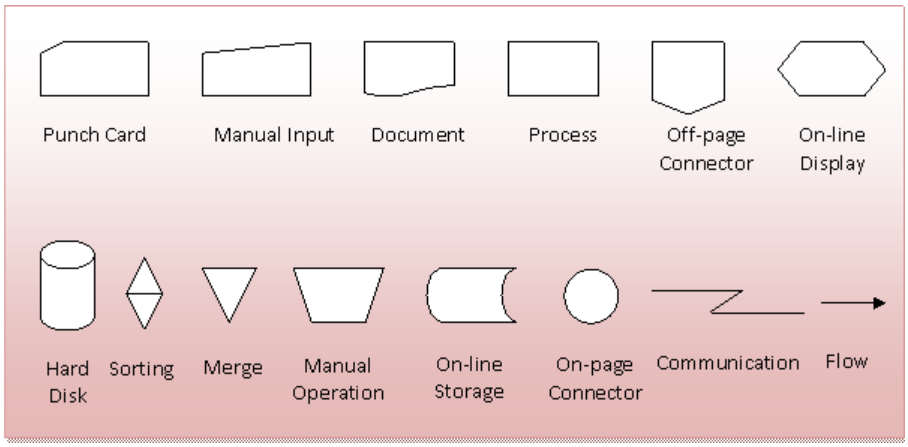


Fig. 6.1 System flow chart symbols

Example: A typical system flow chart for milk receipt and billing system in brief is shown in fig. 6.2 given below. The same concept is shown in two styles.

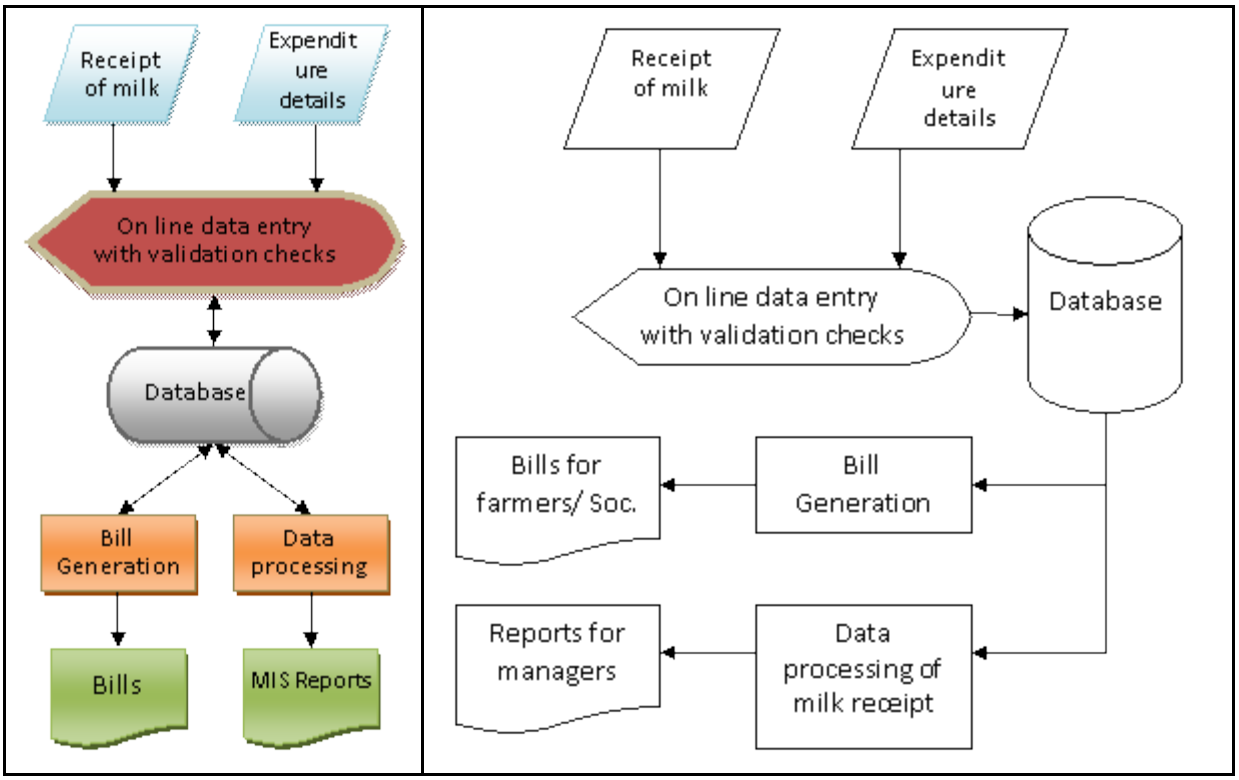


Fig. 6.2 System flow chart for milk receipt and billing system

6.9.2 Data flow diagram

DFD is a graphical or pictorial representation of logical flow of data. This tool describes the movement or flow of data within an information system from beginning up till end regardless of whether the system is manual or automated. DFD shows only data flow not the program logic or processing steps. It

models a system by using external entities from which data flows to a process which transforms data and creates output data flows which go to other processes or external entities or files. Data in files may also flow to processes as inputs. Main advantage of DFD is that it provides a logical model of the system that shows what a system does rather than how it is being done. Aim is to clarify system requirements and identify major processes, transformations of data, data stores and external entities to ensure developed model is accurate and easy to understand. DFD is also known as bubble charts as it consists of a series of bubbles joined by lines.

A DFD uses four pictorial symbols to represent processes, data flows, data store, and external entities. Several versions of DFD symbols exist but all serves the same purpose. Two major thoughts for drawing DFD are shown in fig. 6.3 given below.

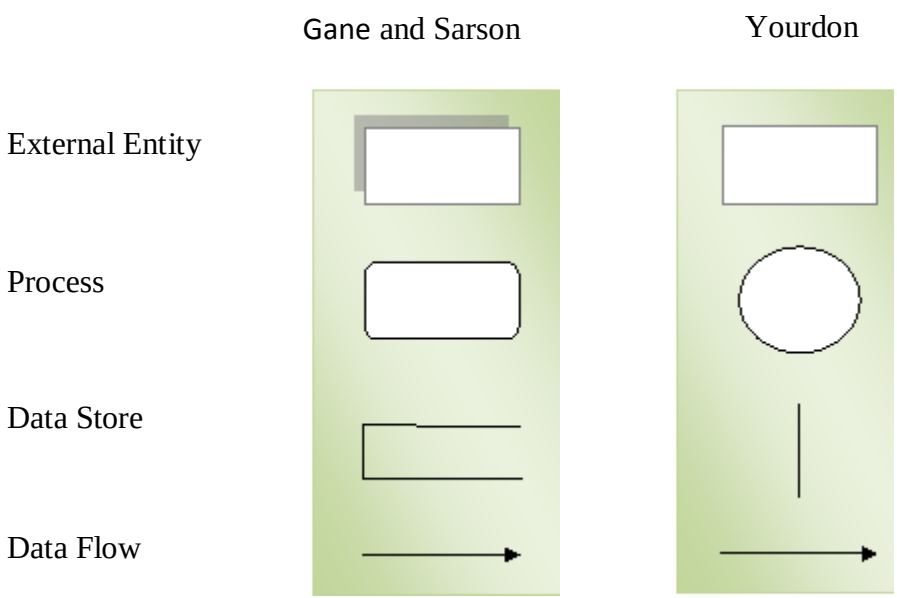


Fig. 6.3 DFD symbols used in two different schemes

a. **External Entity**: External entity is a source or destination of data. It can be files, departments, person, vendors, customer, other information systems etc. External entities show the system boundaries and how a system interacts with outside world. These are also called terminators since they are data origins (source) or final destination (sink). Name of the external entity is written inside the symbol.

b. **Process**: Process is the people or procedures involved in the system that are used to transform data. It receives input data, processes it and produces output. In DFDs a process is just like a black box for which underling details are not shown except input, output and general function of the process. Process name is written inside the symbol.

c. **Data store**: Data store is a data repository in a system. These are used to show the data stored by a process within a system like files, registers, documents, reports, vouchers, etc. The stored data

are used or shared by other processes of the system at later stage. A data store must be connected to a process with a data flow. Name of the data store is written inside the symbol.

d. **Data flow**: Data flow is a path that shows movement of data from one process and to another process in a system from origin to destination. Data flow names appear above, below, or alongside the line. Data flow names represent one or more data items.

6.9.2.1 Procedure for drawing a DFD

Step 1: Represent the whole system by a single process and identify its input and outputs. This is known as context diagram or zero level DFD. A context DFD is a top level view of a system that shows boundaries and scope of the system. To draw context level DFD put a single process symbol in the centre which represents the entire system. Then place external entities around the central process symbol and use data flows to connect entities to it. Data stores are not being used in context diagram because these are internal to the system. Information about names and contents of external entities and data flows are collected during fact gathering process.

Step 2: Context diagram shows the most general view of a system which contains only one process. This is just like a black box. To know more details about the system next level of DFD are drawn. To draw next level of DFD, identify major processes of the system and draw DFD considering these major processes only and also identify inputs and outputs for each and every major process. This is known as top level DFD or first level DFD.

Step 3: After identification of major processes in first level DFD, an analyst has to identify processes from first level DFD which can be further expanded. Analyst then draws DFD of a particular process identified with expansion. This is known as exploded DFD or Expanded DFD,

6.9.2.2 General rules for drawing DFD

- Processes should be uniquely named and numbered
- Data flows within the system must be named
- Context DFD must fit on the single page
- Process name in the context DFD should be the name of the system
- Do not cross the lines

Example: Fig 6.4 shows a context diagram for milk procurement and billing system (MPBS).

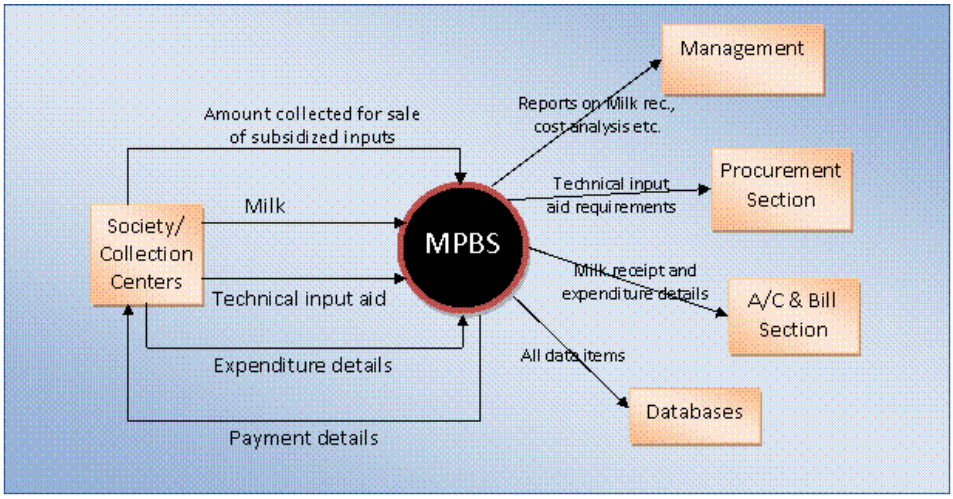


Fig. 6.4 Context diagram or zero level DFD of MPBS

6.9.3 Data Dictionary

A data dictionary is a structured repository of description of data item involved in a system. As discussed earlier, DFDs describe the logical model of a system in which details of input and output data items is not shown. The details of data items are documented separately in data dictionary. A data dictionary is an organized collection of precise and accurate definition of all processes, data elements, data structures, data stores and data flows of a systems. Data structure consists of a group of data elements and data element is simplified unit of data which cannot be further decomposed. For example employee name may consist of three parts as first name, middle name, surname. Thus employee name is data structure and its parts are data elements. Data structures also known as records are meaningful combination of related data elements that is included in a data flow or retained in a data store.

6.9.4 Decision Tree

A decision tree is a graphical representation for describing logical conditions, actions and rules. It describes all actions that result from various combinations of conditions as per logical rules. The graphical representation of conditions and outcomes resembles with the branches of a tree. The branches of a tree depends the logical alternatives. It is drawn horizontally with the root on left side and branches to the right. Though, it is easy to construct, read and update a decision tree for a simple problem but as the complexity increases it becomes tedious to draw and update.

Example: Following figure 6.5 shows the subsidy given to dairy farmers on technical inputs to the farmers by dairy plants for promotion of dairy in the area.

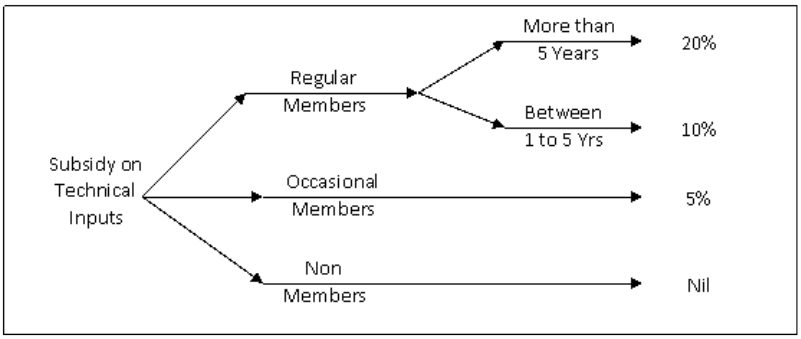


Fig. 6.5 An example of decision tree

6.9.6 Decision Table

Decision table describes a logical structure with all possible combination of conditions, actions and decision rules for initiating an action on occurrence of various conditions. It shows conditions and actions in a simplified and orderly manner in a form of matrix of rows and columns. Decision table describes the same information as discussed in decision tree but in a tabular form. Analysts may also use other tools like decision tree, structured English, pseudo codes to describe a logical process and to ensure to cover all combination of logical possibility. A decision table is a better tool to represent combinations of complex conditions while decision tree is effective to describe relatively simple process. A decision table consists of four sections as described below and shown in table 6.2:

- **Condition stub (upper left part):** All conditions are written in this area.
- **Action stub (lower left part):** All actions are written in this area.
- **Condition entry (upper right part):** All possible combination of conditions are marked either by 'Y' or 'N' where 'Y' is for true and 'N' for falseness of condition.
- **Action entry (lower right part):** In this part the action which will be taken as per the combination of action specified in condition entry is shown by marking 'X' in front of action that will be taken with respect to the condition.

Table 6.2 Structure of a decision table

Condition Stub	Condition Entry
Action Stub	Action Entry

Example: Following example (Table 6.3) shows a decision table constructed to determine rates of fresh milk received from farmers.

Table 6.3 Decision table describing rules for determining rates of fresh milk

Conditions	Decision Rules		
Fat% and SNF% within specified range	Y	N	N
Low Fat% and SNF%		Y	N
High Fat% and SNF%			Y
Normal Rate	X		X
Penalty		X	
Incentive			X

****** ☺ ******

Lesson 7

PROJECT MANAGEMENT TOOLS

7.1 Introduction

This lesson will discuss the concept of project management and various tools for project scheduling and monitoring for timely completion and implementation of information system in the organization. These topics will be useful for monitoring the progress of project as per timelines defined in the beginning of the project.

Project management is the process of defining, planning, organizing, leading, and controlling the development of an information system. Project management is important throughout the entire life cycle of the system development. While developing and conceptualizing the project, managers must consider the project size and scope, IT resources available, prior experience with similar projects, and applicable constraints. Project is divided into number of activities and accordingly the time and budget estimates are prepared for each activity to complete the project within time frame and allotted budget. The management of large projects requires analytical tools for scheduling activities and allocating resources. This lesson describes various tools such as Project Evaluation and Review Technique (PERT), Critical Path Method (CPM), Critical Path Scheduling (CPS) and Gantt chart that has proven valuable to project managers for project management.

7.2 Project management and monitoring

Project management and monitoring is an essential part of information system development to ensure higher success rate and to increase the satisfaction of customers by delivering/implementing the system as per deadlines. Project manager has to plan and control activities within the computer based system and resources like people, equipment, money, time, etc., for successful development of system. Efficient planning and monitoring keeps budget under control and create an environment of success and users satisfaction.

Generally information systems developed for business activities involves data processing. This involves higher risks, giving rise to the expenditure for the system to be developed. On order to reduce the effect of risks or uncertainties, planning and controlling measures are to be taken care of by the analyst. Project manager has to monitor the plan very closely in such a way that even smaller deviations from plans are recognized and corrective actions are taken to avoid or mitigate the effect and deviations. Despite best efforts, projects may fail due to number of factor such as:

- Goals are not understood.
- Escalation in budget.
- Lack of coordination among various teams of project development
- Planning was performed by a separate group who were not part of analysis team.
- User/Customer not co-operating.
- User changes the requirement frequently.
- Frequently shifting of skilled staff from the project team.

- Unforeseen constraints

7.3 Project scheduling

An analyst must plan the project with applicable constraints and should design a yardstick to control the development of system to minimize the ambiguities and impact of risks and their consequences. Each step or activity of the project must be planned and scheduled to monitor the progress of the project. Project manager must know the activities to be performed, duration of each activity, the order in which the activities will be performed, the start and end times for each activity, and who will be assigned to each specific task. Project scheduling is an integral part of project development. Manager has to consider and perform the following actions while preparing the project schedule.

7.3.1 Define activities

A list of activities with their detailed description is prepared for the system. This list is used as a reference by each and every member of the development team.

7.3.2 Estimate duration of each activities

After defining the activities, estimate the time for completing each activities. This will help the manager to estimate the total time required for developing a system. The time should be calculated in man days. To estimate the time, weighted average of three estimates is selected for each activity. The three estimates are as given below:

(i) Optimistic time: How long the activity will take to complete the job, if everything goes as per plan.
(ii) Most likely time: How long the activity will take normally to complete the job.
(iii) Pessimistic time: How long the activity will take to complete the job, if everything does not go as per plan.
The estimated time for each activity will be computed on the basis of the formula (eq. 1) given below. However, manager can device their own criteria for weighted average (if required).

$$Estimated\ time = \left(\frac{O+4M+P}{6}\right) \qquad (1)$$

Where

O = Optimistic time
M = Most likely time
P = Pessimistic time

7.3.3 Activities relationship and priority

Generally, activities are dependent on each other; therefore the manager has to determine which activities are inter-dependent. A dependent activity cannot be started until one or more other related tasks are completed. Hence all activities cannot be accomplished concurrently, but activities that are to be accomplished first should be given higher priority in order to optimally utilize the resources. For example one cannot go for pilot testing of the software until required hardware has been installed. After identifying all the activity dependencies, arrange the activities in a logical sequence.

7.3.4 Define start and end time of each activities

Define the staring and end time of each activity to be performed. An activity cannot start until all preceding activities on which it depends are completed. The ending time for an activity is its start time plus whatever time it takes to complete the task.

7.3.5 Assign man power

On the basis of skill of the members of development team, activities should be assigned. After assignment, analyst calculates the man days with respect to the member who is assigned the activities. The resulting figure is compared with estimated man days, if the comparison is balanced then the analyst will continue with assignment job else the analyst has to reassign the task to avoid delays with respect to time in man days estimated in the previous step. Assignment should not overload or under utilize team members.

After performing the above actions, analyst can represent the activities pictorially for better understanding while scheduling project using graphical tools such as PERT/CPM and Gantt chart.

7.4 PERT/CPM

PERT and CPM are project management techniques, which have been created to plan, schedule and control complex projects. PERT was developed by the US Navy to manage the construction of nuclear submarine during 1958. CPM was the developed approximately during the same period (1957) by private industry to meet similar project management needs. The first test was made in 1958, when CPM was applied to the construction of a new chemical plant.

A PERT/CPM chart shows a project as a network diagram. The project is defined as a collection of interrelated activities with each activity consuming time and resources. PERT/CPM provide analytic means for scheduling the activities. CPM deals with the situations when deterministic time estimates for the activity completion time can be given precisely while PERT can be used in the situations when the activity completion time cannot be given precisely i.e. activity completion time is probabilistic. The distinction between the two methods have disappeared over time and today the technique is called either PERT, CPM or PERT/CPM.

7.4.1 Terms related to PERT/CPM

Commonly used terms to prepare PERT/CPM chats are as follows:

- **Activity:** An activity is a task or job required to be completed within a certain time limit. Activity is represented by an arrow shaft (i.e. Vectors).
- **Event:** It is a point which represents start or finish of an activity. It is represented by circle. The event at the beginning of the activity is called start event and the event at the end of the activity is called as finish event. The activity lines connect one node to another. The length of a line is not related with duration of the activity it represents. Fig.7.1 shows two events connected by an activity line/ vector. Each event is identified by a number event-1 is the beginning of the activity and event-2 marks the end. Each activity is identified by a short description above the vector, or with a letter or code explained in a table. The estimated duration of the activity appears below the vector.

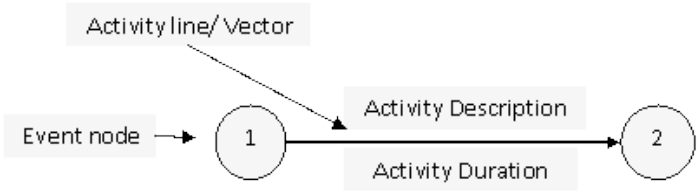


Fig. 7.1 Events connected by activity vector

- **Dummy activity:** A dummy activity in a PERT/CPM chart indicates an event dependency. The dummy activities neither utilize any resources nor does it consume any unit of time. These activities are represented by dotted lines to create sequence of activities. For example, in fig. 7.2 the dummy activity connecting events 3 and 4 identifies that event 4 cannot take place until event 3 occurs.

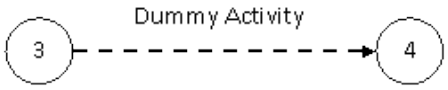


Fig. 7.2 Events connected by dummy activity

- **Dependent activities:** When activities are completed in a sequence, they are called dependent or serial activities. Fig. 7.3(a) shows three dependent activities A, B, and C. One can notice that activity A must end before the start of activity B. Event 3, which marks the end of activity B, must occur before activity C can start.
- **Parallel/ Concurrent activities:** When activities are completed at the same time, they are called concurrent or parallel activities. Fig. 7.3(b) shows activities A and B are parallel that can be done at the same time, but the duration of these two activities may be different. For example activity A may take one week and activity B any take 5 weeks for completion. One can notice that activity C depends on completion of activities A and B i.e. activity C cannot start until both A and B are completed. The dummy activity between events 2 and 3 connects activities A and B into a single path that leads to activity C and shows that they both must be completed before C can begin.

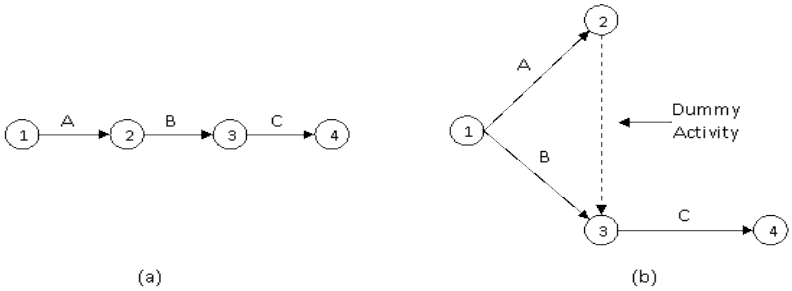


Fig. 7.3 (a) Dependent activities (b) Parallel activities

- **Earliest Completion Time (ECT):** This is the minimum amount of time necessary to complete all activities that precede the event. To determine ECT, work out from left to right.
- **Latest Completion Time (LCT):** LCT for an event is the latest time at which the event can occur without delaying the project. To determine LCTs, you may work backward through the chart, from right to left.

- **Critical Path (CP):** A critical path is a series of events and activities which determine the total length of the project. It is the longest path through the network for which every node has equal ECT and LCT. A critical path is a series of events and activities with no slack time. If any activity along the critical path falls behind schedule, the entire project schedule is similarly delayed. Critical path include all activities that are vital to the project schedule. So the critical path determines the project completion date. Project managers must know what the critical path is, so that they can monitor progress and make prompt decisions, if necessary, to keep the project on track.

7.4.2 Procedure for drawing PERT/CPM chart

1. Define the Project and all of its significant activities or tasks. The Project (made up of several tasks) should have only a single start activity and a single finish activity.
2. Develop the relationships among the activities. Decide which activities must precede and which must follow others.
3. Draw the "Network" connecting all the activities. Each Activity should have unique event numbers. Dummy arrows are used where required to avoid giving the same numbering to two activities.
4. Assign time and/or cost estimates to each activity
5. Determine the earliest completion time for each event.
6. Determine the latest completion time for each event.
7. Compute critical path which is longest time path in the network.
8. Use the network to help plan, schedule, monitor and control the project.

The key concept used by CPM/PERT is that a small set of activities, which make up the longest path through the activity network control the entire project. If these "critical" activities could be identified and assigned to responsible persons, management resources could be optimally used by concentrating on the few activities which determine the fate of the entire project. Non-critical activities can be re-planned, rescheduled and resources for them can be reallocated flexibly, without affecting the whole project.

7.4.3 Rules for construction of PERT

- All activities should either go from left to right or from top to bottom in downward direction.
- Loops within the PERT must be avoided.
- Activities should neither cross each other nor intersection of activities is allowed.
- Usage of dummy activities should be as less as possible.

7.5 Example

Draw PERT chart and determine critical path for given set of activities involved in developing a software project for dairy industry. List of activities along with estimated time and dependency is given below in table 7.1.

Table 7.1 List of activities along with estimated time and dependency

Activity Symbol	Activity Description	Dependency	Time (in days)
A	System Analysis	Start	15
B	Hardware Selection	-	25
C	System Design	A	35
D	Hardware Installation	B	20
E	Program Coding	C	50
F	File Conversion	D	25
G	Documentation	D	35
H	Program Testing	E	25
I	User Training	H	25
J	User Test	I	25

Solution:

As per the given information, there are 10 activities and their dependencies are also shown. Network diagram connecting all activities for drawing PERT Chart is shown in fig. 7.4 given below. The diagram includes 10 activities, 11 events and two dummy activities that are used to reconnect parallel paths. The labels shown above each and every arrow represents the name of the activity and the numbers below the arrow line represent the estimated completion time. Number within circles specifies the events, and arrow lines show the progress of network in forward direction.

After identifying the activities, events and duration, next task is to determine the overall length (i.e. critical path) of the project. First step is to compute ECT and LCT for each and every event from the diagram. ECT is the minimum amount of time necessary to complete all the activities that precede the event. ECT is computed by working out forward from left to right across the chart. LCT is the latest time at which the event can occur without delaying the project. LCT is computed by working backward from right to left across the chart. There will exists at least one complete path through the network for which every node has equal ECTs and LCTs. Computed values of ECT and LCT are shown the table 7.2 as well as in fig. 7.4. ECT values are displayed in a *rectangle* and LCT values in a *triangle* symbol for each event above the nodes in Fig. 7.4. Critical path is a sequence of activities/ events where ECT and LCT are same for the events. From the network diagram, you can see the path **1-2-4-6-9-10-11** (or **A-C-E-H-I-J**) has equal number of ECTs and LCTs. This path is the critical path for the given problem and has been highlighted in the figure.

Table 7.2 ECT and LCT values

Event	Activity	ECT	LCT
1		0	0
2	A	15	15
3	B	25	70
4	C	50	50
5	D	45	90*

6	E	100	100
7	F	70	125
8	G	80	125
9	H	125	125
10	I	150	150
11	J	175	175

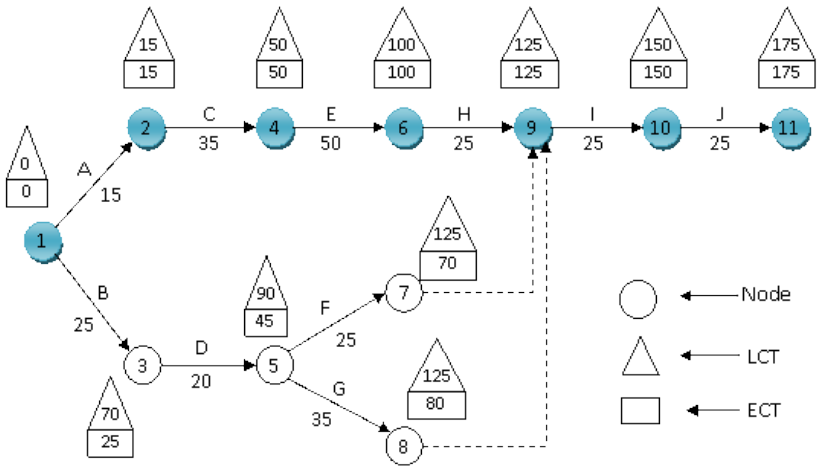


Fig. 7.4 Network diagram showing the PERT/CPM chart

7.6 Gantt Charts

Gantt charts were developed by Henry L. Gantt during 1944, as a production control technique and become very popular among the community. This tool is used to illustrate scheduling of activities involved in a project. Gantt chart is a kind of horizontal bar chart to show the order of activities against time. It uses horizontal bars to show the durations of actions or tasks. The left end of bar represent beginning of activity and the right end represent the end of an activity. Activities are displayed vertically (i.e., on Y axis) and time periods horizontally (i.e., on X axis). The activities may be arranged from top to bottom or bottom to top in the order of their start dates. Gantt chart is shown in fig. 7.5. This chart represents various activities involved in developing an information system in chronological order for the problem discussed in example given above.

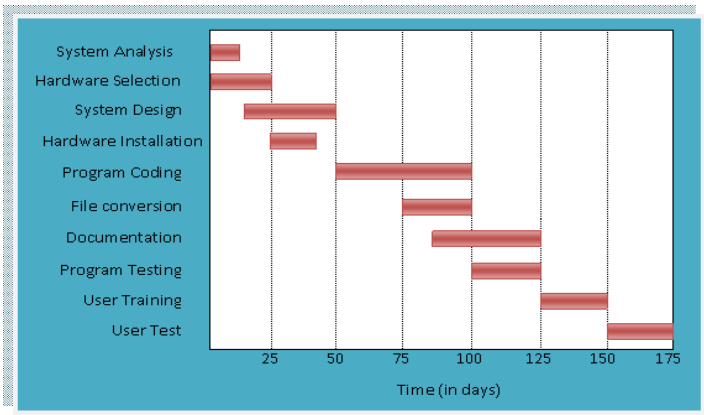


Fig. 7.5 Gantt chart showing the activities involved in system development

7.7 Project Management Software

A large project involving complex relationships and hundreds of tasks/ activities and events is difficult to analyze manually. Therefore, to achieve efficient utilization of people, time, and resources managers are using other project management techniques and software packages to manage large and complex projects. These software packages meet the needs of modern management practices and facilitate the project management process substantially to deal with the complexity of large projects. Project management software offer many features such as estimation and planning of man power, scheduling of activities and resources, PERT and Gantt charts, cost control and budget management, resource allocation, communication, quality management, printed reports, screen displays, graphical plots, and documentation or administration systems.

Project management software can be implemented as a program that runs on the desktop of each user or as a Web application programs accessed through an intranet, or an extranet using a web browser. Desktop version gives the most responsive and graphically-intense style of interface. Web based software packages have all the usual advantages and disadvantages of web applications. Project management software package can be a single-user system or a collaborative system. Single user system is programmed with the assumption that only one person will edit the project plan. A collaborative system is designed to support multiple users modifying different sections of the plan. There are number of project management software available in market, readers may refer the web link for exhaustive list of such software packages *http://en.wikipedia.org/wiki/Comparison_of_project_management_software*. A few commonly used project management software are mentioned below:

1. **Microsoft Project:** This is one of the most popular packages and it now offers a web interface and deep Office, Outlook and SharePoint integration.
2. **Matchware MindView:** MindView has an easy-to-use spreadsheet-like layout, and its mind map option lets you see your project in visuals, reminiscent of brainstorming bubbles.
3. **Project Kickstart:** Project KickStart is an easy-to-use project management package that integrates with other applications like PowerPoint, Outlook, Excel, Word, and Microsoft Project.
4. **RationalPlan Multi Project:** This Project has features to manage resources and budgets as well as multiple projects. It has an interactive Gantt chart, normally available only in higher-end suites.

5. **Basecamp:** Basecamp is a low-cost web-based project management and collaboration package which is gaining momentum.
6. **Milestones Professional:** This is project scheduling software developed by KIDASA Software, Inc., Milestones Professional is a Windows-based program designed to create presentation-ready Gantt Charts and schedules for users.
7. **5pm: 5pm** is an online tool and application service provider for project management. This software is available in 16 languages and provides interactive Gantt chart. Mobile edition is also available.
8. **Collabtive:** It is a web-based project management software published as free software. It provides an open source alternative to proprietary tools likeBasecamp or ActiveCollab. Collabtive is written in PHP5 and uses an AJAX-based user interface. It is multi-language capable, supporting more than 30 languages. Collabtive integrates with other web services via an XML API. It also provides for importing functionality and data from other applications. It allows the user to work with to-do lists, milestones, files and to track the time worked on a task-by-task basis.

****** ☺ ******

Lesson 8

INTRODUCTION TO DATABASES

8.1 Introduction

Students will learn the concept of database, objectives and characteristics of databases, basic terminology, Database Management System (DBMS), components and advantages of DBMS. These topics will be useful for students to build up an understanding of databases and DBMS.

Since the beginning of civilization, data is being collected and processed for drawing conclusions and taking appropriate actions. With the invention of computers during 1950s, data were collected and organized in computer files and processed to generate information and reports. This was generally faster and more accurate than the manual procedures. Group of data and records were stored in separate files based on their applications and usages. Such systems are known as file processing systems. In file processing system, traditionally each department in an organization used to design and maintain their own data files for specific applications. For example, in a dairy plant, personnel section maintains details of all employees working in dairy for various purposes like performance evaluation, timely increments, training etc. Similarly, payroll section responsible for preparation of pay bills of all employees maintains details of employees separately than personnel section, different product manufacturing section may have details of employees working in respective sections for their own purpose.

It can be observed that this kind of arrangement of data in separate files leads to duplication of data and storage space. The details of employees are to be entered many times for different sections/ departments. In this process new errors may be introduced during entry or updation of data in different files. In addition to time consumption for multiple data entry and storage space, when data is allowed to be duplicated, errors can happen if one instance of the data is altered and another instance remains the same. This leads to data inconsistency. In case of duplication, more maintenance and system resources are required to ensure the data integrity. For example change of address or status of an employee in personnel section may not be reflected or performed simultaneously in other similar files maintained in other sections. File formats may be different and incompatible depending on the programming language used to create, read, update and process data from these files. Therefore, even though the data is available in computers but difficult to access from different files due to incompatible formats. File processing system is the simplest approach to process data using computers and easy to manage data for single or personal use.

In view of above mentioned problems and with the advancement in computer hardware, storage media, and software technologies, researchers developed a new approach for organizing data in a different way which was more efficient than the file processing system. A new concept of data management was introduced in form of databases. The term database was first time used in 1963 by System Development Corporation in a symposium and became popular in USA and Europe by late 60s. In simple terms, database is an organized collection of data. According to Charles Bachman, a researcher from General

Electric, Arizona, pioneer in the field of databases, the aim of developing the database systems was to make more effective use of new direct access storage devices. Before that, data processing was made using punched cards and magnetic tapes i.e. serial processing was the dominant activity. Two key data models arose at that time first the network model based on Bachman's ideas, and another (apparently independently) the hierarchical model used in a system developed by North American Rockwell, later adopted by IBM as the cornerstone of their IMS products. E.F. Codd in 1970s proposed new concept of relational data model for data management. He disconnected the schema (logical organization) of a database from the physical storage methods. He criticized the existing models for confusing the abstract description of information structure with descriptions of physical access mechanisms. However relational data model remained of academic research only for a long time. The commercial products based on relational data models Oracle, Ingres and DB2 were appeared during late 1970s. During mid 1980s, the first successful database product launched for microcomputers was dBASE for CP/M and PC-DOS/ MS-DOS operating system. After that a number for software packages appeared in market such as dBASE IV, Foxpro, Visual Foxpro, MS-Access, SQL and MySQL based on the concept of relational models for personal computers.

In the 1990s, attention shifted to Object Oriented Databases. These had some success in fields where it was necessary to handle more complex data than relational systems could easily cope with, such as spatial databases, engineering data, and multimedia data. Some of these ideas were adopted by the relational vendors, who integrated new features into their products, as a result in 2000s, the fascinating area for innovation was the XML database and still is continued. XML database aims to remove the traditional divide between document and data allowing all of an organization's information resources to be held at one place, whether they are highly structured or not.

In future, huge (terabyte) systems are appearing and will require novel means of handling and analyzing large scale databases such as genome project, geological, national security, and space exploration data. Clickstream analysis (the process of collecting, analyzing, and reporting aggregate data for example, in what order the visitors visited the web pages in web sites) is happening now for traffic analysis and e-commerce analysis. Data mining, data warehousing, data marts are a few commonly used techniques today. Mobile database is a product now coming to market in various ways. Distributed transaction processing is becoming the norm for business planning in many arenas.

8.2 Database

The database approach to manage data overcomes many of the shortcomings of the traditional file processing environment. One of the key features of a database system is that data is stored as a single logical unit called database. The data may be spread across multiple physical files in a database but overall it is connected to a single large repository of data. Organizing data in single logical repository allows for easy manipulation and querying of the data, in contrast to traditional file system.

A database may be designed for batch processing, real-time processing, or in-line processing (in which single transactions are processed to completion one at a time but without the tight time constraints of a real-time system). A database can be of any size and varying complexity.

Several authors have defined database in their own way, a few definitions are as follows:

- In most simple terms database is collection of records of different type of data.

- A database consists of an organized collection of data for one or more uses, typically in digital form.
- Database is a mechanized, shared, formally defined and centrally controlled collection of data used in an organization.
- A database is a set of logically related files organized to facilitate access by one or more application programs and to minimize data redundancy. It does not mean that all data relating to a business should be contained in a single database, but simply that all records in a database should be related and that redundant data should be minimized.

More regressively database may be defined as follows:

"A database may be defined as a collection of inter-related data stored together without harmful or unnecessary redundancy to serve multiple applications; the data are stored so that they are independent of programs which use the data; a common and controlled approach is used in adding new data and in modifying and retrieving existing data within the database. The data is structured so as to provide a foundation for future application development. One system is said to contain a collection of databases if they are entirely separate in structure."

8.3 Objectives of a database

Objective of a database system is to make application development easier. Moreover, the development process should be cheap, fast, and flexible. A database system is a repository of data of an organization therefore it should be accurate, private and protected from damage. A database is to handle many related data and files in an integrated form in order to ensure data authenticity, reliability and security with optimum performance rate having least redundant data and permitting multiple users for processing concurrently.

8.4 Characteristics of a database

Some of the desirable characteristics of a database system are as follows:

- **Ability to represent the inherent structure of the data:** The database system should be able to represent the true properties of the data. The implementation procedures should not force the data in to structures which do not represent its inherent nature. For example a system which can only represent tree structures is inadequate.
- **Performance:** Database applications designed for use by a terminal operator must give a response time appropriate for the man-terminal dialogue. In addition the database system must be able to handle an appropriate throughput of transactions. The response time depends on the nature of man-terminal dialogue, kind of applications, and physical storage devices. Response time is of no concern for databases designed for batch processing applications.
- **Minimum cost:** Techniques are used to reduce the cost for storage of data and programming and also minimize the cost of making changes.
- **Minimum redundancy:** Redundancy means duplication of data. In file processing system, lot of duplication occurs. Even with data base techniques, as increasing amounts of information are combined to form integrated databases a great potential exists for the presence of redundant data. As we have seen already redundant data takes more storage space and require more than one updating

operation which may give conflicting results. The objective of database organization is to eliminate redundant data wherever possible (and to control the inconsistency that is caused by redundant data values). Some fields are necessary to keep in multiple files in database to join (or link) two or more files using common fields. Therefore redundancy cannot be completely removed.

- **Search capability:** User of a database may ask a wide variety of questions about the data that are stored in database therefore the database should be able to process anticipated as well as unanticipated queries quickly.
- **Constant growth:** One of the most important characteristics of databases is that they will need to change and grow. Easy restructuring of the database must be possible as new data types and new applications are added. The restructuring should be possible without having to rewrite the application programs and in general should cause as little upheaval as possible.
- **Data integrity:** Data integrity refers to the correctness, completeness, wholeness, soundness and compliance with the standards fixed by the organization and database designers. It is achieved by preventing accidental or deliberate but unauthorized insertion, modification or destruction of data in a database. It is important that the data items and associations between items are not destroyed. Hardware failures and various types of accident may occur occasionally. Storage of data, updating and insertion procedures, must be such that the system can recover from these circumstances without harm to the data. In addition to protecting data from system problems like hardware failure, the integrity checks may also be designed to ensure that data values confirm to certain specified rules. Test checks (or constraints) may be expressed in form of rules to check relationship between data values. Example: Age of student in B. Tech. Program must not be less than 17 years. Similarly milk product code and name sold at milk parlor must be from the list of products manufactured in dairy plant. Constraints are important because they help to ensure data integrity.
- **Privacy and security:** Data in data base systems must be kept secure and private. It must not be stolen or lost since it is vital for the organization. It must be protected from hardware or software failure from catastrophes, and from criminals, incompetent, and people who would misuse it. *Data security* refers to protection of data against accidental or intentional disclosure to unauthorized persons or unauthorized modifications or destruction.*Data privacy* refers to the rights of individuals and organizations to determine for themselves when, how, and to what extent information about them is to be transmitted to others.
- **Interface with the past:** When an organization installs new database software it is important that it can work with the existing programs, procedures, and existing data can be converted. This kind of compatibility can be a major constraint in switching over to new database system.
- **Interface with the future:** This is an important feature since in future the data and its storage media will change in many ways. An organization has to move along with advancement taking place in technology for its survival. Moreover, the needs of an organization also grow over the years due to changes in policies, expansion in business, etc. Therefore under these circumstances the database should be designed in such a way that these changes can be incorporated in it with minimum cost. This can be achieved to some extent by using the concept of logical and physical data independence concept while designing the database.

- **Simplicity:** The overall logical view of data should be conceived in a simple and neat fashion. In many systems pointers are used in the logical representation to show relationships between data items. But, as more and more relationships are added the representation between data items becomes more complicated and difficult to represent overall logical view of data with clarity. Multiple pointer links can be highly misleading.
- **Data integration:** Data integration provides a unified view of data by extracting or combining data available in different files or sources. This process is becoming more important in today's time since many databases are being developed for commercial and scientific applications by different organizations and users need to extract data from these sources for their work. It has become the focus of extensive research work, and numerous open problems remain unsolved. Data integration is also referred as "Enterprise Information Integration" (EII).
- **Data consistency:** It summarizes the validity, accuracy, usability and integrity of related data between applications and across organization. This ensures that each user observes a consistent view of the data, including visible changes made by the user's own transactions and transactions of other users or processes. Data Consistency problems may arise at any time but are frequently introduced during or following recovery situations when backup copies of the data are used in place of the original data or by some other reason multiple copies are created.
- **Data sharing:** Individual pieces of data in database may be accesses and used by multiple applications. This feature makes database different from file processing system.
- **Data Independence:** Data at different levels should be independent of each other so that the changes in one level should not affect the other level. The ability to modify the schema definition in one level without affecting schema at next higher level is called data independence. There are two types of data independence namely logical and physical data independence. Physical data independence is the ability to modify the physical description of data without causing logical description or application programs to be rewritten i.e. physical layout and organization of the data may be changed without changing either the overall logical structure of data or application programs. Modifications at the physical level are occasionally necessary to improve performance of databases and keeping abreast with latest development in storage media. Logical data independence is the ability to modify the logical schema without causing application program to be rewritten i.e. overall logical structure of the data may be changed without changing the application programs. Modifications at the logical level are necessary whenever the logical structure of the database is altered to incorporate new developments occurred in the organization. For example dairy plant started the production of a new milk product or a new branch is opened. Logical data independence is more difficult to achieve than physical data independence, since application programs are mainly dependent on the logical structure of the data that they access.
- **Centralized control:** There is centralized control of data in database. All data is collected at one place and distributed or shared from that place. Thus all control controls and rights can be audited from one central place, which make use of database more reliable, trustworthy and easy to use.

8.5 Basic Terminology

Followings are few commonly used terms in database environment:

I. **Byte:** A byte is the smallest individually addressable group of bits generally eight bits.

II. **Data item:** A data item is the smallest unit of named data. It may consist of any number of bits or bytes. It is also referred as field or data element.

III. **Data aggregate:** A Collection of data items within a record, which is given a name and is referred to as a whole. For example DATE may compose of the data items MONTH, DAY and YEAR. It is also known as group or group item.

IV. **Persistent data:** This data is infrequently accessed and not likely to be modified. It exists from session to session.

V. **Transient data:** Data that is created within an application session. At the end of the session, it is discarded or reset back to its default and not stored in a database.

VI. **Record:** A record is a named collection of data items or data aggregates. For example a record named MILKRECEIPT may be a collection of related data items such as RECEIPT DATE, SOCIETY NUMBER, ROUTE NUMBER, TIME, TYPE OF MILK, QUANTITY, SNF%, FAT% etc.

VII. **Segment:** A segment contains one or more data items and is the basic quantum of data which passes to and from the application programs under control of the data base management software.

VIII. **File:** A file is a named collection of all occurrences of a given type of record.

IX. **Data base:** A data base is a collection of the occurrences of multiple record types, containing the relationship between records, data aggregates, and data items.

X. **Data base system:** Collection of data bases. A data base system should be a repository of the data needed for an organization's data processing. That should be accurate, private and protected from damage. It should be organized so that diverse applications with different data requirements can employ the data. Different views of data must be derived from a common overall data structure. Their methods of accessing or searching the data will differ.

XI. **Data model:** Data model is representation of collection of conceptual tools for describing data, data relationships, data semantics, operations that are performed on data and data constraints. It describes the idea of a complex "real-world" data structure. A data model contains structural, manipulative and data integrity parts. Structural part defines data types, relationships and constraints that hold on data.

XII. **Schema or global logical database description:** A schema is a chart of entire logical database. This is an overall view of the data seen by database administrator. It gives names to entities and attributes, and specifies relationship between them. It is a framework into which values of data items can be filled. Schema is specified during database design and is not expected to change frequently. A displayed schema is called a schema diagram.

XIII. **Subschema:** This refers to an application programmer's view of data he/she uses. This is a portion of the schema which is oriented to the needs of one or more application programs. Many different subschemas can be derived from one schema.

XIV. **Physical database description:** This is concerned with the physical representation, layout and organization of data on storage units. It is concerned with indices, pointers, chains, and other means of physically locating records and with overflow areas and techniques used for inserting and deleting records.

8.6 Entities and Attributes

I. **Entity:** The item about which information is stored is known as entity. It is a "thing" in the real world with an independent existence. An entity may be a tangible object with physical existence like a car, house or employee, student, etc. It may be intangible such as an event, bank account, MS-windows, abstract concepts, etc.

II. **Entity set:** Collection of similar entities that share same properties. For example students of a class B.Tech. (DT) III Year.

III. **Weak entity set:** An entity set that does not have a primary key is referred to as a week entity set. It represented by double outlined box.

IV. **Strong entity set:** An entity set with a primary key is referred to as a strong entity set.

V. **Attributes:** Properties of an entity (e.g. student) such as name, age, class, marks obtained, and address of a student are referred as attributes. Attributes may be atomic or composite. Composite attributes can be divided into smaller subparts which represent more basic attributes with independent meaning. For example, the Address can be divided further into Street, City, State, Zip, etc. Attributes that are not divisible are called as simple or atomic attributes.

VI. **Flat files:** This is an arrangement of attributes and the values of these attributes for each instances of an entity in columns and rows. For example attributes of a student (student is an entity) like name, age, class, marks can be place in columns and values of the attributes for each student are placed row vise. Related set of values of all attributes in one row for one instance of an entity is referred as tuple.

VII. **Primary key:** An attribute which uniquely identify a record from a given set of values of an entity is referred as primary key. For example a student can be uniquely identified by roll number in a class. Some important properties required for a key field are:

- Unique identification: The value of the key should be able to identify a record in a relation uniquely.
- Non redundancy: No attribute in the key can be discarded without destroying the property of unique identification.
- Numeric or text
- Meaningless

VIII. **Candidate key:** A relation in which there is more than one attributes possessing the unique identification property. All these keys are suitable candidate for primary key, therefore called as

candidate key. For example in relation SOCIETY, societies can also be identified by ***Society-Name***instead of ***Society#.***

IX. **Foreign key:** An attribute is called foreign key in a relation if it is a primary key of some other relation. For example an attribute ***Supervisor-ID*** is an ordinary attribute in relation SOCIETY but it is a primary key of relation EMPLOYEE. Thus primary and foreign keys provide a means of representing relationships between tuples.

X. **Alternate key:** Out of all candidate keys only one can be chosen as primary key, and rest all candidate keys are an alternative to primary key, hence such keys are called alternate key.

XI. **Composite key:** Some time a single attribute is not able to uniquely identify a record in a relation, therefore a combination of more than one attributes is used to uniquely identify records. This combination of key is called as composite key. For example in MILKRECEIPT relation three attributes namely ***Receipt-Date + Society# + Milk-Type*** in combination are user to identify a record uniquely instead of single attribute.

XII. **Secondary key:** An attribute which identify a set of records which have certain property. Such attribute is referred as secondary key. For example name of a student.

XIII. **Entity relationship diagram:** This diagram shows the relationship between different entity and their attributes in a database in graphical way.

8.7 Data Base Management System

A database management system consists of a collection of computer programs to manage, store, creation, organization, updation, access of database and processing of queries to produce desired results from database. The primary goal of DBMS is to provide an environment for retrieving and storing information. DBMS serves as an interface between users and database. A DBMS is a system software package that helps to integrate collection of data records and files known as databases and allows different user application programs to easily access the same database. It provides facilities for controlling data access, enforcing data integrity, managing concurrency, and restoring the database from backups. These systems provide a series of privileges and rights to a particular user. So, DBMS can be designated as a system which controls various functions of file management as well as they provide various administration rights to the users.

A DBMS also provides the ability to logically present database information to users and features for application development. There are different types of DBMS ranging from small systems that run on PCs to large mainframe systems. A few important DBMS used in industry are Oracle, Microsoft Access, SQL Server, MySQL, Ingres, Sybase, dbase, FoxPro, FileMaker and DB2. In simple terms a DBMS is the system in which related data is stored in an efficient and compact manner. "Efficient" means that the stored data can be accessed quickly and "compact" means that the data takes up very little space in the computer's memory. The phrase "related data" means that the stored data pertains to a particular topic.

8.8 Components of DBMS

- **DBMS engine:** It accepts logical requests from various other DBMS subsystems / components, converts them into physical equivalents, and actually accesses the database and data dictionary as they exist on a storage device.

- **Data definition language (DDL):** It helps the user to create and maintain the data dictionary and define the structure of the files in a database. A data base schema is specified by set of definitions expressed by special language called DDL.
- **Data manipulation language (DML):** Used to add, change, and delete information in a database and query it for valuable information. The language enables a user to access or manipulate data as organized by appropriate data model.
- **DML compiler:** It translates DML statements in a query language into low-level instruction that the query evaluation engine can understand.
- **Data manager:** It is responsible for converting user queries into physical file system. It maintains data integrity, consistency and security.
- **File manager:** It take care of file space and file structure. It demands for the block of data from disk. It allows creating new files, storing a record into file, retrieving a record from file, deleting a record from file etc.
- **Disk manager:** It transfer the block requested by user command to data manager. It transfers the data from main memory to disk.
- **Query processor:** This is used to interpret any online query and convert it into a series of efficient operations that are carried out one by one by the data manager.
- **Telecommunication system:** This is generally a remote system or an online input device being used by user to send and receive messages and data.
- **Data files:** These can store all the data portions of the database. It transfer data requested by user from data files to main memory.
- **Data dictionary:** It is data about data known as Meta data. It contains information about entities and attributes, their limits checks & validation. Data dictionary stores description of database users and their responsibilities.
- **Report generator:** Extracts information from one or more files and presents the information in a specified format. Most report generator (or writer) allows to select records that meet certain conditions and to display selected fields in rows and columns. Data can be formatted into pie charts, bar charts, and other diagrams.

8.9 Advantages of DBMS

- **Systems integration:** In DBMS, all files are integrated into one system thus reducing redundancies and making data management more efficient. In addition, DBMS provides centralized control of the operational data.
- **Controlled redundancy:** The redundancy is controlled i.e. redundancy is minimized.
- **Easy data retrieval:** Retrieval of data is very easy and efficient. "Efficient" means that the stored data can be accessed quickly.
- **Compactness of data:** Data stored in DBMS takes very little space in the computer's memory.

- **Restricting unauthorized access:** A DBMS provides a security and authorization subsystem, by creating user accounts and to specify restrictions on user accounts before accessing the data. When multiple users share a database, it is likely that some users will not be authorized to access all information in the database. For example, financial data is often considered confidential and hence only authorized persons are allowed to access such data.
- **Persistent storage:** It provides persistent storage for program objects and data structure.
- **Multiple user interface:** It provides multiple user interfaces for the users of varying level of technical knowledge.
- **Enforcing integrity constraints:** The DBMS enforces certain integrity constraints that hold on data.
- **Faster development of applications:** It provides feature like 4GL for development of new applications.
- **Represent complex relationship:** It represent complex relationship of data in very and easy and efficient manner for retrieval and updation.
- **Providing backup and recovery:** A DBMS must provide facilities for recovering from hardware or software failures. The backup and recovery subsystem of the DBMS is responsible for recovery.

****** ☺ ******

Lesson 9

DATABASE STRUCTURES

9.1 Introduction

Students will learn about the concept of data association, entity relationship diagram, Hierarchical (tree), Network (plex), and relational database models. These topics will be useful for students to understand the concept natural association among the data values and different types of database models.

9.2 Data Associations

Data items by themselves do not convey any meaning. For example, a given set of value of dairy plants and another of products supplied by the plants as in the following table 9.1, does not communicate anything:

Table 9.1 Dairy plants and products supplied

Dairy Plants	Product Supplied
Mother Dairy, Delhi	Dairy Whitener
NDRI, Karnal	Ghee
Parag, Gaziabad	Ice-cream
Nestle, Moga	Tonned Milk

However, if it is informed that the dairy plant, Mother dairy manufactures Ice-cream, NDRI manufactures Ghee, Nestle manufactures dairy whitener etc., immediately conveys some information. This information is available because a data association has been established between the values of Dairy plants and the product supplied by them. Data association is a correspondence or mapping between the members of two sets, e.g. the correspondence, in above table between Dairy plants and products supplied is given by writing them on the same line. The data association between the data items is also known as relationship or mapping. Schema and subschema are maps showing the data item types and relationships between them. The data association exists at two levels in database 1) between two or more data items i.e. attributes and 2) between two or more records (entities) / tables. There are various ways of drawing the associations.

The **association between two data items (attributes)** can be of two types as described below:

9.2.1 One-to-one (1 : 1)

In this type of association there is one-to-one mapping from data item A to data item B. This means that each value of A has one and only one value of B associated with it. For example Product name and

its code have one to one relationship, an employee ID number, and his GPF number have 1:1 correspondence. This type of relationship is shown in fig. 9.1 by single arrow.

Fig. 9.1 One-to-one data association

9.2.2 One-to-many (1 : M)

In this type of association there is one-to-many mapping from data item A to data item B. This means that one value of A has one or more values of B associated with it. For example on a particular route there may be more than two or more societies for collection of milk, parent and child relationship have 1:M correspondence. This relationship is shown in fig. 9.2 by double arrow.

Fig. 9.2 One-to-many data association

A one-to-one association is also referred as simple and one-to-many association as complex mapping. Mapping between two data items is always in both directions thus there may be four combination of associations in forward and reverse directions such as **1:1, 1:M, M:1, M:M**. In M:M type of relationship there is many-to-many mapping from data item A to data item B and vice versa. This means that one value of A has one or more value of B associated with it and similarly one value of B has one or more value of A associated with it. For example one product may be packed in different packaging sizes and one packaging size may be used for packing different products, a vendor may supply many products as well as a product may be supplied by many vendors. This relationship is shown in fig. 9.3 by double arrow at both ends.

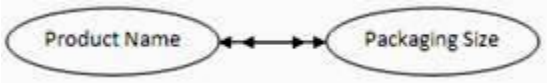


Fig. 9.3 Many-to-many data association

In addition to these associations there is another type of association which is distinguished as conditional mapping or relationship from A to B. This means that each value of A may have one or zero values of B associated with it. For example milk received in the plant may or may not be sanctioned for manufacturing sweats on particular day.

The **association between two records (entities)** can be of same types as described above say 1:1, 1:M, M:1, M:M and conditional relationship. As discussed in previous chapter a record is a collection of related attributes. Record is uniquely identified by an attribute known as primary key. For example the details of milk can be grouped as one record containing date of milk receipt, society number, route number, time of milk collection, type of milk, quantity of milk, fat%, SNF%, pH, acidity, etc. Graphically a record is drawn as an elongated box with number of attributes. Primary key of the record is distinguished from other attributes by underlining it. Primary key may be a combination of more than one attributes. Milk details record is identified by a key which is a combination of attributes i.e. ReceiptDate + SocietyNo + MilkTime + MilkType. Similarly details of routes, societies, different

sections in a dairy plant, details of milk manufactured, sale details of milk products etc. may be grouped in different records. The relationship between society and milk details is 1:M since one society may have many records in milk details record (one record for each time of milking) and may be depicted as shown in fig. 9.4 given below:

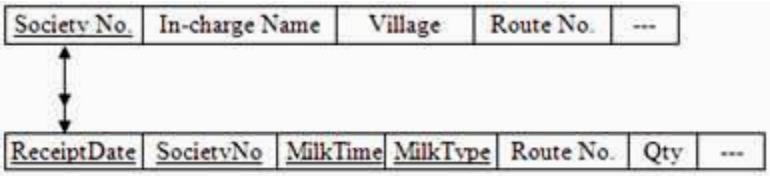


Fig. 9.4 One-to-many data association between records

9.2.3 Procedure for drawing a schema using data association

I. Collect all data items (attributes) and group them into records clearly.
II. Each attribute is given a short name as well as record for reference.
III. The diagram should clearly distinguish between the names of data items and records
IV. Record name is written outside on top and attributes name inside the box.
V. Duplicate name should not be used.
VI. Primary key should be made clear.
VII. Where secondary keys are an important part of the schema, these should be made clear (for example by drawing double arrow headed line from secondary key to primary key of the record)
VIII. Data association should be clearly depicted in the diagram using arrow lines.
IX. If associations between records are given names these should be on the diagram.

Example-1: Consider the HRD department name of a dairy plant where all details pertaining to employees of the plants are being maintained, e.g., Employee ID, name, date of birth, salary, department, skill, number of children etc. All this information may be grouped into different records and the relationship between records as shown in figure 9.5 given below. Relationship from EMPLOYEE to SKILL and CHILD is one sided i.e. only downward link 1:M is shown not the upward link, i.e., not from SKILL to EMPLOYEE and CHILD to EMPLOYEE since for these, this path is not intended to beSused.

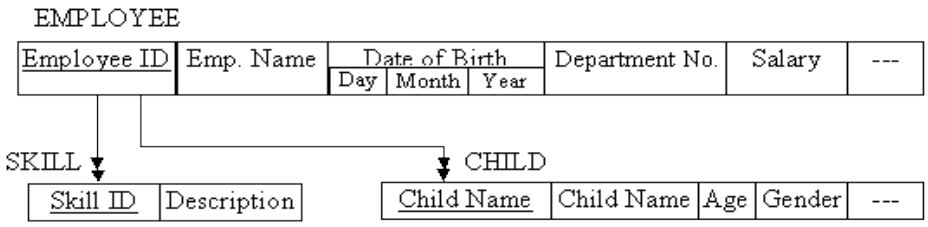


Fig. 9.5 One of drawing schema for employees

Example-2: Consider another example of purchasing items in the store of a plant. Generally a list of items being used in the plant is maintained as well as list of suppliers. Whenever an item is to purchased quotations are invited from the supplier for the item to be purchased. The purchase order is placed with the supplier who is fulfilling the quality constraints and has quoted the lowest rate. Schema for this kind of problem is shown in fig. 9.6.

Example-3: Consider the schema described in fig. 9.5. For users working in account section and users working in HRD and training section may have different sub schemas as per their own requirement as shown in fig. 9.7 (a) and 9.7 (b) respectively given below.

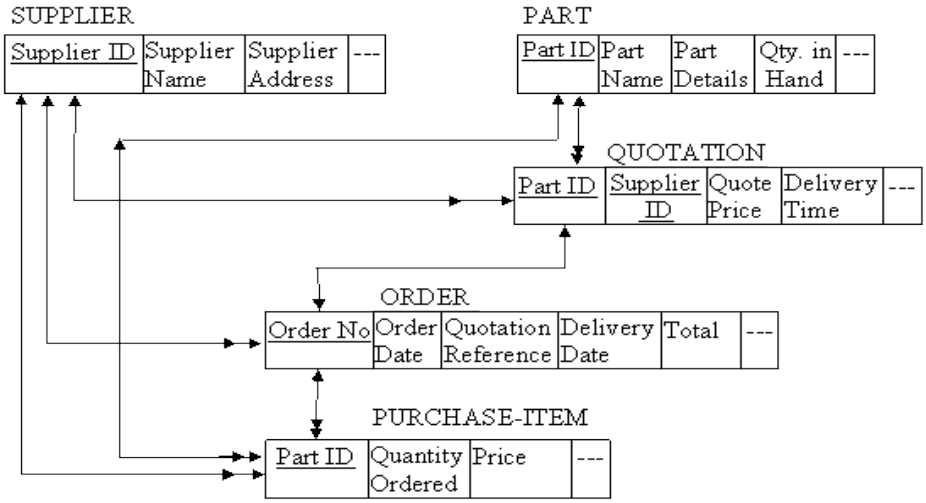


Fig. 9.6 Schema for purchase orders

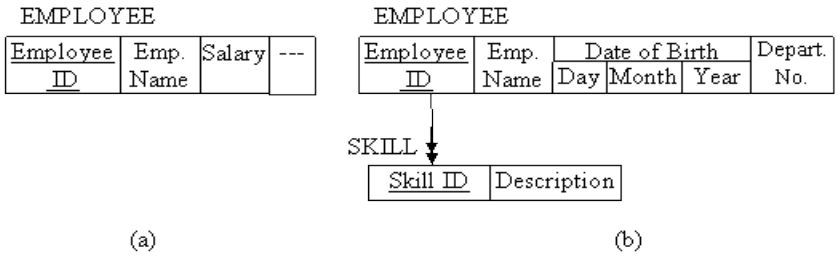


Fig. 9.7 Subschema for (a) account section and (b) HRD and training section

9.3 Entity relation model and diagram

Entity-Relationship (ER) model is a conceptual data model that views the real world as entities and relationship. A basic component of the model is Entity Relationship diagram which is used to visually represent data objects. This model was originally introduced by Peter in 1976 as a way to unify the network and relational database views. Since its origin the model has been extended and is being commonly used for database design by database administrators. It maps the relationship among data items very simply and naturally. ER model can be used as a design plan by database developers to implement a data model in a specific DBMS. The constructs used in ER diagram can easily be transformed into database structure.

9.3.1 Basic components of ER diagram

The ER diagram views the real world situation in form of entities and association between entities. So, let us first understand few important components used in ER diagram:

I. ***Entities:*** Entities are usually recognizable concepts either concrete or abstract such as persons, places, things, or events which have relevance to the database. Some specific examples of entities are **employees, milk products, manufacturing losses, milk societies** etc. An entity set is a set of entities of same type that share the same properties, e.g., set of all employees, dairy plants etc.

II. ***Independent entities:*** The entities that do not rely on others for identification.

III. ***Dependent entities:*** The entities that rely on others for identification.

IV. ***Associative entities:*** These entities are used to associate two or more entities in order to reconcile a M:M relationship and also known as intersection entities.

V. ***Relationships:*** Relationships represent an association between two or more entities. These provide the structure needed to draw information from multiple entities.

VI. ***Degree of a relationship:*** The number of entities associated with the relationship. The n-ary is the general form of representing the relationship of degree n e.g. binary -between 2 entities, ternary – among 3 entities etc.

VII. **Connectivity:** It describes the mapping of associated entities instances in the relationship. There are three types of connectivity 1. One-to-One, 2. One-to-Many, 3. Many-to-Many.

VIII. ***Cardinality:*** Cardinality of a relationship is the number of related occurrences for each of the two entities.

9.3.2 Symbols used in ER diagram

Information in ER diagrams is represented with the help of symbols. Three basic symbols used are Entity, attribute and Relationship. Other are derived from these symbols. There is no standard for representing data objects in ER diagrams. Each modeling methodology uses its own notation. HoweverScommonly used symbols or notations for drawing ER diagram are described in fig. 9.8 given below.

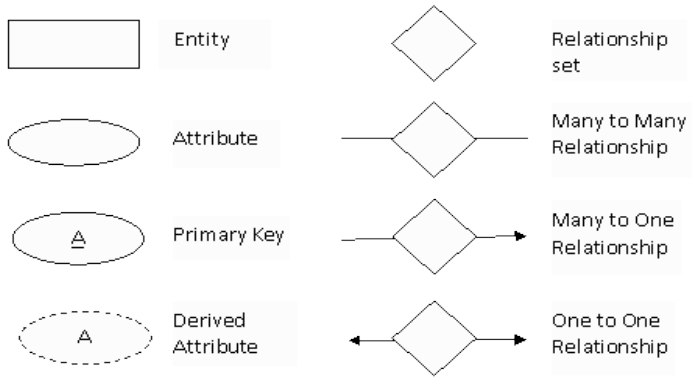


Fig. 9.8 Symbols used in ER diagram

9.3.3 Example for developing ER diagram

Developing an ER diagram requires an understanding of the system and its components. Before discussing the procedure, let's look at the example narrated as follows:

Consider the milk procurement activity in brief:

- Milk procurement area is identified.
- **Societies** are created in each village. These societies are managed by villagers.
- Milk producers/ **Farmers** become ***members*** of the societies.
- **Quality of milk** is checked at the time of taking milk from the members. Members ***contribute milk*** in morning and evening time.
- Optimized **routes** are designed by plant to ***collect milk*** from societies. **Vehicle** is sent to all societies to collect milk.
- Milk is ***delivered*** to the dockyard of the dairy plant at both times or one time as per the policy of plant. Again **milk quality** is checked at the dockyard.
- **Payment** is ***made to*** farmers based on the quality of milk as per the policy of plant (weekly/ fortnightly/ 10 Days etc.)

9.3.4 Steps for drawing ER diagram

Following steps are involved in planning and drawing the ER diagram:

I. **Define Entities:** these are usually nouns used in descriptions of the system, in the discussion of business rules, or in documentation; identified in the narrative (see highlighted/ bold items above).

II. **Define Relationships:** these are usually verbs used in descriptions of the system or in discussion of the business rules (entity to entity); identified in the narrative (see bold italic items above).

III. **Add attributes to the relations and entities**: these are determined by the queries, and may also suggest new entities. List of attributes are decided based on the purpose of creation of database or the problem to be solved.

IV. **Add cardinality to the relations:** Many-to-Many relationship must be resolved to one-to-many relationship with an additional entity called associative entities. Usually this happens automatically but sometimes involves introduction of a link entity (which will be all foreign key) Examples: Route-Vehicle information.

For the above problem ER diagram may be drawn as shown in fig. 9.9 given below.

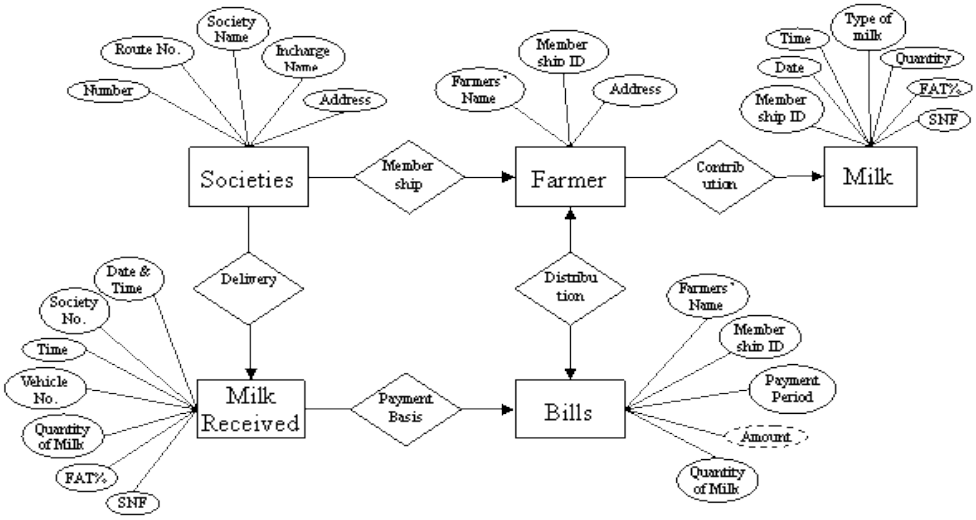


Fig. 9.9 ER diagram for milk procurement activity

9.4 Database Models

A database model provides theoretical foundation for developing a database and basically determines how data can be stored, organized, retrieved, updated and manipulated in a database system. It represents the collection of conceptual tools for describing data, relationships, semantics, operations that are performed on data and data constraints. It describes the idea of a complex "real-world" data structure. A database model contains structural, manipulative and data integrity parts. Based on the organization and relationship of data items related to different entities in real world, database models may be grouped as follows:

- **Record-based models:** In this model, real world entities are represented in form records. Each record has a fixed number of attributes (also known as fields) and each field has a fixed data type and size usually for easy implementation. Thus record-based database is structured in fixed format records of various types. The three most widely accepted models are hierarchical (tree), network (plex) and relational models.
- **Object-based models:** These are the models that are most popular in designing the database. These models involve object oriented concepts for designing the database.
- **Physical models:** Physical database models describe data at lowest level for storage on devices. Very few models are available for this purpose, e.g., Unifying model, Frame memory etc.

9.5 Hierarchical Model

Hierarchical data model organizes data in a tree-like structure. There is a hierarchy of parent and child relationship. All attributes and instances of an entity are listed in form of a table (or record type or entity type) in database. Columns of table are known as fields and rows representing instances of entity are called as records. Hierarchical model uses Parent-Child relationship to create links between these tables (or record type) i.e. 1: M mapping between record types. Hierarchical structures were widely used in the first mainframe database management systems. However owing to their

restrictions, they often cannot be used to relate structure that exists in real world. Hierarchical relationship between different types of data can make it very easy to answer some questions, but very difficult to answer others. Hierarchical DBMSs were popular from the late 1960s to 1970s with the introduction of IBM's Information Management System (IMS) DBMS.

Hierarchical model is based on tree structure. A tree is composed of a hierarchy of elements called node, the uppermost level of the hierarchy has only one node, called the **root**. Except the root node, every node has one node related to it at a higher level called as **parent** node. No element can have more than one parent. Each element can have one or more elements related to it at a lower level. These are called **children**. Elements at the end of the branches are called **leaves**. A tree can be defined as a hierarchy of nodes with binodal relationships such that:

- The highest level in the hierarchy has one node called a **root**.
- All nodes except the root are only one node on a higher level than themselves.

Fig. 9.10 shows a tree structure with 19 elements (nodes) where element 1 is root node and elements 5, 14 and 19 as leaves. Trees are normally drawn upside down with root at top and leaves at bottom. There are four levels in this tree, level 1 represent root, level 2, 3 and 4 represent branches. Trees are used in both logical and physical data descriptions. In logical description these are used to describe relations between tables/ entity type/ record type while in physical description these are used to describe sets of pointers and relations between physical segments of data values.

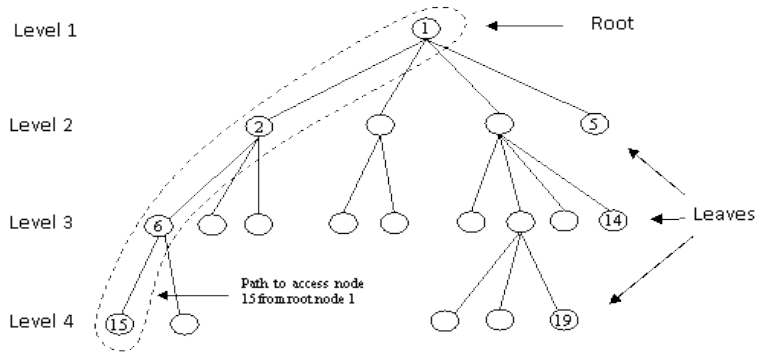


Fig. 9.10 A tree structure

9.5.1 Hierarchical files

This refers to a file with a tree structure relationship between the records. Data tend to break down into hierarchical categories. One category of data is a subset of another. A record may have multiple records subordinate to it, which in turn may have multiple records subordinate to them. In other words, multiple records of a particular type belong to (or subordinate to) a single record of another type higher in the hierarchy. For example in inventory of finished milk products the relationship between Master record type of product and transaction record type of product (receipt and issue) have hierarchical model of two level as shown in fig. 9.11 given below:

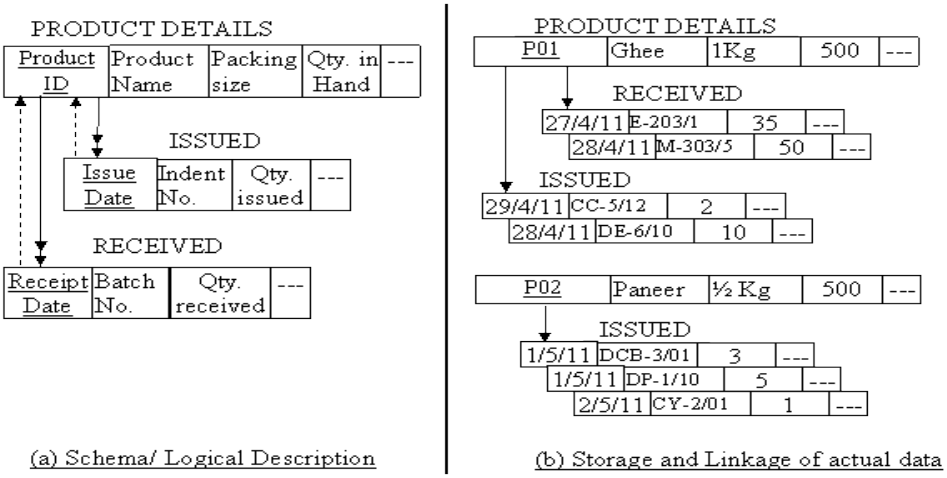


Fig. 9.11 Hierarchical Files at two levels with three record type

In the above fig 9.11(a), master record type named as PRODUCT DETAILS (root node) contains the description of milk products. There are two transaction records namely ISSUED and RECEIVED (branches) at lower level. You can see that in transaction record type ISSUED and RECEIVED, an attribute *Product ID* is not included since this information is drawn from the parent record type i.e. PRODUCT DETAILS. In figure 9.11(b), actual data is linked between the root node and branch node. For root *Product ID* **P01** there are two branches of RECEIVED record type and two branches of ISSUED record type. Another root Product ID P02 has three branches of ISSUED record type.

9.5.2 Path dependency

The lower record in hierarchy (tree structure) may be incomplete in meaning without their parent. In the example discussed above (fig. 9.11 (a)), entries made in record types ISSUED and RECEIVED are incomplete without having a link to which *Product ID* these entries belong. Therefore, in the schema dotted lines are drawn from these record types to master record type PRODUCT DETAILS to show the path dependency.

9.6 Network Model

Hierarchical model allows one parent and one or more children relationship. Some data items and record types are more naturally modeled with more than one parent per child relationship i.e. a given data item or record type may have any number of superiors as well as any number of subordinates. Network model permits the modeling of many-to-many relationships in data. Basic data modeling construct in the network model is set construct. A set consists of an owner record type, a set name, and a member record type. A member record type can have that role in more than one set, hence the multi-parent concept is supported. An owner record type can also be a member or owner in another set. Data model is a simple network, and link and intersection record types may exist, as well as sets between them. In network structure any item or node may be linked to any other item. The network data structure looks like a tree structure, except that a dependent node called a child node may have more than one parent or owner node. So, one or more nodes may have multi-parents. Therefore a network model allows a more natural modeling of relationship between entities. There is no superior

or subordinate relationship in network model as exists in hierarchical models. This kind of structure is also known as ***Plex structure***. Fig. 9.12 given below shows some examples of network or plex structure.

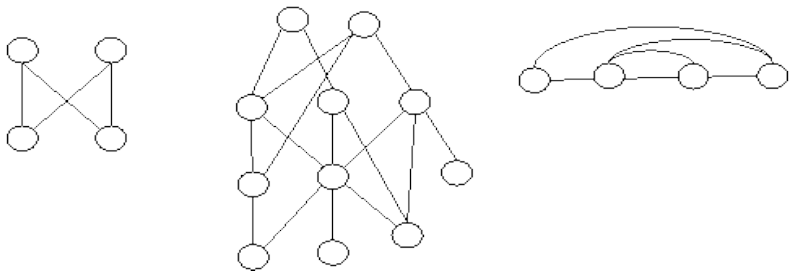


Fig. 9.12 Examples of network (plex) structure

9.6.1 Examples of M:M relationship

In real world there exist many instances including in a dairy plant where the relationship between data items or record type is many to many, for example:

- Relationship between the dairy pant and milk products, a particular plant may manufacture more than one milk product and a particular milk product may be manufactured by more than one plant for example see fig. 9.13 given below.
- Relationship between type of milk and milk products, a particular type of milk may be used in manufacturing different type of milk products and a particular milk product may be manufactured by using different type of milk.
- Relationship between machines and parts, a particular machine uses many parts and one particular part is used in many machines.
- Relationship between supplier and items, a particular supplier may supply many items and one particular item may be supplied by many suppliers.
- Relationship between course and students, a particular course may be offered by many students and a particular student may take more than one course.

The complexity of model increases with more than two entities are related with M:M relationship included in the database.

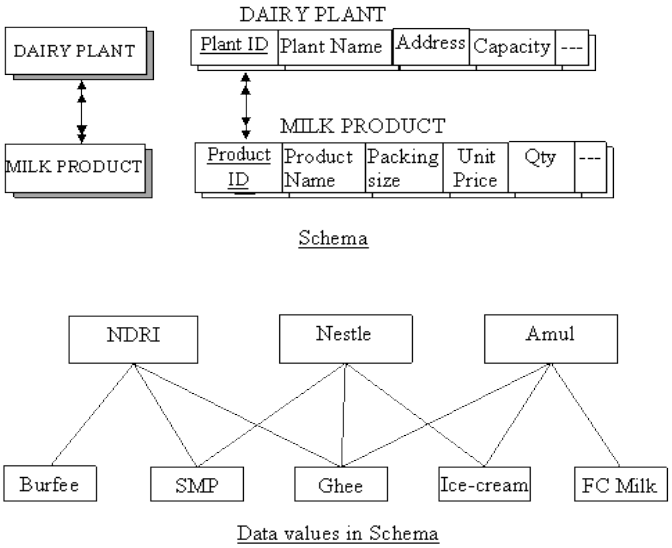


Fig. 9.13 Many to many relationships between dairy plant and milk products

9.7 Relational Model

In previous lesson two database models namely hierarchical and network were discussed. These models are structured, visualized and maintained with the help of links and pointers to show the relationship among data items or record types. These models have some advantages and disadvantages, general problem is that as the size of database grows the problem of maintaining database increases many fold. By using the links and pointers, database systems become too cumbersome, inflexible, problematic and unmanageable. The logical linkages tend to multiply as new applications are added or a user request for a new kind of query. The whole model looks like a cobweb. In 1970s, Dr. E.F. Codd, computer scientist, introduced the concept of normalization and relational model to avoid the messing of data items in hierarchical and network models. Basically, the relational models are related to user's view of the data or logical description of data instead of physical representation of data. The physical representation and the hardware can be changed (if needed) without changing the logical/ user's view. These concepts provided a great flexibility in designing, maintaining, updating, adding new applications and querying the database by users. Because of simplicity of management and understanding of data in relational models, relational database became very popular. Worldwide almost all databases are built up based on relational model. Relational databases are currently the predominant choice in storing data like financial records, medical records, personal information and manufacturing and logistical data.

The most natural way of recording data is in two dimensional tables. Generally at work places, data is recorded in registers and is arranged in columns and rows i.e. in tabular form. User is familiar with this style and can understand, remember and visualize data in two dimensional tables while in hierarchical or network structure relationship among data items seems to be natural but difficult to record and visualize. Therefore, practically network structure is broken down into hierarchical structure for better management of data and hierarchical structure is broken down into two dimensional tables for recording of data. A table is referred as relation in relational model. These relations form the base of relational database model.

A relation based database consists of a set of multiple tables. Each table is a relation and so a relational database can be thought of as a collection of tables. A table usually represents an object and information about that object. Objects may be physical objects or logical concepts. In each table, the rows (called tuples) represent unique entities or records and columns (called fields) represent attributes. Data values are stored at the intersection of row and column. Each named column has a domain, which is the set of values that may appear in that column. Each record in a table contains the same set of fields. Relationships are represented by common fields or attributes in different tables. Relational data structure is a mathematical model defined in terms of predicate logic and set theory. It is based on a formal theory of relational algebra. The fundamental assumption of the relational model is that all data are represented as mathematical ***n-ary*** relations; an *n-ary* relation being a subset of the Cartesian product of *n* sets. Relations have no specific order in relational database.

The relational model allows defining data structure, storage and retrieval operations and integrity constraints. Certain fields are defined as keys to index the records in a relation for fast searching and arranging the records in a particular order on key fields without creating duplicate copy of entire table. This improves the speed of searching records for specific values of that key field. Applications access data by specifying queries, which use operations such as ***select*** to identify records, ***project*** to identify attributes, and ***join*** to combine relations. Relations can be modified using the ***insert***, ***delete***, and ***update*** operators. New records can supply explicit values or be derived from a query. Similarly, queries identify records for updating or deleting. It is necessary for each record of a relation to be uniquely identifiable by some combination (one or more) of its attribute values. This combination is referred to as the primary key.

Relations that store data in a relational database are called "base relations". Some relations are computed by applying relational operations on base relations. Such relations are called as "derived relations". In implementations these are called "**views**" or "**queries**". Views can retrieve information from several relations and act as a single relation.

9.7.1 Example of relational database model

Consider a simple case of milk procurement application where milk is being collected through number of milk societies established in milk procurement area. Each society may supply CM (Cow Milk) or BM (Buffalo Milk) or both type of milk. Each society is being supervised by a staff of dairy plant. One supervisor may supervise more than one society. In this example, there are three entities or relations namely MILKRECEIPT, SOCIETY AND EMPLOYEE. Structures of these relations are described as follows:

MILKRECEIPT(<u>Receipt-Date, Society#, Milk-Type</u>, Quantity)

SOCIETY(<u>Society#</u>, Society-Name, Supervisor-ID, Society-Address)

EMPLOYEE (<u>Supervisor-ID</u>, Supervisor-Name, Designation, Salary)

Underlined attributes are the primary key of these relations. These relations are linked in such a way that societies supply milk to dairy and dairy plant staff supervises the societies. The relationship between MILKRECEIPT and SOCIETY is an attribute *Society#* and relationship between SOCIETY AND

EMPLOYEE is an attribute *Supervisor-ID.* A sample of relational database consisting of these relations (tables) with some data values is shown in fig. 9.14 as given below:

MILKRECEIPT

Receipt-Date	Society#	Milk-Type	Quantity
1/4/11	101	CM	1000
1/4/11	101	BM	2000
1/4/11	102	CM	1500
1/4/11	103	BM	2500
1/4/11	104	CM	2000
1/4/11	105	CM	1700
1/4/11	105	BM	2200
2/4/11	101	CM	1200
2/4/11	101	BM	1800
2/4/11	102	CM	1300
2/4/11	102	BM	1200
2/4/11	103	BM	2200
2/4/11	104	CM	1900
2/4/11	105	CM	1750
2/4/11	105	BM	2100
3/4/11	101	CM	1300
3/4/11	101	BM	1700
3/4/11	102	CM	1400
3/4/11	102	BM	1300
3/4/11	104	BM	2100
3/4/11	104	CM	2000
3/4/11	105	CM	1820
3/4/11	105	BM	2200

SOCIETY

Society#	Society-Name	Supervisor-ID	Society-Address
101	Baldi Milk Society	30011	Baldi Village
102	Dadupur Milk Society	30011	Dadupur Village
103	Phusgarh Milk Society	30012	Phusgarh Village
104	Goghripur Milk Society	30013	Goghripur Village
105	Ballha Milk Society	30013	Ballha Village
106	Kachwa Milk Society	30015	Kachwa Village

EMPLOYEE

Supervisor-ID	Supervisor-Name	Designation	Salary
30011	Ajay Singh	Procurement Officer	50000
30012	Timir Kumar	Asst. Procurement Officer	35000
30013	Manoj Sharma	Procurement Officer	52000

Fig. 9.14 A view of relational database for milk procurement application

9.7.2 Properties of Relational Model

A relation in relational model has following properties:

I. **Values are atomic:** This implies that columns do not have repeating groups and cannot be broken down further. It simplifies data manipulation.

II. **Each row is unique:** This property ensures that each row is distinct from others. Two rows in a relational table are not identical. Rows are uniquely identified by an attribute or field known as primary key. Primary key may be a singly column (field) or set of multiple columns. This property guarantees that every row in a relational table is meaningful and that a specific row can be identified and retrieved by specifying the value of primary key.

III. **Each column has a unique name:** Columns represent the attributes/ characteristics of an object or entity. Each characteristic describes some different property of an object so it should be given a unique name to store data about these characteristics otherwise it may create ambiguity in table. Data in a relational table is retrieved by column names not by position of columns therefore duplicate name will create confusion in the same table. In general, a column name need not be unique within an entire database but only within the table to which it belongs. Moreover in relational table the column size is also fixed.

IV. **Column values are homogeneous:** This means that values in a column must be of the same kind. All values in column come from the same domain. For example type of milk should have values out of CM (cow milk) or BM (buffalo milk) if a dairy plant accept only these two type of milk. GM (goat milk) would be a wrong data value in this column. This property simplifies the data access because developers and users can be certain of the type of data contained in a given column. It also ensures validation of data.

V. **The sequence or order of columns is insignificant:** This property states that the ordering of the columns in relational table has no meaning. Columns are retrieved by name so they can be in any order and in various sequences. This property enables many users to share the same table without concern of how the table is organized. It also permits the physical structure of database to change without affecting the relational tables.

VI. **The sequence or order of rows is insignificant:** This property is analogous to the above property. To retrieve relevant rows or records based on some condition, each row is searched in entire table to meet the specified condition so the order of rows does not matter. Adding new rows in a relational table (at any position, i.e., top, bottom, in between, etc.) does not affect existing queries.

9.7.3 Relational Databases Terminology

I. **Domain:** It is used to organize and describe an attribute's set of possible values. It is pool of values from which the actual values appearing in a column are drawn. For example values appearing in ***Society#*** column of both the MILKRECEIPT and SOCIETY tables are drawn from the underlying domain of all valid society numbers.

II. **Relationship:** An association between entities.

III. **Degree of a relation:** No. of columns in a relation.

IV. **Degree of a relationship:** This is number of entities associated with the relationship. The n-ary relationship is the general form for degree n. Special cases are the binary, and ternary, where the degree is 2 and 3 respectively.

9.7.4 Integrity Rules

Data in a database must adhere to a set of pre-defined rules defined by an enterprise/ organization to maintain consistency in database. For example the fat% of cow milk must be in range 3% to 7%. If there is deviation then error may be either in recording the data or milk may be adulterated. The set of defined rules are called as integrity rules or constraints. In relational model there are two integrity rules as described below:

9.7.5 Entity integrity

No component of primary key value may be NULL. Primary keys perform the unique identification function in a relational database. A identifier that is wholly NULL would be a contradicting in terms, in effect, it would be saying that there was some entity that did not have any unique identification and if two entities are not identifiable from each other, then by definition these are not two, but only one entity exists.

9.7.6 Referential integrity

Let D be a domain of an attribute 'A' in a relation R1, then at any given time, each value of A in R1 must be either NULL or from the given domain. If A is a foreign key than it must be either NULL or any value from the set of values defined in other relation where A is primary key of that relation.

Lesson 10

RELATIONAL DATABASES

10.1 Introduction

In this lesson students will learn about relational database operator for manipulating data in relational databases, and normalization techniques for designing relational databases. These topics will be useful for students to design conceptual model for development of stable database to meet the present and future requirements for an organization.

10.2 Operations on Relations

The concept of relational database is based on the theory of relational algebra therefore various types of algebraic operations can be applied on relations (tables) in relational database. Let us consider the example of relational databases discussed in previous lesson for understanding the functioning of various operations. A few commonly used operations are described as follows:

10.2.1 INSERT

This operator is used to insert records in tables. Some examples of insert operation are:

- **Insert-1:** Insert a record for new society established in a village e.g. **INSERT INTO** *table_name* [column list] **VALUES** (values)
  Action: Record can be easily inserted in relation SOCIETY, even if the society is not started contributing milk to dairy plant. e.g. **INSERT INTO*SOCIETY*** [column list] **VALUES** (values)
- **Insert-2:** Insert a record for receipt of milk in MILKRECEIPT relation.
  Action: Record can be easily inserted in relation MILKRECEIPT, but must exist in SOCIETY relation for maintaining data integrity. For example milk received from ***Society#110*** does not make any sense since this society number does not exist in SOCIETY relation.

10.2.2 DELETE

This operator is used to remove existing records (or rows) from tables. For example:

- **Delete:** Suppose ***Society# 103*** in SOCIETY relation has broken relations with dairy plant and stopped supplying of milk. So, this record can be deleted from the database by giving command e.g. **DELETE FROM** table_name [**WHERE** condition].
  Action: This record can be deleted easily from SOCIETY relation the information of this society will be lost. **DELETE FROM** SOCIETY**WHERE** society# = 103.

10.2.3 UPDATE

UPDATE allows writing queries to change data in the existing records. For example:

- **Update:** Change the address of the ***Society# 102*** in SOCIETY relation.

Action: This record can be updated easily in SOCIETY relation since one change is required at one place only. **UPDATE** SOCIETY **SET**society-address = 'new address' **WHERE** society# = 102.

10.2.4 SELECTION

The SELECT operation retrieves records/ rows that satisfy the given conditions. A temporary view is created which can be saved in a new table or relation. This is the most commonly used operator in relational database model. It is represented by lowercase Greek letter sigma σ. For example:

- **Select:** Execute select operation to retrieve records/ rows of cow milk only (i.e. for milk_type equal to 'CM') from relation MILKRECEIPT
  SELECT Receipt-Date, Society#, Milk-Type, Quantity **FROM** MILKRECEIPT **WHERE** Milk-Type = 'CM'.
  Action: Output of this Select operation is shown in fig. 10.1 given below. The output includes the records those have satisfied the condition from the original relation (fig. 9.14). Sequence of records may be original as in table

MILK

Receipt-Date	Society#	Milk-Type	Quantity
1/4/11	101	CM	1000
1/4/11	102	CM	1500
1/4/11	104	CM	2000
1/4/11	105	CM	1700
2/4/11	101	CM	1200
2/4/11	102	CM	1300
2/4/11	104	CM	1900
2/4/11	105	CM	1750
3/4/11	101	CM	1300
3/4/11	102	CM	1400
3/4/11	104	CM	2000
3/4/11	105	CM	1820

Fig. 10.1 Example of selection operation

10.2.5 PROJECTION

This operator enables to select specified columns from a relation/ table as per requirement to create a new relation/table. The order of columns may be changed. This operator is generally represented by lowercase Greek letter pi 'π'. In SQL it is implemented through SELECT command. For example, consider the projection

MILK = π MILKRECEIPT(Society#, Milk-Type) of relation MILKRECEIPT.

- **Projection:** Create a new relation with fields Society#, Milk-Type from relation MILKRECEIPT. For this purpose execute the following SQL statement
  SELECT Society#, Milk-Type **FROM** MILKRECEIPT

Action: Output of this projection operator is shown in fig. 10.2 given below (from original relation in fig. 9.14). The output does not have any duplicate record and the sequence of records may be as in original table.

MILK

Society#	Milk-Type
101	CM
101	BM
102	CM
103	BM
104	CM
105	CM
105	BM
102	BM
104	BM

Fig. 10.2 Example of projection operation

10.2.6 JOIN

The Join operation is performed to create new table (view) from two existing tables/ relations when they share a common data item. Sometimes to generate reports or queries it may be required to extract columns from two or more than two tables. These tables must be sharing common attributes. When tables are joined on a given attribute, only those records will appear in the output which shares the same value of that attribute. Therefore output may have less number of records than either of the original tables. This is called natural join. There are other types of join such as inner join, outer join etc. The Projection operator splits tables while Join operation puts together columns from different relations. Join operator is represented by '∗'. In SQL it is implemented through SELECT command. For example, consider the Join operation on MILKRECEIPT and SOCIETY relations, MILK-SOCIETY = MILKRECEIPT ∗ SOCIETY (Receipt-Date, Society#, Society-Name, Milk-Type, Quantity) for Milk-Type = 'CM'.

- **Join:** Create a new relation with fields Receipt-Date, Society#, Society-Name, Milk-Type, Quantity from relations MILKRECEIPT and SOCIETY for cow milk only. For this purpose execute the following SQL statement
  SELECT receipt-date, society#, society-name, milk-type, quantity **FROM** MILKRECEIPT, SOCIETY **WHERE** milk-type = 'CM'
  Action: Output of this Select operation is shown in fig. 10.3 given below. Output includes the records those have satisfied the condition. Sequence of records may be original as in first table.

MILK-SOCIETY

Receipt-Date	Society#	Society-Name	Milk-Type	Quantity
1/4/11	101	Baldi Milk Society	CM	1000
1/4/11	102	Dadupur Milk Society	CM	1500
1/4/11	104	Goghripur Milk Society	CM	2000
1/4/11	105	Ballha Milk Society	CM	1700

2/4/11	101	Baldi Milk Society	CM	1200
2/4/11	102	Dadupur Milk Society	CM	1300
2/4/11	104	Goghripur Milk Society	CM	1900
2/4/11	105	Ballha Milk Society	CM	1750
3/4/11	101	Baldi Milk Society	CM	1300
3/4/11	102	Dadupur Milk Society	CM	1400
3/4/11	104	Goghripur Milk Society	CM	2000
3/4/11	105	Ballha Milk Society	CM	1820

Fig. 10.3 Example of Join operation on MILKRECEIPT and SOCIETY

Similarly the result of join operation on SOCIETY and EMPLOYEE relation is shown in fig. 10.4 given below:

SOC-EMP = SOCIETY∗ EMPLOYEE (Society#, Supervisor-ID, Supervisor-Name, Salary)

SELECT society#, supervisor-ID, supervisor-Name, salary **FROM** SOCIETY, EMPLOYEE

SOC-EMP

Society#	Supervisor-ID	Supervisor-Name	Salary
101	30011	Ajay Singh	50000
103	30012	Timir Kumar	35000
104	30013	Manoj Sharma	52000

Fig. 10.4 Example of Join operation on MILKRECEIPT and SOCIETY

10.2.7 Union

This operator combines the records of two relations and removes all duplicate records from the result.

10.2.8 Intersection

This operator produces a set of common records from two relations.

10.2.9 Cartesian product

Cartesian product of two relations is a join that is not restricted by any criteria, resulting in every record of the first relation being matched with every record of the second relation.

Many updates and queries are accomplished by combining multiple tables in various ways, thus providing an extremely powerful access capability. For example, MILKRECEIPT and SOCIETY relations can be joined along with select and project operations. It is evident from the foregoing discussion that relational approach is conceptually straight forward in design as compared to the other two approaches. In addition, the underlying theory is elegant and precise allowing for complex natural relations to be represented clearly.

In a hierarchical or network model the connections and relationships are in the data structure. If a new relationship is to be added, new connections and access paths must be established. In a relational database, access paths need not be pre-determined. Creating new relations simply requires a joining of tables. Relational databases are therefore the most flexible and useful for unplanned and ad-hoc queries. The pre-determined relationships of the hierarchical or the network structures require more complex

data definition language (DDL) and data manipulation language (DML). Maintenance is more difficult. The relational model data definition and data manipulation languages are simple and user oriented. Maintenance and physical storage are fairly simple.

10.3 Conversion of Hierarchical Model (Tree Structure) into Relational Model

A hierarchical model (tree structure) can be easily converted into relational model with minimal redundancy. For example, consider a tree structure as given in fig. 9.11, PRODUCT DETAIL record type is having 1:M relationship with ISSUED and RECEIVED record type. Each record in ISSUED and RECEIVED relation is identified by a link from an attribute ***Product ID*** in PRODUCT DETAIL i.e. for every record in ISSUED and RECEIVED there is path dependency from PRODUCT DETAIL. To convert this trees structure into relational model, path dependency is to be removed by introducing one more attribute ***Product ID*** in relations ISSUED and RECEIVED. There will tree relations in the relational database model as shown in fig. 10.5.

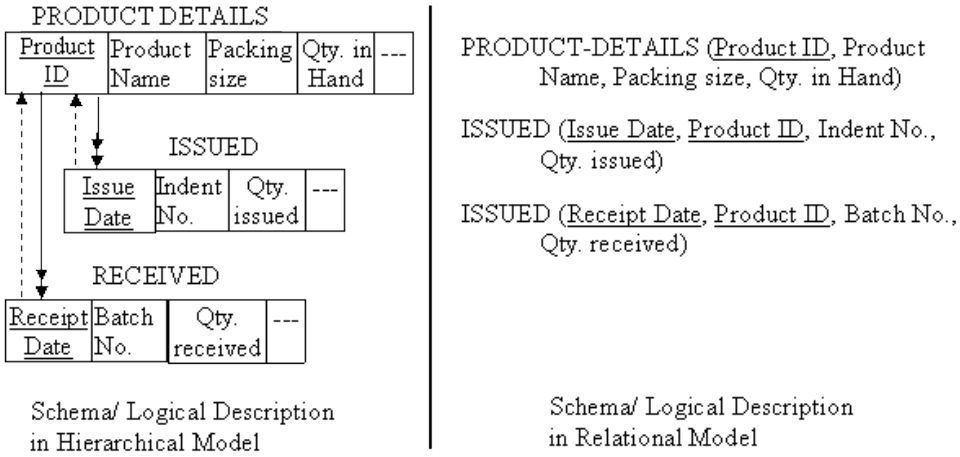
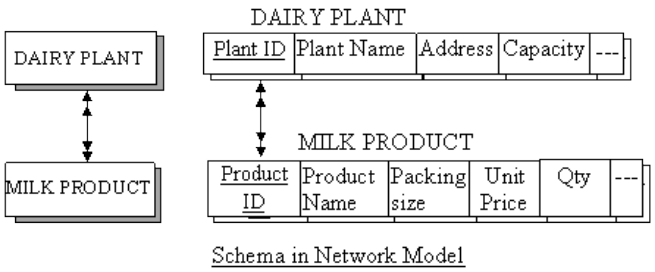


Fig. 10.5 Conversion of hierarchical model into relational model

10.4 Conversion of Network Model (Plex Structure) into Relational Model

A network model (plex structure) can be converted as easily as hierarchical model into relational model. Consider an example of network model shown fig. 9.13 and as discussed in previous lesson, there are two record types namely DAIRY PLANT and MILK PRODUCT. These records are linked to each other with M:M relationship and association between them is ***Unit Price*** of milk product fixed by different diary plants. This attribute is recognized by two items e.g. Product ID and Plant ID. To convert this structure into relational model, a new record type is created to store the association of existing two record types as shown in fig. 10.6.



Schema in Network Model

DAIRY-PLANT (Plant ID, Plant Name, Address, Capacity)

MILK-PRODUCT (Product ID, Product Name, Packing size)

PRODUCT-PLANT (Product ID, Plant ID, Unit Price, Qty)

Schema in Relational Model

Fig. 10.6 Conversion of network model into relational model

10.5 Advantages of Relational Database Model

- **Easy to use and understand:** It very easy to use and understand than other models. It is easy to add information in the form of table.
- **Data independence:** In comparison with Hierarchical or Network model, it is easy to achieve data independence in relational model.
- **Flexible:** Relational model is flexible to add new data items or new tables/ relations etc.
- **Security:** Data security measures can be implemented easily.
- **Ad-hoc query:** Ad-hoc unforeseen queries can be generated and process easily.

10.6 Disadvantages of Relational Database Model

- **Complex:** It is complex to design and develop and then creating and maintaining relationships among relations.
- **Requires more memory:** Extra memory is required to create and store indexes for searching and retrieving data.
- **Hardware cost:** Powerful hardware is required to process data.
- **Time consuming:** Extra time is required to search a record because records are searched sequentially.

10.7 Normalization

Normalization is a design technique that is widely used as a guide in designing relational databases. Normalization is a natural process of grouping related data and then placing in different tables to minimize duplication of data. Normalization process has been defined by different researchers in their own way. Some important descriptions are as follows:

- Normalization is a process by which attributes are grouped together to form a well structured relation.
- Normalization is a process of simplifying the relationship between the data elements in a record.
- Normalization is basically a process of efficiently organizing data in a database with the objectives to eliminate redundant data (for example storing the same data in more than one table) and to ensure data dependencies make sense (only storing related data in a table). These

objectives are important as they reduce the amount of space of database and ensure that data is logically stored.

- Normalization is a step by step process for replacing association (1: M and M: M) between data items in two dimensional tabular forms (or flat-file representation) by storing data in separate files to minimize redundancy and simplify basic database file management.
- The purpose of normalization is to produce a stable set of relations that is a faithful model of the operations of an enterprise. By following the principle of normalization we can achieve a design that is highly flexible, allowing the model to be extended.
- The practice of storing data in separate two dimensional tables or flat files to minimize redundancy and simplify basic database file management is called normalization
- Normalization is a process of breaking a relation into sub relations such that the sub relations so formed are free from insert, delete, and update anomalies.

From above descriptions, finally we can say that normalization is:
"Normalization is a step by step process for replacing association (1:M and M:M) between data items in two dimensional tabular forms (i.e. flat-file representation) by storing data in separate files to minimize redundancy and simplify basic database file management. The overall aim is to produce a stable set of relations that is a faithful model of the operations of an enterprise. Normalization process involves the removing:

- Redundant data
- Partial dependencies
- Transitive dependencies

By following the principles of normalization, one can achieve a database design that is highly flexible; reliable; extendable; free from insert, delete, and update anomalies; and easy to implement".
Normalization theory is based on the concepts of normal forms. A relational table is said to be in a particular normal form if it satisfies a certain set of constraints. There are currently five normal forms termed as 1st, 2nd, 3rd, Boyce-Codd Normal Form (BCNF), 4th and 5th normal form (NF) respectively. In practice, majority of the relations become free from different kind data anomalies when they are in 3rd NF and generally, it has been observed by practitioners that entities that are in 3rd NF are also in 4th and 5th NF. So, first three forms are sufficient to design a good and stable database for an organization. Therefore for simplicity, we will discuss only first three normal forms in this lesson.
Normalization process has following goals:

I. The goal of normalization is to create a set of relational tables that are consistent and free of redundant data.
II. The normalization rules are designed to prevent update anomalies and data inconsistencies.
III. The normalized design enhances the data integrity by minimizing the redundancy and inconsistency but at cost of performance for certain retrieval application.

10.8 Important Terms Used in Normalization Process

Before understanding the normal forms, it is important to understand few key terms which forms the base of normalization concept:

10.8.1 Functional Dependencies

The concept of functional dependencies is the basis for the first three normal forms. Functional dependency describes a relationship between attributes in a single relation. An attribute is functionally dependant on another if we can use the value of one attribute to determine the value of another. So functional dependency may be defined as: "An attribute B of a relation R is functionally dependent on attribute A or R if, at every instant of time, each value in A has no more than one value in B associated with it in relation R". Or in other words it can be defined as

"A column, Y of a relational table R is said to be functionally dependent on column X of R, if and only if each value of X in R is associated with precisely one value of Y at any given time. X and Y may be composite". Saying that column Y is functionally dependent on X is the same as saying the values of column X identify the values of column Y. If column X is a primary key, then all columns in relational table R must be functionally dependent on X.

Symbol → is used to indicate a functional dependency. → is read as functionally determines. One can read X → Y as, "X determines Y". Functionality is shown graphically in fig. 10.7 given below:

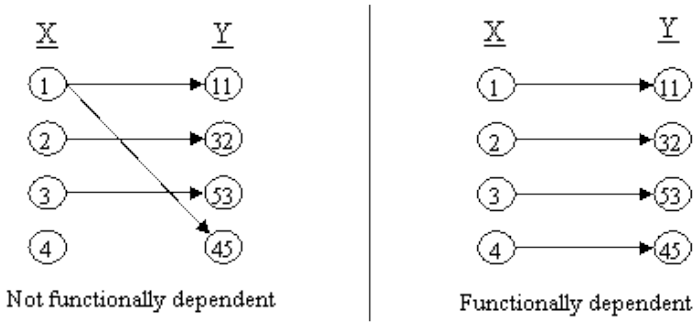


Fig. 10.7 Graphical representation of functional dependencies

Example-1: Consider a relation of containing employee details working in different projects running in a dairy plant.

EMPLOYEE (Employee-ID, Employee-Name, Qualification, Phone-Number, Salary, Project-ID, Completion-Date)

The functional dependencies in this relation are as follows in fig. 10.8:

Attributes	Functional Dependency	Attribute name	Symbolic Trems
*Employee -ID	Dependent on	Employee-Name	Employee-ID → Employee-Name
*Employee-Name	Dependent on	Employee-ID	Employee-Name → Employee-ID
Qualification	Dependent on	Employee-ID	Employee-ID → Qualification
Phone-Number	Dependent on	Employee-ID	Employee-ID → Phone-Number
Salary	Dependent on	Employee-ID or Employee- Name	Employee-ID → Salary (but reverse is not true)
Project-ID	Dependent on	Employee-ID or Employee- Name	Employee-ID →Employee-Name → Project-ID (but reverse is not true)

Completion-Date	Dependent on	Employee-ID or Employee-Name or Project-ID	Employee-ID → Completion-Date Employee-Name → Completion-Date Project-ID → Completion-Date (but reverse is not true)

Fig. 10.8 Description of functional dependencies for relation EMPLOYEE

Employee-Name is functionally dependant on Employee-ID because Employee-ID can be used to determine the value of Employee-Name and one Employee-ID will be related to only one name (Employee-ID → Employee-Name). Employee-ID is not functionally dependent on Salary because more than one employee could have the same salary. Similarly, Employee-ID is not functionally dependent on Project-ID or Completion-Date. Completion-Date is functionally dependent on Project-ID only. No other attributes in this relation is fully dependent on project-ID except Completion-Date. The asterisks indicate prime attributes (member for candidate key)

Example-2: An attribute can be functionally dependent on group of attributes rather on a single attribute. Consider a relation of employees where a particular employee is working in more than one section of a dairy plant.

EMP-SECTION (Emp-ID, Section-ID, Emp-Name, Section-Name, Total-Work-Hrs)

The functional dependencies in this relation are as follows in fig. 10.9:

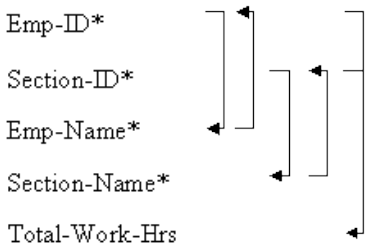


Fig. 10.9 Description of functional dependencies for relation EMP-SECTION

Here Total-Work-Hrs is functionally dependant on Emp-ID and Section-ID. Either Emp-ID or Section-ID individually is not competent to identify the value of Total-Work-Hrs. The asterisks indicate prime attributes (member for candidate key)

Example-3: Consider a relation having details of milk product in a dairy plant

PRODUCT (Product-ID, Product-Name, Packing-Size, Unit-Price, FAT%, SNF%)

The functional dependencies in this relation are as follows in fig. 10.10:

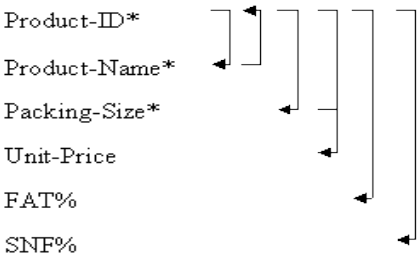


Fig. 10.10 Description of functional dependencies for relation PRODUCT

Product-Name is functionally dependant on Product -ID because Product -ID can be used to determine the value of Product -Name and one Product ID will be related to only one product (Product-ID → Product-Name). Unit price is dependent on Product-ID as well as packing size. The asterisks indicate prime attributes (member for candidate key)

Example-4: Consider a relation having details of supply source of various machine parts used in a dairy plant. These parts are supplied by different suppliers.

SUPPLY-SOURCE (Supplier-ID, Part-ID, Supplier-Name, Supplier-Address, Price)

The functional dependencies in this relation are as follows in fig. 10.11:

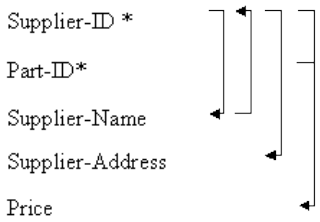


Fig. 10.11 Description of functional dependencies for relation PRODUCT

Here Price is functionally dependant on Supplier-ID and Part -ID because different suppliers may supply the same part at different price. So both Supplier-ID and Part -ID attributes are used to identify the attribute Price. The asterisks indicate prime attributes (member for candidate key)

10.8.2 Full Functional Dependency

This applies to tables with composite keys and can be defined as "An attributes or a collection of attributes B of a relation R can be said to be fully functionally dependent on another collection of attributes A; of relation R if B is functionally dependent on the whole of A but not on any subset of A". Or in other words as

"Column Y in relational table R is fully functional dependent on X of R if it is functionally dependent on whole X and not functionally dependent on any subset of X. Full functional dependency means that when a primary key is composite, made of two or more columns, then the other columns must be identified by the entire key and not just some of the columns that make up the key".

For example, consider the relation EMP-SECTION given in example-2 above (fig. 10.9), an attribute ***Total-Work-Hrs*** is fully functional dependent of the combination of key (***Emp-ID*** + ***Section-ID)*** because to know that how many hours a particular worker has worked in a particular section, information from both of these key fields is required. Neither ***Emp-ID***, alone nor ***Section-ID*** alone can identify the attribute ***Total-Work-Hrs.*** Other fields such as***Emp-Name***, is fully functional dependent on ***Emp-ID*** alone and Section-Name is fully functional dependent on ***Section-ID*** alone. Similarly you see in example-3 and example-4 discussed above.

10.8.3 Partial Dependencies

Partial dependencies might occur in a relation that contains a composite primary key and some attributes are dependent on part of the primary key rather than entire key. In that situation, any data item that is not dependent on all key fields should be removed into a separate database file. For example consider the

relation SUPPLY-SOURCE as discussed in example-4 above (fig. 10.11). The primary key of this relation is combination of ***Supplier-ID*** +***Part-ID*** attributes. As seen in fig. 10.11, attributes ***Supplier-Name*** and ***Supplier-Address*** are functionally dependent on part of the key attribute i.e.***Supplier-ID*** not on the whole key. Only an attribute ***Price*** of the relation is totally dependent on the whole key i.e. ***Supplier-ID*** + ***Part-ID***. Therefore, attributes ***Supplier-Name*** and ***Supplier-Address*** must be removed from relation SUPPLY-SOURCE and stored in some other relation with key attribute ***Supplier-ID*** only.

10.8.4 Partial key

It is a set of attribute that can be uniquely identifying weak entities and that are related to same owner entity. It is sometime called as Discriminator.

10.8.5 Artificial key

If no obvious key either stands alone or composite is available, then the last resort is to simply create a key, by assigning a unique number to read each record or occurrence. Then it is known as developing an artificial key.

10.9 First normal form (1NF)

The normalization process starts with natural instinct of arranging data items into two dimensional tables by breaking up of data items into smallest possible units. Repeating data items are removed from the table and are placed in a new separate table. A key field is used to link these two partitioned tables. After that a database is said to be in first normal form (1NF). More precisely, "A relation R is said to be in first normal form (1NF) if and only if all underlying domains contain atomic values only". Thus 1NF describes the tabular format in which:

- All occurrences of a record type must contain the same number of fields.
- All attributes are atomic.
- All the key attributes are defined.
- There are no repeating groups in the table.
- All attributes are dependent on primary key.

For example consider the relation EMPLOYEE as discussed above in example-1 (figure-26.1). At first glance this relation seems to be in 1NF, if we assume that the attributes ***Employee-Name***, ***Phone-Number*** and ***Qualification*** are atomic i.e. these cannot be braked down further. Similarly if we assume that one employee may have only value for attributes ***Phone-Number*** and ***Qualification*** respectively. In real word such assumptions are not realistic. Practically, employee name may be considered as a combination of three fields namely ***First-Name, Second-Name,*** and ***Last-Name***.; employee phone number may consist of ***Country-Code, STD-Code,*** and ***Actual-Number***; and qualification may be broken into ***Exam-Passed, Year,*** and ***Marks-Obtained***. Moreover one employee can have more than one qualification and phone number. Therefore as per the above definition of 1NF this relation can be redesigned as follows:

EMPLOYEE (Employee-ID, First-Name, Middle-Name, Last-Name, Salary, Project-ID, Completion-Date)
QUALIFICATION (Employee-ID, Sr-No, Exam-Passed, Year, Marks-Obtained)

PHONES (Employee-ID, Sr-No, Country-Code, STD-Code, and Actual-Number)

In the relations QUALIFICATION and PHONES, a new dummy field namely Sr-No is added and concatenated with key field Employee-ID to identify a record uniquely in these tables.

10.10 Second normal form (2NF)

The definition of second normal form states that only the data items those are directly relevant to the entire key fields (in case of composite key) should be in relational table and other data items must be split into separate tables. This means that in 2NF each field in a table with composite primary key must be directly related to the entire key not on a part of primary key. More precisely, "A relation R is in 2NF if it is in 1NF and every non-prime attributes of R is fully functionally dependent on primary key of R".

For example consider the relation EMPLOYEE as discussed above in 1NF. The partitioned tables are already in 2NF since the primary key ***Employee-ID*** of relation EMPLOYEE consists of only one attributes, and every non-prime attributes are fully dependent on the primary key. In other relations QUALIFICATION and PHONES all non prime attributes are fully dependent on the composite primary key, i.e., ***Employee-ID*** + ***Sr-No*** respectively.

The relation EMP-SECTION in fig. 10.9, it is already in 2NF since the attribute ***Total-Work-Hrs*** is fully functionally dependant on whole primary key i.e. on the combination of ***Emp-ID*** + ***Section-ID.*** Similarly it is dependent on combination of candidate keys like ***Emp-Name, Section-Name***.

The following relation described in example-4 (fig. 10.11) is not in 2NF:

SUPPLY-SOURCE (Supplier-ID, Part-ID, Supplier-Name, Supplier-Address, Price)

The relation has a composite primary key ***Supplier-ID*** + ***Part-ID***. Attributes ***Supplier-Name*** and ***Supplier-Address*** are functionally dependent only on part of the key i.e. ***Supplier-ID*** not on whole primary key. The attribute ***Price*** is fully functional dependent on combined key. When a relation is not in 2NF then there will be few problems related to data updates in such relations as discussed below:

- Details of a supplier cannot be entered in database until that supplier supplies some parts.
- Suppose a supplier is supplying only a single part and due to some reasons he temporarily stops the supply of that part, then deletion of records for that part will also remove the information of supplier. But it would be desirable to preserve the details of supplier in an organization for future use.
- Suppose a supplier changes his address then all records in the database are to be updated for that supplier. This would be required to be done at many places (redundancy), time consuming and may lead to inconsistency in the database.

These anomalies can be solved by splitting the relation into two relations as follows:

SUPPLY-SOURCE (Supplier-ID, Supplier-Name, Supplier-Address)

SUPPLY-SOURCE (Supplier-ID, Part-ID, Price)

In these relations all non prime attributes are fully functionally dependent on primary key of the relations. Similarly the relation as given below and discussed in example-3 (fig. 10.10) is not in 2NF

PRODUCT (Product-ID, Product-Name, Packing-Size, Unit-Price, FAT%, SNF%)

This relation can be brought into 2NF by splitting into two relations as follows:

PRODUCT (Product-ID, Product-Name, FAT%, SNF %)
PRODUCT-PACKING (Product-ID, Packing-Size, Unit-Price)

10.11 Third Normal Form (3NF)

The problems related to data updates as discussed in above section can occasionally occur even if a relation is in second normal form. To remove such anomalies database relation is further refined by using the third normalization step. The third normal form requires that all columns in a relational table should be dependent only on the primary key. A more formal definition is: "A relation R is in third normal form (3NF) if it is already in 2NF and every non-key attributes of R is non-transitively dependent on each candidate key of R". When all the transitive dependencies have been removed from the relation; the database is said to be in the third normal form.

10.11.1 Transitive dependency

A transitive dependency occurs when non-key attribute that is a determinant of primary key is also the determinant of other attribute. This means that transitive dependency in a relation occurs due to those attributes that are occasionally (though not always) dependent on some other non key field in the same relation.

In symbolic term it can be defined as "suppose that A, B and C are three attributes or distinct collections of attributes of a relation R, If C is functionally dependent on B and B is functionally dependent on A, then C is functionally dependent on A. If the inverse mapping is non-simple (i.e., if A is not functionally dependent on B or B is not functionally dependent on C), then C is said to be transitively dependent on A. The transitive dependency may be removed by splitting the relation into two parts as shown in fig. 10.12 given below:

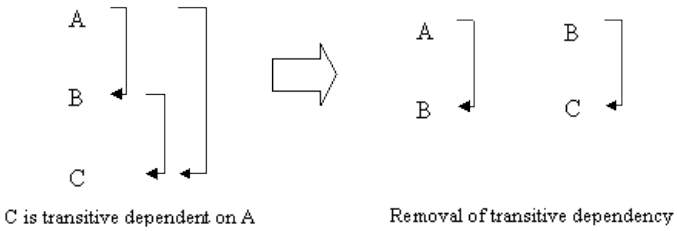


Fig. 10.12 Representation of transitive dependency and its removal

For example consider the relation EMPLOYEE as discussed above in example-1 (fig. 10.8). This relation was further converted into three relations after passing through the first normalization form as given below:

- EMPLOYEE (Employee-ID, First-Name, Middle-Name, Last-Name, Salary, Project-ID, Completion-Date)
- QUALIFICATION (Employee-ID, Sr-No, Exam-Passed, Year, Marks-Obtained)
- PHONES (Employee-ID, Sr-No, Country-Code, STD-Code, and Actual-Number)

Automatically this relation is also in 2NF as it fulfills the requirement of second normalization form. But there is a problem when the conditions of 3NF are applied on these relations especially in relation EMPLOYEE. In this relation, attribute ***Project-ID*** is functionally dependent on ***Employee-ID*** and attribute***Completion-Date*** is functionally dependent on ***Project-ID*** which is a non key attribute. Further on analyzing the reverse relationship it is observed that is***Employee-ID*** in not functional dependent

on ***Project-ID*** similarly ***Project-ID*** is not functional dependent on ***Completion-Date.*** It means the inverse mapping is non-simple. Therefore, ***Completion-Date*** is transitively dependent on ***Employee-ID***. To remove the transitive dependency the relation EMPLOYEE may be split into two relations as follows:

- EMPLOYEE (Employee-ID, First-Name, Middle-Name, Last-Name, Salary, Project-ID)
- PROJECT (Project-ID, Completion-Date)

Thus after going through 3 phases of normalization process i.e. 1NF, 2NF and 3NF the original relation would be converted into four relations as given below:

- EMPLOYEE (Employee-ID, First-Name, Middle-Name, Last-Name, Salary, Project-ID)
- QUALIFICATION (Employee-ID, Sr-No, Exam-Passed, Year, Marks-Obtained)
- PHONES (Employee-ID, Sr-No, Country-Code, STD-Code, and Actual-Number)
- PROJECT (Project-ID, Completion-Date)

When a relation is not in 3NF then there will be few problems related to data updates in such relations as discussed below:

- Unless an employee is hired in a project, the project detail can be entered in database. This means that for a new project if employee is not recruited, the details of the project like completion date cannot be entered because this information is a part of the employee relation.
- If all employees working in a project left the job then their records will also be deleted from the database. The deletion of records may lead to the loss of an important information about the project i.e. completion date because this information is a part of the employee relation.
- Suppose a particular project is extended by another one year more, then this change has to be incorporated in all relevant records. This may lead to inconsistency in the database.

After removing transitive dependency the original relation is converted into two new relations. In each relation the attributes are fully functional dependent on the prime attribute of the relation. These converted relations will not have the data anomalies as discussed above. Thus, the database emerging out after passing through the three steps of normalization will be flexible; reliable; extendable; free from insert, delete, and update anomalies; and easy to implement.

****** ☺ ******

Lesson 11

COMMUNICATION PRINCIPLES

11.1 Introduction

This lesson will introduce the need and concept of communication systems. Students will learn about the capabilities, elements and limitations of communication systems, representation of signals, analog and digital signals. These topics will be useful for students to develop an understanding about the basics of communication process and intricacies involved in transmitting data.

Communication is the transfer of data/ information from source to destination. Basically through communication information/ views are being shared among interested user groups. Communication has always been the requirement of human being since the origin. The sharing can be local or remote. Local communication takes place between two persons face to face while remote communication takes place over distance. In today's dynamic world the tools and methods of communication are changing very fast. Computer networks, Internet and World Wide Web (WWW) are the most commonly used communication tools to interact people at global level. These tools have changed the life style of people and business environment. The development of computers had brought tremendous changes for business, industry, science and education. Computers gave the new processing power to the scientific and business world but still an efficient way of sharing information was missing. Development of computer networks during 1970s proved the revolutionary point in communication era. This laid the foundation of foot prints of Internet. Finally Internet became a commercial success in the 1990s and now a day it is household name. Internet is one of the most interesting and exciting phenomena of computer networking. Internet was the key technology of the 20thcentury and will also play a lead role in 21st century.

The success of Internet is based on the technological innovations made in the networking of computers and the communication facilities. Reduction in the prices of computing and networking devices has also played a significant role in bringing Internet and computing devices to common man. Networking of computers was only the matter of interest of defense and academic personnel twenty years ago. But latter on it gain popularity among the business houses and now it has become the requirement of each and every household. Computer networks especially the Internet has changed the lifestyle of a common man. Networks and Internet is being used in all sphere of life for example research, business, banking transactions, governance, military, advertisements, etc. therefore, it is beyond the imagination to survive without Internet in the new century. Communication networks are being used for diversified applications. Various applications and use of network can be broadly grouped in the following groups:

- **Access to remote information:** Applications under this group involves interaction between a user and a remote database such as file sharing, payment of bills, banking and financial

services, investments, shopping, electronic data interchange (EDI), news papers, information services, marketing and sales, WWW, etc.

- **Person to person communication:** Applications of this group are email, chat, voice mail, voice conferencing, video conferencing, newsgroups, cellular phone etc.
- **Interactive entertainment:** Applications in this group are cable TV network, games playing, 3-dimensional real time images, photographic quality moving images etc.

11.2 Communication process

Communication process is the way of conveying data or messages (voice/ video/ numbers/ text etc.) between two points and this information or messages may be verbal or non-verbal. Data refers to facts, concepts and instructions presented in whatever the form is agreed upon by the concerned parties involved in creating and using data. Data is represented by binary information units or bits in the form of 0 and 1. Data communication can be defined as exchange of data (in the form of 0's and 1's) electronically between two devices via some form of transmission medium (wire cable or wireless). The effectiveness of data communication depends on: Delivery, Accuracy, and Timeliness. A communication process requires a sender, a message, receiver, communication channel or medium through which information travels between two points and protocols that govern the communication. The communication process is completed once the receiver completely understood the message sent by sender.

11.3 Communication Principles

Operations of any communication system are governed by the following principles:

I. Infrastructure to transfer information from one point to another point. All services supported by the network are designed to facilitate the exchange of information.

II. The receiver must understand the message. Without understanding, no meaningful communication takes place.

Any communication system is useful and effective when a system is able to accept source data; structure the source data so that it can be sent quickly and accurately; transmit the data to specific destination; able to detect the error during transmission (if any); and once the data is received, reconvert it to a form understandable by the destination.

11.4 Components (Elements) of Communication

A data communication system is made up of the following components (see figure 11.1):

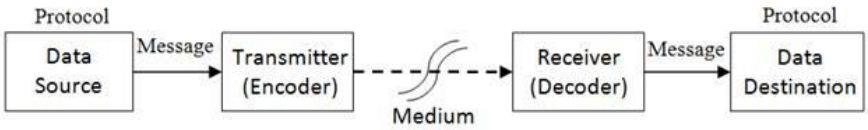


Fig. 11.1 Components of a data communication system

- **Data source:** A communication system serves to communicate data between two locations. A data source generates the data to be transmitted.

- **Message:** A message is the information (data) to be communicated. A message can contain text, numbers, voice, video, images or any combination of these.
- **Transmitter:** The main function of the transmitter is to convert the message obtained from the data source into electrical signal with the help of transducer if it is not in the electrical form. It convert low frequency signal to high frequency signal for long distance transmission.
- **Medium or channel:** Channel means the medium through which message signals travels from transmitter to receiver. In other words, the function of the channel is to provide a physical connection (wired or wireless) between the sender and destination.
- **Receiver:** Accept signal from transmission system and converts (reproduce) it into a form that can be handled by destination. This reproduction is accomplished by a process known as demodulation. Demodulation is the process of extracting back the original signal from the modulated signal.
- **Destination**: Destination is the final stage where an electrical signal is converted into original form for further use.
- **Protocols:** A protocol is a set of rules that govern all aspects of information communication. A protocol defines what is communicated, how it is communicated, and when it is communicated.

11.5 Limitation on Communication System

A communication system consists of a number of components and devices. Therefore, one may face some constraints or limitations mainly due to imperfection of transmission medium used in communication system. Some of the important limitations are:

- Transmission impairments
- Limited bandwidth

11.5.1 Transmission impairments

Generally transmission mediums are not perfect and this cause impairment in the signal sent through the medium. This means the signal received at the end is not the same as sent originally. There are three types of impairments in the signals:

I. ***Attenuation:*** This is loss of energy. After travelling through the transmission medium a signal loses some of its energy to overcome the resistance of the medium. To compensate this loss, signal is amplified using amplifiers.

II. ***Distortion:*** Distortion means that the signal changes its form or shape. Generally distortion occurs in composite signal made up of different frequencies. Each signal component has its own propagation speed through a medium and therefore, its own delay in arriving at the final destination. At the destination, received signal is found different from the original signal.

III. ***Noise:*** Noise is an unwanted signal which interferes with the transmitted signals in a communication system. Noise limits the ability to identify the desired message correctly. Noise is random in nature it cannot be eliminated completely but its effect can be minimized using several techniques. Several type of noise may corrupt the signal. Few are described below:

- **Thermal noise:** It is the random motion of electrons in a wire that creates an extra signal not originally sent by the transmitter.
- **Crosstalk:** It is effect of one wire on the other. One wire acts as sending antenna and other as the receiving antenna.
- **Impulse noise:** It is a spike (a signal with high energy in a very short period of time) that comes from power lines, lightening etc.
- **Induces noise:** It comes from sources such as motors and appliances. These devices act as sending antenna and the transmission medium acts as the receiving antenna.

11.5.2 Limited bandwidth

Bandwidth for a particular transmission medium is allocated by International regulatory agency to avoid interference among the signals having same frequency. But, for a given transmission medium this bandwidth may not be sufficient to transmit entire information. Thus puts a major impact on communication system.

11.6 Representation of signal

In general term, anything which carries information is a signal. Signals may be verbal or nonverbal. Daily we receive number of signals and process them to extract information for taking appropriate action. For example human voice, vocalization of cattle, chirping of birds, lighting, body gestures, ring tones etc. can be treated as signals since they convey something.

In electronic communication, a signal is an electric current or electromagnetic field used to convey data trough transmission medium from source to destination. Data can be any message that is readable and understandable to the sender and receiver such as voice, image, text, numbers, or code. As such this information (say image) cannot be transmitted over a transmission medium. This image has to be encoded in digital form using 0 and 1 bit. This encoding must be understandable to the destination to decode the data. Further, even the string of 0s and 1s cannot be sent across links because transmission media as such do understand 0 and 1. These codes must be converted in to a form that transmission media can accept. Transmission media work by conducting energy along a physical path. Therefore the string of 0s and 1s must be converted into energy in the form of electrical signals.

11.6.1 Analog

Analog refers to something that is continuous - a set of specific points of data and all possible points in between. Analog data refers to continuous data e.g. human voice – when someone speaks a continuous wave is created in the air, atmospheric temperature, real numbers etc. Analog describes any fluctuating, evolving, or continually changing process.

11.6.2 Digital

Digital refers to something that is discrete - a set of specific points of data and no other points in between. Digital data refers to discrete data e.g. data stored in computer memory, integer numbers, etc.

11.6.3 Signal

An electromagnetic wave that is propagated along a transmission media. Data is transmitted through signals in a media. A signal can take either analog or digital form. Both analog and digital signals can be of two forms periodic and aperiodic.

Periodic Signal: A Periodic signal consists of a continuously repeated pattern. It completes a pattern within a measurable time frame called a period and repeats the identical pattern over subsequent periods as shown in figure 11.2. The completion of full pattern is called as cycle.

Aperiodic signal: An aperiodic signal has no repetitive pattern. It changes constantly without exhibiting a pattern or cycle that repeats over time.

11.6.4 Analog signal

Analog signal refers to a continuous wave form which changes smoothly over a time. As the wave moves from value A to value B, it passes through and includes an infinite number of values along its path. In other word, an analog signal will have all possible (infinites) points within a given set of specific points of data. An analog signal can be represented as a series of sine waves. Analog signal can be classified as simple or composite. A simple analog signal or a sine wave cannot be decomposed into simpler signals. A composite analog signal is composed of multiple sine waves.

11.6.5 Simple analog signal

A simple analog signal is visualized as a simple oscillating curve, which is continuous, smooth and consistent. A simple analog signal can be represented by a sine wave (composed of single frequency). Three characteristics: amplitude, period or frequency, and phase describing a sine wave are as shown in figure 11.2 given below.

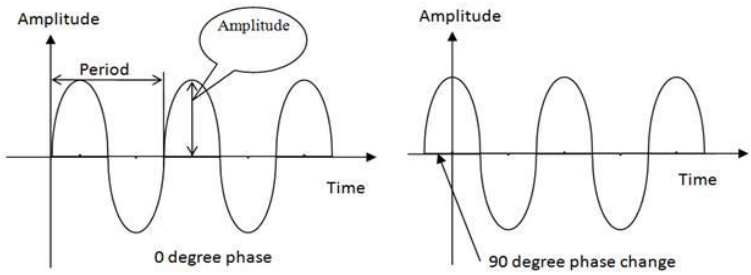


Fig. 11.2 A sine wave with different phases

11.6.6 Amplitude

It is a value of the signal at any point of time on the wave. It is equal to the vertical distance from a given point on the wave from the x-axis. The minimum and maximum amplitude of a sine wave are same. Amplitude is measured in volts, amperes, or watts depending on the type of signals. This also represents the strength of a signal.

11.6.7 Period and frequency of a signal

Period refers to the amount of time, in seconds, a signal needs to complete one cycle. Frequency refers to the number of periods in one second i.e. number of cycles per second. In other words frequency is rate

of change with respect to time. Frequency is expressed in hertz (Hz). Relationship between period and frequency (*f*): *f = 1/ Period*. Let T be a period and *f* be frequency then *T = 1/f*.
Period is expressed in seconds:

Seconds(s)	=	1s
Milliseconds (ms)	=	10^{-3}s
Microseconds (µs)	=	10^{-6}s
Nanoseconds (ns)	=	10^{-9}s
Picoseconds (ps)	=	10^{-12}s

Example: A sine wave has a frequency of 8KHz, then its period will be

$$T = {}^{1}/_{8000} = 0.000125$$

$$= 125 * 10^{-6}\ seconds = 125\mu s$$

11.6.8 Wavelength of a signal

Wavelength (λ) refers to the distance covered in one period. It binds the period or frequency of signal to the propagation speed of the medium. Wavelength can be calculated if the propagation speed (c) and the period of the signal are given. Wavelength = C x Period. If c is propagation speed of a signal in the medium and *f* is frequency then, Wavelength can be express as

11.6.9 Composite analog signal

In real world life an electromagnetic wave is a periodic composite signal that contains more than one sine wave of different amplitude, frequency and phase. Elements of a composite signal can be decomposed into simple analog signals through Fourier analysis. A composite signal is shown in figure 11.3 where two simple signal (a) frequency 5 and amplitude 6 and (b) DC current of frequency 0 and amplitude of 9 are forming a composite signal (c). The resultant signal is similar to a sine wave but time axis is shifted downward and amplitude is 15.

11.6.10 Frequency spectrum and bandwidth of a signal

Frequency spectrum of a signal is a collection of all component frequencies it contains. Bandwidth is a path required for transmission of signals and is expressed in frequencies. Bandwidth is a width of the frequency spectrum. In other words bandwidth refers to the range of component frequencies. And frequency spectrum refers to the elements within that range.

Example: A composite periodic signal having three sine waves with frequencies 200, 500, and 800MHz then bandwidth of this signal is 800-200MHz = 600MHz

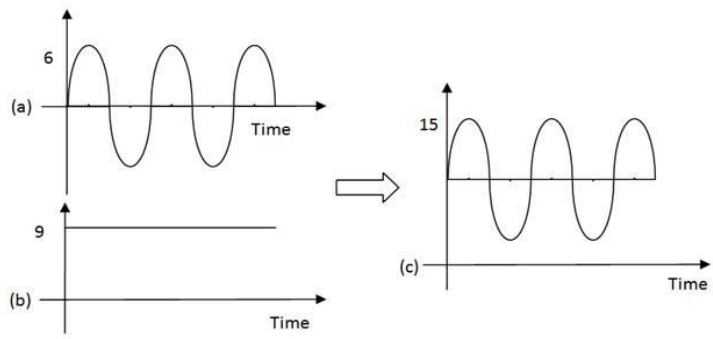


Fig. 11.3 A composite analog signal

11.6.11 Digital signal

Digital signal refers to discrete change of the wave form into pre-defined levels. Data is represented in two states as 1 and 0. In digital signal 1 can be coded as positive voltage and 0 as zero voltage. Digital signals are generally aperiodic and therefore, period and frequency is not appropriate terms for digital signals. In digital signals, the terms **bit interval** (instead of period) and **bit rate** (instead of frequency) are used to describe digital signal as shown in figure 11.4. The bit interval is the time required to send one single bit and bit rate is the number of bits sent in one second expressed in bits per second (bps).

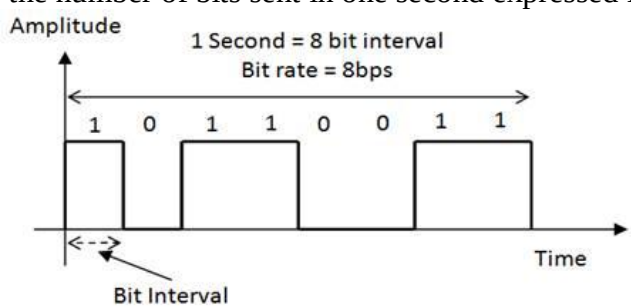


Fig. 11.4 A digital signal with bit rate and bit interval

11.7 Fourier Analysis

Mathematically, equation for a basic sine-wave with frequency f (frequency measured Hertz, i.e. cycles/second) can be defined as:

$$y(t) = A.sin(2\pi ft + \theta)$$

where: y(t) is the signal at time t
A is the maximum amplitude of the signal
f is the number of cycle per second
θ is the phase of the signal

If the phase shift is 90 degree (π/2 radians), the same signal can be expressed as cosine wave instead of sine wave:

$$y(t) = A.sin(2\pi ft + \pi/2) = A.cos(2\pi ft)$$

Example: The human voice is a composite signal i.e. summation of number of sine waves, each having different amplitude, frequency and phase. The bandwidth is normally between 300Hz and 3300Hz. The general equation for this signal will be like:

$$y(t) = A_1.sin(2\pi f_1 t + \theta_1) + A_2.sin(2\pi f_2 t + \theta_2) + \cdots + A_n.sin(2\pi f_n t + \theta_n)$$

11.7.1 Fourier series

Fourier series were introduced by Joseph Fourier (1768–1830). A Fourier series decomposes any periodic signal into the sum of a (possibly infinite) set of simple oscillating sine and cosine waves each having different frequency and phase. A periodic composite signal can be decomposed as follows:

$$x(t) = C_0 + C_1.sin(2\pi f_1 t + \theta_1) + C_2.sin(2\pi f_2 t + \theta_2) + \cdots + C_n.sin(2\pi f_n t + \theta_n) + \cdots$$

The coefficients c_0, c_1, …, c_n are the amplitude of the simple signals. c_0 is the amplitude of the signal with frequency 0 (i.e. DC component).

11.8 Fourier transformation

The Fourier transformation decompose a composite a periodic signal into the sum of a (possibly infinite) set of simple oscillating sine and cosine waves each having different frequency and phase.
Note: Mathematical explanation of Fourier series and transformation is beyond the scope of this lesson.

****** ☺ ******

Lesson 12

COMMUNICATION CHANNELS

12.1 Introduction

This lesson will introduce the need and concept of communication systems. Students will learn about Modulation and Encoding; Analog Versus Digital Communications; Electromagnetic Spectrum (Frequency Spectrum) – Frequency bands and their application areas; Communication Channels and characteristics of transmission media. These topics will be useful for students to develop an understanding about the basics of communication process and intricacies involved in transmitting data.

A signal by itself does not carry any information. The signal is modified so that it contains identifiable changes that are recognizable to the sender and receiver as representing the information intended. In computer, data is stored in form of 0s and 1s i.e. in digital form. This digital data must be converted to either into digital signal or analog signal for further transmission to destination. During transmission, four type of conversion of data may occur depending on the nature of signals as shown in the following table 12.1.

Table 12.1 Conversion schemes

Source	Converted to Signal	Type of Conversion	Purpose/ usage
Digital data	Digital	Digital to digital conversion or encoding digital data into a digital signal	For example communication between computer and printer
Analog data	Digital	Analog to digital conversion or digitizing an analog signal	To digitize human voice for sending over long distances.
Digital data	Analog	Digital to analog conversion or modulating a digital signal	Transmission of data from one computer to another through public telephone network which carry analog signals.
Analog data	Analog	Analog to analog conversion or modulating an analog signal	Transmission of voice through radio system

12.2 Modulation and encoding

Encoding is a process of transforming source information into signals. Source data may be in either digital or analog form is encoded into a digital signal using an encoding technique to optimize the usage of transmission media. Some important encoding schemes used for digital data to digital signal

conversion are Non return to zero-level, Non return to zero-Invert, Manchester, Differential Manchester, Bipolar 8-zero substitution (B8ZS), and High density bipolar 3 (HDB3).

Modulation is a process of encoding source data onto analog signal of a continuous constant high-frequency periodic waveform called as ***carrier signal***. Modulation technique involves operation of varying one or more properties of *carrier signal*, like amplitude, frequency, and phase with respect to a***modulating signal*** (*which contains information to be transmitted*). Typically a high-frequency sinusoid waveform or square wave pulse is used as carrier signal. A device that performs modulation is known as a modulator. Modulation is of two types:

12.2.1 Continuous wave modulation (Digital/ analog to analog conversion)

When carrier wave is continuous in nature then modulation process is known as continuous or analog communication, .e.g.,

- ***Amplitude modulation*:** In amplitude modulation process amplitude of carrier wave is varied in accordance to the instantaneous value of modulating wave or message signal.
- ***Frequency modulation:*** In frequency modulation process, frequency of carrier signal is varied accordance to the instantaneous value of modulating signal or message signal.
- ***Phase modulation:*** In this process phase of carrier signal is varied in accordance to the instanttaneous vlaue of modulating signal.

12.2.2 Pulse modulation (analog to digital conversion)

When carrier wave is a digital pulse train, the modulation process is known as pulse modulation. This is called analog to digital conversion or digitizing an analog signal. Examples of pulse modulations are:

Pulse code modulation (PCM): A PCM is a method to represent analog signal in digital form, in which magnitude of analog signal is sampled regularly at uniform intervals, with each sample being quantized to the nearest value within a range of digital steps.

12.3 Analog verses digital communication

I. The set up required for digital communication systems are simple and cheaper compared to analog communication systems because of the advancement made in the IC technologies.
II. In digital communication data is binary in nature; therefore a large amount of noise interference may be tolerated as compare to analog communication.
III. Since channel coding is used in digital communication that makes the error detection and correction easier at the receiver side compare to analog communication.
IV. In digital communication transmission is digital and channel encoding is used, therefore the noise does not accumulate from repeater to repeater in long distance communications.
V. Privacy can be maintained in digital communication using data encryption technique which allow only permitted receiver to detect the transmitted data.
VI. In digital communication, speech, video and other data may be merged and transmitted over a common channel using multiplexing.
VII. In digital communication data is in binary form means have a higher data rate which require more bandwidth for transmission as compare to analog communication

12.4 Frequency/ Electromagnetic spectrum

Computers and other telecommunication devices use signals to represent data. These signals are transmitted from one device to another in form of electromagnetic energy which may travel through vacuum, through air, or through other transmission media. The electromagnetic energy includes power, voice, radio waves, infrared light, visible light ultraviolet light, and X, gamma, and cosmic rays. Not all of these forms of electromagnetic energy are usable for telecommunication. Electromagnetic spectrum is the range of all possible frequencies of electromagnetic radiations from gamma rays to the longest radio waves. A complete range of frequencies and their usages in communication is given the following table -12.2:

Table 12.2 Electromagnetic spectrum

Radio band	Frequency Range (Hz)	Transmission Media	Propagation Mode	Applications
Audio Frequency	0K – 3K	Twisted pairs	Surface (air, water)	Telephone, telegraph
Very Low Frequency (VLF)	3K – 30K			Long radio navigation, submarine communication, aeronautical communication
Low Frequency (LF)	30K– 300K			
Medium Frequency (MF)	300K-3M	Coaxial cable	Line of Sight, Sky wave	AM radios, maritime radio, radio direction finding (RDF), emergency frequencies
High Frequency (HF)	3M-30M		Sky wave	Citizen's band (CB) radio, business, Civil defence, military communication, international broadcasting, amateur radio
Very High Frequency (VHF)	30M-300M		Line of Sight	VHF television, FM radio, aeronautical, mobile radio
Ultra High Frequency (UHF)	300M-3G	Wave guide		UHF television, mobile telephone, cellular radio, paging, microwave link beyond 1GHz
Super High Frequency (SHF)	3G-30G			Experimental, Terrestrial and satellite microwave communication, radar communication,
Extremely High Frequency (EHF)	30G-300G			
Infrared	300G-200T	Optical fiber	Line of sight, Laser beams	Lasers beams, Guided missiles, Range finders, wireless communication between to devices inside a lab, remote control, wireless local area networks

				(WLANs),
Visible Light	200T-1000T	Optical fiber	Laser beams	Optical fibers
Ultraviolet, X-ray, Gama rays	> 1000T		Laser Beam	Under experiments

12.5 Communication Channels

Communication channel in refers to a transmission medium to transport data from one or several senders (or transmitters) to one or several receivers. A wide variety of communication channels are used in data communication. These channels can be classified into two groups:

- **Wire-line:** Copper wires (Twisted pairs, coaxial cables) and Optical fibers.
- **Wireless:** Radio waves, Micro Waves and Infrared waves.

12.5.1 Wire-line media

Wire-line (or wired) media consists of a solid medium generally of copper metal or glass fiber. A signal in form of electromagnetic wave travels through this solid medium from source to destination. A signal traveling along any of these media is directed and contained by the physical limits of the medium. Therefore, a wired media is also called as guided media. Voice-band frequencies are generally transmitted as current over metal cables, such as twisted-pair or co-axial cable. Visible light used for communications, is harnessed using fiber-optic cable. Different types of cables used in wired media are twisted pair, coaxial cable, and fiber optic cable.

12.5.1.1 Copper wires

The primary medium to connect communicating devices in a network is copper cables because of its low resistance to electric current which means signals can travel for a longer distance. Basically two types of copper cables are used in networks – twisted pair and coaxial cable.

a. Twisted pair

A twisted pair consists of two insulated copper wires of about 1 mm thick. The pair of wires is twisted to each other. This pair of wire is treated as a single unit in a communication link. In the twisted pair cable one wire is used for signal (a voltage level or current level) and another for ground reference. These two wires may be separate and run parallel but are twisted to each other to reduces the effects of external electrical noise by using differential signaling. A number of pairs are bundled together and encased in a protective sheath.

The twisting minimizes the crosstalk interference between adjacent pairs in multi pair cable. Twisted pair cables are the most commonly used transmission media for both analog and digital communication. These cables are used to connect telephone handset with local exchange and local area network (LAN). The bandwidth depends on the quality of wire and the distance traveled. The data rate varies from 64kbps to 100mbps. Twisted pair cable is cost effective and easy to install. Twisted pair cables are available in two varieties unshielded twisted pair (UTP) and shielded twisted pair (STP).

In UTP, twisted pairs of insulated wires are grouped or placed in a flexible plastic jacket. Un-shielding the cable reduces the cost, size, and installation time of the cable and connecters. But it increases the

effect of outside the interferences from nearby twisted pair and from the noise generated in the environment. The UTP cables are available in different varieties. The CAT3, CAT5, CAT6 are the most common varieties of UTP cables. The major difference between category -3 and category -5 is the number of twists per unit distance. In category -3 there are 3-4 twists per foot while category -5 is has 3-4 twists per inch.

In STP, twisted pairs of wires are covered by a protective metal shield. The shielding reduces the effects of interference signals and provides better performance but increases the cost and are difficult to handle in comparison to UTP cables.

b. Coaxial cable

Coaxial cable consists of an inner insulated core of stranded or solid wire (generally of copper) surrounded by an outer insulated flexible wire braid. The inner wire carries the signals and outer metal braid works as conductor for ground reference. There are four parts in this cable. First the inner wire which works as inner conductor, second this wire is covered by a hard plastic material for providing insulation to wire. Third, the inner wire with its plastic cover is surrounded by a net of metal wires which works as outer conductor. Fourth, the whole wire is again encased with insulated plastic material. Both the conductors share the same axis therefore it is called coaxial cable.

The shielding of cable provides a better protection from external interferences and crosstalk than twisted pair. The coaxial cable provides a higher bandwidth by carrying high frequency signals of range from 10MHz to 500MHz. The latest wires are able to carry 1GHz frequency signals. The diameter of the cable varies from 0.4 to 1 inch. It can be used for transmitting the analog and digital signals for long distances up to 1Km without any repeaters. It is used for a wide range of applications in data communication such as cable television network, Long distance telephone networks, local area network. The commonly used coaxial cables are thin coaxial cable and thick coaxial cable. Thin coaxial cables are generally used in thin Ethernet. The diameter of cable is around 0.5 inch. These cables are flexible and easy to handle. Thick Coaxial cables are generally used in thick Ethernet. The diameter of cable is around 1 inch. These cables are very stiff and difficult to handle.

12.5.1.2 Glass fibers

Computer networks also use flexible glass fibers to transmit data. This new transmission media is known as “optical fiber” and uses light to transmit a signal rather than voltages used in copper media. The optical fiber is made up of glass or plastic material. Usually silica glass optical fiber used for high data speed networks where more accuracy is required in data transmission over long distances. Plastic optical fiber is also used as low cost solution for limited distance applications. These cables can be used in LAN, telephone systems, long distance communication, and establishing the backbone of country wide network. The optical fiber consists of the following:

- ***Inner core****:* Inner core is made up of very thin strands or fibers of glass or plastic material. The diameter of fiber is generally in the range of 2-125 microns. It must be of good quality material and smooth in size and shape. Usually silica glass is used for better performance.
- ***Cladding****:* Cladding means the process of protecting one metal by bonding a second metal to its surface. Each fiber is surrounded by its own cladding which is a glass or plastic material of lower density than inner core.

- ***Outer Jacket:*** This is the outer most layer of the optical fiber cable. It contains one or more strands of fibers with cladding. It is made up of different kind of materials to protect the inner core form moisture, abrasion, crushing and other damages.

The optical fiber uses the light to transmit the data. Generally infrared light is used which has a longer wavelength and therefore invisible to humans. A presence of a light pulse can be used to represent a 1 bit and absence to represent a 0 bit. For transmitting the data a transmitter is required to transmit the data. This transmitter uses light emitting diode (LED) or laser diode (LD) to converts the electrical data signals into light pulses and transmits the signals in optical fiber. At the other end of the cable a receiver is required to receive and sense the light pulses. The receiver uses the light sensitive photodiode or photo transistor to convert the light pulses into electrical voltages. The transmitter is also known a light source and receiver as detector.

12.5.1.3 Advantages of optical fiber

- Higher data rate can be up to 2Gbps or more for longer distances (tens of km).
- Smaller size and lighter weight than coaxial cables.
- Since optical fibers use light therefore these cable are not susceptible to electrical interference, cross talk.
- These cables do cause the electrical interference to other nearby cables.
- The glass fibers are designed in such a way that whole light is reflected inside the core, so a fiber can carry signal for a long distance than the copper wire.
- Lower attenuation.
- The Light can be encoded with more information than electrical voltage therefore an optical fiber can carry more information.
- Light can travel in a single fiber while in electricity it requires a pair of wires to complete a circuit.

12.5.1.4 Disadvantages of optical fibers

- Installation of optical fiber is difficult than coaxial cables. It requires special care at the terminal ends to pass through the light.
- Fiber glass is fragile. Therefore it can not be bent like copper wire other wise it will break..
- Identifying the location of breaks in fiber is difficult.
- Repair of broken fiber is difficult. It needs special instruments to join the two fibers so that the light can pass through the joint.
- Optical fiber cable and associated equipments are more expensive than copper wires.
- It requires high quality manufacturing standards because even small impurity or imperfection in core glass can defeat the purpose.

12.5.2 Wireless media

In wireless media the signals are transmitted in the form of electromagnetic waves through air (it may be water also in some cases). A special device, antenna is used to receive and transmit signals in wireless communication. For transmission antenna radiates the electromagnetic waves into medium. At the other end antenna receives the electromagnetic waves from the surrounding medium. Signals in the medium can be transmitted in two ways broadcasting (omni-directional) and directional. In broadcasting, electromagnetic wave (signals) spreads out in all directions and can be received by more than one

antenna. In directional, the electromagnetic wave is directed in particular direction through antenna. The receiving device must be aligned in that direction to receive the signals. In general, the higher, the frequency of a signal, the more it is possible to focus it into a directional beam. Three general ranges of frequencies are common for wireless transmission and are identified as radio waves, microwaves and infrared waves. Radio frequencies can travel through air or space, but require specific transmitting and receiving mechanisms. The mode of propagation of electromagnetic waves in the atmosphere and in free space may be subdivided into following categories:

12.5.2.1 Surface

The radio waves travel through the lowest portion of the atmosphere, just touching the surface. The distance travel depends on the power in the signal greater the power greater the distance. Surface propagation can also take place in seawater. VLF and LF waves are propagated as surface waves.

12.5.2.2 Tropospheic propagation

The signal can be directed in a straight line from antenna to antenna (line of sight). Another way is that waves can be broadcasted at an angle so that the waves get back on the earth after reflecting from troposphere. Troposphere is the portion of atmosphere extending outward approximately up to 30 miles from earth surface. MF signals are propagated in the troposphere since these frequencies are absorbed by the ionosphere. Absorption increases during day time. Therefore, most MF transmission relies on light-of-sight antennas. Atmospheric noise, man-made noise, and thermal noise from electronic components at the receiver are main disturbances for signal transmission of MF.

12.5.2.3 Ionospheric propagation or sky-wave propagation

Sky-wave propagation results from transmitted signals being reflected from the ionosphere, which consists of several layers of charged particles ranging in altitude from 30–250 miles above the surface of the earth. This type of transmission allows for greater distances to be covered with lower power output.

12.5.2.4 Line-of-sight (LOS) propagation

In LOS propagation very high frequency (VHF) signals are transmitted in straight line directly from antenna to antenna. Antennas must be directional, facing each other and separated from each other so the LOS should not be affected by curvature of the earth. This propagation is being used for important services like TV, FM radio, aircraft navigation etc.

12.5.2.5 Space propagation

This utilizes satellite relays in place of atmospheric refraction. A broadcast signal is received by an orbiting satellite, which rebroadcasts the signal to the receiver back on the earth. It acts like a super high antenna and increase the distance remarkably between sender and receiver. Basically it is LOS communication used to transmit signals of SHF and EHF for covering longer distances because SHF and EHF signals can cross without absorbance in the ionosphere layers.

12.6 Characteristics of Communication Channels

The following are the main characteristics of communication channels:

12.6.1 Bandwidth

The bandwidth refers to the total capacity of a communication channel to transmit data. In **analogous communication**, it is the difference between the highest and lowest frequencies capable of being carried over a channel. The greater the bandwidth, the more signals that can be carried over a given frequency range. For example voice grade lines transmit frequencies from 300 Hz to 3400 Hz. Thus the bandwidth is 3400 Hz - 300 Hz = 3100 Hz or 3.1 KHz.

In **digital communication** and networking, bandwidth refers to data rate – the amount of data that can be transferred over a communication medium in a given period. Data rate is measured in bits per second (bps) and can vary considerably from one type of channel to another. For example LAN has data rates ranging from 4 Mbps (megabits per second) to 1000Mbps (or 1Gbps). Now a days it is up to 10Gbps also on optical media. The bandwidth of dialup connections using modem ranges from 33.6Kbps to 56Kbps or 1Mbps.

12.6.2 Data transmission speed – bps and baud

Serial data speed is expressed as the number of bits transmitted per second (bps). This is also often referred to as the baud rate, but actually baud and bps not necessarily the same. When data is given in bps, the actual number of bits transmitted per second is specified, whereas, baud rate is the number of signals transmitted per second.

12.6.3 Baseband and broadband

Baseband is a signaling technique in which the signal is transmitted in its original form and not changed by modulation. Broadband makes use of multiple channels over the same medium by frequency division of the bandwidth.

12.6.4 Narrowband

In this bandwidth, data is transmitted in a range of 300 to 1200bps. They are mainly used for low speed terminals.

12.6.5 Voice band or medium band

This channel is the standard telephone lines wich allows transmission rates from 300 to 2400bauds or upto 9600bauds or more.

12.6.6 Broadband or high speed

Transmission rate varies from 19200bps to billion bps. It is being used for transmission of large volume of data such as TV network satellite communication etc.

12.6.7 Transmission technology

Broadcast networks: Two or more communicating devices on the network share a single communication channel. This is also known as **multipoint** ormultidrop line configuration. A message sent by a device is received by all devices but only the concerned device process the message and other discard the message.

Point-to-point networks: This type of network consists of a dedicated link between communicating devices. This is also known as point-to-point line configuration. The entire capacity of the channel is reserved for transmission between those two communicating devices.

12.6.8 Transmission mode

This term is used to define the direction of signal flow between two linked devices. Different types of transmission mode are described as follow:

Simplex: In this mode, the communication is unidirectional. Only one of the two stations on a link can transmit; the other can only receive. Data flows in one direction only. For example Radio system, TV transmission, Printers, Keyboards, PAGER system etc.

Half duplex: In half duplex mode, each station can both transmit and receive, but not at the same time. Data flows in both directions but at a time only in one direction. For example voice communication (telephone talk), Transmission of data to and from hard disk etc.

Full duplex: In full duplex mode, both stations can transmit and receive simultaneously. Data flows in both directions. For example data transmission, Internet access, Chatting etc.

12.6.9 Digital data transmission

Data transmission across the media can be either in serial or parallel. In **parallel** mode multiple bits each on separate channel, are sent simultaneously at a time. For example, for 8 bits transmission 8 channels are required. Advantage of parallel channels is high speed but cost is too high because of multiple channels. Therefore it is generally used for short distances. In **serial** transmission one bit is transmitted at one time over single channel. Thus it reduces the cost but also speed. Serial communication is generally used for long distance communication. Communication within the devices is parallel therefore conversion devices are required at the interfaces between sender and line and between line and receiver.

Serial transmission may either be asynchronous or synchronous. **Asynchronous transmission** transmits one character (8 bits) at a time, with each character preceded by a start bit and followed by a stop bit because the timing of signal is unimportant. Sender and receiver are not alert to send and receive the signals with time binding. Asynchronous transmission is inefficient because of additional bit are required to indicate start and stop bit and an ideal time between the transmission of characters. It is normally used for low speed data transmission at rates below 2400 bps. In **synchronous transmission** a group of characters is sent at a time. The start and end of a character is determined by a timing signal indicated by the sending device. Thus it eliminates the need for that start and stop bit as well as gaps between characters. However the sender and receiver must be in perfect coordination to avoid the loss or gain of data. A unique string of bits called sync bits is used to synchronize the timing of sender and receiver. Synchronous transmission is generally used for high speed data communication.

12.6.10 Transmission error control

A communication channel may be subject to noise which may corrupt the original signals. This necessitates accuracy controls for data transmission. These controls consist of bits known as parity bits that are similar to check sums added to data by the sender. Parity bits are checked at the receiving end to find whether bits were lost during data transmission. If errors are detected, these may be rectified by

using backward or forward error correction methods. In backward error correction method, the sender is requested to retransmit the entire data or a particular part as per requirement. Forward error correction method makes use of knowledge of data stream and mathematical algorithms (for example Cyclic redundancy checks (CRC), Hamming code, Burst error correction (BEC) etc.) to allow the receiver to correct the received data without going back to sender. However this method is more complex but preferred over long distances where retransmission is costly.

****** ☺ ******

Lesson 13

COMMUNICATION SYSTEMS

13.1 Introduction

This lesson will discuss various communication systems used for communication purpose and their applications. These topics will be useful to the students for selecting communication mediums based on their characteristics and systems used for various purposes.

13.2 Communication Systems

Basically communication systems are for smooth flow of information between two or more parties. These may be categorized based on the primary medium used for data transmission. Communication systems may also be classified as one-way, two-way, or multiple-way systems, depending on how many parties can exchange information through its various components. A few important communication systems are described as follows:

13.2.1 Radio communication system

In the radio communication system, information is transmitted with the help of a radio. Radio is a small part of the electromagnetic spectrum ranging from 3 KHz to 300GHz but in radio communication system waves ranging from 3 KHz to 300MHz (i.e. VLF, LF, MF, HF and VHF bands) are used because wave above 100MHz behave in different manner therefore, require different technology for transmission. In radio communication, signals are transmitted by modulation of electromagnetic waves in this range. Radio communication system works with the help of a transmitter and a receiver both equipped with an antenna. On one end of these radio systems is a transmitter that will take the information and electronically convert it into radio waves. These radio waves travel to the other end of the radio communication system, which is designed to detect and decode the waves and convert them to recognizable information.

Radio waves can travel for long distances along earth curvature and can penetrate buildings easily, so they are widely used for communication, both indoors and outdoors. Radio waves are Omni directional, meaning that they travel in all directions from the source, so that the transmitter and receiver do not have to be carefully aligned physically. The properties of radio waves are frequency dependent. At low frequencies, radio waves pass through obstacles well, but the power falls off sharply with distance from the source. At high frequencies, radio waves tend to travel in straight lines and bounce off obstacles. They are also absorbed by rain. At all frequencies, radio waves are subject to interference from motors and other electrical equipment. Radio system is popular for wireless data

services and is being widely used for Audible Radio through transistors and radio instruments at home; Wireless local area networks; Mobile phone communication etc.

Advantages:	Disadvantages:
• Easy to generate • Can travel long distances, • Can penetrate buildings • Omini Directional • No proper alignment of transmitter and receiver is required	• Range is limited • Not very reliable • Prone to electrical interference

13.2.2 Power line communication systems

These systems are used to transmit signals electronically from a source to destinations. A type of electronic system that often is referred to is cable television, widely known for its transmission of a video channels to houses in addition to their use for providing Internet access. Coaxial cable is generally used in cable TV network. Power line communication systems are often used because of their relatively low cost, even though there are other systems which are better in quality and efficiency. Low bandwidth is main limitation of this system.

13.2.3 Optical communication system

Optical communication system has revolutionized the wire-line communication system and offers many improvements over other type of systems. This system uses light as the medium of communication. These signals are faster, clearer, and more reliable than electrical or radio signals. These signals travel through optical fiber cables. Data/ information from source device is converted into an optical signal and reaches to the recipient. The receiver then decodes the signal and responds accordingly. Optical communication is being used extensively in data communication in Local Area Network, Public and Private communication networks (WAN, MAN) for providing backbone of the network etc. It is also being used in helicopters and aircraft for safe landing. The pilots receive light signals from the base and decide their next movements. Visible lights (Red, Yellow, Green) are used to control traffic on roads.

13.2.4 Terrestrial microwave communication system

Electromagnetic waves above 100MHz (i.e. microwaves) do not follow the curvature of earth and therefore travel in a straight line through air. Because of Earth's curvature and the microwave transmission being in a straight line, this system requires relay stations (repeaters) with a dish antenna at suitable distance for the reception, amplification and transmission of data to increase the distance served by terrestrial microwave communication system. Transmitting and receiving antennas must be accurately aligned with each other. The higher the towers are, the further apart relay stations can be. The distance between relay stations goes up very roughly with the square root of the tower height. For 100-m high towers, repeaters can be spaced 80km apart. Terrestrial microwave with repeaters provides the basis for telephone systems worldwide.

Unlike radio waves at lower frequencies, microwaves do not pass through buildings well. In addition, even though the beam may be well focused at the transmitter, there is still some divergence in space.

Some waves may be refracted off low-lying atmospheric layers and may take slightly longer to arrive than direct waves.

Microwave signals propagate in one direction at a time, which means that two frequencies are necessary for two-way communication such as a telephone conversation. One frequency is reserved for transmission in one direction and the other for transmission in the other. Each frequency requires its own transmitter and receiver. Today, both pieces of equipment usually are combined in a single piece of equipment called a transceiver, which allows a single antenna to serve both frequencies and functions.

Advantages:	Disadvantages:
• Less expensive to install than cables for moderate distances. • Offers high data rates in comparison to Radio system • Require little or no maintenance • Easy to install	• Limited by the curvature of the earth • LOS is subject to environmental and atmospheric conditions. • Noise problem • Prone to electrical interference

13.2.5 Satellite communication system

Satellite communication is another mode of data communication system for line-of-sight microwave transmission. It uses an artificial satellite positioned in the space to facilitate the communication among the various points on the earth. SHF and EHF microwaves are generally used for satellite communication since these waves can travel at faster speed in straight line and can penetrate into the space crossing ionosphere. The principle is same as in terrestrial microwave communication with a satellite acting as a super tall antenna and repeater. Satellite communication allows a microwave signal to covers a large span of the earth with a single bounce. Satellite contains several transponders, each of which listens to some portion of the spectrum, amplifies the incoming signal, and then rebroadcasts it at another frequency, to avoid interference with the incoming signal. Satellite majorly works on the solar power which is continuously received by the satellite's solar panels.

There are two components in satellite communication namely Ground station or the Earth base station and space component (satellite). A signal is transmitted from earth base station towards satellite stationed in space. Satellite receives signal with the help of antenna. Received signals are amplified to an optimum level and then with the help of transponders they are retransmitted back to the earth. The Earth station then receives the signal from the satellite, and re-amplifies it and helps in the communication. Transmission of signals from Earth station to satellite is called ***uplink*** and transmission from satellite to the Earth station is called ***down link***. Common frequency bands used for satellite communication are as follows (table-13.1):

Table 13.1 Satellite frequency bands

Band	Downlink (GHz)	Uplink (GHz)
C	3.7 - 4.2	5.925 – 6.425
Ku	11.7 – 12.2	14.0 – 14.5
Ka	17.7 – 21.7	27.5 – 31.0
Q	36	46
V	46	56

Different type of satellite systems are being used for different purpose and requirement of connectivity for communication. Some important types of satellite system are described as given below:

13.2.5.1 Geosynchronous satellites (or Fixed satellite system)

Satellites that are placed in the Earth's orbit and move with the same speed of Earth to provide continuous communication at all time during 24hours in a day are called geosynchronous satellites. These satellites seem to remain fixed above a certain spot. One satellite in orbit has line-of-sight contact with a vast number of Earth stations, but the curvature of the Earth still keeps much of the planet out of sight. Therefore, one geosynchronous satellite cannot cover the whole earth. It takes a minimum of three satellites equidistant from each other in geosynchronous orbit to provide full global transmission. The fixed satellite system helps in the transfer of numerous data and information across the countries through fixed point on the earth's surface.

13.2.5.2 Mobile satellite system

It is helpful in connecting ships, aircrafts at distant and remote places.

13.2.5.3 Research satellite system

Research Satellite System is primarily helpful in various research processes for the scientists. The scientists can gather all the necessary and useful data through the research satellite system.

13.2.5.4 VSAT

It is a low-cost micro station, sometime called VSATs (Very Small Aperture Terminals). These tiny terminals have 1-meter antenna and can put out about 1 watt of power. In many VSAT systems, the micro stations do not have enough power to communicate directly with one another (via the satellite). Instead, a special ground station, the hub with a large, high-gain antenna is needed to relay traffic between VSATs.

13.2.5.5 Applications

- Television Distribution
- Long distance Telephone Transmission
- Private Business Networks

Advantages:	Disadvantages:
• Can cover large area of the earth • Cost is independent of distance covered • Mobility can be easily achieved • Ideal for TV broadcast	• Very expensive technology • Less secure • Vulnerable to interference • Effected by atmospheric conditions.

13.2.6 Infrared communication system

Infrared communication systems use the infrared (IR) electromagnetic waves (300GHz-200THz) for communication. IR behaves in different ways therefore require different technology to use IR in communication. Unguided infrared and millimeter waves are widely used for short-range communicationamong computer peripherals. IR enabled devices (like remote controls, printers, PDAs, etc.) use infrared light-emitting diodes (LEDs) to emit infrared radiation which is focused by a plastic lens into a narrow beam. The beam is modulated, i.e. switched on and off, to encode the data. The receiver uses a silicon photodiode to convert the infrared radiation to an electric current. It responds only to the rapidly pulsing signal created by the transmitter, and filters out slowly changing infrared radiation from ambient light. Infrared communications are useful for indoor use in areas of high population density. Infrared is the most common way for remote controls to command appliances. Infrared remote control protocols like ***RC-5, SIRC***, are used to communicate with infrared.

They are relatively directional, cheap and easy to build, but have a major drawback: they do not pass through solid objects. An advantage of this is that an infrared system in one room of a building will not interface with a similar system in adjacent rooms. Security of infrared systems against eaves dropping is better than radio systems. No government license is needed to operate an infrared system, in contrast to radio systems.

13.2.6.1 Applications

The remote controls used on televisions, VCRs and stereos all use infrared communication. Infrared imaging is used extensively for military and civilian purposes. Military applications include target acquisition, surveillance, night vision, homing and tracking. Non-military uses include thermal efficiency analysis, remote temperature sensing, short-ranged wireless communication, spectroscopy, and weather forecasting

IR can be used in indoor wireless LANs. The computers and offices in a building can be equipped with relatively unfocused (somewhat omni directional) infrared transmitters and receivers. In this way, portable computers with infrared capability can be on the local LAN without having to physically connect to it. When several people show up for a meeting with their portables, they can just sit down in the conference room and be fully connected without having to plug in. Infrared communication cannot be used outdoors because the sun shines as brightly in the infrared as in the visible spectrum.

13.4 Computer Networks

A Computer Network may be defined as an interconnected collection of autonomous computers. Two computers are said to be interconnected if they are able to exchange information. The connections may be through copper wires, optical fibers, and wireless electromagnetic or optical media. Autonomous means that there is no master/slave relationship between the connected devices.

Typically, each device in a network serves a specific purpose for one or more individuals. For example, a PC can provide access to information or software. On the other hand another PC may be a file server devoted to managing a disk drive containing shared files.

13.4.1 Network relationship types

This term refers to different concepts about how one computer makes use of another computer's resources over network. There are two types of network relationship as described below:

- **Peer-to-pear network relationship**: In this relationship, each device or computer on the network communicates with each other as equals and channel is also shared equally. Each computer is responsible for making its resources available to other computers on the network.
- **Client/Server network relationship**: In this relationship a distinction is made between the computers on the network. A computer is made responsible for providing services and is called as a server while other computers use these services and called as clients.

13.4.2 Categories of networks

The category of a network is determined by its size, its ownership, the distance it covers, and its physical architecture. A computer network is generally divided into three categories:

- Local area network (LAN): It may cover a small geographic area connecting devices in a single building or group of buildings.
- Metropolitan area network (MAN): It covers medium size area such as a city.
- Wide area network (WAN): It covers a large area such as a state, country or the world

13.4.3 Internetworks and intranet

A collection of interconnected networks is called an internetwork or just internet. The term internet (lower i) should not be confused with Internet (uppercase I). The first is generic term used to mean an interconnection of networks. The second is the name of a specific worldwide network. Intranet is a closed network set up for exclusive use and managed by an organization.

13.4.4 Goals of computer networks

Network goals can be summarized in terms of the uses of networks for companies, organizations, people etc. These uses can be viewed as the facilities provided by computer networks. Some of the goals or objectives can be summarized as:

I. **Resource sharing:** Goal is to make all programs, data, and equipment available to anyone on the network without regard to the physical location of the resource and the user. This provides a high availability of resources to users.

II. **Load sharing:** This is another aspect of resource sharing. Sharing load between multiple computers connected together can reduce the delays for carrying out time intensive applications.

III. **High reliability:** High reliability can be achieved due to alternative sources of supply. For example, all files could be replicated on two or three machines. So, if one of them is

unavailable (due to a hardware failure), the other copies could be used. In addition, the presence of multiple CPUs means that if one goes down the others may be able to take over its work.

IV. **Cost effectiveness:** Small computers have a much better price/performance ratio than large ones. Therefore linking of small size computer gives high performance just equivalent to large systems.

V. **Scalability:** A closely related point is the ability to increase system performance gradually as the workload increases just by adding more PCs.

VI. **Powerful communication medium:** A real time communication can be possible between two persons sitting on-line and far apart (distant geographical locations).

13.4.5 Local area networks

Generally called LANs, are privately - owned networks within a single office single building or campus of up to a few kilometers in size. They are widely used to connect personal computers and workstations in company offices and factories to share resources (e.g., printers) and exchange information. LANs with their emphasis on low cost and simplicity have been based on the broadcast approach. LANs are distinguished from other kinds of networks by three characteristics: (1) their *size*, (2) their *transmission technology*. (3) their *topology*.

13.4.5.1 Size

LANs are restricted in size, which means that the worst-case transmission time is bounded and known in advance.

13.4.5.2 Transmission Technology

LANs often use a transmission technology consisting of a single cable to which all the machines are attached. Traditional LANs run at speeds of 10 to 100 Mbps, have low delay (tens of microseconds), and make very few errors. Newer LANs may operate at higher speeds, up to hundreds of megabits sec.

13.4.5.3 Topology

Topology of a network is the geometric representation of the relationship (physical and logical) of all links and linking devices to each other. Topologies used in LAN are:

- **Star**: Each device has a dedicated point-to-point link only to a central controller called a hub or switch. The devices are not directly linked to each other. It is less expensive (less cables required), easy to install and reconfigure and robust. Link failure of one device will isolate that device from the network. The whole network will be down only when central device (switch/ hub) is faulty. This topology is scalable but not infinitely.
- **Tree**: It is a variation of star topology. The devices are connected to a hub and that hub is further connected to another hub or a central hub. Thus it makes a tree like shape. The advantages and disadvantages are same as star topology. As the tree size, it will adversely affect throughput of the network.

- **Bus**: All the devices in the network are connected to a single long cable that acts as a backbone of the network. A bus topology is a kind of broadcast network i.e. multipoint line configuration. It is easy to install but difficult to reconfigure and fault isolation. Scalability is also an issue as only limited number of devices can be attached to one patch of cable. If there is fault in the backbone then whole network will not be functional.
- **Ring**: Each device is connected with the two devices on either side of it through a dedicated point-to-point link. Data is transferred from one computer to another in sequential order. It is easy to install and reconfigure. Each device is linked to its immediate neighbors therefore addition and removal of computers is easy from the network. The physical placement of devices is immaterial. Device can be placed in rectangular, triangular or in any shape. If one link is breakdown than network will not be in working order.

These topologies are graphically depicted in the figure 13.1 given below:

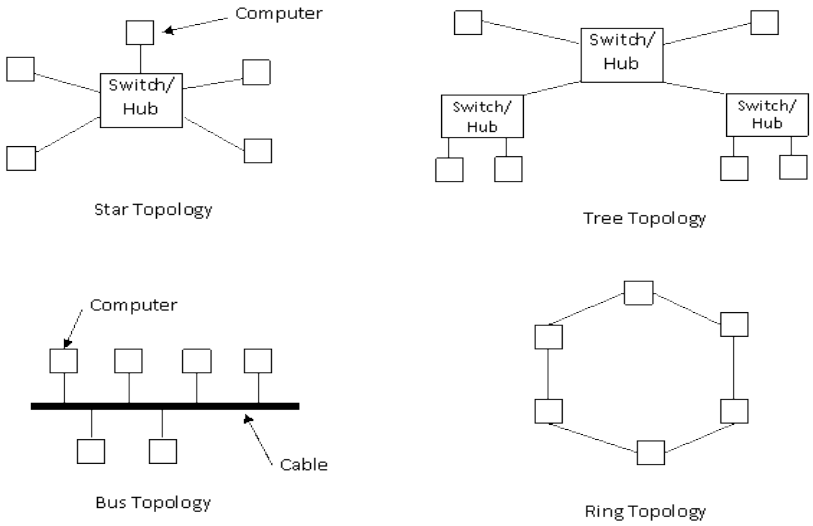


Fig. 13.1 LAN topologies

Lesson 14
PROCESS CONTROL SYSTEM

14.1 Introduction

This lesson will discuss the need and importance of process control system and Programmable Logic Controller (PLC) in dairy industry. Various topics related to the concept of process control system such as types of process control systems; open loop; closed loop; closed loop with feed back; cascade types; and concept of PLCs; system scale; user interface; communication and programming are elaborated here. The topics discussed in this lesson and next lesson will be useful to the students for understanding the state of the art technology being used in dairy plants for automation.

Process control system facilitate mass production of continuous process such pasteurization of milk, milk powder, ghee production, oil refining, cleaning in place (CIP), etc. On detection of an out-of-control situation by process monitoring, the person responsible for the process makes a change to bring the process back into control. Process control enables automation with which a small staff of operating personnel can operate a complex process from a central control room. Therefore it is being used extensively in many industries including dairy plants. Process control is a synergy of statistics and engineering discipline that deals with architectures, mechanisms and algorithms for maintaining the output of a specific process within a desired range.

For example, heating up the temperature in pasteurization of milk a process that has the specific, desired outcome to reach and maintain constant over time at 70°C. Here, the temperature is the **controlled variable**. At the same time, it is the **input variable** since it is measured by a thermometer and used to decide whether to heat or not to heat. The desired temperature (70°C) is the setpoint. The state of the heater (e.g. the setting of the valve allowing hot water to flow through it) is called the **manipulated variable** since it is subject to control actions.

Two of the most important signals used in process control are called:

I. Process variable or PV
II. Manipulated variable or MV

At this time it is necessary to define some terms used in the field of automatic process control. The first term is ***controlled variable***, which is the variable that must be maintained, or controlled, at some desired value. If the process outlet temperature has to be maintained at some constant value, then $T(t)$, is the controlled variable. Sometimes the term ***process variable*** and/or measurement are also used to refer to the controlled variable. In industrial process control, the PV is measured by an instrument in the field, and acts as an input to an automatic controller which takes action based on the value of it. Alternatively, the PV can be an input to a data display so that the operator can use the reading to adjust the process through manual control and supervision.

The ***set point*** is the desired value of the controlled variable. Thus the job of a control system is to maintain the controlled variable at its set point. The ideal value of the PV is often called the target value, and in the case of an automatic control, the term ***set point*** (SP) value is preferred.

Following are the process control equipments.

1. Transmitter: It sense the parameter being controlled and transmits the corresponding signal
2. Controller: It is the device that serves to maintain the process variable value at the set point. It receives the signal of measured variable and compares with that of preset value.
3. Actuator: It physically responds to the signal received from controller. Controller sends a signal to actuator if the difference between measured value and desired value is higher than the limit.

General working principle of process control equipment involving the functioning of process control loops is shown in the fig. 14.1 given below

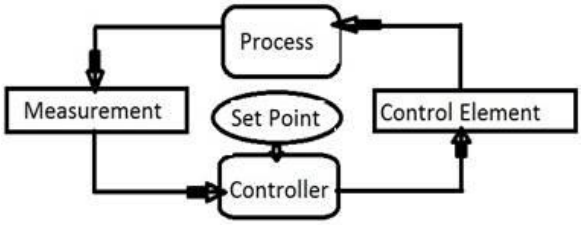


Fig. 14.1 Diagram shows the working principle of process control loop

14.2 Types of Control Systems

Control system will be designed based on the process. Most of the process will have dynamic variables, pneumatic, electrical or electronic type signals are being used to measure the process variables. The controller compares the measured signal with that of preset value and initiates the corrective actions if it is deviating. The control element exerts a direct influence on the process by accepting the input signal from controller and transforms into some action to regulate the process.

There are four major process control loops, they are as follows.

1. Feed-back control loop
2. Feed-Forward control loop
3. Sequence control loop
4. Cascade Control

14.2.1 Feed-back control loop

This is very commonly used control loop. Here the sensor measures the actual value of controlled variable and compares with reference value (set point). The difference between desired and measured value is used as input to the feedback controller to regulate the controlled variable. This loop corrects the error between the controlled output and its corresponding reference value. It tries to maintain the controlled variable to set point. Schematic diagram of feedback control loop is shown in diagram (fig.14.2) given below.

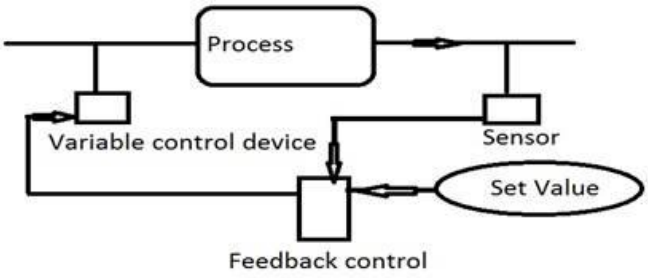


Fig. 14.2 Feed-back control loop

14.2.2 Feed –Forward control loop

This type of control is different from the feedback control. This has more advance technology in it, unlike the feedback control, which aims at eliminating the errors, feed-forward control aims at minimizing the errors. Feed-forward control acts before the output is disturbed and the instrumentation will take corrective action before a deviation in the process variable. The objective of feed-forward control is to measure the disturbances and compensate for them before the controlled variable deviates from the set point. If applied correctly, the controlled variable deviation would be minimal.

A concrete example of feed-forward control is the heat exchanger. Suppose, "major" disturbances are the inlet temperature and the process flow. To implement feed-forward control these two disturbances must first be measured and then a decision made as to how to manipulate the steam valve to compensate for them. The feed-forward controller makes the decision about how to manipulate the steam valve to maintain the controlled variable at set point, depending on the inlet temperature and process flow.

14.2.3 Sequence control loop

This loop performs the step by step execution of timely ordered events. Each step may be simple actions such as opening of valve, starting a pump or stopping a motor drive. Here combinations of feed-forward and feed-back control can be made to avail the best suited solution to control the process variable.

14.2.4 Cascade control

In this loop, two or more control loops are nested. Example, feedback control loops is a secondary control loop located inside a primary control loop (Feed-forward control loop). The primary loop controller is used to calculate the set point for the inner (secondary) control loop.

14.3 Programmable logic controller

A programmable logic controller or programmable controller is a microprocessor-based device used to control industrial processes or machines used for automation of electromechanical processes, such as control of machinery on factory assembly lines, amusement rides, or light fixtures. PLCs are used in many industries and machines. They provide advanced functions, such as analog monitoring, high speed motion control as well as share data over communication networks. Unlike general-purpose computers, PLC is designed to work for multiple inputs and output arrangements. PLCs are robust devices which are not affected by extreme temperature ranges, electrical noise, and resistance to vibration and impact. Programs to control machine operation are typically stored in battery-backed-up or non-volatile memory. A PLC is an example of a hard real time system since output results must be produced in

response to input conditions within a bounded time, otherwise unintended operation will result. In simple terms PLCs are digital computer to control the functioning of machines automatically.

14.3.1 Components of PLC

Main components of PLCs are described below and shown in fig. 14.3.

1. **Input Module:** Main function is to converts incoming signals into signals which can be processed by PLC, and to pass these to the central control unit. This module is connected/embedded with hardware/software drivers for the industrial process sensors such as switch status sensors, proximity detectors, interlock settings and so on.
2. **Output Module:** Main function is to convert the PLC signal into signals suitable for the actuators. This module is connected with hardware/software drivers for the industrial process actuators, such as solenoid switches, motors and valves which perform some action.
3. **Central Control Unit:** Actual processing of input signals is performed in this unit in accordance with the logic of program stored in memory.
4. **Data Memory:** Status of switches, interlocks, past values of data and other working data is stored here
5. **Program Memory:** Instructions for the logical control sequence are stored here in program memory.

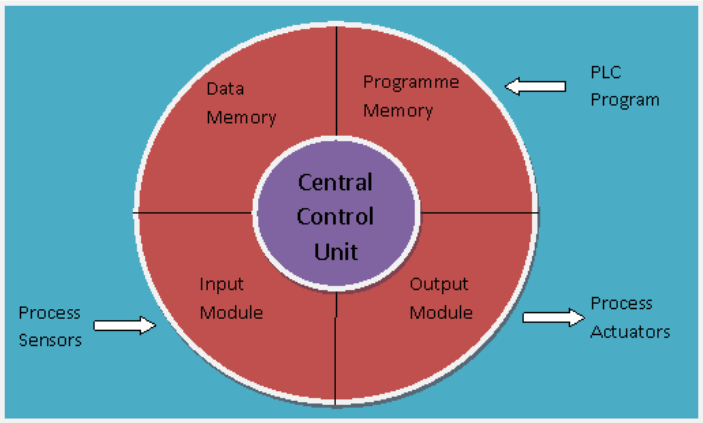


Fig. 14.3 Schematic diagram of PLC

14.3.2 Advantages of PLC

- Cost effective for controlling complex systems.
- Flexible and can be reapplied to control other systems quickly and easily.
- Computational abilities allow more sophisticated control.
- Trouble shooting aids make programming easier and reduce downtime.
- Reliable components make these likely to operate for years before failure.

14.3.3 System scale

Normally a small PLC will have a fixed number of connections built in for inputs and outputs but expansions are available if the base model has insufficient I/O. Modular PLCs have a chassis (also called a rack) in which different modules are placed with different functions. The processor and selection of I/O modules may be customized for the particular application.

14.3.4 User interface

As discussed in previous chapters, user interface are required for human and machine interface. PLCs are interactive systems therefore needs to interact with users for the purpose of configuration, alarm reporting or everyday control. A human-machine interface is used for this purpose. These are also referred as graphical user interface. A simple system may use radio button, text box, check box, lights, command buttons etc. to interact with the user. Advanced GUI with text displays as well as graphical touch screens is also available.

14.3.5 Communication

PLCs need to communicate with input device to receive data; GUI to display inputs/ processed data and output devices to act finally based on results. PLCs have built in communications ports for this purpose, usually 9-pin RS-232 is used for communication, but optionally EIA-485 or Ethernet ports are also available. Commonly used communication protocols are Modbus, BACnet or DF1. Other options include various fieldbuses such as DeviceNet orProfibus. Modern PLCs can communicate over a network to some other system, such as a computer running a SCADA (Supervisory Control and Data Acquisition) system or web browser.

14.3.6 Programming

PLCs are programmed to perform certain predefined tasks as per input signals received from input devices. PLC programs are typically written in a special application on a personal computer and then downloaded by a direct-connection cable or over a network to the PLC. The program is stored in the PLC either in battery-backed-up RAM or some other non-volatile flash memory. At basic level, PLCs are programmed in a simple form of assembly code. PLCs can also be programmed using standards-based programming languages, including the IEC 61131-3 standard, Function Block Diagrams etc. A graphical programming notation called Sequential Function Charts is available on certain programmable controllers. However, a well established standard programming language for programmable logic, called 'Ladder Logic', is universally understood by PLC programmers. Fundamental concepts of PLC programming are common to all manufacturers. However some differences exist in I/O addressing, memory organization and instruction sets. It means that PLC programs are never perfectly interchangeable between different makers. Even within the same product line of a single manufacturer, different models may not be directly compatible.

****** ☺ ******

Lesson 15
PLANT AUTOMATION

15.1 Introduction

This lesson will discuss the concept plant automation. Various topics related to automation such as advantages and disadvantages; architecture of plant automation; automation tools are elaborated here. The topics discussed in this lesson will be useful to the students for understanding the state of the art technology being used for dairy plant automation.

Over the last few decades unit operations pertaining to procurement, processing and distribution in the dairy and food sector is inclining towards mechanization and automation very rapidly. The main driving force at the beginning for mechanization was economic reasons, such as labour costs. Gradually, safety considerations, mandated by legislatures, entered the scene. Nowadays food safety policies became the major factor for the mechanization and automation of the process of food production.

15.2 Automation Meaning and Definition

The word 'Automation' is derived from Greek words "Auto" (self) and "Matos" (moving). Automation is a technology concerned with the application of mechanical, electronic and computer based systems to operate and control production. In general, automation could be considered as a process that started with machines doing the heavy work, while humans maintained the function of telling the machines what to do. However, automated systems achieve significantly superior performance than what is possible with manual systems, in terms of power, precision and speed of operation.

"Automation may be defined as a set of technologies that results in operation of machines and systems without significant human intervention and achieves performance superior to manual operation"

15.3 Types of Automation Systems

Based on the flexibility and level of integration in manufacturing process automation systems can be categorized as follows

15.3.1 Fixed automation

Fixed automation is a system in which the sequence of processing operations is fixed by the equipment configuration. This type of automation is very commonly used in case of process using mechanized machinery to perform fixed and repetitive operations in order to produce a high volume of similar parts or products.

Typical features of fixed automation are:

1. High initial investment for custom-engineered equipment
2. High production rates
3. Relatively inflexible for the accommodation of product variety

The economic justification for fixed automation is found in products that are produced in very large quantities and at high production rates. Example of fixed automation includes Distillation Process, Conveyors, Transfer lines etc.

15.3.2 Programmable automation

In programmable automation, the equipment is designed with the capability to change the sequence of operations to accommodate different product configurations using electronic controls. Set of instructions in the form of programmes loaded into the system to control the sequence of operation. This type of automation is being used where the parts or products are typically made in batches. The system must be reprogrammed with the set of machine instructions to produce each new batch of a different product. In Cheese blenders, Paper Mills etc. Programmable automation is widely used.

Some of the features that characterize programmable automation include:

1. High investment in general purpose equipment
2. Lower production rates than fixed automation
3. Flexibility to deal with variations and changes in product configuration

15.3.3 Flexible automation

It is an extension of programmable automation. This system is widely used for the production of varieties of products from the same equipment or system. This system is capable of producing a variety of products with virtually no time lost for changeover between the products. After processing, products are automatically transferred to next machine.

The features of flexible automation can be summarized as follows:

1. High investment for a custom engineered system
2. Continuous production of variable mixture of products
3. Medium production rates
4. Flexibility to deal with product design variations

15.3.4 Integrated automation

Integrated automation system is a complete automation of the manufacturing plant. The process functioning is controlled through computer system and under coordination through digital information processing. It includes technologies such as computer-aided design and manufacturing, computer-aided process planning, flexible machining systems, automated storage and retrieval systems, automated material handling systems such as robots and automated cranes and conveyors, computerized scheduling and production control. It may also integrate a business system through a common database. In other words, it symbolizes full integration of process and management operations using information and communication technologies. Typical examples of such technologies are seen in Advanced Process Automation Systems and Computer Integrated Manufacturing (CIM)

15.5 Necessity of Automation

Even though food safety and security concerns are the major driving force for implementation of automated system in the food processing plants; the industry faces other additional emerging drivers that further accentuate the importance of implementing automation as soon as possible. Integrated automation can help food and beverage manufacturers address these below emerging trends while providing important safety benefits.

1. Customer or consumer demands
2. Labour availability and reliability issues
3. Labour laws
4. Regulatory requirements
5. Supply-chain management

15.6 Advantages of Automated systems

Automated systems have got the following advantages over the manual controlling systems

1. Production of consistent quality goods
2. Reduction in production costs
3. Flexibility to meet market demands
4. Adoption to constantly changing legal demands
5. Ensures high and consistent product quality
6. Increases production
7. Reduces losses
8. Guarantees a safe operation
9. Ensures manpower savings
10. Allows for efficient production planning, execution and reporting

Overall, the aim of automation in a dairy or food processing plant or any production unit is to meet the company need to produce goods safely, in a cost-effective manner, with high and consistent quality and in the appropriate quantities to meet market demand. A typical automated system in a diary plant different level of automation is shown graphically in fig. 15.1 given below.

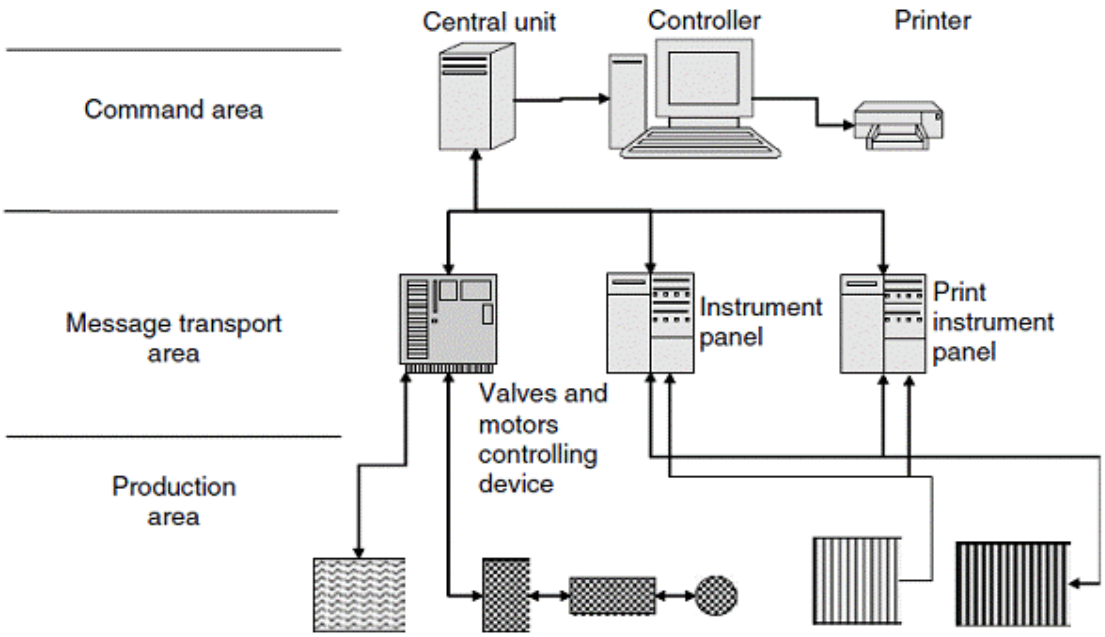


Fig. 15.1 Typical automated system in the dairy plant

15.7 Architecture of Automation system

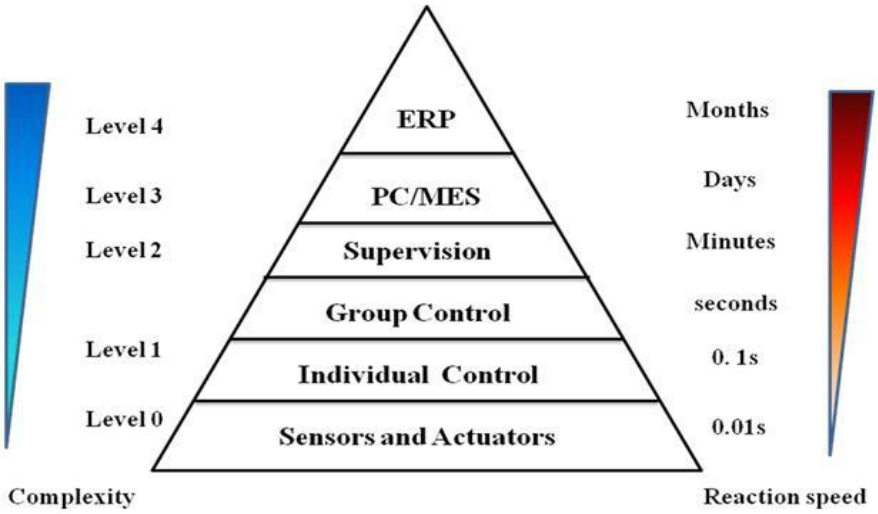


Fig. 15.2 Architecture of automation system

ERP – Enterprise Resource Planning

MES – Manufacturing Execution System

PC – Production Control

Various components in an industrial automation system can be explained using the automation pyramid as shown above. Here, various layers represent the wideness (in the sense of no. of devices) and fastness of components on the time-scale.

15.7.1 Sensors and actuators layer

This layer is closest to the processes and machines, used to translate signals so that signals can be derived from processes for analysis and decisions and hence control signals can be applied to the processes. This forms the base layer of the pyramid also called 'level 0' layer.

15.7.2 Automatic control layer

This layer consists of automatic control and monitoring systems, which drive the actuators using the process information given by sensors. This is called as 'level 1' layer.

15.7.3 Supervisory control layer

This layer drives the automatic control system by setting target/goal to the controller. Supervisory Control looks after the equipment, which may consist of several control loops. This is called as 'level 2' layer.

15.7.4 Production control layer

This solves the decision problems like production targets, resource allocation, task allocation to machines, maintenance management etc. This is called 'level 3' layer.

15.7.5 Enterprise control layer

This deals less technical and more commercial activities like supply, demand, cash flow, product marketing etc. This is called as the 'level 4' layer.

15.8 Conclusions

Automation can address the critical issues of product safety and security both proactively and reactively with its integrated logistics, material handling and warehousing systems. Proactively, automated systems reduce the number of people with direct access to products. This alone can improve product reliability significantly while reducing the risk of catastrophe. Reactively, automated systems include up-to-the minute track-and-trace functionality. This key feature of automation can easily facilitate any recalls that may be required. But beyond enhancing food safety and security, a fully integrated automation system can provide many compelling business benefits. Advanced automation systems increase visibility into food-manufacturing operations by improving the transparency and traceability of vital business information. The result is lower labour costs, increased productivity, reduced scrap and waste, and meeting, and even anticipating, the continually increasing societal and legal demands.

Lesson 16

DAIRY PROCESS MODELLING AND SIMULATION

16.1 Introduction

This lesson introduces the concept of process modelling and simulation of dairy products, application of software packages such as SCADA, INTELLUTION, CAD and CAM for product design and manufacturing. These topics will be useful for students to develop an idea about advancement taken places in dairy industry with respect to design and manufacturing of high quality dairy products.

16.2 Process Modelling

Modelling in dairy and food science helps in prediction of food materials' behaviour in varying product and process conditions. This involves data collection, analysis, interpretation and mathematical modelling. Use of computers is prevailed to run large programs for the modelling purpose. The conceptual models are then used to perform quantitative analyses and predict behaviour of food material. Food process modelling requires study of the inherent/naturally occurring processes, processes applied to the food product and also of the processes occurring in model food systems as a response to its quality. The applications of the process modelling cover production, processing and storage of food products.

16.3 Fundamentals of Process modelling

Two approaches of process modelling are used, i.e. deductive and inductive.

16.3.1 Deductive modelling

In order to solve complex problems one has to decompose the complex process into its constituent processes, then study them individually, develop their models and combine their parameters to develop complete process models. The deductive modelling of food processing therefore, starts with the general laws of (chemical) physics, and uses them to build realistic mathematical models of food processing operations. The models for the individual processes are deduced from existing theories, principles and fundamentals. Deductive approach utilises principle of 'inference from general to particular'. Advantages of deductive models are:

1. Reusable and possibility of parameter transfer
2. Widely applicable within the space of variables and systems described by the model
3. Involves basic fundamentals of food and related science
4. If model is valid on a simple system it can serve as a tool for computational exercise to adapt it to different processing conditions and more complex product geometry.

Applications of deductive modelling in dairy and food science include;

1. Process optimization: this involves modelling and simulation of all processing steps starting from milk production, processing, packaging, storage and handling. Some processes are modelled realistically; however, others are based on assumptions and theoretical considerations. The process optimization has to be governed by integrated approach which involves shelf life as

well as sensory aspects of the food products. Thus it helps in prediction of keeping quality, shelf life, microbial growth, chemical reactions and phase changes.

2. Modelling keeping quality and shelf life of foods: Optimizing the drying or freezing operations and shelf life estimation of the food stuff considering rates of various deteriorative changes helps in modelling keeping quality and shelf life.
3. Studies on migration of microbial metabolites into the food product.
4. Studying the effect of food microstructure and the microscopic water distribution on survival of micro-organisms.
5. Modelling of flavour release from the food materials: This predicts mathematically the effect of varying food composition, food structure and mastication behaviour on the perceived time intensity flavour release profile.

16.3.2 Inductive or empirical modelling

Sometimes decomposition of the complex problems and processes result in more complicated forms which are not possible to be studied through fundamental approaches. Similarly if there is no or a little information available about the processes then inductive/ empirical modelling is preferred. The term inductive refers to 'inference of the general law from particular instances'. Inductive modelling is typically data driven and requires minimal knowledge of the products or processes involved. This involves plenty of data collections and its analysis by fitting statistical models using tools such as regression and correlation. In every step an additional parameter is added to the objective function and the response is statistically analysed to see its significance. The resulting model converts model inputs into outputs and is often referred to as a 'black box' as there is no relation whatsoever with the real underlying mechanism.

Advantages:

- Models can be developed relatively quickly
- Show good results even when there is insufficient understanding of the processes involved

Disadvantages:

- Not be able to increase the understanding or generate new knowledge on the underlying mechanism.

16.3.3 Kinetic modelling

Kinetic modelling is useful from two perspectives: one, changes in physical-chemical, biochemical and microbiological reactions that occur during processing as also storage lead to loss of quality of the processed dairy foods. Kinetic modelling allows us to describe these changes as a function of processing variables and storage time-temperature so as to ascertain the rates of progress of these reactions and establish quantatively the progress of these reactions. To accomplish these tasks, understanding of classical concept of thermodynamics and kinetics are necessary.

Reaction kinetics: Kinetic evaluation of reactions involves the study of rates and mechanisms by which one chemical compound converts to another. Kinetic Equation of nth order is:

- Simplified Zero order kinetic eq. (n=0): $Q/Q_0 = kt$
- First Order Kinetic eq. (n=1): $ln(Q/Q_0) = kt$
- Second Order Kinetic eq. (n=2): $Q_0/Q = kt$

The rate constant is generally temperature dependent and their relationship can be modeled as: $k = k_0 exp\left[-\frac{E_a}{RT}\right]$

Where, R is universal gas constant, T is temperature in Kelvin, E_a is activation energy (j/mol), k_0 is Arrhenius constant

Simplified form of above equation is: $ln(k) = ln(k_0) - (E_a/R)(1/T)$

16.3.4 Heat and mass transfer modelling

Heat and mass transfer processes are the most common physical phenomena that occur during processing of foods. As a result several important variables such as the temperature and the moisture concentration inside the food depend on time as well as on their position inside the food system. Several of the product properties and quality indices of foods viz., microbial load, nutritional value, texture and organoleptic quality, are affected by these variables. An example of such a situation is very common processing intervention: thermal processing which may include retort processing or UHT sterilization. In both these processes, temperature of the boundary layer is likely to be higher than that at the centre of the can and therefore the dynamics of spore destruction will be different at the boundary and the centre of the can. The heat processing consideration may have to be designed keeping in mind the slowest heating point in the can. The famous Fourier equation considers only heat conduction processes in which heat is transported by molecular diffusion processes. Therefore the assumption is that the heat conduction in transient heat conduction in an isotropic object Ω with boundary Γ is governed by the following Fourier equation.

$$\rho c \frac{\partial T}{\partial t} = \nabla k \nabla T + Q \ on_{\Omega}$$

Where,

ρ= density (kgm^{-3})
c= heat capacity ($Jkg^{-1o}C^{-1}$)
k=thermal conductivity ($Wm^{-1o}C^{-1}$)
Q=volumetric heat generation (Wm^{-3})
T= temperature (^oC)
t= time (s)

The thermo physical parameters k, and c may be temperature dependent and therefore the problem becomes nonlinear. In general, both the heat capacity and the density can be calculated with sufficient accuracy but the models for the thermal conductivity require some assumptions to be made with respect to the direction of heat flow. In conventional thermal processing conditions, the heat generation Q is zero. However, in the case of ohmic or volumetric heating like microwave and ohmic heating, Q is the driving force of the heat transfer. The modelling of these techniques is a very active research area

The phenomena of diffusion also occur during the transport of water etc in the food system and are governed by Fick's second law of diffusion.

$$\frac{\partial C_a}{\partial t} = \nabla D_a \nabla C_a + r_a \ on_{\Omega}$$

Where,

C_a = molar concentration of component a (mol m^{-3})
D_a = mass diffusion coefficient of component a (m^2s^{-1})
r_a = rate of production of component a (mol $m^{-3}s^{-1}$)
T= time (s)

The above equation is valid only for diffusion in solids or stationary liquids with the assumptions of constant density ρ and zero mass velocity encountered in pure diffusion of a single species with negligible changes in the total density. For other conditions, more complex transport equations may be used. D_a is not constant but depends on temperature and the concentration of the components in the mixture. r_a depends on the metabolic activity of the product, which is a function of temperature and composition. In liquids, transport of heat and mass is much more complicated than in solid foods, as besides diffusion also convective transport of liquid particles may take place. Therefore different models and mathematical considerations may have to be used for such situations. Many heat transfer processes in food operations may also involve turbulent flow of water. Turbulence enhances heat transfer rates considerably and therefore turbulence should be incorporated in the models. Software tools are available for modelling heat and mass transfer problems. The reliability of the numerical solution, however, largely depends on the availability of suitable thermo physical properties and the complexity of the governing models.

16.4 Simulation

Simulation or mathematical simulation refers to a mathematical model that permits us to work out a reasonable approximation of the corresponding real-life system. It is used as a powerful tool in food processing applications for understanding the behaviour of complex interactive systems, predicting operation results, developing process control systems or optimizing the performance of a system. Mathematical aspect of simulation has been made much easy after the advent of computers. Using computers one can go from a spreadsheet where a model is built from scratch to different types of available software packages. The packages to model processes using computational methods are made of a block of generic models of unit operations and balance equations and a block of physical properties of the materials being processed. The two blocks interact to describe changes along a flow description of the process.

The steps involved in building a process simulation model are as under:

- Define the problems
- Identify the objectives
- Understand the system and collect data
- Select software
- Draw process flow diagram
- Create a rough model
- Verify and validate the initial results
- Refine the model
- Modify the base model with alternatives
- Analyze results and draw conclusions

Some of the important software systems used in process modelling and simulation studies is:

16.4.1 Supervisory control and data acquisition (SCADA)

Supervisory Control and Data Acquisition (SCADA) system is a type of industrial control system for supervisory control and data acquisition of industrial processes. It is a computer software package which is interfaced with hardware system through PLCs to monitor and control wide range of industrial activities, infrastructure and processes. It gathers and analyzes real time data to monitor and control a plant or equipment in industries at supervisory level only. Some important features of SCADA systems are:

- Dynamic process Graphic
- Alarm summery and history
- Real time trend
- Historical time trend
- Security (Application Security)
- Data base connectivity
- Device connectivity
- Recipe management

SCADA system consists of following subsystems:

1. Human-machine interface (HMI) - presents processed data to the plant operator enabling him to control the process. Frequently, it refers existing SCADA databases and provides relevant information to the operator for necessary action e.g. maintenance schedule, trouble shooting, particulars of the instruments etc. the HMI presents information generally in the form of graphical/pictorial/mimic diagrams, process flow charts which are interactive in nature.
2. Supervisory system/master station (computer) - receives digital signal and sends commands to the process controller. With the help of servers and software, it communicates with field instruments i.e. Remote terminal units (RTU)/PLC and the HMI softwares.
3. Remote terminal units – converts analogue signal to digital one and sends it to computer. It connects computer physically to the instruments.
4. Programmable logical controller - it also performs control actions by actuating feedback control loop
5. Network connecting various RTU/PLC to the supervisory system- it utilizes both wired and radio connections. Remote management of SCADA system is known as telemetry. SCADA protocols are very compact.
6. Processes and instruments

The role of SCADA system in controlling process is explained through a diagram given below (fig. 16.1, Source: http://en.wikipedia.org/wiki/SCADA). The SCADA system reads the measured flow and the level in the tank and sends the set points (given by human operator) to the PLCs. PLC1 compares the measured flow to the setpoint, controls the pump speed as required to match flow to the set point and the PLC2 controls the flow control valve to match the level to the setpoint.

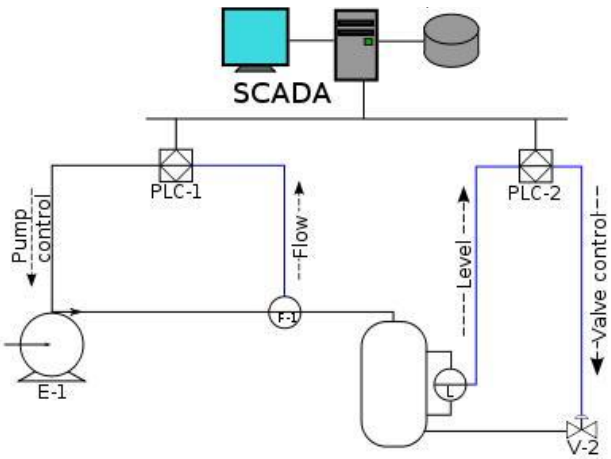


Fig. 16.1 Schematic overview of SCADA system

16.4.2 Intellution

Intellution Inc. is a company which was founded in 1980 with the aim to develop and supply software for industrial automation and very soon, it become the leader in industrial automation software. It developed and supplied the software under brand name Intellution which, provides a software platform that collects, distributes, controls and visually presents information from the plant floor throughout the enterprise, supplying customers with scalable solutions. In 2002, GE Fanuc International, Inc. purchased this company and recently released a new version of Intellution software named as Intellution iFIX 5.5 with following additional capabilities and features:

- Flexibility and reliability of connecting and presenting data
- Best-in-class information analysis, real-time data management and control with enhanced full-featured SCADA solution
- Enables faster, better intelligent control, and visibility into plant operations
- Offers open architecture, highly scalable and distributed networking model
- Scalability from isolated sensor to company-wide integration
- Adherence to industry standards for improved consistency, quality & compliance
- Completely hardware independent
- E-signature functionality
- User-friendly interface
- Provision for developing and integrating new modules using VBA Scripting and .NET

16.4.3 CAD and CAM in dairy industry

Computer-aided design (CAD) and Computer-aided manufacturing (CAM) is a concept of using suitable computer application software for designing, analyzing and manufacturing the product. The nature and use of these software is relative to what is being designed and manufactured. Formally, the efforts for developing specialized CAD/CAM software were started in piecemeal at different places for solving problems related to drafting, testing of designs by simulation, flow of design data directly to the manufacturing process, etc. These developments were pioneered by the General Motors Research Laboratories in early 1960s. During early 1970s, the large scale production of electronic chips and microprocessors has made powerful computers affordable at low prices to all. This leads to development

of more sophisticated CAD/ CAM software with advanced features for designing and manufacturing products.

Computer-aided design is the use of computers to assist in drafting, creation, communication, analyses, and optimization of 2D design or 3D models for a product or components of a product. The design of computer models is stored in computer files in form of geometrical parameters using either vector based graphics or raster graphics. Design of a product can be visualized from different angles on computer screens in two or three dimensional representations and the parameters of the product can be changed if required. The design can be tested and evaluated (e.g., physical shape, size and volume, aesthetic, attributes, fluid dynamics, material suitability, conductivity, etc.) by simulating with real-world conditions.

Initially, the CAD systems were used for by electronic and mechanical engineering industries for designing machines (say cars, airplanes, etc.) to improve the quality of designs. But, now days CAD applications are used almost in all fields such as civil engineering, garment design and production, food industries, oil industry, etc. In food and dairy industry, CAD software enables the manufacturers to design and develop novel food products of high quality at reasonable prices. This covers all areas of product development, from idea generation and research, investigating the functionality and interactivity of different ingredients, producing manufacturing flow-charts to assess food safety issues, knowledge based systems for food legislation and innovative imaging systems for product quality and fault diagnostic purposes, study of shelf life, etc. CAD system offer number of advantages such as:

- Easy modification in design
- Zooming to magnify certain elements of a model to facilitate further inspection
- 3D models give the feeling of real objects
- Simulation by changing the parameter values
- Increases the productivity of designers
- Improve the quality of design
- Improve communications through documentation
- Create a database for manufacturing.

Computer-aided manufacturing is the use of computer to assist in manufacturing process by controlling machine tools and related machinery automatically in the production of components. In broad sense this term is used when one or more manufacturing processes such as process control, robotics,measuring, monitoring and controlling production are carried out at one time aided by a computer. CAM uses 2D or 3D design data of components generated by CAD software in geometrical coded form to control the computer numerical control (CNC) machines automatically for manufacturing of components. This system differ from older forms of numerical control (NC) in which, design data are encoded mechanically. Main purpose of CAM is to monitor and control different processes more precisely during manufacturing process to produce consistent and high quality products at faster speed by minimizing the waste and energy consumption.

Generally the CAM systems were used by mechanical engineering industries. Because of its potential uses, it is being used extensively in other industries also including dairy and food industries. In diary and food industries, CAM systems are linked to manufacturing lines monitoring and controlling manufacture to produce consistent and high quality end products. For example, 'dedicated control systems' monitor single unit operations, e.g. controlling the temperature of a heat exchanger; they do not share the information with other computers. However, 'centralised control systems' monitor and control complete operations throughout manufacture; providing feedback about the entire process, e.g. milk processing plant, fish finger production. Other applications of CAM include production line robots deboning meat, decorating cakes, picking mushrooms and packaging chocolates. A few advantages of CAM within the dairy and food industry are:

- Reduced food wastage through efficient manufacture;
- Improved product consistency;
- Avoidance of downtime;
- Reduction in overheads, e.g. labour costs;
- Increased production capacity;
- No fatigue from repetitive manufacturing demands;
- Improved food safety and hygiene standards;
- Enhanced quality control.

The term CAD/CAM in combination implies to a system that can be used both for designing a product and for controlling manufacturing processes. CAD/CAM is highly integrated system since both CAD and CAM are computer-based methods and share numerical information of design data directly from CAD designs to control manufacturing of the products. The integration of CAD/CAM gave the designer much more direct control over the production process and created the possibility of completely integrated design and manufacturing processes within short duration.

Lesson 17

SPECIAL INSTRUMENTS FOR THE DAIRY INDUSTRY

17.1 Introduction

This lesson will discuss some the special instruments used in dairy industry such as e-tongue, e-nose, robotics, etc. The topics discussed in this lesson will be useful to students to know about the state of the art technology used in industry of processing of food products.

For increasing efficiency and optimizing various processing operations in either dairy or the food industry, many special instruments with state of art design and applications are increasingly finding place today. E-nose, E-tongue and robotics are few of such devices which amalgamate advances in electronic hardware and software to bring out novel applications.

17.2 E-nose

E-Nose (Electronic-Nose) or artificial olfaction use multisensory arrays and neural computing to analyse odour or aroma as a whole as perceived by human nose. The system mimics the human olfactory system and thus finds routine application for odour and aroma analysis in the food, beverage, chemical and packaging industries. The modular sensing system comprising of the array of different sensors transduce the chemical quality into electric signals followed by appropriate signal conditioning and processing to classify known odours or identify unknown odours. The sample identification is carried out through the use of pattern recognition routines operated by artificial neural networks (ANNs) to evaluate data from sensor arrays. Once the ANNs are trained for odour or volatile compound recognition, the operation consists of propagating the sensor data through the network. Briefly, sensors in E-Nose act in response of a number of organic and inorganic compounds in the ppm range. The response can be highly reproducible. Application of E-Nose permits many sorts of diverse sample to be examined.

17.2.1 Principle of E-nose

The functions of human olfactory system comprises of operations such as sniffing, reception, detection and cleansing. The odour sensation depends upon the interaction of volatile molecules with a group of specialized nerve cells called olfactory receptors, situated just above the nose out of the main air stream. The first step in the olfactory recognition is the dissolution of the chemicals in the sample in an aqueous mucous layer covering the olfactory receptor cells. The interaction of odourant molecules with olfactory receptors results in excitation of the receptors cells thus producing electrical signals which pass to the olfactory bulb. The electrical signals are further processed and finally transferred to the brain.

The E-Nose mimics the human olfactory system and each chemical sensor functions like a group of olfactory receptors resulting in a time dependent electrical signal in response to an odour. The number of odorous substances the system can recognize depends on the number of sensors used. Electronic noses usually utilize a data pre-processor, which is analogous to the olfactory bulb in the human olfactory system. The pre-processor compresses the signals and amplifies the output in order to reduce noise and improve the sensitivity of the sensor. The digital signals are then ready for analysis by the computer.

17.2.2 Sensors

The chemical sensors are detectors that measure changes in the conductance when they come in contact with volatile chemical compounds. These compounds react with the sensitive materials on the surface of the sensors interfaced to a transducer (signal transforming device). The interaction brings about some physical changes that are sensed by the transducer and converted into output signals which can be captured and interpreted. The measurement principles to detect the chemicals are: electrical, thermal, optical and mass changes. Some of the popular sensors for the e-noses are:

17.2.2.1 Chemoresistor sensors

17.2.2.2 Metal oxide semiconductors (MOS)

MOS measures changes in electrical-resistance occurring as a result of vapours of the odorous substances getting adsorbed onto the semiconductor surface. When oxygen is adsorbed on to the sensor surface, electrons are removed from semiconductor and thus its electrical resistance is increased. When reducing gases interact with the surface adsorbed oxygen, it decreases trapped electron leading to increase in electrical conductivity of the sensor. MOS are generally operated at higher temperatures (up to 400 ^{O}C). Metal-oxide sensors are fairly sensitive. Metal oxides such as SnO_2, ZnO, $Fe_2O_3$and WO_3 respond to reducible gases such as H_2, CH_4, CO, C_2H_5 or H_2S and increase their conductivity. Contrary to these semiconductors such asCuO, NiO and CoO respond to oxidizable gases such as O_2, NO_2 and Cl_2. The major limitation of these conductors is that they operate at elevated temperatures and hence energy requirements are very high.

17.2.2.3 Conducting polymers (CP)

Conducting polymers are usually synthesized by chemically or electrochemically oxidizing of the corresponding monomers. There is reversible adsorption of molecules to the films inducing a temporary change in the electrical conductance of the film by altering the population of active charge carriers in the polymer structure. When these sensors are exposed to a particular vapor, each polymer layer undergoes a characteristic swelling, drawing the conducting particles away from one another and thus increasing the measured resistance across the capacitor. Compared with metal oxides, organic polymers are much more diverse. CPs has high sensitivities, short response time and can operate at room temperature. Furthermore, conducting polymers have good mechanical properties which allow flexibility in fabrication of sensors.

17.2.2.4 Chemocapacitors (CAP)

The CAP sensors work on the following principals. In the first state, only air present in the sample environment is incorporated into the polymer and the baseline value of capacitance is obtained. In the second stage, gaseous analyte molecules that are present in the sampling environment are adsorbed onto the sensitive polymer layer resulting in change in its electrical (e.g. dielectric constant) and physical properties (e.g. volume). This deviation over the baseline value is measured.

17.2.3 Electrochemical sensors

17.2.3.1 Metal oxide semiconductor field effect transistors (MOSFET)

The micro-chemosensors are based on the structure of a MOSFET. Catalytic metals or metal alloys (e.g., Pd, Pt or Ir alloys) are used as materials of construction of gates of the transistor which are then left exposed to air. The adsorbed gases interact with the sensor and change the surface-charge density resulting in change of the potential of the device. Selectivity of MOSFET sensors is achieved by the choice of the operation temperature, the metal on the gate and by varying the microstructure of the metal.

17.2.3.2 Amperometric sensors

An amperometric sensor has a working, counter, and reference electrodes that are dipped in an electrolyte. The sensors record the current in the electrochemical cell between the working and counter electrodes as a function of the analyte concentration. The molecules of the analyte diffuse into the electrochemical cell and to the working electrode surface through a porous membrane. Then, the analyte reacts electrochemically, and through either oxidation or reduction process produces or consumes electrons at the working electrode.

17.2.4 Optical odor sensors

17.2.4.1 Surface plasmon resonance (SPR)

SPR works based on an optical phenomenon. The incident light excites a charge-density wave at the interface between a highly conductive metal and a dielectric material. The conditions for excitation are determined by the characteristics of the metal and the dielectric material. Optical SPR sensors are sensitive to the change in the refractive index of a sample surface and thus measures small changes in the refractive index of a thin region adjacent to the metal surface.

17.2.4.2 Fluorescent odor sensors

The optical sensing element is composed of a reagent phase immobilized at the fiber tip by either physical entrapment or chemical binding. This reagent phase contains a chemical indicator which interacts with analyte gases or vapours and experiences the change in optical properties. The responses depend upon the nature of the organic vapor and the strength of its interaction with the different polymer systems used. The sensitivity of optical chemosensor is generally not high (detection limits is 1000 ppm).

17.2.4.3 Gravimetric odour sensors

Gravimetric odour sensors use acoustic wave devices which operate by detecting the effect of sorbed molecules on the propagation of acoustic wave. Two types of acoustic wave odour sensors viz., bulk acoustic wave (BAW) sensor and Surface acoustic wave (SAW) sensor are used. In both the types, the basic device consists of a piezoelectric substrate such as quartz, lithium niobate and ZnO coated with a suitable sorbent coating. The sorption of vapourmolecules onto the sorbent membrane coated with the substrate affect propagation of the acoustic wave leading to changes in the resonant frequency and the wave velocity.

17.2.5 Data analysis in E-Nose

For sample identification, different pattern recognition techniques are used to analyze data derived from an E-Nose. Pattern recognition is a process of identifying structure in data by comparing it to known structure. The different pattern-recognition modules are used to obtain instrumental correlation to sensory tests for evaluating the data recorded from the array of sensors. These techniques could be statistical methods, proprietary pattern recognition methods or Artificial Neural Networks (ANNs). The statistical or chemometric approach, complementary to ANNs includes principal components analysis (PCA), partial least squares (PLS), discriminate analysis (DA), discriminate factorial analysis (DFA) and cluster analysis (CA).

17.2.6 Application of E-Nose in food industry

E-Nose has many applications which include its uses in evaluation of sensory quality of foods, detection of environmental toxins, carcinogens and pollutants, air quality as well as in medicine. Some specific functions related to food industry are listed below:

- Identification of adulteration of many raw and processed foods
- Evaluation of aging processes related to cheese maturity, fish freshness, shelf life of ground coffee etc.
- Presence of contaminants such as diacetyl detection in orange juice, identification of milk from mastitis-affected cows etc.
- Quality control applications like recognizing sensorily acceptable and rejectable samples to evaluate raw materials and finished products.
- Product and formula matching for comparing one's product with a competitors' sample
- Classify various types of cheese, various strains of bacteria, and to determine the odour from paper based packing material for foods.
- Detecting aroma differences among samples of dairy products like ice cream

17.3 E-Tongue

E-tongue refers to an array of sensors that are immersed in liquids in order to identify their "tastes". It has widespread applications in dairy and food industries to monitor the quality of products. The major advantage of e-tongue is that unlike human beings the sensitivity does not decrease even during prolonged exposure. It also permits evaluation of toxic substances which can not be tasted by human beings.

17.3.1 Sensors

Depending on the task in hand, the number of sensors in the sensor arrays may vary from 4 to 40. Sensors based on different principles of signal transduction (e.g., potentiometric and amperometric) may be used simultaneously in the same sensor array. Potentiometric ion-selective electrode (ISE) sensors are more popular sensors in electronic tongue systems. Potentiometric measurement is carried

out using a multichannel voltmeter with high input impedance. Values of the sensor potential is measured against a conventional Ag/AgCl reference electrode and stored as computer data files. Sample pretreatment prior to measurement is not necessary. Sensors made of chalcogenide glass and PVC-based polymer membranes with enhanced cross-sensitivity can be incorporated into a sensor array. A number of conventional ISEs, e.g., pH glass electrode, sodium- and chloride-selective electrodes, may also be included in the system. An array of sensors made of ultrathin films of polymers and composite films of several polymers are also used in the e-tongue system. Such films are deposited on top of a glass substrate that holds interdigitized microelectrodes. Sensors prepared from different materials produce different electric responses and their variation allows a "fingerprint" of the samples. Like e-nose, the e-tongue also consists of hardware and software components. The hardware is used for the capacitance measurements of sensorial units and the software controls the data acquisition, perform the calculations and analyze the electrical signals. The main hardware components are: signal generator; signal amplifier; multiplexer; data acquisition board and a lap-top computer. The software component deals with electrical signals and provides the capacitance values, which are stored into files using a pre-defined format. The software interface allows the definition of parameters to control the data acquisition. The data analysis is similar to that in case of the e-nose.

17.3.2 Data processing and pattern recognition

Data processing is the second most important part of an E-Tongue. Since the number of sensors in the array of an E-Tongue may reach 40, each of them may produce a complex response in the multicomponent environment and relevant multidimensional data processing will be a stupendous task. This can be accomplished by different pattern recognition methods such as Artificial Neural Networks (ANNs) or multivariate calibration tools. The various pattern-recognition analysis modules for evaluating the data may include principal component analysis (PCA), discrimination function analysis (DFA), Soft Independent Model Clam Analogy (SIMCA), and Partial Least Square (PLS). The modules of the E-Tongue are the same and/or very similar to those used for the E-Nose.

17.3.3 Applications in the food industry

The wide range of applications of the e-tongue in dairy and food processing industry are:

- Continuous control on product quality
- Detection of pollutants in water (environmental applications)
- Detection of analytes in low concentration solutions– difficult to be distinguished by human being or even impossible.
- Quantitative analysis and recognition (identification, classification) of a very wide range of liquids
- Quality control and identification of the conformity to standards for different food stuffs-juices, coffee, beer, wine, spirits, etc.
- Taste quality of a food to be monitored continuously from the raw material stage right through to final product.

17.4 Robotics

Robotics is a concept designed at creating a machine equivalent to humans. It was the famous Czech writer Karel Capek who coined the word robot. The word robot comes from the word robota, meaning "drudgery" or "hard work" in modern. Robot Institute of America (1979) defined a robot as a reprogrammable, multifunctional manipulator designed to move material, parts, tools or specialized devices through variable programmed motions for the performance of a variety of tasks. Robotics is a system that comprises of sensors, control systems, manipulators, power supplies and software all working together to perform a task. Designing, building, programming and testing a robot involves combining frontier areas of different sciences viz., physics, mechanical engineering, electrical engineering, structural engineering, mathematics and computing. In some cases biology, medicine, chemistry might also get involved.

A robot has the following essential characteristics: A robot is to be first able to sense its surroundings. It could accomplish this task by first using its sensors: light sensors (eyes), touch and pressure sensors (hands), chemical sensors (nose), hearing and sonar sensors (ears), and taste sensors (tongue). A robot then needs to be able to move around its environment. It could achieve this by rolling on wheels, walking on legs or propelling by thrusters. The robot can either move whole of itself or just parts of it (arm, neck etc). A robot can be powered through either solar energy or electrical energy. A robot also needs intelligence which can be supported by appropriate software programs.

17.4.1 Application of robots in dairy and food processing operations

Robots are finding applications in many food processing activities because of their versatile features, some of which are being listed below. In high speed repetitive operations in modern day food processing and packaging where human operators begin to show onset of fatigue thereby increasing risk of mental errors causing quality or hygiene problems. Furthermore, in food processing operations, where long durations of performing repetitive tasks may lead to error or injury thereby leading to loss of quality and work.

17.4.2 Features of robots

Features	Benefits
Better process control	Easy to clean robot, minimum retention areas, connection protection
High reliability, high speed	Increased productivity
High dexterity, several mounting positions	Compact cell, less room required, simpler mechanical solution
Cleanliness	Better hygiene
Flexibility	Marketing innovative products and packaging
Vision and conveyor tracking	Product picked and controlled in process, in any position

17.4.3 Some of the areas of robot applications are

- Precision curd slicing and ultrasonic cheese slicing
- Food packaging and stacking operations
- The sophisticated control system with a built-in palletizing function makes it possible to load and unload the objects with precision and accuracy.
- Robotics are used to reduce the chances of contamination and consequently eliminate many safety concerns associated with human contact..

****** ☺ ******

Lesson 18

ROLE OF COMPUTERS IN OPTIMIZATION

18.1 Introduction

This lesson will discuss the concept of Operations Research (OR), computer oriented algorithmic approaches, queuing systems and waiting models, and applications of OR in dairy industry. The topics discussed in this lesson will be useful to the students for understanding the importance of OR in decisions making and solving the complex problems.

The formal beginning of operational research as discipline was initiated in England during World War II when a team of British scientists set out to make decisions regarding the best utilization of war material. In the decades after the war, the techniques began to be applied more widely to problems in business, industry and society to improve efficiency and productivity. Since that time, operational research has expanded into a field widely used in industries ranging from petrochemicals to airlines, finance, logistics, dairy plants and government, moving to a focus on the development of mathematical models that can be used to analyse and optimize complex systems. Today, OR is a dominant and indispensable decision making tool and has become an area of active academic and industrial research. This subject is also known as Industrial engineering.

Operations research or Operational research (also referred to as decision science or management science) is an interdisciplinary mathematical science that encompasses a wide range of problem-solving techniques and methods applied to improve decision-making and efficiency. Some of the tools used by operations researchers are statistics, graph theory, probability theory, decision science, queuing theory, linear programming, mathematical optimization, simulation, and other stochastic-process models. Almost all of these techniques involve the construction of mathematical models that attempt to describe the system. Though, the solution of mathematical modelling provides a basis for making decision, intangible factors such as human behaviour which cannot be quantified, must be taken into account before reaching to final decision. Because of the computational and statistical/ mathematical nature of most of these fields, OR also has strong ties to computer science and analytics. Researchers try to determine the most appropriate techniques suitable for a given system, goals for improvement, and constraints on time and computing power.

The major sub-disciplines in modern operational research are:

- Computing and information technologies
- Environment, energy, and natural resources

- Financial and marketing Engineering
- Manufacturing, service sciences, and supply chain management
- Policy modelling and public sector work
- Revenue management
- Simulation
- Stochastic models
- Transportation

18.2 Operations Research

Operations research is a discipline that deals with the application of advanced analytical methods and information technology to make better decisions. The complex problem is decomposed into basic components and solved with mathematical analysis by employing analytical methods from interdisciplinary mathematical sciences, such as multi-criteria decision analysis, network analysis, queuing theory, statistical analysis, optimization, conjoint analysis, neural networks, etc., to arrive at optimal or near-optimal solutions to complex decision-making problems. Psychology and management sciences also play important role in final decision making process. The goal is to provide rational bases for decision making to improve the system performance by understanding the complexities involved in the system. Non managerial level personals are also benefitted from the improved and streamed decision making process. OR techniques applied for solving problems can be broadly classified into following three basic groups of methods. Within each of the three basic groups many probabilistic methods provide the ability to assess risk and uncertainty factors.

- **Optimization methods:** The goal is to enable the decision maker to search best choice based on certain criteria's from the available various possible choices in an efficient and effective manner.
- **Simulation method:** These methods are used to observe the real systems. It is used to estimate the measures of performance of modelled systems and provides the ability to conduct sensitivity analysis to search for improvements.
- **Data analysis method:** The goal is to aid the decision maker in finding actual patterns, forecasting, and inter connections in the data set.

Major steps involved in solving a problems using OR tools are:

I. Problem Definition

II. Construction of mathematical model

III. Solution of model

IV. Validation of model

V. Implementation of the solution

Some of the benefits offered by OR are:

- Decrease cost or investment
- Increase revenue or return on investment
- Manage and reduce risks
- Improve quality of services
- Increase throughput while decreasing delays
- Improve utilization from limited resources

18.3 Role of computers in optimization

There is no single technique that can solve all mathematical models in OR. The type and complexity of mathematical models used for solution depends on the nature of problem. Sometimes models are simple and straight forward involving few steps to solve the problem. But in practical situation, the models are complex and involve lot of computations. A large number of OR tools are available to deal with different kind of problems and complexities, a few commonly used techniques are:

- **Linear programming** techniques are used to find optimal solutions for models with strict linear objectives and constraints functions from the limited resources available.
- **Network programming** techniques are used where the problem can be modelled as a network. These techniques are used to find shortest path, project scheduling, computing critical path, PERT/CPM/CPS, etc.
- **Queuing and simulation** techniques are used to study the behaviour of real-time systems. These techniques measure the performance of systems such as average waiting time in queues, average waiting time for services and utilization of services.

The distinguish characteristics of most OR techniques is that solutions are not generally obtained in closed form just like a one or two line formula instead these are determined in several step by using algorithms. An algorithm provides fixed computational rules that are applied repetitively to the problem with each repetition (called iteration) moving the solution closer to optimum. Because the computations associated with each iteration are typically tedious and voluminous, it is imperative that these algorithms are executed on the computers. Moreover, some mathematical models are so complex such as simulation models that it is just impossible to solve them manually. In these situations, it has become necessary to convert the algorithms into computer programs to find solutions in efficient and convenient way. Therefore, use of a digital computer has become an

integral part of the OR approach for decision making. The computers are required due to the complexity of the model, volume of data required and the computations to be made.

Specialized software packages have been developed by experts for solving general OR problems as well particular problems related to specific domain. General purpose software such as SAS, MATLAB, MATHMATICA, EXCEL, etc., have module for solving OR problems. One of such computer software for academic purpose is TORA and can also be used for commercial and industrial real time applications. This is Window based software distributed along with operation research book (Hamdy A. Taha). TORA offers solutions for almost all mathematical models used in OR for problem solving such as matrix inversion, solutions of simultaneous equations, linear programming, transportation models, network models, queuing models, project planning with CPM and PERT, integer programming, non-linear programming, etc. It is easy to operate and contains self explanatory instructions and menus. It can be operated in tutorial or automated mode. Tutorial mode is particularly useful for beginners.

18.4 Queuing systems and waiting models

Waiting for services is a part of our life since the servers/ facilities providing the services are limited and in flow of persons in search of services are more. Waiting in lines to obtain services is most difficult and it seems like wastage of our precious time. We wait to eat in restaurants, to withdraw/ deposit amount in banks/ ATM counter etc. Similarly lactating cows wait for milking at milking parlour. Jobs wait for processing by machines, congestion in computer networks, etc. In general a queue is formed when either units requiring services (commonly referred as customers) wait for service or the service facilities, stand idle and wait for customers. Unfortunately, we cannot eliminate waiting without increasing service facilities which, incurs extra ordinary cost on the system. We can only try to reduce the adverse impact to tolerable levels. Queuing theory has applications in diverse filed such as computing, telecommunications, dairy industry, designing of factories, shops, milk collection centres, hospitals etc.

Mathematical study of waiting lines or queues is called queuing theory. It enables mathematical analysis of several related processes of queue such as arriving, waiting, and being served at the front of the queue. The theory, deals with quantifying the phenomenon of waiting in lines using representative measures of performance such as average queue length, average waiting time in

queue, and average facility utilization. Various analytical models are used to study queuing system which, help the managers to design cost effective service facilities. The objective of queuing analysis is to evaluate the service and the cost of a facility so as to maximize its usefulness. This result in minimizing the total cost associated with the idle time to facilities or services versus the waiting time costs of employees or customers.

18.4.1 Types of queuing systems

As the number of queues and service facilities, there are four variations in the queuing system as given below:

- Single queue – Single server
- Single queue – Multiple servers (either in parallel or series)
- Multiple queues – Single server
- Multiple queues – Multiple servers

18.4.2 Elements of a queuing system

The queuing process mainly revolves around the **customers** requiring services and **server** which provide one or more service facilities. Customers are generated from a source which may be from **finite** or **infinite** population. On the arrival at the facility, they can start service immediately or wait in a queue if the facility is busy. When a facility completes a service, it automatically takes next customer waiting in queue. If there is no customer in the queue (i.e. queue is empty), the facility becomes idle until a new customer arrives.

From analysis point of view, arrival process is represented by the **interarrival time** between successive customers, and the service is described by the**service time** per customer. Generally the interarrival and service times can be probabilistic as number of calls in telephone exchange or deterministic as the arrival of cows for milking in a milk parlour.

Queue size plays a role in analysing the queues. It may have finite size or infinite as in mailing order facilities. **Queue discipline** is an important factor in the analysis of queuing models. It represents the order in which customers are selected from queue. The most common disciplines are First-Come-First-Served (FCFS), Last-Come-First-Served (LCFS) and Service in Random Order (SIRO). Customers may also be selected from the queue based on some order of **priority**.

The **queuing behavior** of customers plays a role in waiting line analysis. Customers may **jockey** from one queue to another in the hope of reducing their waiting time. They may also **balk** from joining the queue because of long waiting time or they may **renege** from a queue because they have been waiting too long.

Queuing system takes into account of the whole situation considering from arrival of the customer into the system till exit of the customer. The time in system is generally taken to be the queuing plus service time. The variations in the elements of a queuing situation give rise to a variety of queuing models.

18.4.3 Interarrival and service time distributions

In most situations the arrival of customers occurs in a totally random fashion. Randomness means that the occurrence of an event (e.g., arrival of customer or completion of service is not influenced by the length of time that has elapsed since the occurrence of the last event. The random interarrival and service time are approximated by one of the well known mathematical distributions such as:

- **Poisson distribution**: The density function of Poisson probability distribution is given by:

$$P_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t} \text{ for } n \geq 0$$

- **Exponential distribution**: The density function of exponential distribution is given by:

$$P(t) = \lambda e^{-\lambda t};\ 0 < t < \infty$$

Where,

n = Number of customers in the system (waiting and in service)

$P_n(t)$ = Transient state probability that n calling units are in the queuing system at time t

P_n = Steady state probability of having n units in the system

λ = Average (expected) customer arrival rate or average number of arrivals per unit of time in the queuing system

μ = Average (expected) service rate or average number of customers served per unit time at the place of service

e = Constant, equal to 2.71828

Note: Above functions can also used for service time, in that case λ is replaced with μ.

18.4.4 Queuing Models

A queuing model is used to approximate a real queuing situation or system, so the queuing behavior can be analyzed mathematically. Queuing models allow a number of useful steady state performance measures to be determined, including:

- the average number in the queue, or the system,
- the average time spent in the queue, or the system,
- the statistical distribution of those numbers or times,
- the probability the queue is full, or empty, and
- the probability of finding the system in a particular state.

Generally queuing model can be specified by the symbolic representation (a|b|c) : (d|e|f) using Kendall's notation. The meaning of symbols used in representing queuing model and the standard notations used in place of these symbols are described in table 18.1 given below:

Table 18.1 Symbols used queuing model and their standard notations

Meaning of symbols		Standard notations used in place of symbols
a :	Arrival distribution (or inter-arrival)	• M for a Markovian (poisson, exponential) distribution • E_k for an Erlang distribution with $_k$ phases • D for constant (or deterministic) time • GI for general distribution of interarrival time • G for general distribution of service time
b :	Service time (departure) distribution	
c :	Number of parallel servers or service stations	Finite or infinite (any positive number 1, 2, 3, …, or)
d :	Queue discipline	FCFS, LCFS, SIRO, GD (general discipline i.e. any type of discipline)
e	Capacity of the system	Finite or infinite (any positive number 1, 2, 3, …, or)
f :	Size of the calling source	Finite or infinite (any positive number 1, 2, 3, …, or)

Examples:

i. (M|M|1): (FCFS|∞|∞) indicates a queuing system when the inter-arrival times and service times are exponentially distributed having one server in the system with first come first served discipline and the number of customers allowed in the system i.e. capacity of the system can be infinite as well as size of the calling source.

ii. (M|D|5): (GD|20|∞) indicates a queuing system where the inter-arrival times is poisson (or Exponential) distribution, constant service time, and 5 parallel server in the system. Queue discipline is GD (i.e. general discipline), and there is a limit of 20 customers on the entire system. The size of the source from which customers arrive is infinite.

18.5 Applications of OR in dairy industry

A large number of mathematical models are used in OR to solve problems related to optimization, simulation, network programming, forecasting and data analysis. Some prominent techniques such as linear programming, queuing theory, simulation, etc., are used extensively in dairy industry for efficient management of dairy operations. Following are few common applications of these techniques in dairy industry:

18.5.1 Linear programming applications

Linear programming technique is used to find the optimum solutions for given number of constraints and limited resources. This technique has numerous applications in dairy industry as explained below with real time examples.

A. A dairy plant is interested in developing a least cost ice-cream mix. The ingredients to be used for making ice-cream are cow milk, skim milk powder (SMP) roller, white butter, sugar and gelatin. With these ingredients an ice-cream mix of 100 kgs. may be prepared having a minimum of 10 per cent fat and 10.5 per cent SNF. Similarly the minimum quantity of sugar required is 15 kgs. And that of gelatin should be exactly equal to 0.5 kg. The following table gives the fat and SNF content in cow milk, skim milk powder and white butter as well as net cost per unit of all the ingredients.

	Cow milk	SMP roller	W. Butter	Sugar	Gelatin
Fat %	4.5	1	84	-	-
SNF %	8.8	94.5	1.5	-	-
Net cost(Rs/Kg)	15.00	110.00	150.00	20.00	200.0

Formulate a linear programming problem to minimize the cost of ice-cream mix. Also develop the matrix of coefficients for solving LP problem using L.P. software package.

B. A farm manager is interested in developing a least cost concentrate mixture for their milch stock. The ingredients to be used for formulation of concentrate mixture are ground nut cake (GNC), mustard cake (MC), wheat bran (WB), rice bran (RB), maize (MZ), barley (BL), mineral mixture (MM), salt (SL). With these ingredients a concentrate mixture of 100 kgs. with minimum 70% TDN and 14% DCP is to be prepared which should contain a maximum of 25 kgs. of GNC, 10 kgs. MC, 15 kgs. WB, and 10 kgs RB. However the minimum requirement of maize and barley in the concentrate mixture is 30 kgs each. The quantity of mineral mixture and salt should be exactly equal to 2 kgs and 1 kg respectively. The following table gives the TDN and DCP composition of various feed ingredients as well as net cost per unit for all the ingredients.

Sl. No.	Ingredients	Nutritive value/kg DM		Price (Rs/Kg)
		TDN	DCP	
1	G.N.Cake	0.78	0.40	9.50
2	Mustard Cake	0.79	0.28	7.90
3	Wheat Bran	0.72	0.10	4.85
4	Rice Bran	0.68	0.054	4.00
5	Maize	0.80	0.08	6.00
6	Barley	0.75	0.075	5.70
7	Mineral Mixture	-	-	10.00
8	Salt	-	-	3.00

Formulate a linear programming problem to minimize the cost of concentrate mixture. Also prepare a matrix to solve this L.P. problem through LP software package.

18.5.2 Queuing theory applications

Queuing theory has numerous applications in dairy industry for the situations where waiting time is to be reduced for improving the efficiency of system A few real time applications of queuing theory in dairy industry are:

A. The author S. Pimpicki, through his research work has demonstrated the application of queuing theory for reducing the vehicle waiting time for raw milk collection by a dairy. A method is

presented for defining and solving practical problems of delivery of raw materials to food processing factories, based on the queuing theory. The method is demonstrated by the example of raw milk collection by a dairy (optimization of number of milk reception points so as to reduce vehicle waiting time while at the same time achieving max. utilization of reception capacity). A monogram is given for calculating average queuing time as a function of number of reception points and movement rate. The hypothesis that vehicle arrival follows a Poisson distribution was verified.

{Source: Pimpicki, S., "Application of queueing theory to organization of supply of raw agricultural materials illustrated by a dairy industry example", Zeszyty Naukowe Akademii Rolniczo-Technicznej w Olsztynie No. 148 (Ekonomika 2) , 1975, 3-12.}

B. The author Piotr Cegiełła et. al., in an another study have beautifully shown the applications of queuing theory in dairy industry. This article deals with an example of the real problem of the Vehicle Collection optimization, which enables the plants to carry on their production. The problem has been solved in Poland in the delivery of milk to dairy plants. The milk collection optimization, however, is not a simple inversion of dairy product deliveries in town. Therefore we have the Vehicle Collection Problem (VCP) as an extension of the Vehicle Scheduling Problem (VSP). Next we present the idea of the Multicolmilk procedure, which satisfied several technical and organizational constraints of the milk collection process. To solve the VCP we have developed a modular Collection Optimization System (Colos). The basic component of the Colos is an optimization module based on the Multicolmilk procedure. The last part of the article contains implementation results in the dairy plants of Konin, Gniezno,Zielona Góra, and Cz¸estochowa, together with the presentation of difficulties which have cropped up in the course of preparation and implementation processes.

{Source: Piotr Cegiełła, Wacław Szymanowski, and Jerzy Prokuratorski, "Colos: A collection optimization system as an application of the vehicle scheduling problem", European Journal of Operational Research, Volume 27, Issue 1, 1986, Pages 82–90}

C. Another important application is how to reduce the waiting of milking for lactating cows standing in queue for milking. Students can think and try to solve this problem.

****** ☺ ******

Appendix A. System analysis for milk procurement and billing system

CASE STUDY - I

SYSTEM ANALYSIS FOR MILK PROCUREMENT AND BILLING SYSTEM

A.1 Introduction

The case study for preparing system analysis report for milk procurement and billing system of a dairy plant is discussed in this section. This sample report may be used for preparing reports for other case studies like Milk marketing system, Milk product manufacturing system. These topics will be useful to understand the minor issues involved in preparing report on system analysis.

A report prepared by system analyst on the system study and analysis should have following contents:

Contents

1. Problem Definition and Introduction.
2. Introductory Investigation.
3. System Analysis.
   - 3.1 Detailed Investigation
     - 3.11 Organizational Chart.
     - 3.12 Review the present system.
     - 3.13 Analyze Inputs.
     - 3.14 Review methods, procedures and data communication.
     - 3.15 Analyze Outputs.
     - 3.16 Model the existing logical system.
     - 3.17 Other analysis and considerations.
   - 3.2 Concluding Investigation
     - 3.21 Context diagram of MPBS.
     - 3.22 Expected volume of data.
     - 3.23 Feasibility study.
       - 3.231 New MIS alternatives.
       - 3.232 Advantages, disadvantages and cost comparison.
       - 3.233 Comparative statement.
   - 3.3 Final Survey Report for top management
     - 3.31 MPBS in brief.
     - 3.32 Major Problems of MPBS.
     - 3.33 Redefined Objectives.
     - 3.34 Main Benefits expected - qualified and quantitative.
     - 3.35 Recommendation.

A.2 Problem Definition

A dairy plant is procuring milk from the nearby villages through a well established network of societies/ collection centers in the radius of 120-200km. Milk producers are paid for the milk based on the prices decided by the management. The management of dairy plant is interested in developing general purpose computerized milk procurement and billing system considering all the procedures of milk procurement and billing.

A.2.1 Problem introduction

A large number of dairy plants are operating in the business of milk procurement and manufacturing milk products. Some of these are owned by the federations/societies registered at state/district level and some are managed by private personnel.

Success of a dairy plant primarily depends on a well panned and organized system of milk procurement. In cases where procurement system is not well established, dairy plants remain underutilized.

The dairy plant gets more than sufficient milk during the flush season (October to March) and it generally fails to meet the demand of the consumers during the lean season (April to September). Hence the management of the dairy has to take up measures to procure sufficient quantity of milk through-out the year. A well planned milk procurement system has the following advantages:

I. An assured market to the milk producers' through-out the year at reasonable price.
II. Full capacity utilization of the dairy plant.
III. Increase in the milk production and collection.

The cost of procurement of milk is sometimes as high as 10 to 15% of the cost of its production. For an efficient determination of this component of cost, the use of computer has been found imperative. Further a quick billing system for the procurement of milk from different agencies has been possible with the employment of computer for maintenance of various accounting systems.

A.2.2 Organizing an MIS project

The task was given to MIS section of the dairy plant and asked to prepare a report for improving the efficiency in present manual method of milk procurement & billing system.

The management of dairy plant established an MIS executive steering committee, consisting of heads of Procurement section, Account & bill section and General Manager (operations). The committee will be responsible for computerizing the activity on behalf of management. The committee will review priorities, objective of the milk procurement and billing systems and will be giving directions to the concerned users of such a system. The committee will also coordinate with the MIS task force, consisting of In-charge of MIS section and two system analysts. The MIS task force is to study the feasibility of upgrading and computerizing the current milk procurement and billing system.

A.3 Introductory Investigation

A.3.1 Defining the scope of MIS project

The MIS task force interviewed various officials like MD of plant, chief milk procurement officer, and field staff of procurement section, supervisor of chilling centre, chairman and employees of collection

centre/society. MIS task force also visited some of the societies/collection centers and discussed the possible improvement in the system. In particular the task force was especially interested in identifying the effective and ineffective aspects of the existing system. The main questions asked are as under:

i). Why a computer based system being is considered?
ii). What are the basic objectives of the proposed system?
iii). What is the volume of data to be processed?
iv). What is the frequency of processing?

The task force also discussed with the members of the society/collection centers about the kind of services they expected. All the members were satisfied with the concept and kind of activities performed by societies/centers. But they were unhappy over the unusual delay in payments and in providing technical input services. Moreover sometimes mistakes/errors occur in payments and measurement of quantity and quality of milk. Presently the payment is made to members thrice in a month i.e. upto 10th, 20th and 30th day of the month. Further the plant takes at least one week to process the data manually and the payment is done about 8 or 9 days after the due date and sometimes even later than that one more complaint made by the members, supervisors of societies/collection centers and chilling centers was that the management takes a lot of time in solving the disputes over incorrect payments and other irregularities. The same is time with the providing of technical input aids/services. These problems are causing dissatisfaction among the members and discouraging the milk procurement activities.

MIS task force held several rounds of meetings with MIS executive committee and users to present the facts they had gathered. At this time it is difficult to work out the monetary benefit but definitely the computerization of milk procurement and billing system will increase the membership of societies/collection centers which in turn will increase the revenue of the plant. The most important factor is the improvement in the quality of services given to the members. Together, they developed a set of objectives for computerization of existing system as described below. These objectives are thought to be very challenging, but of course attainable.

- Providing better services to the members of society/collection centers.
- Improving the payment system.
- Monitoring the performance of milk procurement.
- Monitoring technical input services programme.
- Providing timely information to the management.
- Enhancing the milk procurement.

A.3.2 Schedule of MIS

The MIS executive steering committee strongly emphasized the importance of this project in improving the efficiency of the system and urged the MIS task force to set a completion date as early as possible. The MIS task force in turn prepared a list of the major activities to be performed during the system analysis and system design phases and estimated the weeks required to complete each task as shown in fig. A.1 using Gantt chart. MIS task force concluded that the system analysis study could be completed by two system analysts in about seven weeks.

Steps	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Introductory Investigation	<->																
System Analysis																	
Detailed Investigation		<---------->															
Concluding Investigation				<------->													
Final survey report						<-->											
System Specifications							<------>										
System design																	
Input design(data files)									<---------->								
Output design										<---------->							
Input menus(modules)												<---->					
Programming & testing												<---------->					
Equipment selection										<---------->							

Fig. A.1 Activities involved in analyzing and developing MPBS

A.3.3 Presentation of result of introductory investigation

After reviewing the system with user group, the MIS task force and the MIS executive steering committee met the managing director of plant to present the result of the introductory investigation. The main focus of the presentation was that poor services to the members of societies/collection centre led to the decrease in the strength of membership which is causing direct financial losses to the plant. The MIS task force then explained the concept of integrated milk procurement from society/collection centers, transportation to main dairy dock yard, processing of bills, monitoring of technical input services. They also told that with reduction of time gap to 1/2 days or few hours for payment will satisfy the members and will be effective in increasing the membership of society/collection centers.

The MIS task force also presented the time schedule to complete this work in most efficient way, beginning with complete system analysis of present system, summarized in a study report for presenting to top management for reviewing and approval, system designing and implementation.

The managing director was satisfied with the presentation and gave his approval to go ahead but also asked them to go strictly by the time schedule to complete this project. He further asked the committee members to design the system in such a way that in future it can be integrated with processing of milk, inventory of finished milk products, marketing and distribution of products etc. He further recommended that written communication may be sent to all the concerned staff members regarding ***project description*** and to help them in conducting the system analysis of milk procurement and billing system. The project description is given below:

A.4 Project Description

A.4.1 Project title

System analysis study of milk procurement and billing systems (MPBS).

A.4.2 Project Objectives

- To identify the problems faced by staff and members of the society/collection centre.
- To evaluate the economic and technical feasibility of a new milk procurement and billing system.

A.4.3 Approach to be taken

- Interview the concerned staff and members of society/collection centre.
- Analyze inputs, methods, procedures, data registers/files, outputs and internal controls of the current system.
- Identify and evaluate alternative systems.
- Prepare recommendation and present Final survey report to top management.

A.4.4 Potential system benefits

- Reduce the time and percentage of error in bill processing, this improvement in customer service may enable the plant to increase the quantity of milk procurement.
- Timely generation of MIS reports will help manager to take immediate decision.
- Computerization will improve work efficiency.
- Provide an efficient MPBS as basic module to an integrated network system encompassing billing, monitoring of technical input services, marketing, processing, Inventory of finished milk products etc.

A.4.5 Schedule and efforts required

- Start date.
- Expected completion date.
- This project will require two full time system analysts for seven weeks.

Approved by: Managing director.

A.5 Structured Systems Analysis

As indicated in introductory investigation, the managing director desires to have new milk procurement and billing system integrated with other areas. Basically, the MPBS would be designed to interface effectively with milk processing, finished milk product inventory, marketing of products, inventory of engineering spare parts etc. These systems must take second priority since they require data from the MPBS. Though these systems could be partially designed concurrent with the design of MPBS, but the current number of MIS department personnel is insufficient to undertake the simultaneous development of all systems. Overall, the new MPBS is to be the initial major module or building block for adding new integrated system later.

A structured approach to systems analysis of present MPBS involves analysis the highest-level data processing activities at central level as well as detailed activities at the society/collection centre and

chilling centre level. This means that analyze all processing activities that have bearing on MPBS. Exceptions, irregularities and problems in procurement and billing are also noted for analysis. In turn milk procurement and bill processing activities are analyzed in a top-down manner that permits functional decomposition of procurement and bill processing modules to the lowest level.

A functional decomposition approach allows system analysts to examine both the functional activity and data aspects of procurement and billing system from the highest level to the lowest level. This approach provides a means of identifying inputs, constraining factors, and control information which affect procurement processing activities and outputs. As such functional decomposition provides a means for the system analysts to understand the present milk procurement and billing system. Functional decomposition diagram and context level data flow diagram representing the existing milk procurement and billing system is shown in figure 11.3 and 11.5 respectively in lesson 11. Finally, functional decomposition provides an effective basis for comparing the new MPBS system to the present one. This approach, then, plays an integral part in the detailed and concluding investigation of system analysis.

A.5.1 Detailed investigation of present MPBS

In order, for the MIS task force to comprehend what become apparent in the milk procurement and billing system, there is need for a detailed investigation using a structured approach. This includes the following modules:

- Review the present system and analyze inputs.
- Review methods, procedures and data communications.
- Analyze outputs.
- Model the existing logical system.
- Undertake other analysis and considerations.

Each module is covered in a sufficiently comprehensive manner to enable any knowledgeable reader to envision the scope of the milk procurement and billing system and the relationship of its detailed parts.

A.5.2 Organizational structure

The plant is managed by one managing director for its day to day activities. He is assisted by a team of directors (Technical Director, Procurement Director, Marketing Director and Administrative Director). For technical matters the plan is looked after by a chairman appointed by the board of directors of the plant. An organizational chart of dairy plant can be drawn.

A.5.3 Review the present system

Can be attached as Annexure.

A.5.4 Analyze inputs

The input data is generated at four levels as described below :

A.5.4.1 Level I : Society/collection centre

I. Society/collection centre.
II. Membership data.
III. Requirement of stationary/equipment/chemicals etc.
IV. Requirement of technical input aids (indents). .

V. Data of milk quality & quantity received from members.
VI. Distribution of technical input material.
VII. Beneficiaries' details.
VIII. Truck sheet data (i.e. composite milk statement) dispatched to chilling centre/main dairy.
IX. Expenditure data on society/collection centre.
X. Revenue generated through the sale of technical input services.

A.5.4.2 Level II : Chilling centre

I. Statement of milk received (include quantity & quality parameters).
II. Statement of milk despatched.
III. Expenditure incurred on chilling centre.
IV. Transportation data.

A.5.4.3 Level III : Main dairy

I. Statement of milk received at dock yard.
II. Expenditure data for maintaining the procurement section.
III. Expenditure on technical input services.
IV. Transportation data.
V. Expenditure on equipment/chemicals etc.

A.5.4.4 Level IV : Management

I. Milk and transportation rates are decided by the management.
II. Other policy matters.

A.5.5 Methods, procedures and data/document transfer communication

Major source of transferring data/documents between society/collection centre and main dairy is the transport vehicle. All kind of requirements indents, statements are passed on to main dairy through vehicle at the time of collection of milk in morning/evening from society/collection centre. Description of document flow is as follows.

The milk compiled statements of milk received at collection centre/society is transferred to chilling centre. If the society/collection centre is nearby main dairy then these can be passed directly to main dairy. The indents for technical input services required and expenditure statements are also passed on to main dairy. From the main dairy dockyard milk is passed on to processing section and other statements are passed on to milk procurement section. Milk statements are sent to account and bill section where the draft alongwith details are generated and drafts are issued for the society/chilling centre, those who have supplied the milk at dock yard. The societies which supplied the milk at chilling centre, are paid by chilling centre based on the milk received at chilling centre. Further the farmers are paid by the societies/collection centers. The milk rate is decided by dairy management. The indents for technical input services are processed by procurement section and material is supplied to the societies/collection centers. The daily statement of milk received, procurement cost analysis, and transportation cost analysis etc. is also prepared as and when required by managing director/management.

A.5.6 Analysis of output -Various outputs generated at different levels are as follows:

A.5.6.1 Level 1 : Account and bill Section

i. Payment detail for chilling centre & collection centre/societies
No. & Name of society/collection centre/chilling centre : Date :

Date	Time	Milk type	Quality	Qty	Fat%	SNF%	Rate/litre	Amount
.								
.								
Total								

Total Fat in kg =
Total SNF in kg =

ii. Bills for private contractors (if any) in same format as above.

A.5.6.2 Level 2 : Main dairy reception dock yard

i. Daily statement of milk received : Date :

Date	Time	Chilling Centre No.	Society/Collection Centre No.	Quality	Qty	Fat%	SNF%
.							
.							
Average							
Total							

A.5.6.3 Level 3 : Procurement section

i). Total milk received route wise in a specific period.

Milk procured route wise for the period:_____to _____

Route No.	Type of milk	Fresh Qty	Fresh Fat%	Fresh SNF%	Sour Qty	Sour Fat%	Sour SNF%	Curdled Qty	Curdled Fat%	Curdled SNF%	Total Qty	Total Fat%	Total SNF%
.													
.													
Total													

ii). Cost analysis statement for the period_____to _____

Route No.	Quantity	Fat%	SNF%	Procurement cost/lit	Transportation centre/lit	Chilling cost. lit
.						
.						
Total						

iii). Summery statement of important activities performed under technical input aid program in a given period

Activity Name : Cattle feed/fodder Seed/Vet. Medicine/AIS

Soc./C.C. No.	Name	No. of beneficiaries	Qty/No.	Amount	Subsidiary Given
.					
.					
Total					

iv). Bills for transport contractors

Transport contractor name:

Period :_____to_____

Vehicle No. : Vehicle name :

Date	Time	Route No,	Distance travelled	Amount
.				
.				
Total				

A.5.6.4 Level IV : Chilling centre

Bills for societies/CC in the same format as in level 1.

A.5.6.5 Level V : Society/collection centre

Payment details for members same format at level 1.

A.5.7 Model of the existing logical system

Based upon detailed investigation of existing milk procurement and billing system just completed, students are advised to draw data flow diagram. It should be noted that neither the equipment nor the type of manual and computer processing are shown in this diagram. The total concentration should be on the logical flow of data. All data needed in above processing steps serve as a basis for developing model of the system and data dictionary.

A.5.8 Undertake other analysis and consideration

The dairy plant is in a sound financial condition. It is noted for its strong management and its high profit. It is highly regarded in the society and is procuring more than 50% of milk production from the procurement area. Overall, the plant committee management is strongly committed to help the company move ahead. Because the plant has sufficient resources, it can adopt a new technology to remain in the competition with the other plants. Therefore, the Managing Director and MIS steering committee have agreed to computerize the activity without going into the details of exact or estimated returns from this project (in terms of amount). Just to improve the quality of services for its members of the societies/collection centers.

A.6 Concluding Investigations

A.6.1 Context diagram of MPBS

The working of entire system is depicted in one data flow diagram, i.e., context diagram in fig. 6.4. (in lesson 6). This diagram gives little detail and shows only the overview of the system. Detailed DFD can be drawn further for each component.

A.6.2 Volume of data

The volume of data varies according to the area covered for milk procurement and the number of farmers supplying the milk to the society/collection centers which varies from time to time. The plant under study is procuring milk from approximately 900-1000 societies/collection centers and each society/collection centre has 200-500 members and milk is procured once a day in the morning. Thus the number of transactions per day will be up to 500,000 per day, which is a very high volume application. Beside this there will be thousands of entries related to expenditure on technical input aids, number of AIs performed, detail of sale of fodder seeds, cattle feed, details of collection centers/societies, chilling centers and transport vehicles, etc.

A.6.3 Feasibility study of new MPBS

Now that each area of the present MPBS has been carefully analyzed, a feasible set of MIS alternatives must be developed in order to select the best one we will examine the following area before selecting the best feasible MIS solution:

A.6.4 Financial feasibility study

After discussing the matter with MD, it was found that dairy in good financial health therefore finance will not be any problem at all. The management wants a solution to be effective and beneficial to milk producers.

A.6.5 Social feasibility study

During discussions held with sectional heads of procurement and accounts, they shown keen interest in developing and implementing new system and extended their full support. Users of the system like operational staff at doc yard, field staff at society level and other were cordial and ready to contribute in developing and implementing the system since new system will save lot of time in preparing bills and reducing errors.

A.6.6 Technical feasibility study

Based on an analysis of the existing logical system of the MPBS and available hardware technology, MIS task force has identified the following basic MIS alternatives:

A.6.7 Centralized data processing

One powerful computer server with latest specifications with 8-10 terminals with windows network system connected through LAN, and high speed line/ laser printer will be installed in the main dairy plant. All data records will be transferred to the main computer centre from collection centre. Data entry & further processing will be made on the main computer system and bills may be prepared and distributed to society for further distribution to members. One/ two automatic milk collection units will also be required at central dairy to automatically feed the quantity and quality parameters into computer.

- **Cost :** The cost will be around Rs. 10-11 lakh including high speed line printer.
  Cost involved in employing trained man power.

A.6.8 Distributed data processing with microcomputers at the user end

All societies may be given an automatic milk collection system equipped with milk analyzer, online weighing balance connected with PC and GUI based software installed on PC to carry out all activities ate collection centre. In this case, data (Member ID, Quantity, Fat%, SNF%, etc) is entered automatically without human intervention. Price of milk will also computed automatically based on quality parameters and bills can be generated immediately. All processing will made at society level. In this case, at central level bills will be generated only for society/ collection centers not for individual members.

- **Cost** One complete automatic milk collection units costs around 1.5 Lakhs which includes computer and printer and necessary software
- Benefits and disadvantages of both approaches are discussed in table A.1
- Comparison of current manual system with computerized systems on certain parameters is shown in table A.2.

Table A.1 Benefits / disadvantages and cost of each feasible MIS alternatives

Type of Systems	Benefits	Disadvantages
Centralized data processing	Limited trained man power is required to complete the job.	Duplication of work at society & chilling centre level.
	Supervision of work is easy at one place.	Data communications.
	Maintenance of CS will not be difficult at one place.	More powerful computer system is required at the centre.

	More applications of dairy plant may also be taken up.	If the main computer is down then whole work will be stopped
Distributed data processing	Each society will have full control over activities to be performed at society level	Costly solution
	Data will be entered automatically and bill can be generated immediately	Trained manpower required at each society
	This will reduce the burden of data entry at central level.	Some societies may not be able to perform up to the mark.

Table A.2 Comparison table of current MPBS with feasible MIS alternatives

Sl. No.	Options	Current System	Centralized data Processing	Distributed data processing
1	Time taken in bill generation(days)	8-9	2-3	Same day
2	Data communication	Transport	Transport	Not required
3	% of errors in payment statement	5	<5	<5
4	Time taken in solving disputes over payments(days)	7	2	Immediately
5	Trained manpower required (This includes system analysts, two programmers, two console operators, 8-10 data entry operators)	Nil	15	Nil
6	Capability of processing data if main computer is down	Not Applicable	No	Yes
7	Capability of providing wide range of facilities	No	Yes	Yes
8	Initial cost (in lakhs)	Nil	10-12	1.5 Lakhs for each society
9	Software development	Nil	In house	Ready made

A.7 Final Survey Report

The MIS task force with the assistance of the MIS user group prepared final survey report summarizing their findings to the MIS executive steering committee.

A.7.1 MPBS in brief

The dairy plant procures the milk from the societies or by setting up collection centers at the village level. Some private agents also supply the milk directly at dock yard. The milk is tested for quality and quantity at collection centre, chilling centre and at the dock yard of main dairy plant. Route network is designed by the procurement officer to collect the milk from societies/collection centers and transport vehicle is hired on contract basis to transport milk. Payment is made to members according to details entered in the pass book at the time of milk collection. Similarly societies/collection centers and chilling centers are paid on the statements of milk received at chilling centre/main dairy. If there is delay in transporting the milk then suitable deduction is made from his payments.

Dairy plant also supplies the subsidized cattle feed, fodder seed, veterinary aid, A.I. assistance and other technical input services/aids to enhance the milk production. Periodically the management is interested to know the cost of milk procurement/ transportation/ chilling per liter of milk, daily average milk procured route wise, average Fat% and SNF%, handling losses etc. The management is also interested to know the expenditure incurred on technical input services and the no. of beneficiaries. Bills are also to be generated periodically for payment to farmers against the milk supplied. The whole process is very time consuming and complex.

A.7.2 Major problems found in the system

The major problems found in the present system are described as follows:

- First major requirement of MPBS is to reduce the time period in generating bills by improving internal processing of data. This will also help in reducing errors in bill.
- Second major problem is that of errors in testing of milk at the society/collection centre level, chilling centre level and main dairy dock yard.
- Third major problem is data communication. Physically the data is moved through the transport vehicle. Therefore the data reaches the destination either at the end of day or on the next day. So a lot of time is wasted in data communication.
- Fourth major problem is time taken in solving the disputed cases of payment among societies, chilling centers and main dairy plant.

A.7.3 Redefined objectives

- The basic objectives of the new milk procurement and billing system will be :
- Timely generation of bills and payment to milk producers.
- Timely solution of disputed cases over payments.
- To determine handling losses.
- Cost analysis to monitor the performance of milk procurement.
- Monitoring of technical input services.

A.7.4 Main benefit expected from new MPBS

- Error free and prompt generation of bills for payment to milk producers.
- Prompt reply to queries regarding daily procurement of milk.
- Ability to analyze the procurement, chilling and transportation cost etc.
- Easy determination of procurement losses.
- Ability to calculate expenditure incurred on technical input services/aids.
- Easy to calculate the commission paid to societies.
- Prompt reply to the queries of management from time to time.
- Improvement in the quality of service.
- Better quality work by producing legible outputs.
- Ability to reprocess the data without much extra efforts.

A.7.5 Recommendations

- The report concluded with the following recommendations for management approval:
- Immediately computerize the activity to at society level to generate bills well in time with minimum errors.

- Authorize the MIS task force to proceed with the design phase of new MPBS,
- Develop new with distributed data processing.
- Outputs may be designed in consultations with users and officials.
- Continue to examine the feasibility of other related fields of dairy plant like milk processing, inventory of finished milk product, marketing of milk etc.
- Automatic machines may be provided to all the societies/collection centres to weigh and test the quality of milk.

Upon listening to an oral presentation of the foregoing recommendations by MIS task force, MIS executive steering committee concurred with all recommended items. The executive committed observed that there is a lot of difference in the cost of feasible MIS alternative cases. But the efficiency wise distributed processing system is more effective.

MIS task force next met with the managing director to present their final report. Having been kept informed of the project status by the MIS executive committee, he was well prepared to discuss and review their recommendations. He was very much satisfied with the progress made by the MIS executive steering committee and MIS task force in the project. He also thanked all the members for taking keen interest. Further he assured that within one week you will get the final approval to proceed with the next step.

****** ☺ ******

CASE STUDY - II

SYSTEM ANALYSIS FOR INVENTORY CONTROL

B.1 Problem Definition

A dairy plant is processing 100000 liter milk per day and its capacity is expected to increase. A wide range of milk products are being manufactured by the plant. A number of machines are installed in different sections for manufacturing products. To keep all machines in running state, inventory of large number of engineering spare parts are being maintained in the plant by store section. The basic purpose is to provide related spare part at the time of need to avoid longer breakdown of machines. The term INVENTORY is defined as "the systematic control and regulation of purchase, storage and usage of materials in such a way so as to maintain an even flow of production and at the same time avoiding excessive investment in inventories". The store manager is interested to computerize the operations of inventory to keep track on items available in the store.

B.2 Introductory Investigation

The system analyst interviewed various officials like manager of the store, dealing hand, and users to discuss present way of working and possible improvement in the system. During the discussions it was found that first user has to place a request on preprinted form to store manager through the concerned head of the section. Depending on the availability, item will be issued to user otherwise NOC is issued. Quantity of item issued is subtracted from the balance and when item is received from the supplier quantity is added in the balance. A special ledger is maintained to records issue and receipt entries of items. When the particular item reaches below reorder level procurement procedure starts to replenish the item. ABC analysis of inventory items is performed on regular basis which is time consuming process. The management also finds fast moving and slow items from time to time so that sufficient stock of fast moving items can be maintained and slow moving items can be reduced from the inventory. To facilitate inventory management, ABC analysis classifies the inventory items into following three classes based on the consumption value:

I. Class "A": These items constitute the most important class of inventories so far as the proportion in the total value of inventory .The "A" items consists of approximately 10% of the total inventory items, accounts for 70% of the total consumption value of items. These items need a tightly controlled inventory system with constant attention to the purchase & stores management. A larger effort per item on only a few items will cost only moderately, but the effort can result in larger savings.

II. Class "B": These items constitute an intermediate position, which constitute approximately 20% of the total items, accounts for approximately 20% of the total material consumption value. These items merit a formalized inventory system & periodic attention but the purchase & the stores management.

III. Class "C": It consists remaining 70% items, accounting only 10% of the monetary value of total material usage. Quite relaxed inventory procedures are used. Low cost items that are typically commercially available and require little management control.

Some of the problems faced in existing manual system are:

i. A lot of paper work
ii. Occasionally register entries found misplaced
iii. Items have been ordered prior to the reorder level
iv. Orders have not been placed well in time therefore items run out of stock
v. Items issued are more than the available items in stock.
vi. Difficult to check the availability of items in store.
vii. Not able to perform ABC analysis regularly.
viii. Difficulty in finding fast and slow moving items.

The main concern of the store manager was that it is becoming difficult to manage large number of spare parts. He was also interested to know the availability of quantity in stock at the click of mouse. At the same time he wants to reduce the inventory cost as plant is spending large amount on inventory. After, thorough discussions with the concerned users a set of objectives for computerization of existing system were developed as described below:

I. Providing better services to users and inventory managers.
II. To reduce inventory cost avoiding under stocking.
III. Efficient management of items in stock.

B.2.1 Expected outputs from the system

I. Current balance of item in stock.
II. List of items below reorder level.
III. Item wise cost of inventory.
IV. List of spoiled items.
V. List of items received and issued within a specific period.
VI. Ledger preparations as required by management
VII. ABC analysis of inventory items.
VIII. Placement of orders for replenishment of items

B.2.2 Input requirements

I. Details of Inventory items
II. Details of items received
III. Details of items issued
IV. Details of issue form
V. Details of items ordered for procurement
VI. Details of potential vendors'

B.2.3 Methods and procedure

Computing Current Balance: *Current balance = Current Balance + Receipt - Issue*
Computing Reorder Level: *Reorder Level = DailyDemand * ProcurementTime*
DailyDemand = YearlyDemand / *No.of WorkingDays*
Procurement time is generally constant

B.2.4 Model the existing logical system

Based upon detailed investigation of the present manual system, modular chat or functional decomposition diagram of the existing system is drawn in figure B.1 as given below. It should be noted that neither the equipment nor the type of manual and computer processing are shown in this diagram. The total concentration is on the logical flow of data. All data needed in the above processing steps serve as a basis for developing model of the system and data dictionary.

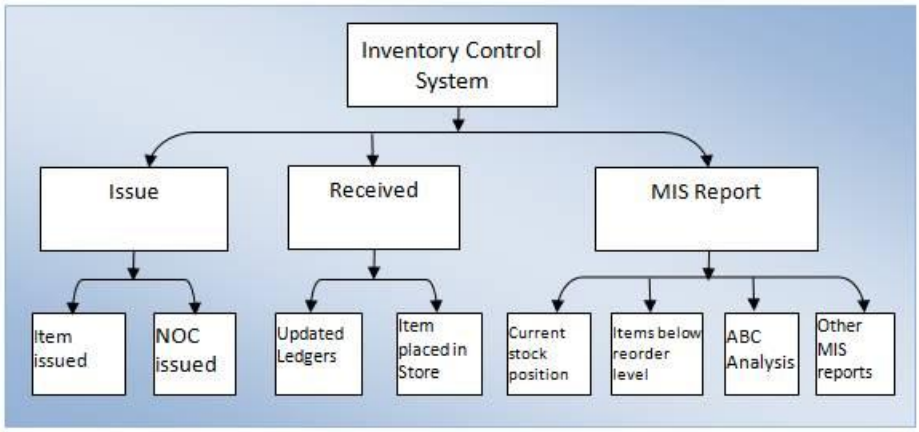


Fig. B.1 Functional decomposition of components of inventory control system.

****** ☺ ******

CASE STUDY - III

DATABASE DESIGN FOR MILK PROCUREMENT AND BILLING SYSTEM

C.1 Description of the problem

A Model dairy plant is procuring milk from milk shed area near the plant. Dairy plant has established a number of milk collection centers in the area. Each collection center is covered by a specified route. Milk is collected only in the morning and is directly transported to dairy plant. Only fresh milk of either buffalo or cow is accepted and payment is made on the basis of Fat% and SNF% content of milk. Collection centers are maintained by dairy plant and various technical inputs in the form of veterinary aid, artificial insemination facilities, feed subsidy, etc. are provided to the societies for smooth collection of milk. Management is interested to computerize the milk procurement and billing system in order to bring efficiency in the system and to provide better services to societies/ members. Following reports are generated for the societies and management of the plant:

1. Milk collection report for each collection center.
2. Bill for each collection center for given period.
3. Summary report of total milk procured and Transportation cost route wise.
4. Milk collection cost collection center wise for given period.
5. Milk procurement cost route wise and overall procurement cost. Procurement cost includes all kind of expenditures incurred to maintain collection centers and transportation.

After studying the system and in order to produce above mentioned reports, various input data has been collected and placed in different files/tables as given below:

Table C.1: Collection Centers

	Field Name	Data Type	Field Size / Format	Description/ Validation Checks
*	ccno	Text	3	Collection Centre Number
	ccName	Text	20	Collection centre name
	inchargeName	Text	20	Incharge / Supervisor name
	ccAddress	Text	20	Collection centre address

Table C.2: Expenditure Details

	Field Name	Data Type	Field Size/ Format	Description/ Validation Checks
*	ccno	Text	3	Values for this field should be taken from

				Table **Collection Centers**
*	expenditureDate	Date/Time	Short Date	Date of Expenditure Incurred
	salary	Number	Integer	Salary of Employee
	maintenance	Number	Integer	Expenditure on chemicals, consumables, fixed cost, etc.
	vetaid	Number	Integer	Expenditure on veterinary aid
	ai	Number	Integer	Expenditure on artificial insemination
	feedSubsidy	Number	Integer	Expenditure on feed subsidy given
	extensionActivity	Number	Integer	Expenditure on ext. activity like educational shows, trade fairs, visits etc.
	otherExpenditure	Number	Integer	Other Expenditure on unforeseen activities

Table C.3: Milk Receipts

	Field Name	**Data Type**	**Field Size/ Format**	**Description/ Validation Checks**
*	ccno	Text	3	Values for this field should be taken from Table **Collection Centers**
*	receiptDate	Date	Short Date	Higher than 1/1/2012
*	milkType	Text	2	CM- Cow Milk; BM-Buffalo Milk
	quantity	Number	Integer	Between 0 to 2000, Default value 0
	fat	Number	Single (Decimal: 1)	Between 3.0 equal to 10.0
	snf	Number	Decimal (1 digit)	Between 7.0 equal to 12.0
	qualityNormal	Logical	Yes/ No	Quality of milk is normal or not (Yes/No)

Table C.4: Route Details

	Field Name	**Data Type**	**Field Size/ Format**	**Description/ Validation Checks**
*	routeno	Text	2	Route number
	description	Text	20	Route description
	rate	Number	Single (Decimal: 2)	Running expenditure per km for running a vehicle on the route

Table C.5: Distance Covered				
	Field Name	**Data Type**	**Field Size/ Format**	**Description/ Validation Checks**
*	routeno	Text	2	Values for this field should be taken from table**Route Details**
*	rtDate	Date/Time	Short Date	Date of transporting milk. Transportation date be checked from milk receipt date.
	distance	Number	Integer	Distance traveled. It cannot be more than 100 km on single route in one day.

Table C.6: Route CC				
	Field Name	**Data Type**	**Field Size/ Format**	**Description/ Validation Checks**
*	routeno	Text	2	Values for this field should be taken from table **Route Details**
*	ccno	Text	3	Collection centre falling on the specific route. Values for this field should exist in table **Collection Centers**

1. Design a database using above mentioned tables with proper labels for each input data and following validation checks:
   a) Collection centre and Route numbers should be valid numbers i.e. before entering these numbers in transaction files they must exist in master files.
   b) Type of milk should be either BM or CM.
   c) Fat% and SNF%, rate per kilometer and distance traveled must be in a specific range.
2. Define suitable relationship among the created tables based on key field.
3. Create forms for data entry for each table. Use the concept of sub-form if needed for display of data/ data entry.
4. Enter data in tables for at least 15 days of the receipt of milk from 7 societies distributed on 3 routes. There is one entry of expenditure in a month for each society.
5. Sort data (from milk receipt table) on collection centre number, date of receipt of milk etc.
6. Filter data (from milk receipt table) by selection/ form for a particular collection centre number or date of receipt of milk.
7. Create following queries for retrieving data from above tables:

a) Ddisplay data on the fields **ccno, inchargeName, receiptDate, milkType, quantity, fat, snf, *fat-kg, snf-kg*** for given dates and/or ccno.

b) Display data on fields **ccno, ccName, inchargeName**, **routeNo,** and **description** from ***Collection Centers*** and ***Route Details*** tables for which In-charge name starts with alphabet "R". Arrange the In-charge names alphabetically (i.e. ascending order) in the output.

c) Summation/ summary of milk received Collection Center wise.

d) Summation/ summary of milk received route wise.

e) Total expenditure report collection center wise with center name, address and expenditure date for given period.

f) Cross tabulate **quantity of milk** received by collection centre wise and type of milk.

g) Create new table with fields **ccno, ccname, routeno, route description** from existing tables.

h) Create a delete query for deleting records from milk receipt table for a given ccno.

i) Update the field **ratekm** with **ratekm +2.0** for all routes in *Route Details* table.

8. Generate the above said reports with suitable formats.

****** ☺ ******

CASE STUDY - IV

DATABASE DESIGN FOR MILK MARKETING SYSTEM

D.1 Problem description

Management of a dairy plant is interested to computerize its milk marketing operations to bring efficiency in the system. After discussion with the plant management and milk marketing section in-charge, a report of system investigation prepared by system analyst is given below:

Plant is manufacturing and selling 10 types of milk products in addition to liquid milk of two types namely full cream milk and toned milk. There are two types of sale outlets one owned and maintained by employees of the plant and others maintained by agents on commission basis. The agents have to deposit some token amount as security money in advance to the plant for running sale outlets. Few sale outlets are selling only liquid milk while others are selling liquid milk as well as milk products. Milk and products are dispatched to sale outlets daily in the morning through plant owned vehicles. Unsold liquid milk is collected back from sale outlets for reprocessing. Vehicles operate on different specified routes defined by plant covering all sale outlets. Milk market section produces following reports for management of the plant:

I. Daily sale proceeds of liquid milk and milk products route wise and sale outlet wise.
II. Commission to be paid to agents.
III. Product wise sale proceeds in a given time period.

As a database designer, designed a suitable and effective database to meet the present and future requirements of the plant. A sample of database design for milk marketing section is given below considering the requirement of dairy plant. The rest process is same as followed in previous case study.

Table D.1: Detail of sale point employees

Fields	Data type	Data size
Sempno	Text	3
Sempname	Text	50
Amntdipos	Number	Integer
Empstatus	Yes/No	Logical
Empaddrs1	Text	50
Empaddrs2	Text	20
Empphone	Text	20

Table D.2: Product information

Pcode	Text	3
Pname	Text	30
Prate	Number	Real 2 decimal place
Pcommission	Number	Integer

Table D.3: Route information

Rcode	Text	3
Rdesc	Text	50

Table D.4 Sale point information

Spcode	Text	3
Sempno	Text	3
Spname	Text	50
Spstatus	Yes/No	Logical
Prdsale Products	Number	Type of sale 1-Milk; 2-Milk
Spadd fields)	Text	100 (can be broken into 2/3
Spphone	Text	20

Table D.5: Route and sale point interaction

Rcode	Text	3
Spcode	Text	3

Table 27.12 Detail of per day sale

Saledt	Date	Short
Spcode	Text	3
Pcode	Text	3
Qtysupplied	Number	Integer
Qtyreturned	Number	Integer
Qtysold	Number	Integer

Students are expected to identify the primary key for each table, data validation checks for different attributes, reports, queries etc. themselves to make this database workable and useful to users.