ALAGAPPA UNIVERSITY

[Accredited with 'A+' Grade by NAAC (CGPA:3.64) in the Third Cycle
and Graded as Category–I University by MHRD-UGC]

(A State University Established by the Government of Tamil Nadu)

KARAIKUDI – 630 003

Directorate of Distance Education

BA [Public Administration]

VI - Semester

106 61

COMPUTER AND DATA PROCESSING

Authors:

Sanjay Saxena, Managing Director, Total Synergy Consulting Pvt Ltd., New Delhi
Units: (1.0-1.2, 1.5-1.6, 2.0-2.2)

S. Mohan Naidu, Academic Author
Units: (5.5, 6.0-6.2)

Vikas Publishing House: Units: (1.3-1.4, 1.7-1.11, 2.3-2.10, 3, 4, 5.0-5.4, 5.6-5.10, 6.3-6.10, 7-14)

VIKAS® PUBLISHING HOUSE PVT. LTD.
E-28, Sector-8, Noida - 201301 (UP)
Phone: 0120-4078900 • Fax: 0120-4078999
Regd. Office: A-27, 2nd Floor, Mohan Co-operative Industrial Estate, New Delhi 1100 44
• Website: www.vikaspublishing.com • Email: helpline@vikaspublishing.com

Work Order No.AU/DDE/DE12-27/Preparation and Printing of Course Materials/2020 Dated 12.08.2020 Copies - 500

SYLLABI-BOOK MAPPING TABLE

Computer and Data Processing

Syllabi	Mapping in Book
BLOCK I: FUNDAMENTALS OF COMPUTER SYSTEM **Unit-1** Computer System Fundamentals: Hardware and Software - Evolutions of Computers - Classification. **Unit-2** Basic Components of a Digital Computer - Internal and Auxiliary Storages - Remote Data Entry Devices. **Unit-3** Characteristics of Internal Storage - Auxiliary Storage - Processing Methods - Batch, Real-Time and Time Shared Processing.	**Unit 1:** Basics of Computer **(Pages 1-31);** **Unit 2:** Components of a Digital Computer **(Pages 32-52);** **Unit 3:** Internal Storage and Processing Methods **(Pages 53-73)**
BLOCK II: DATA PROCESSING SYSTEM **Unit-4** Data Processing Systems and Tools: Types of Data Processing System - Extent of Data Processing Systems - Data Processing Cycle **Unit-5** Components of a Data Processing Systems - Problem Definition - Planning - Algorithms. **Unit-6** Program Flowcharts - Decision Table - Top-Down Programming Techniques - Structure Programming. **Unit-7** Batch Processing: A Typical Batch Processing Application - Master File - Transaction File - File Update.	**Unit 4:** Fundamentals of Data Processing Tools **(Pages 74-86);** **Unit 5:** Data Processing Components **(Pages 87-122);** **Unit 6:** Flowcharts and Programming Techniques **(Pages 123-158);** **Unit 7:** Batch Processing: Techniques and Applications **(Pages 159-168)**
BLOCK III: DATA STORAGE **Unit-8** Direct Access Storage and Retrieval - File Organization Techniques - Report Generation. **Unit-9** Examples of Flow Charts - Programs for the Functions. **Unit-10** Applications: Inventory Control and Accounting - Payroll - Production, Planning and Control.	**Unit 8:** Direct Access Storage, Retrieval and File Organization **(Pages 169-181);** **Unit 9:** Flowcharts and Programs **(Pages 182-195);** **Unit 10:** Applications: Inventory Control and Accounting **(Pages 196-211)**
BLOCK IV: ONLINE PROCESSING AND CONTROL **Unit-11** Online Processing Controls - Examples of Specific Online Applications. **Unit-12** Air Line Reservation - Railway Reservation.	**Unit 11:** Online Applications **(Pages 212-218);** **Unit 12:** Online Reservation Systems **(Pages 219-225)**
BLOCK V: MANAGEMENT OF BUSINESS APPLICATIONS **Unit-13** Management of Stores - Query Package. **Unit-14** Real Time Business Applications - On Line Business Transactions.	**Unit 13:** Store Management and Query Package **(Pages 226-245);** **Unit 14:** Online Business Transactions **(Pages 246-256)**

CONTENTS

BLOCK II: DATA PROCESSING SYSTEM

INTRODUCTION

NOTES

A computer is a machine that manipulates data according to a set of instructions called a computer program. The program has an executable form that the computer can use directly to execute the instructions.

Data processing is defined as, 'The collection and manipulation of items of data to produce meaningful information'. In this sense it can be considered a subset of information processing, 'The change (processing) of information in any manner detectable by an observer'. The term 'Data Processing (DP)' has also been used to refer to a department within an organization responsible for the operation of data processing applications.

Online transaction is a payment method in which the transfer of fund or money happens online over Electronic Fund Transfer (EFT). OnLine Transaction Process (OLTP) is secure and password protected. Three steps involved in the online transaction are Registration, Placing an Order, and Payment. Online transactions occur when a process of buying and selling takes place through the Internet. When a consumer purchases a product or a service online, he/she pays for it through online transaction.

This book, *Computers and Data Processing*, is divided into five blocks, which are further subdivided into fourteen units. This book provides a basic understanding of the subject and helps to grasp its fundamentals. In a nutshell, it explains various aspects, such as computer system fundamentals, hardware, software, basic components of a digital computer, internal and auxiliary storages, remote data entry devices, data processing methods, batch, real-time and time shared processing, data processing tools, types of data processing system, data processing cycle, components of a data processing system, algorithm, program flowcharts, top-down programming techniques, structured programming, batch processing, master file, transaction file, file organization techniques, report generation, flowcharts and programs, applications, inventory control and accounting, payroll, production planning and control, online processing controls, airline reservation, railway reservation, management of stores, query packages, real-time business applications, and on line business transactions.

The book follows the Self-Instructional Mode (SIM) wherein each unit begins with an 'Introduction' to the topic. The 'Objectives' are then outlined before going on to the presentation of the detailed content in a simple and structured format. 'Check Your Progress' questions are provided at regular intervals to test the student's understanding of the subject. 'Answers to Check Your Progress Questions', a 'Summary', a list of 'Key Words', and a set of 'Self-Assessment Questions and Exercises' are provided at the end of each unit for effective recapitulation.

NOTES

BLOCK - I
FUNDAMENTALS OF COMPUTER SYSTEM

UNIT 1 BASICS OF COMPUTER

Structure

1.0 INTRODUCTION

Computers are electronic devices that perform the basic operations of input, process, output and storage under the direction and control of a program. Computers have become an integral part of our lives. Most of the work that is done these days is performed by computers in some way or other. In this unit, you will learn about the functions of a computer system, hardware, software, their evolution and types.

In this unit, you will learn about the computer systems fundamentals, hardware and software, evolutions of computers, and classification of computer.

1.1 OBJECTIVES

After going through this unit, you will be able to:

- Define computer systems fundamentals
- Explain the concept of hardware and software
- Discuss the evolution of computers
- Understand the classification of computer

NOTES

1.2 COMPUTER SYSTEM FUNDAMENTALS

A computer is a programmable machine designed to automatically process a sequence of various arithmetic or logical operations. The interface between the computer and the human operator is known as the user interface. A computer consists of memory which stores information and data in the form of text, images and graphics, and audio and video files. CPU or Central Processing Unit performs the arithmetic and logic operations with the help of sequencing, and control unit that can change the order of operations based on the information that has been stored in memory. Peripheral devices allow information to be entered from an external source and allow the results of operations to be sent out. A Central Processing Unit or CPU executes a series of instructions to read, manipulate and store the data. The control unit, Arithmetic Logic Unit or ALU, memory registers and basic Input/Output or I/O devices are collectively known as a Central Processing Unit or CPU. Devices that provide input or output to the computer are known as peripherals. On a Personal Computer or PC, peripherals include input devices, such as the keyboard and mouse, and output devices, such as visual display unit or monitor and printer. Hard disk drives, floppy disk drives and optical disk drives serve as memory devices. A graphics processing unit is used to display 3-Dimensional or 3-D graphics. Modern desktop computers contain various smaller computers that assist the main CPU in performing I/O operations. Memory refers to the physical devices which are used to store programs, sequences of instructions or data, such as programs in a computer. Data is stored either in hard disk or in secondary memory devices, such as tape, magneticdisks and optical disks, Compact Disk Read Only Memory or CD-ROM and Digital Versatile/Video Disc or DVD-ROM. Memory is associated with addressable semiconductor memory, i.e., integrated circuits consisting of silicon based transistors.

Basic Functions of a Computer

There are three basic functions of a computer are as follows:

- **Data Processing:** A computer must be able to process data.
- **Data Storage:** A computer must be able to store data. Even if data is supplied to a computer on the fly, for processing and producing the result immediately, the computer must be able to store that data temporarily. Apart from short term data storage, it is equally important for a computer to perform a long term storage function to store different files.
- **Data Movement:** A computer must be able to move data between itself and the outside world. The computer operating environment consists of devices that serve as data sources or destinations. When data is received from or delivered to a machine that is directly linked to a computer, the process is known as input/output and the devices used for this purpose are referred as input/output devices. When data moves over longer distances to or from a remote machine the process is known as data communication.

Capabilities and Limitations of Computers

NOTES

The increasing popularity of the computer has proved that it is a powerful and useful tool. Its usefulness is due to its following features:

- **Speed**: Computers are very fast. They can process millions of instructions every second. The speed is related to the amount of data it processes and the time it takes to complete the processing task.
- **Storage**: Computers can store vast amounts of information in the form of files, which can be recalled at any time. These files help in easy and speedy retrieval of information. This type of storage is known as electronic storage system.
- **Accuracy**: In addition to being fast, computers are also accurate. The degree of accuracy for a particular computer depends upon its design. Most errors in computers are non-technical. Generally, programmers are responsible for these errors.
- **Diligence**: Computers are diligent as they can perform any complicated task accurately without making any error. Computers do not suffer from carelessness, boredom or tiredness. Moreover, their efficiency does not decrease with age.
- **Versatility**: Computers perform various tasks depending upon the instructions fed into them and their hardware characteristics. They are capable of performing any task, provided the task is reduced to a series of logical steps. A computer can be used to prepare a Word document and in between called to search for another document that is stored in its memory. It can perform both tasks simultaneously.
- **No IQ**: Computers do not have their own intelligence and their I.Q. (Intelligence Quotient) is zero. Hence, the user can and has to decide what tasks a computer should perform.
- **No Feelings**: Computers have no feelings because they are machines. They cannot make judgements as they process on the basis of a set of instructions, called programs, provided by the users.

Though computers can do better than human beings in terms of accuracy, speed and memory, but even then there are certain limitations of computer systems because they depend on human beings for their operations and functions. The following are some of the limitations of computers:

- Human beings program them for efficient, accurate and fast functioning.
- Computers cannot think intelligently and work independently like human beings.
- They follow instructions given by programs or by users.
- They can neither take decisions nor can correct wrong instructions.
- Programmers or users maintain and update them.
- As with many other modern appliances, computers also need electric power to run.

1.3 HARDWARE

NOTES

Computer hardware refers to the physical components that make up a computer system. They are the building blocks of personal computers. These are typically installed into a computer case, or attached to it by a cable or through a port. In the latter case, they are also referred to as peripherals. Various hardware components are as follows:

1. **Case:** A computer case (also known as a computer cabinet, tower, system unit or simply case) is the box that contains most of the components of a computer (excluding the monitor, keyboard and mouse).
2. **Power Supply Unit (PSU):** It converts AC electric power to low voltage DC power for the internal components of the computer. A power supply unit provides regulated power at the several voltages required by the motherboard and accessories, such as disk drives and cooling fans.
3. **Motherboard:** The motherboard is the main component inside the case. It connects the other parts of the computer including the CPU, the RAM, the disk drives (CD, DVD, hard disk, etc.) as well as any peripherals connected via the ports or the expansion slots. Components directly attached to the motherboard include:
    - **CPU:** It performs most of the calculations which enable a computer to function. It is also known as brain of the computer.
    - **Chipset:** It mediates communication between CPU and other components of system.
    - **RAM:** RAM stands for Random Access Memory. It stores the data and commands that is actively used by CPU.
    - **ROM:** ROM stands for Read Only Memory. It stores the BIOS (Basic Input Output System) instructions. It includes process of booting of the system.
    - **Buses:** Bus connects the CPU to other internal components and also to expansion cards slot for graphics and sound card.
    - **Ports**: It is used for connecting external peripherals.

Secondary Storage Devices

Computer data storage refers to computer components and recording media that retain digital data.

Fixed Media

- **Hard Disk Drives:** A Hard Disk Drive (HDD) is a device for storing and retrieving digital information, primarily computer data. It consists of one or more rigid rapidly rotating discs, coated with magnetic material and with magnetic heads arranged to write data to the surfaces and read it from them.

- **RAID Array Controller:** It is a device to manage several internal or external hard disks and optionally some peripherals in order to achieve performance or reliability improvement in what is called a RAID array.

Removable Media

- **Optical Disc Drives:** Optical disc drives are used for reading from and writing to various kinds of optical media, including Compact Discs, such as CD-ROMs, DVDs, DVD-RAMs and Blu-ray Discs. Optical discs are the most common way of transferring digital video, and are popular for data storage as well.
- **Floppy Disk Drives**: Floppy disk drives are used for reading and writing to floppy disks, an outdated storage media consisting of a thin disk of a flexible magnetic storage medium.
- **Zip Drives:** They are an outdated medium-capacity removable disk storage system, for reading from and writing to Zip disks.
- **USB:** USB flash drives are plugged into a USB port and do not require a separate drive. USB flash drive is a small, lightweight, removable and rewritable flash memory data storage device integrated with a USB interface.
- **Memory Card Readers:** They are used for reading from and writing to memory card. Memory card is a flash memory data storage device used to store digital information.
- **Tape Drives:** They are drives to read and write data on a magnetic tape and are used for long term storage and backups.

Check Your Progress

1. Define CPU.
2. List three basic functions of a computer.
3. What is the function of PSU?
4. What is RAID array controller?

1.4 SOFTWARE

A computer cannot operate without any instructions and is based on a logical sequence of instructions in order to perform a function. These instructions are known as a 'computer program', and constitute the computer software. The sequences of instructions are based on algorithms that provide the computer with instructions on how to perform a function. Thus, it is impossible for a computer to process without software, a term attributed to John W. Tukey in 1958.

Different kinds of software designs have been developed for particular functions. Popular computer software include interpreter, assembler, compiler,

NOTES

operating systems, networking, word processing, accounting, presentation, graphics, computer games, etc. The computer software is responsible for converting the instructions in a program into a machine language facilitating their execution.

Software engineers develop computer software depending on basic mathematical analysis and logical reasoning. Before implementation, the software undergoes a number of tests. Thus, the programming software allows you to develop the desired instruction sequences, whereas in the application software the instruction sequences are predefined. Computer software can function from only a few instructions to millions of instructions; for example, a word processor or a Web browser. Figure 1.1 shows how software interacts between user and computer system.

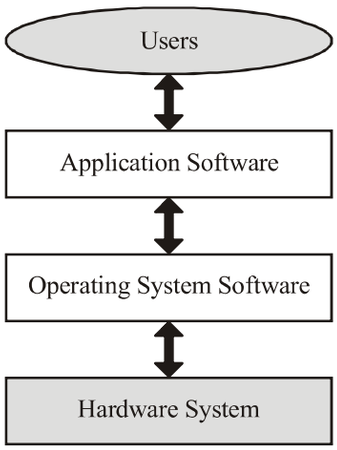


Fig. 1.1 *Interaction of Software between User and a Computer System*

On the functional basis, software is categorized as follows:

- **System Software:** It helps in the proper functioning of computer hardware. It includes device drivers, operating systems, servers and utilities.
- **Programming Software**: It provides tools to help a programmer in writing computer programs and software using various programming languages. It includes compilers, debuggers, interpreters, linkers, text editors and an Integrated Development Environment (IDE).
- **Application Software:** It helps the end users to complete one or more specific tasks. The specific applications include industrial automation, business software, computer games, telecommunications, databases, educational software, medical software and military software.

Types of Computer Software

Today, software is a significant aspect of almost all fields including business, education, medicine, etc. The basic requirement for software is a distinct set of procedures. Thus, software can be used in any domain that can be described in logical and related steps and every software is developed with the aim of catering to a particular objective, such as data processing, information sharing,

communication, etc. Software is based on the type of applications that are as follows:

- **System Software:** This type of software is involved managing and controlling the operations of a computer system. System software is a group of programs rather than one program and is responsible for using computer resources efficiently and effectively. Operating system, for example, is system software, which controls the hardware, manages memory and multitasking functions and acts as an interface between applications programs and the computer.
- **Real-Time Software:** This is based on observing, analysing and controlling real life events as they occur. Manually, a real-time system guarantees a response to an external event within a specified period of time. The real-time software, for instance, is used for navigation in which the computer must react to a steady flow of new information without interruption. Most defence organizations all over the world use real time software to control their military hardware.
- **Business Software:** This kind of software is functional in the domain of management and finance. The basic aspect of a business system comprises payroll, inventory, accounting and software that permits users to access relevant data from the database. These activities are usually performed with the help of specialized business software that facilitates efficient framework in the business operation and in management decisions.
- **Engineering and Scientific Software:** This software has developed as a significant tool used in the research and development of next generation technology. Applications, such as study of celestial bodies, study of undersurface activities and programming of orbital path for space shuttle, are heavily dependent on engineering and scientific software. This software is designed to perform precise calculations on complex numerical data that are obtained during real-time environment.
- **Artificial Intelligence (AI) Software:** Certain problem solving techniques are non-algorithmic in nature and primarily require this type of software. The solutions to such problems normally cannot be arrived at using computation or straightforward analysis. Such problems need particular problem solving techniques including expert system, pattern recognition and game playing. Also, it constitutes various kinds of searching techniques, such as the application of heuristics. The function of AI is to add certain degree of intelligence into the mechanical hardware to have the desired work done in an agile manner.
- **Web-based Software:** This category of software performs the function of an interface between the user and the Internet. There are various forms in which data is available online, such as text, audio or video format, linked with hyperlinks. For the retrieval of Web pages from the Internet a Web

NOTES

browser is used, which is a Web-based software. The software incorporates executable instructions written in special scripting languages, such as Common Gateway Interface (CGI) or Active Server Page (ASP). Apart from providing navigation on the Web, this software also supports additional features that are useful while surfing the Internet.

- **Personal Computer (PC) Software:** This software is primarily designed for personal use on a daily basis. The past few years have seen a marked increase in the personal computer software market from normal text editor to word processor and from simple paintbrush to advance image editing software. This software is used mostly in almost every field, whether it is database management system, financial accounting package or a multimedia based software. It has emerged as a versatile tool for daily life applications.

Software can also be classified in terms of the relationship between software users or software purchasers and software development.

- **Commercial Off-The-Shelf (COTS):** This comprises the software without any committed user before it is put up for sale. The software users have less or no contact with the vendor during development. It is sold through retail stores or distributed electronically. This software includes commonly used programs, such as word processors, spreadsheets, games, income tax programs, as well as software development tools, such as software testing tools and object modelling tools.
- **Customized or Bespoke:** This software is designed for a specific user, who is bound by some kind of formal contract. Software developed for an aircraft, for example, is usually done for a particular aircraft making company. They are not purchased 'off-the-shelf' like any word processing software.
- **Customized COTS:** In this classification, a user can enter into a contract with the software vendor to develop a COTS product for a special purpose, that is, software can be customized according to the needs of the user. Another growing trend is the development of COTS software components—the components that are purchased and used to develop new applications. The COTS software component vendors are essentially parts stores which are classified according to their application types. These types are listed as follows:
    - **Stand-Alone Software:** A software that resides on a single computer and does not interact with any other software installed in a different computer.
    - **Embedded Software:** A software that pertains to the part of unique application involving hardware like automobile controller.
    - **Real-Time Software:** In this type of software the Operations are executed within very short time limits, often microseconds, e.g., radar software in air traffic control system.
    - **Network Software:** In this type of software, software and its components interact across a network.

Figure 1.2 illustrates the various types of customized COTS.

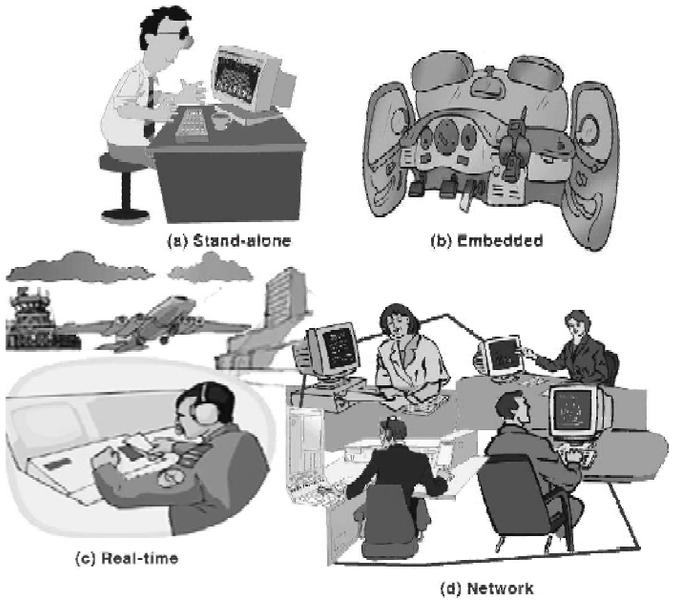


Fig. 1.2 *Types of Customized COTS*

System Software

System software constitutes all the programs, languages and documentation provided by the manufacturer in the computer. These programs provide the user with an access to the system so that he can communicate with the computer and write or develop his own programs. The software makes the machine user-friendly and makes an efficient use of the resources of the hardware. Systems software are permanent programs on a system and reduce the burden of the programmer as well as aid in maximum resource utilization. MS DOS (Microsoft Disk Operating System) was one of the most widely used systems software for IBM compatible microcomputers. Windows and its different versions are popular examples of systems software. Systems software are installed permanently on a computer system used on a daily basis.

Classification of an Operating System

An Operating System (OS) is the main control program for handling all other programs in a computer. The other programs, usually known as 'application programs', use the services provided by the OS through a well-defined Application Program Interface (API). Every computer necessarily requires some type of operating system that instructs the computer about operations and use other programs installed in the computer. The role of an OS in a computer is similar to the role of the manager in an office for the overall management of the college.

NOTES

Any computer system can be broadly classified in terms of four component dimensions:

(i) Hardware
(ii) Operating system
(iii) Application programs (like MS Word, Games, Calculator).
(iv) Users (people who work on the computer).

Figure 1.3 displays the various components of the computer system.

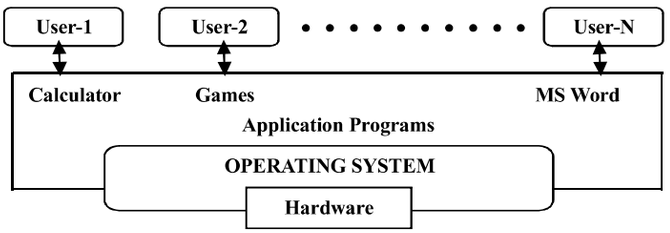


Fig. 1.3 *Components of a Computer System*

Application Software

Users install specific software programs based on their requirements; for instance, accounting software (like Tally) used in business organizations and designing software used by architects. All programs, languages and utility programs constitute software. With the help of these programs, users can design their own software based on individual preferences. Software programs aid in achieving efficient application of computer hardware and other resources.

1. Licensed Software

Although there is a large availability of open source or free software online, not all software available in the market is free for use. Some software falls under the category of **Commercial Off-The-Shelf (COTS).** COTS is a term used for software and hardware technology which is available to the general public for sale, license or lease. In other words, to use COTS software, you must pay its developer in one way or another.

Most of the application software available in the market need a software license for use.

Software is licensed in different categories. Some of these licenses are based on the number of unique users of the software while other licenses are based on the number of computers on which the software can be installed. A specific distinction between licenses would be an Organizational Software License, which grants an organization the right to distribute the software or application to a certain number of users or computers within the organization, and a Personal Software License which allows the purchaser of the application to use the software on his or her computer only.

NOTES

2. Free Domain Software

To understand this, let us distinguish between the commonly used terms Freeware and Free Domain software. The term 'freeware' has no clear accepted definition, but is commonly used for packages that permit redistribution but not modification. This means that their source code is not available. Free domain software is software that comes with permission for anyone to use, copy, and distribute, either verbatim or with modifications, either gratis or for a fee. In particular, this means that the source code must be available. Free domain software can be freely used, modified, and redistributed but with one restriction: the redistributed software must be distributed with the original terms of free use, modification and distribution. This is known as 'copyleft'. Free software is a matter of freedom, not price. Free software may be packaged and distributed for a fee. The 'Free' here refers to the ability of reusing it — modified or unmodified, as a part of another software package. The concept of free software is the brainchild of Richard Stallman, head of the GNU project. The best known example of free software is Linux, an operating system that is proposed as an alternative to Windows or other proprietary operating systems. Debian is an example of a distributor of a Linux package.

Free software should, therefore, not be confused with freeware, which is a term used for describing software that can be freely downloaded and used but which may contain restrictions for modification and reuse.

A few types of application programs that are widely accepted these days, are:

1. Word Processing

A word processor is an application program used for the production of any type of printable text document including composition, editing, formatting and printing. It takes the advantage of a Graphical User Interface (GUI) to present data in a required format. It can produce any arbitrary combination of images, graphics and text. Microsoft Word is the most widely used word processing system.

Microsoft Word can be used for the simplest to the most complex word processing applications. Using Word, you can write letters and reports, prepare bills and invoices, prepare office stationery, such as letterheads, envelopes and forms, design brochures, pamphlets, newsletters and magazines, etc.

2. Spreadsheet

Excel is ideal for a task that needs a number of lists, tables, financial calculations, analysis and graphs. Excel is good for organizing different kinds of data, however it is numerical data that is best suited. Thus, Excel can be used when you not only need a tool for storing and managing data, but also analysing and querying it. In addition to providing simple database capabilities, Excel also allows you to create documents for the World Wide Web (WWW).

The menus, toolbars and icons of MS Excel are very similar (though not the same) to MS Word. This is in keeping with Microsoft's much hyped philosophy

NOTES

and strategy of offering users a totally integrated office suite pack. From the user's point of view, this means less time spent in learning the second package once you know the first, and almost effortless and seamless exchange of data between various components.

3. Presentation Graphics

PowerPoint is a presentation tool that helps create eye-catching and effective presentations in a matter of minutes. A presentation comprises of individual slides arranged in a sequential manner. Normally, each slide covers a brief topic. The term 'Free' software specifies the freedom of using the software by various computer users (private individuals as well as organizations and companies) granting them freedom and control in running and adapting the computing and data processing as per their needs. The key objective of free software is to grant freedom rights to users so that the users are free to run, copy, distribute, study, change and improve the software. For example, you can use PowerPoint software for preparing presentations and adding notes to the specific slides. Similarly, you have the option of either printing the slides—in case you want to use an overhead projector—or simply attach your computer to an LCD display panel that enlarges the picture several times and shows the output on a screen.

You have three options for creating a new presentation:

(i) Begin by working with a wizard (called the **AutoContent Wizard**) that helps you determine the theme, contents and organization of your presentation by using a predefined outline, or

(ii) Start by picking out a PowerPoint **Design Template** which determines the presentation's colour scheme, fonts and other design features, or

(iii) Begin with a completely blank presentation with the colour scheme, fonts and other design features set to default values.

If you decide to choose the third option, PowerPoint designers have provided a wide assortment of predefined slide formats and Clip Art graphics libraries. Through these predefined slide formats, you can quickly create slides based on standard layouts and attributes.

PowerPoint shares a common look and feel with other MS Office components, and having once mastered Word and Excel, learning PowerPoint is almost like playing a game. And it is also easy to pick up data from Word and Excel directly into a PowerPoint presentation and vice versa.

Database Management Software

Nowadays, all large businesses require database management. When managing a large customer base, it is important to examine vital information like the busying pattern, cheap suppliers and the number of orders being received. In order to efficiently manage all these functions, MS Access is required.

NOTES

As a first step, plan and create your database structure, identifying the required fields based on the type of data (numbers, alphanumeric, data, etc.), and the maximum width of each field. After determining the structure, you can create a table either in the design mode (which is customized) or you can use the table wizard and any of the predefined tables, with the required modifications.

Creating the tables through the table wizard is much faster and easier than through the design mode. However, if you use wizards you are somewhat restricted with the predefined settings already available.

Once you have created the table you can then use the form's wizard to create user friendly and aesthetically pleasing layouts for data entry. Creating forms for data entry also ensures that the user inputs only the right kind of information and both data entry errors as well as typing work is minimized.

Once the forms have been created and relevant data has been entered, using these you can then use the report wizard to generate any kind of report. Using reports, you can not only organize and present your data in a more meaningful manner, but you can also use various standard functions like subtotals, totals, sorting to summarize your data.

Now to really fine-tune this Access application, you can create data access pages to enable people spread over a large geographical area to share and compile information using the Internet.

Computer Languages

A computer language essentially implies a language that is understandable to the computer. It is the computer's native language. Computer languages serve the same purpose as human languages. They are a means of communication. Let us understand the similarities and differences between computer languages and human languages.

Languages that we speak daily, such as English, Hindi, French or German are known as material or human languages. It constitutes words and rules known as lexicon and syntax, respectively. These words are joined to make meaningful phrases according to the rules of the syntax. A computer language also consists of lexicon and syntax, i.e., characters, symbols and rules of usage that allow the user to communicate with the computer.

The primary difference between a natural language and computer language is that natural languages have a large set of words (vocabulary) to choose from while computer languages have a limited or restricted set of words. Thus, fewer words but more rules characterize a computer language.

All problems to be solved by the computer needs to be broken down into discrete logical steps before the computer can execute them. The process of writing such instructions in a computer or programming language is called programming or coding.

NOTES

Since as computer hardware has improved over the years, programming languages have also moved from machine-oriented languages (that used strings of binary 0s and 1s) to problem-oriented languages (that use common English terms). All computer languages can, however, be classified under the following categories:

- Machine Language (First Generation Language).
- Assembly Language (Second Generation Language).
- High-Level Language (Third Generation Language).

Classification of Computer Languages

Computer languages are classified as follows:

The computer can understand only a binary-based language. This is a combination of 0s and 1s. Instructions written using sequences of 0s and 1s constitute a are known as machine language. First-generation computers used programs written in machine language.

A major drawback of machine language is that it is highly complex and difficult to use. Also, it consumes a lot of time and requires a substantial effort on the part of the programmer. Thousands of machine language instructions are needed to carry out simple tasks, such as listing a few addresses for mails. Any instruction in machine language is divided into two components:

(i) **Command:** Also called the 'operation code' or opcode including addition, multiplication, etc.

(ii) **Operand:** Refers to the address of the date on which the function has to be performed.

A general machine language instruction is presented as follows:

OP Code	Operand
001	010001110

The number of operands varies with each computer and is therefore computer dependent.

It can be concluded that in order to develop computer programs in machine language, the programmer will be required to remember a lot of operation codes and addresses of the data items based on the storage location and also information regarding the internal structure of the computer. Thus, using machine language can be highly complicated and liable to errors. Identifying these errors and introducing changes had become increasingly difficult leading programmers to seek better options.

Assembly Language

The development of assembly language marked the beginning of the evolution of programming languages. In assembly language mnemonics (symbolic codes) were used to present operation codes as well as strings of characters to represent addresses. Instructions in assembly language may appear as follows:

Opration	Opration address
READ	M
ADD	L

NOTES

Certain important facts about assembly language are as follows:

- Assembly language was designed to replace each machine code by an understandable mnemonic and each address with a simple alphanumeric string. It was matched to the processor structure of a particular computer and was therefore (once again) machine dependent. This meant that programs written for a particular computer model could not be executed on another one. In other words, an assembly language program lacked portability.
- A program written in assembly language needs to be translated into machine language before the computer can execute it. This is done by a special program called 'Assembler' which takes every assembly language program and translates it into its equivalent machine code.
- The assembly language program is known as the source program, while the equivalent machine language program is known as the object program. It may be useful to know that the assembler is a system program supplied by the computer manufacturer. Second-generation computers used assembly language.
- The lack of portability of programs (written using machine or assembly languages) between various computer systems led to the development of high level languages. Since they allowed a programmer to overlook a lot of low-level particulars of the hardware of the computer system, they were called high level language programs.
- It was obvious that if the syntax, mnemonics and rule and regulations of the programming language were closer to the natural language, it would be easier for the programmer to program and the lesser the possibility of introducing errors (or bugs) into the program. Hence, third generation languages, which were algorithmic and procedural, came into being in the mid-1950s. They were designed to solve a particular problem. They contained commands that are particularly suited to one type of application. For example, a number of languages were designed to process scientific or mathematical problems. Others emphasized on commercial applications. These languages varied very little between different computer systems, unlike machine or symbolic languages. But a complier or an interpreter program was required to translate these machine codes. Once again, the high level program is called the source code while its equivalent machine language program is referred to as the object code.
- Easy-to-learn feature, machine independence, easier maintenance and portability contributed to the popularity of high level languages. Slow program execution was the main disadvantage since programs needed to be converted

NOTES

into machine language (by an interpreter or a compiler) before they could be executed.

The assembly language program is known as the source program, while the equivalent machine language program is known as the object program. It may be useful to know that the assembler is a system program supplied by the computer manufacturer. Second generation computers used assembly language.

The lack of portability of programs (written using machine or assembly languages) between various computer systems led to the development of high level languages. Since they allowed a programmer to overlook a lot of low-level particulars of the hardware of the computer system, they were called high level language programs. And it was obvious that if the syntax, mnemonics and rule and regulations of the programming language were closer to the natural language, it would be easier for the programmer to program and the lesser the possibility of introducing errors (or bugs) into the program. Hence, third generation languages, which were algorithmic and procedural, came into being in the mid-1950s. They were designed to solve a particular problem. They contained commands that are particularly suited to one type of application, for example, a number of languages were designed to process scientific or mathematical problems. Others emphasized on commercial applications. These languages varied very little between different computer systems, unlike machine or symbolic languages. But a complier or an interpreter program was required to translate these machine codes. Once again, the high level program is called the source code while its equivalent machine language program is referred to as the object code.

Easy-to-learn features, machine independence, easier maintenance and portability contributed to the popularity of high level languages. Slow program execution was the main disadvantage since programs needed to be converted into machine language (by an interpreter or a compiler) before they could be executed.

High Level Languages

Some of the high level languages have been discussed as follows:

The third generation programming language (3GL) is a refinement of a second generation programming language. The 3GL made the languages more programmers friendly. High level language is a synonym for third generation programming language. First introduced in the late 1950s, FORTRAN (FORmula TRANslation), ALGOL (ALGOrithmic Language) and COBOL (COmmon Business Oriented Language) are early examples of 3GL. Most of popular and general purpose languages today, such as C, C++, C#, Java, BASIC and Pascal are also third generation languages. Most 3GLs support structured programming. The following are the examples of 3GL.

1. FORTRAN

FORTAN (FORmula TRANslation) was the first high -level language developed by John Backus at IBM in 1956.

FORTRAN has a number of versions with FORTRAN IV being one of the earlier popular versions. In 1977, the American National Standards Institute (ANSI) published standards for FORTRAN with a view to standardizing the form of the language used by manufacturers. This standardized version is called FORTRAN 77.

2. COBOL

COBOL (COmmon Business Oriented Language), the first language used for commercial applications, was developed under the leadership of Grace Hopper, a US Navy programmer, with a group of computer manufacturers and users in 1959. The maintenance and further growth of the language was handed over to a group called CODASYL (COnference on DAta SYstems Languages).

It is written using statements that resemble simple English and can be understood easily; for example, to add two numbers (stored in variables A and B), a simple statement in COBOL would be: ADD A TO B GIVING C.

COBOL was standardized by ANSI in 1968 and in 1974. COBOL became the most widely used programming language for business and data processing applications.

3. BASIC

BASIC (Beginner's All-purpose Symbolic Instruction Code) was developed as a teaching tool for undergraduate students in 1966 by John Kemeny and Thomas Kurtz, two professors at Dartmouth College. Eventually BASIC was used as the main language amongst the personal computer users.

A minimum version of BASIC was standardized by ANSI and is so simple that it has been incorporated in every subsequent version of BASIC. Some versions of BASIC include MBASIC (Microsoft BASIC) and CBASIC (Compiler based BASIC).

One of the newer versions of BASIC, commonly known as Visual Basic, has also evolved from the original BASIC language. It contains various statements and functions that can be used to create applications for a Windows or GUI environment.

4. PASCAL

PASCAL was designed by Nicholas Wirth, a Swiss professor, in 1971. It was developed as a more structured language used for teaching which Wirth named after the French mathematician Blaise Pascal, who also designed the first successful mechanical calculator. His primary aim was to provide a language that supported beginners learning good problem solving and programming techniques.

In addition to manipulation of numbers, PASCAL supports manipulation of vectors, matrices, strings of characters, records, files and lists, thereby supporting non-numeric programming. Hence, it has proved to be an attractive language for professional computer scientists.

PASCAL has been standardized by ISO (International Standards Organization) and ANSI.

5. PL/1

NOTES

PL/1 (Programming Language 1) was developed by IBM in the 1960s and was the first language that was attempted to be used for a variety of applications rather than one particular area like business or science or artificial Intelligence.

6. LISP

LISP was developed in the early 1950s but was implemented in the 1959 by John McCarthy at the Massachusetts Institute of Technology. It became a standard language with the artificial intelligence community and was a program that could easily handle recursive.

7. C

This language was developed by Dennis Ritchie of Bell Laboratories in order to implement the operating system UNIX.

8. C ++

This language was developed by the Bjarne Stroustrup of Bell Laboratories by enhancing C. C++ is also used to write procedural programs like C but the reason for its increased popularity is perhaps because of its capability to handle the rigours of object-oriented programming. C and C++ are the most extensively used general-purpose languages amongst programming experts.

9. JAVA

Java is again an object-oriented language like the C++ but is a simplified version with extra features. It is less prone to programming errors. It was developed for writing programs that could be safely and easily executed through the Internet. It is free from any kind of common virus threats. It is basically a network-oriented language that can develop Website pages with enhanced multimedia features using small java programs known as java applets, Java is a secure to use over the Internet and is a platform independent language.

1.5 EVOLUTIONS OF COMPUTERS

History of computer stretches back to more than 2500 years to the Abacus. Abacus is a simple calculator made from beads and wire. The difference between an ancient Abacus and a modern computer seems vast, but the principle of making repeated calculations more quickly than the human brains is exactly the same. Table 1.1 will help you understand the generation of computers.

NOTES

Table 1.1 *Generation of Computers*

Generation	Time	Hardware	Software	Features	Examples
I	**1942-1955**	**Vacuum Tubes**	Machine Language (Binary Language)	High-speed electronic switching device; memory type was electromagnetic; bulky in size; generated a large amount of heat; frequent technical faults; required constant maintenance; used for scientific purposes; air-conditioning required	ENIAC, EDVAC, EDSAC, UNIVAC I
II	**1955-1964**	**Transistors**	High-level languages FORTRAN, COBOL, ALGOL, SNOBOL	Better electronic switching devices than vacuum tubes; made of germanium semiconductors; memory type was magnetic cores; powerful and more reliable; easy to handle; much smaller than vacuum tubes; generated less heat as compared to vacuum tubes; used for business and industries for commercial data processing; air-conditioning required	Livermore Atomic Research Computer (LARC), IBM
III	**1964-1975**	**Integrated Circuits (ICs)** made up of transistors, resistors and capacitors fixed on single silicon chip	High-level languages PL/1, PASCAL, BASIC, VISUAL BASIC, C, C++, C#, Java	ICs were smaller than transistors; consumed less power; dissipated less heat as compared to transistors; more reliable and faster than earlier generations; capable of performing about 1 million instructions per second; large storage capacity; used for both scientific and commercial purposes; air-conditioning required	Mainframe, Minicomputers
IV	**1975-1989**	**Microprocessor** made up of Large Scale Integration Circuits (LSI) and Very Large Scale Integration Circuits (VLSI)	Advanced Java (J2EE, JDO, JavaBeans), PHP, HTML, XML, SQL	Microprocessor had control on logical instructions and memory; semiconductor memories; personal computers were assembled; used in LAN and WAN to connect multiple computers at a time; used graphical user interface; smaller, more reliable and cheaper than third-generation computers; larger primary and secondary storage memories; had Computer Supported Cooperative Working (CSCW); air-conditioning not required	Personal Computers (PCs), LAN, WAN, CSCW
V	**1989-Present**	Ultra Large Scale Integration (ULSI), Optical Disks	Artificial Intelligence, PROLOG, OPS5, Mercury	PCs were assembled – portable and non-portable, powerful desktop PCs and workstations; less prone to hardware failure; user-friendly features – Internet, e-mailing; air-conditioning not required	Portable PCs, Palmtop Computers, Laptop

1.6 CLASSIFICATION OF COMPUTER

Computers can be classified on the basis of their size, processing speed and cost. The various types of computers are:

- Personal computers
- Workstations
- Notebook/laptop computers
- Tablet PC
- PDA
- Mainframe computers
- Supercomputers

NOTES

Analog

Analog computers are generally used in industrial process controls and to measure physical quantities, such as pressure, temperature, etc. An analog computer does not operate on binary digits to compute. It works on continuous electrical signal inputs and the output is displayed continuously. Its memory capacity is less and can perform only certain type of calculations. However, its operating speed is faster than the digital computer as it works in a totally different mode.

Analog computers perform computations using electrical resistance, voltage, etc. The use of electrical properties signifies that the calculations can be performed in real time or even faster at a significant fraction of the speed of light. Typically, an analog computer can integrate a voltage waveform using a capacitor which ultimately accumulates the charge. The basic mathematical operations performed in an electric analog computer are summation, inversion, exponentiation, logarithm, integration with respect to time, differentiation with respect to time, multiplication and division. Hence in the analog computers, an analog signal is produced which is composed of DC and AC magnitudes, frequencies and phases. The starting operations in an analog computer are done in parallel. Data is represented as a voltage that is a compact form of storage.

Digital

Digital computers are commonly used for data processing and problem solving using specific programs. A digital computer stores data in the form of digits (numbers) and processes. It is in the discrete form from one state to the next. These processing states involve binary digits which acquire the form of the existence or nonexistence of magnetic markers in a standard storage devices, on-off switches or relays. In a digital computer, letters, words, symbols and complete texts are digitally represented, i.e., using only two digits 0 and 1. It processes data in discrete form and has a large memory to store huge quantity of data.

The functional components of a typical digital computer system are input-output devices, main memory, control unit and arithmetic logic unit. The processing of data in a digital computer is done with the help of logical circuits, which are also termed as digital circuits. All the circuits processing data in side a computer function in an extremely synchronized mode; which is further controlled using a steady oscillator acting as the computer's 'clock'. The clock rate of a typical digital computer ranges from several million cycles per second to several hundred million cycles, whereas the clock rate of fastest digital computers are about a billion cycles per second. Hence, the digital computers operate on very high speed and are able to perform trillions of logical or arithmetic operations per second to provide quick solution to problems, which is not possible for a human being to do manually.

NOTES

Hybrid

Hybrid computers are the combination of digital and analog computers. A hybrid computer uses the best features of digital and analog computers. It helps the user to process both continuous and discrete data. Hybrid computers are generally used for weather forecasting and industrial process control.

The digital component basically functions as a controller to provide logical operations, whereas the analog component functions as a solver to provide solutions of differential equations. Remember that the hybrid computers are different from hybrid systems. The hybrid system is a digital computer equipped with an analog-to-digital converter for input and a digital-to-analog converter for output. The term 'hybrid computer' signifies a mixture of different digital technologies to process specific applications with the help of various specific processor technologies.

General Purpose

Workstations are high-end, general-purpose computers designed to meet the computing needs of engineers, architects and other professionals who need computers with greater processing power, larger storage and better graphic display facilities. These are commonly used for Computer-Aided Design (CAD) and for multimedia applications such as creating special audio-visual effects for television programmes and movies. A workstation looks like a PC and can be used by only one person at a time. The characteristics of a workstation, which are often used to differentiate it from a PC are as follows:

- **Display facility**: Most workstations have a large-screen monitor (21 inches or more) capable of displaying high-resolution graphics as compared to PCs, which have a small-screen monitor (19 inches or less).
- **Storage capacity**: Workstations have a larger main memory than PCs, which have only a few hundred MB of main memory. The hard disk capacity of workstations is also more than that of PCs.
- **Processing power**: The processing power of workstations is several times greater than that of PCs.
- **Operating system**: PCs can run any of the five major operating systems—MS-DOS, MS-Windows, Windows-NT, Linux and Unix—but all workstations generally run the Unix operating system or a variation of it such as AIX (used in IBM workstations), Solaris (used in SUN workstations) and HPUX (used in HP workstations).
- **Processor design**: PCs normally use CPUs based on the Complex Instruction Set Computer (CISC) technology, whereas workstation CPUs are based on the Reduced Instruction Set Computer (RISC) technology.

NOTES

Special Purpose

A special purpose computer is a digital or an analog computer specifically designed to perform desired specific task. These are high-performance computing systems with special hardware architecture, which is dedicated to solve a specific problem. This is performed with the help of specially programmed FPGA chips or custom VLSI chips. They are used for special applications, for example, astrophysics computations, GRAPE-6 (for astrophysics and molecular dynamics), Hydra (for playing chess), MDGRAPE-3 (for protein structure computations), etc.

Micro, Mini, Mainframe and Supercomputers

Microcomputers

Microcomputers are developed from advanced computer technology. They are commonly used at home, classroom and in the workplace. Microcomputers are called home computers, personal computers, laptops, personal digital assistants, etc. They are powerful and easy to operate. In recent years, computers were made portable and affordable. The major characteristics of a microcomputer are as follows:

- Microcomputers are capable of performing data processing jobs and solving numerical programs. Microcomputers work rapidly like minicomputers.
- Microcomputers have reasonable memory capacity which can be measured in megabytes.
- Microcomputers are reasonably priced. Varieties of microcomputers are available in the market which can be as per the requirement of smaller business companies and educational institutions.
- Processing speed of microcomputers is measured in megahertz. A microcomputer running at 90MHz works approximately at 90 MIPS.
- Microcomputers have drives for floppy disk, compact disk and hard disks.
- Only one user can operate a microcomputer at a time.
- Microcomputers are usually dedicated to one job. Millions of people use microcomputers to increase their personal productivity.
- Useful accessory tools, such as clock, calendar, calculator, daily schedule reminders, scratch pads, etc., are available in a microcomputer.
- Laptop computers, also called notebook computers, are microcomputers. They use the battery power source. Laptop computers have a keyboard, mouse, floppy disc drive, CD drive, hard disk drive and monitor. Laptop computers are expensive in comparison to personal computers.

Minicomputers

Minicomputers are a scaled-down version of mainframe computers. The processing power and cost of a minicomputer are less than that of the mainframe. The

minicomputers have big memory sizes and faster processing speed compared to the microcomputer. Minicomputers are also called workgroup systems because they are well suited to the requirements of the minor workgroups within an organization. The major characteristics of a minicomputer are as follows:

- Minicomputers have extensive problem solving capabilities.
- Minicomputers have reasonable memory capacity which can be measured in megabytes or gigabytes.
- Minicomputers have quick processing speeds and operating systems facilitated with multitasking and network capabilities.
- Minicomputers have drives for floppy disk, magnetic tape, compact disk, hard disks, etc.
- Minicomputers can serve as network servers.
- Minicomputers are used as a substitute of one mainframe by big organizations.

Mainframe Computers

Mainframe computers are generally used for handling the needs of information processing of organizations like banks, insurance companies, hospitals and railways. This type of system is placed in a central location with several user terminals connected to it. The user terminals act as access stations and may be located in the same building Figure 1.4.

Fig. 1.4 *Mainframe Computer*

Mainframe computers are bigger and more expensive than workstations. They look like a row of large file cabinets and need a large room with closely monitored humidity and temperature levels. A mainframe system of lower configuration is often referred to as a minicomputer system. The various components of a mainframe computer are as follows:

- **Host, front-end and back-end computers**: A mainframe system consists of several computers, such as a host computer that carries out most of the computations and has direct control over all other computers. The front-end portion is used for handling communications to and from all the user terminals connected to the mainframe computer. The back-end

NOTES

portion is used to handle data input/output operations. The host computer and other computers are located in the systems room, to which entry is restricted to system administrators and maintenance staff only.

- **Consoles**: Console terminals are directly connected to the host computer and are mainly used by the system administrator to perform certain administrative tasks like installing new software on the system, taking system backups and changing the configuration of the system.
- **Storage devices**: A mainframe computer has several magnetic disk drives directly connected to the back-end computer. The host computer, via the back-end computer, accesses all data to and from these magnetic disks. In addition, a mainframe computer also has a few tape drives and a magnetic tape library (located in the systems room) for restoration and backup of data. The tape drives are located in the users' room, so that users' tapes can be used for input and output.
- **User terminals**: User terminals are used to access the required stations, which may be located at different locations. Since mainframe computers support multiprogramming with time-sharing, they can run different operating systems and can be accessed by multiple users simultaneously.
- **Output devices**: A mainframe computer has several output devices like printers and plotters, connected to the back-end computer, so that these devices are accessible to the user for taking their outputs. A plotter is a device that prints vector graphics on paper using ink pens and pencils on mechanical arms mainly used for large size printouts of architectural and engineering drawings.

Supercomputers

Supercomputers are the most powerful and expensive computers available today. They are primarily used for processing complex scientific applications that involve tasks with highly complex calculations and solving problems with mechanical physics, such as weather forecasting and climate research systems, nuclear weapon simulation and simulation of automated aircrafts. Military organizations, major research and development centres, universities and chemical laboratories are major users of supercomputers.

Supercomputers use multiprocessing and parallel processing technologies to solve complex problems promptly. They use multiprocessors, which enable the user to divide a complex problem into smaller problems. A parallel program is written in a manner that can break up the original problem into smaller computational modules. Supercomputers also support multiprogramming, which allows simultaneous access to the computer by multiple users. Some of the manufacturers of supercomputers are IBM, Silicon Graphics, Fujitsu and Intel.

NOTES

Personal Computers

A PC is a small single-user microprocessor-based computer that sits on your desktop and is generally used at homes, offices, and schools. As the name implies, PCs were mainly designed to meet the personal computing needs of individuals. Personal computers are used for preparing normal text documents, spreadsheets with predefined calculations and business analysis charts, database management systems, accounting systems and also for designing office stationary, banners, bills and handouts. Children and youth love to play games and surf the Internet, communicate with friends via e-mail and net telephony and do many other entertaining and useful tasks.

The configuration varies from one PC to another depending on its usage. However, it consists of a CPU or system unit, a monitor, a keyboard and a mouse. It has a main circuit board or motherboard (consisting of the CPU and the memory), hard disk storage, floppy disk drive, CD-ROM drive and some special add-on cards (like Network Interface Card) and ports for connecting peripheral devices like printers.

PCs are available in two models—desktop and tower. In the desktop model, the monitor is positioned on top of the system unit, whereas in the tower model the system unit is designed to stand by the side of the monitor or even on the floor to save desktop space. Due to this feature, the tower model is very popular

Some popular operating systems for PCs are MS-DOS, MS-Windows, Windows-NT, Linux and Unix. Most of these operating systems have the capability of multitasking, which eases operation and saves time when a user has to switch between two or more applications while performing a job. Some leading PC manufacturers are IBM, Apple, Compaq, Dell, Toshiba and Siemens.

Types of Personal Computers

Notebook/Laptop Computers

Notebook computers are battery-operated personal computers. Smaller than the size of a briefcase, these are portable computers and can be used in places like libraries, in meetings or even while travelling. Popularly known as laptop computers, or simply laptops, they weigh less than 2.5 kg and can be only 3 inches thick. Notebook computers are usually more expensive as compared to desktop computers though they have almost the same functions, but since they are sleeker and portable they have a complex design and are more difficult to manufacture. These computers have large storage space and other peripherals such as serial port, PC card, modem or network interface card, CD-ROM drive and printer. They can also be connected to a network to download data from other computers or to the Internet. A notebook computer has a keyboard, a flat screen with Liquid Crystal Display (LCD) display (Refer Figure 1.5), and can also have a trackball and a pointing stick.

NOTES

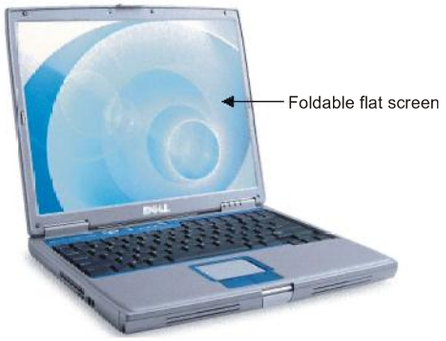


Fig. 1.5 *Laptop Computer*

A notebook computer uses the MS-DOS or WINDOWS operating system. It is used for making presentations as it can be plugged into an LCD projection system. The data processing capability of a notebook computer is as good as an ordinary PC because both use the same type of processor, such as an Intel Pentium processor. However, a notebook computer generally has lesser hard disk storage than a PC.

Tablet PC

Tablet PC is a mobile computer that looks like a notebook or a small writing slate but uses a stylus pen or your finger tip to write on the touch screen. It saves whatever you scribble on the screen with the pen, as shown in picture in the same way as you have written it. The same picture can than be converted to text with the help of a HR (hand recognition) software.

PDA

A Personal Digital Assistant (PDA) is a small palm sized hand-held computer which has a small colour touch screen with audio and video features. They are nowadays used as smart phones, web enabled palmtop computers, portable media players or gaming devices.

Most PDAs today typically have a touch screen for data entry, a data storage/ memory card, bluetooth, Wi-Fi or an infrared connectivity and can be used to access the Internet and other networks.

Check Your Progress

5. What are the different types of software on the basis of functionality?
6. Define OS.
7. What are the three categories of computer languages?
8. What are the different types of computer on the basis of size, processing speed and cost?
9. What are mainframe computers?

1.7 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

NOTES

1. The control unit, Arithmetic Logic Unit or ALU, memory registers and basic I/O devices are collectively known as a Central Processing Unit or CPU.
2. The three basics function of a computer is data processing, data storage and data movement.
3. Power Supply Unit (PSU) converts AC electric power to low voltage DC power for the internal components of the computer.
4. RAID Array Controller is a device to manage several internal or external hard disks and optionally some peripherals in order to achieve performance or reliability improvement in what is called a RAID array.
5. There are three types of software on the basis of functionality i.e. system software, programming software and application software.
6. An Operating System (OS) is the main control program for handling all other programs in a computer. The other programs, usually known as 'application programs', use the services provided by the OS through a well-defined Application Program Interface (API).
7. Machine language, assembly language and high-level languages are the three categories of computer languages.
8. The various types of computers are:
   - Personal computers
   - Workstations
   - Notebook/laptop computers
   - Tablet PC
   - PDA
   - Mainframe computers
   - Supercomputers
9. Mainframe computers are generally used for handling the needs of information processing of organizations like banks, insurance companies, hospitals and railways. This type of system is placed in a central location with several user terminals connected to it.

1.8 SUMMARY

- A computer is a programmable machine designed to automatically process a sequence of various arithmetic or logical operations.

NOTES

- CPU or Central Processing Unit performs the arithmetic and logic operations with the help of sequencing, and control unit that can change the order of operations based on the information that has been stored in memory.
- Computers are very fast. They can process millions of instructions every second.
- Computers can store vast amounts of information in the form of files, which can be recalled at any time.
- The degree of accuracy for a particular computer depends upon its design. Computers are diligent as they can perform any complicated task accurately without making any error.
- Computers perform various tasks depending upon the instructions fed into them and their hardware characteristics.
- Computers do not have their own intelligence and their I.Q.
- Computers have no feelings because they are machines.
- Computer hardware refers to the physical components that make up a computer system.
- Computer data storage refers to computer components and recording media that retain digital data.
- A computer cannot operate without any instructions and is based on a logical sequence of instructions in order to perform a function. These instructions are known as a 'computer program', and constitute the computer software.
- Different kinds of software designs have been developed for particular functions. Popular computer software include interpreter, assembler, compiler, operating systems, networking, word processing, accounting, presentation, graphics, computer games, etc.
- System software is a group of programs rather than one program and is responsible for using computer resources efficiently and effectively.
- Real-time software is based on observing, analysing and controlling real life events as they occur. Manually, a real-time system guarantees a response to an external event within a specified period of time.
- An Operating System (OS) is the main control program for handling all other programs in a computer. The other programs, usually known as 'application programs', use the services provided by the OS through a well-defined Application Program Interface (API).
- COTS is a term used for software and hardware technology which is available to the general public for sale, license or lease. In other words, to use COTS software, you must pay its developer in one way or another.

NOTES

- Free domain software is software that comes with permission for anyone to use, copy, and distribute, either verbatim or with modifications, either gratis or for a fee.
- All computer languages can, however, be classified under the following categories:
  1. Machine Language (First Generation Language)
  2. Assembly Language (Second Generation Language)
  3. High-Level Language (Third Generation Language)
- Analog computers are generally used in industrial process controls and to measure physical quantities, such as pressure, temperature, etc. An analog computer does not operate on binary digits to compute.
- Digital computers are commonly used for data processing and problem solving using specific programs. A digital computer stores data in the form of digits (numbers) and processes.
- Hybrid computers are the combination of digital and analog computers. A hybrid computer uses the best features of digital and analog computers. It helps the user to process both continuous and discrete data.
- Microcomputers are called home computers, personal computers, laptops, personal digital assistants, etc. They are powerful and easy to operate.
- Minicomputers are a scaled-down version of mainframe computers. The processing power and cost of a minicomputer are less than that of the mainframe.
- Mainframe computers are generally used for handling the needs of information processing of organizations like banks, insurance companies, hospitals and railways. This type of system is placed in a central location with several user terminals connected to it.
- Supercomputers are the most powerful and expensive computers available today. They are primarily used for processing complex scientific applications that involve tasks with highly complex calculations and solving problems with mechanical physics, such as weather forecasting and climate research systems, nuclear weapon simulation and simulation of automated aircrafts.
- A PC is a small single-user microprocessor-based computer that sits on your desktop and is generally used at homes, offices, and schools.

1.9 KEY WORDS

- **Hardware**: It refers to the physical components that make up a computer system.

NOTES

- **Power Supply Unit (PSU)**: It converts AC electric power to low voltage DC power for the internal components of the computer.
- **RAM**: RAM stands for Random Access Memory. It stores the data and commands that is actively used by CPU.
- **ROM**: ROM stands for Read Only Memory. It stores the BIOS (Basic Input Output System) instructions.
- **Ports**: It is used for connecting external peripherals.
- **Operating System (OS)**: It is the main control program for handling all other programs in a computer.
- **Personal Digital Assistant (PDA)**: It is a small palm sized hand-held computer which has a small colour touch screen with audio and video features.

1.10 SELF-ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. State the basic functions of a computer.
2. What is hardware?
3. What is software? Explain its types.
4. What are the different types of computer languages?
5. What is supercomputer?
6. What is personal computer?
7. Define hybrid computer.
8. What is mainframe computer?

Long-Answer Questions

1. Explain the capabilities and limitations of a computer system giving examples.
2. What is the difference between hardware and software? Explain giving examples.
3. Explain the various types of computer software.
4. Differentiate between system software and application software giving examples.
5. What is free domain software? Explain giving examples.
6. Discuss in detail about the various generation of computers.

7. Differentiate between minicomputer and macro computer.
8. Briefly discuss about the analog, digital and hybrid computers.
9. Elaborate on the different types of personal computers.

NOTES

1.11 FURTHER READINGS

Price, Wilson T. 1981. *Introduction to Computer Data Processing*, 3rd Edition. United States: Holt-Saunders International Editions.

Verzello, Robert J. and John Reutter. 1983. *Data Processing: Systems and Concepts*. New York: The McGraw-Hill Publishing Company, Inc.

Beech, Carol and Janice Burn. 1985. *Applications in Business Data Processing*. London (England): Pitman Publisher.

Fuori, William M. and Lawrence J. Aufiero. 1987. *Computers and Information Processing*. United States: Prentice Hall.

Sanders, Donald H. 1988. *Computers Today*. New York: McGraw-Hill.

Sinha, Pradeep K. and Priti Sinha. 2004. *Computer Fundamentals: Concepts, Systems & Applications*, 8th Edition. New Delhi: BPB Publications.

Rajaraman, V. 1999. *Fundamentals of Computers*. New Delhi: Prentice-Hall of India.

Lipschutz, Martin M. and S. Lipschultz. 1982. *Theory and Problems of Data Processing, Schaum's Outline Series*. New Delhi: Tata McGraw Hill Publishing Co. Ltd.

Stallings, William. 1996. *Data and Computer Communications*. New Jersey (USA): Prentice-Hall PTR.

UNIT 2 COMPONENTS OF A DIGITAL COMPUTER

NOTES

Structure

2.0 INTRODUCTION

Digital computer is a device or machine used to process any kind of information. A digital computer has four basic functional components that include input/output devices, memory, control unit and arithmetic-logic unit.

Storage in computer hardware is typically about holding information that are entered by the user, permanently or temporarily. The storage device are mainly of two types either internal or external storage. Memory is used for storage and retrieval of instructions and data in a computer system. Every computer thus requires storage space where instructions and data of a program can reside temporarily when the program is being executed. This temporary storage area is built into the computer hardware and is known as the primary storage or main memory. Devices that provide backup storage, (such as magnetic tapes and disks) are called secondary storage or auxiliary memory. A Remote Data Entry (RDE) system is a computerized system designed for the collection of data in electronic format.

In this unit, you will learn about the basic components of a digital computer, internal storages, auxiliary storage, and remote data entry devices.

2.1 OBJECTIVES

After going through this unit, you will be able to:

- Explain the various components of a digital computer
- Define the internal storage

- Explain the concept of auxiliary storage
- Describe the remote data entry devices

NOTES

2.2 BASIC COMPONENTS OF A DIGITAL COMPUTER

A digital computer has four basic functional components i.e., input/output devices, main memory, control unit and arithmetic-logic unit.

In order to transfer data to the memory of the computer input devices are used. The Arithmetic Logic Unit (ALU) is responsible for calculations, to which this data from the memory is stored. Once the calculations are done, the data is transferred back to the memory. The memory is responsible for storing data, according to which different functions are carried out. This memory is also known as the main memory or the Immediate Access Store (IAS).

The control unit is responsible for controlling various computer operations, which involves accepting instructions, interpreting and processing of this information in the correct parts of the computer. The main function of the control unit is to make sure that the instructions are correctly followed and all operations are done exactly according to the correct instructions at the correct time. This process leads to outcomes that are stored in memory. Figure 2.1 displays a computer system.

Fig. 2.1 *A Computer System*

(i) Motherboard

The main PCB (Printed Circuit Board) is sometimes alternatively known as a logical board or a main board of a Personal Computer. In fact, any complex electronic system is known as a motherboard. It includes a flat fibreglass platform which hosts the CPU (Central Processing Unit), the main electronic components, device controller chips, main memory slots, slots for attaching the storage devices and other subsystems. Figure 2.2 displays a motherboard.

NOTES

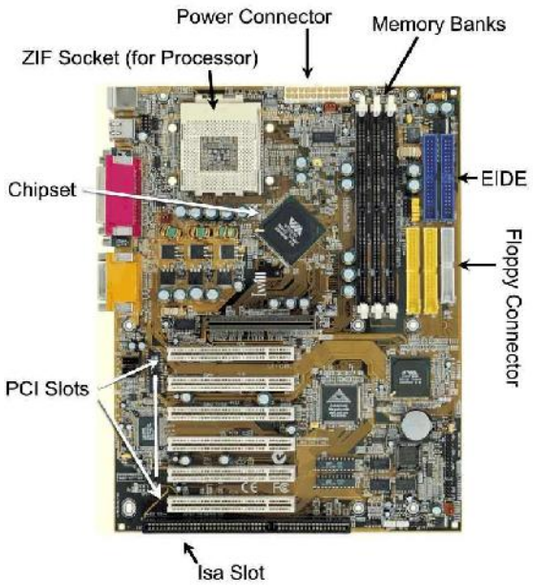


Fig. 2.2 *A Motherboard*

(ii) Sockets and Ports

- **Main Power Socket:** The top part of the rear of the computer locates the main power cable socket, which supplies power from the electric mains to the computer system. This socket is the part of the main power supply unit of the computer.
- **Monitor Power Socket:** The socket that supplies the power from the computer system to the computer monitor and is located below the main power cable socket. However, you might not find this socket in all computers and you can plug in the monitor directly in main power supply.
- **PS/2 Mouse Port:** Next you will find a small, round, green colored port with seven holes and a small logo of the mouse printed next to it. This is where your PS/2 mouse will be plugged in.
- **PS/2 Keyboard Port:** Right next to it you will find another similar purple colored port with the keyboard logo printed next to it. This is where your PS/2 keyboard will be plugged in.
- **Fan Housings:** You will notice two fan housings at the back of your computer. One fan housing is a part of the power supply unit and the other will be somewhere below it to cool off the heat generated by the CPU.
- **Serial Ports:** It is a 9-pin connector normally used to attach the old serial port mouse, hand-held scanners, modems, joysticks, game pads and other such devices.

NOTES

- **Parallel Port:** It is a 25-pin connector used to attach parallel port printers, modems, external hard disk drives, etc.
- **Audio Jacks:** There are three audio jacks in your computer system. One jack is used for connecting your speakers or headphones, the second is used to connect the microphone and the third to connect to another audio device, such as a music system.
- **LAN Port:** The LAN port is where the RJ45 connector of your LAN cable is plugged in to connect your computer to other computers or the Internet.
- **USB Ports:** The USB port is designed to connect multiple peripheral devices in a single standardized interface and has a plug and play option that allows devices to be connected and disconnected without having to restart or turning off the computer. It has replaced many serial and parallel port devices, such as mouse, printers, modems, joysticks, game pads, scanners, digital cameras and other such devices.
- **VGA Port:** This is a 15-pin connector that connects the signal cable of the monitor to the computer.

Figure 2.3 displays monitor and CPU power cables and sockets.

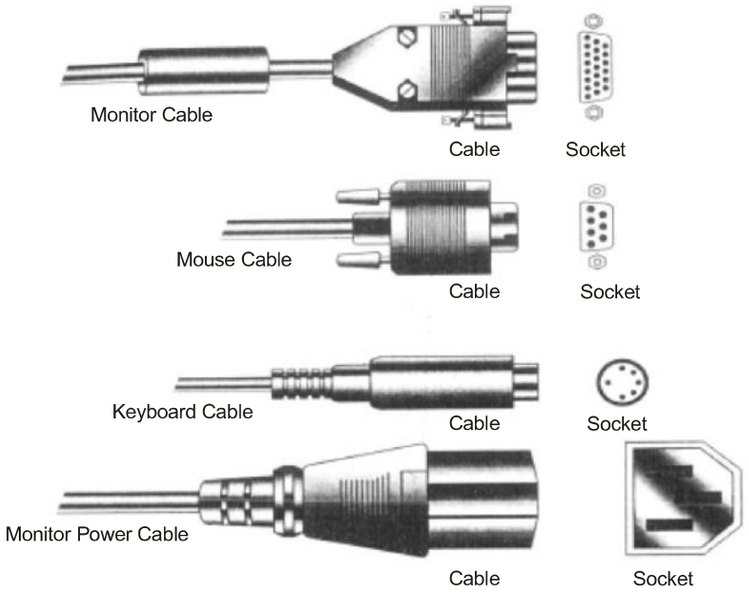
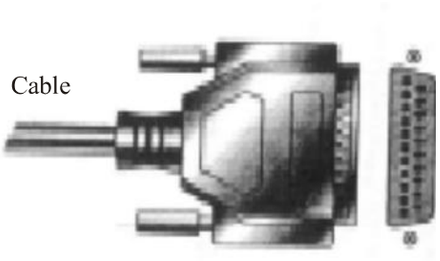


Fig. 2.3 *Monitor and CPU Power Cable and Sockets*

NOTES

Figure 2.4 displays a LAN cable and a printer cable and its socket.

LAN Cable



Printer Cable and Socket

Fig. 2.4 *LAN Cable and Printer Cable with Socket*

CPU

The primary function of the computer is executing programs. The programs or the set of instructions are stored in the computer's main memory and are executed by the CPU. The CPU processes the set of instructions along with any calculations and comparisons to complete the task. Additionally, the CPU controls and activates various other functions of the computer system. It also activates the peripherals to perform input and output functions. Figure 2.5 displays major components of the CPU.

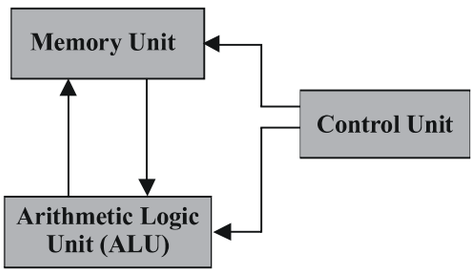


Fig. 2.5 *Major Components of a CPU*

Memory

Storage and retrieval of instructions and data in a computer system is the responsibility of the *memory*. In order to store data and instructions, the CPU constitutes many registers, though these are capable of storing very few bytes. All computers need storage space for temporarily storing instruction and data during

the execution of the program as the CPU can process data at a speed that is much faster than the speed at which data can be transferred from disks to registers. This could lead to the CPU remaining free most of the time if the data was located in secondary storage including magnetic tapes and disks. The primary or the main memory is the temporary storage located in the computer hardware. Secondary storage or auxiliary memory constitutes devices that can give backup storage, such as magnetic tapes and disks. The memory is classified as follows:

(i) **Internal Processor Memory:** A small set of high-speed registers placed inside a processor and used for storing temporary data while processing.

(ii) **Primary Storage Memory:** The main memory of the computer which communicates directly with the processor. This memory is large in size and fast, but not as fast as the internal memory of the processor. It comprises a couple of integrated chips mounted on a printed circuit board plugged directly on the motherboard. Random Access Memory (RAM) is an example of primary storage memory.

(iii) **Secondary Storage Memory:** This stores all the system software and application programs and is basically used for data backups. It is much larger in size and slower than primary storage memory. Hard disk drives, floppy disk drives and flash drives are a few examples of secondary storage memory.

(iv) **Memory Capacity:** Capacity, in computers, refers to the number of bytes that it can store in its main memory. This is usually stated in terms of Kilobytes (KB) which is 1024 bytes or Megabytes (MB) which is equal to 1024 KB (10,48,576 bytes). The rapidly increasing memory capacity of computer systems has resulted in defining the capacity in terms of Gigabytes (GB) which is 1024 MB (1,07,37,41,824 bytes). Thus a computer system having a memory of 256 MB is capable of storing (256 × 1024 × 1024) 26,84,35,456 bytes or characters.

Processors Used in PCs

The most significant part of the computer is the Central Processing Unit or the CPU. The CPU is mostly a microprocessor-based chip located on a single or sometimes a multiple printed circuit boards and is an internal component of the system. It is directly connected to the motherboard; however, the compatibility of the mother board and the CPU depends on the specific series of the latter. Due to the tremendous amount of heat generated by the CPU, it contains a heat sink and a cooling fan.

Popular microprocessors include Intel and AMD, which manufacture IBM compatible CPUs.

The brands of CPUs listed are not the only differentiating factors, between different processors. There are various technical aspects to these processors which allow us to differentiate between CPUs of different power, speed and processing capability. Accordingly, each of these manufacturers sells numerous product lines

NOTES

offering CPUs of different architecture, speed, price range, etc. The following are the most common aspects of modern CPUs that enable us to judge their quality or performance:

- **32 or 64-Bit Architecture:** A bit is the smallest unit of data that a computer processes. 32 or 64-bit architecture refers to the number of bits that the CPU can process at a time.
- **Clock Rate:** The speed at which the CPU performs basic operations, measured in Hertz (Hz) or in modern computers Megahertz – MHz or Gigahertz – GHz.
- **Number of Cores:** CPUs with more than one core are essentially multiple CPUs running in parallel to enable more than one operation to be performed simultaneously. Current ranges of CPUs offer up to eight cores. Currently, the Dual core (i.e., two cores) CPU is most commonly used for standard desktops and laptops and Quad core (i.e., four cores) is popular for entry level servers.
- **Additional Technology or Instruction Sets:** These refer to unique features that a particular CPU or range of CPUs offer to provide additional processing power or reduced running temperature. These range from Intel's MMX, SSE3 and HT to AMD's 3DNOW and Cool n Quiet.

hese technical factors are the basic way to judge how a CPU will perform. It is important to consider multiple factors when looking at a CPU rather than just the clock speed or any one specification on its own. It is easy for a single-core processor to run music videos, Internet applications or games individually, but when multiple applications are run together, it starts to slow down. A system running on a dual-core processor would be able to multitask better then a single-core processor, while it is very easy for an 8-core processor to run all these applications plus a lot more without showing any signs of slowing down. However, Intel's 4-core processors are actually two dual-core processors combined in a single processor, whereas AMD's 4-core processors are actually four processors built in a single chip.

A combination of the above mentioned specifications, along with the operating systems that the processor supports and the specific purpose for which the computer is to be used, are the factors to be considered when deciding which CPU is the most suitable for your needs.

2.3 INTERNAL STORAGE

Storage in computer hardware is typically about holding information that are entered by the user, permanently or temporarily. The storage device are mainly of two types either internal or external storage. Internal is basically inside the computer, such as hard disk this is the most basic level in the computer system need to hold operating system so that it can access input and output devices. It allow to store

data and application to load into the memory and ready for use. Data in the internal memory accessed faster than data stored on the external storage.

Internal storage refers to a computer's internal hard drive. This is the primary storage device used to store a user's files and applications. If a computer has multiple internal hard drives, then they are all considered part of the computer's internal storage.

Another popular type of internal storage is flash memory. It serves the same purpose as a hard drive, but stores data electronically rather than magnetically. Flash memory is the most common type of internal storage used by portable electronic devices, such as mobile phones and portable music players. Some computers now use flash drives rather than hard drives.

Most computers have some form of internal storage. The storage devices are categorised as 'Internal' which means that they are inside the computer case. The most common type of internal storage is the hard disk. At the most basic level, internal storage is needed to hold the operating system so that the computer is able to access the input and output devices, and also to store the applications software.

Internal storage allows the data and applications to be loaded very rapidly into memory, ready for use. The data can be accessed much faster than data which is stored on an external storage device. This is because internal storage devices are connected directly to the motherboard and its data bus whereas external devices are connected through a hardware interface, such as USB (Universal Serial Bus), which means they are considerably slower to access.

The main disadvantage of internal storage is that when the hard disk fails (and it will), all the data and applications may be lost.

This can be avoided to some extent by using more than one hard disk within the machine. Each hard disk has a copy of all the data, so if one fails the other can carry on. This is called a RAID (Redundant Array of Inexpensive Disks) array. An alternative is to use external drives for backup.

A storage device is also known as a storage medium or storage media. Digital storage is measured in megabytes (MB), gigabytes (GB), and, these days, terabytes (TB). Some computer storage devices are able to hold information permanently while others can only hold information temporarily. Every computer has both primary and secondary storage, with primary storage acting as a computer's short-term memory, and secondary as a computer's long-term memory.

Primary Storage: Random Access Memory (RAM)

Random Access Memory, or RAM, is the primary storage of a computer. While working on a file on the computer, it will temporarily store data in the RAM. RAM allows you to perform everyday tasks like opening applications, loading webpages, editing a document or playing games. It also allows you to jump from one task to another without losing your progress. Basically, the larger the RAM of your computer, the smoother and quicker it is for you to multitask.

NOTES

RAM is a volatile memory, meaning it cannot hold onto information once the system turns off. For example, if you copy a block of text, restart your computer, and then attempt to paste that block of text into a document, you will find that your computer has forgotten the copied text. This is because it was only stored temporarily in your RAM.

RAM makes it possible for a computer to access data in a random order, and thus reads and writes much faster than a computer's secondary storage.

Secondary Storage: Hard Disk Drives (HDD) and Solid State Drives (SSD)

In addition to RAM, every computer also has another storage drive that is used for storing information on a long-term basis. This is secondary storage. Any file you create or download saves to the computer's secondary storage. There are two types of storage device used as secondary storage in computers: Hard Disk Drives (HDD) and Solid State Drives (SSD). Secondary storage devices are often removable, so you can replace or upgrade your computer's storage, or move your storage drive to a different computer.

Hard Disk Drives (HDD): The Hard Disk Drive (HDD) is the original hard drive. These are magnetic storage devices that have been around since the 1950s, though they have evolved over time. A HDD is comprised of a stack of spinning metal disks known as platters. Each spinning disk has trillions of tiny fragments that can be magnetized in order to represent bits (1s and 0s in binary code). An actuator arm with a read/write head scans the spinning platters and magnetizes fragments in order to write digital information onto the HDD, or detects magnetic charges to read information from it. HDDs are used for TV and satellite recorders and servers as well as laptop and PC storage.

Solid State Drives (SSD): Solid State Drives (SDD) emerged far more recently, in the '90s. SSDs do not rely on magnets and disks, instead they use a type of flash memory called NAND. In an SSD, semiconductors store information by changing the electrical current of circuits contained within the drive. This means that unlike HDDs, SSDs do not require moving parts to operate. Because of this, SSDs not only work faster and smoother than HDDs (HDDs take longer to gather information due to the mechanical nature of their platters and heads), they also generally last longer than HDDs (with so many intricate moving parts, HDDs are vulnerable to damage and wear). Other than PCs and high-end laptops, you can find SSDs in smartphones, tablets, and sometimes video cameras.

Flash Memory Devices

A flash memory device contains trillions of interconnected flash memory cells that store data. These cells hold millions of transistors that when switched on or off represent 1s and 0s in binary code, allowing a computer to read and write information.

One of the most recognizable type of flash memory device is the USB flash drive. Also known as a thumb drive or a memory stick, these small, portable storage devices have long been a popular choice for extra computer storage. USB flash drives were essential for easily moving files from one device to another. However, they can only be used on devices with a USB port.

These days, a USB flash drive can hold up to 2 TB of storage. They are more expensive per gigabyte than an external hard drive, but they have prevailed as a simple, convenient solution for storing and transferring smaller files.

In addition to storage media contained within a computer, there are also digital storage devices that are external from computers. These are commonly used to expand storage capacity on a computer runs low on space, allow more portability, or provide easy file transfers from one device to another.

2.4 AUXILIARY STORAGE

The storage devices that provide backup storage are called auxiliary memory.

RAM is a volatile memory and thus a permanent storage media is required in a computer system. Auxiliary memory devices are used in a computer system for the permanent storage of information and hence, are the devices that provide backup storage. They are used for storing system programs, large data files and other backup information. The auxiliary memory has a large storage capacity and is relatively inexpensive, but has low access speed as compared to the main memory. The most common auxiliary memory devices used in computer systems are magnetic disks, floppy disks and tapes. Now optical disks are also used as auxiliary memory.

Magnetic Disk

Magnetic disks are circular metal plates coated with magnetized material on both sides. Several disks are stacked to a spindle one below the other with a read/write head to make a disk pack. The disk drive consists of a motor and all disks rotate together at very high speed. Information is stored on the surface of a disk along concentric sets of rings called tracks. These tracks are divided into sections called sectors. A set of corresponding tracks in all surfaces of a disk pack is called cylinder. Thus, if a disk pack has n plates, there are $2n$ surfaces, hence the number of tracks per cylinder is $2n$. The minimum quantity of information, which can be stored is a sector. If the number of bytes to be stored in a sector is less than the capacity of the sector, the rest of the sector is padded with the last type recorded. Figure 2.6 shows a magnetic disk memory.

NOTES

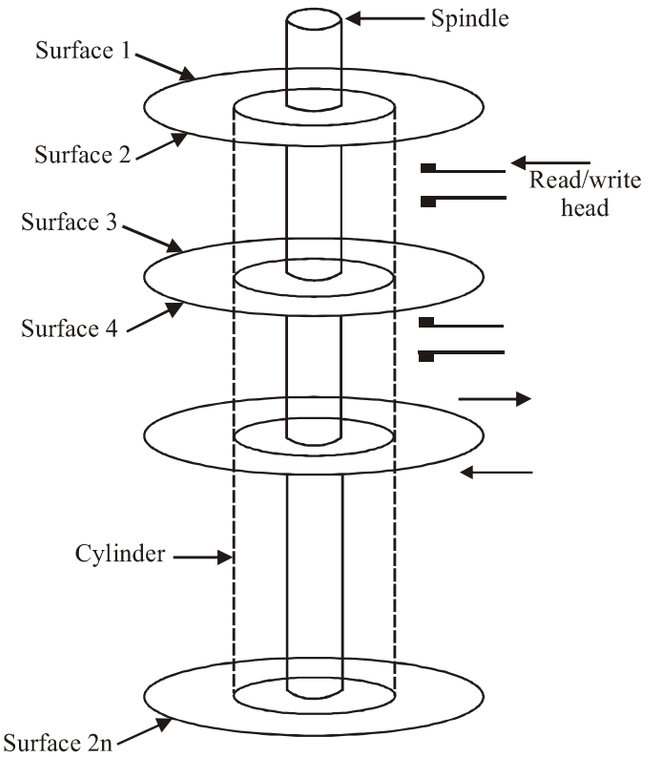


Fig. 2.6 *Magnetic Disk*

The subdivision of a disk surface into tracks and sectors is shown in Figure 2.7.

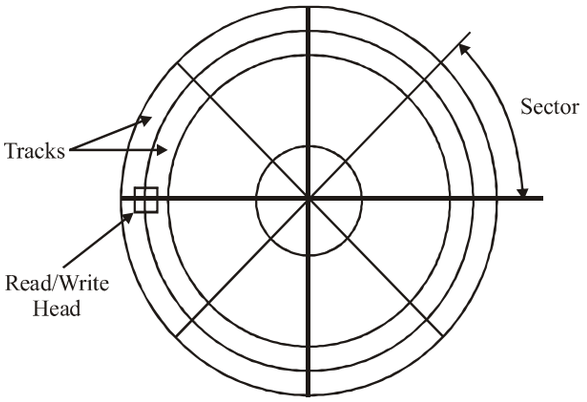


Fig. 2.7 *Surface of a Disk*

Suppose *s* bytes are stored per sector, there are *p* sectors per track, *t* tracks per surface and *m* surfaces. Then, the capacity of disk will be defined as:

$$\text{Capacity} = m \times t \times p \times s \text{ bytes}$$

If *d* is the diameter of the disk, the density of recording is:

$$\text{Density} = \frac{(p \times s)}{(\pi \times d)} = \text{bytes/inch}$$

NOTES

A set of disk drives are connected to a disk controller. The disk controller accepts commands and positions the read/write heads for reading or writing. When the read/write command is received by the disk controller, the controller first positions the arm so that the read/write head reaches the appropriate cylinder. The time taken to reach the appropriate cylinder is known as *Seek time* (*Ts*). The maximum seek time is the time taken by the head to reach the innermost cylinder from the outermost cylinder or vice versa. The minimum seek time will be 0 if the head is already positioned on the appropriate cylinder. Once the head is positioned on the cylinder, there is further delay because the read/write head has to be positioned on the appropriate sector. This is rotational delay also known as *Latency time* (*Tl*). The average rotational delay equals half the time taken by the disk to complete one rotation.

Floppy Disk

A floppy disk, also known as diskette, is a very convenient bulk storage device and can be taken out of the computer. It can be either 5.25" or 3.5" size, the 3.5" size being more common. It is contained in a rigid plastic case. The read/write heads of the disk drive can write or read information from both sides of the disk. The storage of data is in the magnetic form, similar to that in hard disk. The 3.5" floppy disk has storage up to 1.44 Mbytes. It has a hole in the centre for mounting it on the drive. Data on the floppy disk is organized during the formatting process. The disk is organized into sectors and tracks. The 3.5" high density disk has 80 concentric circles called tracks and each track is divided into 18 sectors. Tracks and circles exist on both sides of the disk. Each sector can hold 512 bytes of data plus other information like address, etc. It is a cheap read/write bulk storage device.

Magnetic Tapes

Magnetic disk is used by almost all computer system as a permanent storage device; however, magnetic tape is still a popular form of low cost magnetic storage media and it is primarily used for backup storage purposes. The standard backup magnetic tape device used today is Digital Audio Tape (DAT). These tapes provide approximately 1.2 Gbytes of storage on a standard cartridge size cassette tape. These magnetic tapes memories are similar to that of audio tape recorders.

A magnetic tape drive consists of two spools on which the tape is wounded. Between the two spools, there is a set of nine magnetic heads to write and read information on the tape. The nine heads operate independently and record information on nine parallel tracks, parallel to the edge of the tape. Eight tracks are used to record a byte of data and the ninth track is used to record a parity bit for each byte. The standard width of the tape is half an inch. The number of bits per

NOTES

inch (bpi) is known as *recording density*.

Normally, when data is recorded into the tape, a block of data is recorded and then a gap is left and then another block is recorded and so on. This gap is known as Inter Block Gap (IBG). The blocks are normally 10 times long as that of IBG. The Beginning Of the Tape (BOT) is indicated by a metal foil known as marker and the End Of Tape (EOT) is also indicated by a metal foil known as end of tape marker.

The data on the tape is arranged as blocks and cannot be addressed. They can only be retrieved sequentially in the same order in which they are written. Thus, if a desired record is at the end of the tape, earlier records have to be read before it is reached and hence, the access time is very high as compared to magnetic disks.

Optical Disks

Optical disk storage technology provides the advantage of high volume and economical storage with somewhat slower access times than traditional magnetic disk storage.

CD-ROM

Compact Disk-Read Only Memory (CD-ROM) optical drives are used for the storage of information that is distributed for read only use. A single CD-ROM can hold up to 800 MB of information. Software and large reports distributed to a large number of users are good candidates for this media. CD-ROM is also more reliable for distribution than floppy disks or tapes. Nowadays, almost all software and documentations are distributed only on CD-ROM.

In CD-ROMs the information is stored evenly across the disk in segments of the same size. Therefore, in CD-ROMs, data stored on a track increases as we go towards the outer surface of disk and hence, CD-ROMs are rotated at variable speeds for the reading process.

Information in a CD-ROM is written by creating pits on the disk surface by shining a laser beam. As the disk rotates, the laser beam traces out a continuous spiral. When 1 is to be written on the disk, a circular pit of around 0.8 micrometer diameter is created by the sharply focused beam and no pit is created if a zero is to be written. The pre recorded information on the CD-ROM is read with the help of a CD-ROM reader, which uses a laser beam for reading. For this, the CD-ROM disk is inserted into a slot of CD drive. Then the disk is rotated by a motor. A laser head moves in and out to the specified position. As the disk rotates, the head senses pits and land, which is converted to 1's and 0's by the electronic interface and sent to the computer. Figure 2.8 depicts a CD-ROM.

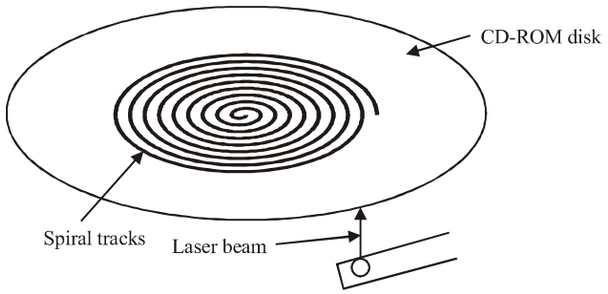


Fig. 2.8 *Tracks on a Disk Surface*

The speed of the disk is indicated by nx, where n is an integer indicating the factor by which the original nominal speed of 150 Kb/s is multiplied. Thus, a 52 × CD-ROM disk speed will be 52 × 150 = 7800 Kb/s. CD-ROM has a buffer size of 256 Kilobytes to keep data temporarily. It is connected to the computer system by a Small Computer System Interface (SCSI) adapter.

The main advantages of CD-ROMs are:

- Large data/information storage capacity.
- Mass replication is inexpensive and fast.
- These are removable disks.

Disadvantages of CD-ROMs are:

- It is read only and hence cannot be up dated.
- Access time is longer than that of a magnetic disk.

Erasable Optical Disk

Recent development in optical disks is the erasable optical disks. They are used as an alternative to standard magnetic disks when speed of the access is not important and the volume of the data stored is large. They can be used for image, multimedia, a high volume and low activity backup storage. Data in these disks can be changed as repeatedly as in a magnetic disk. The erasable optical disks are portable and highly reliable and have longer life. They use format that makes semi random access feasible.

Check Your Progress

1. List the various components of a digital computer.
2. Write the types of internal storage.
3. Write a note on control unit.
4. What is auxiliary memory?
5. What are tracks?

NOTES

2.5 REMOTE DATA ENTRY DEVICES

A Remote Data Entry (RDE) system is a computerized system designed for the collection of data in electronic format. The term is most commonly applied to a class of software used in the life sciences industry for collecting patient data from participants in clinical research studies—research of new drugs and/or medical devices.

Typically, RDE systems provide:

- A graphical user interface component for data entry
- A validation component to check user data
- A reporting tool for analysis of the collected data

RDE software was started in the mid- to late-1980s as software installed locally on portable computers with modems. It has largely been replaced by a newer generation of software called Electronic Data Capture or EDC that provides the same type of functionality over the Internet using web pages.

Fundamentally, the Remote Data Entry (RDE) means computer software or programs that reside in master computer and the remote data is read from remote terminals with the help of computer devices. Therefore, the RDE system is a computerized system.

The remote data entry devices are devices that read the data from a source that is provided and then transfer it directly to the specific computer system. Following are some examples of remote data entry devices.

1. Magnetic Stripe Reader: Magnetic stripe reader is used to read data from magnetic stripes on mostly banking cards.

Advantages: Information can be put into very quickly, more accurate than typing.

Disadvantages: Not secure as data stored on the chip, very limited storage capacity in the stripe.

NOTES

2. Chip Readers: Chip reader's reads data from the chip on bank cards.

Advantages: More secure than magnetic stripe system, can hold lots of data, data protected with PIN (Personal Identification Number).

Disadvantages: It is not widely available as magnetic stripe readers.

3. PIN Pads: PIN pads are used to enter data into Automated Teller Machines (ATM), EFTPOS system, entry doors and handheld devices.

Advantages: PIN codes help prevent unauthorised access, secure entry doors can be accessed without physical key cards.

Disadvantages: People can be seen entering the code, anyone with the correct code can use the device and PIN code can be forgotten.

4. Optical Mark Reader (OMR): Optical Mark Reader (OMR) is used to read and input information from a form.

Advantages: Whole pages of information can be read very quickly, small marks required.

Disadvantages: Incorrectly completed forms can result in errors, characters or numbers cannot be read.

5. Barcode Reader: Barcode reader is used to scan codes directly from the products, books and membership cards.

Advantages: Faster, more accurate than manually inputting data, barcodes are used internationally to identify the product.

Disadvantages: Barcode only contains numerical code, barcodes can be easily damaged.

It is essential to select the best remote data entry device based on the user's needs and considering the advantages and the disadvantages of it.

Check Your Progress

6. What are remote data entry devices?
7. What is the use of pin pads?
8. What is the use of optical mark reader?

2.6 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. The main components of a digital computer are:
   i. Input/output devices
   ii. Control unit
   iii. Memory
   iv. Arithmetic-logic unit

2. The storage device are mainly of two types either internal or external storage.
3. The control unit is responsible for controlling various computer operations, which involves accepting instructions, interpreting and processing of this information in the correct parts of the computer. The main function of the control unit is to make sure that the instructions are correctly followed and all operations are done exactly according to the correct instructions at the correct time.
4. The storage devices that provide backup storage are called auxiliary memory.
5. Information is stored on the surface of a disk along concentric sets of rings called tracks.
6. A Remote Data Entry (RDE) system is a computerized system designed for the collection of data in electronic format.
7. PIN pads are used to enter data into Automated Teller Machines (ATM), EFTPOS system, entry doors and handheld devices.
8. Optical Mark Reader (OMR) is used to read and input information from a form.

2.7 SUMMARY

- Input/output devices, control unit, memory and arithmetic-logic unit are the major components of a digital computer.
- Storage in computer hardware is typically about holding information that are entered by the user, permanently or temporarily.
- The storage device are mainly of two types either internal or external storage.
- Internal storage allows the data and applications to be loaded very rapidly into memory, ready for use. The data can be accessed much faster than data which is stored on an external storage device.
- RAM makes it possible for a computer to access data in a random order, and thus reads and writes much faster than a computer's secondary storage.
- The Hard Disk Drive (HDD) is the original hard drive. These are magnetic storage devices that have been around since the 1950s, though they have evolved over time. A HDD is comprised of a stack of spinning metal disks known as platters.
- Solid State Drives (SDD) emerged far more recently, in the '90s. SSDs do not rely on magnets and disks, instead they use a type of flash memory called NAND.
- The control unit is responsible for controlling various computer operations, which involves accepting instructions, interpreting and processing of this information in the correct parts of the computer.

- The memory is responsible for storing data, according to which different functions are carried out.
- The Arithmetic Logic Unit (ALU) is responsible for calculations, to which this data from the memory is stored.
- The storage devices that provide backup storage are called auxiliary memory.
- The most common auxiliary memory devices used in computer systems are magnetic disks, floppy disks and tapes.
- Magnetic disks are circular metal plates coated with magnetized material on both sides. Several disks are stacked to a spindle one below the other with a read/ write head to make a disk pack.
- The disk drive consists of a motor and all disks rotate together at very high speed.
- The average rotational delay equals half the time taken by the disk to complete one rotation
- A floppy disk, also known as diskette, is a very convenient bulk storage device and can be taken out of the computer
- Magnetic disk is used by almost all computer system as a permanent storage device; however, magnetic tape is still a popular form of low cost magnetic storage media and it is primarily used for backup storage purposes
- A magnetic tape drive consists of two spools on which the tape is wounded. Between the two spools, there is a set of nine magnetic heads to write and information on the tape.
- Optical disk storage technology provides the advantage of high volume and economical storage with somewhat slower access times than traditional magnetic disk storage.
- Compact Disk-Read Only Memory (CD-ROM) optical drives are used for the storage of information that is distributed for read only use. A single CD-ROM can hold up to 800 MB of information.
- The remote data entry devices are devices that read the data from a source that is provided and then transfer it directly to the specific computer system.

2.8 KEY WORDS

- **Internal Storage:** Internal storage allows the data and applications to be loaded very rapidly into memory, ready for use.
- **Auxiliary Storage**: The storage devices that provide backup storage are called auxiliary memory.
- **Magnetic Disk**: They are circular metal plates coated with magnetized material on both sides.

- **Tracks**: Information is stored on the surface of a disk along concentric sets of rings called tracks.
- **Magnetic Stripe Reader**: It is used to read data from magnetic stripes on mostly banking cards.
- **Barcode Reader**: It is used to scan codes directly from the products, books and membership cards

2.9 SELF-ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. What are the major components of a CPU?
2. Write the main disadvantage of internal storage.
3. Define auxiliary storage.
4. Write short note on magnetic disk.
5. What is magnetic tape?
6. Define magnetic stripe reader.
7. Explain about the chip readers.
8. What are the advantages of Optical Mark Reader (OMR)?

Long-Answer Questions

1. What are the basic components of a digital computer? Explain.
2. Briefly explain the internal storage giving appropriate examples.
3. Elaborate on auxiliary storage giving example.
4. What is optical disk? Explain.
5. What are the advantages and disadvantages of magnetic stripe reader?
6. What are the advantages and disadvantages of pin pads?
7. Explain about the barcode reader.

2.10 FURTHER READINGS

Price, Wilson T. 1981. *Introduction to Computer Data Processing*, 3rd Edition. United States: Holt-Saunders International Editions.

Verzello, Robert J. and John Reutter. 1983. *Data Processing: Systems and Concepts*. New York: The McGraw-Hill Publishing Company, Inc.

Beech, Carol and Janice Burn. 1985. *Applications in Business Data Processing*. London (England): Pitman Publisher.

NOTES

Fuori, William M. and Lawrence J. Aufiero. 1987. *Computers and Information Processing*. United States: Prentice Hall.

Sanders, Donald H. 1988. *Computers Today*. New York: McGraw-Hill.

Sinha, Pradeep K. and Priti Sinha. 2004. *Computer Fundamentals: Concepts, Systems & Applications*, 8th Edition. New Delhi: BPB Publications.

Rajaraman, V. 1999. *Fundamentals of Computers*. New Delhi: Prentice-Hall of India.

Lipschutz, Martin M. and S. Lipschultz. 1982. *Theory and Problems of Data Processing, Schaum's Outline Series*. New Delhi: Tata McGraw Hill Publishing Co. Ltd.

Stallings, William. 1996. *Data and Computer Communications*. New Jersey (USA): Prentice-Hall PTR.

UNIT 3 INTERNAL STORAGE AND PROCESSING METHODS

NOTES

Structure

3.0 INTRODUCTION

Computer data storage is a technology consisting of computer components and recording media that are used to retain digital data. It is a core function and fundamental component of computers. Data processing is defined as, 'The collection and manipulation of items of data to produce meaningful information'. In this sense it can be considered a subset of information processing, 'The change (processing) of information in any manner detectable by an observer'. The term 'Data Processing (DP)' has also been used to refer to a department within an organization responsible for the operation of data processing applications.

Computerized data processing, or Electronic Data Processing (EDP) represents a later development, with a computer used instead of several independent pieces of equipment. Data processing is the rules by which data is converted into useful information. A data processing system is an application that is optimized for a certain type of data processing. For example, a timesharing system is typically designed to run timesharing processing optimally. It can also be used to run batch processing.

The term 'Batch Processing' originates in the traditional classification of methods of production as job production (one-off production), batch production (production of a 'Batch' of multiple items at once, one stage at a time), and flow production (mass production, all stages in process at once). Computerized batch processing is the running of 'Jobs that can run without end user interaction, or can be scheduled to run as resources permit'.

In this unit, you will learn about the characteristics of internal storage, auxiliary storage, processing methods, batch processing, real-time processing, and time-sharing processing.

NOTES

3.1 OBJECTIVES

After going through this unit, you will be able to:

- Discuss the characteristics of internal storage
- Define the auxiliary storage
- Understand the processing methods
- Explain batch processing
- Analysis the real-time processing
- Elaborate the time-sharing processing

3.2 CHARACTERISTICS OF INTERNAL STORAGE

Computer data storage is a technology consisting of computer components and recording media that are used to retain digital data. It is a core function and fundamental component of computers.

The Central Processing Unit (CPU) of a computer manipulates data by performing computations. In practice, almost all computers use a storage hierarchy, which puts fast but expensive and small storage options close to the CPU and slower but less expensive and larger options further away. Generally the fast volatile technologies (which lose data when off power) are referred to as 'Memory', while slower persistent technologies are referred to as 'Storage'. Historically, memory has been called core memory, main memory, real storage or internal memory. Meanwhile, non-volatile storage devices have been referred to as secondary storage, external memory or auxiliary/peripheral storage.

Primary storage (also known as main memory, internal memory or prime memory), often referred to simply as memory, is the only one directly accessible to the CPU. The CPU continuously reads instructions stored there and executes them as required. Any data actively operated on is also stored there in uniform manner.

Main memory is directly or indirectly connected to the central processing unit via a memory bus. It has two buses, an address bus and a data bus. The CPU firstly sends a number through an address bus, a number called memory address that indicates the desired location of data. Then it reads or writes the data in the memory cells using the data bus. Additionally, a Memory Management Unit (MMU) is a small device between CPU and RAM recalculating the actual memory address, for example to provide an abstraction of virtual memory or other tasks.

As the RAM (Random Access Memory) types used for primary storage are volatile (uninitialized at start up), a computer containing only such storage would not have a source to read instructions from, in order to start the computer. Hence, non-volatile primary storage containing a small start-up program BIOS (Basic Input/Output System) is used to bootstrap the computer, that is, to read a larger program from non-volatile secondary storage to RAM and start to execute it. A non-volatile technology used for this purpose is called ROM, for Read-Only Memory (the terminology may be somewhat confusing as most ROM types are also capable of random access).

Many types of ROM are not literally read only, as updates to them are possible; however it is slow and memory must be erased in large portions before it can be re-written. Some embedded systems run programs directly from ROM (or similar), because such programs are rarely changed. Standard computers do not store non-rudimentary programs in ROM, and rather, use large capacities of secondary storage, which is non-volatile as well, and not as costly.

Without a significant amount of memory, a computer would merely be able to perform fixed operations and immediately output the result.

3.3 AUXILIARY STORAGE

Auxiliary storage is all addressable data storage that is not currently in a computer's main storage or memory. Synonyms are external storage and secondary storage.

Auxiliary memory units are among computer peripheral equipment. They provide slower access rates for greater storage capacity and data stability. Auxiliary memory holds programs and data for future use, and, because it is non-volatile (like ROM or Read Only Memory), it is used to store inactive programs and to archive data. Early forms of auxiliary storage included punched paper tape, punched cards, and magnetic drums. Since the 1980s, the most common forms of auxiliary storage have been magnetic disks, magnetic tapes, and optical discs.

Auxiliary storage is any storage that is made available to the system through input/output channels. This term refers to any addressable storage that is not within the system memory (RAM or random Access Memory). These storage devices hold data and programs for future use and are considered non-volatile storage that retains information even when power is not available. They trade slower read/write rates for increased storage capacity.

Auxiliary storage may also be referred to as secondary storage.

Magnetic Disk Drives: Magnetic disks are coated with a magnetic material, such as iron oxide. There are two types: hard disks made of rigid aluminum or glass, and removable diskettes made of flexible plastic.

Hard drives generally have several disks, or platters, with an electromagnetic read/write head for each surface; the entire assembly is called a comb. A

NOTES

microprocessor in the drive controls the motion of the heads and also contains RAM to store data for transfer to and from the disks. The heads move across the disk surface as it spins up to 15,000 revolutions per minute; the drives are hermetically sealed, permitting the heads to float on a thin film of air very close to the disk's surface. A small current is applied to the head to magnetize tiny spots on the disk surface for storage; similarly, magnetized spots on the disk generate currents in the head as it moves by, enabling data to be read.

RAID (Redundant Array of Inexpensive Disks) combines multiple disk drives to store data redundantly for greater reliability and faster access. They are used in high-performance computer network servers.

Magnetic Tape: Magnetic tape, similar to the tape used in tape recorders, has also been used for auxiliary storage, primarily for archiving data. Tape is cheap, but access time is far slower than that of a magnetic disk because it is sequential access memory, i.e., data must be sequentially read and written as a tape is unwound, rather than retrieved directly from the desired point on the tape. Servers may also use large collections of tapes or optical discs, with robotic devices to select and load them.

Optical Discs: Another form of largely Read Only Memory (ROM) is the optical compact disc, developed from videodisc technology during the early 1980s. Data are recorded as tiny pits in a single spiral track on plastic discs that range from 3 to 12 inches (7.6 to 30 cm) in diameter, though a diameter of 4.8 inches (12 cm) is most common. The pits are produced by a Laser or by a stamping machine and are read by a low-power Laser and a photocell that generates an electrical signal from the varying light reflected from the pattern of pits. Optical discs are removable and have a far greater memory capacity than diskettes; the largest ones can store many gigabytes of information.

A common optical disc is the CD-ROM (Compact Disc - Read Only Memory). CD-ROMs are used to distribute software, encyclopaedias, and multimedia text with audio and images. CD-R (CD-Recordable), or WORM (Write-Once Read-Many), is a variation of CD-ROM on which a user may record information but not subsequently change it.

3.4 PROCESSING METHODS

Data processing is generally defined as, 'The collection and manipulation of items of data to produce meaningful information'. In this sense it can be considered a subset of information processing, 'The change (processing) of information in any manner detectable by an observer'. The term 'Data Processing (DP)' has also been used to refer to a department within an organization responsible for the operation of data processing applications.

Computerized data processing, or Electronic Data Processing (EDP) represents a later development, with a computer used instead of several independent

pieces of equipment. Data processing is the rules by which data is converted into useful information. A data processing system is an application that is optimized for a certain type of data processing. For example, a timesharing system is typically designed to run timesharing processing optimally. It can also be used to run batch processing.

Before data can be processed and analysed, it must be organised systematically, so it can be easily read by algorithms. Raw or unprocessed data are essentially organised using the ETL (Extract, Transform, Load) method for preparing and organising the data for further processing.

The term 'Batch Processing' originates in the traditional classification of methods of production as job production (one-off production), batch production (production of a 'Batch' of multiple items at once, one stage at a time), and flow production (mass production, all stages in process at once). Computerized batch processing is the running of 'Jobs that can run without end user interaction, or can be scheduled to run as resources permit'.

Batch data processing is an efficient way of processing high volumes of data where a group of transactions is collected over a period of time. Data is collected, entered, processed and then the batch results are produced, such as Hadoop focuses on batch data processing. Batch processing requires separate programs for input, process and output. An example is payroll and billing systems.

In contrast, real time data processing involves a continual input, process and output of data. Data must be processed in a small time period (or near real time), for example the Radar systems, customer services and bank ATMs (Automated Teller Machines).

Definition: Collection, manipulation, and processing of collected data for the required use is known as **Data Processing (DP)**. This technique is typically performed by means of a computer; the process includes retrieving, transforming, or classification of information.

Fundamentally, the processing of data principally depends on the following:

- The volume of data that need to be processed
- The complexity of data processing operations
- Capacity and inbuilt technology of respective computer system
- Technical skills
- Time constraints

Commercial data processing involves a large volume of input data, relatively few computational operations, and a large volume of output. For example, an insurance company needs to keep records on tens or hundreds of thousands of policies, print and mail bills, and receive and post payments.

Data in its raw form is not useful to any organization. Data processing is the method of collecting raw data and translating it into usable information. It is usually

NOTES

performed in a step-by-step process by a team of data scientists and data engineers in an organization. The raw data is collected, filtered, sorted, processed, analysed, stored and then presented in a readable format.

Data processing steps include the following methods or procedures for efficient and systematic data processing:

- **Validation** – Ensuring that supplied data is correct and relevant.
- **Sorting** – Arranging items in some sequence and/or in different sets.
- **Summarization** – Reducing detailed data to its main points.
- **Aggregation** – Combining multiple pieces of data.
- **Analysis** – Collection, organization, analysis, interpretation and presentation of data.
- **Reporting** – List detail or summary data or computed information.
- **Classification** – Separation of data into several relevant or required categories.

Data processing is essential for organizations to create enhanced business strategies and to increase their competitive edge. By converting the data into a readable format, such as graphs, charts and documents, the employees throughout the organization can easily understand and use the data.

A data processing system is a combination of machines, people, and processes that for a set of inputs produces a defined set of outputs. The inputs and outputs are interpreted as data, facts, information, etc., depending on the interpreter's relation to the system.

A term commonly used synonymously with data or storage (codes) processing system is Information System (IS). With regard particularly to electronic data processing, the corresponding concept is referred to as Electronic Data Processing (EDP) system.

Cycle of Data Processing

The data processing cycle comprises of a sequences of steps where raw data (input) is fed into a process (CPU or Central Processing Unit) to produce perceptions (output). Each step of data processing cycle is taken in a precise and definite order and the whole process is repeated in a specified cyclic method. The output of first data processing cycle is stored and fed into the computer which can be used as the input for the next cycle. Figure 3.1 illustrates the six steps of the data processing cycle

NOTES

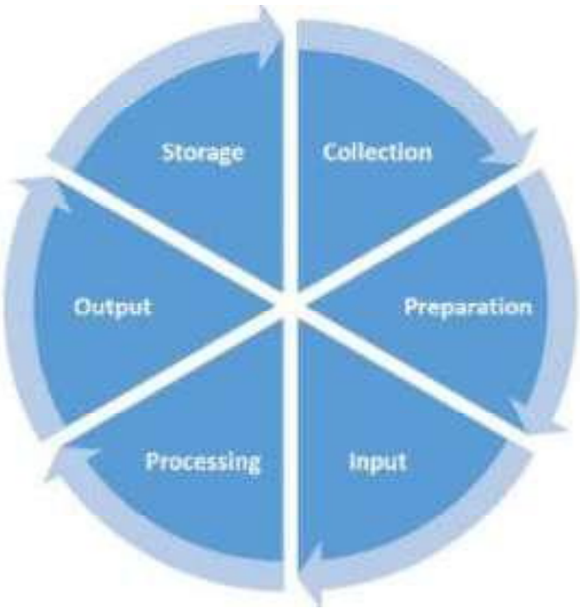


Fig. 3.1 *Data Processing Cycle*

Normally, there are following six key steps in the data processing cycle:

Step 1: Collection of Raw Data

The first step of the data processing cycle is collection of relevant and appropriate raw data. The output or the processed result depends on the type of raw data collected. Therefore, raw data must be collected from well-defined, precise and accurate sources so that the subsequent results and outcomes are valid and correct. Raw data includes monetary figures, website cookies, profit/loss statements of a company, user behaviour, etc.

Step 2: Preparation of Raw Data

Data preparation refers to the process of sorting and filtering the raw data to eliminate or exclude superfluous, unnecessary, excessive and inaccurate data. Raw data is then checked for errors, duplication, miscalculations or missing data, and transformed into an appropriate form for further analysis and processing. This ensures that only the utmost quality of data is fed into the processing unit.

Step 3: Input of Prepared Data

In input step, the raw data is first converted or transformed into the machine readable form using the precise computer software and then this transformed data is fed into the processing unit. The input of data or data entry is done through a keyboard, scanner or any other input source.

NOTES

Step 4: Data Processing

In data processing step, the raw data is subjected to several data processing methods using machine learning and artificial intelligence algorithms for generating or producing a required and appropriate output. This step may be somewhat different from process to process depending on the source of data being processed, such as data banks, online databases, connected devices, etc., and also on the planned use of the output.

Step 5: Output

The processed data is finally transmitted/transferred and displayed/presented to the data user in a readable form, such as graphs, tables, vector files, audio, video, documents, etc. This output or processed data is stored and further processed or used in the next data processing cycle.

Step 6: Storage

The sixth step of the data processing cycle is storage. The output data and metadata is stored for further use. This permits for quick access and retrieval of information whenever required. It also permits the data user to use the output or processed data as input in the next data processing cycle directly.

Applications of Data Processing

Following are the key applications of data processing.

Commercial Data Processing: Commercial data processing involves a large volume of input data, relatively few computational operations, and a large volume of output. For example, an insurance company needs to keep records on tens or hundreds of thousands of policies, print and mail bills, and receive and post payments.

Data Analysis: In science and engineering, the terms data processing and information systems are considered too broad, and the term data processing is typically used for the initial stage followed by a data analysis in the second stage of the overall data handling.

Data analysis uses specialized algorithms and statistical calculations that are less often observed in a typical general business environment. For data analysis, software suites like SPSS (Statistical Package for the Social Sciences) or SAS (Statistical Analysis System).

Following are the different types of data processing methods based on the source of data and the steps taken by the processing unit to generate an output.

NOTES

Check Your Progress

1. Explain the term magnetic disk drives.
2. Define the term data processing.
3. How data is organised systematically?
4. Explain the principal on which the processing of data depends.
5. What are the steps for efficient and systematic data processing?

3.4.1 Batch Processing

Batch processing can be defined as a type of data processing in which chunks of data that is stored over a period of time are analysed or processed in batches simultaneously. Fundamentally, the data is collected and processed in the form of batches, and it is typically used when the data type is homogenous and in huge quantities. Batch processing, therefore, can be defined as concurrent, simultaneous, or sequential execution of the data processing action. Simultaneous batch processing is used when the data is executed or processed using the similar resource for all the data processing activities instantaneously. Sequential batch processing, however, takes place when the data is executed or processed using the similar resource for different data processing activities either instantaneously or instantaneously after one another.

Concurrent batch processing means when the data is executed or processed using the similar resources not simultaneously but partially overlapping in time. Generally, it is used in financial applications or in the specific conditions where additional levels of security are essential. In the concurrent batch processing, the computational time is comparatively less because a function is applied to the whole data altogether for extracting or producing the output. It is capable or proficient to complete work with a very less amount of human intervention. By processing the data in batches, saves on computational resources.

Batch processing is preferred over real-time processing when accuracy is more important than speed. Additionally, the efficiency of batch processing is also measured in terms of throughput. Throughput is the amount of data processed per unit time.

Batch processing jobs are run on regularly scheduled times, for example overnight or on an as-needed basis. The common example includes bills for utilities and other services received by consumers which can be typically generated by batch processing each month. Batch processing is beneficial because it is a cost-effective means of handling large amounts of data simultaneously. One limitation is that the inputs for the processing must be correct or else the results of the whole batch will be faulty, which would cost time and money.

NOTES

Components of Batch Processing Architecture

The typical batch processing architecture consists of the following components:

- **Data Sources:** The digital properties where actions or events are being created.
- **Data Storage:** A distributed file store that serves as a repository for high volumes of data in various formats.
- **Batch Processing System:** Solution that processes data files using long-running batch jobs to filter, aggregate, and prepare data for analysis.
- **Analytical Data Store:** System that can store and serve processed data so that it can be queried by analytics tools.
- **Analysis, Reporting and Customer Engagement:** Downstream systems that enable data consumers to access processed data within a friendly interface.

Common Batch Processing Usage

- Efficient bulk database updates and automated transaction processing, as contrasted to interactive OnLine Transaction Processing (OLTP) applications. The Extract, Transform, Load (ETL) step in populating data warehouses is inherently a batch process in most implementations.
- Performing bulk operations on digital images, such as resizing, conversion, watermarking, or otherwise editing a group of image files.
- Converting computer files from one format to another. For example, a batch job may convert proprietary and legacy files to common standard formats for end-user queries and display.

Advantages of Batch Processing

- Batch processing is a technique for automating and processing multiple transactions as a single group.
- Batch processing helps in handling tasks like payroll, end-of-month reconciliation, or settling trades overnight.
- Batch processing systems can save money and human intervention.

Notable Batch Scheduling and Execution Environments

The IBM mainframe z/OS (Operating System) or platform has possibly the most highly refined and evolved set of batch processing facilities owing to its origins, long history, and continuing evolution. Today such systems commonly support hundreds or even thousands of concurrent online and batch processing tasks within a single operating system. Technologies that aid concurrent batch and online processing include Job Control Language (JCL), Scripting Languages, such as REXX (Restructured Extended Executor), Job Entry Subsystem (JES2 and JES3),

WorkLoad Manager (WLM), Automatic Restart Manager (ARM), Resource Recovery Services (RRS), DB2 Data Sharing, Parallel Sysplex, and Unique Performance Optimizations, such as HiperDispatch, I/O (Input/Output) Channel Architecture, and several others.

3.4.2 Real-Time Processing

Real-time data processing system is capable and proficient for taking input of fast changing data and then providing the output nearby instantaneously or promptly so that the change over time can be freely observed in such a data processing system. For example, a radar system depends on a continuous flow of input data which is processed by a computer to reveal the location of various aircraft flying within the range of the radar and then display it on a screen so that anyone looking at the screen can know the actual location of an aircraft at that moment.

Real-time data processing is also termed as the stream processing because it essentially uses the continuous stream of input data for yielding output for that specific action or process. Significant examples of real-time data processing systems are bank ATMs, traffic control systems, GPS-tracking applications, modern computer systems (such as, the PC and mobile devices), e-commerce order processing, online booking and reservations, and credit card real-time fraud detection. The principal advantage of real-time data processing is that it provides instantaneous results from input data that ensures and guarantees that entire stored and processed data is up-to-date.

Definition: Real-time processing is defined as the processing of unbounded stream of input data, with very short latency requirements for processing — measured in milliseconds or seconds. This input data is typically entered in an unstructured or semi-structured format and has the similar processing requirements as batch processing, but with shorter turnaround times to support real-time functions and utilities.

Generally, the processed data is frequently written to an analytical data storage system, which can be optimized for analytics and visualization. The processed data can also be directly utilized into the analytics and reporting layer for analysis, business intelligence, and real-time dashboard visualization.

Real-time processing computes incoming data as fast as possible. If it encounters an error in the incoming data, then it ignores the error and moves to the next chunk of data coming in, for example the GPS-tracking applications are the most commonly used real-time data processing system.

Components of Real-Time Processing Architecture

The typical real-time processing architecture consists of the following components:

- **Real-Time Message Processing:** A specific system is built for data capture and storage of real-time messages to be processed by means of a stream processing.

NOTES

- **Stream Processing:** A typical computerised system is used for capturing and processing real-time messages, which filters, aggregates and then prepares the data for analysis.
- **Data Store:** System stores the processed data in a structured format so that it can be queried using business data processing tools.
- **Downstream Activation Systems:** This refers to tools that are managed by data consumers for activating data and providing insights through analytics and reporting.

Real-Time Computing (RTC)

Real-Time Computing (RTC) or reactive computing is the computer science term for hardware and software systems subject to a 'Real-Time Constraint', for example from event to system response. Real-time programs must guarantee response within specified time constraints, often referred to as 'Deadlines'.

Real-time responses are often understood to be in the order of milliseconds, and sometimes microseconds. A system not specified as operating in real-time cannot usually guarantee a response within any timeframe, although typical or expected response times may be given. Real-time processing fails if not completed within a specified deadline relative to an event; deadlines must always be met, regardless of system load.

A real-time system has been described as one which 'Controls an environment by receiving data, processing them, and returning the results sufficiently quickly to affect the environment at that time'. The term 'Real-Time' is also used in simulation to mean that the simulation's clock runs at the same speed as a real clock, and in process control and enterprise systems to mean 'without significant delay'.

Real-time software may use one or more of the following:

- Synchronous Programming Languages
- Real-Time Operating Systems
- Real-Time Networks

Each of the above mentioned components provide essential frameworks on which the real-time software application is build.

3.4.3 Time-Sharing Processing

In data processing, the time-sharing method of operation is specific method in which multiple users with different programs interact nearly simultaneously with the Central Processing Unit (CPU) of a large-scale digital computer. Since the central processor functions or operates considerably faster as compared to most peripheral devices (such as, Video Display Units (VDUs), tape drives, and printers), it has sufficient time to solve several discrete problems during the input/output process. Even though the central processor addresses the problem of each user in

NOTES

sequence, access to and retrieval from the time-sharing system seems instantaneous from the standpoint of remote terminals since the solutions are available to them the moment the problem is completely entered.

Time-sharing was developed during the late 1950s and early '60s to make more efficient use of expensive processor time. The first time-sharing project was implemented by John McCarthy in the closing months of 1957 using a modified IBM 704 and later a modified IBM 7090. The first commercially successful time-sharing system was the Dartmouth Time Sharing System.

Commonly used time-sharing techniques include multiprocessing, parallel operation, and multiprogramming. Also, many computer networks organized for the purpose of exchanging data and resources are centred on time-sharing systems.

Definition: In computing, time-sharing is the sharing of a computing resource among many users at the same time by means of multiprogramming and multitasking.

The emergence of time-sharing data processing as the prominent model of computing in the 1970s represented a major technological shift in the history of computing. By allowing many users to interact concurrently with a single computer, time-sharing dramatically lowered the cost of providing computing capability, made it possible for individuals and organizations to use a computer without owning one, and promoted the interactive use of computers and the development of new interactive applications.

This multiprogramming and multitasking allowed many people to use a computer. Users were able to access the same computer through different terminals and were prompted when it was their turn. This is a historical method of using computers as there is no need to queue up users since modern computers, even the smallest ones, are able to cater to multiple users because of fast processors and multitasking operating systems.

Time Shared Operating System

A time shared operating system allows multiple users to share computers simultaneously. Each action or order at a time the shared system becomes smaller, so only a little CPU time is required for each user. As the system rapidly switches from one user to another, each user is given the impression that the entire computer system is dedicated to its use, although it is being shared among multiple users.

A time shared operating system uses CPU scheduling and multiprogramming to provide each with a small portion of a shared computer at once. Each user has at least one separate program in memory. A program loaded into memory and executes, it performs a short period of time either before completion or to complete I/O. This short period of time during which user gets attention of CPU is known as **time slice**, **time slot** or **quantum**. It is typically of the order of 10 to 100 milliseconds. Time shared operating systems are more complex than multiprogrammed operating systems. In both, multiple jobs must be kept in memory simultaneously, so the system must have memory management and security. To

NOTES

achieve a good response time, jobs may have to swap in and out of disk from main memory which now serves as a backing store for main memory. A common method to achieve this goal is virtual memory, a technique that allows the execution of a job that may not be completely in memory.

Time sharing is a logical extension of multiprogramming. The CPU performs many tasks by switches are so frequent that the user can interact with each program while it is running.

Figure 3.2 illustrates a time shared operating system, in which the User 5 is **active state** but User 1, User 2, User 3 and User 4 are in **waiting state** whereas User 6 is in **ready state**.

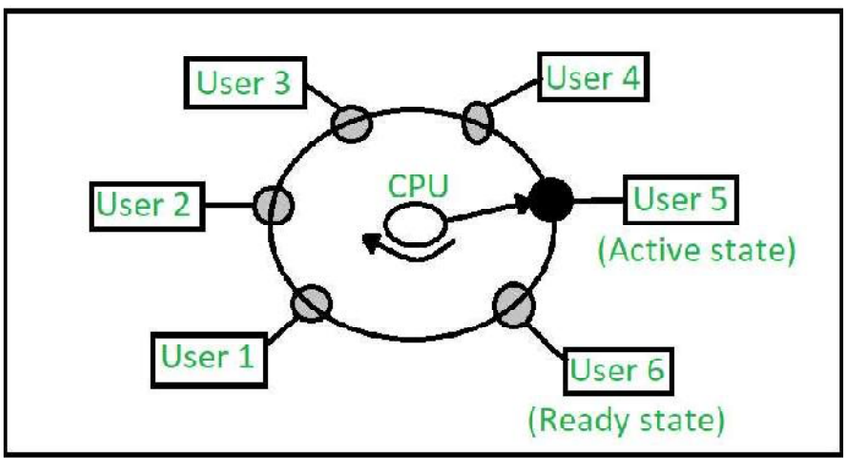


Fig. 3.2 *Time Shared Operating System*

Active State: The user's program is under the control of CPU. Only one program is available in this state.

Ready State: The user program is ready to execute but it is waiting for its turn to get the CPU. More than one user can be in ready state at a time.

Waiting State: The user's program is waiting for some Input/Output (I/O) operation. More than one user can be in a waiting state at a time.

Requirements for Time-Sharing Operating System

For a time-sharing operating system, an alarm clock mechanism is essential to send an interrupt signal to the CPU after every time slice. Memory protection mechanism is used to prevent one job's instructions and data from interfering with other jobs.

Advantages of Time-Sharing Operating System

- Each job or task gets an equal opportunity.
- Less chances of duplication of software.
- CPU idle time can be reduced.

NOTES

Check Your Progress

6. Define the term batch processing.
7. What are the advantages of batch processing?
8. Define the term real-time processing.
9. Explain the time-sharing method and techniques of data processing.

3.5 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. Magnetic disks are coated with a magnetic material, such as iron oxide. There are two types: hard disks made of rigid aluminium or glass, and removable diskettes made of flexible plastic.
2. Data processing is, generally, 'The collection and manipulation of items of data to produce meaningful information'. In this sense it can be considered a subset of information processing, 'The change (processing) of information in any manner detectable by an observer'.
3. Before data can be processed and analysed, it must be organised systematically, so it can be easily read by algorithms. Raw or unprocessed data are essentially organised using the ETL (Extract, Transform, Load) method for preparing and organising the data for further processing.
4. Fundamentally, the processing of data principally depends on the following:
   - The volume of data that need to be processed
   - The complexity of data processing operations
   - Capacity and inbuilt technology of respective computer system
   - Technical skills
   - Time constraints
5. Data processing steps include the following methods or procedures for efficient and systematic data processing:
   - Validation – Ensuring that supplied data is correct and relevant.
   - Sorting – Arranging items in some sequence and/or in different sets.
   - Summarization – Reducing detailed data to its main points.
   - Aggregation – Combining multiple pieces of data.
   - Analysis – Collection, organization, analysis, interpretation and presentation of data.
   - Reporting – List detail or summary data or computed information.

NOTES

- Classification – Separation of data into several relevant or required categories.

6. Batch processing can be defined as a type of data processing in which chunks of data that is stored over a period of time are analysed or processed in batches simultaneously. Fundamentally, the data is collected and processed in the form of batches, and it is typically used when the data type is homogenous and in huge quantities.
7. Following are the advantages of batch processing:
   - Batch processing is a technique for automating and processing multiple transactions as a single group.
   - Batch processing helps in handling tasks like payroll, end-of-month reconciliation, or settling trades overnight.
   - Batch processing systems can save money and human intervention.
8. Real-time processing is defined as the processing of unbounded stream of input data, with very short latency requirements for processing — measured in milliseconds or seconds. This input data is typically entered in an unstructured or semi-structured format and has the similar processing requirements as batch processing, but with shorter turnaround times to support real-time functions and utilities.
9. In data processing, the time-sharing method of operation is specific method in which multiple users with different programs interact nearly simultaneously with the Central Processing Unit (CPU) of a large-scale digital computer.

   Commonly used time-sharing techniques include multiprocessing, parallel operation, and multiprogramming. Also, many computer networks organized for the purpose of exchanging data and resources are centred on timesharing systems.

3.6 SUMMARY

- The Central Processing Unit (CPU) of a computer manipulates data by performing computations. In practice, almost all computers use a storage hierarchy, which puts fast but expensive and small storage options close to the CPU and slower but less expensive and larger options further away.
- Auxiliary storage is all addressable data storage that is not currently in a computer's main storage or memory. Synonyms are external storage and secondary storage.
- Auxiliary memory units are among computer peripheral equipment. They provide slower access rates for greater storage capacity and data stability.
- Magnetic disks are coated with a magnetic material, such as iron oxide. There are two types: hard disks made of rigid aluminium or glass, and removable diskettes made of flexible plastic.

NOTES

- Magnetic tape, similar to the tape used in tape recorders, has also been used for auxiliary storage, primarily for archiving data.
- Data processing is defined as, 'The collection and manipulation of items of data to produce meaningful information'. In this sense it can be considered a subset of information processing, 'The change (processing) of information in any manner detectable by an observer'.
- The term 'Data Processing (DP)' has also been used to refer to a department within an organization responsible for the operation of data processing applications.
- Computerized data processing, or Electronic Data Processing (EDP) represents a later development, with a computer used instead of several independent pieces of equipment. Data processing is the rules by which data is converted into useful information.
- A data processing system is an application that is optimized for a certain type of data processing. For example, a timesharing system is typically designed to run timesharing processing optimally. It can also be used to run batch processing.
- Before data can be processed and analysed, it must be organised systematically, so it can be easily read by algorithms. Raw or unprocessed data are essentially organised using the ETL (Extract, Transform, Load) method for preparing and organising the data for further processing.
- Computerized batch processing is the running of 'Jobs that can run without end user interaction, or can be scheduled to run as resources permit'.
- Batch data processing is an efficient way of processing high volumes of data where a group of transactions is collected over a period of time.
- Collection, manipulation, and processing of collected data for the required use is known as Data Processing (DP). This technique is typically performed by means of a computer; the process includes retrieving, transforming, or classification of information.
- Data in its raw form is not useful to any organization. Data processing is the method of collecting raw data and translating it into usable information.
- Data processing is usually performed in a step-by-step process by a team of data scientists and data engineers in an organization. The raw data is collected, filtered, sorted, processed, analysed, stored and then presented in a readable format.
- The data processing cycle comprises of a sequences of steps where raw data (input) is fed into a process (CPU or Central Processing Unit) to produce perceptions (output).
- Each step of data processing cycle is taken in a precise and definite order and the whole process is repeated in a specified cyclic method. The output

NOTES

of first data processing cycle is stored and fed into the computer which can be used as the input for the next cycle.

- Batch processing can be defined as a type of data processing in which chunks of data that is stored over a period of time are analysed or processed in batches simultaneously.
- Simultaneous batch processing is used when the data is executed or processed using the similar resource for all the data processing activities instantaneously.
- Sequential batch processing, however, takes place when the data is executed or processed using the similar resource for different data processing activities either instantaneously or instantaneously after one another.
- Concurrent batch processing means when the data is executed or processed using the similar resources not simultaneously but partially overlapping in time. Generally, it is used in financial applications or in the specific conditions where additional levels of security are essential.
- Batch processing is preferred over real-time processing when accuracy is more important than speed. Additionally, the efficiency of batch processing is also measured in terms of throughput. Throughput is the amount of data processed per unit time.
- Real-time data processing system is capable and proficient for taking input of fast changing data and then providing the output nearby instantaneously or promptly so that the change over time can be freely observed in such a data processing system.
- Real-time data processing is also termed as the stream processing because it essentially uses the continuous stream of input data for yielding output for that specific action or process.
- Real-time processing is defined as the processing of unbounded stream of input data, with very short latency requirements for processing — measured in milliseconds or seconds. This input data is typically entered in an unstructured or semi-structured format and has the similar processing requirements as batch processing, but with shorter turnaround times to support real-time functions and utilities.
- Real-Time Computing (RTC) or reactive computing is the computer science term for hardware and software systems subject to a 'Real-Time Constraint', for example from event to system response.
- Real-time responses are often understood to be in the order of milliseconds, and sometimes microseconds. A real-time system has been described as one which 'Controls an environment by receiving data, processing them, and returning the results sufficiently quickly to affect the environment at that time'.

- In data processing, the time-sharing method of operation is specific method in which multiple users with different programs interact nearly simultaneously with the Central Processing Unit (CPU) of a large-scale digital computer.
- Commonly used time-sharing techniques include multiprocessing, parallel operation, and multiprogramming. Also, many computer networks organized for the purpose of exchanging data and resources are centred on timesharing systems.
- In computing, time-sharing is the sharing of a computing resource among many users at the same time by means of multiprogramming and multitasking.
- A time shared operating system allows multiple users to share computers simultaneously. Each action or order at a time the shared system becomes smaller, so only a little CPU time is required for each user.

3.7 KEY WORDS

- **Data processing**: It is defined as, 'The collection, manipulation and processing of items of data to produce meaningful information'.
- **Batch processing**: It can be defined as a type of data processing in which chunks of data that is stored over a period of time are analysed or processed in batches simultaneously.
- **Data storage**: A distributed file store that serves as a repository for high volumes of data in various formats.
- **Analytical data store**: System that can store and serve processed data so that it can be queried by analytics tools.
- **Real-time processing**: It is defined as the processing of unbounded stream of input data, with very short latency requirements for processing — measured in milliseconds or seconds.
- **Stream processing**: A typical computerised system is used for capturing and processing real-time messages, which filters, aggregates and then prepares the data for analysis.
- **Time-sharing processing**: In computing, time-sharing is the sharing of a computing resource among many users at the same time by means of multiprogramming and multitasking.

3.8 SELF-ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. Define the term optical discs.

NOTES

2. Explain the term data processing.
3. What are the different types of data processing methods?
4. How the raw or unprocessed data is organised?
5. What does data processing steps include?
6. Define the term batch processing.
7. Elaborate on the components of batch processing architecture.
8. What is real-time processing?
9. State about Real-Time Computing (RTC).
10. When was time-sharing developed?
11. What are the commonly used time-sharing techniques?

Long-Answer Questions

1. Describe the characteristic features of internal storage.
2. Discuss about the auxiliary storage in details.
3. Briefly explain the concept of data processing giving definition and basic features.
4. How commercial data processing is done? Explain giving appropriate examples.
5. Discuss in detail about the six significant steps of data processing cycle giving examples of each type,
6. Explain the key applications of data processing.
7. Describe the batch processing technique of data processing giving definition and characteristic features.
8. Explain the concurrent, simultaneous, or sequential execution of the data processing action in a batch processing system.
9. Explain the applications, usage and components of batch processing architecture.
10. Why real-time data processing is also termed as the stream processing? Explain giving significant features.
11. What is the significance of Real-Time Computing (RTC) and real-time software?
12. Briefly discuss the time-sharing method of data processing giving significant features and examples.
13. Discuss in detail about the time shared operating system giving suitable example on different user's states.

3.9 FURTHER READINGS

Price, Wilson T. 1981. *Introduction to Computer Data Processing*, 3rd Edition. United States: Holt-Saunders International Editions.

Verzello, Robert J. and John Reutter. 1983. *Data Processing: Systems and Concepts*. New York: The McGraw-Hill Publishing Company, Inc.

Beech, Carol and Janice Burn. 1985. *Applications in Business Data Processing*. London (England): Pitman Publisher.

Fuori, William M. and Lawrence J. Aufiero. 1987. *Computers and Information Processing*. United States: Prentice Hall.

Sanders, Donald H. 1988. *Computers Today*. New York: McGraw-Hill.

Sinha, Pradeep K. and Priti Sinha. 2004. *Computer Fundamentals: Concepts, Systems & Applications*, 8th Edition. New Delhi: BPB Publications.

Rajaraman, V. 1999. *Fundamentals of Computers*. New Delhi: Prentice-Hall of India.

Lipschutz, Martin M. and S. Lipschultz. 1982. *Theory and Problems of Data Processing, Schaum's Outline Series*. New Delhi: Tata McGraw Hill Publishing Co. Ltd.

Stallings, William. 1996. *Data and Computer Communications*. New Jersey (USA): Prentice-Hall PTR.

NOTES

BLOCK - II
DATA PROCESSING SYSTEM

NOTES

UNIT 4 FUNDAMENTALS OF DATA PROCESSING TOOLS

Structure

4.0 INTRODUCTION

Data processing can be defined as a process of manipulating collected data and converting it into meaningful information. Data usually contain raw facts and figures about some specific topic or subject while information is the meaningful form of that data and is typically used for decision-making. Fundamentally, data is the input and information is the output. Several software packages and tools are used for processing the data efficiently.

Data can be processed manually using a pen and paper, mechanically using simple devices, such as typewriter and electronically using modern data processing tools, such as computers. There are three types of data processing systems, namely manual data processing system, mechanical data processing system and electronic data processing system.

The data processing cycle is the set of operations used to transform data into useful information. The intent of this processing is to create actionable information that can be used to enhance a business. Now-a-days in challenging business environment, the effective data processing method has a significant role for business organizations, corporation or enterprises who are using new and innovative methods or techniques for surviving, flourishing and thriving against multiple competitors.

In this unit, you will learn about the data processing systems and tools, types of data processing system, extent of data processing system and data processing cycle.

4.1 OBJECTIVES

NOTES

After going through this unit, you will be able to:

- Understand the importance of data processing systems
- Explain the types of data processing system
- Define the tools and significance of data processing system
- Elaborate on the data processing cycle

4.2 DATA PROCESSING SYSTEMS AND TOOLS

Data processing can be defined as a process of manipulating collected data and converting it into meaningful information. Data usually contain raw facts and figures about some specific topic or subject while information is the meaningful form of that data and is typically used for decision-making. Fundamentally, data is the input and information is the output.

Data can be processed manually using a pen and paper, mechanically using simple devices, such as typewriter or electronically using modern data processing tools, such as computers.

Data processing can be defined as "A process of converting data into meaningful information".

4.2.1 Types of Data Processing Systems

Following are the three key types of data processing systems.

1. Manual Data Processing

In the manual data processing method, the data is processed manually using paper sheets, registers, pens or pencils to obtain required results. In this method no device, calculator or any other tool is used. Therefore, in manual data processing methodology all the calculations and logical operations are manually performed on the data. Additionally, the processed data is also physically transferred from one place to another manually in the form of hardcopies. But this method of data processing is extremely slow and there may be errors in the manually processed outputs.

With the advancement in technology, this method is hardly used now-a-days because it has very high probability of errors, cost effective and very time consuming.

2. Mechanical Data Processing

In the mechanical data processing method, the data is processed using the different types of simple devices, such as typewriters, mechanical printers or any other

NOTES

mechanical device. The mechanical data processing method is comparatively faster and more accurate as compared to manual data processing method. With the advanced technology and with the invention and evolution of advanced complex machines having improved computing power, the mechanical data processing method is seldom used now-a-days.

3. Electronic Data Processing (EDP)

The Electronic Data Processing or EDP method of data processing is the modern method or technique used for processing the data. In the EDP method, the data is processed using the computer and the relevant data processing software. Data is input into the computer system and then the set of instructions are also input into the computer so that the computer can automatically process the data as per the given set of instructions. The computer system is also termed as the Electronic Data Processing or EDP machine.

The EDP method of data processing is considered as extremely fast, precise and accurate. For example, the banks uses the computerized system with specific software type in order to accurately maintain or process the accounts of customers.

Following are the types of EDP methods.

Batch Processing: In the batch processing method, the information is organized and sorted into groups for efficient and sequential processing of data.

Online Processing: In the online processing method, the Internet connections and equipment are directly attached to a computer system. In this method, the data is stored in one place and can be used at different places. For example, cloud computing uses online processing of data. In addition, it is also generally used for information recording and research.

Real-Time Processing: The real-time processing method or technique responds almost instantaneously to various signals for acquiring and processing the information. The real-time processing requires high maintenance and upfront cost accredited to advanced computing technology. The output is observed in real-time and it also saves lot of time. For example, the real-time output is very useful in banking transactions. Examples of real-time processing include airline reservation systems, theatre (cinema) booking, hotel reservations, banking systems, police enquiry systems, chemical processing plants, hospitals to monitor the progress of a patient and missile control systems.

Distributed Processing: The distributed processing method is usually utilized by remote workstations which are connected to one big central workstation or server. ATMs (Automated Teller Machines) are examples of distributed data processing method, because all the end machines run on a fixed software located at a particular place and makes use of exactly same information and sets of instruction.

4.2.2 Tools of Data Analysis and Data Processing

Data analytics tools are specific applications and software used by data analysts for developing and performing the necessary data analytical processes that help the business organizations and companies for making enhanced, more informed business decisions however lowering costs and increasing profits. Following are some significant tools for data analysis and data processing which are used by various business organizations and enterprises.

Hadoop

Hadoop is an open-source Big Data analytics tools which provides huge storage for all kinds of data. Basically, Hadoop has amazing processing power and proficiency for handling numerous tasks and is comparatively safe from hardware failure. Hadoop's core strength is its HDFS (Hadoop Distributed File System), which holds all types of data, video, images, JSON, XML and plain texts across the same file system. It is very useful for research and development purposes. The data can be accessed easily.

Apache Hadoop

The Apache Hadoop software library is a framework that allows for the distributed processing of large data sets across clusters of computers using simple programming models. It is specifically designed to scale up from single servers to thousands of machines, each offering local computation and storage.

Sequentum Enterprise

Sequentum Enterprise is an advanced data extraction tool. Basically, the Sequentum Enterprise is an excellent tool for large-scale web data extractions. Specifically designed for corporations that depend on structured web data and legal compliance. Users can control and debug the crawler with C# or VB.NET, or they can write their scripts.

Cloudera

Cloudera provides a data storage and processing platform based on the Apache Hadoop ecosystem, along with a proprietary system and data management tools for design, deployment, operations and production management.

Google BigQuery

Google provides a fully-managed enterprise data warehouse for data analytics through its BigQuery invention. It is server less and enables organizations to analyse any data by creating a logical data warehouse over managed, columnar storage, and data from object storage and spreadsheets. BigQuery holds data in real-time through its streaming assimilation feature and it is built atop the Google Cloud Platform. The BigQuery provides users the capability to share perceptions via datasets, queries, spreadsheets, and reports.

NOTES

MongoDB

MongoDB is an open source NoSQL database which is an advanced alternative to modern databases. Typically, it is a document oriented database for storing large volumes of data. Instead of rows and columns used in traditional databases, in MongoDB the documents and collections are used.

Documents comprise of key-value pairs and the collections include function and document sets. MongoDB is ultimate for companies who want to take quick decisions and need to work with real-time data. The Big Data technology is commonly used for storing data obtained from mobile applications, product catalogues and content management systems.

Cassandra

Cassandra is a distributed database management system specifically used for handling large volumes of data across several servers. This is considered as one of the most standard Big Data technologies which is ideal for processing structured data sets. It was first developed by Facebook as a NoSQL solution and now it is used by big corporates, such as Netflix, Twitter and Cisco.

OpenRefine

OpenRefine is a powerful and dominant Big Data tool specifically used for cleaning the data and then converting it into different required formats. The huge data sets can be explored using OpenRefine tool securely.

Xplenty

Xplenty is an ETL (Extract, Transform, Load) platform that requires no coding or deployment. Xplenty has a point-and-click interface that allows simple data integration, processing, and preparation. It also connects with a large variety of data sources to perform data analytics.

ZOHO Office Suite

ZOHO Office Suite is an Indian web-based online office suite containing word processing, spreadsheets, presentations, databases, note-taking, wikis, web conferencing, Customer Relationship Management (CRM), project management, invoicing and other applications. It is developed by ZOHO Corporation.

KNIME

KNIME, the Konstanz Information Miner, is a free and open source data analytics, reporting and integration platform. KNIME integrates various components for machine learning and data mining through its modular data pipelining 'Lego of Analytics' concept. A Graphical User Interface (GUI) and use of JDBC (Java DataBase Connectivity) allows assembly of nodes blending different data sources, including preprocessing (ETL: Extraction, Transformation, Loading), for modeling, data analysis and visualization without, or with only minimal, programming.

Calculation and Analysis Tools

The calculation and analysis of numerical data is significant for data processing. The spreadsheets are used for the processing of numerical data for organizing the data in the required manner specifying the formulae and functions processes that are predefined in the spreadsheet. Applications, such as MS EXCEL from Microsoft and CALC from Sun Microsystem's Open Office are typically used which simplify the calculation process and analysis of numerical data.

DataBase Management Tools

Data processing is defined as an organized digital repositories (databases) to generate aggregate information with distinct 'Views' (specifying the content and level details) defined as per the requirements of the user. Using the predefined functions of DataBase (DB) tools the data can be accessed by means of comprehensive request queries in the stored segments of the database. Tools, such as Microsoft Office's tool MS ACCESS and BASE from Sun Microsystem's Open Office are considered as extremely user friendly and powerful tools for data processing. These application programs helps the users to manage, operate and control the database in an efficient and organized digital form.

Statistical Data Analysis Tools

Statistical methods includes data collection, data interpretation, and data validation. Statistical analysis technique are used to perform numerous statistical operations to quantify the data applying statistical analysis. Quantitative data includes descriptive data type, such as surveys and observational data, hence also termed as descriptive analysis. Descriptive analysis includes various tools for performing statistical data analysis, such as SAS (Statistical Analysis System), SPSS (Statistical Package for the Social Sciences), StatSoft, etc.

SAS (Statistical Analysis System): SAS (Statistical Analysis System) is a statistical software suite developed by SAS Institute for data management, advanced analytics, multivariate analysis, business intelligence, criminal investigation, and predictive analytics.

SPSS (Statistical Package for the Social Sciences): SPSS Statistics is a software package used for interactive, or batched, statistical analysis. Long produced by SPSS Inc., it was acquired by IBM in 2009. Current versions (post 2015) have the brand name: IBM SPSS Statistics. SPSS is software for editing and analysing all sorts of data. Basically, these data may come from any source, such as scientific research, a customer database, Google Analytics or even the server log files of a website. SPSS can open all file formats that are commonly used for structured data, such as spreadsheets from MS Excel or OpenOffice, plain text files (.txt or .csv), relational (SQL or Sequential Query Language) databases, Stata and SAS. Stata is a general purpose statistical software package developed by StataCorp for data manipulation, visualization, statistics, and automated reporting.

NOTES

StatSoft: StatSoft is the original developer of Statistica. Statistica is an advanced analytics software portfolio that provides enterprise and desktop software for statistics, data analysis, data management, data visualization, data mining (also called predictive analytics), and quality control.

4.2.3 Data Processing Cycle

Data processing method refers to the transformation of raw data into meaningful information as output. The data processing cycle is the set of processes and operations used for transforming data into significant information. The basic notion of the data processing cycle is to produce authentic information that can be used by the business organisations and enterprises for enhancing their business. Figure 4.1 illustrates the five significant steps of data processing cycle.

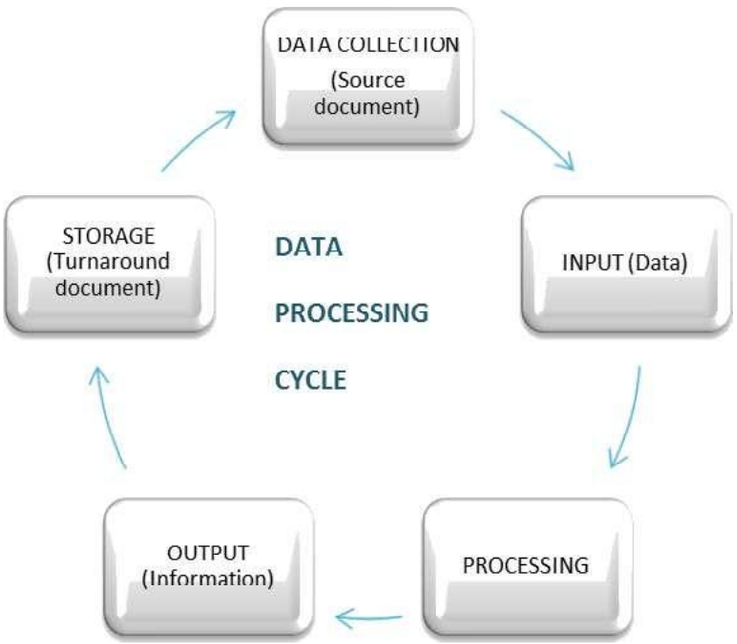


Fig. 4.1 *Five Significant Steps of Data Processing Cycle*

The five fundamental steps of the data processing cycle are given below.

Step 1: Data Collection: This step includes obtaining the data facts and figures from the source document for processing using the computer or digital system.

Step 2: Data Input: In this step the collected data is converted into machine-readable form by means of an input device and then send into the machine for processing.

Step 3: Processing: In this step the machine-readable form of input data is transformed into a more significant form, termed as information, in the CPU (Central Processing Unit).

Step 4: Output Information: In this step the processed data is obtained in the form of required output as information. This output information can also be used as input for some cases.

Step 5: Storage: In this step the processed data or the output information is stored in the digital storage devices, such as computer Hard Disk Drive (HDD), Compact Disc (CD), portable hard disk drive, pen drive, etc.

Procedures of Data Processing

Data processing is the significant step of data processing cycle. Following steps are included in the process of data processing:

- **Data Recording:** The data recording refers to the transferring of data into some specific form or document.
- **Data Duplication:** Data duplication means reproducing the data into various forms or documents.
- **Data Verification:** Data verification is done for carefully checking the recorded data for errors, if any.
- **Data Classification:** Data classification refers to sorting and sequentially organising data into specific categories.
- **Data Sorting:** Data sorting means arranging data in a specific order.
- **Data Merging:** Data merging means merging of two or more sets of data. All the data sets are sorted by means of similar key and putting them together to form a single sorted merged set of data.
- **Data Calculation:** Data calculation means performing numerical calculations on the input data using computing devices and specific software, such as MS Excel.

Check Your Progress

1. Define data processing concept.
2. How is the data processed?
3. Explain manual data processing method.
4. Elaborate on Electronic Data Processing or EDP method of data processing.
5. What is data processing cycle?

4.4 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. Data processing can be defined as a process of manipulating collected data and converting it into meaningful information. Data usually contain raw facts

NOTES

and figures about some specific topic or subject while information is the meaningful form of that data and is typically used for decision-making. Fundamentally, data is the input and information is the output.

2. Data can be processed manually using a pen and paper, mechanically using simple devices, such as typewriter or electronically using modern data processing tools, such as computers.
3. In the manual data processing method, the data is processed manually using paper sheets, registers, pens or pencils to obtain required results. In this method no device, calculator or any other tool is used. Therefore, in manual data processing methodology all the calculations and logical operations are manually performed on the data.
4. The Electronic Data Processing or EDP method of data processing is the modern method or technique used for processing the data. In the EDP method, the data is processed using the computer and the relevant data processing software. Data is input into the computer system and then the set of instructions are also input into the computer so that the computer can automatically process the data as per the given set of instructions.
5. Data processing method refers to the transformation of raw data into meaningful information as output. The data processing cycle is the set of processes and operations used for transforming data into significant information. The basic notion of the data processing cycle is to produce authentic information that can be used by the business organisations and enterprises for enhancing their business. The five significant steps of data processing cycle include data collection, data input, data processing, output information and data storage.

4.5 SUMMARY

- Data processing can be defined as a process of manipulating collected data and converting it into meaningful information.
- Data usually contain raw facts and figures about some specific topic or subject while information is the meaningful form of that data and is typically used for decision-making. Fundamentally, data is the input and information is the output.
- Data can be processed manually using a pen and paper, mechanically using simple devices, such as typewriter or electronically using modern data processing tools, such as computers.
- Data processing can be defined as “A process of converting data into meaningful information”.
- In the manual data processing method, the data is processed manually using paper sheets, registers, pens or pencils to obtain required results. In this method no device, calculator or any other tool is used.

NOTES

- In manual data processing methodology all the calculations and logical operations are manually performed on the data.
- The processed data is also physically transferred from one place to another manually in the form of hardcopies. But this method of data processing is extremely slow and there may be errors in the manually processed outputs.
- In the mechanical data processing method, the data is processed using the different types of simple devices, such as typewriters, mechanical printers or any other mechanical device.
- The mechanical data processing method is comparatively faster and more accurate as compared to manual data processing method.
- The Electronic Data Processing or EDP method of data processing is the modern method or technique used for processing the data. In the EDP method, the data is processed using the computer and the relevant data processing software.
- Data is input into the computer system and then the set of instructions are also input into the computer so that the computer can automatically process the data as per the given set of instructions. The computer system is also termed as the Electronic Data Processing or EDP machine.
- In the online processing method, the Internet connections and equipment are directly attached to a computer system. In this method, the data is stored in one place and can be used at different places.
- The distributed processing method is usually utilized by remote workstations which are connected to one big central workstation or server. ATMs (Automated Teller Machines) are examples of distributed data processing method, because all the end machines run on a fixed software located at a particular place and makes use of exactly same information and sets of instruction.
- Data analytics tools are specific applications and software used by data analysts for developing and performing the necessary data analytical processes that help the business organizations and companies for making enhanced, more informed business decisions however lowering costs and increasing profits.
- Hadoop is an open-source Big Data analytics tools which provides huge storage for all kinds of data. Basically, Hadoop has amazing processing power and proficiency for handling numerous tasks and is comparatively safe from hardware failure.
- Google provides a fully-managed enterprise data warehouse for data analytics through its BigQuery invention. It is server less and enables organizations to analyse any data by creating a logical data warehouse over managed, columnar storage, and data from object storage and spreadsheets.

- BigQuery holds data in real-time through its streaming assimilation feature and it is built atop the Google Cloud Platform.
- ZOHO Office Suite is an Indian web-based online office suite containing word processing, spreadsheets, presentations, databases, note-taking, wikis, web conferencing, Customer Relationship Management (CRM), project management, invoicing and other applications. It is developed by ZOHO Corporation.
- The spreadsheets are used for the processing of numerical data for organizing the data in the required manner specifying the formulae and functions processes that are predefined in the spreadsheet.
- Applications, such as MS EXCEL from Microsoft and CALC from Sun Microsystem's Open Office are typically used which simplify the calculation process and analysis of numerical data.
- SAS (Statistical Analysis System) is a statistical software suite developed by SAS Institute for data management, advanced analytics, multivariate analysis, business intelligence, criminal investigation, and predictive analytics.
- SPSS (Statistical Package for the Social Sciences) Statistics is a software package used for interactive, or batched, statistical analysis. SPSS is software for editing and analysing all sorts of data.
- Data processing method refers to the transformation of raw data into meaningful information as output.
- The data processing cycle is the set of processes and operations used for transforming data into significant information. The basic notion of the data processing cycle is to produce authentic information that can be used by the business organisations and enterprises for enhancing their business.
- The five significant steps of data processing cycle include data collection, data input, data processing, output information and data storage.

4.6 KEY WORDS

- **Online processing:** In the online processing method, the Internet connections and equipment are directly attached to a computer system.
- **Distributed processing:** The distributed processing method is usually utilized by remote workstations which are connected to one big central workstation or server.
- **SAS (Statistical Analysis System):** SAS (Statistical Analysis System) is a statistical software suite developed by SAS Institute for data management, advanced analytics, multivariate analysis, business intelligence, criminal investigation, and predictive analytics.

- **SPSS (Statistical Package for the Social Sciences):** SPSS Statistics is a software package used for interactive, or batched, statistical analysis. SPSS is software for editing and analysing all sorts of data.
- **Data processing cycle:** The data processing cycle is the set of processes and operations used for transforming data into significant information.
- **Data verification:** Data verification is done for carefully checking the recorded data for errors, if any.
- **Data classification:** Data classification refers to sorting and sequentially organising data into specific categories.
- **Data sorting:** Data sorting means arranging data in a specific order.
- **Data merging:** Data merging means merging of two or more sets of data. All the data sets are sorted by means of similar key and putting them together to form a single sorted merged set of data.
- **Data calculation:** Data calculation means performing numerical calculations on the input data using computing devices and specific software, such as MS Excel.

4.7 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. Explain the term data processing systems.
2. What are the different types of data processing system?
3. Describe the online processing and distributed processing methods.
4. Elaborate the term data processing tools.
5. Explain the statistical data analysis tools.
6. Why a data processing cycle is essential?
7. Define the significant steps of data processing cycle.

Long-Answer Questions

1. Briefly explain the significance of data processing systems giving examples.
2. Discuss in detail about the different types of data processing systems giving appropriate examples.
3. Explain briefly the different tools of data analysis and data processing giving their features.
4. Discuss briefly the concept of data processing cycle giving appropriate examples.
5. Explain the five fundamental steps of the data processing cycle giving example and illustration.

4.8 FURTHER READINGS

Price, Wilson T. 1981. *Introduction to Computer Data Processing*, 3rd Edition. United States: Holt-Saunders International Editions.

Verzello, Robert J. and John Reutter. 1983. *Data Processing: Systems and Concepts*. New York: The McGraw-Hill Publishing Company, Inc.

Beech, Carol and Janice Burn. 1985. *Applications in Business Data Processing*. London (England): Pitman Publisher.

Fuori, William M. and Lawrence J. Aufiero. 1987. *Computers and Information Processing*. United States: Prentice Hall.

Sanders, Donald H. 1988. *Computers Today*. New York: McGraw-Hill.

Sinha, Pradeep K. and Priti Sinha. 2004. *Computer Fundamentals: Concepts, Systems & Applications*, 8th Edition. New Delhi: BPB Publications.

Rajaraman, V. 1999. *Fundamentals of Computers*. New Delhi: Prentice-Hall of India.

Lipschutz, Martin M. and S. Lipschultz. 1982. *Theory and Problems of Data Processing, Schaum's Outline Series*. New Delhi: Tata McGraw Hill Publishing Co. Ltd.

Stallings, William. 1996. *Data and Computer Communications*. New Jersey (USA): Prentice-Hall PTR.

UNIT 5 DATA PROCESSING COMPONENTS

NOTES

Structure

5.0 INTRODUCTION

Data processing is done when data is collected and translated into usable information. Usually performed by a data scientist or team of data scientists, it is important for data processing to be done correctly as not to negatively affect the end product, or data output. Data processing starts with data in its raw form and converts it into a more readable format (graphs, documents, etc.), giving it the form and context necessary to be interpreted by computers and utilized by employees throughout an organization.

The problem statement stage is referred as the first and most significant phase of solving an analytics database problem. It can create or disrupt the entire project. Planning is the process of thinking about the activities required to achieve a desired goal. It is the first and foremost activity to achieve desired results. It involves the creation and maintenance of a plan. As such, planning is a fundamental property of intelligent behaviour. Algorithm is a sequence of well-defined steps that can be implemented in a computer system to solve a class of problems or to perform a computation.

In this unit, you will learn about the components of data processing, definition and planning, and algorithms.

5.1 OBJECTIVES

After going through this unit, you will be able to:

- Discuss the components of data processing system
- Understand the importance of problem definition and planning

- Define algorithm
- Discuss the properties of algorithm
- Explain the criteria for algorithm design

NOTES

5.2 COMPONENTS OF A DATA PROCESSING SYSTEMS

Data processing is done when data is collected and translated into usable information. Usually performed by a data scientist or team of data scientists, it is important for data processing to be done correctly as not to negatively affect the end product, or data output.

Data processing starts with data in its raw form and converts it into a more readable format (graphs, documents, etc.), giving it the form and context necessary to be interpreted by computers and utilized by employees throughout an organization.

Fundamentally, a data processing system is a combination of machines, people, and processes that for a set of inputs produces a defined set of outputs. The inputs and outputs are interpreted as data, facts, information, etc.

Data processing term is commonly used synonymously with data or storage (codes) processing system is information system. With regard particularly to Electronic Data Processing (EDP), the corresponding concept is referred to as electronic data processing system.

A data processing system may involve some combination of:

- **Conversion** – Converting data to another form or Language.
- **Validation** – Ensuring that supplied data is clean, correct and useful.
- **Sorting** – Arranging items in some sequence and/or in different sets.
- **Summarization** – Reducing detail data to its main points.
- **Aggregation** – Combining multiple pieces of data.
- **Analysis** – Collection, organization, analysis, interpretation and presentation of data.
- **Reporting** – List detail or summary data or computed information.

Following are the basic components of a data processing system:

1. Connecting API (JDBC, ODBC, CLI).
2. Session Controller.
3. Tokenizer and Syntax Analyzer.
4. Symbol Resolver.
5. Semantic Analyzer.
6. Rule Based Optimizer.

7. Cost Based Optimizer.
8. Query Access Planner.
9. Step Dispatcher and Scheduler.
10. Step Executor.
11. File System.
12. Operating System Interface.
13. Memory Manager.
14. Performance Booster Designs.

NOTES

Types of Data Processing Systems

Following are the various types of data processing systems which are commonly used in different organizations. The data processing system is broadly categorised on the basis of 'Application Area' and 'Service Type'.

By Application Area

Scientific Data Processing: Scientific data processing usually involves a great deal of computation (arithmetic and comparison operations) upon a relatively small amount of input data, resulting in a small volume of output.

Commercial Data Processing: Commercial data processing involves a large volume of input data, relatively few computational operations, and a large volume of output. Accounting programs are the prototypical examples of data processing applications. Information Systems (IS) is the field that studies such organizational computer systems.

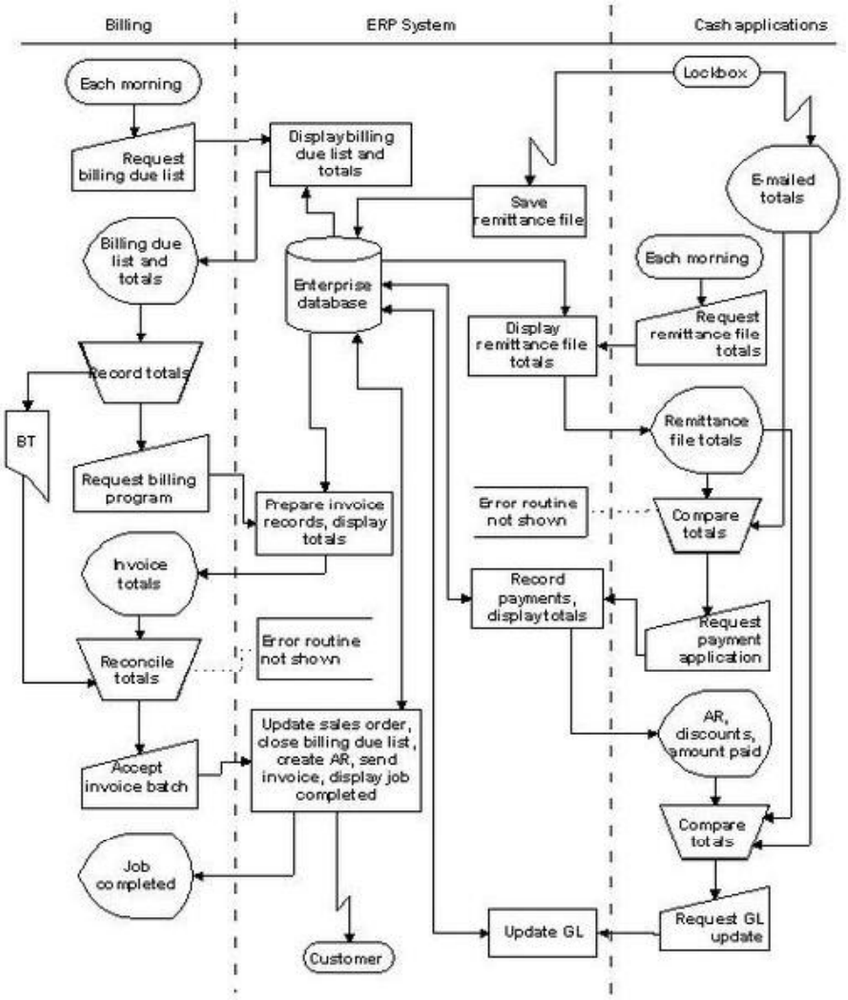
Data Analysis: Data analysis is a body of methods that help to describe facts, detect patterns, develop explanations, and test hypotheses, for example, data analysis might be used to look at sales and customer data to identify connections between products to allow for cross selling campaigns.

By Service Type

- Transaction Processing Systems
- Information Storage and Retrieval Systems
- Command and Control Systems
- Computing Service Systems
- Process Control Systems
- Message Switching Systems

Following flowchart is a real-world example of a data processing system combining manual and computerized processing to handle accounts receivable, billing, and general ledger.

NOTES



Six Stages of Data Processing

1. **Data Collection:** Collecting data is the first step in data processing. Data is pulled from available sources, including data lakes and data warehouses. It is important that the data sources available are trustworthy and well-built so the data collected (and later used as information) is of the highest possible quality.
2. **Data Preparation:** Once the data is collected, it then enters the data preparation stage. Data preparation often referred to as 'Pre-Processing'is the stage at which raw data is cleaned up and organized for the next/following stages of data processing. During preparation, raw data is diligently checked for any errors. The purpose of this step is to eliminate bad data (redundant, incomplete, or incorrect data) and begin to create high-quality data for the best business intelligence.

NOTES

3. **Data Input:** The clean data is then entered into its destination (perhaps a CRM like Salesforce or a data warehouse like Redshift), and translated into a language that it can understand. Data input is the first stage in which raw data begins to take the form of usable information.
4. **Processing:** During the processing stage, the data inputted to the computer in the previous stage is actually processed for interpretation. Processing is done using machine learning algorithms, though the process itself may vary slightly depending on the source of data being processed (data lakes, social networks, connected devices, etc.) and its intended use (examining advertising patterns, medical diagnosis from connected devices, determining customer needs, etc.).
5. **Data Output/Interpretation:** The output/interpretation stage is the stage at which data is finally usable to non-data scientists. It is translated, readable, and often in the form of graphs, videos, images, plain text, etc.). Members of the company or institution can now begin to self-serve the data for their own data analytics projects.
6. **Data Storage:** The final stage of data processing is storage. After all of the data is processed, it is then stored for future use. While some information may be put to use immediately, while some can be used later. Additionally, properly stored data is a necessity for compliance with data protection legislation like GDPR (General Data Protection Regulation). When data is properly stored, it can be quickly and easily accessed by members of the organization when required.

5.3 PROBLEM DEFINITION

The problem statement stage is referred as the first and most significant phase of solving an analytics database problem. It can create or disrupt the entire project. When a business approaches a data scientist with a problem they want to solve, they will always define the problem in layman's terms. This means the problem will not be clear enough, from an analytics point of view, to begin solving it right away. The problem needs to be well framed.

The data scientist analyses the problem statement in mathematical terms. A good data science problem should be relevant, specific, and unambiguous. It should align with the business strategy.

A problem statement is a concise description of an issue to be addressed or a condition to be improved upon. It identifies the gap between the current (problem) state and desired (goal) state of a process or product. Focusing on the facts, the problem statement should be designed to address the Five Ws. The first condition of solving a problem is understanding the problem, which can be done by way of a problem statement.

NOTES

Problem statements are widely used by most businesses and organizations to execute process improvement projects. A simple and well-defined problem statement will be used by the project team to understand the problem and work toward developing a solution. It will also provide management with specific insights into the problem so that they can make appropriate project-approving decisions. As such, it is crucial for the problem statement to be clear and unambiguous.

The main purpose of the problem statement is to identify and explain the problem. This includes describing the existing environment, where the problem occurs, and what impacts it has on users, finances, and ancillary activities. Additionally, the problem statement is used to explain what the expected environment looks like. Defining the desired condition provides an overall vision for the process or product. It makes clear the purpose for initiating the improvement project and the goals that it is meant to accomplish.

Another important function of the problem statement is to be used as a communication device. A problem statement helps with obtaining buy-in from those involved in the project. Before the project begins, the stakeholders verify the problem and goals are accurately described in the problem statement. Once this approval is received, the project team reviews it to ensure everyone understands the issue at hand and what they are trying to accomplish. This also helps define the project scope, which keeps the project concentrated on the overall goal.

The problem statement is referenced throughout the project to establish focus within the project team and verify they stay on track. At the end of the project, it is revisited to confirm the implemented solution indeed solves the problem. A well-defined problem statement can also aid in performing root-cause analysis to understand why the problem occurred and ensure measures can be taken to prevent it from happening in the future.

Remember that the problem statement does not define the solution or methods of reaching the solution. The problem statement simply recognizes the gap between the problem and goal states. It can be said that, "A problem well stated is half solved".

However, there are often multiple, viable solutions to a problem. Only after the problem statement is written and agreed upon should the solution(s) be discussed and the resulting course of action determined. Problem statements can vary in length, depending on the complexity of the problem.

5.4 PLANNING

Planning is the process of thinking about the activities required to achieve a desired goal. It is the first and foremost activity to achieve desired results. It involves the creation and maintenance of a plan. As such, planning is a fundamental property of intelligent behaviour. An important further meaning, often just called 'Planning', is

NOTES

the legal context of permitted building developments. Also, planning has a specific process and is necessary for multiple occupations (particularly in fields, such as management, business, etc.). In each field there are different types of plans that help companies achieve efficiency and effectiveness.

Planning a database correctly is essential to the success of your application and its usefulness to the people managing it and the organisation it supports.

Effective database planning means that the software is proficient of managing and consolidating all the data that is specifically generated and relied upon by the organisation or the business. A good database plan will permit the organisation to develop a clear structure for the technique using which data is stored and managed by the staff or the application using it. This process will also help to identify a base for future requirements and could also provide the framework for both an immediate and long term growth strategy of the organization or business.

Relational Database Management System (RDBMS) permit for multiple types of data 'Objects' to be stored with keys and indexes to link related data across multiple tables. For example, one table holds customer details and assigns each customer a numeric ID, while another table holds order details, including the customer ID - removing the need to duplicate customer details if there are multiple orders from the same customer. This methodology is referred to as 'Data Normalisation' and in its simplest form, attempts to minimise data redundancy. Normalised database structure ensures concise database architecture with the elimination of unnecessary data duplication, as well as the potential for optimised database operations and balancing workload between the 'Database Server' and the 'Web Server'. Ensuring the correct plan and structure for the database also has arguably the most essential benefit; thereby greatly reducing the potential for human error, making the data flowing through the bespoke software robust, reliable and scalable.

Database planning must be integrated with the overall IS (Information Strategy) strategy of the organization.

Following are three main issues involved in formulating an IS strategies:

- Identification of enterprise plans and goals with the subsequent purpose of information systems requirements.
- Evaluation of current information systems to find out existing strengths and weaknesses.
- Appraisal of IT opportunities that might yield aggressive advantage.

An essential first step in database planning is to define the mission statement for the database system. The mission statement describes the primary aims of the database system. Those are driving the database project within the organization that generally defines the mission statement. A mission statement helps to simplify the purpose of the database system and provide a clearer path towards the efficient and effective creation of the required database system.

NOTES

5.5 ALGORITHM

In general, an algorithm can be defined as a well-defined procedure for computation that takes an input of some values and produces an output. Hence, an algorithm is a sequence of steps which transform an 'input' into 'output'.

It can also be considered as a tool for solving a well-specified 'computational problem'. The problem statement defines the input/output relationship. The algorithm lists a specific computational procedure to reach that input/output relationship.

To put it simply, 'An algorithm is a step-by-step procedure for performing some task in a finite amount of time.'

Definition

An algorithm is composed of a finite set of steps each of which may require one or more operations. Every operation may be characterized as either a simple or complex. Operations performed on scalar quantities are termed simple, while operations on vector data normally termed as complex.

Properties of Algorithm

The following are the five important properties (features) of algorithm:

- Finiteness
- Definitiveness
- Input
- Output
- Effectiveness
- **Finiteness:** An algorithm must always terminate after a finite number of steps. If we trace out the instructions of an algorithm, then for all cases, the algorithm terminates after a finite number of steps.
- **Definitiveness:** Each operation must have a definite meaning and it must be perfectly clear. All steps of an algorithm need to be precisely defined. The actions to be executed in each case should be rigorously and clearly specified.
- **Inputs:** An algorithm may have zero or more 'inputs' quantities. These inputs are given to the algorithm initially prior to its beginning or dynamically as it runs. An input is taken from a specified set of objects. Also, it is externally supplied to the algorithm.
- **Output:** An algorithm has one or more 'output' quantities. These quantities have a specified relation to the inputs. An algorithm produces at least one or more outputs.
- **Effectiveness:** Each operation should be effective, i.e. the operation must be able to carry out in a finite amount of time.

NOTES

An algorithm is usually supposed to be 'effective' in the sense that all its operations needs to be sufficiently basic so that they can in principle be executed exactly in the same way in a finite length of time by someone using pencil and paper.

Criteria for Algorithm Design

Several active areas of research are included in the study of algorithms. The following four distinct areas can be identified:

1. **Devising algorithms:** The creation of an algorithm is an art. It may never be fully automated. A few design techniques are especially useful in fields other than computer science, such as operations research and electrical engineering. All of the approaches we consider have applications in a variety of areas including computer science. However, some important design techniques such as linear, non-linear and integer programming are not covered here as they are traditionally covered in other courses.
2. **Validating algorithms:** Once you have devised an algorithm, you need to show that it computes the correct answer for all possible legal inputs. This process is referred to as algorithm validation. It is not necessary to express the algorithm as a program. If it is stated in a precise way, it will do. The objective of the validation is to assure the user that the algorithm will work correctly and independently of the issues concerning the programming language in which it will eventually be written. After the validity of the method is shown, it is possible to write the program. On the completion of writing the program, a second phase begins. This phase is called program providing or sometimes as program verification. A proof of correctness requires that the solution be stated in two forms. One form is usually as a program, which is annotated by a set of assertions about the input and output variables of the program. The second form is called specification and this may also be expressed in the predicate calculus. A proof consists of showing that these two forms are equivalent in that for every given legal input, they describe the same output. A complete proof of program correctness requires that each statement of the programming language be precisely defined and all basic operations be proved correct. All these details may cause a proof to be very much longer than the program.
3. **Analysing algorithms:** As an algorithm is executed, it uses computer's central processing unit (CPU) for performing operation. It also uses the memory for holding the program and its data. Analysis of algorithms is the process to determine the computing time and storage required by an algorithm.
4. **Testing a program:** Testing of a program comprises two phases: (i) debugging and (ii) profiling.

NOTES

Debugging refers to the process of carrying out programs on sample data sets for determining if there are any faulty results. If any faulty result occurs, it is corrected by debugging. A proof of correctness is much more valuable than a thousand tests since it guarantees that the program will work correctly for a possible input.

Profiling refers to the process of the execution of a correct program on data sets and the measurement of the time and space it takes in computing the results. It is useful in the sense that it confirms a previously done analysis and points out logical places for performing useful optimization.

An example of profiling:

If we wish to measure the worst-case performance of the sequential search algorithm, we need to do the following:

- Decide the values of n for which the times are to be obtained
- Determine for each of the above value of n the data that exhibits the worst-case behaviour

Algorithm for sequential search:

```
1. Algorithm seqsearch (a, x, n)
2. //search for x in a[l: n] . a[0] is used as additional
   space
3. {
4. i := n; a[0] := x;
5. while(a[i] * x) do i := i " 1;
6. return i;
7. }
```

The decision on which the values of n to use is based on the amount of timing we wish to perform and also on what we expect to do with the times once they are obtained. Assume that for algorithm, our interest is to simply predict how long it will take, in the worst case, to search for `x`, given the size `n` of `a`.

Algorithms as Technology

If computers were infinitely fast and computer memory was free, you would be in a position to adopt any correct method to solve a problem. In all likelihood, you would like your implementation to be adhering to good software engineering practice. However, you would use the method which is the easiest to implement.

However, computers may be fast, but they cannot be infinitely fast. Similarly, memory may be cheap, but it cannot be free. Thus, computing time and space in memory are bounded resources . You need to use these resources wisely. Such algorithms which are efficient in terms of time or space will be helpful.

NOTES

Efficiency

It has been found that algorithm devices used for solving the same problem usually differ considerably in their efficiency. These differences are more significant than those due to hardware and software.

Algorithms and Other Technologies

Algorithms are indeed very important for contemporary computers considering other advanced technologies:

- Hardware with high clock rates, pipelining and super scalar architectures
- Easy to use, intuitive graphical user interfaces (GUIs)
- Local area networking (LAN) and wide area networking (WAN)

A truly skilled programmer possesses a solid algorithmic knowledge and technique. It separates him/her from a novice. It is true that with modern computing technology, you can perform some tasks even if you do not have much knowledge of algorithms. However, if you are with a good background in algorithms, you can perform much more.

Types of Algorithms

- Approximate algorithm
- Probabilistic algorithm
- Infinite algorithm
- Heuristic algorithm

Approximate Algorithm

An algorithm is said to approximate if it is infinite and repeating., e.g. $\sqrt{2} = 1.414$, $\sqrt{3} = 1.732$, $\pi = 3.14$, etc.

Probabilistic Algorithm

If the solution of a problem is uncertain, then it is called probabilistic algorithm, e.g. tossing of a coin

Infinite Algorithm

An algorithm, which is not finite, is called infinite algorithm, e.g. a complete solution of a chessboard, division by zero.

Heuristic Algorithm

Giving less inputs and getting more outputs is called heuristic algorithm.

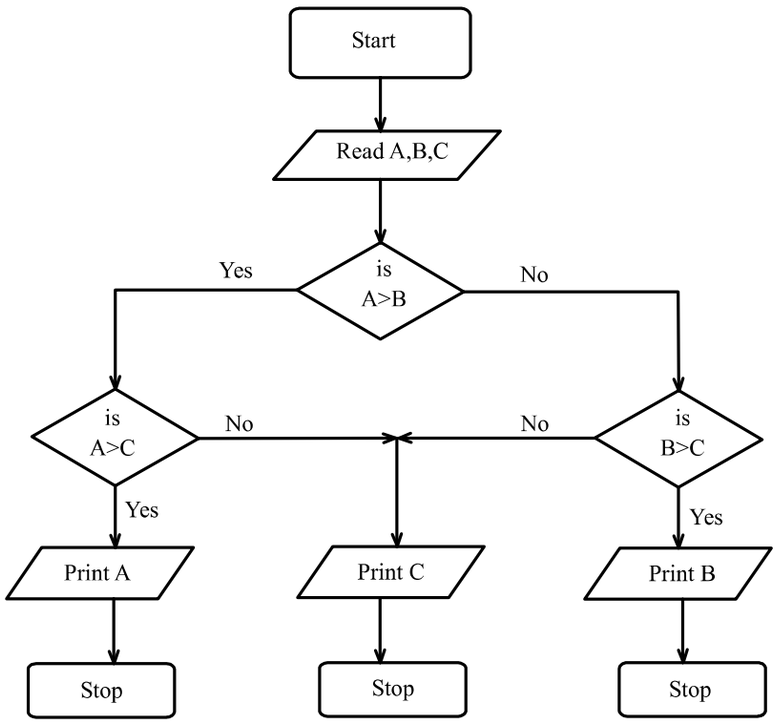
Design of an Algorithm

Example 5.1: Algorithm to pick the largest of three numbers.

```
Step 1: Read A, B, C.
Step 2: If A > B, go to Step 3.
     Else go to Step 5.
```

NOTES

Step 3: If A > C
    Print A as the largest number.
    Else
    Print C as the largest number.
Step 4: Stop.
Step 5: If B > C
    Print B as the largest number.
    Else
    Print C as the largest number.
Step 6: Stop.



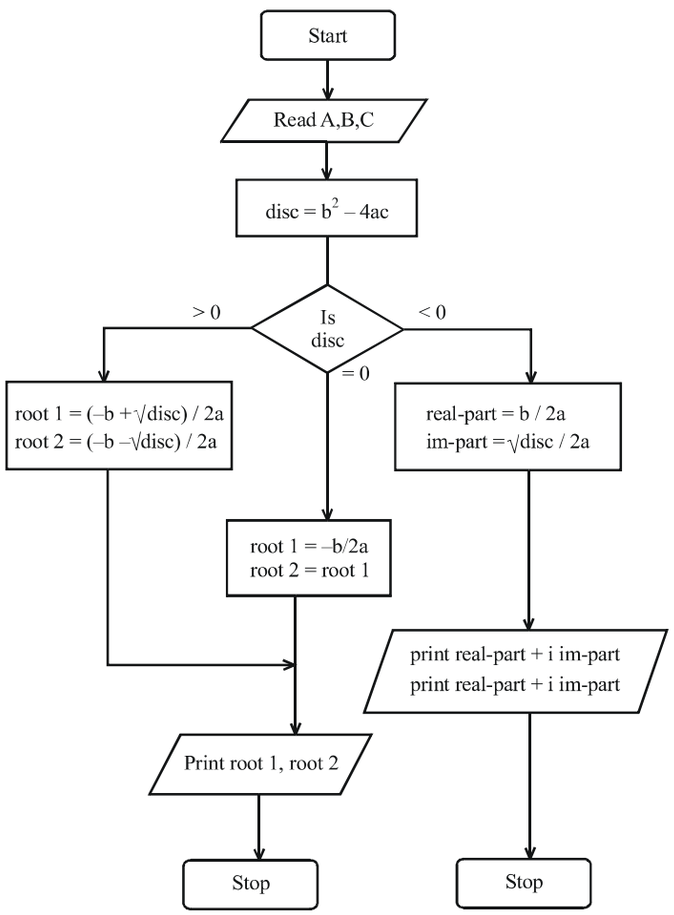
Explanation: Read the three numbers A, B and C. A is compared with B. If A is larger, then it is compared with C. If A turns out to be the largest number again, then A is the largest number; otherwise, C is the largest number. If in the second step, A is less than or equal to be B, then B is compared with C. If B is larger, then B is the largest number; otherwise, C is the largest number.

Example 5.2: Algorithm to find the roots of a quadratic equation $ax^2 + bx + c = 0$ for all cases.

Step 1: Read a, b, c.
Step 2: disc ← b^2 - 4ac.
Step 3: If disc ← 0, go to Step 4.

NOTES

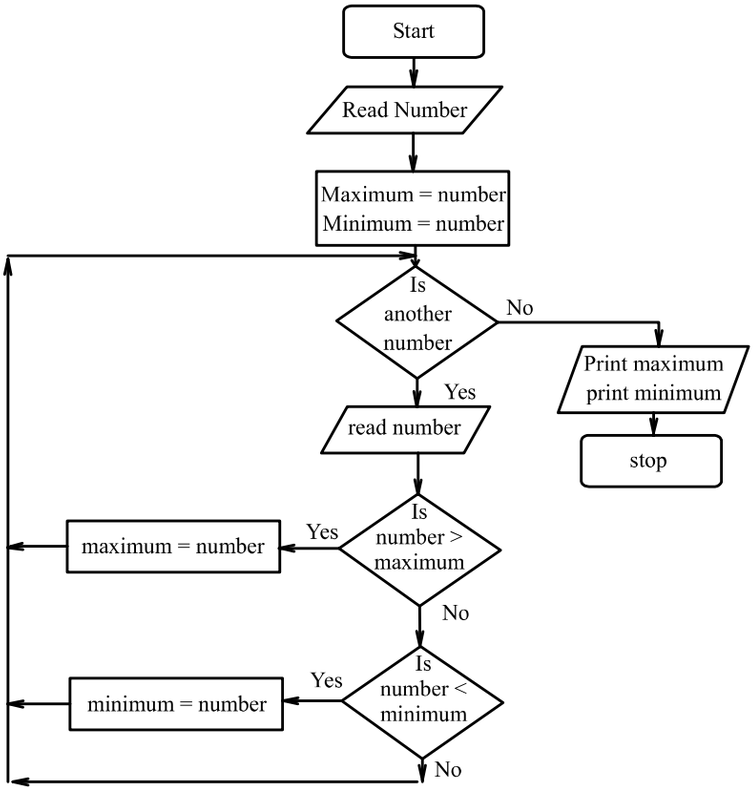
```
     Else, if disc > 0, go to Step 5.
     Else, go to Step 6.
Step 4: root 1 ← – b/2a.
     root 2 ← root 1.
     go to Step 7.
Step 5: Root 1 ← (–b + sqrt (disc)) / 2a.
     Root 2 ← (–b – sqrt (disc)) / 2a.
     go to Step 7.
Step 6: real-part ← –b / 2a.
     im-part ← sqrt(–disc) / 2a.
     Print real-part + i im-part.
     Print real-part – i im-part.
     Stop.
Step 7: Print root 1, root 2.
     Stop.
```



NOTES

Example 5.3: Algorithm for finding maximum and minimum numbers.

```
Step 1: Read number.
Step 2: Maximum ← number. Minimum ← number.
Step 3: If (another number) go to Step 4.
          Else go to Step 7.
Step 4: Read number.
Step 5: If number > Maximum Maximum = number.
          Else if number < Minimum
          Minimum = number.
Step 6: go to Step 3.
Step 7: Print Maximum.
          Print Minimum.
Step 8: Stop.
```



Example 5.4: Algorithm for finding maximum and minimum of given n numbers.

```
Step 1: Read N.
Step 2: Counter ← 1.
     Read number.
```

Maximum ← number.

Minimum ← number.

Step 3: If Counter < N go to Step 4.

Else go to Step 7.

Step 4: Read number.

Counter ← Counter + 1.

Step 5: If number > Maximum

Step 6: Maximum ← number.

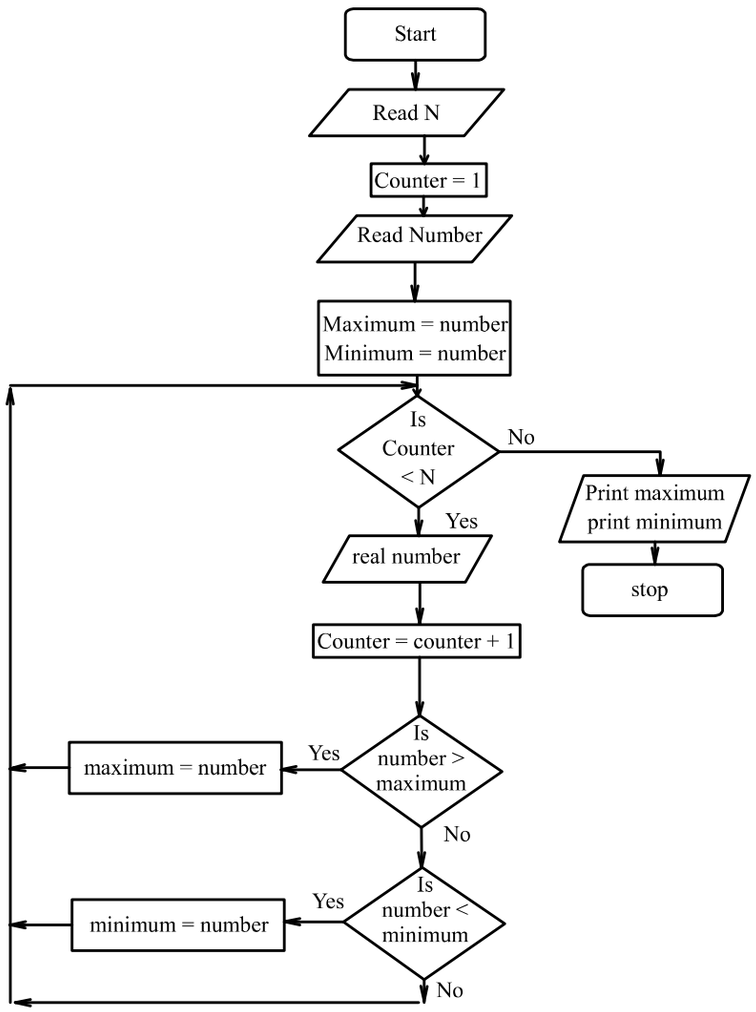
Step 7: Else If number < Minimum

Step 8: Minimum ← number.

Step 9: go to Step 3.

Step 10: Print Maximum. Print Minimum.

Step 11: Stop.



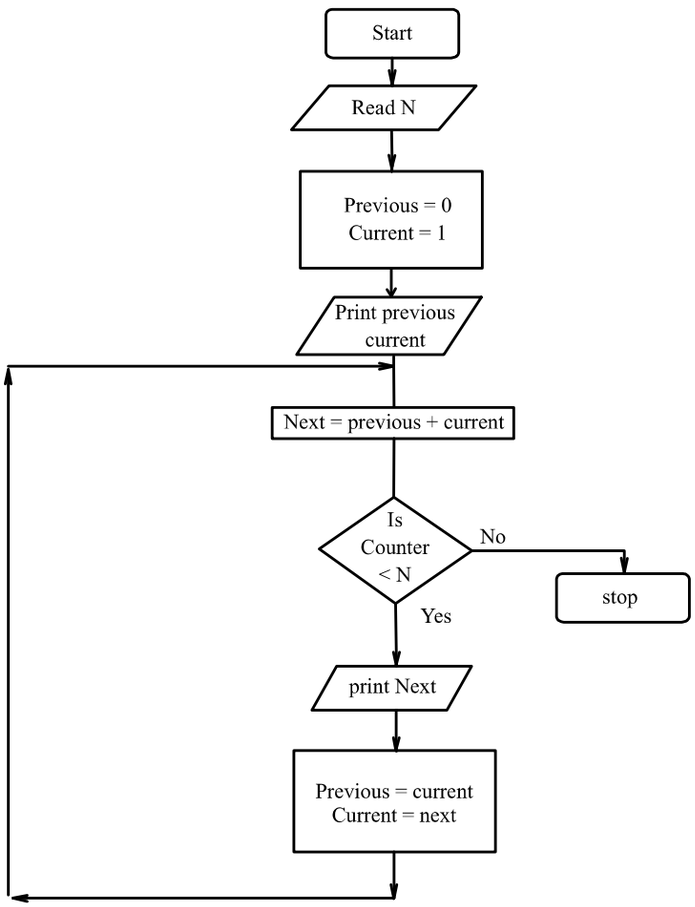
NOTES

NOTES

Example 5.5: Algorithm for generating Fibonacci numbers up to N.

The first and second terms in the Fibonacci series are 0 and 1. The third and subsequent terms in the sequence are found by adding the preceding two terms in the series. The Fibonacci series is: 0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, ...

```
Step 1: Read N.
Step 2: Previous ← 0. Current ← 1.
    Print Previous, Current.
Step 3: Next ← Previous + Current.
Step 4: If Next < N
    Print Next.
    Previous ← Current.
    Current ← Next,
    go to Step 3.
    Else go to Step 5.
Step 5: Stop.
```



NOTES

Example 5.6: Algorithm for generating first k Fibonacci numbers.

Step 1: Read k.

Step 2: Counter ← 2.

Previous ← 0.

Current ← 1.

Print 'First k Fibonacci numbers are:'

Print Previous, Current.

Step 3: Next ← Previous + Current.

Counter ← Counter + 1.

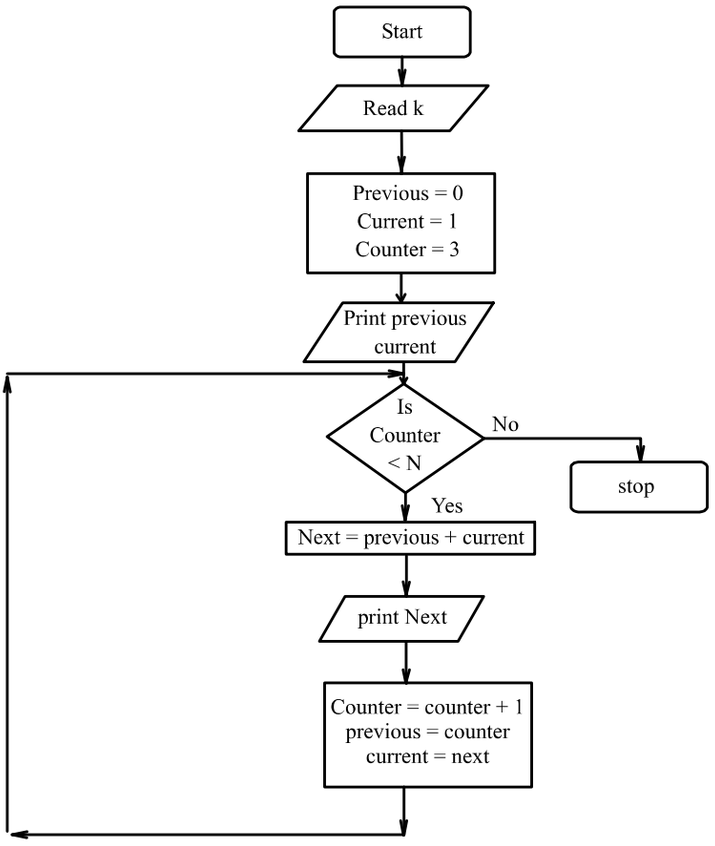
Print Next. Previous ← Current.

Current ← Next.

Step 4: If (Counter < k) go to Step 3.

Else go to Step 5.

Step 5: Stop.



NOTES

Example 5.7: Sum of first n Factorials

The factorial of a non-negative integer n is the product of all positive integers less than or equal to n and is denoted by n ! It is defined as follows:

```
n! = n(n-1) … 2*1
```

For example, 5!=5*4*3*2*1. Its older notation was $\lfloor n$. In factorial number system where the denominations are 1, 2, 6, 24, 120, …, etc. the nth digit is in the range 0 to n. This identity works as shown in this example:

```
1*1!+2*2!+3*3!+ … +k*k! = (k+1)! – 1
Sum of 2!+3! = (2*1) + (3*2*1) = (2) + (6) = 8
```

The following algorithm is used to find the sum of n factorials:

Algorithm of sum of n factorials:

```
Step 1: integer n, factorial, i, j, sum;
Step 2: sum←0;
Step 3: print 'Enter the number';
Step 4: read n;
Step 5: for i←1 to n //Running outer loop till n value
{
factorial←1
for j←1 to i
//Inner loop to calculate the sum of n factorial values
{
factorial ←factorial*j;
//Calculating n factorial values
}
sum←sum+factorial;
//Calculating sum of n factorial values
}
Step 6: print 'Sum of n Factorials';
Step 7: print sum;
//Print the sum value of n factorials
```

Implementation to Find the Sum of First n Factorials

```
/*———— START OF PROGRAM ————*/
#include <stdio.h> //Declaration of Header files
#include <conio.h>
/*——-1/1! + 2/2! + 3/3! + 4/4! ...-——*/
void main()
{
int factorial,sum=0,i,j,n;
//Declaring and assigning the variables
```

NOTES

```
printf("Enter a value for [n] value = ");
scanf("%d", &n);
//Accept input value for n term
for(i=1;i<=n;i++) //For outer loop till n value
{
factorial=1;
for(j=1;j<=i;j++)
//Using inner for loop to calculate the sum of n factorial
{
factorial = factorial *j;
}
sum=sum+ factorial;
}
Printf("\n sum of %d factorial = %d", n, sum);
}
getch();
}
```

The result of the preceding program is as follows:

```
Enter a value for [n] value = 4
sum of 4 factorial = 33
```

How this program works it is explained stepwise in the following:

```
1!+2!+3!+4! = 33
The value of 1! = 1
The value of 2! = 2
The value of 3! = 6
The value of 4! = 24
1+2+6+24= 33
```

Example 5.8: *To find largest value and the second largest value of the list:* The largest and second largest values in the given list are determined by array implementation. Array can contain the various elements of the list. The algorithm to find the largest and second largest of given list is as follows:

Algorithm to find the largest value and second largest value of the given list

```
Step 1: integer M, a[M], i, largest, t, second_largest;
Step 2: print 'Enter a value for array';
Step 3: read M;
Step 4: for i←1 to M
print 'Enter values:';
read a[i];
Step 5: if i==1
largest←t←second_largest←a[i];
```

NOTES

```
Step 6: else if a[i]>largest
second_largest←largest;
largest←a[i];
Step 7: else if a[i]>second_largest && a[i]<largest
second_largest←a[i];
Step 8: else if a[i]<t
t←a[i];
Step 9: print 'Largest value in the given list =';
Step 10: print largest;
Step 11: 'Second largest value in the given list =';
Step 12: print second_largest;
```

The result of this algorithm is as follows:

```
Enter a value for array
5
```

Then `array A[M]` is assigned a value 5 as `A[5].`

The input values are entered in the following way:

```
Enter values
45
90
112
4
35
Largest value in the given list = 112
Second largest value in the given list = 90
```

The first element of the array is `45` which is assumed to be the largest value and it is kept in the temporary location where it is temporarily stored in variable `t`. All the remaining values are checked from this number. Now, the `A[i]` value is assigned as `45`. At second step, the condition is satisfied so largest value is `90`. Now, `90` is checked with the next entered value `112`. As the condition is not satisfied, `112` is assumed as greater value. The values `4` and `35` are less than `90`, so the condition for less than largest is not satisfied. The checking process of second largest value `'45'` is done after checking the rest four values and declaring `112` as first largest value. Further, the statement `'second_largest=largest'` is used. The first element of the array is again taken as largest among the four values. Now, `45` is checked step-by-step in `if else if` conditional statement to find the second largest value.

Program to find the largest value and second largest value in a given list

```
/*———————— START OF PROGRAM ————————*/
#include <stdio.h>
#include <conio.h>
```

NOTES

```
#define M 5 //Define preprocessor directive that assigns
M = 5
void main()
{
int a[M], i, largest, t, second_largest;
clrscr(); //Clear the screen of previous
for(i=1; i<=M); i+)
{
printf("Enter %d value");
scanf("%d", &a[i]);
if(i==1)
largest=t=second_largest=a[i];
//The largest, t, second_largest values are assigned as
the value of a[i].
if(a[i]>largest)
{
second_largest=largest;
largest=a[i];
}
if(a[i]>second_largest && a[i]<largest)
second_largest=a[i];
if(a[i])<t)
t=a[i];
}
printf("\nLargest Vvalue in the given list = %d", largest);
printf("\nSecond largest value in the given list = %d",
second_largest);
getch();
}
```

The result of the preceding program is as follows:

```
Enter 1 value = 45
Enter 2 value = 90
Enter 3 value = 112
Enter 4 value = 4
Enter 5 value = 35
Largest value in the given list =112
Second largest value in the given list=90
```

In this program, `#define M 5` statement defines preprocessor directive that works as a macro. It means wherever `M` comes in the program, its value `5` is changed automatically. The `#define` statement can not be terminated by a semicolon (;) because the preprocessor is a program that comes before `main()` statement.

NOTES

Example 5.9: Determining *n*th root of a number.

The *n*th root of a number *a* is a number where *n* is positive integer. The nth roots are taken with the following iteration where *a* is the input values and *n* is the root value to be taken. The equation is arranged in the following ways:

$$b = \sqrt[n]{a}$$

where *a* is the number, *n* is the *n*th root and *b* is the value that retains the nth root of number a. For example, nth root is equal to 3 and number *a* is equal to 2 can be written as $2^3 = 8$. The following algorithm is used to find out the *n*th root of a given value:

Algorithm to find the *n*th root of a number:

```
Step 1:   double calculate_root(double,double);
//Declare a calculate_root function having two parameters
Step 2:   double Find_nth_Root(double,double,double);
//Declare a calculate_root function having three parameters
Step 3: double number(double,double);
//Declare a number function having two parameters
Step 4: double x←1, NUMBER_OF_ITERATIONS←40, n;
//Assign value 1 to x variable and NUMBER_OF_ITERATIONS=40
Step 5: double N;
//Declare a variable N as double data type
Step 6: double root;
//Declare a variable root as double data type
Step 7: x_label:
//Assign a label named as x_label
Step 8: print 'Enter root do want [2,3, …5] ?'
Step 9: read n;
//Accept input value n
Step 10: if n<=0
//Check the condition where n is less than 0
Step 11: print 'Number should be Greater than 0';
Step 12: print n;
Step 13: goto x_label;
//Go to label on x_label
Step 14: y_label:
//Assign a label named as y_label
Step 15: print 'Enter the number for Root';
Step 16: read N;
Step 17: if N<=0
Step 18: print 'Number should be greater than 0';
```

NOTES

```
Step 19: print 'PRESS ANY KEY TO ENTER AGAIN';
Step 20: goto y_label;
//Go to label on y_label
Step 21: x←calulate_root(n,N);
//x retains the returned value of function calculate_root
Step 22: print 'The first assumed root is',x;
Step 23: root←Find_nth_Root(N,n,x);
//root retains the Find_nth_Root returned value
Step 24: print 'Root of n',n;
Step 25: print N;
Step 26: print root;
Step 27: double calculate_root(double n,double N)
Step 28: integer i,xr;
//integer i and xr are declared
Step 29: xr←1;
//xr is assigned as 1
Step 30: double j←1;
//double j is assigned as 1
Step 31: while(1)
Step 32: for i←0 to n //Running for loop
    {
    xr←xr*j; //xr retains the value of xr*j
    }
Step 33: if xr>N
    Return j-1; //Returns j-1
Step 34: j←j+1; //j value is increased by 1
Step 35: xr←1; //xr value is increased by 1
Step 36: double Find_nth_Root(double NUM,double n,double
X0) //Function Find_nth_Root starts from here.
Step 37: int i;
Step 38: double d←1.0;
Step 39: double first_term, second_term, root←X0;
Step 40: for i←1 to NUMBER_OF_ITERATIONS
//Body of for loop starts that calculates first term and
second term value of enter values of NUMBER_OF_ITERATIONS
Step 41: d←number(root,n);
//d retains the n th value of given number.
Step 42: first_term←((n-1)/n)*root;
// first_term retains the value of let say 5 (5-1)/5)*
root value
```

NOTES

```
Step 43: second_term←(1/n)*(NUM/d);
Step 44: root←first_term+second_term;
Step 45: print first_term,second_term,root;
Step 46: return root;
Step 47: double number (double x,double n)
Step 48: double d←1;
Step 49: integer i;
Step 50: for i←1 to n-1
Step 51:  d←d*x; //Printing the final nth root value of
given number n
Step 52: return d;
//Returns the resulted value to d
```

The preceding algorithm can work in the following way:

The odd nth root let say cube root of a real number b can not be identified with the fractional power a^{1/n}, although so has been done in the entries nth root and cube root. The fractional power with a negative base is not uniquely determined therefore, it depends not only on the value of the exponent but also on the form of the exponent, e.g.

(–1)^{1/3} = the 3rd root of –1, i.e. = –1

(–1)^(2/6) = the 6th root of (–1)^2, i.e. = 1

Implementation to Find the nth Root of a Number

```
/*————————— START OF PROGRAM —————————*/

#include<stdio.h>
#include<conio.h>
#define NUMBER_OF_ITERATIONS 40
//Preprocessor directive where NUMBER_OF_ITERATIONS is
defined as macro
double calculate_root(double,double);
double Find_nth_Root(double,double,double);
double number(double,double);
void main()
{
     double x=1,n;
     double N;
     double root;
x_label:
printf("\n Enter root value [2,3, …5] ?");
     scanf("%f",&n);
     if(n<=0)
```

NOTES

```
    {
    printf("\nNumber should be Greater than 0");
printf("Press any key to enter again");
    getch();
    goto x_label;
    }
y_label:
printf("\n\rEnter a number = ");
scanf("%f",&N);
    if(N<=0)
    {
    printf("\nNumber should be greater than 0");
printf("\n PRESS ANY KEY TO ENTER AGAIN …");
    getch();
goto y_label;
    }
x = calulate_root(n,N);
printf("\n\nThe first assumed root is calculated as
%f\n",x);
    root=Find_nth_Root(N,n,x);
printf("\n\n%f Root of %f = ",n,N);
    printf("Root value is = %f",root);
    getch();
   }

double calculate_root(double n,double N)
   {
    int i, xr=1;
    double j=1;
    while(1)
    {
    for(i=0;i<n;i=i+1)
    {
    xr=xr*j;
    }
    if(xr>N)
    {
    return(j-1);
    break;
    }
```

NOTES

```
        j=j+1;
        xr=1;
        }
        }
double Find_nth_Root(double NUM,double n,double X0)
{
        int i;
        double d=1.0;
        double first_term,second_term,root=X0;
for(i=1;i<=NUMBER_OF_ITERATIONS;i++)
{
d=number(root,n);
first_term=((n-1)/n)*root;
second_term=(1/n)*(NUM/d);
root=first_term+second_term;
printf("\n%f\t%f\t%f",first_term,second_term,root);
}
        return(root);
}
double number(double x,double n)
{
    double d=1;
    int i;
    for(i=1;i<=n-1;i++)
    d=d*x;
    return(d);
}
```

The result of the preceding program is as follows:

```
Enter root value [2,3, …5] ? 3
Enter a number = 64
Root value is = 4
```

In this program, the syntax of `#define` is as follows:

```
#define macro-name replacement-string
```

The `#define` command is used to make substitutions throughout the program file in which it is located. It causes the compiler to go through the file, replacing every occurrence of macro-name with replacement-string. The replacement-string stops at the end of the line. The above program calculates the nth root of any number a. This program uses the NEWTON_RAPTION_ ITERATION method for calculation. For example, you have to calculate the square root of 16, then n=2 (square root), a=16 (the number). The following examples show how nth root of the given number can be written:

Enter a Number = 32, Enter a Root = 5. The (n^{th}) 5^{th} root of 32 is 2.

Enter a Number = 11, Enter a Root = 4. The 4^{th} root of 11 is 1.82116.

Example 5.10: Greatest Common Divisor (GCD).

NOTES

The GCD of two integers is the largest integer value that divides both integer values where both the values are not zero. The basic identities of GCD are as follows:

```
GCD(A,B)=GCD(B,A)
GCD(A,B)=GCD(-A,B)
GCD(A,0)=ABS(A)
```

Both the integer values can be assumed as nonnegative integers. The GCD procedure extracts the greatest common divisor A because the common divisor B divides to get the remainder until finally B divides A. The result A is in fact a greatest common divisor because it contains every other common divisor B.

GCD Algorithm

```
Step 1: integer m, n, q, r; //Variables are defined
Step 2: print'Enter two values:';
Step 3: read m,n;
//Input two values for m and n variables
Step 4: if m==0 OR n==0
//Checking the condition whether m is equal to 0 or n is
equal to 0
print'One number is Zero';
else
reach: //Label reach is defined for loop
q←m/n;
//Get the value of q after dividing m by n
r←m – q*n; //Gets remainder value
Step 5: if r==0
print 'GCD Value is :'; //Prints message
print n; //Prints GCD value
goto end; //Got to end label
else
m←n←r;
//Assigning m is equal to n that is also equal to r
goto reach; //Go to reach label
end:; //Label end is defined
```

If the two given values are 10, 12 then the greatest common factor is the number that divides both the values 10 and 12.

The GCD of two given integers (a and b) is the largest positive integer which divides both integers a and b, for example, gcd (10,12)=2. The

NOTES

following table shows the step-by-step procedure to get resultant GCD value:

Let the two values be, m =15 and n = 18.

div	quo	%Quo	%div	Resultant value
0				1
1	15	False	True	1
2	7	False	False	1
3	5	False	True	3
4	3			

The loop exits and returns 3. So, the resultant GCD value of the two given values 15 and 18 is 3.

Program to find GCD of given values:

```
/*————————— START OF PROGRAM —————————*/
#include <stdio.h>
#include <conio.h>
void main()
{
int m, n, q, r;
clrscr();
printf("Enter two values:");
scanf("%d%d", &m,&n);
if (m==0||n==0)
printf("One number is Zero");
else
reach:
{
q=m/n;
r=m - q*n;
}
if(r==0)
{
printf("GCD Value is : %d", n);
goto end; //Go to end label
}
else
{
m=n=r;
goto reach;
}
end:;
}
```

NOTES

The GCD can also be calculated applying Euclidean algorithm. If the integers a and b are two positive integers and n is the remainder, then (a, b) = (b, r).

```
Euclidean_gcd(a,b)
Step 1: integer x, y, f, d;
Step 2: x←f; y←d;
Step 3: if y=0 return x
Step 4: r←x mod y;
Step 5: x←y;
Step 6: y←r;
Step 7: goto Step 2.
```

The above algorithm works in the following way:

Small value (x) = 10, Large value (y) =12.

Large	Small	Remainder
12	10	2
10	2	0

Result: 2 is the GCD of 10, 12.

The above algorithm is known as Euclid's GCD algorithm that extracts the greatest common divisor x. The common divisor y divides x and keeps remainder as value n. This process is continued until y divides x finally. Therefore, value assigned for x is the greatest common divisor if it contains every other common divisor y.

Example 5.11: Base Conversion (Decimal to Binary).

The base of a binary number is 2 and of decimal number is 10 (denary). Binary numbers have only two numerals (0 and 1), whereas decimal numbers have 10 numerals (0, 1, 2, 3, 4, 5, 6, 7, 8, 9). An example of a binary number is 10011100 and decimal number is 0.012345679012. The decimal numeral system is the one that is the most widely used. Computer operations are performed with number base conversion.

The following algorithm is an example of printing an integer value into binary format:

Algorithm

```
Step 1: integer number, binary_val,temp_val, counter,d_val;
Step 2: binary_val←0; //Assigning value 0 to binary_val
Step 3: temp_val←number; // Assigning temp_val is equal
to number
Step 4: counter←0; //Assigning value 0 to counter
Step 5: print 'Enter the number';
Step 6: read number; //Accept input values to number
```

NOTES

```
Step 7: if temp_val>0
{
d_val ← mod(temp_val,2)
binary_val ← binary_val + d_val*10^counter;
//10^counter means power(10,counter)
d_val← d_val + a_val*p_val;
temp_val ← int(temp_val/2)
//Change the fraction values as integer data types.
counter ← counter + 1;
//Increase the counter value by one
}
Step 7: print 'Binary Value';
Step 8: print binary_val; // Prints resultant binary
value
```

How this algorithm works is explained in the following.

Let us take a decimal value 6.

```
d_val = 6 mod 2 that returns 0
binary value=0+3*10^0 returns 0
d_val = 3 mod 2 returns 1
counter = 0 + 1 = 1
```

The decimal number **6** is equal to binary number **110**. This conversion is explained in the following way:

number	number/2	number % 2
6	3	0
3	1	1
1	0	1

Implementation of Base Conversion (Decimal to Binary)

```
/*————————— START OF PROGRAM —————————*/
#include <stdio.h> //Declaring Header files
#include <conio.h>
#include <math.h>
void main() //Start main() function
{
int number, binary_val, temp_val, counter, d_val, p_val;
binary_val=0;
// Declaring integer data types variables
temp_val=number; //Assigning temp_val is equal to number
counter=0; //Initailizing 0 to counter
printf("\n Enter a number");
scanf("%d", &number); //Accept input value
```

NOTES

```
if (temp_val>0)
{
d_val = temp_val%2;
//Returns remainder to d_val
p_val=power(10,counter);
binary_val= binary_val+ d_val*p_val; //The value of binary_val is added to d_val by d_val by 10 'raise to the power' counter value
temp_val = int(temp_val/2)
//if temp_val contains fraction value, int() function changes the integer type value
counter = counter +1;
//Counter variable is increased
}
printf("Binary Value = %d", binary_val); //Printing the binary value
getch();
}
```

Base Conversion (Binary to Decimal)

Algorithm

Step 1: integer number,d_val,temp_val,counter,a_val;

Step 2: d_val←0; //Assigning 0 to d_val

Step 3: temp_val←number; //Assigning temp_val is equal to number

Step 4: counter←0; //Assigning 0 to counter

Step 5: print 'Enter the number';

Step 6: read number; //Accept input value for number

Step 7: if temp_val>0 //Body of if control statement

```
{
a_val ← mod(temp_val,10);
p_val←power(2,counter);
d_val← d_val+ a_val*p_val;
temp_val ← int(temp_val/10);
counter ← counter +1;
}
```

Step 8: print 'Decimal value';

Step 9: print d_val;

How the preceding algorithm works is explained in the following:

Let us take binary number 1011.

$=1*2^3+0+2^2+1*2^1+1*2^0$

$=8+0+2+1 = 11$

NOTES

The binary number **1011** is equal to decimal number **11**.

```
/*————————— START OF PROGRAM —————————*/
#include <stdio.h> //Declaring Header files
#include <conio.h>
#include <math.h>
void main() //Start main() function
{
int number,d_val,temp_val,counter,a_val;
// Declaring integer data types variables
temp_val=number;
//Assigning temp_val is equal to number
d_val=0;
counter=0; //Initailizing 0 to counter
printf("\n Enter a number");
scanf("%d", &number); //Accept input value
if (temp_val>0)
{
a_val = temp_val%10;
//Returns remainder to d_val
p_val = pow(2,counter);
//Returns counter value raise to the power 2 to p_val
variable
d_val=d_val+ a_val*p_val;
//The value of d_val is added to multiplied value of
a_val and p_val
temp_val = int(temp_val/10)
//if temp_val contains fraction value, int() function
changes the integer type value
counter = counter +1;
//Counter variable is increased
}
printf("Decimal Value = %d", d_val); //Printing the binary
value
getch();
//Pressing key to return the program
}
```

This program is able to convert the binary number into decimal number. The result of the program is as follows:

```
Enter a number = 1011
Decimal Value = 11
```

When a theoretical algorithm design is combined with the real-world data, it is called **algorithm engineering**. When you take an algorithm and combine it with a hardware device that is connected to the real-world, you can verify and validate the algorithm results and behaviour more precisely and accurately. A simple data acquisition or stimulus device may be considered as the real-world device. Alternatively, you can implement an algorithm on some embedded platform, such as a **field-programmable gate array (FPGA)** or microprocessor which can be similar to the final system design.

The first specific use of the term, 'algorithm engineering' was at the inaugural Workshop on Algorithm Engineering (WAE) in 1997.

It has of late been used for describing the steps in a graphical system design: 'A modern approach to design, prototype and deploy the embedded systems which combine open graphical programming with the **commercial off-the-shelf** (**COTS**) hardware for dramatically simplifying development, bringing higher-quality designs with a migration to custom design.'

With the help of algorithm engineering, you can transform a pencil-and-paper algorithm into a robust, efficient, well-tested and easily usable implementation. It covers various topics, from modelling cache behaviour to the principles of good software engineering. However, experimentation is its main focus.

Check Your Progress

1. What is the main purpose of problem statement?
2. Write a note on planning.
3. What is an algorithm?
4. List the important properties of an algorithm.

5.6 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. The main purpose of the problem statement is to identify and explain the problem.
2. Planning is the process of thinking about the activities required to achieve desired goal. It is the first and foremost activity to achieve desired results. It involves the creation and maintenance of a plan.
3. An algorithm can be defined as a well-defined procedure for computation that takes an input of some values and produces an output
4. The following are the five important properties (features) of algorithm:
   i. Finiteness
   ii. Definitiveness

iii. Input
iv. Output
v. Effectiveness

NOTES

5.7 SUMMARY

- Data processing is done when data is collected and translated into usable information.
- Scientific data processing usually involves a great deal of computation (arithmetic and comparison operations) upon a relatively small amount of input data, resulting in a small volume of output.
- Commercial data processing involves a large volume of input data, relatively few computational operations, and a large volume of output.
- The main purpose of the problem statement is to identify and explain the problem. This includes describing the existing environment, where the problem occurs, and what impacts it has on users, finances, and ancillary activities.
- Planning is the process of thinking about the activities required to achieve a desired goal. It is the first and foremost activity to achieve desired results. It involves the creation and maintenance of a plan.
- An algorithm is composed of a finite set of steps each of which may require one or more operations. Every operation may be characterized as either a simple or complex.
- An algorithm may have zero or more 'inputs' quantities. These inputs are given to the algorithm initially prior to its beginning or dynamically as it runs.
- Analysis of algorithms is the process to determine the computing time and storage required by an algorithm.
- An algorithm, which is not finite, is called infinite algorithm, e.g. a complete solution of a chessboard, division by zero.
- The first specific use of the term, 'algorithm engineering' was at the inaugural Workshop on Algorithm Engineering (WAE) in 1997.

5.8 KEY WORDS

- **Data processing system**: Data processing is done when data is collected and translated into usable information.
- **Conversion**: Converting data to another form or Language.
- **Aggregation**: Combining multiple pieces of data.

- **Data storage**: The final stage of data processing is storage. After all of the data is processed, it is then stored for future use.
- **Algorithm**: It can be defined as a well-defined procedure for computation that takes an input of some values and produces an output.

NOTES

5.9 SELF-ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. What are the various components of data processing?
2. What is problem definition? Why it is required?
3. What do you understand by planning?
4. State the properties of algorithm.
5. Write an algorithm for sequential search.

Long-Answer Questions

1. What are the different types of data processing systems? Discuss giving examples.
2. Explain the six stages of data processing giving examples.
3. Explain the criteria for algorithm design.
4. What are the different types of algorithm? Explain giving examples.
5. Write an algorithm for base conversion from binary to decimal.

5.10 FURTHER READINGS

Price, Wilson T. 1981. *Introduction to Computer Data Processing*, 3rd Edition. United States: Holt-Saunders International Editions.

Verzello, Robert J. and John Reutter. 1983. *Data Processing: Systems and Concepts*. New York: The McGraw-Hill Publishing Company, Inc.

Beech, Carol and Janice Burn. 1985. *Applications in Business Data Processing*. London (England): Pitman Publisher.

Fuori, William M. and Lawrence J. Aufiero. 1987. *Computers and Information Processing*. United States: Prentice Hall.

Sanders, Donald H. 1988. *Computers Today*. New York: McGraw-Hill.

Sinha, Pradeep K. and Priti Sinha. 2004. *Computer Fundamentals: Concepts, Systems & Applications*, 8th Edition. New Delhi: BPB Publications.

NOTES

Rajaraman, V. 1999. *Fundamentals of Computers*. New Delhi: Prentice-Hall of India.

Lipschutz, Martin M. and S. Lipschultz. 1982. *Theory and Problems of Data Processing, Schaum's Outline Series*. New Delhi: Tata McGraw Hill Publishing Co. Ltd.

Stallings, William. 1996. *Data and Computer Communications*. New Jersey (USA): Prentice-Hall PTR.

UNIT 6 FLOWCHARTS AND PROGRAMMING TECHNIQUES

Structure

6.0 INTRODUCTION

A flowchart is a type of diagram that represents a process or workflow. It is also defined as a diagrammatic representation of an algorithm. Decision table is a brief visual representation for specifying which actions to be performed based on given conditions. Top-down programming techniques and structured programming. Top-down approach is a programming style in which design starts with complex pieces and then differentiating or dividing them into smaller pieces successively. In other words, we can say that top down approach focuses on dividing a complex problem into smaller pieces to find a complete solution.

Structured programming is in fact, a sub-discipline of procedural programming. Small programs are better written in procedure-oriented programming. Bigger and complex programs are better written in structured manner. Structured programming is also known as modular programming. This enforces a logical structure for the program, making it more efficient, easy to understand and easier to modify.

In this unit, you will learn about the program flowchart, decision table, top-down programming techniques, and structure programming.

6.1 OBJECTIVES

After going through this unit, you will be able to:

- Explain the significance of flowchart
- Discuss the types of flowchart

NOTES

- Define decision table
- Understand the concept of top-down programming techniques
- Explain the procedural and structured programming
- Elaborate on the benefits of structured programming
- Discuss the features of procedural oriented programming

6.2 PROGRAM FLOWCHARTS

In the beginning, the use of flowcharts was restricted to electronic data processing for representing the conditional logic of computer programs. The1980s witnessed the emergence of structured programming and structured design. As a result of this, in database programming, data flow and structure charts began to replace flowcharts. With the widespread adoption of such ALGOL-like computer languages as Pascal, textual models like pseudocode are being used frequently for representing algorithms. Unified modelling language (UML) started the synthesis and codification these modelling techniques in the 1990s.

A flowchart refers to a graphical representation of a process which depicts inputs, outputs and units of activity. It represents the whole process at a high or detailed (depending on your use) level of observation. It serves as an instruction manual or a tool to facilitate a detailed analysis and optimization of workflow as well as service delivery.

Flowcharts have been in use since long. Nobody can be specified as the 'father of the flowchart'. It is possible to customize a flowchart according to need or purpose. This is why flowcharts are considered a very unique quality improvement method for representing data.

Symbols

A typical flowchart has the following types of symbols:

- **Start and end symbols**: They are represented as ovals or rounded rectangles, normally having the word 'Start' or 'End'.
- **Arrows:** They show the 'flow of control' in computer science. An arrow coming from one symbol and ending at another symbol shows the transmission of control to the symbol the arrow is pointing to.
- **Processing steps**: They are represented as rectangles.

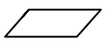
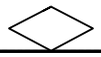
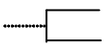
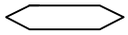
  Example: Add 1 to X.
- **Input/output symbol**: It is represented as a parallelogram.

  Examples: Get X from the user; display X.
- **Conditional symbol**: It is represented as a diamond (rhombus). It has a Yes/No question or True/False test. It contains two arrows coming out of it, normally from the bottom and right points. One of the arrows corresponds

to Yes or True, while the other corresponds to No or False. These two arrows make it unique.

There are also other symbols in flowcharts may contain, e.g. connectors. Connectors are normally represented as circles. They represent converging paths in the flowchart. Circles contain more than one arrow. However, only one arrow goes out. Some flowcharts may just have an arrow point to another arrow instead. Such flowcharts are useful in representing an iterative process, what is known as a loop in terms of computer science. A loop, for example, comprises a connector where control first enters processing steps, a conditional with one arrow exiting the loop, and another going back to the connector. These are listed in Table 6.1.

Table 6.1 *Symbols Used in Flowcharts*

Shape	Symbol	Symbol Name	Purpose
Oval		Terminator	To represent the begin/end or start/stop of a flow chart
Rectangle		Process	To represent calculations and data manipulations
Parallelogram		Data	To represent Input/Output data
Diamond		Decision	To represent a decision or comparison control flow
Double sided Rectangle		Predefined Process	To represent Modules or set of operations or a function
Bracket with broken line		Annotation	To represent descriptive comments or explanations
Document		Print out	To represent output data in the form a document
Multiple documents		Print outs	To represent output data in the form of multiple documents
Circle		Connector	To connect different parts of the flow chart
Hexagon		Repetition/ Looping	To represent a group of repetitive statements
Trapezoid		Manual Operation	To represent an operation which is done manually
Card		Card	To represent a card, e.g., punched card
Arrows		Flows of control	To represent the flow of the execution

NOTES

It is now used at the beginning of the next line or page with the same number. Thus, a reader of the chart is able to follow the path.

Instructions

The following is the step-by-step process for developing a flowchart:

Step 1: Information on how the process flow is gathered. For this, the following tools are used:

- Conservation
- Experience
- Product development codes

Step 2: The trial of process flow is undertaken.

Step 3: Other more familiar personnel are allowed to check for accuracy.

Step 4: If necessary, changes are made.

Step 5: The final actual flow is compared with the best possible flow.

Construction/Interpretation Tips for a Flowchart

- The boundaries of the process should be defined unambiguously.
- The simplest symbols should be used.
- It should be ensured that each feedback loop contains an escape.
- It should be ensured that there is only one output arrow out of a process box. Otherwise, it would require a decision diamond.

Types of Flowcharts

A flowchart is common type of chart representing an algorithm or a process and showing the steps as boxes of different kinds and their order by connecting these with arrows. We use flowcharts to analyse, design, document or manage a process or program in different fields.

There are many different types of flowcharts. On the one hand, there are different types for different users, such as analysts, designers, engineers, managers or programmers. On the other hand, those flowcharts can represent different types of objects. Sterneckert (2003) divides four more general types of flowcharts:

- Document flowcharts showing a document flow through system
- Data flowcharts showing data flows in a system
- System flowcharts showing controls at a physical or resource level
- Program flowchart showing the controls in a program within a system

However, there are several of these classifications. For example, Andrew Veronis named three basic types of flowcharts: the system flowchart, the general flowchart, and the detailed flowchart. Marilyn Bohl (1978) stated 'in practice, two kinds of flowcharts are used in solution planning: system flowcharts and

program flowcharts...'. More recently, Mark A. Fryman (2001) stated that there are more differences. Decision flowcharts, logic flowcharts, systems flowcharts, product flowcharts and process flowcharts are just a few of the different types of flowcharts that are used in business and government.

Interpretation

- Analyse flowchart of the actual process.
- Analyse flowchart of the best process.
- Compare both charts looking for areas where they are different. Most of the time, the stages where differences occur are considered to be the problem area or process.
- Take appropriate in-house steps to correct the differences between the two separate flows.

Example: Process flowchart—finding the best way home.

This is a simple case of processes and decisions in finding the best route home at the end of the working day.

A flowchart provides the following:

- **Communication:** Flowcharts are excellent means of communication. They quickly and clearly impart ideas and descriptions of algorithms to other programmers, students, computer operators and users.
- **An overview:** Flowcharts provide a clear overview of the entire problem and its algorithm for solution. They show all major elements and their relationships.
- **Algorithm development and experimentation:** Flowcharts are a quick method of illustrating program flow. It is much easier and faster to try an idea with a flowchart than to write a program and test it on a computer.
- **Check program logic:** Flowcharts show all major parts of a program. All details of program logic must be classified and specified. This is a valuable check for maintaining accuracy in logic flow.
- **Facilitate coding:** A programmer can code the programming instructions in a computer language with more ease with a comprehensive flowchart as a guide. A flowchart specifies all the steps to be coded and helps to prevent errors.
- **Program documentation:** A flowchart provides a permanent recording of program logic. It documents the steps followed in an algorithm.

Advantages of Flowcharts

- Clarify the program logic.
- Before coding begins, a flowchart assists the programmer in determining the type of logic control to be used in a program.

NOTES

- Serve as documentation.
- Serve as a guide for program coding of program writing.
- A flowchart is a pictorial representation that may be useful to the businessperson or user who wishes to examine some facts of the logic used in a program.
- Help to detect deficiencies in the problem statement.

Limitations of Flowcharts

- Program flowcharts are bulky for the programmer to write. As a result many programmers do not write the chart until after the program has been completed. This defeats one of its main purposes.
- It is sometimes difficult for a business person or user to understand the logic depicted in a flowchart.
- Flowcharts are no longer completely standardized tools. The newer structured programming techniques have changed the traditional format of a flowchart.

Differences between Flowcharts and Algorithms

Flowchart

- It is the graphical representation of the solution to a problem.
- It is connected with the shape of each box indicating the type of operation being performed. The actual operation, which is to be performed, is written inside the symbol. The arrow coming out of symbol indicates which operation to perform next.

Algorithm

- It is a process for solving a problem.
- It is constructed without boxes in a succession of steps.

An algorithm can be written in the following three ways:

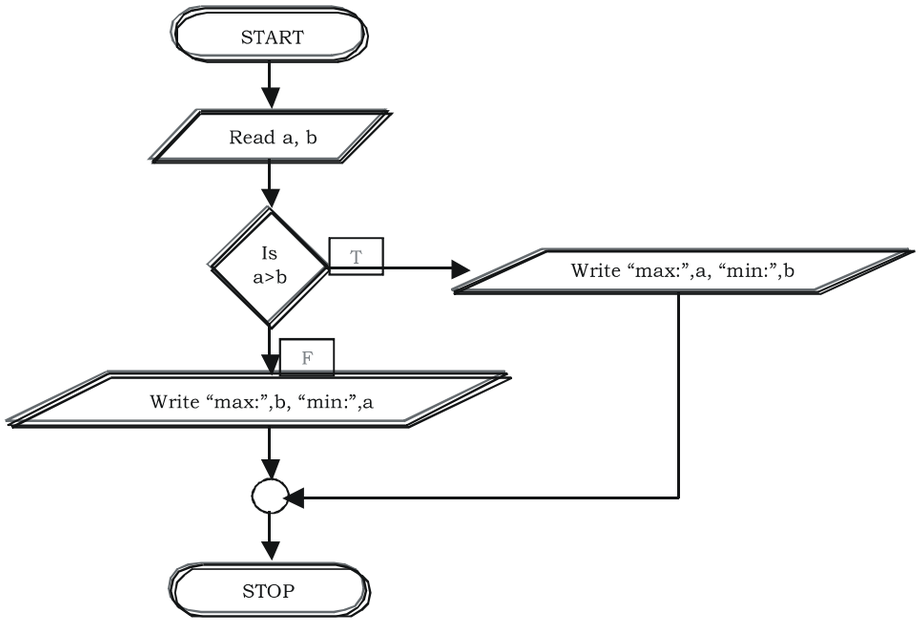
- **Straight sequential:** A series of steps that can be performed one after the other.
- **Selection or transfer of control:** Making a selection of a choice from two alternatives of a group of alternatives
- **Iteration or looping:** Performing repeated operations.

The following are the examples of algorithms and flowcharts for some different problems:

NOTES

Examples of Straight Sequential Execution

Example 6.1: Write a flowchart to find the maximum and minimum of given numbers.



Example 6.2: Write the various steps involved in executing a 'C' program and illustrate it with the help of a flowchart.

Executing a program written in C involves a series of steps. They are as follows:

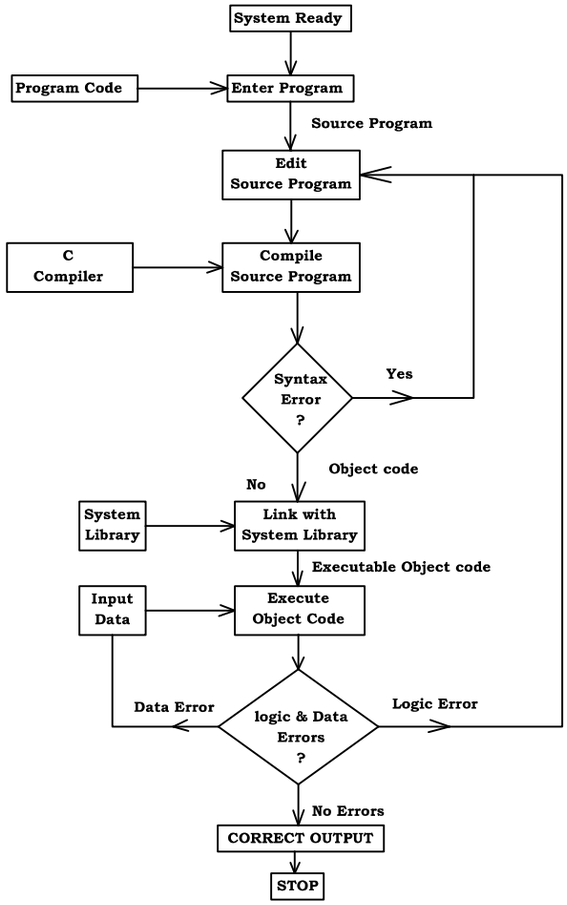
- Creating the program.
- Compiling the program.
- Linking the program with functions that are needed from the C library.
- Executing the program.

Although these steps remain the same irrespective of the operating system, system commands for implementing the steps and conventions for naming files may differ on different systems.

An operating system is a program that controls the entire operation of a computer system. All input/output operations are channelled through the operating system. The operating system, which is an interface between the hardware and the user, handles the execution of user programs.

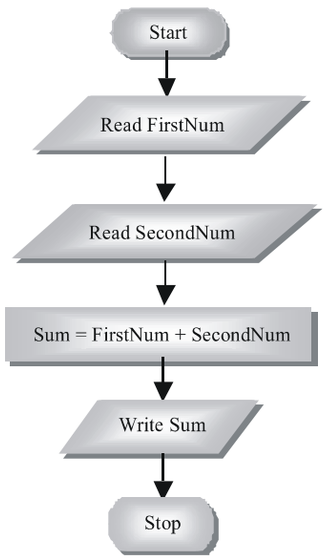
The two most popular operating systems today are UNIX (for minicomputers) and MS-DOS (for microcomputers).

NOTES



Examples for Flowcharts with Algorithms

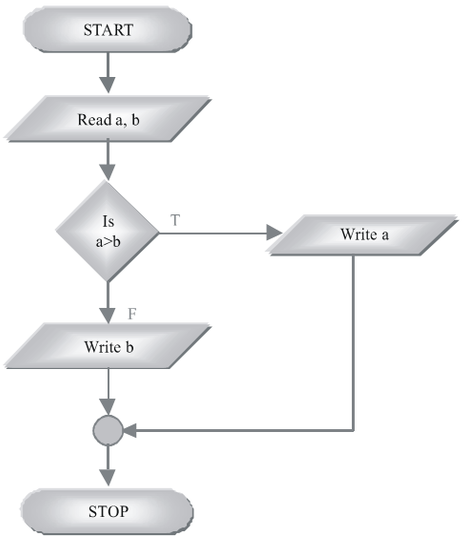
a. Draw a flowchart for adding two numbers and write an algorithm for it.



Algorithm for addition of two numbers:

```
Step 1: Start
Step 2: Read FirstNumber
Step 3: Read SecondNumber
Step 4: Sum= FirstNumber + SecondNumber
Step 5: Write (Sum)
Step 6: Exit
```

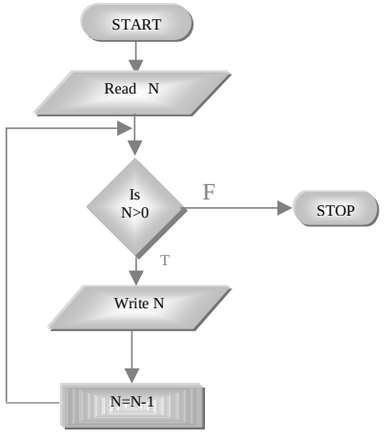
b. Draw a flowchart to find the larger number between two numbers and write an algorithm for it.



Algorithm for finding large number between two numbers

```
Step 1: Start
Step 2: Read a and b
Step 3: IF a > b THEN Write (a)
        ELSE Write(b)
Step 5: Stop
```

c. Draw a flowchart to display natural numbers between 1 and N in reverse order.



NOTES

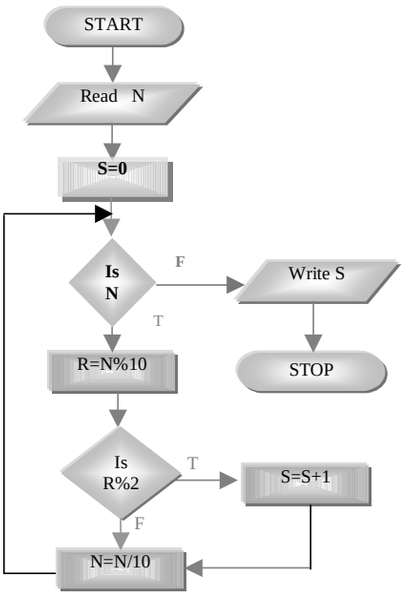
d. Draw a flowchart to display number of odd digits in a given number.

Algorithm for displaying Natural numbers between 1 and N in Reverse Order.

```
Step 1: Start
Step 2: Read  N
Step 3: Repeat while N>0
                Write (N)
                 N=N-1
Step 4:Exit
```

Algorithm to display number of odd digits exist in a given number.

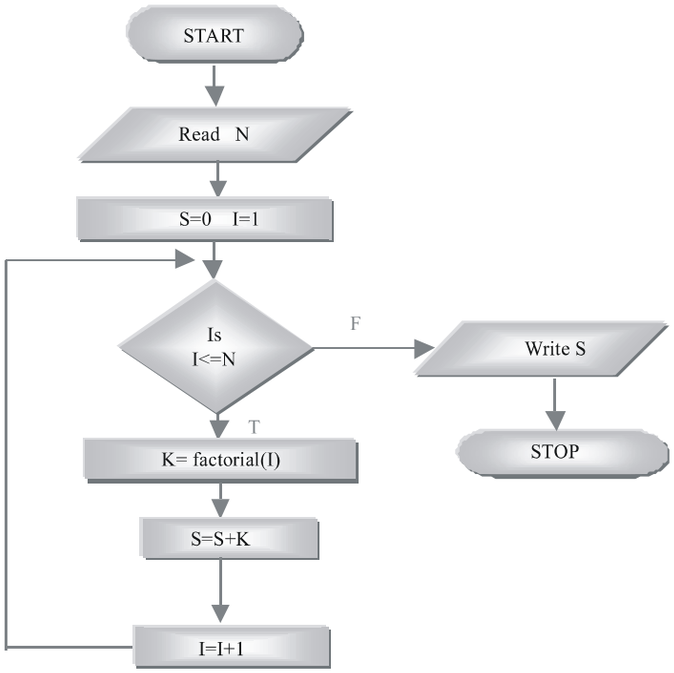
```
Step 1: Start
Step 2: Read  N
Step 3: S=0
Step 4: REPEAT while N>0
               R=N mod 10
               IF  R mod 2 THEN
                      S=S+1
               N =N/10
Step 5. Write(s)
Step 6: Exit
```



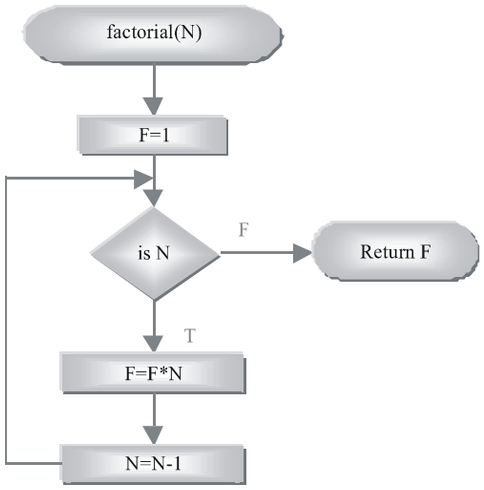
e. Draw a flowchart to evaluate the series `1! + 2!+ 3!+ ··· +N!`

Algorithm for evaluating the series 1!+2!+…..+N!

```
Step 1: Start
Step 2: Read  N
Step 3: S=0,I=1
Step 4: Repeat while I<=N
                K=factorial(I)
                S=S+K
                I=I+1
Step 5. Write(S)
Step 6: Exit
```



f. Flowchart to evaluate N!

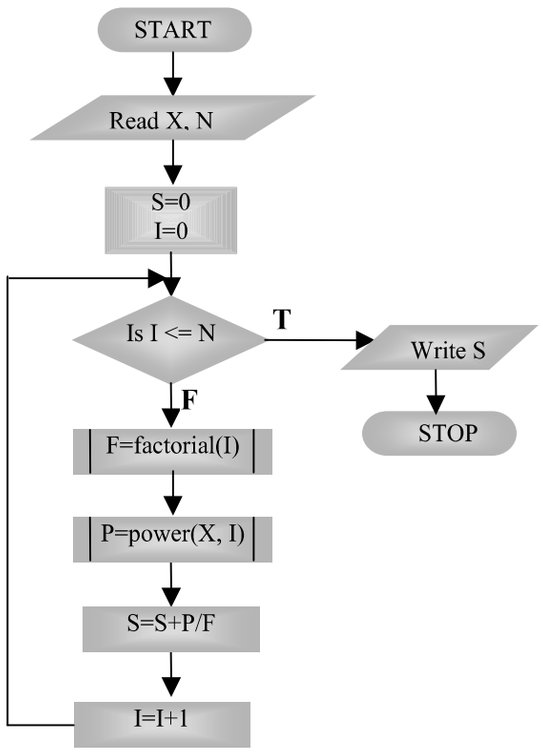


NOTES

Algorithm to find factorial(N). Where N is a value and function returns Factorial value for N

```
Step 1: F=1
Step 2: Repeat while N <> 0
                F=F*N
                N=N-1
Step 3: Return F
```

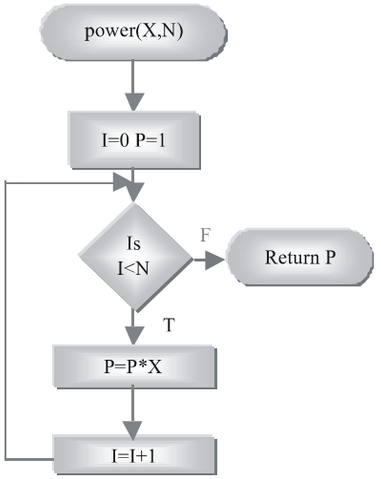
g. Draw a flowchart to evaluate the series $1+x+ x^2/2! +...+x^n /N!$



Algorithm to evaluate $1 + X + X^2/2! +...+ X^n/N!$

```
Step 1: Read X,N
Step 2: S=0,I=0
Step 3: Repeat while I<=N
                F=factorial(I)
                P=power(X,I)
                S=S+P/F
                I=I+1
Step 4: Writes(S)
Step 5: Exit
```

h. Flowchart to evaluate `Power(X, N)`



Algorithm to evaluate Power(X, N). Where X and N are values

```
Step 1: I=0,P=1
Step 2: Repeat while I<N
                P=P*X
                I=I+1
Step 3: Return P
```

6.3 DECISION TABLE

Decision table may be defined as a table that contains all the possible conditions for a specific problem and the corresponding results using condition rules that connect conditions with results. These tables are used to specify the complicated decision logic. A decision table is composed of rows and columns, which contain four separate quadrants, as shown in Figure 6.1.

Conditions	**Condition Alternatives**
Actions	**Action Entries**

Fig. 6.1 *Four Quadrants of a Decision Table*

- **Conditions** (upper left quadrant): Contains a list of all the possible conditions pertaining to a problem.
- **Condition Alternatives** (upper right quadrant): Contains the condition rules (set of possible values for each condition) of alternatives.
- **Actions** (lower left quadrant): Contains actions, which can be a procedure or operation to be performed.
- **Action Entries** (lower right quadrant): Contains the action rules, which specify the actions to be performed on the basis of the set of condition alternatives corresponding to that action entry.

Note that each column in the quadrant can be interpreted as a *processing rule*. The basic four-quadrant structure is common for all the decision tables; however, they may differ in the way of representation of the condition alternatives and action entries. For example, in some decision tables, condition alternative are represented using true/false values while in other tables, Yes/No, numbers (0 or 1), fuzzy logic or probabilistic representations may be used. Similarly, in some decision tables, action entries are represented by placing check marks, cross marks, or numbering the actions to be performed.

Developing a decision table

In order to build decision tables, first the maximum size of the table is determined. Any impossible situations, inconsistencies, or redundancies are then eliminated from the table. To develop a decision table, follow these steps:

1. Determine all actions that can be associated with a specific module or procedure.
2. Determine the number of conditions that may affect the decision.
3. Associate the specified conditions with specific actions. In the simplest form of decision table, there are two alternatives (Y or N) for each condition.
4. Define rules by indicating actions that occur for the specified conditions.

Let us consider an example to understand the concept of decision table by designing a program for a phone card company that sends monthly bills to its customers. If the customers pay their bills within two weeks, they are offered discount as per the scheme given below:

- If the amount > 350, then the discount is 5%.
- If the amount >= 200 and amount <= 350, then the discount is 4%.
- If the amount < 200, there is no discount.

The decision table for this example is shown in Figure 6.2.

Conditions	**Rules**			
	1	**2**	**3**	**4**
Paid within 2 weeks	Y	Y	Y	N
Order over Rs 350	Y	N	N	-
Rs 200 < = order < = Rs 350	N	Y	N	-
Order < 200	N	N	Y	-
Actions				
5 per cent discount	X			
4 per cent discount		X		
No discount			X	X

Fig. 6.2 *A Decision Table*

6.4 TOP-DOWN PROGRAMMING TECHNIQUES

NOTES

Top-down and bottom-up are both strategies of information processing and knowledge ordering, used in a variety of fields including software, humanistic and scientific theories, and management and organization.

A top-down approach, also known as stepwise design and stepwise refinement and in some cases used as a synonym of decomposition, is essentially the breaking down of a system to gain insight into its compositional sub-systems in a reverse engineering fashion. In a top-down approach an overview of the system is formulated, specifying, but not detailing, any first-level subsystems. Each subsystem is then refined in yet greater detail, sometimes in many additional subsystem levels, until the entire specification is reduced to base elements. A top-down model is often specified with the assistance of 'Black Boxes', which makes it easier to manipulate. However, black boxes may fail to clarify elementary mechanisms or be detailed enough to realistically validate the model.

Top-down approaches emphasize planning and a complete understanding of the system. It is inherent that no coding can begin until a sufficient level of detail has been reached in the design of at least some part of the system. Top-down approaches are implemented by attaching the stubs in place of the module. This, however, delays testing of the ultimate functional units of a system until significant design is complete.

Top-down design was promoted in the 1970s by IBM researchers Harlan Mills and Niklaus Wirth. The engineering and management success of this project led to the spread of the top-down approach through IBM and the rest of the computer manufacturing units. Among other achievements, Niklaus Wirth, the developer of Pascal programming language, wrote the influential paper Program Development by Stepwise Refinement. Since Niklaus Wirth went on to develop languages, such as Modula and Oberon (where one could define a module before knowing about the entire program specification), one can infer that top-down programming was not strictly what he promoted. Top-down methods were favoured in software engineering until the late 1980s, and object-oriented programming assisted in demonstrating the idea that both aspects of top-down and bottom-up programming could be utilized.

Modern software design approaches usually combine both top-down and bottom-up approaches.

Top-down is a programming style, the mainstay of traditional procedural languages, in which design begins by specifying complex pieces and then dividing them into successively smaller pieces. The technique for writing a program using top–down methods is to write a main procedure that names all the major functions it will need. Later, the programming team looks at the requirements of each of those functions and the process is repeated. These compartmentalized sub-routines eventually will perform actions so simple they can be easily and concisely coded.

When all the various sub-routines have been coded the program is ready for testing. By defining how the application comes together at a high level, lower level work can be self-contained. By defining how the lower level abstractions are expected to integrate into higher level ones, interfaces become clearly defined.

The 'Stepwise' refinement or top-down approach can be defined as,

- First the problem is broken or fragmented into parts.
- Then break the fragmented parts into parts again and now each of parts will be easy to do.

Advantages of Top-Down Approach

Following are the advantages of top-down approach:

- Breaking or fragmenting the problems into parts helps to identify what needs to be done.
- At each step of refinement, the new fragmented parts will become less complex and therefore easier to solve.
- Parts of the solution may turn out to be reusable.
- Breaking or fragmenting problems into parts allows more than one person to solve the problem.

6.5 STRUCTURE PROGRAMMING

A procedural program has a list of instructions telling computer what to do, step-by-step. It may instruct the computer for opening a file, then reading a number, multiplying it to another number and displaying the result. Programming units have main block with other blocks for programs, functions, precedures, subrouting etc. It also gives scope for files and may include modules and/or libraries.

Procedural programming is most convenient when program is small. It is a natural way of instructing a computer to perform a task. Language of computer processor, the machine code, is itself procedural. Thus, translating high-level procedural programs into machine readable code. A procedural program can split a lengthy list containing many instructions into shorter ones by using functions.

Procedural programming is also an imperative programming specifying steps to solve a problem. But such programming environment makes use of 'procedure call'. These procedures are in form of functions, methods, subroutines or routines. Steps are enumerated as these contain many computational steps in series. A procedure can be called from any where in the program body during execution of a program.

Concept of modularity as well of scope of program codes form the base of procedure programming. A main procedural program contains modules. These modules are also called package(s) or unit(s).

Structured programming is in fact, a subdiscipline of procedural programming. Small programs are better written in procedure-oriented programming. Bigger and complex programs are better written in structured manner. Structured programming is also known as modular programming. This enforces a logical structure for the program, making it more efficient, easy to understand and easier to modify.

Languages like ALGOL, ADA, PL/1, Pascal, and dBASE have features that enforce a logical structure for the program that is to be written. Structured programming employs top-down design model. In this design model, developers create a structure for the program structure by breaking it into separate sections and subsections. A modular approach is followed. Similar functions are coded as a separate module or submodule. Thus code can be easily loaded and since each module performs one specific task, these are reused in other programs too where similar task is performed.

These modules are tested individually, before integrating with other modules to create the overall program structure. These modules are arranged in a hierarchical manner using looping constructs, available in almost every programming language. These are, 'for', 'repeat' and 'while'.

Procedural programming discourages GOTO statement. But in structured programming 'GOTO' is used.

There are different methodologies or techniques to create structured programs. There are three most common approaches:

1. Structured programming of Edsger Dijkstra: In this methodology a program logic results in a structure made of few similar sub-structures. This makes it easy to understand.
2. Derived view of Dijkstra: This view advocates splitting of whole program into many sub-sections with each having one entry point. But there is difference of opinion and a single exit point is considered better.
3. Data Structured Programming or Jackson Structured Programming: This is based on alignment and synchronization of program structures with data structures. This approach applies fundamental structures of Dijkstra.

Low-level Structured Programming

Structured programs, at low level, contain simple and hierarchical program flow structures having sequence, selection, and repetition.

Sequence is an ordered execution of statements. Selection is one out of many statements whose execution depends on the condition and state of the program. Keywords used are, `if..then..else..endif`, `switch`, or `case`. In repetition a statement is executed repeatedly until the program reaches a certain condition or state of operations that are applied to elements of a set. Keywords for this are: `while`, `repeat`, `for` or `do..until`. These are looping statements.

It is recommended that every loop has one entry point and few approaches also advocate only one exit point. Few languages enforce this condition.

A language is 'block-structured' if it has a syntax that encloses structures between bracketed keywords. For example, an if-statement bracketed by `if..fi`. `This is done in` ALGOL 68. In PL/1 , a code section is bracketed by `BEGIN..END`. A language is 'comb-structured' if syntax tells to enclose structures inside an ordered series of keywords. Such a language has multiple structure keywords and each keyword define separate sections inside a block. In ADA, a block is a 4-pronged comb with keywords, `DECLARE`, `BEGIN`, `EXCEPTION`, `END`. The `if`-statement is a 4-pronged comb with keywords `IF`, `THEN`, `ELSE`, `END IF`.

Structured programming is mostly associated with a 'top-down' approach to design. This is based on work-breakdown structure.

Structured programming can be done in any programming language. It is always advisable to use a procedural programming language. After 1970 structured programming gained popularity as a technique of programming. Due to this shift in programming paradigm, new procedural programming languages have features that have encouraged structured programming.

Procedures and Modularity

Modularity is desirable in large and complicated programs. Inputs are usually specified syntactically in the form of *arguments* and the outputs delivered as *return values*.

Scoping is a technique that keeps procedures modular as it prevents access to the variables of one procedure by other. Procedures which are less modular are used in small programs. These programs interact with a large number of variables in the execution environment.

Procedures are convenient means to make pieces of code reusable. These codes may be written as programming libraries. It has the ability to create a simple interface which is self-contained and can be reused.

Basic Control Structure

In computer science **control flow** tells about an order of execution of program instructions in an imperative programming paradigm.

In such programming language execution of a **control flow statement** results in a choice between two or more paths. In non-strict functional languages, functions and language constructs achieve same result, but they control flow statements.

Control flow statements of different languages are different, and these are categorized by their effects:

- Unconditional branch or jump

- Execution of statements only on meeting some condition (i.e., conditional branching)
- Execution of statements zero or more times, on meeting some condition (i.e. loop - same as conditional branch)
- Execution of distant statements, returning flow of control (subroutines)
- Halting program, stopping further execution (unconditional halt).

Interrupts and signals are known as low-level mechanisms that altors flow of control, similar to a subroutine. This normally occurs in response to some external event which may occur asynchronously.

In low level programming, control flow instructions usually alters program counter. In some CPUs only control flow instructions result in conditional or unconditional branching which are also termed jumps.

Primitives

a. Labels

A label denotes an explicit name or number for a fixed location within a source code, referenced by control flow statements that appears somewhere in the source code. Label marks a location inside the source code and has no other effect.

Some languages use line numbers a named label instead of Fortran and BASIC, whole numbers use at the beginning of each line of text in source code. Such languages put constraint on line numbers to increase in value in and these need not be consecutive. For example, in BASIC:

```
10 LET X = 3
20 PRINT X
```

In languages such as C and Ada a label is used as an identifier, normally appearing at the start of a line followed by a colon. For example, in C:

Result: printf ("The operation was successful.\n");

b. Goto

The **goto** statement is most basic that transfers control unconditional by.

Keyword may be written in upper or lower case depending on implementation of the language normally it is written as:

```
goto label
```

When a goto statement is used next statement is executed from the indicated label.

Goto statements is discouraged by many computer scientists, notably Dijkstra.

c. Subroutines

The terminology for subroutines is not standard and it varies. Alternatively subroutines are known as procedures, routines, functions or methods.

NOTES

In the early stage of development in1950s, computer memories used to be very small hence subroutines were used for reducing program size. For this reason small codes were written and reused at many places in the program.

These days, subroutines make a program more structured, and provide modularity.

Minimal structured control flow

In May 1966, Böhm and Jacopini showed that program with **goto**s could be transformed into a goto-free form using choice IF THEN ELSE and loops such as WHILE condition DO xxx. At a later stage authors suggested the use of loops in place of choice use of Bollean variables with looping and choice statements using true/false flags could make program totally goto-free.

Such minimization is possible, but it does not mean it is desirable in all circumstances. But Böhm and Jacopini had shown that it was possible to make programs goto-free. Research has shown that control structures having one entry and one exit easier in comparison to any other form. They can be used anywhere in a program without disrupting the control flow.

Control structures in practice

Most programming languages use control structures with an initial keyword indicating type of control structure involved. Languages then divide as to whether or not control structures have a final keyword. Given below are language that have no final keyword or have to final keyword.

- No final keyword: Algol 60, C, C++, Haskell, Java, Pascal, Perl, PHP, PL/I, Python, PowerShell. Such languages require ways of grouping statements together:
  - Algol 60 and Pascal : begin ... end
  - C, C++, Java, Perl, PHP, and PowerShell: curly brackets { ... }
  - PL/1: DO ... END
  - Python: uses indentation level (see Off-side rule)
  - Haskell: either indentation level or curly brackets can be used, and they can be freely mixed
- Final keyword: Ada, Algol 68, Modula-2, Fortran 77, Visual Basic. The forms of the final keyword vary:
  - Ada: final keyword is end + *space* + initial keyword e.g. if ... end if, loop ... end loop
  - Algol 68: initial keyword spelled backwards e.g. if ... fi, case ... esac
  - Fortran 77: final keyword is end + initial keyword e.g. IF ... ENDIF, DO ... ENDDO
  - Modula-2: same final keyword end for everything

- Visual Basic: every control structure has its own keyword. If ... End If; For ... Next; Do ... Loop

NOTES

Loops

A loop contains a sequence of statements, which is specified once but used several times in succession. The code "inside" the *body* of the loop, shown below as *xxx* is executed a specified number of times, until some condition is met.

a. Count-controlled loops

Most programming languages use costructs to repeat a loop a specified number of times. If N is less than 1 in examples given below then the body is skipped completely, and if N = 1 the body is executed just once. In most cases counting goes downwards instead of upwards and step sizes other than 1 are used.

```
FOR I = 1 TO N      for I := 1 to N do begin
  xxx                 xxx
NEXT I              end;
DO I = 1,N          for ( I=1; I<=N; ++I ) {
  xxx                 xxx
END DO              }
```

Many programming languages, use integers for a reliable count-controlled loop. Use of floating point numbers is not reliable due to hardware constraints. A loop such as

```
for X := 0.1 step 0.1 to 1.0 do
```

might be repeated 9 or 10 times that depends on rounding errors and/or the hardware and/or the compiler version.

b. Condition-controlled loops

Most programming languages have constructs that repeats a loop until some condition changes. Some variations test conditions at the start of the loop, whereas others test at the end. In former case the body may be skipped completely, but in latter case body is executed at least once.

```
DO WHILE (test)          repeat
  xxx                    xxx
LOOP                     until test;
while (test) {             do
  xxx                    xxx
}                        while (test);
```

c. Collection-controlled loops

Several programming languages (e.g. Ada, Smalltalk, Perl, Java, C#, Visual Basic, Ruby, Python, JavaScript) special constructs allowing implicit looping through every element of an array, or every member of a set or collection.

```
someCollection do: [:eachElement |xxx].
```

NOTES

```
foreach someArray { xxx }
Collection<String> coll; for (String s : coll) {}
foreach (string s in myStringCollection) { xxx }
$someCollection | ForEach-Object { $_ }
```

d. General iteration

General iteration constructs in C is **for** statement and in Common Lisp it is **do** used for expressing any of the above loops, as well as others. Looping is also possible over a number of collections in parallel. Specific looping construct is usually preferred over general iteration construct, as it makes the purpose of the expression more clear.

e. Infinite loops

In certain situations, infinite loops are desired. It is desired that looping should continue and terminate only when some exception, such as error occurs. An event-driven program (such as a server) may be required to loop forever handling events, stopping only when the process is killed by the operator.

An infinite loop is mostly due to a programming error in a condition-controlled loop, wherein the loop condition is never changed within the loop.

f. Continuation with next iteration

Sometimes it is desired to skip the remainder of the loop body and continue with the next iteration. Some languages provide statements such as **continue**, **skip**, or **next** for this. This, prematurely terminates the innermost loop body, normal with next iteration. If the iteration is the last one in the loop, it terminates the entire loop.

g. Redo current iteration

Some languages, such as Perl and Ruby, provide a **redo** statement that restarts the current iteration from the beginning.

h. Early exit from loops

In a count-controlled loop that searches through a table, it is desired to stop when the desired item is found. Statement such as **break** or **exit** are provided by some programming languages.Using break or exit the current loop is immediately terminated and control is transferred to the statement immediately following that loop. Things can get a bit messy while searching a multi-dimensional table using nested loops.

The following example is done in Ada which supports both *early exit from loops* and *loops with test in the middle*. Both these features have similarity but codes are different. Codes for *early exit* needs combination with an **if** statement whereas a *condition in the middle* is a self contained construct.

```
with Ada.Text IO;
with Ada.Integer Text IO;
procedure Print_Squares is
```

```
  X : Integer;
begin
  Read_Data : loop
    Ada.Integer Text IO.Get(X);
  exit Read_Data when X = 0;
    Ada.Text IO.Put (X * X);
    Ada.Text IO.New_Line;
  end loop Read_Data;
end Print_Squares;
```

Python provides support for conditional execution of code and this depends on whether a loop was exited early using a break statement or not by using an else-clause with the loop. For example,

```
for n in set_of_numbers:
  if isprime(n):
    print "Set contains a prime number"
     break
else:
  print "Set did not contain any prime numbers"
```

The else clause in the example above is attached to the 'for' statement, and not to the inner if statement. The for and while loops of Python support an else clause and is executed only if early exit of the loop did not occur.

Structured non-local control flow

Many programming languages, favouring dynamic styles of programming, have constructs for **non-local control flow**. Using such constructs execution jumps out of a given context and then resume at some predeclared point. *Common non-local control constructs are, exceptions*, *conditions*, and *continuations*.

a. Conditions

PL/I contains 22 standard conditions (e.g. ZERODIVIDE SUBSCRIPTRANGE ENDFILE) which can be RAISEd and which can be intercepted by: ON *condition* action. Programmers can also define and use their own named conditions.

Like *unstructured if* only one statement can be specified, so in many cases, a GOTO is needed for deciding the location from whole flow of control should resume.

Some implementations need substantial overhead in both space and time (especially SUBSCRIPTRANGE), so these conditions are avoided by many programmers.

Common Syntax examples:

ON *condition* **GOTO** *label*

NOTES

b. Exceptions

Modern languages provide a structured construct for handling exception and do not rely on the use of GOTO:

```
    try {
xxx1                          // Somewhere in here
xxx2                          // use: ''throw'' someValue;
xxx3
} catch (someClass& someId) {  // catch value of someClass
                                 actionForSomeClass
} catch (someType& anotherId) {  // catch value of someType
                                  actionForSomeType
} catch (...) {        // catch anything not already caught
actionForAnythingElse
}
```

Catch can have many numbers and varieties of clauses for use. In D, Java, C#, and Python a finally clause is added to the try construct. It does not matter how the control comes out of the 'try' clause, the code inside the finally clause is guaranteed to execute. This is useful when writing code that must relinquish an expensive resource when finished processing:

```
    FileStream stm = null;  // C# example
    try {
      stm = new FileStream ("logfile.txt", FileMode.Create);
      return ProcessStuff(stm); // may throw an exception
    } finally {
      if (stm != null)
        stm. Close();
    }
    Since this pattern is fairly common, C# has a special syntax:
    using (FileStream stm = new FileStream ("logfile.txt",
FileMode.Create)) {
      return ProcessStuff(stm);   // may throw an exception
    }
```

On leaving the block, named using, the compiler guarantees the release of the object stm. With statement of Python and block argument to File.open of Ruby are used to get similar effect.

All these languages provide definition for standard exceptions and circumstances under which these are thrown. Users may throw exceptions of their own. C++ and Python permit users to throw and catch of almost any type.

In case there is no catch that matches a particular throw, then control moves back through subroutine calls and/or nested blocks to find a matching catch or

reaching the end of the main program and at this, point program is forcibly stopped giving suitable error message.

Proposed control structures

NOTES

In a spoof Datamation article in 1973, a suggestion was put by R. Lawrence Clark that the COMEFROM statement could replace GOTO and this provides some interesting examples. This found actual implementation in programming language INTERCAL, which is a language designed to make programs as obscure as possible.

Donald Knuth in his article 'Structured Programming with go to Statements', published in 1974, identified two situations that remained uncovered by the control structures listed above. He produced examples of control structures capable of handling these situations. These constructions inspite of their utility failed to find their way into mainstream programming languages.

Loop with test in the middle

Dahl proposed this in 1972.

```
loop                loop
  xxx1                read(char);
while test;           while not atEndOfFile;
  xxx2                write(char);
repeat;             repeat;
```

If *xxx1* is omitted we find a loop that tests at the top. If *xxx2* is omitted we find a loop that tests at the bottom. If **while** is omitted, an infinite loop is found. Hence this single construction is capable of replacing several constructions in most programming languages. A possible variant is one that allows more than one **while** test within the loop, but using **exitwhen** covers this case in a better way.

Multiple early exit/exit from nested loops

Zahn proposed this in 1974. Following codes show its modified version.

```
exitwhen EventA or EventB or EventC;
  xxx
exits
  EventA: actionA
  EventB: actionB
  EventC: actionC
endexit;
```

exitwhen is used for specifying events that is likely to occur within *xxx* and their occurrence is indicated by use of name of the event as a statement. On occurrence of some event, relevant action is carried out, and then control moves immediately after **endexit**. Such a construct provides very clear separation between determining situation that applies, and the action corresponding to that situation.

NOTES

exitwhen has conceptual similarity with exception handling, and exceptions or similar constructs find use in many languages for this purpose.

Following simple example searches a two-dimensional table for a particular item.

```
exitwhen found or missing;
  for I := 1 to N do
    for J := 1 to M do
      if table[I,J] = target then found;
  missing;
exits
  found: print ("item is in table");
  missing: print ("item is not in table");
endexit;
```

Structured Programming

Structured programming may be viewed as a subset or subdiscipline of procedural programming. It is a major programming paradigm. Its popularity is due to the fact that it reduces reliance on GOTO statement.

Historically, there are different methodologies or techniques adopted for structured programming. Three most common techniques are:

1. Structured programming of Edsger Dijkstra in which logic of a program has a structure made up of limited number of similar sub-structures. This reduces efforts to understand a program as each substructure is small and entire program is made up of many such similar structures.
2. A view that has been derived from Dijkstra that advocates splitting a program into many sub-sections with each having one entry point but there is another view that support the concept of one exit point instead of entry points.
3. Data Structured Programming or Jackson Structured Programming, which is based on alignment of program structure with data structures. This approach used fundamental structures proposed by Dijkstra, but used high-level structure to be modeled on the underlying data structures. There are at least three major approaches to data structured program design proposed by Jean-Dominique Warnier, Michael A. Jackson, and Ken Orr.

The latter two are more common. Years after the concept given by Dijkstra in 1969, Object-Oriented Programming (OOP) was developed that could handle complex programs.

Low-Level Structure: At low level, structured programs consisted simple program flow that followed hierarchical structures. These made use of sequence, selection, and repetition:

- 'Sequence' indicates an ordered execution of statements.

- 'Selection' means execution of one statement from a number of statements depending on the state of the program usually expressed using keywords if..then..else..endif, switch, or case.
- 'Repetition' means execution of statement when program reaches a certain state or operations that applies to every element of a collection. This is mostly expressed using keywords while, repeat, for or do..until. Often it is preferred that each loop should only have one entry point and also only one exit point. Few languages enforce this.

Dijkstra's original Guarded Command Language put emphasis on unity of these structures using a syntax that completely encloses the structure, as in if..fi. This is not in C, and risk of misunderstanding and incorrect modification is associated.

A language is known as 'block-structured' when there is a syntax to enclose structures between bracketed keywords. For example, an if-statement bracketed by if..fi as in ALGOL 68, or a code section bracketed by BEGIN..END, as in PL/I. However, a language is 'comb-structured' having a syntax for enclosing structures within an ordered series of keywords. A 'comb-structured' language has multiple structure keywords for defining separate sections within a block, analogous to multiple teeth or prongs in a comb separating sections of the comb. For example, in Ada, a block is a 4-pronged comb with keywords DECLARE, BEGIN, EXCEPTION, END, and the if-statement in Ada is a 4-pronged comb with keywords IF, THEN, ELSE, END IF.

Design: Structured programming often follows a 'top-down' approach but not always.

Structured Programming Languages

Structured programming is possible in any programming language, although a procedural programming language is preferred. After 1970 structured programming gained popularity and, most of the new procedural programming languages included features for encouraging structured programming. Better known structured programming languages are ALGOL, Pascal, PL/I and Ada.

Towards the end of 20th century most of the computer scientists favoured the concepts of structured programming. High-level programming languages such as FORTRAN, COBOL, and BASIC that originally lacked features of structured programming now have such features.

Common Deviations: Exception Handling

Multiple entry to a subprogram is never considered a good practice, yet multiple exists are mostly used in a subprogram. This is to ensure that there is exit without much work when there are circumstance that do not allow execution to continue further.

A simple procedure to read data from a file and to process it, is being given below:

```
open file;
while (reading not finished) {
 read some data;
 if (error) {
  stop the subprogram and inform rest of the program about
the error;
 }
}
process read data;
finish the subprogram;
```

The 'stop and inform' is achieved by throwing an exception, second return from the procedure, with a label loop break, or a goto. Since this procedure contains two exit points, rules of structured programming of Dijkstra is not obeyed. Single point exit rule, will make the coding cumbersome in this case. If possible error conditions, are many, having different rules for clean up, a procedure with single-exit-point will become very difficult to read and understand. This may be more difficult to read and understand than program when control is handled by goto statements. Structural programming without single-exit-point, in such cases, will give very clean and readable code.

For this reason, many programming languages adopted multiple points of exit in structural programming. C allowed multiple exits using 'continue', 'break', and 'return'. Other languages, after C have also adopted 'break' with a label and exceptions.

State Machines

Parsers and communications protocols, define many states following each other that it is not easily reduced to basic structures. These systems can be structured by making each state-change a separate subprogram using a variable for indicating active state. Some programmers implement state-changes using a jump to the new state.

Object-Oriented Comparison

Design of language during sixties was mostly small and based on examples of text book programs, but this changed when large programs were written. Most common statements in small programs, are assignment statements, whereas in large programs having more than 10 k lines, most common statements are procedure-calls to subprograms.

Small programs are handled by writing codes for hierarchy of structures. In large programs, organization contains network of structures. Undue emphasis on hierarchical structuring for data and procedures produces cumbersome code having large amounts of 'tramp data.' For example, a program for displaying text, allowing

dynamic change of font size of the entire screen becomes very combursome if font-size data is passed through a hierarchy. As an alternative to this, a system may be used for controlling font data using functions to retrieve data from a common area that is controlled by font-data sub-system.

FORTRAN uses labelled COMMON-blocks for separating global program data into subsystems for permitting program-wide, network-style access to data, such as font-size, by specifying particular COMMON-block name. Confusion is likely in FORTRAN by coding alias names and changing data-types while referencing the same labelled COMMON-block yet mapping alternate variables for over laying the same memory region. Labelled-COMMON concept was extremely valuable while organizing huge software systems. This has led to the use of object-oriented programming for defining subsystems of centralized data controlled by use of accessor functions.Making changes into other data-types were performed by explicitly converting, or casting, data from original variables.

Global subprogram names were found misleading in comparision to global variables or blank COMMON, and subsystems were kept under limit into subprogram names, such as naming using unique prefixes or using name, as used in Java package.

The object-oriented is flexible, as it separates a program into a network of subsystems own data, algorithms, or devices in the whole program, accessible only by specifying named access to the subsystem object-class.Object-oriented programming required a call-reference index for tracing subsystems or classes that are accessed from other locations.

Modern structured systems move toward 'event driven' architectures. Procedural events are designed as independent tasks.

Structured programming led to the concept of object-oriented programming. Memory leaks a program in causes consumption of huge amount of memory and this is due to failure is observing a single exit-point in a subprogram needing memory deallocation.

Structured programming led to a recognition of top-down approach in branching.

Various concepts of structured programming helps in understanding many facets of object-oriented programming.

Benefits of Structured Programming

Structural programming is also called modular programming which is based on work break down structure. It offers lot of benefits.

1. Easy to Write

A modular design is more productive. Programmer or developer looks at the overall picture first and then focuses on details, one-by-one. Every module performs one task and this enables several programmers to work on a single, large program

NOTES

in which each works on a particular module that is different from the other. Structured programs take less time in writing when compared to standard programs.

If a procedure has been written for one program, then it can be reused in other programs too, if it performs the same task. Such a procedure can be used in other programs too and is reusable. Such a program can be adapted to similar tasks by making some changes.

Structured programming follows top down design. In this design developer starts with the complete item first and then breaks it down into smaller parts and sub-parts. Thus, a difficult task is broken down into several small tasks. This is also known as 'divide and conquer' strategy. Once a big task is divided into several small tasks, solving these individual pieces makes it easy to control these small programs independently until every step can easily be implemented. This is done with successive refinement.

2. Easy to Debug

Each procedure, in a program, performs just one task. Thus individual procedure can be checked. An unstructured program has a sequence of instructions, not grouped task wise. Such programs are cluttered with details and creates problem in understandability. A structured program, being modularized, it is easy to locate the fault and change or modify it, without the need of looking into the other parts of the program.

3. Easy to Understand

A modular design shows the relationship between the procedures. Procedure names are mostly, meaningful and a clear documentation identifies the task of each module. It is always advisable to give meaningful variable names so as to identify the purpose of each variable.

4. Easy to Change

In a large program which is procedural makes it difficult to understand when other programmers or developers are working on that. Since a structured program is self-documenting, it poses no difficulty for other programmer to understand it.

According to E. Dijkstra, hierarchical systems have a property that something considered as an undivided entity on one level is considered as a composite object on the next lowest level of greater detail. Thus natural grain of space or time, applicable at each level decreases by an order of magnitude when we shift our attention to the next lower one. A wall can be understood as composed of bricks. Bricks can be thought of as composed of crystals and crystals made up of molecules, etc.

Below we give an example of a small house that is to be made for a pet.

Pet House

We first see the overall image. A house is composed of floor, walls and roof. We move down and form details of each. A floor may be of something on which a person can stand. Walls will be front, back, left and write. Every wall will have top section, lower section and door section. Roof also will be left side and right side. Details are given below.

- **1. Floor or Base**
- **2. Walls**
  - o 2.1 Back Wall
    - – 2.1.1 Top section
    - – 2.1.2 Lower section
  - o 2.2 Left Wall
  - o 2.3 Right Wall
  - o 2.4 Front Wall
    - – 2.4.1 Top section
    - – 2.4.2 Door section
- **3. Roof**
  - o 3.1 Left Side
  - o 3.2 Right Side

Once each part has been detailed one may prepare all sections. These are described below. Join these sections using nails and glue as needed.

- **1. Floor or Base -** cut to size as specified and smooth edges
- **2. Walls**
  - o 2.1 Back Wall
    - – 2.1.1 Top section - cut triangular shape
    - – 2.1.2 Lower section - cut square shape and glue and staple to upper part of wall
  - o 2.2 Left Wall - cut to specified dimensions
  - o 2.3 Right Wall - cut to specified dimensions
  - o 2.4 Front Wall
    - – 2.4.1 Top section - cut triangular piece to specified dimensions
    - – 2.4.2 Door section - use jigsaw to cut U shape for the door in square shape, staple and glue to top section
- **3. Roof**
  - o 3.1 Left Side - cut to specified dimensions
  - o 3.2 Right Side - cut to specified dimensions

A procedural program is a list of actions as input to computer. When program becomes large and complex, it becomes difficult to organize and control these. Also at a later stage when another programmer has to work on that, it becomes difficult to understand it. A structural programming that way is good since it is composed of small routines or tasks organized in structured manner that is called modular too.

After the concept of structured programming only, the concept of object oriented programming came. In procedural programming data is not controlled, but in object oriented programming data belongs to method and to access data, method has to be accessed. Structured program has great importance in managing large and complex programs.

Procedure Oriented Programming

Procedural programming, at times is used as a synonym for imperative programming. Procedures are also known as routines, subroutines, methods, or functions. These contain a series of steps for computation to be carried out. A procedure is called from any point during execution of a program. In a **procedural programming language** each step is defined precisely for performing a task. Benefits of a procedure programming are:

- Re-usability to code with copying.
- Easy tracking of program flow without using GO TO or 'JUMP' statements.
- Creation of structured or modular programs.

Features of Procedure Oriented Programming

1. A procedure oriented program consists of instructions in groups, known as functions. High level languages like Fortran, Pascal and 'C' are commonly known as procedure oriented programming languages.
2. Programs are organised in the form of subroutines and the data items are freely accessible.
3. Data in procedure oriented language is open and can be accessed by any function.
4. Function overloading and operator overloading are not possible in procedure oriented language.
5. In procedure oriented languages local variable can be declared only at the beginning of the block.
6. Program controls are through jumps and calls to sub-routines.
7. Polymorphism, Encapsulation, and inheritance are not possible in procedure oriented languages.
8. For solving the problems, the problem is divided into a number of modules. Each module in procedure oriented language is a sub-program.
9. Data abstraction property is not supported by procedure oriented languages.

Check Your Progress

1. What is flowchart?
2. What is decision table?
3. Write a note on top-down approach.
4. List the benefits of structured programming.

6.6 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. A flowchart refers to a graphical representation of a process which depicts inputs, outputs and units of activity. It represents the whole process at a high or detailed (depending on your use) level of observation
2. Decision table may be defined as a table that contains all the possible conditions for a specific problem and the corresponding results using condition rules that connect conditions with results.
3. A top-down approach, also known as stepwise design and stepwise refinement and in some cases used as a synonym of decomposition, is essentially the breaking down of a system to gain insight into its compositional sub-systems in a reverse engineering fashion.
4. The benefits of structured programming are:
   (i) Easy to write
   (ii) Easy to debug
   (iii) Easy to understand
   (iv) Easy to change

6.7 SUMMARY

- The use of flowcharts was restricted to electronic data processing for representing the conditional logic of computer programs. The1980s witnessed the emergence of structured programming and structured design.
- A flowchart refers to a graphical representation of a process which depicts inputs, outputs and units of activity. It represents the whole process at a high or detailed (depending on your use) level of observation.
- Flowcharts are excellent means of communication. They quickly and clearly impart ideas and descriptions of algorithms to other programmers, students, computer operators and users.

NOTES

- Program flowcharts are bulky for the programmer to write. As a result many programmers do not write the chart until after the program has been completed.
- It is sometimes difficult for a business person or user to understand the logic depicted in a flowchart.
- Flowcharts are no longer completely standardized tools. The newer structured programming techniques have changed the traditional format of a flowchart.
- It is connected with the shape of each box indicating the type of operation being performed. The actual operation, which is to be performed, is written inside the symbol.
- Decision table may be defined as a table that contains all the possible conditions for a specific problem and the corresponding results using condition rules that connect conditions with results.
- A top-down approach, also known as stepwise design and stepwise refinement and in some cases used as a synonym of decomposition, is essentially the breaking down of a system to gain insight into its compositional sub-systems in a reverse engineering fashion.
- Modern software design approaches usually combine both top-down and bottom-up approaches.
- Procedural programming is most convenient when program is small. It is a natural way of instructing a computer to perform a task. Language of computer processor, the machine code, is itself procedural.
- Modularity is desirable in large and complicated programs. Inputs are usually specified syntactically in the form of arguments and the outputs delivered as return values.
- Procedures are convenient means to make pieces of code reusable. These codes may be written as programming libraries. It has the ability to create a simple interface which is self-contained and can be reused.
- A label denotes an explicit name or number for a fixed location within a source code, referenced by control flow statements that appears somewhere in the source code.
- A loop contains a sequence of statements, which is specified once but used several times in succession. The code "inside" the body of the loop, shown below as xxx is executed a specified number of times, until some condition is met.
- Procedural programming, at times is used as a synonym for imperative programming. Procedures are also known as routines, subroutines, methods, or functions.

- In a procedural programming language each step is defined precisely for performing a task.

6.8 KEY WORDS

- **Debugging**: It refers to the process of carrying out programs on sample data sets for determining if there are any faulty results.
- **Profiling**: It refers to the process of the execution of a correct program on data sets and the measurement of the time and space it takes in computing the results.
- **Decision table**: It is a table that contains all possible conditions for a specific problem and the corresponding results using condition rules that connect conditions with results.
- **Scoping**: It is a technique that keeps procedures modular as it prevents access to the variables of one procedure by other.
- **Loop**: It contains a sequence of statements, which is specified once but used several times in succession.

6.9 SELF-ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. Define the different types of symbols used in a flowchart.
2. What are the steps for developing a flowchart?
3. Explain the benefits of top-down approach.
4. Write a short note on procedural programming.
5. Explain the methodologies for creating structured programs.
6. Define the term modularity.

Long-Answer Questions

1. What are the advantages and disadvantages of flowchart?
2. Differentiate between the flowchart and the algorithm.
3. Elaborate on decision table.
4. What do you understand by top-down programming techniques?
5. What are the different types of loops? Explain.
6. What is structured programming? Explain giving examples.

7. Explain the benefits of structured programming.
8. Discuss the features of procedural oriented programming.

NOTES

6.10 FURTHER READINGS

Price, Wilson T. 1981. *Introduction to Computer Data Processing*, 3rd Edition. United States: Holt-Saunders International Editions.

Verzello, Robert J. and John Reutter. 1983. *Data Processing: Systems and Concepts*. New York: The McGraw-Hill Publishing Company, Inc.

Beech, Carol and Janice Burn. 1985. *Applications in Business Data Processing*. London (England): Pitman Publisher.

Fuori, William M. and Lawrence J. Aufiero. 1987. *Computers and Information Processing*. United States: Prentice Hall.

Sanders, Donald H. 1988. *Computers Today*. New York: McGraw-Hill.

Sinha, Pradeep K. and Priti Sinha. 2004. *Computer Fundamentals: Concepts, Systems & Applications*, 8th Edition. New Delhi: BPB Publications.

Rajaraman, V. 1999. *Fundamentals of Computers*. New Delhi: Prentice-Hall of India.

Lipschutz, Martin M. and S. Lipschultz. 1982. *Theory and Problems of Data Processing, Schaum's Outline Series*. New Delhi: Tata McGraw Hill Publishing Co. Ltd.

Stallings, William. 1996. *Data and Computer Communications*. New Jersey (USA): Prentice-Hall PTR.

UNIT 7 BATCH PROCESSING: TECHNIQUES AND APPLICATIONS

Structure

7.0 INTRODUCTION

Batch processing is a method of running high-volume, repetitive data jobs. The batch method allows users to process data when computing resources are available, and with little or no user interaction. Batch processing is the processing of transactions in a batch or group. Beverage processing, dairy processing biotech products manufacturing, dairy processing, food processing, pharmaceutical formulations and soap manufacturing are some of the examples of batch processing.

A file is realized as a stream of bytes or it can be seen as collection of records. A master file contains descriptive data. A transaction file contains a set of transaction records. It keeps a record of day-to- day transactions occurring during business hours of an enterprise or a business firm. It is then used to update master files.

File updating refers to changing or modifying values in one or more records of a file, especially a data file, without changing the organization or semantics of the file.

In this unit, you will learn about the batch processing, typical batch processing application, and file update.

NOTES

7.1 OBJECTIVES

After going through this unit, you will be able to:

- Define the term batch processing
- Discuss the usage of batch processing
- Understand the parallel batch processing
- Define master and transaction file in file processing
- Explain how to update the file

7.2 BATCH PROCESSING

The term 'Batch Processing' originates in the traditional classification of methods of production as job production (one-off production), batch production (production of a 'Batch' of multiple items at once, one stage at a time), and flow production (mass production, all stages in process at once). Computerized batch processing is the running of "Jobs that can run without end user interaction, or can be scheduled to run as resources permit".

Batch processing is a method of running high-volume, repetitive data jobs. The batch method allows users to process data when computing resources are available, and with little or no user interaction.

With batch processing, users collect and store data, and then process the data during an event known as a 'Batch Window'. Batch processing improves efficiency by setting processing priorities and completing data jobs at a time that makes the most sense.

The batch processing method was first used in the 19th century by Herman Hollerith, an American inventor who created the first tabulating machine. This device became the precursor to the modern computer, capable of counting and sorting data organized in the form of punched cards. The cards and the information they contained could then be collected and processed together in batches. This innovation allowed large amounts of data to be processed more quickly and accurately than by manual entry methods.

Batch processing plays a critical role in helping companies and organizations manage large amounts of data efficiently. It is especially suited for handling frequent, repetitive tasks, such as accounting processes. In every industry and for every job, the basics of batch processing remain the same. The essential parameters include the following:

- Who is submitting the job.
- Which program will run.

- The location of the input and outputs.
- When the job should be run.

In other words, the who, what, where, and why.

Common Batch Processing Usage

Following are the common batch processing usages:

- Efficient bulk database updates and automated transaction processing, as contrasted to interactive OnLine Transaction Processing (OLTP) applications. The Extract, Transform, Load (ETL) step in populating data warehouses is inherently a batch process in most implementations.
- Performing bulk operations on digital images, such as resizing, conversion, watermarking, or otherwise editing a group of image files.
- Converting computer files from one format to another. For example, a batch job may convert proprietary and legacy files to common standard formats for end-user queries and display.

Batch processing jobs are run on regularly scheduled times (e.g., overnight) or on an as-needed basis. As an example, bills for utilities and other services received by consumers are typically generated by batch processing each month. Batch processing is beneficial because it is a cost-effective means of handling large amounts of data at once. One caveat is that the inputs for the processing must be correct or else the results of the whole batch will be faulty, which would cost time and money.

7.3 TYPICAL BATCH PROCESSING APPLICATION

Computerized batch processing is the running of jobs that can run without end user interaction, or can be scheduled to run as resources permit. The term **'Batch Processing'** originates in the traditional classification of methods of production as job production (one-off production), batch production (production of a **'Batch'** of multiple items at once, one stage at a time), and flow production (mass production, all stages in process at once).

Batch applications are still critical in most organizations in large part because many common business processes are amenable to batch processing. While online systems can also function when manual intervention is not desired, they are not typically optimized to perform high-volume, repetitive tasks. Therefore, even new systems usually contain one or more batch applications for updating information at the end of the day, generating reports, printing documents, and other non-interactive tasks that must complete reliably within certain business deadlines.

Some enterprise applications contain tasks that can be executed without user interaction. These tasks are executed periodically or when resource usage is low, and they often process large amounts of information, such as log files, database

NOTES

records, or images. Examples include billing, report generation, data format conversion, and image processing. These tasks are called **batch jobs**.

Batch processing refers to running batch jobs on a computer system. Java EE includes a batch processing framework that provides the batch execution infrastructure common to all batch applications, enabling developers to concentrate on the business logic of their batch applications. The batch framework consists of a job specification language based on XML, a set of batch annotations and interfaces for application classes that implement the business logic, a batch container that manages the execution of batch jobs, and supporting classes and interfaces to interact with the batch container.

A batch job can be completed without user intervention. For example, consider a telephone billing application that reads phone call records from the enterprise information systems and generates a monthly bill for each account. Since this application does not require any user interaction, it can run as a batch job.

Some applications are amenable to flow processing, namely those that only want data from a single input at once (not totals, for instance): start the next step for each input as it completes the previous step. In this case flow processing lowers latency for individual inputs, allowing them to be completed without waiting for the entire batch to finish. However, many applications require data from all records, notably computations, such as totals. In this case the entire batch must be completed before one has a usable result: partial results are not usable.

Modern batch applications make use of modern batch frameworks, such as JemThe Bee, Spring Batch or implementations of JSR 352 written for Java, and other frameworks for other programming languages, to provide the fault tolerance and scalability required for high-volume processing. In order to ensure high-speed processing, batch applications are often integrated with grid computing solutions to partition a batch job over a large number of processors, although there are significant programming challenges in doing so. High volume batch processing places particularly heavy demands on system and application architectures as well. Architectures that feature strong input/output performance and vertical scalability, including modern mainframe computers, tend to provide better batch performance than alternatives.

Scripting languages became popular as they evolved along with batch processing.

Batch Window

A batch window is referred as a period of less-intensive online activity, when the computer system is able to run batch jobs without interference from, or with, interactive online systems.

A bank's End-Of-Day (EOD) jobs require the concept of cutover, where transaction and data are cut off for a particular day's batch activity, such as 'Deposits after 3 PM will be processed the next day'.

As requirements for online systems uptime expanded to support globalization, the Internet, and other business needs, the batch window shrank and increasing emphasis was placed on techniques that would require online data to be available for a maximum amount of time.

Batch Size

The batch size refers to the number of work units to be processed within one batch operation. Some examples include,

- The number of lines from a file to load into a database before committing the transaction.
- The number of messages to de-queue from a queue.
- The number of requests to send within one payload.

Common Batch Processing Usage

Following are the usages of common batch processing:

- Efficient bulk database updates and automated transaction processing, as contrasted to interactive OnLine Transaction Processing (OLTP) applications. The Extract, Transform, Load (ETL) step in populating data warehouses is inherently a batch process in most implementations.
- Performing bulk operations on digital images, such as resizing, conversion, watermarking, or otherwise editing a group of image files.
- Converting computer files from one format to another. For example, a batch job may convert proprietary and legacy files to common standard formats for end-user queries and display.

Parallel Batch Processing

Batch jobs often process large amounts of data or perform computationally expensive operations. Batch applications can be benefited from parallel processing in the two circumstances, such as,

- Steps that do not depend on each other can run on different threads.
- Chunk-oriented steps where the processing of each item does not depend on the results of processing previous items can run on more than one thread.

Batch frameworks provide mechanisms for developers to define group

7.4 MASTER FILE AND TRANSACTION FILE

A file is realized as a stream of bytes or it can be seen as collection of records. A record is a collection of related fields. A field is the smallest meaningful unit of information. Files store data/information required for data processing or information storage. Fields are also called attributes. Usually two types of operations are

NOTES

performed on files: retrieval and update. The update operation modifies the contents of the files, whereas the retrieval operation just gives a snapshot of the database. In the information processing environments, the creation and maintenance of different files is an important task. Two types of files are generally created and maintained: master file and transaction file.

7.4.1 Master File

A master file contains descriptive data. For example, the account master file for a bank will hold data, such as account number, name of the account holder, account type, balance amount, address, telephone, etc. A master file holds relatively permanent and summarized data. All transactions that take place in an enterprise or a business firm for over a period of time are stored in master files. Periodically, these master files are updated using transaction files. Summary reports are prepared using these master files.

7.4.2 Transaction File

A transaction file contains a set of transaction records. It keeps a record of day-to-day transactions occurring during business hours of an enterprise or a business firm. It is then used to update master files. For example, banks keep records of day-to-day transaction that occur in a day. At the end of the day, transaction files are used to update master files.

7.5 FILE UPDATE

File updating refers to changing or modifying values in one or more records of a file, especially a data file, without changing the organization or semantics of the file. File updating may be done in one of two techniques. The first technique is commonly used in data processing when the updating process is carried out independently from the entry of modifications or improvements and invisibly from any human operator. The second technique is when records are displayed on an interactive device, and an operator can then amend or improve or modify a record while able to see it, this method is sometimes also called file editing.

Data is not static as it is constantly changing and these changes need to be reflected in their files. The function that keeps files current is known as updating.

Update Files

Three specific files are associated while updating a file.

- The permanent data file, called the master file contains the most current file data.
- The transaction file contains changes to be applied to the master file.

NOTES

- The third file required in an update program is an error report file. The error report contains a listing of all errors discovered during the update process and is presented to the user for corrective action.

Three basic types of changes occur in all file updates, adding new data, deleting old data, and modify data containing revisions.

To process any of these transactions, we need a key. A key is one or more fields that uniquely identify the data in the file. For example, in a student file, the key would be student ID. In an employee file, the key would be Social Security number.

File updates are of following two types:

- In a **batch update**, changes are collected over time and then all changes are applied to the file at once.
- In an **online update**, the user is directly connected to the computer and the changes are processed one at a time-often as the change occurs.

Sequential File Update

Assume a batch in sequential file environment. It is a file that must be processed serially starting at the beginning without any random processing abilities. A sequential file update actually has two copies of the master file, the old master file and the new master file.

Update Errors

Two general classes of errors can occur in an update program. The first being bad data implying that attributes which is not a part of the data. The second class of errors is file errors. File errors occur when the data on the transaction file are not in synchronization with the data on the master file.

Following three different situations can occur:

- **Situation 1:** An add transaction matches a record with the same key on the master file. Master files do not allow duplicate data to be present. When the key on an add transaction matches a key on the master file, therefore, the transaction is rejected as invalid and it is reported on the error report.
- **Situation 2:** A revise transaction's key does not match a record on the master file. In this case, user is trying to change data that do not exist. This is also a file error and must be reported on the error report.
- **Situation 3:** A delete transaction's key does not match a record on the master file. In this case, user is trying to delete data that do not exist, and this must also be reported as an error.

NOTES

Check Your Progress

1. What is batch processing?
2. What are the two types of files that are generally created and maintained in file processing?
3. What do you understand by the term file update?
4. What are the two types of file update?

7.6 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. Batch processing refers to running batch jobs on a computer system.
2. Master file and transaction file are the two types of files which are created and maintained.
3. File updating refers to changing or modifying values in one or more records of a file, especially a data file, without changing the organization or semantics of the file.
4. The two types of file update are batch and online update.

7.7 SUMMARY

- The term 'Batch Processing' originates in the traditional classification of methods of production as job production (one-off production), batch production (production of a 'Batch' of multiple items at once, one stage at a time), and flow production (mass production, all stages in process at once).
- Batch processing is a method of running high-volume, repetitive data jobs. The batch method allows users to process data when computing resources are available, and with little or no user interaction.
- The batch processing method was first used in the 19th century by Herman Hollerith, an American inventor who created the first tabulating machine.
- Batch processing plays a critical role in helping companies and organizations manage large amounts of data efficiently.
- Batch applications are still critical in most organizations in large part because many common business processes are amenable to batch processing.
- Batch processing refers to running batch jobs on a computer system.
- A batch job can be completed without user intervention. For example, consider a telephone billing application that reads phone call records from

the enterprise information systems and generates a monthly bill for each account.

- A batch window is referred as a period of less-intensive online activity, when the computer system is able to run batch jobs without interference from, or with, interactive online systems.
- The batch size refers to the number of work units to be processed within one batch operation.
- A master file contains descriptive data.
- A transaction file contains a set of transaction records.
- File updating refers to changing or modifying values in one or more records of a file, especially a data file, without changing the organization or semantics of the file.
- In a batch update, changes are collected over time and then all changes are applied to the file at once.
- In an online update, the user is directly connected to the computer and the changes are processed one at a time-often as the change occurs.

7.8 KEY WORDS

- **Batch processing**: It is the processing of transactions in a group or batch.
- **Batch size**: It refers to the number of work units to be processed within one batch operation.
- **File update**: It refers to changing or modifying values in one or more records of a file, especially a data file, without changing the organization or semantics of the file.

7.9 SELF-ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. What is batch size?
2. Define the term batch window.
3. What is master file?
4. State about the transaction file.
5. What do you understand by parallel batch processing?
6. What are the two types of file update?

Long-Answer Questions

1. What is batch processing? Explain giving examples.
2. Discuss the usage of common batch processing.
3. Explain the term file update.

7.10 FURTHER READINGS

Price, Wilson T. 1981. *Introduction to Computer Data Processing*, 3rd Edition. United States: Holt-Saunders International Editions.

Verzello, Robert J. and John Reutter. 1983. *Data Processing: Systems and Concepts*. New York: The McGraw-Hill Publishing Company, Inc.

Beech, Carol and Janice Burn. 1985. *Applications in Business Data Processing*. London (England): Pitman Publisher.

Fuori, William M. and Lawrence J. Aufiero. 1987. *Computers and Information Processing*. United States: Prentice Hall.

Sanders, Donald H. 1988. *Computers Today*. New York: McGraw-Hill.

Sinha, Pradeep K. and Priti Sinha. 2004. *Computer Fundamentals: Concepts, Systems & Applications*, 8th Edition. New Delhi: BPB Publications.

Rajaraman, V. 1999. *Fundamentals of Computers*. New Delhi: Prentice-Hall of India.

Lipschutz, Martin M. and S. Lipschultz. 1982. *Theory and Problems of Data Processing, Schaum's Outline Series*. New Delhi: Tata McGraw Hill Publishing Co. Ltd.

Stallings, William. 1996. *Data and Computer Communications*. New Jersey (USA): Prentice-Hall PTR.

BLOCK - III
DATA STORAGE

UNIT 8 DIRECT ACCESS STORAGE, RETRIEVAL AND FILE ORGANIZATION

Structure

8.0 INTRODUCTION

Direct access is also sometimes referred to as machine access or random access, direct access is a term used to describe a computer's ability to immediately locate and retrieve data from a storage device. In computing, a file system or filesystem (often abbreviated to fs) controls how data is stored and retrieved. Without a file system, data placed in a storage medium would be one large body of data with no way to tell where one piece of data stops and the next begins. By separating the data into pieces and giving each piece a name, the data is easily isolated and identified. Taking its name from the way paper-based data management system is named, each group of data is called a 'File'. The structure and logic rules used to manage the groups of data and their names is called a 'File System'.

There are many different kinds of file systems. Each one has different structure and logic, properties of speed, flexibility, security, size and more. Some file systems have been designed to be used for specific applications. For example, the ISO 9660 file system is designed specifically for optical discs. A computer file is a computer resource for recording data in a computer storage device. Just as words can be written to paper, so can data be written to a computer file. Files can be edited and transferred through the Internet on that particular computer system.

File organization refers to the way data is stored in a file. File organization is very important because it determines the methods of access, efficiency, flexibility and storage devices to use. There are four methods of organizing files on a storage media. A report generator is a computer program whose purpose is to take data from a source, such as a database, XML stream or a spreadsheet, and use it to produce a document in a format which satisfies a particular human readership.

In this unit, you will learn about the direct access storage and retrieval, file organization techniques and report generation.

8.1 OBJECTIVES

After going through this unit, you will be able to:

- Explain the direct access storage and retrieval
- Understand the importance of file organization techniques
- Explain about the report generation

8.2 DIRECT ACCESS STORAGE AND RETRIEVAL

Direct access is also sometimes referred to as machine access or random access, direct access is a term used to describe a computer's ability to immediately locate and **retrieve data** from a storage device.

A Direct Access Storage Device (DASD) is a secondary storage device in which each physical record has a discrete location and a unique address. IBM coined the term DASD as a shorthand describing hard disk drives, magnetic drums, and data cells. Later, optical disc drives and flash memory units are also classified as DASD. The term DASD contrasts with sequential storage media, such as magnetic tape, and unit record equipment, such as card devices like card readers and punches.

Access methods for DASD include sequential, indexed, and direct. Direct access contrasts with the sequential access method used in tape drives. A record on a DASD can be accessed without having to read through intervening records from the current location, whereas reading anything other than the 'Next' record on tape requires skipping over intervening records, and requires a proportionally long time to access a distant point in a medium.

The DASD storage class includes both fixed and removable media.

Direct Access Storage Devices (DASDs) are fixed or removable storage devices. Typically, these devices are rotating disk drives or solid state disks. A fixed storage device is any storage device defined during system configuration to be an integral part of the system DASD. The operating system detects an error if a fixed storage device is not available at some time during normal operation.

A removable storage device is any storage device defined by the person who administers your system during system configuration to be an optional part of the system DASD. The removable storage device can be removed from the system at any time during normal operation. As long as the device is logically unmounted first, the operating system does not detect an error.

The following types of devices are not considered DASD and are not supported by the Logical Volume Manager (LVM):

- CD-ROM (Compact Disk Read-Only Memory)
- DVD (Digital Video Disc or Digital Versatile Disc)
- WORM (Write-Once Read-Many)

8.3 FILE ORGANIZATION TECHNIQUES

In computing, a file system or filesystem (often abbreviated to fs) controls how data is stored and retrieved. Without a file system, data placed in a storage medium would be one large body of data with no way to tell where one piece of data stops and the next begins. By separating the data into pieces and giving each piece a name, the data is easily isolated and identified. Taking its name from the way paper-based data management system is named, each group of data is called a 'File'. The structure and logic rules used to manage the groups of data and their names is called a 'File System'.

There are many different kinds of file systems. Each one has different structure and logic, properties of speed, flexibility, security, size and more. Some file systems have been designed to be used for specific applications. For example, the ISO 9660 file system is designed specifically for optical discs. A computer file is a computer resource for recording data in a computer storage device. Just as words can be written to paper, so can data be written to a computer file. Files can be edited and transferred through the Internet on that particular computer system.

In computing, a file sequence is a well-ordered, finite collection of files, usually related to each other in some way. File sequences should ideally obey some kind of locality of reference principle, so that not only all the files belonging to the same sequence ought to be locally referenced to each other, but they also obey that as much as is their proximity with respect to the ordering relation. Explicit file sequences are, in fact, sequences whose filenames all end with a numeric or alphanumeric tag in the end (excluding file extension).

File

A file is a collection of or log of records. The type and frequency of access required determines the type of file organization to be used for a given set of records. A file is organized logically as a sequence of records. These records are mapped onto disk blocks.

File Organisation

The term 'File Organization' refers to the logical relationships among various records that constitute the file, particularly with respect to the means of identification and access to any specific record. In simple terms, storing the files in certain order is termed as 'File Organization'.

Fundamentally, file organization refers to the method in which the data is stored in a file. File organization is very significant because it determines the methods of access, efficiency, flexibility and storage devices to use. Following are four methods of organizing files on a storage media:

- Sequential File Organization
- Random or Direct File Organization
- Serial File Organization
- Indexed-Sequential File Organization

Sequential File Organization

Sequential access is a term describing a group of elements (such as, data in a memory array or a disk file or on magnetic tape data storage) being accessed in a predetermined, ordered sequence. It is the opposite of random access, the ability to access an arbitrary element of a sequence as easily and efficiently as any other at any time. Figure 8.1 illustrates the sequential access compared to random access.

Sequential access is sometimes the only way of accessing the data, for example if it is on a tape. It may also be the access method of choice, for example if all that is wanted is to process a sequence of data elements in order.

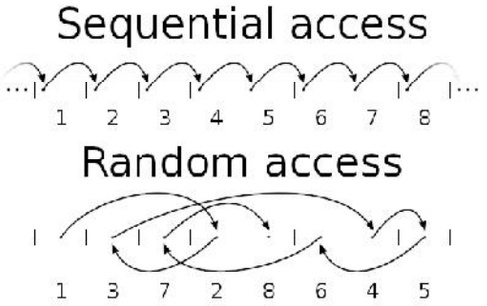


Fig. 8.1 *Sequential Access Compared to Random Access*

In the sequential file organisation:

- Records are stored and accessed in a particular order sorted using a key field.
- Retrieval requires searching sequentially through the entire file record by record to the end.

Since the record in a file are sorted in a particular order, therefore improved file searching methods, such as the binary search technique can be used to reduce the time used for searching a file .

Subsequently, because the records are sorted, hence it is possible to know in which half of the file a particular record being searched is located. Therefore, this method repeatedly divides the set of records in the file into two portions and searches only the portion on which the records is located or found.

For example, of the file has records with key fields 10, 20, 30, 40, 50, 60, 70 and the computer is searching for a record with key field 50, then it starts searching at 40 upwards in its search and ignores the first portion of the set.

Advantages of Sequential File Organization

Following are the advantages of sequential file organization:

- In the sequential file organization the sorting makes it easy to access records.
- The sequential file organization is efficient and process faster for the large volume of data.
- The binary search uses the split technique to reduce record search time by as much as half the time taken.
- The sequential file organization method can be implemented using economic storage devices, such as magnetic tapes.
- The sequential file organization requires less efforts to store and maintain data elements.
- The sequential file organization technique is useful for report generation and statistical computation process.

Random or Direct File Organization

Random access, more precisely and more generally called direct access, is the ability to access an arbitrary element of a sequence in equal time or any datum from a population of addressable elements roughly as easily and efficiently as any other, no matter how many elements may be in the set. In computer science, it is typically contrasted to sequential access which requires data to be retrieved in the order it was stored.

For example, data might be stored notionally in a single sequence like a row, in two dimensions like rows and columns on a surface, or in multiple dimensions. However, given all the coordinates, a program can access each record about as quickly and easily as any other. In this sense, the choice of datum is arbitrary in the sense that no matter which item is sought, all that is needed to find it is its address, i.e., the coordinates at which it is located, such as its row and column (or its track and record number on a magnetic drum). At first, the term 'Random Access' was used because the process had to be capable of finding records no matter in which sequence they were required. However, soon the term

NOTES

'Direct Access' gained favour because one could directly retrieve a record, no matter what its position might be. The operative attribute, however, is that the device can access any required record immediately on demand. The opposite is sequential access, where a remote element takes longer time to access.

In random or direct file organization.

- Records are stored randomly but accessed directly.
- To access a file stored randomly, a record key is used to determine where a record is stored on the storage media.
- Magnetic and optical disks permit data to be stored and accessed randomly.

Advantages of Random File Access

Following are the advantages of random file access:

- The random file access supports quick retrieval of records.
- In random file access, the records can be of different sizes.

Serial File Organization

Serial files are stored in chronological order that is as each record is received it is stored in the next available storage position. In general it is only used on a serial medium, such as magnetic tape. This type of file organisation means that the records are in no particular order and therefore to retrieve a single record the whole file needs to be read from the begging to end.

Serial file organisation is the simplest file organisation method. In serial files, records are entered in the order of their creation. As such, the file is unordered, and is at best in chronological order. Serial files are primarily used as transaction files in which the transactions are recorded in the order that they occur.

In serial file organization,

- Records in a file are stored and accessed one after another.
- The records are not stored in any way on the storage medium this type of organization is mainly used on magnetic tapes.

Advantages of Serial File Organization

Following are the advantages of serial file organization:

- The serial file organization is simple and easy method.
- The serial file organization is economical.

Indexed-Sequential File Organization

Indexed-sequential access file combines both sequential file and direct access file organization. In indexed-sequential access file, records are stored randomly on a direct access device, such as magnetic disk by a primary key.

Indexed-sequential file have multiple keys. These keys can be alphanumeric in which the records are ordered is called primary key.

The data can be accessed either sequentially or randomly using the index. The index is stored in a file and read into memory when the file is opened.

In the indexed-sequential file organization method,

- In the indexed-sequential file organization, an index is used to enable the computer to locate individual records on the storage media. For example, on a magnetic drum, records are stored sequential on the tracks.
- In the indexed-sequential file organization, each record is assigned an index that can be used to access it directly.

Advantages of Indexed-Sequential File Organization

Following are the advantages of indexed-sequential file organization:

- In indexed-sequential file organization, sequential file and random file access is possible.
- The indexed-sequential file organization accesses the records very fast if the index table is properly organized.
- In the indexed-sequential file organization records can be inserted in the middle of the file.
- The indexed-sequential file organization provides quick access for sequential and direct processing.
- The indexed-sequential file organization reduces the degree of the sequential search.

8.3.1 Report Generation

A report generator is a computer program whose purpose is to take data from a source, such as a database, XML stream or a spreadsheet, and use it to produce a document in a format which satisfies a particular human readership.

Report generation functionality is almost always present in database systems, where the source of the data is the database itself. Standalone report generators may work with multiple data sources and export reports to different document formats.

The report generation provides the capability to customize the report by selecting the essential information from the processed data set.

Creating or Generating a Report in Microsoft Excel

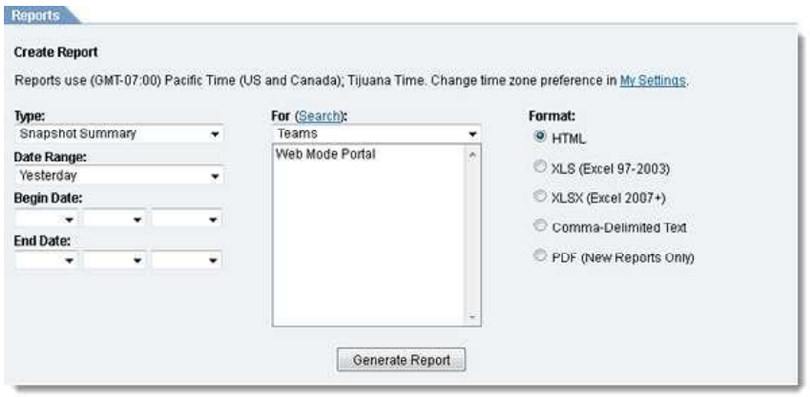
For creating and generating reports in Microsoft Excel, follow the steps given below.

1. For opening and creating a report in **Microsoft Excel** click **Controller → Reports → Open Report…**

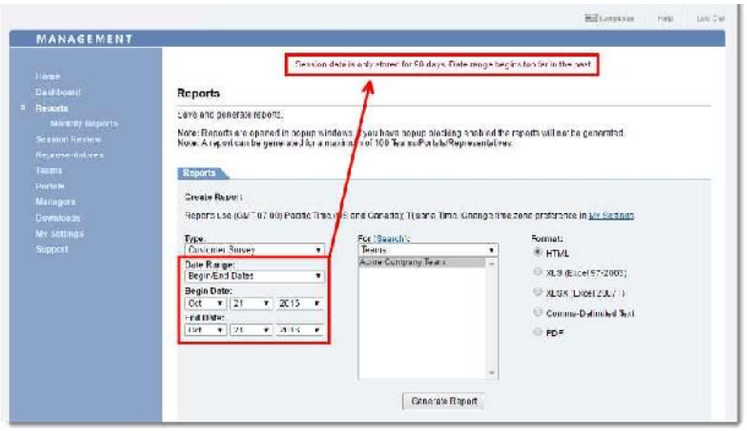
The Controller menu can be found on the Add-In Tab (Microsoft Excel 2007) when Microsoft Excel runs. When opening a form for data entry in Microsoft Excel, you have access to a Forms Toolbar, in addition to the Microsoft Excel functionality.

2. For running a report in Microsoft Excel click **Controller → Reports → Run Report...**
   - Enter the actual fact, period and forecast certainty for which you want to generate the report.
   - Enter the consolidation type and company for which you want to generate the report.
3. On the Reports tab, configure the following options:

   Type – Select the type of report to generate.
   - **Date Range** – Select an automatic time span or set specific start and end dates.
   - **For** – Indicate the focus of the report by selecting Representatives, Teams or Portals from the drop-down menu. Make multiple selections by holding down CTRL on your keyboard or clicking Search to open a new window.
   - **Format** – Select the type of file to generate, for example .html, .xls, .xlsx, .txt or .pdf as shown below in the screenshot.
   - Click on **Generate Report**.



4. If you try to generate or create a report using the 'Date Range: Begin/End Dates' drop-down menu, then after selecting a 'Begin Date' and 'End Date' for the session data outside of the date range permit, then you will come across a message, 'Session data is only stored for 90 days. Date range begins too far in the past'.



5. Saving a Report
   - In Microsoft Excel click **Controller → Reports → Save ReportSave** report. When you click on the Office Button Save As (Microsoft Excel 2007) converts the report to a Controller Link worksheet.
   - Enter a **Name** for the report and click on **Save Report**.

The Save Reports section permits you to save customized parameters for reports that are regularly or frequently run in the folder My Reports. To define and save a report, first create and define the report settings.

Check Your Progress

1. Define the term file sequence in computing.
2. What is a file?
3. Explain the term file organisation.
4. What are the methods of organizing files on a storage media?
5. What is the command for opening and creating a report in Microsoft Excel?

8.4 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. In computing, a file sequence is a well-ordered, finite collection of files, usually related to each other in some way. File sequences should ideally obey some kind of locality of reference principle, so that not only all the files belonging to the same sequence ought to be locally referenced to each other, but they also obey that as much as is their proximity with respect to the ordering relation. Explicit file sequences are, in fact, sequences whose filenames all end with a numeric or alphanumeric tag in the end (excluding file extension).

2. A file is a collection of or log of records. The type and frequency of access required determines the type of file organization to be used for a given set of records. A file is organized logically as a sequence of records. These records are mapped onto disk blocks.
3. The term 'File Organization' refers to the logical relationships among various records that constitute the file, particularly with respect to the means of identification and access to any specific record. In simple terms, storing the files in certain order is termed as 'File Organization'.
4. Following are four methods of organizing files on a storage media:
   - Sequential File Organization
   - Random or Direct File Organization
   - Serial File Organization
   - Indexed-Sequential File Organization
5. For opening and creating a report in Microsoft Excel click Controller →Reports →Open Report…

8.5 SUMMARY

- A Direct Access Storage Device (DASD) is a secondary storage device in which each physical record has a discrete location and a unique address. IBM coined the term DASD as a shorthand describing hard disk drives, magnetic drums, and data cells. Later, optical disc drives and flash memory units are also classified as DASD.
- The term DASD contrasts with sequential storage media, such as magnetic tape, and unit record equipment, such as card devices like card readers and punches.
- In computing, a file system or file system (often abbreviated to fs) controls how data is stored and retrieved. Without a file system, data placed in a storage medium would be one large body of data with no way to tell where one piece of data stops and the next begins.
- By separating the data into pieces and giving each piece a name, the data is easily isolated and identified. Taking its name from the way paper-based data management system is named, each group of data is called a 'File'.
- The structure and logic rules used to manage the groups of data and their names is called a 'File System'.
- There are many different kinds of file systems. Each one has different structure and logic, properties of speed, flexibility, security, size and more. Some file systems have been designed to be used for specific applications.
- In computing, a file sequence is a well-ordered, finite collection of files, usually related to each other in some way.

- File sequences should ideally obey some kind of locality of reference principle, so that not only all the files belonging to the same sequence ought to be locally referenced to each other, but they also obey that as much as is their proximity with respect to the ordering relation.
- Explicit file sequences are, in fact, sequences whose filenames all end with a numeric or alphanumeric tag in the end (excluding file extension).
- A file is a collection of or log of records. The type and frequency of access required determines the type of file organization to be used for a given set of records. A file is organized logically as a sequence of records. These records are mapped onto disk blocks.
- The term 'File Organization' refers to the logical relationships among various records that constitute the file, particularly with respect to the means of identification and access to any specific record. In simple terms, storing the files in certain order is termed as 'File Organization'.
- Sequential access is a term describing a group of elements (such as, data in a memory array or a disk file or on magnetic tape data storage) being accessed in a predetermined, ordered sequence. It is the opposite of random access, the ability to access an arbitrary element of a sequence as easily and efficiently as any other at any time.
- Random access, more precisely and more generally called direct access, is the ability to access an arbitrary element of a sequence in equal time or any datum from a population of addressable elements roughly as easily and efficiently as any other, no matter how many elements may be in the set.
- Serial files are stored in chronological order that is as each record is received it is stored in the next available storage position. In general it is only used on a serial medium, such as magnetic tape.
- In serial files, records are entered in the order of their creation. As such, the file is unordered, and is at best in chronological order.
- Serial files are primarily used as transaction files in which the transactions are recorded in the order that they occur.
- Indexed-sequential access file combines both sequential file and direct access file organization. In indexed-sequential access file, records are stored randomly on a direct access device, such as magnetic disk by a primary key.
- The data can be accessed either sequentially or randomly using the index. The index is stored in a file and read into memory when the file is opened.
- A report generator is a computer program whose purpose is to take data from a source, such as a database, XML stream or a spreadsheet, and use it to produce a document in a format which satisfies a particular human readership.

NOTES

8.6 KEY WORDS

- **File:** A file is a collection of or log of records, organized logically as a sequence of records. These records are mapped onto disk blocks.
- **File organisation:** The term 'File Organization' refers to the logical relationships among various records that constitute the file, particularly with respect to the means of identification and access to any specific record. i.e., storing the files in certain order.
- **Sequential access:** It describes a group of elements (such as, data in a memory array or a disk file or on magnetic tape data storage) being accessed in a predetermined, ordered sequence.
- **Random access:** Random access, more precisely and more generally called direct access, is the ability to access an arbitrary element of a sequence in equal time or any datum from a population of addressable elements roughly as easily and efficiently as any other, no matter how many elements may be in the set.
- **Serial file organisation**: It is the simplest file organisation method in which the records are entered in the order of their creation. Serial files are primarily used as transaction files in which the transactions are recorded in the order that they occur.
- **Indexed-sequential access file:** It combines both sequential file and direct access file organization. In indexed-sequential access file, records are stored randomly on a direct access device, such as magnetic disk by a primary key. The data can be accessed either sequentially or randomly using the index.

8.7 SELF-ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. Elaborate the terms file and file organisation.
2. Define the file organization techniques.
3. Differentiate between sequential file organization and random or direct file organization.
4. What is serial file organization?
5. Define the term indexed-sequential file organization.
6. Explain the term report generation.

Long-Answer Questions

1. What do you understand by direct access storage and retrieval?
2. Briefly explain the significance of file organization techniques giving appropriate examples.
3. Explain briefly the four methods of organizing files on a storage media giving advantages and appropriate examples of each type.
4. Discuss in detail the report generation process giving examples.
5. How the report can be generated in Microsoft Excel? Explain giving the important steps.

8.8 FURTHER READINGS

Price, Wilson T. 1981. *Introduction to Computer Data Processing*, 3rd Edition. United States: Holt-Saunders International Editions.

Verzello, Robert J. and John Reutter. 1983. *Data Processing: Systems and Concepts*. New York: The McGraw-Hill Publishing Company, Inc.

Beech, Carol and Janice Burn. 1985. *Applications in Business Data Processing*. London (England): Pitman Publisher.

Fuori, William M. and Lawrence J. Aufiero. 1987. *Computers and Information Processing*. United States: Prentice Hall.

Sanders, Donald H. 1988. *Computers Today*. New York: McGraw-Hill.

Sinha, Pradeep K. and Priti Sinha. 2004. *Computer Fundamentals: Concepts, Systems & Applications*, 8th Edition. New Delhi: BPB Publications.

Rajaraman, V. 1999. *Fundamentals of Computers*. New Delhi: Prentice-Hall of India.

Lipschutz, Martin M. and S. Lipschultz. 1982. *Theory and Problems of Data Processing, Schaum's Outline Series*. New Delhi: Tata McGraw Hill Publishing Co. Ltd.

Stallings, William. 1996. *Data and Computer Communications*. New Jersey (USA): Prentice-Hall PTR.

UNIT 9 FLOWCHARTS AND PROGRAMS

NOTES

Structure

9.0 INTRODUCTION

A flowchart is a type of diagram that represents a workflow or process. A flowchart can also be defined as a diagrammatic representation of an algorithm, a step-by-step approach to solving a task. The flowchart shows the steps as boxes of various kinds, and their order by connecting the boxes with arrows. This diagrammatic representation illustrates a solution model to a given problem. Flowcharts are used in analysing, designing, documenting or managing a process or program in various fields.

In this unit, you will learn about the flowcharts and programs for the functions.

9.1 OBJECTIVES

After going through this unit, you will be able to:

- Elaborate on the flowcharts for the file organization
- Write programs to read data from the text files

9.2 EXAMPLES OF FLOWCHARTS

A flowchart is a type of diagram that represents a workflow or process. A flowchart can also be defined as a diagrammatic representation of an algorithm, a step-by-step approach to solving a task. The flowchart shows the steps as boxes of various

kinds, and their order by connecting the boxes with arrows. This diagrammatic representation illustrates a solution model to a given problem. Flowcharts are used in analysing, designing, documenting or managing a process or program in various fields.

The two most common types of boxes in a flowchart are:

- A processing step is denoted as a rectangular box.
- A decision step is denoted as a diamond box.

A flowchart is described as 'cross-functional' when the chart is divided into different vertical or horizontal parts, to describe the control of different organizational units. A symbol appearing in a particular part is within the control of that organizational unit. A cross-functional flowchart allows the author to correctly locate the responsibility for performing an action or making a decision, and to show the responsibility of each organizational unit for different parts of a single process.

Types of Flowchart

Sterneckert (2003) suggested that flowcharts can be modelled from the perspective of different user groups (such as, managers, system analysts and clerks), and that there are four general types as defined below.

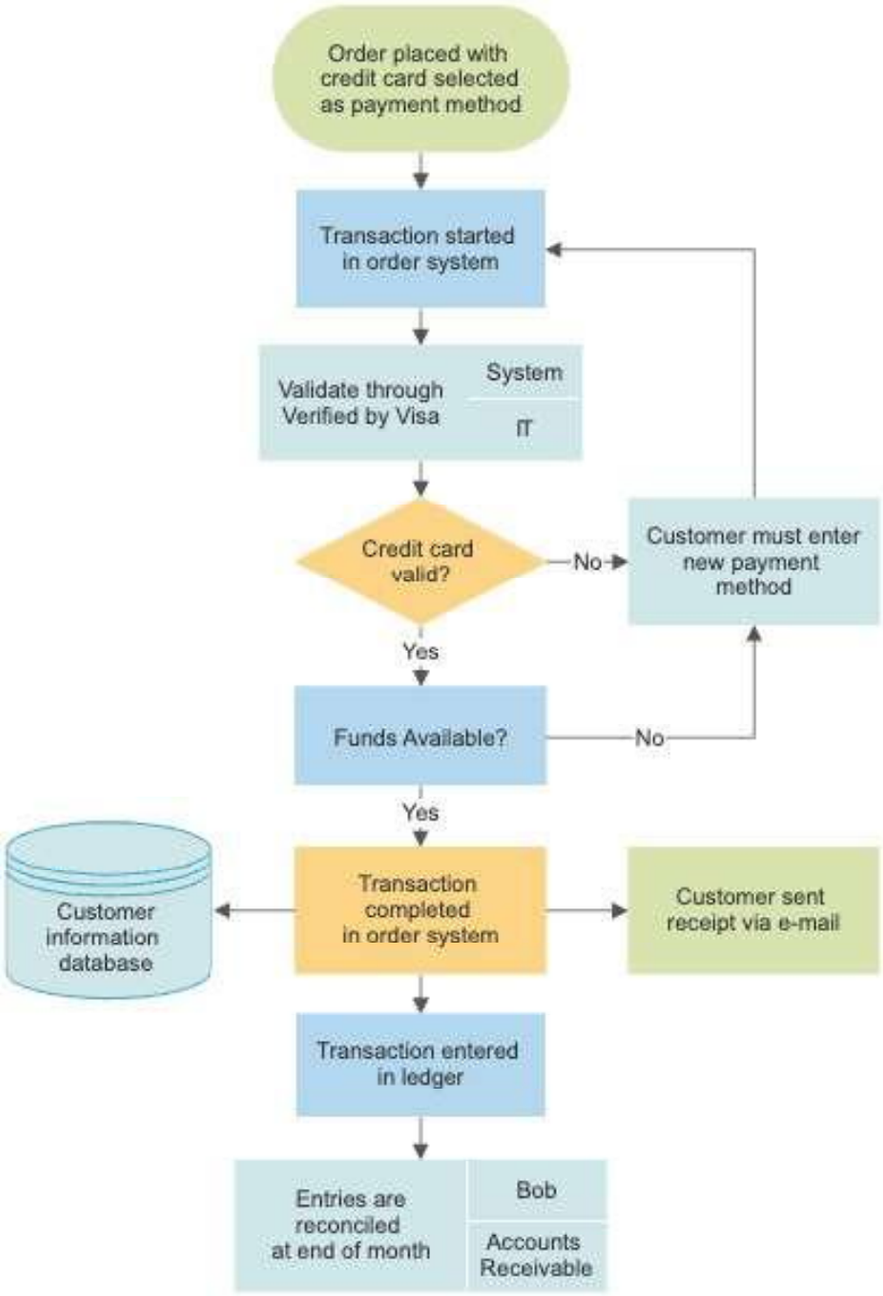
- **Document Flowcharts:** It shows the controls over a document-flow through a system.
- **Data Flowcharts:** It shows controls over a data-flow in a system.
- **System Flowcharts:** It shows controls at a physical or resource level.
- **Program Flowchart:** It shows the controls in a program within a system.

Every type of flowchart focuses on some kind of control, rather than on the particular flow itself. Following are some examples of flowchart types.

Transaction Processing Flowchart

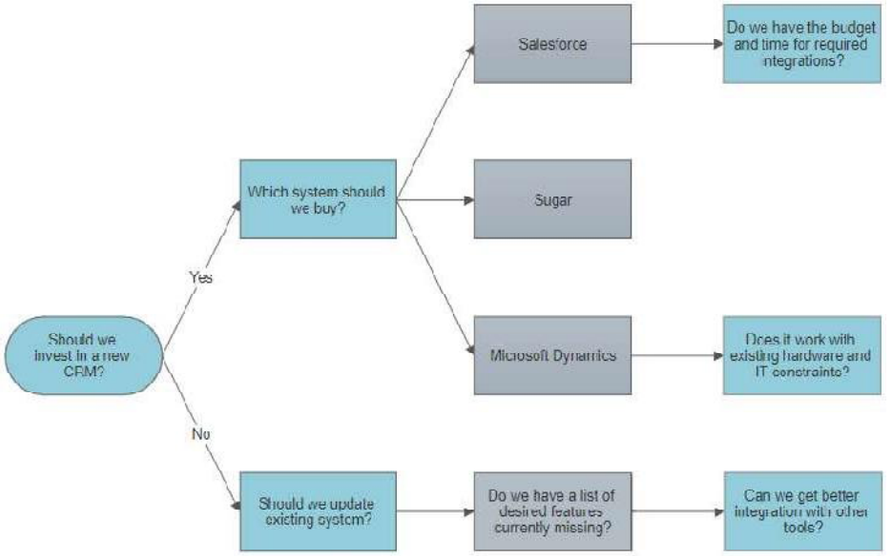
One of the most frequent uses of flowcharts is to map out a new project. They are particularly helpful when the project involves a sequence of steps that include decisions. Following is the example of the basic flowchart that shows the 'Online Transaction Processing' method when the customer places an order and the payment mode is selected as 'Credit Card'. This flowchart checks the validity and fund availability of the customer's credit card from the customer information database.

NOTES



Decision Flowchart

A decision flowchart helps in visualizing the options in any significant business decision. Consider all the questions in a flowchart helps to enhance the strategies before making a business decision. Following is an example of decision flowchart for investing a new CRM (Customer Relationship Management) for the enhancement of business revenue.



Process Map Flowchart

Flowcharts can be used for identifying and analysing a malfunction or to troubleshoot the problems, specifically in the fields of software and electronics. Business organizations also use flowcharts for process enhancement. This is performed by breaking down processes into smaller steps, then examining them closely for revealing the areas of both operating inefficiency and chance for enhancement.

A process map is a detailed or comprehensive flowchart specifically defined as an important tool for auditing a process. The creation of process map includes the following four steps:

Step 1: Identifying and understanding the steps in a process.

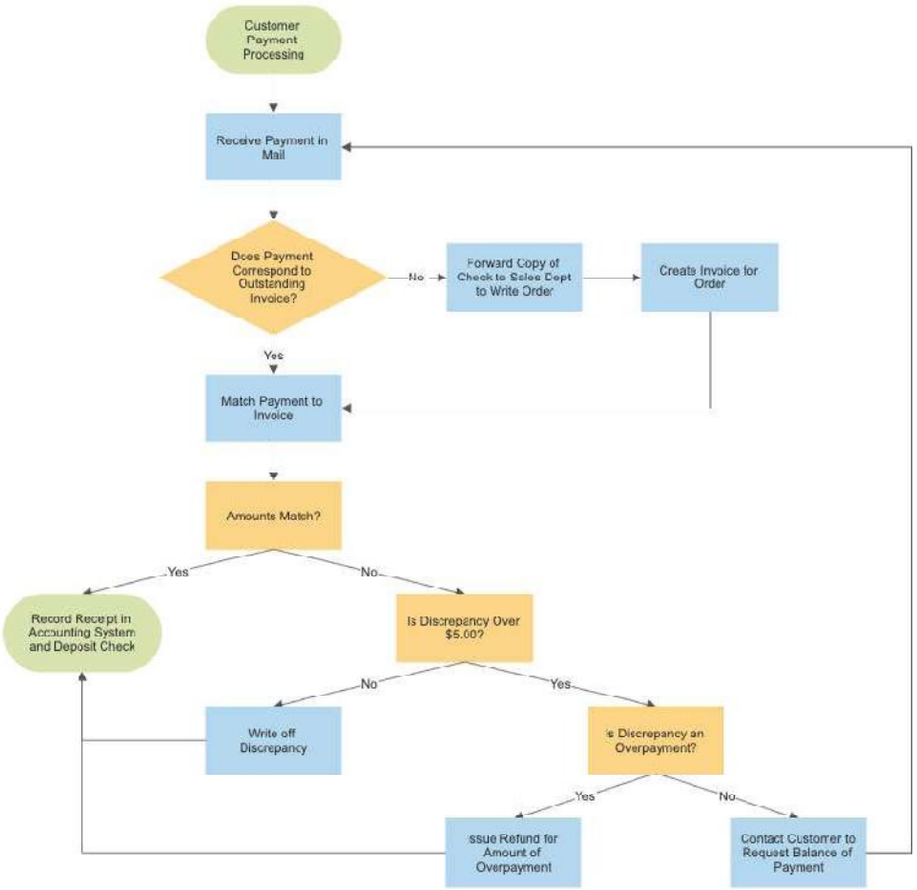
Step 2: Collecting information for identifying the objectives, risks and controls in a process.

Step 3: To interview the persons involved for creating the process map.

Step 4: Analysing and effecting changes for improving the process.

Following is an example of process map flowchart for E-Commerce - Customer Payment Processing.

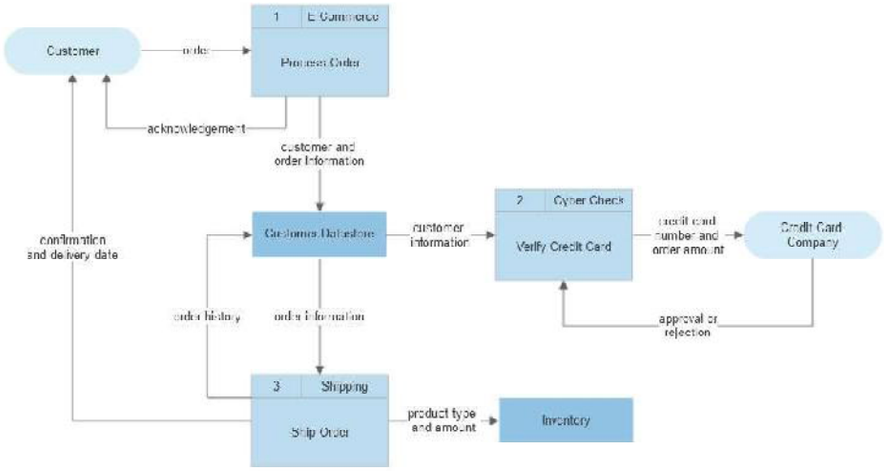
NOTES



Data Flow Diagram (DFD)

Data Flow Diagrams (DFDs) are considered an efficient method for bridging the communication gap between the system developers and the users. The DFDs are specialized flowcharts that filter a substantial amount of information into a relatively few symbols and connectors.

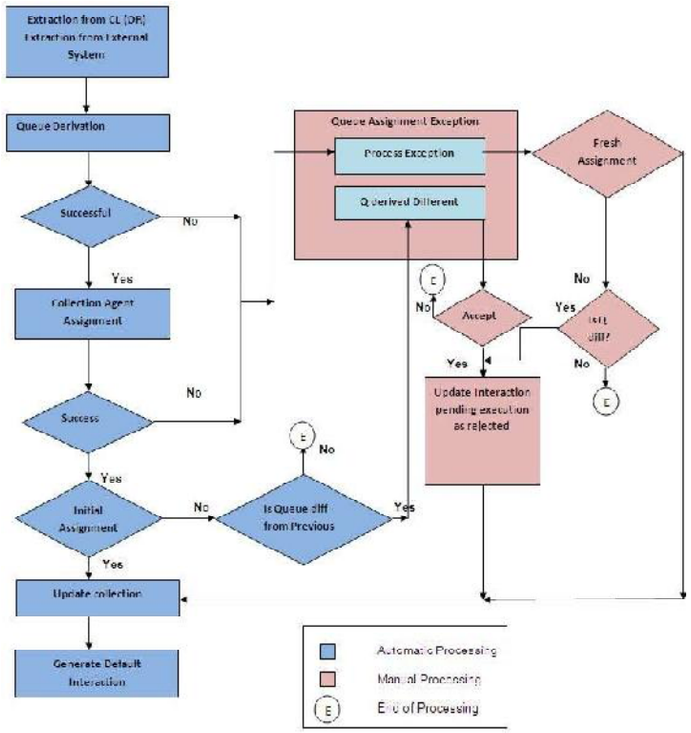
Fundamentally, a Data Flow Diagram (DFD) typically maps the flow of information for any process or system. It uses defined symbols, such as rectangles, circles and arrows, along with short text labels, to show data inputs, outputs, storage points and the routes between each destination. Following is the Data Flow Diagram (DFD) flowchart of E-Commerce – Online Order System.



Flowchart on Batch Processing

Batch processing is defined as the processing of transactions in a group or batch. No user interaction is required once batch processing is in progress.

The following flowchart shows an example of batch processing. It has small batches as Collections module or Collection process, Loan batch, etc. The batch processing is carried out in different phases, for example the Collection batch is processed once the Loans batch checks the amount to be paid by the customer.



(*Source: Oracle Help Centre*)

In the above flowchart, the Loan accounts with amount dues are extracted to Collection module. Extracted data from the Loans module or the Asset details are then typically used for further processing.

NOTES

9.3 PROGRAMS FOR THE FUNCTIONS

For writing programs to read data from text files, an input stream can be connected to a text file. With character-oriented input streams, characters are automatically translated from the external (disk file) format to the internal (Java char) format.

There are numerous ways of writing and reading a text file or a plain text file in Java. The `FileReader`, `BufferedReader` and `Scanner` classes can be used to read a text file. The `BufferedReader` provides buffering of data for fast reading while the `Scanner` provides parsing capability.

Both `BufferReader` and `Scanner` can be used to read a text file line by line in Java. The Java SE 8 introduces another Stream class **java.util.stream.Stream** which provides a more efficient way to read a file.

Reader Streams

Reader is the ancestor class of all character-oriented input streams. Figure 9.1 illustrates the classes in the **Reader** hierarchy.

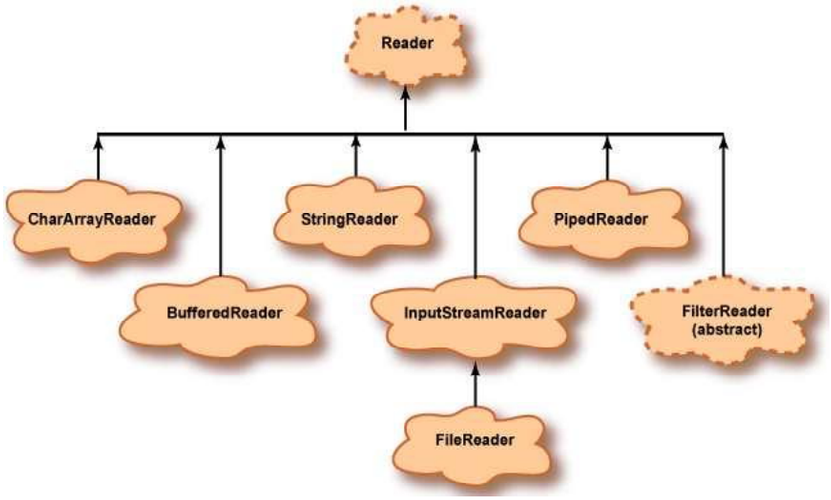


Fig. 9.1 *Classes in the Reader Hierarchy*

`FileReader` and `BufferedReader`

In Java, `FileReader` is used for input of character data from a disk file. The input file can be an ordinary ASCII, one byte per character text file. A **Reader** stream automatically translates the characters from the disk file format into the internal Java **char** format.

The characters in the input file might be from other alphabets also which is supported by the UTF (Unicode Transformation Format) format, in this case there will be up to four bytes per character. In this case also the characters from the file are translated into Java **char** format. Figure 9.2 illustrates the concept of `FileReader` and `BufferedReader`, specifying 8-bit characters and 16-bit characters, respectively.

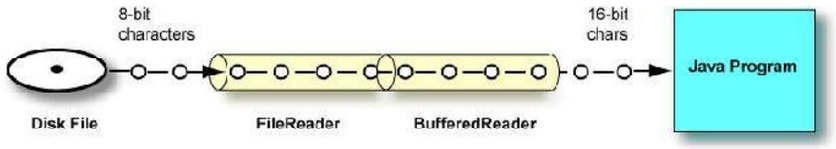


Fig. 9.2 *`FileReader` and `BufferedReader`*

The `readLine()` method reads a line of characters from a `BufferedReader` class.

Example Program 1: Reading line of characters from a `BufferedReader` class using `readLine()` method.

Java Program

```
import java.io.*;
class ReadTextFile
{
public static void main(String[] args)
 {
 String fileName = "reaper.txt";
 String line;

  try
  {
  BufferedReader in = new BufferedReader(new FileReader(
fileName ));

   line = in.readLine();
   while (line != null) // while not end of file
    {
    System.out.println(line);
    line = in.readLine();
    }
   in.close();
  }
 catch (IOException iox)
```

```
        {
        System.out.println("Problem reading " + fileName);
        }
      }
    }
```

NOTES

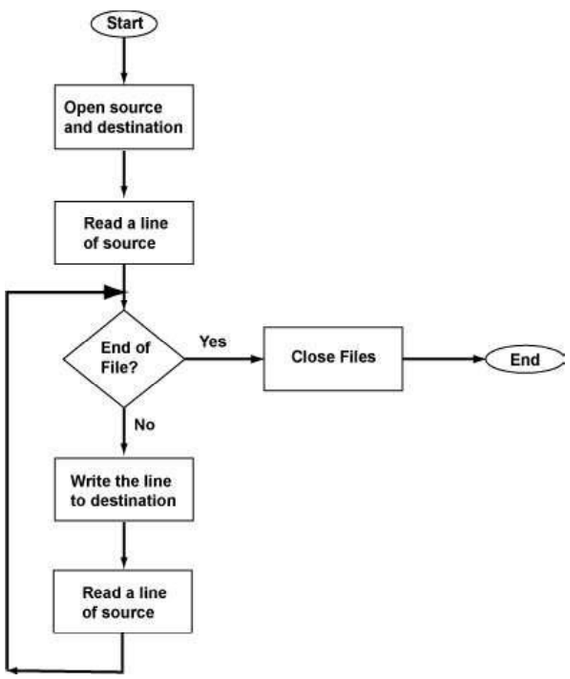
In the above Java code, the `readLine()` method reads a line of text from a character-oriented input stream, and puts it into a new String object which it returns as a reference. If there is no more data in the file, it returns `null`.

This program opens a file for reading. The already created file `reaper.txt` is opened for reading when the `FileReader` stream is constructed. If the file does not exist in the current directory, an `IOException` is thrown.

Subsequently, the program reads each line of the file and writes it to the monitor. When end-of-file is detected the program quits. This is a simple and common programming pattern for reading and processing data while not end-of-file. Usually in documentation, End-Of-File is abbreviated as EOF.

Flow Chart of the Program

The general or common scheme of the execution of any program is given in the following flowchart. It does not contain the detailed of each and every step defined in the program. The key aim of the flowchart is to demonstrate and analyse the overall logic of the program. For example, the box 'Read a line of source' appears in two places, i.e., before the loop gets started and in the body of the loop. The loop terminates when it encounters End-Of-File.



Example Program 2: Java program to illustrate reading from `FileReader` using `BufferedReader`.

Java Program

NOTES

```
Java Program to illustrate reading from FileReader
//using BufferedReader
import java.io.*;
public class ReadFromFile2
{
  public static void main(String[] args) throws Exception
  {
  // We need to provide file path as the parameter:
  // double backquote is to avoid compiler interpret words
  // like \test as \t (ie. as a escape sequence)
   File file = new File("C:\\Users\\pankaj\\
Desktop\\test.txt");

   BufferedReader br = new BufferedReader(new
FileReader(file));

  String st;
  while ((st=br.readLine()) !=null)
    System.out.println(st);
  }
}
```

Program on Getting File Information

To get detailed information about a file, use any of the `File` methods.

Example Program 3: Java program to get the detailed information on a specific file.

Java Program

```
import java.io.File; // Import the File class

public class GetFileInfo {
 public static void main(String[] args) {
  File myObj = new File("filename.txt");
  if (myObj.exists()) {
   System.out.println("Filename: " + myObj.getName());
       System.out.println("Absolute path: " +
myObj.getAbsolutePath());
```

NOTES

```
        System.out.println("Writeable: " + myObj.canWrite());
        System.out.println("Readable " + myObj.canRead());
          System.out.println("File size in bytes " +
    myObj.length());
        } else {
        System.out.println("The file does not exist.");
        }
       }

}
```

The output will be:

```
File name: filename.txt
Absolute path: C:\Users\MyName\filename.txt
Writeable: true
Readable: true
File size in bytes: 0
```

Check Your Progress

1. Define the term flowchart.
2. What are the four general types of flowchart?
3. How the Java programs read data from text files?

9.4 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. A flowchart is a type of diagram that represents a workflow or process. A flowchart can also be defined as a diagrammatic representation of an algorithm, a step-by-step approach to solving a task. The flowchart shows the steps as boxes of various kinds, and their order by connecting the boxes with arrows. This diagrammatic representation illustrates a solution model to a given problem.
2. Following are the four general types of flowchart.
   - Document Flowcharts: It shows the controls over a document-flow through a system.
   - Data Flowcharts: It shows controls over a data-flow in a system.
   - System Flowcharts: It shows controls at a physical or resource level.
   - Program Flowchart: It shows the controls in a program within a system.

NOTES

3. For writing programs to read data from text files, an input stream can be connected to a text file. With character-oriented input streams, characters are automatically translated from the external (disk file) format to the internal (Java char) format.

   There are numerous ways of writing and reading a text file or a plain text file in Java. The `FileReader, BufferedReader` and `Scanner` classes can be used to read a text file. The `BufferedReader` provides buffering of data for fast reading while the Scanner provides parsing capability. Both `BufferReader` and `Scanner` can be used to read a text file line by line in Java.

9.5 SUMMARY

- A flowchart is a type of diagram that represents a workflow or process. A flowchart can also be defined as a diagrammatic representation of an algorithm, a step-by-step approach to solving a task.
- The flowchart shows the steps as boxes of various kinds, and their order by connecting the boxes with arrows. This diagrammatic representation illustrates a solution model to a given problem. Flowcharts are used in analysing, designing, documenting or managing a process or program in various fields.
- Sterneckert (2003) suggested that flowcharts can be modelled from the perspective of different user groups (such as, managers, system analysts and clerks).
- Document flowcharts shows the controls over a document-flow through a system.
- Data flowcharts shows controls over a data-flow in a system.
- System flowcharts shows controls at a physical or resource level.
- Program flowchart shows the controls in a program within a system.
- For writing programs to read data from text files, an input stream can be connected to a text file. With character-oriented input streams, characters are automatically translated from the external (disk file) format to the internal (Java char) format.
- There are numerous ways of writing and reading a text file or a plain text file in Java. The `FileReader, BufferedReader` and `Scanner` classes can be used to read a text file. The `BufferedReader` provides buffering of data for fast reading while the Scanner provides parsing capability.
- Both `BufferReader` and `Scanner` can be used to read a text file line by line in Java. The Java SE 8 introduces another Stream class java.util.stream. Stream which provides a more efficient way to read a file.
- Reader is the ancestor class of all character-oriented input streams.

NOTES

9.6 KEY WORDS

- **Document flowcharts**: It shows the controls over a document-flow through a system.
- **Data flowcharts**: It shows controls over a data-flow in a system.
- **System flowcharts**: It shows controls at a physical or resource level.
- **Program flowchart**: It shows the controls in a program within a system.
- **Reader**: It is the ancestor class of all character-oriented input streams.

9.7 SELF-ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. What is a flowchart?
2. Elaborate on the system flowcharts.
3. Define `Reader` class of Java.

Long-Answer Questions

1. What are flowcharts? Explain the different types of flowcharts giving appropriate examples.
2. Draw and analyse a flowchart for smooth workflow process in an organisation.
3. Draw and analyse a flowchart for event-driven process chain in an organisation.
4. Draw a process flow flowchart on oil and petroleum refining.
5. Discuss in detail on the Java `FileReader`, BufferedReader or `Scanner` classes giving appropriate examples.
6. Write three Java programs to read a text file using Java `FileReader`, `BufferedReader` and `Scanner` classes.

9.8 FURTHER READINGS

Price, Wilson T. 1981. *Introduction to Computer Data Processing*, 3rd Edition. United States: Holt-Saunders International Editions.

Verzello, Robert J. and John Reutter. 1983. *Data Processing: Systems and Concepts*. New York: The McGraw-Hill Publishing Company, Inc.

NOTES

Beech, Carol and Janice Burn. 1985. *Applications in Business Data Processing*. London (England): Pitman Publisher.

Fuori, William M. and Lawrence J. Aufiero. 1987. *Computers and Information Processing*. United States: Prentice Hall.

Sanders, Donald H. 1988. *Computers Today*. New York: McGraw-Hill.

Sinha, Pradeep K. and Priti Sinha. 2004. *Computer Fundamentals: Concepts, Systems & Applications*, 8th Edition. New Delhi: BPB Publications.

Rajaraman, V. 1999. *Fundamentals of Computers*. New Delhi: Prentice-Hall of India.

Lipschutz, Martin M. and S. Lipschultz. 1982. *Theory and Problems of Data Processing, Schaum's Outline Series*. New Delhi: Tata McGraw Hill Publishing Co. Ltd.

Stallings, William. 1996. *Data and Computer Communications*. New Jersey (USA): Prentice-Hall PTR.

NOTES

UNIT 10 APPLICATIONS: INVENTORY CONTROL AND ACCOUNTING

Structure

10.0 INTRODUCTION

Data processing refers to manipulation of data by a computer. It includes the conversion of raw data to machine-readable form, flow of data through the CPU

(Central Processing Unit) and memory to output devices, and formatting or transformation of output. Any use of computers to perform defined operations on data can be included under data processing.

Data processing is, generally, 'The collection and manipulation of items of data to produce meaningful information'. In this sense it can be considered a subset of information processing, 'The change (processing) of information in any manner detectable by an observer'. Data processing is the re-structuring or re-ordering of data by people or computers to increase their usefulness and add values for a specific purpose.

In this unit, you will learn about the various data processing applications, i.e., inventory control and accounting, payroll, production planning and control.

10.1 OBJECTIVES

After going through this unit, you will be able to:

- Discuss the role of IT in data processing
- Understand the inventory control and accounting software

- Describe the payroll and its software
- Explain the objectives, elements and stages of production planning and control

NOTES

10.2 APPLICATIONS

Information Technology (or IT as it is popularly called) has significant role in data processing, as the computerised data techniques provide accurate results and saves time. IT has dramatically changed the way of our life. After the discovery of electricity, the computer ranks as one of the most important breakthroughs of the modern era. Like electricity, IT has impacted all facets of life and, in fact, its usage is so ubiquitous that it is hard for today's generation to even visualize how our ancestors lived without computers. From medicine to transportation, from banking to the entertainment industry, there is hardly any industry or sector that does not deploy IT in a fundamental manner to achieve one of the three goals:

(a) Reducing the cost of operations by increasing operational efficiency and staff productivity.

(b) Improving revenues and bottom lines by helping management in informed decision-making and focusing on priority areas.

(c) Improving customer satisfaction by providing better, faster and value-added services.

IT has opened up several allied industries and employment opportunities which never existed before. Whether it is Business Process Outsourcing or BPO (remote data processing) or web-enabled services (medical transcription, call centres, etc.), IT has opened up new avenues for jobs. Thanks to the Internet, developing nations today can also participate in the global economy and help bridge the divide between haves and have-nots.

10.3 INVENTORY CONTROL AND ACCOUNTING

Inventory control or stock control can be broadly defined as, 'The activity of checking a shop's stock'. It is the process of ensuring that the right amount of supply is available within a business. However, a more focused definition takes into account the more science-based, methodical practice of not only verifying a business' inventory but also maximising the amount of profit from the least amount of inventory investment without affecting customer satisfaction. Other facets of inventory control include forecasting future demand, supply chain management, production control, financial flexibility, purchasing data, loss prevention and turnover, and customer satisfaction.

An extension of inventory control is the inventory control system. This may come in the form of a technological system and its programmed software used for

NOTES

managing various aspects of inventory problems, or it may refer to a methodology (which may include the use of technological barriers) for handling loss prevention in a business. The inventory control system allows for companies to assess their current state concerning assets, account balances, and financial reports.

Inventory control typically refers to the processes employed to maximize a company's usage of inventory. The aim of inventory control is to generate the maximum profit from the least amount of inventory investment without intruding upon customer satisfaction levels.

Accounting or Accountancy is the measurement, processing and communication of financial and non-financial information about economic entities, such as businesses and corporations. Accounting, which has been called the 'Language of Business', measures the results of an organization's economic activities and conveys this information to a variety of users, including investors, creditors, management, and regulators. Practitioners of accounting are known as accountants. The terms 'Accounting' and 'Financial Reporting' are often used as synonyms.

Accounting can be divided into several fields including financial accounting, management accounting, external auditing, tax accounting and cost accounting. Accounting information systems are designed to support accounting functions and related activities. Financial accounting focuses on the reporting of an organization's financial information, including the preparation of financial statements, to the external users of the information, such as investors, regulators and suppliers; and management accounting focuses on the measurement, analysis and reporting of information for internal use by management. The recording of financial transactions, so that summaries of the financials may be presented in financial reports, is known as bookkeeping, of which double-entry bookkeeping is the most common system.

Inventory Control System

Inventory is a very crucial issue for any manufacturing unit, warehouse, clearing and forwarding agents or a distributor. Stacked inventory means blocked money that reduces liquidity and indirectly profits. Cash discounts are available from almost all manufactures because liquidity means more rotation of goods and more profits. On the other hand, lower inventory leads to losses. So, optimum stock levels optimize operational efficiency.

Most large manufacturing units usually need a variety of raw material for production. And manually managing such a large number of raw material and then keeping track of the finished products is not an easy task. IT plays a useful role here; a computerized inventory management system provides the facility of specifying the right amount of inventory in hand and determines the time at which, and the amount of inventory required.

The computerized inventory control system maintains a date-wise list of all items along with the maximum, minimum, reorder and inventory in hand. It keeps

on automatically updating the same list as and when material is despatched or received making things much easier and simpler for the production unit. The computerized inventory control system is also capable of preparing many other useful MIS (Management Information System) reports, such as aging analysis, goods movement analysis, slow and fast moving stock report, valuation report and it assists the storekeeper and accountants.

ERP software like ORACLE Financials, BAAN, SAP are some examples of sophisticated inventory control packages which can also generate purchase orders automatically whenever the minimum level of stocks is reached, provide automatic posting of accounting entries as soon as any purchase or sale is carried out and generate analytical reports which (itemize) show the previous and future trends in inventory consumption.

Many organizations nowadays have inventory control systems connected through the Internet or the intranet whereby the request for purchase or the purchase order is instantly delivered to the department or the vendor through e-mail, the moment they receive an order or request for an item which is out of or low in stock. Some inventory control software packages have a feature of automatically generating a purchase or a supply order electronically and mailing it to the preferred supplier, without any human intervention.

Some interesting IT-based innovations used to simplify and improve inventory management are as follows:

Barcode Readers: Barcodes are a series of black and white parallel and adjacent bars with spaces which represent a string of character. Barcode printers are used to encode and print the barcodes and then a barcode reader, which is actually a barcode scanner, is used to decode the barcodes. Bar coding is a quick and easy method for automatically reading barcodes from the products identifying their batch numbers, manufacturing and expiry dates, etc., without having to manually read and type it in the computer to generate bills and track inventory.

Hand-Held Terminals (HHTs): HHTs are microprocessor-based simple devices used to communicate with any type of microprocessor-based device. HHTs' standard input device is basically a calculator-like device with a small LCD display output. It is a compact, easy to use and a rugged device designed for collecting data from large warehouses.

Following are some of the additional common areas in which the implementation of inventory control is done:

Availability of Raw Materials: There must be sufficient raw materials inventory on hand to ensure that new jobs can be launched in the production process in a timely manner, but not so much that the company is investing in an inordinate amount of inventory. The key control planned to check the balance and to order the stock frequently in small lots from the suppliers engaging them in just-in-time deliveries.

NOTES

Availability of Finished Goods: A company may charge a higher price for its products if it can be reliably shipped to customers simultaneously. Consequently, there may be a special pricing associated with having high levels of finished goods on hand. Though, the cost of investing in the inventory may exceed the expected profits, hence inventory control involves balancing the proportion of acceptable backorders with a reduced level of on-hand finished goods.

Work in Process: The amount of inventory can be reduced by checking the amount of work in the production process, which can reduce the investment in the inventory. This involves comprehensive actions, such as using production units to work in smaller groups, shifting the work area to nearby location to reduce the amount of inventory travel time, reducing machine setup times to switch to new jobs, and minimizing job sizes.

Computerised Accounting System

The computerised accounting system is an accounting data system that processes the financial transactions and transactions as per GAAP (Generally Accepted Accounting Principles) to produce reports as per user demands. Every accounting system, either computerised or manual, has two aspects. First, it has to work under a set of well-defined theories known as accounting principles. Another, that there is a user-defined structure for the maintenance of records and the creation of reports.

In a computerised accounting system, the structure of storage and processing of data is known as an operating environment that comprises of hardware as well as software in which the accounting system, works. The kind of accounting system used defines the operating environment. Both hardware and software are interdependent. The type of software defines the formation of the hardware. Further, the choice of hardware is reliant upon many factors, such as the number of users, level of privacy and the nature of multiple activities of operational departments in an organisation.

Computerized Accounting Software

Computerized accounting software is used for evaluating the profitability and financial performance of different organizations. Manual accounting has become redundant nowadays, as they are prone to human error. Thus, businesses of all sizes rely on the best accounting software in India to make the complex accounting task more manageable and error-free.

Computerized accounting software is a software solution that can be installed on your computer for seamlessly managing your company's financial dealings, reports, data, and statements. The software can also be accessed remotely via your mobile device with the help of mobile data or a strong Wi-Fi connection.

Computerized accounting solutions acts as the primary tool for recording, tracking and generating accounting transaction reports of small and medium-sized

enterprises. The modules include bookkeeping for tracking and managing payments and collections; invoicing for generating invoices for your customers; general ledger for summarizing all the transactions taking place in the organization; inventory module for tracking products from purchase to sale; accounts payable for processing vendor invoices; accounts receivable for generating customer invoices; purchase order module for tracking and managing all inventory orders; and debt collection module for collecting late invoices. These modules help in controlling the different aspects of accounting processes efficiently.

Features of Computerized Accounting Software

A robust accounting software solution is proficient for expanding the business to new heights of success. The invoicing software not only provides the overview of the financial transactions but it also helps to maintain the legal obligations required for running the business smoothly.

Following are some of the useful features of computerised accounting software:

Invoicing: This is the basic module and component of all accounting software that automates and systemises the process of collections so that the invoice can be sent timely. An invoicing system stores all the data of the customer and also timely generates or produces accurate invoices for clients, in addition it reminds them about outstanding bills as per the industry standards.

Reporting: The features of computerized accounting software provides a broad array of reporting options, such as balance sheet, income statement highlighting the profit and loss, cash flow statement, payroll summary, etc. It has a customizable reporting feature that helps to generate reports as per the preferences or requirements. The computerized accounting software systems convert the financial data of the company into graphical format so that it can be comprehend that where the money is being spent.

Banking: Computerized accounting software keeps track of the banking accounts of the company and also imports data from the respective bank into the system. The automatic payment module prints checks, and schedules bank payments and direct deposits, therefore, automating or systematising the bank payment processes.

Budgeting and Forecasting: Accurate computerized accounting software calculates and depicts the financial performance of the company for the current year and upcoming financial year. It is capable of generating balance sheet summaries to describe and interpret how well the business is performing at that moment so that the realistic sales targets can be set.

Inventory Management: An efficient computerized billing software solution typically manages the stock at hand which helps to keep checks about the availability of the products so that the delivery issues can be speedily managed.

Payroll: Some of the computerized accounting software also provides the sophisticated payroll modules for perfectly managing all aspects of payroll. This

NOTES

module enables the company's account department to generate employee payments and process their checks on time. In addition, it also simplifies or streamlines the process of managing variable pay schedules along with bonuses, commissions, profit-sharing, etc. The computerized accounting software also gives permission to observe the tax commitments with perfection.

Fund Accounting: Accounting system is specifically designed for NGOs (Non-Governmental Organizations) and public agencies since it comes with fund accounting module for grant management, tracking donations, and managing GASB (Governmental Accounting Standards Board) regulations. GASB is the source of Generally Accepted Accounting Principles (GAAP) used by state and local governments. The GASB is subject to oversight by the Financial Accounting Foundation (FAF), which selects the members of the GASB and the Financial Accounting Standards Board (FASB), and it funds both organizations. Its mission is to establish and improve standards of state and local governmental accounting and financial reporting that will result in useful information for users of financial reports and guide and educate the public, including issuers, auditors, and users of those financial reports.

Ability to Collaborate: Computerized billing systems facilitate to authorize contractual accountants to access the financial data. The data can be easily synchronized with the bank account and can import the financial reports in seconds.

Multi-Functionality: An efficient computerized accounting billing solution is also capable of recording and delineating financial dealings in multiple currencies.

Simplifies Tax Compliance: Automated computerized accounting software system provides a fast overview of the amount of taxes that the company has paid till date. It also computes the amount of tax that the company or the clients have to pay on each invoice.

Computerized accounting software streamlines accounting tasks and offers several benefits to its users. It not only automates the accounting responsibilities but also helps to chalk out a results-driven investment strategy.

Financial Accounting

In most cases, computerization of business organizations started with the computerization of financial accounting systems.

The importance of financial accounting for any business organization cannot be overemphasized, but the degree of transactions, the recurring and structured nature of the data and the sheer volumes involved in the case of large corporate make for an ideal case for computerization. Computerization of accounts made life much more easier for accountants as they no longer had to maintain manual books, filled vouchers, update registers, maintain long ledgers and then spend days cross-checking the manual entries. Now they can concentrate more on analysing information.

Features like ledger database, automatic calculation, figures tally systems, checks and validations, automatic posting of entries from the voucher to the profit and loss account and the balance sheet, inventory systems, invoicing, creation of challans and purchase orders, creation of relevant reports, interconnectivity between users at different geographical locations and many more, make the computerized accounting systems an error-free, time saving and a fairly simple system to use.

According to the complexity of the operation and size of the organization, various ready-made financial accounting software are available in the market .Tally, EX and Busy are financial accounting software which are good enough for most small and medium scale organizations.

On the other hand, larger organizations with offices in multiple locations and multiple operations use Enterprise Resource Planning (ERP) software like ORACLE Financials, BAAN and SAP. Though ERP software cost millions of rupees and are comparatively much more complex to implement, they provide an excellent platform for ensuring that the company's system and procedures are consistently followed at multiple locations at the same time. Since ERP software provide complete integrated solutions for all functions of business such as financial accounting, payroll management and inventory control, they make it easy for the companies with office at multiple locations to consolidate huge information at a centralized place, thus, enabling the top management to get effective real-time management information.

Check Your Progress

1. What is the aim of inventory control?
2. Why is fund accounting designed?

10.4 PAYROLL

In treasury management, a payroll is the list of employees of some company that is entitled to receive payments as well as other work benefits and the amounts that each should receive. Along with the amounts that each employee should receive for time worked or tasks performed, payroll can also refer to a company's records of payments that were previously made to employees, including salaries and wages, bonuses, and withheld taxes, or the company's department that deals with compensation. One way that payroll can be handled is in-house, meaning that a company handles all aspects of the payroll process on its own, including timesheets, calculating wages, producing paychecks, sending the ACH (Automated Clearing House), for any direct deposits, and remitting any tax payments essential. Payroll can also be outsourced to a full-service payroll processing company. When a company chooses to outsource their payroll, timesheets, wage calculations, creating paychecks, direct deposits, and tax payments can be handled all, or in part, by the payroll company.

NOTES

Payroll plays a major role in the internal operations of a business for several reasons. From the perspective of accounting, payroll and payroll taxes are subject to laws and regulations. Payroll also plays a large role from the human resources point of view. Payroll errors, such as late or incorrect paychecks, are a sensitive topic that can cause tension between employees and their employers. One requirement for maintaining high employee morale is that payroll must be paid accurately and promptly because employees are very sensitive to any payroll errors.

Companies typically process payroll at regular intervals. This interval varies from company to company and will often differ within the company for different employees for larger companies. According to the research conducted in 2019, the four most common pay frequencies are weekly, biweekly, semi-monthly and monthly. The other, much less common payroll frequencies include daily, four-weekly, bimonthly, quarterly, semi-annually and annually.

Weekly payrolls have 52 40-hour pay periods per year and include one 40 hour workweek for overtime calculations. Biweekly payrolls consist of 26 80-hour pay periods per year and consist of two 40 hour work weeks for overtime calculations. Weekly and biweekly payrolls are the most common for non-exempt employees because they are the two that allows for the easiest and most transparent overtime calculations.

Semi-monthly payrolls have 24 pay periods per year with two pay dates per month. These pay dates are commonly paid on either the 1st and the 15th day of the month or the 15th and the last day of the month and consist of 86.67 hours per pay period. Monthly payrolls have 12 pay periods per year with a monthly payment date. Each monthly payroll consists of 173.33 hours. Both semi-monthly and monthly payrolls are more common for exempt employees that are earning a set salary each payroll. This is because overtime can be confusing and difficult to follow as it may be earned in one week but then falls under a different pay period.

Daily payroll is paid to the employee on a day-to-day basis. Four-weekly payrolls are paid once every four weeks or 13 times per year. Bimonthly payrolls are paid once every two months or six times per year. These three payroll options are very rarely used but are used for regular payroll when used. Quarterly payrolls are paid once at the end of each quarter, or four times per year. Semi-annually payrolls are paid twice per year and annual payrolls are paid once per year. These three payroll options are mostly used for payments, such as bonuses or owner profit sharing or capital gain payrolls.

Each employee views their individual payroll on either a physical check stub or via an online portal. When viewing this stub, different components are separate from one another yet all tie together to complete the transaction. The first component is gross pay, or gross income. Gross pay is simply the total amount of compensation that an employee will receive before any deductions or reimbursements are made, including, but not limited to, regular wages, overtime pay, commissions, and bonuses. The next part of a paycheck is any pretax

deductions that may be applicable. These could include insurances, such as health, dental, or life insurance, deductions for certain retirement accounts, or deductions for FSA (Flexible Spending Account) or HSA (Health Savings Account) accounts.

A paycheck also includes taxes. Taxes taken out of the paycheck are collected by the employer and then paid to either the federal, state or local governments. After the taxes are taken out of payroll, additional adjustments are made in the form of deductions, reimbursements, and garnishments. After the payroll is adjusted for the different components, the final total that the employee takes home is known as the net pay, or net amount, of the check. Net pay is the amount that the employee gets to keep for them and spend however they want.

In payroll, the gross pay is 'The Big Number' on an employee's paycheck. Gross pay, also known as gross income, is the total payment that an employee earns before any deductions or taxes are taken out. For employees that are hourly, gross pay is calculated when the rate of hourly pay is multiplied by the total number of regular hours worked. If the employee has overtime hours, these are multiplied by the overtime rate of pay, and the two amounts are added together. Also included in gross pay is any other type of earnings that an employee may have. These may include holiday pay, vacation or sick pay, bonuses, and any miscellaneous pay that the employee may receive. Gross pay for salaried employees is an amount that is set for every payroll unless there are bonuses or other pay involved.

Deductions and Reimbursements

There are two types of payroll deductions that are taken out of gross pay. The first type is mandatory deductions. These deductions are simply the taxes taken out. The other type of deductions is then considered to be voluntary deductions. There are a wide array of voluntary deductions that can be taken out of an employee's gross pay, some of which are taken out before taxes and some being taken out after taxes. Pre-tax deductions are deductions that are taken out of an employee's gross pay amount before it is subject to tax. Since pretax deductions are taken from an employee's gross pay before taxes are applied, it also reduces the amount of taxable income for that employee.

After-tax deductions are just as they seem. They are deductions that are taken out after taxes have been taken out of an employee's payroll. Unlike pre-taxed deductions, after-tax deductions do not reduce the amount of taxable income for an employee. Like pretax deductions, there are numerous forms of after-tax deductions that can be taken out of an employee's check each payroll.

Net Pay

Net pay is simply the total amount that an employee receives from their paycheck after all required and voluntary deductions are taken out.

The formula for determining an employee's net pay would be:

Net Pay = Gross Pay - Pretax Deductions - Taxes + Reimbursements - Garnishments - After-Tax Deductions

NOTES

Payroll is the total of all compensation a business must pay to its employees for a set period of time or on a given date. It is usually managed by the accounting or human resources department of a business; small-business payrolls may be handled directly by the owner or an associate.

Payroll Software Programs

In lieu of using specialized payroll services, some companies opt to rely on payroll software programs. Once the company purchases the software, there are no additional monthly fees. Software programs usually include printable tax forms and withholding tables. In addition to the financial savings, internal payroll systems help companies keep confidential financial information private. However, software programs can be time-consuming, which can pose a problem for companies with a small staff.

10.5 PRODUCTION PLANNING AND CONTROL

Production planning is the planning of production and manufacturing modules in a company or industry. It utilizes the resource allocation of activities of employees, materials and production capacity, in order to serve different customers.

Different types of production methods, such as single item manufacturing, batch production, mass production, continuous production etc. have their own type of production planning. Production planning can be combined with production control into production planning and control, or it can be combined with Enterprise Resource Planning (ERP).Production control is the activity of controlling the workflow in the production. It is partly complementary to production planning.

Production planning is the future of production. It can help in efficient manufacturing or setting up of a production site by facilitating required needs. A production plan is made periodically for a specific time period, called the planning horizon. It can comprise the following activities:

- Determination of the required product mix and factory load to satisfy customer's needs.
- Matching the required level of production to the existing resources.
- Scheduling and choosing the actual work to be started in the manufacturing facility.
- Setting up and delivering production orders to production facilities.

In order to develop production plans, the production planner or production planning department needs to work closely together with the marketing department

and sales department. They can provide sales forecasts, or a listing of customer orders. The work is usually selected from a variety of product types which may require different resources and serve different customers. Therefore, the selection must optimize customer-independent performance measures, such as cycle time and customer-dependent performance measures, such as on-time delivery.

A critical factor in production planning is 'The accurate estimation of the productive capacity of available resources, yet this is one of the most difficult tasks to perform well'. Production planning should always take into account material availability, resource availability and knowledge of future demand.

Production Planning and Control (PPC) is a predetermined process which includes the use of human resource, raw materials, machines, etc. PPC is the technique to plan each and every step in a long series of separate operation. It helps to take the right decision at the right time and at the right place to achieve maximum efficiency.

Objectives of Production Planning and Control (PPC)

Following are the objectives of Production Planning and Control (PPC):

1. To ensure safe and economical production process.
2. To effectively utilize plant to maximize productivity.
3. To maximize efficiency by proper coordination in production process.
4. To ensure proper delivery of goods.
5. To place the right man for the right job, at right time for right wages.
6. To minimize labour turnover.
7. To reduce the waiting time.

Main Elements of Production Planning and Control

The following are main elements of Production Planning and Control (PPC).

- Routing
- Loading
- Scheduling
- Dispatching
- Follow Up
- Inspection
- Corrective

Stages of Production Planning and Control (PPC)

Production Planning and Control (PPC) is done in following three stages:

Pre-Planning: In this phase of production planning, basic ground work on the product design, layout design and work flow are prepared. The operations relating

to the availability scope and capacity of men, money materials, machines, time are estimated.

Planning: In this phase complete analysis on routing, estimating and scheduling is done. It also tries to find out the areas of concern for short time scheduling and long-time scheduling so that prominent planning can be prepared.

Control: In this phase, the functions include dispatching, follow up, inspection and evaluation. It tries to analyse the expedition of work in progress. This is one of the important phases of the Production Planning and Control (PPC).

Production Planning and Control is, therefore, a significant and essential task of Production Manager, since it has to be determined that production process is properly decided in advance and it is carried out as per the plan. Production is related to the conversion of raw materials into finished goods. This conversion process involves a number of steps, such as deciding what to produce, how to produce, when to produce, etc. These decisions are a part, of production planning. Merely deciding about the task is not sufficient.

The entire process should be carried out in a best possible way and at the lowest cost. Production Manager has to target that the things proceed as per the plans. This is a control function and has to be carried as meticulously as planning. Both planning and control of production are essential to produce better quality goods at reasonable prices and in a most systematic manner.

Check Your Progress

3. What are the two types of payroll deductions?
4. What do you understand by production planning?

10.6 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. The aim of inventory control is to generate the maximum profit from the least amount of inventory investment without intruding upon customer satisfaction levels.
2. Fund accounting system is specifically designed for NGOs (Non-Governmental Organizations) and public agencies since it comes with fund accounting module for grant management, tracking donations, and managing GASB (Governmental Accounting Standards Board) regulations.
3. There are two types of payroll deductions that are taken out of gross pay. The first type is mandatory deductions. These deductions are simply the taxes taken out. The other type of deductions is then considered to be voluntary deductions.

4. Production planning is the planning of production and manufacturing modules in a company or industry. It utilizes the resource allocation of activities of employees, materials and production capacity, in order to serve different customers.

10.7 SUMMARY

- Information Technology (or IT as it is popularly called) has significant role in data processing, as the computerised data techniques provide accurate results and saves time. IT has dramatically changed the way of our life.
- Inventory control or stock control can be broadly defined as, 'The activity of checking a shop's stock'. It is the process of ensuring that the right amount of supply is available within a business.
- The inventory control system allows for companies to assess their current state concerning assets, account balances, and financial reports.
- Accounting or Accountancy is the measurement, processing and communication of financial and non-financial information about economic entities, such as businesses and corporations.
- Computerized inventory control system is also capable of preparing many other useful MIS (Management Information System) reports, such as aging analysis, goods movement analysis, slow and fast moving stock report, valuation report and it assists the storekeeper and accountants.
- Barcodes are a series of black and white parallel and adjacent bars with spaces which represent a string of character.
- Computerized accounting software is used for evaluating the profitability and financial performance of different organizations. Manual accounting has become redundant nowadays, as they are prone to human error.
- Computerized accounting software keeps track of the banking accounts of the company and also imports data from the respective bank into the system.
- Payroll can also refer to a company's records of payments that were previously made to employees, including salaries and wages, bonuses, and withheld taxes, or the company's department that deals with compensation.
- Production planning is the future of production. It can help in efficient manufacturing or setting up of a production site by facilitating required needs. A production plan is made periodically for a specific time period, called the planning horizon.

NOTES

10.8 KEY WORDS

- **Inventory:** It refers to the goods and materials that a business holds for the ultimate goal of resale, production or utilisation.
- **Accounting**: It is the measurement, processing and communication of financial and non-financial information about economic entities, such as businesses and corporations.
- **Paycheck:** It is traditionally a paper document (a cheque) issued by an employer to pay an employee for services rendered.

10.9 SELF-ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. Define the role of IT in data processing.
2. Elaborate on computerised accounting system.
3. Why computerised accounting software is used?
4. What are the objectives of production planning and control?
5. What are the main elements of production planning and control?

Long-Answer Questions

1. What is inventory? Explain the inventory control system giving examples.
2. What are some of the interesting IT-based innovations used to simplify and improve inventory management? Explain giving examples.
3. Explain the features of computerized accounting software.
4. What is payroll? Explain the process of calculating the net pay.
5. Discuss the stages of production planning and control.

10.10 FURTHER READINGS

Price, Wilson T. 1981. *Introduction to Computer Data Processing*, 3rd Edition. United States: Holt-Saunders International Editions.

Verzello, Robert J. and John Reutter. 1983. *Data Processing: Systems and Concepts*. New York: The McGraw-Hill Publishing Company, Inc.

Beech, Carol and Janice Burn. 1985. *Applications in Business Data Processing*. London (England): Pitman Publisher.

Fuori, William M. and Lawrence J. Aufiero. 1987. *Computers and Information Processing*. United States: Prentice Hall.

Sanders, Donald H. 1988. *Computers Today*. New York: McGraw-Hill.

Sinha, Pradeep K. and Priti Sinha. 2004. *Computer Fundamentals: Concepts, Systems & Applications*, 8th Edition. New Delhi: BPB Publications.

Rajaraman, V. 1999. *Fundamentals of Computers*. New Delhi: Prentice-Hall of India.

Lipschutz, Martin M. and S. Lipschultz. 1982. *Theory and Problems of Data Processing, Schaum's Outline Series*. New Delhi: Tata McGraw Hill Publishing Co. Ltd.

Stallings, William. 1996. *Data and Computer Communications*. New Jersey (USA): Prentice-Hall PTR.

NOTES

BLOCK - IV
ONLINE PROCESSING AND CONTROL

NOTES

UNIT 11 ONLINE APPLICATIONS

Structure

11.0 INTRODUCTION

Online processing is a process or method of transaction where companies can use an interface using Internet to receive and execute orders and handle payments from customers. Online processing is considered faster and it also provides updated continuous data to management.

In this unit, you will learn about the online processing controls and examples of specific online applications.

11.1 OBJECTIVES

After going through this unit, you will be able to:

- Define online processing
- Discuss the various examples of specific online applications

11.2 ONLINE PROCESSING CONTROL

Online processing is an automated way to enter and process data or reports continuously as use as the source documents are available. The online processing system continuously updates the entire accounting system.

Online Processing Definition: Online processing is the ongoing entry of transactions into a computer system in real time. The opposite of this system is batch processing, where transactions are allowed to pile up in a stack of documents, and are entered into the computer system in a batch.

NOTES

The term 'Online' processing is the term used to describe when a user is connected to a computer or network (they are logged in) and are processing the data files at the same time as using the input, output and storage devices. Content is stored in a database which is a mandatory requirement, especially if the user wishes to view, amend or upgrade that content.

Before the use of computers in business accounting systems, most of the companies were processing the data or reports in batches. For example, the invoices had to be collected, entered and then processed periodically by the employees. Batch processing was done either on daily basis or even weekly. This was referred as the most efficient and manual technique to process data, but it also provides outdated information since it is not updated regularly. At present the computers and mainframe servers are being used by almost all businesses, and hence most business moved from batch processing to online processing.

The online processing system has many significant advantages over the batch processing system. Online processing is considered faster and it also provides updated continuous data to management. For example, when the batch processing was used, then the inventory reports would have only been updated periodically. This specifies that management would only have useful inventory data on the day when a batch process was run. Since online processing updates the data continuously, hence the managers can run an inventory report any day at any time to get an accurate up-to-date information.

For example, the staff at warehouse uses the online processing for scanning the bar codes that are attached to the items in the warehouse in order to document the replacement of these items from one location to another location in the warehouse. This helps the staff who are looking and tracing the inventory for items, since they can rely upon this information for determining the current/present location of the inventory.

Another example of online processing is bar code scanning. When the customer purchases a shirt at any store, then the store staff scans the bar code and get it registered. The information about the sale of this shirt (source document) is updated immediately in the store inventory system as being 'Sold'. In addition, it is also updated in the cost and sales reports.

11.3 EXAMPLE OF SPECIFIC ONLINE APPLICATION

An 'On-Line' or 'Internet' application provides services for end users. It is generally not concerned with how data is actually transmitted between the hosts. Following are some distributed on-line or the Internet applications that require well-defined application level protocols:

- Sending and Receiving E-Mail

NOTES

- Searching and Browsing Information Archives
- Copying Files between Computers
- Conducting Financial Transactions
- Navigating (in Car, Smart Scooter, Smart Bike, etc.)
- Playing Interactive Games
- Video and Music Streaming
- Chat or Voice Communication (Direct Messaging, Video Conferencing)

Additionally, the following network services support on-line or the Internet applications to run efficiently:

- Name Servers
- Configuration Servers
- Mail Gateways, Transfer Agents, Relays
- File and Print Servers

All the on-line or Internet applications function over the exact equivalent transport layers. The Internet does not specify that how these applications should function or work, basically the on-line or the Internet protocol provides IP (Internet Protocol), TCP (Transmission Control Protocol) and UDP (User Datagram Protocol) for its efficiently working. TCP is a standard that defines how to establish and maintain a network conversation through which application programs can exchange data. TCP works with the IP, which defines how computers send packets of data to each other.UDP is a communications protocol that is primarily used for establishing low-latency and loss-tolerating connections between applications on the Internet. It speeds up transmissions by enabling the transfer of data before an agreement is provided by the receiving party.

Applications must essentially know the following:

(1) The IP address of the other party, i.e., on what host the other party is running — which is a network layer concept.

(2) The port number of the application running at the other end, because the other machine might be running multiple services— which is a transport layer concept.

These two essential information is passed to the transport layer to make the communication happen.

A **web application** (or **web app**) is application software that runs on a web server, unlike computer-based software programs that are run locally on the Operating System (OS) of the device. Web applications are accessed by the user through a web browser with an active network connection. These applications are programmed using a client–server modelled structure—the user ('Client') is provided services through an off-site server that is hosted by a third-party.

Examples of commonly-used web applications include web-mail, online retail sales, online banking, and online auctions.

Following are some of the on-line Internet applications that are widely used all over the world:

NOTES

1. **LiveTimer:** It is an online service used by professionals all over the world to track their time. People who want to track time for billing purposes, or to improve accountability or personal productivity are using LiveTimer.
2. **NomaDesk:** It is document collaboration software for geographically dispersed professionals who need constant and secure access to shared files. NomaDesk automatically syncs data across any computer – wherever, whenever required.
3. **Wrike:** It is an integrated online project management solution that helps to manage projects, teams and businesses.
4. **Tagzania:** It is about the locations or places. When signed up, one can add points, lanes or shapes to create and document the maps.
5. **Goowy:** It develops innovative online products and services that provide simple, intuitive tools for managing the digital life and expressing about the company or individual on the web.
6. **Tempo:** It is a cross between project management and time tracking, keeps track via SMS, Mobile Web, and even twitter.
7. **LogMeIn:** It provides organizations and individuals with secure, easy-to-use and cost effective solutions for remotely supporting, connecting and accessing digital information, applications and Internet-enabled devices. LogMeIn aims at improving mobility, business productivity and connectivity through on-demand software-as-a-service solutions.
8. **Zoho:** It is a suite of online business applications that are free for individuals and some have a subscription fee for organizations. The applications include CRM, Mail, project management, invoicing, and much more.
9. **vSocial:** It provides a platform that allows marketers to add social media to their promotions.
10. **iPhone VPN:** It connects your iPhone to your company Virtual Private Network with these handy instructions from Apple.

Check Your Progress

1. Give the definition of online processing.
2. What are the online or Internet application level protocols?
3. What is NomaDesk?

11.4 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

NOTES

1. Online processing is the ongoing entry of transactions into a computer system in real time. The opposite of this system is batch processing, where transactions are allowed to pile up in a stack of documents, and are entered into the computer system in a batch.
2. Following are some distributed on-line or the Internet applications that require well-defined application level protocols:
   - Sending and Receiving E-Mail
   - Searching and Browsing Information Archives
   - Copying Files between Computers
   - Conducting Financial Transactions
   - Navigating (in Car, Smart Scooter, Smart Bike, etc.)
   - Playing Interactive Games
   - Video and Music Streaming
   - Chat or Voice Communication (Direct Messaging, Video Conferencing)
3. NomaDesk is document collaboration software for geographically dispersed professionals who need constant and secure access to shared files. NomaDesk automatically syncs data across any computer – wherever, whenever required.

11.5 SUMMARY

- Online processing is the ongoing entry of transactions into a computer system in real time. The opposite of this system is batch processing, where transactions are allowed to pile up in a stack of documents, and are entered into the computer system in a batch.
- The online processing system has many significant advantages over the batch processing system. Online processing is considered faster and it also provides updated continuous data to management.
- 'Internet' application provides services for end users. It is generally not concerned with how data is actually transmitted between the hosts.
- The IP address of the other party, i.e., on what host the other party is running — which is a network layer concept.
- The port number of the application running at the other end, because the other machine might be running multiple services— which is a transport layer concept.

NOTES

- A web application (or web app) is application software that runs on a web server, unlike computer-based software programs that are run locally on the Operating System (OS) of the device. Web applications are accessed by the user through a web browser with an active network connection.
- LiveTimer is an online service used by professionals all over the world to track their time. People who want to track time for billing purposes or to improve accountability or personal productivity are using LiveTimer.

11.6 KEY WORDS

- **Online processing**: It is the ongoing entry of transactions into a computer system in real time. The opposite of this system is batch processing, where transactions are allowed to pile up in a stack of documents, and are entered into the computer system in a batch.
- **LiveTimer**: It is an online service used by professionals all over the world to track their time. People who want to track time for billing purposes, or to improve accountability or personal productivity are using LiveTimer.
- **Tempo**: It is a cross between project management and time tracking, keeps track via SMS, Mobile Web, and even twitter.

11.7 SELF-ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. What is online processing?
2. State the advantages of online and batch processing.
3. Name some online Internet applications.
4. What is web application?

Long-Answer Questions

1. Explain online processing with the help of examples.
2. Discuss about the online Internet applications.

11.8 FURTHER READINGS

Price, Wilson T. 1981. *Introduction to Computer Data Processing*, 3rd Edition. United States: Holt-Saunders International Editions.

Verzello, Robert J. and John Reutter. 1983. *Data Processing: Systems and Concepts*. New York: The McGraw-Hill Publishing Company, Inc.

NOTES

Beech, Carol and Janice Burn. 1985. *Applications in Business Data Processing*. London (England): Pitman Publisher.

Fuori, William M. and Lawrence J. Aufiero. 1987. *Computers and Information Processing*. United States: Prentice Hall.

Sanders, Donald H. 1988. *Computers Today*. New York: McGraw-Hill.

Sinha, Pradeep K. and Priti Sinha. 2004. *Computer Fundamentals: Concepts, Systems & Applications*, 8th Edition. New Delhi: BPB Publications.

Rajaraman, V. 1999. *Fundamentals of Computers*. New Delhi: Prentice-Hall of India.

Lipschutz, Martin M. and S. Lipschultz. 1982. *Theory and Problems of Data Processing, Schaum's Outline Series*. New Delhi: Tata McGraw Hill Publishing Co. Ltd.

Stallings, William. 1996. *Data and Computer Communications*. New Jersey (USA): Prentice-Hall PTR.

UNIT 12 ONLINE RESERVATION SYSTEMS

NOTES

Structure

12.0 INTRODUCTION

Airline Reservation Systems (ARS) are part of the so-called Passenger Service

Systems (PSS), which are applications supporting the direct contact with the passenger. The airline business is one of the largest users of computers. Computers have been deployed in almost all aspects of the airline business for increasing revenues, reducing cost and enhancing customer satisfaction.

Indian Railways (IR) is a government agency under the ownership of Ministry of Railways that operates India's national railway system. Rail transport is one of the important modes of transport in India. The railway database also plays significant role in the smooth functioning of system. The Online Railway Ticket Booking System helps the passenger to reserve or book the railway tickets online for travelling from a particular source to the destination,

In this unit, you will learn about the airline reservation and railway reservation.

12.1 OBJECTIVES

After going through this unit, you will be able to:

- Define how airline reservation is done
- Discuss about the railway reservation

12.2 AIRLINE RESERVATION

NOTES

Airline Reservation Systems (ARS) are part of the so-called Passenger Service Systems (PSS), which are applications supporting the direct contact with the passenger.

ARS eventually evolved into the Computer Reservations System (CRS). A computer reservation system is used for the reservations of a particular airline and interfaces with a Global Distribution System (GDS) which supports travel agencies and other distribution channels in making reservations for most major airlines in a single system.

Airline reservation systems incorporate airline schedules, fare tariffs, passenger reservations and ticket records. An airline's direct distribution works within their own reservation system, as well as pushing out information to the GDS. The second type of direct distribution channel are consumers who use the Internet or mobile applications to make their own reservations. Travel agencies and other indirect distribution channels access the same GDS as those accessed by the airline reservation systems, and all messaging is transmitted by a standardized messaging system that functions on two types of messaging that transmit on SITA's High Level Network (HLN).

Users access an airline's inventory through an availability display. It contains all offered flights for a particular city-pair with their available seats in the different booking classes. This display contains flights which are operated by the airline itself as well as code share flights which are operated in co-operation with another airline. If the city pair is not one on which the airline offers service, it may display a connection using its own flights or display the flights of other airlines. The availability of seats of other airlines is updated through standard industry interfaces. Depending on the type of co-operation, it supports access to the last seat (last seat availability) in real-time. Reservations for individual passengers or groups are stored in a so-called Passenger Name Record (PNR).

The airline business is one of the largest users of computers. Computers have been deployed in almost all aspects of the airline business for increasing revenues, reducing cost and enhancing customer satisfaction.

Imagine the trouble airline companies will have booking air tickets across thousands of booking counters spread all over the world without computerized booking systems and interconnectivity of these systems. The airline industry is using a vast web-based online system that can be accessed by anyone from anywhere in the world. Some of the major IT initiatives taken by the airline industry are as follows:

NOTES

Online Ticket Booking through the Internet: Almost all the airline companies, may it be domestic or international are selling air tickets online. Air tickets can be booked online by paying through credit card and e-tickets can be printed on your printer.

Flight and Seat Availability Information: Flight and seat availability information along with the cost of the ticket is now easily available online with an option to compare it with other available airlines, making it easier to choose the airlines according to the time and price that best suits an individual. The facility of choosing the seat position and the meals (veg or non-veg) if offered, is also available. yatra.com, makemytrip.com and travelguru.com are a few websites available for online booking other than the official website of each airline.

Last Minute Deals and Auctions: To recover the lost revenue opportunity on unsold seats, most airlines have started bidding for last minute tickets in online auction. Several specific airline ticket auctions sites like razorfinish.com are also available. This option is beneficial both for the passenger as well as the airline company since the airline company gets some revenue on un-booked seats and the passenger gets a good deal by paying much less than the regular price.

All these facilities would not have been possible for airline companies without the use of computers.

12.3 RAILWAY RESERVATION

Indian Railways (IR) is a government agency under the ownership of Ministry of Railways that operates India's national railway system. Rail transport is one of the important modes of transport in India. The railway database also plays significant role in the smooth functioning of system. The Online Railway Ticket Booking System helps the passenger to reserve or book the railway tickets online for travelling from a particular source to the destination, hence the passenger do not have to wait in long queues to get the ticket.

The Tatkal Scheme is a ticketing program established by Indian Railways. The scheme is used for booking journeys at very short notice. The Indian Railways introduced it in all forms of reserved classes on almost all trains in India. It was introduced in 1997, when Nitish Kumar was the Railway Minister of India. Bookings can be made online and offline. Online tickets can be booked on the Internet at IRCTC (Indian Railway Catering and Tourism Corporation).

IRCTC has revolutionizing train travel system in India. It pioneered Internet-based rail ticket booking through its website, as well as from the mobile phones via WiFi, GPRS or SMS. It also provides SMS facility to check PNR (Passenger Name Record) status and 'Live Train Status' as well. In addition to e-tickets,

Indian Railways Catering and Tourism Corporation also offers I-tickets that are basically like regular tickets except that they are booked online and delivered by post. The tickets PNR status is also made available. Commuters on the suburban rail can also book season tickets through the website. It has also launched a loyalty program called ShubhYatra for frequent travellers. Through this program, passengers can avail discounts on all tickets booked round the year by paying an upfront annual fee.

The railroad industry is the backbone of the country's economy. So for better economic growth, the rail network has to be effective, efficient and timely. In case of the Indian Railways, which is one of the largest rail networks of the world, it has to manage 11000 trains everyday covering around 108,706 kms, connecting 6,853 stations and a 1.54 million workforce. Since it was very difficult and complex to manage and operate on such a large scale, Indian Railways decided to go for computerization to simplify their operations and have a better control over the management. Some of the major IT initiatives taken by Indian Railways are as follows:

- All India centralized reservation system provides the facility for the passenger to book tickets from any destination and is one of the most successful examples of computerization in the country.
- IRCTC is an online railway ticket booking portal which enables the passenger to book railway tickets for any destination in India from anywhere in the world. It also provides very useful information like computerized reservation-related enquiries about passenger status, train schedule and trains between pairs of stations. Booking a railway ticket is now easy with an option to print your tickets from your printer or get it couriered within twenty-four hours. Apart from this, the railway timetables, network maps and freight charges are also available on the Indian railways website.
- Indian Railways has also launched the online computerized system for unreserved passengers. This service caters to almost 14 to 15 million people travelling with unreserved tickets. These unreserved tickets being available from locations other than boarding stations reduce the long queues and chaos at the railway station ticket counters.
- Indian Railways has also started a pilot project for issuing periodical season tickets through the ATM (Automated Teller Machines) and another pilot project for buying tickets through smart cards.
- The Interactive Voice Response System (IVRS) has also been introduced to update passengers with railway inquiry and other related information.

This national enquiry system is capable of providing train running positions in a real-time system through the IVRS and other output devices.

- Indian Railways (IR) is a government agency under the ownership of Ministry of Railways that operates India's national railway system. Rail transport is one of the important modes of transport in India. The railway database also plays significant role in the smooth functioning of system.

Check Your Progress

1. Explain the term airline reservation.
2. What is tatkal scheme?

12.4 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. Airline Reservation Systems (ARS) are part of the so-called Passenger Service Systems (PSS), which are applications supporting the direct contact with the passenger.
2. The Tatkal Scheme is a ticketing program established by Indian Railways. The scheme is used for booking journeys at very short notice.

12.5 SUMMARY

- Travel agencies and other indirect distribution channels access the same GDS as those accessed by the airline reservation systems, and all messaging is transmitted by a standardized messaging system that functions on two types of messaging that transmit on SITA's High Level Network (HLN).
- The airline business is one of the largest users of computers. Computers have been deployed in almost all aspects of the airline business for increasing revenues, reducing cost and enhancing customer satisfaction.
- Airline Reservation Systems (ARS) are part of the so-called Passenger Service Systems (PSS), which are applications supporting the direct contact with the passenger.
- Airline reservation systems incorporate airline schedules, fare tariffs, passenger reservations and ticket records. An airline's direct distribution works within their own reservation system, as well as pushing out information to the GDS. The second types of direct distribution channel are consumers who use the Internet or mobile applications to make their own reservations.

NOTES

- Flight and seat availability information along with the cost of the ticket is now easily available online with an option to compare it with other available airlines, making it easier to choose the airlines according to the time and price that best suits an individual.
- Indian Railways has also launched the online computerized system for unreserved passengers. This service caters to almost 14 to 15 million people travelling with unreserved tickets.

12.6 KEY WORDS

- **Airline Reservation Systems (ARS):** Airline Reservation Systems (ARS) are part of the so-called Passenger Service Systems (PSS), which are applications supporting the direct contact with the passenger.
- **Indian Railways (IR):** Indian Railways (IR) is a government agency under the ownership of Ministry of Railways that operates India's national railway system.

12.7 SELF-ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. Explain about the flight and seat availability information.
2. Define some of the major IT initiatives for railway reservation.

Long-Answer Questions

1. Briefly explain the airline reservation with the help of examples.
2. Discuss about the railway reservation giving appropriate examples.

12.8 FURTHER READINGS

Price, Wilson T. 1981. *Introduction to Computer Data Processing*, 3rd Edition. United States: Holt-Saunders International Editions.

Verzello, Robert J. and John Reutter. 1983. *Data Processing: Systems and Concepts*. New York: The McGraw-Hill Publishing Company, Inc.

Beech, Carol and Janice Burn. 1985. *Applications in Business Data Processing*. London (England): Pitman Publisher.

Fuori, William M. and Lawrence J. Aufiero. 1987. *Computers and Information Processing*. United States: Prentice Hall.

Sanders, Donald H. 1988. *Computers Today*. New York: McGraw-Hill.

Sinha, Pradeep K. and Priti Sinha. 2004. *Computer Fundamentals: Concepts, Systems & Applications*, 8th Edition. New Delhi: BPB Publications.

Rajaraman, V. 1999. *Fundamentals of Computers*. New Delhi: Prentice-Hall of India.

Lipschutz, Martin M. and S. Lipschultz. 1982. *Theory and Problems of Data Processing, Schaum's Outline Series*. New Delhi: Tata McGraw Hill Publishing Co. Ltd.

Stallings, William. 1996. *Data and Computer Communications*. New Jersey (USA): Prentice-Hall PTR.

NOTES

NOTES

BLOCK - V
MANAGEMENT OF BUSINESS APPLICATIONS

UNIT 13 STORE MANAGEMENT AND QUERY PACKAGE

Structure

13.0 INTRODUCTION

Store is an important component of material management since it is a place that keeps the materials in a way by which the materials are well accounted for, are maintained safe, and are available at the time of requirement. Storage is an essential and most vital part of the economic cycle and store management is a specialized function, which can contribute significantly to the overall efficiency and effectiveness of the materials function. Literally store refers to the place where materials are kept under supervision. Typically a store has a few processes and a space for storage. The supplementary process of store is the stock control also known as inventory control. In a manufacturing organization, this process of receiving, keeping in custody, and issuing forms a cyclic process which runs on a continuous basis. The organizational set up of the store depends upon the requirements of the organization and is to be tailor made to meet the specific needs of the organization.

The main objective of the store is to provide essential inputs to the production and operations departments at a minimum cost. Store functions concern receiving, movement, storage and issue of items, such as raw materials, bought out parts, tools, spares, consumables, etc., required for production, maintenance and operation of the plant and finished goods until is dispatch to customers. Store, therefore, is the custodian of all goods that are received in the company until they are consumed or sold and naturally it assumes the responsibility of receiving, storage, preservation, issue, and accounting functions.

Stores management play a vital role in the operations of company. It is in direct touch with the user departments in its day-to-day activities. The most important purpose served by the stores is to provide uninterrupted service to the manufacturing divisions. Further, stores are often equated directly with money, as money is locked up in the stores. Traditional database queries follow a simple model, they define constraints that each tuple in the result must satisfy. The package queries is a new query model that extends traditional database queries to handle complex constraints and preferences over answer sets.

In this unit, you will learn about the management of stores, and query packages.

13.1 OBJECTIVES

After going through this unit, you will be able to:

- Understand the significance of management of stores
- Explain about the store functions and materials management
- Elaborate on the query and packages

13.2 MANAGEMENT OF STORES

Store management refers to the efficient management of materials. Store management is concerned with ensuring that all the activities involved in storekeeping are carried out efficiently and economically by those employed in the store.

Stores form the basis of material management. Stores play a vital role in the operations of a company. In an organization, stores are mainly intended to provide staff activity in the production of goods or services. No industrial unit or public undertaking of any size can be managed efficiently without it. The basic objective of storekeeping is to provide services to the operating functions in the most economical manner.

Fundamentally, store is an important component of material management since it is a place that keeps the materials in a way by which the materials are well accounted for, are maintained safe, and are available at the time of requirement. Storage is an essential and most vital part of the economic cycle and store management is a specialized function, which can contribute significantly to the overall efficiency and effectiveness of the materials function. Literally store refers to the place where materials are kept under supervision. Typically a store has a few processes and a space for storage. The supplementary process of store is the stock control also known as inventory control. In a manufacturing organization, this process of receiving, keeping in custody, and issuing forms a cyclic process which runs on a continuous basis. The organizational set up of the store depends

NOTES

upon the requirements of the organization and is to be tailor made to meet the specific needs of the organization.

The main objective of the store is to provide essential inputs to the production and operations departments at a minimum cost. Store functions concern receiving, movement, storage and issue of items, such as raw materials, bought out parts, tools, spares, consumables, etc., required for production, maintenance and operation of the plant and finished goods until is dispatch to customers. Store, therefore, is the custodian of all goods that are received in the company until they are consumed or sold and naturally it assumes the responsibility of receiving, storage, preservation, issue and accounting functions.

Stores management also play a vital role in the operations of company. It is in direct touch with the user departments in its day-to-day activities. The most important purpose served by the stores is to provide uninterrupted service to the manufacturing divisions. Further, stores are often equated directly with money, as money is locked up in the stores.

According to **Afford and Beatty**, "Store management is that aspect of material control concerned with the physical storage of goods".

According to **Maynard**, "Store management is to receive materials, to protect them while in storage from damage and unauthorized removal, to issue the materials in the right quantities, at the right time to the right place and to provide these services promptly and at least cost".

Objectives of Store Management

Following are the key objectives of efficient store management:

- Minimizing the cost of production by minimizing the cost of materials.
- Reducing the cost of storage.
- Maintaining the value of materials.
- Providing services to user departments.
- Establishing coordination with other departments.
- Safeguarding uninterrupted supply of materials without delay to various users of the organization.
- Preventing overstocking and under stocking of the materials.
- Ensuring safe handling of materials and preventing their damage.
- Protecting materials from pilferage, theft, fire and other risks.
- Confirming proper and continuous control over the materials.
- Guaranteeing most effective utilization of available storage space.
- Optimizing the efficiency of the personnel engaged in the store.

NOTES

Functions of Store Department

The store department performs the following functions:

- **Receipt:** It receives and accounts for inventories.
- **Storage:** It stores and preserves the inventories protecting them from damage, pilferage and deterioration.
- **Retrieval:** It helps easy access to materials and ensures optimum space utilization. Materials can be located and retrieved with ease.
- **Issue:** It satisfies the demands of consuming departments by the proper issue of inventories on receiving the requisitions.
- **Records:** It keeps proper records of the issue and receipts.
- **HouseKeeping:** Space is kept neat and clean so that material handling, preservation, storage, issue and receipt is done satisfactorily.
- **Surplus Stock:** Scrap and surplus disposal management is a function of stores.
- **Verification:** Physical verification and purchase initiation to avoid stock-outs.
- **Coordination and Cooperation:** To interface with the production and inspection department.

Classification of Stores

Store can be of permanent nature or of temporary nature. Stores are basically classified in the following broad categories.

Functional Stores: Functional stores are named based on the function of the materials stored. Examples are fuels store, chemicals store, tools store, raw materials store, spare parts store, equipment store, refractories store, electric store, explosives store, finished goods store, etc.

Physical Stores: Physical stores can be centralized stores or decentralized stores. These stores are named based on the size and location of the store. Examples are central store, sub-store, department store, site store, transit stores, receipt store, intermediate store, open yard store, covered store, etc.

Specific Stores: Stores are also classified by naming them after the departments to which they serve. Examples are construction stores, operation stores, rolling mill stores, blast furnace stores, steel melting shop stores, etc.

Supplementary Stores: Stores are sometimes classified based on the nature of materials stored in them. Examples are general store, bonded store, perishable store, inflammable store, salvage store, reject store, quarantine store, etc.

NOTES

Types of Stores

Main or Centralized Stores: A central store is generally a 'Wholesale' supplier to other units, departments, or sub-stores that operate on a retail basis issuing goods directly to users. All material is received and issued by one central store.

Branch or Decentralized Stores: Decentralized or branch stores are provided in considerably large plants and where one main store cannot meet the requirements of the plant without a waste of time and inconvenience.

Central Store with Sub-Stores: A very big factory having a large number of product lines may have this type of storage system. It has a main store that can serve as a base with sub-stores for each unit of production, preferably located as near the unit as possible. The sub-stores draw their requirements from the main store for a certain period, say, a fortnight or a month. This fixed quantity of material to the particular department is known as Float. After the completion of the determined period, the storekeeper of the sub-store will describe the material consumed and will issue the quantity of material equal to the material consumed to bring the level to the replenishment level. This system of issuing and controlling materials is known as a periodic system of store control.

Warehouse: Warehouses are the godowns that take the responsibility of keeping and storing goods and providing ancillary services to help the small and medium-size traders and manufacturers who, because of technical and economic reasons, may not like to have their storehouses. These warehouses undertake to preserve the goods scientifically and systematically to maintain their original value, quality, and usefulness. They charge a certain prescribed rent at a fixed rate in advance.

Tools and Miscellaneous Stores: Tools and miscellaneous stores are equipped with all the necessary tools needed by the productions and other shops. The stock of tools must be maintained with due regard to the requirements of the work. This store is responsible for issuing tools, spare parts, and other accessories to different departments.

Processes of Stores

Typically a store management inccludes a few key processes and a space for material storage. The main processes of store are,

- To receive the incoming materials (receiving)
- To keep the materials provided that they are required for use (keeping in custody)
- To move them out of store for use (issuing)

The auxiliary process of store is the stock control also known as inventory control. In a manufacturing organization, this process of receiving, keeping in custody, and issuing forms a cyclic process which runs on a continuous basis. The organizational set up of the store depends upon the requirements of the organization and is to be tailor made to meet the specific needs of the organization.

Store follows certain activities which are managed through use of various resources. Store management ensures that all the activities involved in storekeeping and stock control are carried out efficiently and economically by the store personnel.

The basic responsibilities of store are to act as custodian and controlling agent for the materials to be stored, and to provide service to users of these materials. Proper management of store systems provide flexibility to absorb the shock variation in demand and enable purchasing to plan ahead.

Since the materials have a cost, the organization is to manage the materials in store in such a way so that the total cost of maintaining materials remains optimum.

Store needs a secured space for storage. It needs a proper layout along with handling and material movement facilities, such as cranes, forklifts, etc., for safe and systematic handling as well as stocking of the materials in the store with an easy traceability and access.

Store is to preserve the stored materials and carry out their conservation as required to prevent deterioration in their qualities. Also store is to ensure the safety of all items and materials whilst in the store, i.e., protecting them from pilferage, theft, damage, deterioration, and fire.

Materials Management

Materials management is a core supply chain function and includes supply chain planning and supply chain execution capabilities. Specifically, materials management is the capability firms use to plan total material requirements. The material requirements are communicated to procurement and other functions for sourcing. Materials management is also responsible for determining the amount of material to be deployed at each stocking location across the supply chain, establishing material replenishment plans, determining inventory levels to hold for each type of inventory (raw material, WIP, finished goods), and communicating information regarding material needs throughout the extended supply chain.

Typical roles in materials management include Materials Manager, Inventory Control Manager, Inventory Analyst, Material Planner, Expediter and emerging hybrid roles like 'Buyer Planner'.

The primary business objective of materials management is assured supply of material, optimum inventory levels and minimum deviation between planned and actual results.

The goal of materials management is to provide an unbroken chain of components for production to manufacture goods on time for customers. The materials department is charged with releasing materials to a supply base, ensuring that the materials are delivered on time to the company using the correct carrier. Materials is generally measured by accomplishing on time delivery to the customer, on time delivery from the supply base, attaining a freight, budget, inventory shrink management, and inventory accuracy. The materials department is also charged with the responsibility of managing new launches.

NOTES

In some companies materials management is also charged with the procurement of materials by establishing and managing a supply base. In other companies the procurement and management of the supply base is the responsibility of a separate purchasing department. The purchasing department is then responsible for the purchased price variances from the supply base.

In large companies with multitudes of customer changes to the final product there may be a separate logistics department that is responsible for all new acquisition launches and customer changes. This logistics department ensures that the launch materials are procured for production and then transfers the responsibility to the plant materials management.

Check Your Progress

1. Explain the term store management.
2. What are the main objective and functions of the store?
3. Define the term store management as per Afford and Beatty, and Maynard.
4. How the stores are broadly classified?
5. What are the main processes of store?

13.3 QUERY PACKAGE

A query is a request to retrieve data from the respective 'Store'. The request is in the form of an **SQL SELECT** statement. When you run a query, the query SQL statement is sent to the database that contains the Store details, and the database returns the results of the query. A '**Package**' is a public view of the results of the query. Following are the different types of queries.

Generic Queries

A generic query is a type of query that is defined by using a query wizard and building blocks. No SQL knowledge is required. The **Queries Tool** generates an **SQL SELECT** statement from the selected building blocks. The generated SELECT statement works with all supported databases.

Custom Queries

A custom query is a type of query where you can define your own SQL SELECT statement. When you want to use database-specific SQL syntax and grammar, create a custom query.

Query Groups

A query group is a user-defined container for queries. Use query groups to organize your queries so that they are easier to find and run.

13.3.1 Packages

A '**Package**' is referred as a public view of the results of a query. Data stewards use packages with the **Data Manager** and **Merge Manager** tools. You can also use packages with external applications.

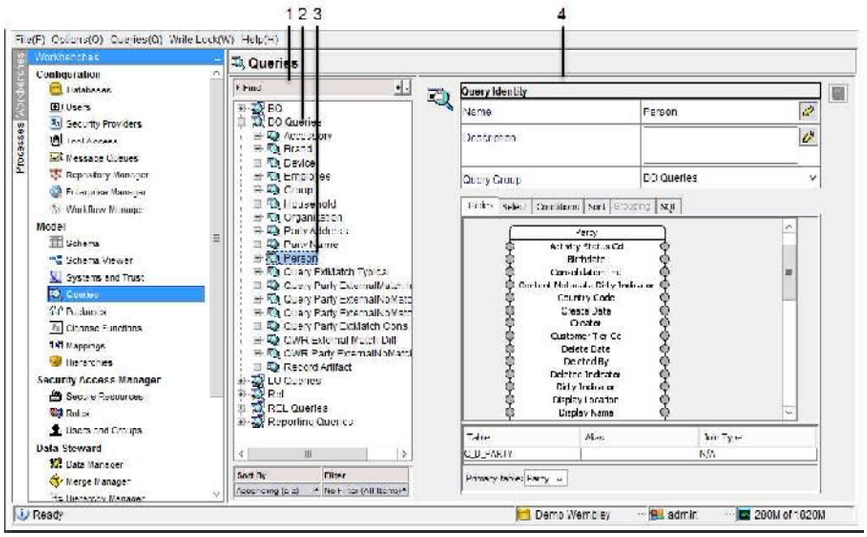
NOTES

Queries Tool

Use the **Queries Tool** for adding, editing and deleting generic queries, custom queries, and query groups. You can also view the results of a query or view the packages that depend on the query.

The following illustration shows the Queries Tool with a Query selected. In the illustration,

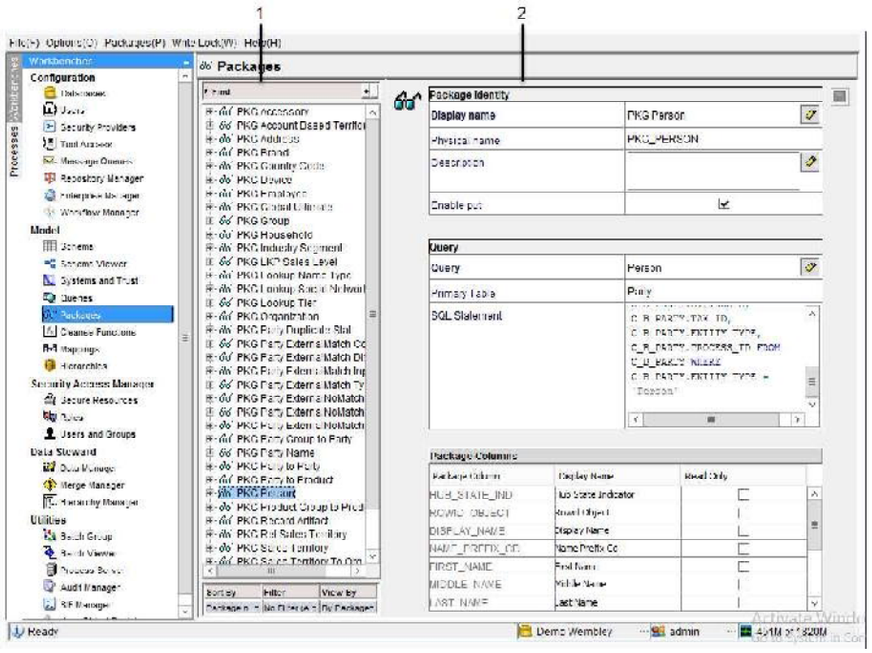
1. Navigation Pane. Contains user-defined query groups and queries.
2. Query Group.
3. Query.
4. Properties Pane. Contains the properties for the selected query group or query.



Packages Tool

The **Packages Tool** is used for adding, editing and deleting packages. The following illustration shows the Packages tool with a package selected. In the illustration,

1. Navigation Pane. Contains user-defined packages.
2. Properties Pane. Contains the properties for the selected package.



Query Groups

A **query group** is a **user-defined** container for queries. Use query groups to organize your queries so that they are easier to find and use. For example, you can create separate groups for the following types of queries:

- Generic queries that are suitable for use in update pacakges.
- Queries that select from more than one base object table.
- Queries that select from lookup tables.

Adding a Query Group: Use query groups to organize your queries.

- In the Model workbench, click **Queries**.
- Acquire a write lock.
- In the navigation pane, right-click and click **New Query Group**.
- In the **Add Query Group** window, enter a name for the query group and, optionally, enter a description.
- Click **OK**.

  The query group appears in the navigation pane in alphabetical order.

Generic Queries

The **generic queries** can be built by using the building blocks in the **Queries Tool.** You do not need to know SQL to create generic queries.

The building blocks that you select specify the criteria to use to retrieve data, including table names, column names, and a set of conditions. Queries can

also include instructions about how to sort the results and group the results. The SQL statement works with any database.

Adding a Generic Query: Use the building blocks to construct a query that can be understood by any of the supported databases for adding a generic query.

If you want to code an SQL statement, add a custom query instead.

- In the Model workbench, click **Queries**.
- Acquire a write lock.
- Optionally, in the navigation pane, select the query group to which you want to add the query.
- Right-click in the navigation pane and click **New Query**.
- The **New Query Wizard** opens.
- If you see a **Welcome** screen, click **Next**.

Specify the following general query properties, and click **Next**:

Property	Description
Query Name	Type a descriptive name for the query.
Description	Optionally, type a description for the query.
Query Group	Optionally, select a different query group.
Select Primary Table	Select the table from which you want to retrieve data.

To retrieve a subset of columns, select the columns.

- In the **Select query columns** screen, select the columns to include and clear the check boxes for all other columns.
- If you intend to use the query in a PUT-enabled package, select the Object column.
- The Object is a required column for PUT-enabled packages.
- Click **Finish.**
- The query appears within the selected query group.
- To view the results of the query, in the navigation pane, expand the query and click **View.**
- The **Queries tool** displays the results of the query.

To further refine the query, edit the query and use the query building blocks.

Packages

A **package** is a public view of one or more queries. You use packages with the data **steward** tools in the **Core Console** and with external applications that use services to access master data.

NOTES

You can create two kinds of packages: **Display Packages** and **Update Packages**.

Display packages define read-only views of master data.

Update packages define views from which authorized users can make changes to master data.

To restrict a package to authorized users, make the package a secure resource. You can manage all secure resources from the **Roles Tool**.

Display Packages

If you want users to view the query results from the data steward tools or from an external application, create display packages. Users can view the query results, but cannot make changes to the data.

For each authorized role, set read privileges for the display packages that the role can use. Do not enable any other privileges.

Update Packages

An update package is also known as a PUT-enabled package, a putable package or a merge package. When you create an update package, you basically enable the PUT API within the package. A data steward can use the update package to add, change or merge master data.

Create update packages when you want authorized users to be able to perform any of the following actions:

- **Update** data in records from the **Data Manager** or from an external application.
- **Insert** records from the **Data Manager** or from an external application.
- **Merge** records from the **Merge Manager** or from an external application.

For each authorized role, set privileges for the update packages that the role can use. Grant the read privilege and at least one of the create, update or merge privileges.

Requirements for Update Packages

Update packages have some specific requirements at the package level and at the query level.

Package Level Requirements

An update package must adhere to the following requirements:

- Select a generic query for the update package. Custom queries and other packages are not supported with update packages.

- Select the **Enable Put** option.
- Access only the tables and relationships defined within the generic query. The update package cannot contain joins to other tables.

NOTES

Query Level Requirements

The generic query must adhere to the following requirements:

- Select only a primary table. The generic query cannot contain additional tables.
- If you select individual columns, you must add the ROWID_OBJECT column, where ROWID_OBJECT is the column name.
- The generic query cannot contain system tables, constant columns, aggregate functions or groupings.

Adding a Package

You can **create** either a **display package** or an **update package**.

To create an update package, before you begin, create a generic query that meets the requirements for update packages.

- In the Model workbench, click **Packages.**
- Acquire a write lock.
- Right-click in the navigation pane and click **New Package**.
- The **New Package Wizard** opens.
- If you see a **Welcome** screen, click **Next**.
- Set the following properties:

Properties	Description
Display Name	Type a descriptive name for the package. This name appears in the navigation pane.
Physical Name	Optionally, change the suggested physical name. The wizard suggests a physical name that is based on the display name you specified.
Description	Optionally, type a description for the query.
Query Group	Optionally, select a different query group.
Enable PUT	To create an update package, select this option. To create a display package, clear this option.
Secure Resource	To restrict who can use the package, select this option. You use the Roles tool to assign user roles to secured packages.

- Click **Next**.
- The **New Package Wizard** displays the **Select Query** dialog box.

For update packages, you must select a generic query and for display packages, you can select a generic query or a custom query.

NOTES

- Select the query that you want to use in the package.
  1. To use an existing query, select the query from the list.
  2. To create a query, click **New Query** and create the query.
  3. To create a query group, click **New Query Group** and create the query group.
- Click **Finish**.
- To preview the package results, in the navigation pane, expand the package and click **View**.
- The **Packages** tool displays a preview of the package.

If the package fails to generate the selected custom query, then authenticate that the custom query adheres to the SQL syntax constraints.

Editing a Package

- You can change package properties and the underlying query.
- In the Model workbench, click **Packages**.
- Acquire a write lock.
- In the navigation pane, select the package.
- The properties appear in the properties pane.
- To edit a text property, click its **Edit** icon, edit the text, and click the **Accept Edit** icon.
- Click the **Save** icon.
- Optionally, you can edit the query for this package.
  1. In the navigation pane, expand the package.
  2. Click **Query**.
  3. **Edit** the query.
- To preview the package results, in the navigation pane, expand the package and click **View**.
- The **Packages** tool displays a preview of the package.

Refreshing a Package After Changing Queries

If you change a query, **refresh all packages** that use the query.

After a refresh, if a package is not in synch with the query, then select or clear columns to match the query.

- In the Model workbench, click **Packages.**
- Acquire a write lock.
- In the navigation pane, expand the package.
- Click **Refresh**.

Specifying Join Queries

You can create a package that permits data stewards to view base object information, along with information from the other tables, in the Data Manager or Merge Manager.

- Create an update package that queries a base object.
- Create a query to join the update package with the other tables.
- Create a display package based on the query you just created.

Deleting a Package

When the created package is no longer required, then delete the package to remove the dependency from the underlying query.

- In the Model workbench, click **Packages**.
- Acquire a write lock.
- In the **navigation pane**, right-click the **package** and click **Delete Package**.

Check Your Progress

6. Explain the terms query and package.
7. What is generic query?
8. How a package is deleted?

13.4 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. Store management refers to the efficient management of materials. Store management is concerned with ensuring that all the activities involved in storekeeping are carried out efficiently and economically by those employed in the store. Stores form the basis of material management and play a vital role in the operations of a company. In an organization, stores are mainly intended to provide staff activity in the production of goods or services.
2. The main objective of the store is to provide essential inputs to the production and operations departments at a minimum cost. Store functions concern receiving, movement, storage and issue of items, such as raw materials,

NOTES

bought out parts, tools, spares, consumables, etc., required for production, maintenance and operation of the plant and finished goods until is dispatch to customers. Store, therefore, is the custodian of all goods that are received in the company until they are consumed or sold and naturally it assumes the responsibility of receiving, storage, preservation, issue and accounting functions.

3. Store management according to Afford and Beatty, and Maynard is, According to Afford and Beatty, "Store management is that aspect of material control concerned with the physical storage of goods". According to Maynard, "Store management is to receive materials, to protect them while in storage from damage and unauthorized removal, to issue the materials in the right quantities, at the right time to the right place and to provide these services promptly and at least cost".
4. Stores are basically classified in the following broad categories.
   - Functional Stores: Functional stores are named based on the function of the materials stored. Examples are fuels store, chemicals store, tools store, raw materials store, spare parts store, equipment store, refractories store, electric store, explosives store, finished goods store, etc.
   - Physical Stores: Physical stores can be centralized stores or decentralized stores. These stores are named based on the size and location of the store. Examples are central store, sub-store, department store, site store, transit stores, receipt store, intermediate store, open yard store, covered store, etc.
   - Specific Stores: Stores are also classified by naming them after the departments to which they serve. Examples are construction stores, operation stores, rolling mill stores, blast furnace stores, steel melting shop stores, etc.
   - Supplementary Stores: Stores are sometimes classified based on the nature of materials stored in them. Examples are general store, bonded store, perishable store, inflammable store, salvage store, reject store, quarantine store, etc.
5. The main processes of store are,
   - To receive the incoming materials (receiving)
   - To keep the materials provided that they are required for use (keeping in custody)
   - To move them out of store for use (issuing)
6. A query is a request to retrieve data from the respective 'Store'. The request is in the form of an SQL SELECT statement. When you run a query, the query SQL statement is sent to the database that contains the Store details, and the database returns the results of the query. A 'Package' is a public

view of the results of a query. Data stewards use packages with the Data Manager and Merge Manager tools. You can also use packages with external applications.

7. A generic query is a type of query that is defined by using a query wizard and building blocks. No SQL knowledge is required. The Queries Tool generates an SQL SELECT statement from the selected building blocks.

   The generated SELECT statement works with all supported databases.

8. Deleting a Package

   When the created package is no longer required, then delete the package to remove the dependency from the underlying query.

   - In the Model workbench, click Packages.
   - Acquire a write lock.
   - In the navigation pane, right-click the package and click Delete Package.

13.5 SUMMARY

- Store management refers to the efficient management of materials. Store management is concerned with ensuring that all the activities involved in storekeeping are carried out efficiently and economically by those employed in the store.
- Stores form the basis of material management. Stores play a vital role in the operations of a company. In an organization, stores are mainly intended to provide staff activity in the production of goods or services. No industrial unit or public undertaking of any size can be managed efficiently without it.
- The basic objective of storekeeping is to provide services to the operating functions in the most economical manner.
- Storage is an essential and most vital part of the economic cycle and store management is a specialized function, which can contribute significantly to the overall efficiency and effectiveness of the materials function.
- The organizational set up of the store depends upon the requirements of the organization and is to be tailor made to meet the specific needs of the organization.
- The main objective of the store is to provide essential inputs to the production and operations departments at a minimum cost.
- Store functions concern receiving, movement, storage and issue of items, such as raw materials, bought out parts, tools, spares, consumables, etc., required for production, maintenance and operation of the plant and finished goods until is dispatch to customers.

NOTES

- According to Afford and Beatty, "Store management is that aspect of material control concerned with the physical storage of goods".
- According to Maynard, "Store management is to receive materials, to protect them while in storage from damage and unauthorized removal, to issue the materials in the right quantities, at the right time to the right place and to provide these services promptly and at least cost".
- Functional stores are named based on the function of the materials stored. Examples are fuels store, chemicals store, tools store, raw materials store, spare parts store, equipment store, refractories store, electric store, explosives store, finished goods store, etc.
- Physical stores can be centralized stores or decentralized stores. These stores are named based on the size and location of the store. Examples are central store, sub-store, department store, site store, transit stores, receipt store, intermediate store, open yard store, covered store, etc.
- Stores are also classified by naming them after the departments to which they serve. Examples are construction stores, operation stores, rolling mill stores, blast furnace stores, steel melting shop stores, etc.
- Stores are sometimes classified based on the nature of materials stored in them. Examples are general store, bonded store, perishable store, and inflammable store, salvage store, reject store, quarantine store, etc.
- A central store is generally a 'Wholesale' supplier to other units, departments, or sub-stores that operate on a retail basis issuing goods directly to users. All material is received and issued by one central store.
- Decentralized or branch stores are provided in considerably large plants and where one main store cannot meet the requirements of the plant without a waste of time and inconvenience.
- Warehouses are the godowns that take the responsibility of keeping and storing goods and providing ancillary services to help the small and mediumsize traders and manufacturers who, because of technical and economic reasons, may not like to have their storehouses.
- Store is to preserve the stored materials and carry out their conservation as required to prevent deterioration in their qualities. Also store is to ensure the safety of all items and materials whilst in the store, i.e., protecting them from pilferage, theft, damage, deterioration, and fire.
- Materials management is responsible for determining the amount of material to be deployed at each stocking location across the supply chain, establishing material replenishment plans, determining inventory levels to hold for each type of inventory (raw material, WIP, finished goods), and communicating information regarding material needs throughout the extended supply chain.

- A query is a request to retrieve data from the respective 'Store'. The request is in the form of an SQL SELECT statement. When you run a query, the query SQL statement is sent to the database that contains the Store details, and the database returns the results of the query.
- A generic query is a type of query that is defined by using a query wizard and building blocks. No SQL knowledge is required. The Queries Tool generates an SQL SELECT statement from the selected building blocks. The generated SELECT statement works with all supported databases.
- A custom query is a type of query where you can define your own SQL SELECT statement. When you want to use database-specific SQL syntax and grammar, create a custom query.
- A query group is a user-defined container for queries. Use query groups to organize your queries so that they are easier to find and run.
- A 'Package' is a public view of the results of a query. Data stewards use packages with the Data Manager and Merge Manager tools. You can also use packages with external applications.
- You can create two kinds of packages: Display Packages and Update Packages.
- Display packages define read-only views of master data.
- Update packages define views from which authorized users can make changes to master data.

13.6 KEY WORDS

- **Store management**: It refers to the efficient management of materials. Store management is concerned with ensuring that all the activities involved in storekeeping are carried out efficiently and economically by those employed in the store.
- **Functional stores**: Functional stores are named based on the function of the materials stored.
- **Physical stores**: Physical stores can be centralized stores or decentralized stores. These stores are named based on the size and location of the store.
- **Specific stores**: Stores are also classified by naming them after the departments to which they serve.
- **Warehouse:** Warehouses are the godowns that take the responsibility of keeping and storing goods and providing ancillary services to help the small and medium-size traders and manufacturers who, because of technical and economic reasons, may not like to have their storehouses.
- **Generic queries:** A generic query is a type of query that is defined by using a query wizard and building blocks. No SQL knowledge is required.

- **Custom queries:** A custom query is a type of query where you can define your own SQL SELECT statement. When you want to use database-specific SQL syntax and grammar, create a custom query.
- **Query groups:** A query group is a user-defined container for queries. Use query groups to organize your queries so that they are easier to find and run.
- **Package:** A 'Package' is a public view of the results of a query. Data stewards use packages with the Data Manager and Merge Manager tools. You can also use packages with external applications.

13.7 SELF-ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. What is store? Why it is required?
2. Explain the term store management.
3. How are stores categorised?
4. Name the different types of stores.
5. Elaborate on the importance of materials management.
6. What are query packages?
7. Why queries tools are used?
8. Differentiate between Display Packages and Update Packages.

Long-Answer Questions

1. Briefly explain the significance of store and store management giving appropriate examples.
2. Discuss in detail about the main objectives and functions of stores with the help of suitable examples.
3. Explain briefly about the classification of stores giving examples of each category.
4. Briefly explain the different types of stores giving appropriate examples of each type.
5. What is the role of materials management in store management and business processes? Explain giving examples.
6. Explain the terms query and the different types of queries giving examples.
7. What is a package? What are the types of packages? Explain with the help of examples.

8. Give commands and steps for:
    - Display Packages
    - Update Packages
    - Adding a Package
    - Editing a Package
    - Refreshing a Package after Changing Queries
    - Deleting a Package

13.8 FURTHER READINGS

Price, Wilson T. 1981. *Introduction to Computer Data Processing*, 3rd Edition. United States: Holt-Saunders International Editions.

Verzello, Robert J. and John Reutter. 1983. *Data Processing: Systems and Concepts*. New York: The McGraw-Hill Publishing Company, Inc.

Beech, Carol and Janice Burn. 1985. *Applications in Business Data Processing*. London (England): Pitman Publisher.

Fuori, William M. and Lawrence J. Aufiero. 1987. *Computers and Information Processing*. United States: Prentice Hall.

Sanders, Donald H. 1988. *Computers Today*. New York: McGraw-Hill.

Sinha, Pradeep K. and Priti Sinha. 2004. *Computer Fundamentals: Concepts, Systems & Applications*, 8th Edition. New Delhi: BPB Publications.

Rajaraman, V. 1999. *Fundamentals of Computers*. New Delhi: Prentice-Hall of India.

Lipschutz, Martin M. and S. Lipschultz. 1982. *Theory and Problems of Data Processing, Schaum's Outline Series*. New Delhi: Tata McGraw Hill Publishing Co. Ltd.

Stallings, William. 1996. *Data and Computer Communications*. New Jersey (USA): Prentice-Hall PTR.

UNIT 14 ONLINE BUSINESS TRANSACTIONS

NOTES

Structure

14.0 INTRODUCTION

With high consumer expectations in the competitive market, decisions that are based on the most real time data available improve customer relationships, increase revenue, increased efficiency for business reporting and maximize operational efficiency. Real-time business intelligence systems mainly provide the information necessary to tactical take advantage of events as they occur.

In this unit, you will learn about the real-time business applications and on line business transactions.

A business transaction, in the context of electronic commerce, is any monetary transaction that is made between consumers or businesses via the Internet. Business transactions free up time when conducted online since each party does not need to be physically present in order to make the transaction.

In this unit, you will learn about the real time business applications and On Line business transactions.

14.1 OBJECTIVES

After going through this unit, you will be able to:

- Discuss the real-time business applications
- Explain the on line business transaction

14.2 REAL TIME BUSINESS APPLICATIONS

A Real-Time Application (RTA) is an application program that functions within a time frame that the user senses as immediate or current. The latency must be less

than a defined value, usually measured in seconds. Whether or not a given application qualifies as an RTA depends on the Worst-Case Execution Time (WCET), the maximum length of time a defined task or set of tasks requires on a given hardware platform. The use of RTAs is also called Real-Time Computing (RTC).
Examples of RTAs include videoconference applications, VoIP (Voice over Internet Protocol), online gaming, community storage solutions, some e-commerce transactions, chatting IM (Instant Messaging), etc.

The term '**Real-Time Business**' refers to processes that permit the companies or business organisations to conduct a range of business activities instantaneously. The activities that make up real-time business can, therefore, include all aspects of business, including gathering and acting on business intelligence, developing promotional and marketing tactics, controlling and adjusting production processes, managing inventory, identifying and managing business risks, closing and fulfilling sales, and meeting customer needs.

Real-time operations depend on in-memory analytics, which takes a different approach as compared to traditional business intelligence systems. Rather than storing information on various external disks and caching bits of data in a computer's Random Access Memory (RAM), real-time intelligence puts the data directly into RAM.

Implementing a real-time business approach can provide companies with numerous benefits at both the operational level and management level as defined below.

Operational Level: At an operational level, the real-time business can be done by speeding up data capture and simplifying processes, executives can reduce inventories, minimize business risks, lower operational costs, accelerate speed to market, foster productivity and using improved methods to accomplish customer needs.

Management Level: At the management level, the real-time business can be done by accelerating decision-making and planning, executives can exploit market opportunities faster, identify competitive threats sooner, cope with market shifts more quickly and transform stagnating businesses.

The real-time business applications can be defined on the basis of Real-Time Business Intelligence (RTBI) concept which typically describes the process of delivering Business Intelligence (BI) or information about business operations as they occur. Real-time means near to zero latency and access to information whenever it is required.

The speed of today's processing systems has allowed typical data warehousing to work in real-time. The result is real-time business intelligence. Business transactions as they occur are fed to a real-time BI system that maintains the current state of the enterprise. The RTBI system not only supports the classic strategic functions of data warehousing for deriving information and knowledge from past enterprise activity, but it also provides real-time tactical support to drive

enterprise actions that react immediately to events as they occur. As such, it replaces both the classic data warehouse and the Enterprise Application Integration (EAI) functions.

NOTES

Applications of Real-Time Business System

Real-time business system is a system that is put through real-time, i.e., the response is obtained within a specified timing constraint or system meets the specified deadline. Real-time system has applications in various fields. Following are some significant applications of real-time business system.

1. **Industrial Applications:** Real-time system has a significant and very prominent role in the recent businesses and industries. The systems are specifically made real-time based so that maximum and accurate output can be obtained, hence the real-time systems are used in maximum industrial organizations. The real-time systems provide improved performance and high productivity in less time. The examples of industrial applications include Automated Car Assembly Plant, Chemical Plants, etc.
2. **Medical Science Applications:** Real-time systems are also used in the field of medical sciences, since it has vast and significant impact on the human health care, diagnosis and treatment. Because of the introduction of real-time system in medical sciences, several lives are now can be saved using the advanced medicines, technology and medical devices and equipment, the treatment of complex diseases have become easy. The examples of medical science applications include MRI (Magnetic Resonance Imaging) Scan, CT (Computed Tomography) Scan, Coronary CT Angiography (CCTA), Electron Beam Tomography (EBT), Computed Tomography Angiography (CTA), CT Pulmonary Angiogram (CTPA), Radiation therapy, etc.
3. **Peripheral Equipment Applications:** Real-time system has made the printing of books, newspapers, magazines, catalogues, booklets, leaflets, large banners, etc., very easy. Peripheral equipment are used for various purposes. These systems are embedded with microchips and perform accurately in order to get the desired response. The examples of peripheral equipment applications include Laser printer, fax machine, digital camera, etc.
4. **Telecommunication Applications:** Real-time telecommunication system map the world in such a way that it can be connected within a short time. Fundamentally, the real-time telecommunication systems have enabled the whole world to connect by means of a medium across the Internet. These telecommunication systems connect the people with each other in fraction of seconds. The examples of telecommunication applications of real-time systems include video conferencing, cellular system, etc.

5. **E-Commerce Applications:** Real-time e-commerce is the most distinguished and recognised applications based on data mining and analytics. Numerous e-commerce companies use data mining and business intelligence to offer cross-sells and up-sells through their websites. One of the most well-known and common example of real-time e-commerce application is 'Amazon'.

Check Your Progress

1. Define the term Real-Time Application (RTA) giving examples.
2. What is a real-time business system?

14.3 ON LINE BUSINESS TRANSACTIONS

A business transaction, in the context of electronic commerce, is any monetary transaction that is made between consumers or businesses via the Internet. Business transactions free up time when conducted online since each party does not need to be physically present in order to make the transaction.

'Online Business' or 'e-business' is any kind of business or commercial transaction that includes sharing information across the Internet. Commerce constitutes the exchange of products and services between businesses, groups, and individuals and can be seen as one of the essential activities of any business. Electronic commerce focuses on the use of information and communication technology to enable the external activities and relationships of the business with individuals, groups, and other businesses, while e-business refers to business with help of the Internet. Electronic business differs from electronic commerce as it does not only deal with online transactions of selling and buying of a product and/or service but also enables to conduct of business processes (inbound/outbound logistics, manufacturing and operations, marketing and sales, customer service) within the value chain through internal or external networks. The term 'E-Business' was coined by IBM's marketing and Internet team in 1996.

E-commerce (short for 'Electronic Commerce') is trading in products or services using computer networks, such as the Internet. Electronic commerce draws on technologies, such as mobile commerce, Electronic Funds Transfer (EFT), supply in chain management, Internet marketing, OnLine Transaction Processing (OLTP), Electronic Data Interchange (EDI), inventory management systems, and automated data collection. Modern electronic commerce typically uses the World Wide Web (WWW) for at least one part of the transaction's life cycle, although it may also use other technologies, such as e-mail.

Security

E-business systems naturally have greater security risks than traditional business systems, therefore it is important for e-business systems to be fully protected against

NOTES

these risks. A far greater number of people have access to e-businesses through the Internet than would have access to a traditional business. Customers, suppliers, employees, and numerous other people use any particular e-business system daily and expect their confidential information to stay secure. Hackers are one of the great threats to the security of e-businesses. Some common security concerns for e-Businesses include keeping business and customer information private and confidential, the authenticity of data, and data integrity. Some of the methods of protecting e-business security and keeping information secure include physical security measures as well as data storage, data transmission, anti-virus software, firewalls, and encryption to list a few.

Privacy and Confidentiality

Confidentiality is the extent to which businesses makes personal information available to other businesses and individuals. With any business, confidential information must remain secure and only be accessible to the intended recipient. However, this becomes even more difficult when dealing with e-businesses specifically. To keep such information secure means protecting any electronic records and files from unauthorized access, as well as ensuring safe transmission and data storage of such information. Tools, such as encryption and firewalls manage this specific concern within e-business.

Authenticity

E-business transactions pose greater challenges for establishing authenticity due to the ease with which electronic information may be altered and copied. Both parties in an e-business transaction want to have the assurance that the other party is who they claim to be, especially when a customer places an order and then submits a payment electronically. One common way to ensure this is to limit access to a network or trusted parties by using a Virtual Private Network (VPN) technology. The establishment of authenticity is even greater when a combination of techniques are used, and such techniques involve checking "Something You Know" (i.e., Password or PIN), "Something You Need" (i.e., Credit Card), or "Something You Are" (i.e., Digital Signatures or Voice Recognition Methods). Many times in e-business, however, "Something You Are" is pretty strongly verified by checking the purchaser's "Something You Have" (i.e., Credit Card) and "Something You Know" (i.e., Card Number).

Data Integrity

Data integrity answers the question, "Can the information be changed or corrupted in any way?" This leads to the assurance that the message received is identical to the message sent. A business needs to be confident that data is not changed in transit, whether deliberately or by accident. To help with data integrity, firewalls protect stored data against unauthorized access, while simply backing up data allows recovery should the data or equipment be damaged.

NOTES

Non-Repudiation

This concern deals with the existence of proof in a transaction. A business must have the assurance that the receiving party or purchaser cannot deny that a transaction has occurred, and this means having sufficient evidence to prove the transaction. One way to address non-repudiation is using digital signatures. A digital signature not only ensures that a message or document has been electronically signed by the person, but since a digital signature can only be created by one person, it also ensures that this person cannot later deny that they provided their signature.

Access Control

When certain electronic resources and information is limited to only a few authorized individuals, a business and its customers must have the assurance that no one else can access the systems or information. There are a variety of techniques to address this concern including Firewalls, Access Privileges, User Identification and Authentication Techniques (such as, Passwords and Digital Certificates), Virtual Private Networks (VPN), and much more.

Online Transaction

Online transaction is a payment method in which the transfer of fund or money happens online over electronic fund transfer. OnLine Transaction Processing (OLTP) is secure and password protected. Three steps involved in the online transaction are Registration, Placing an Order, and Payment.

OLTP is information systems that facilitate and manage transaction-oriented applications, typically for data entry and retrieval transaction processing. So online transaction is done with the help of the Internet. It cannot work properly without a proper Internet connection.

Online transactions occur when a process of buying and selling takes place through the Internet. When a consumer purchases a product or a service online, he/she pays for it through online transaction.

Stages of Online Transaction

Following are the three stages of Online Transactions:

Pre-Purchase/Sale: In this stage, the product or service is advertised online with some details for the customers.

Purchase/Sale: When a customer likes a particular product or service, he/she buys it and makes the payment online.

Delivery Stage: This is the final stage where the goods bought are delivered to the consumer.

NOTES

Steps Involved in Online Transaction

The following steps are involved in Online Transaction:

1. Registration

The consumer has to register online on the particular website to buy a particular good or service. The Customer's E-Mail ID, Name, Address, and other details are saved and are safe with the Website.

For security reasons, the Buyer's 'Account' and his/her 'Shopping Cart' is password protected.

2. Placing an Order

When a customer likes a product or a service, he/she puts the product in the 'Shopping Cart'.

The shopping cart gives a record of all the items selected by the buyer to be purchased, the number of units or quantity desired to be bought per item selected and the price for each item.

The buyer then proceeds to the payment option after selecting all the products.

3. Payment

The buyer then has to select the payment option, he/she has various payment options. These payment pages are secured with very high-level encryptions so that the personal financial information that you enter (Bank/Card Details) stay completely secure. Following are some ways in which you can make the payment:

Cash on Delivery: The Cash on Delivery option lets the buyer pay when he/she receives the product. Here, the payment is made at the doorstep. The customer can pay in cash, or by debit or credit card.

Cheque: In this type of payment, the buyer sends a cheque to the seller and the seller sends the product after the realization of the cheque.

Net Banking Transfer: Here, the payment is transferred from the buyer's account to the seller's account electronically, i.e., through the Internet. After the payment is received by the seller, the seller dispatches the goods to the buyer.

Credit or Debit Card: The buyer has to send his debit card or credit card details to the seller, and a particular amount will be deducted from his/her account.

Digital Cash: Digital Cash is a form of electronic currency that exists only in cyberspace and has no real physical properties. Here the money in buyer's bank account is converted into a code that is saved on a microchip, a smart card or on the hard drive of his computer. When he makes a purchase, he needs to mention that particular code to the Website and thereafter the transaction is duly processed.

NOTES

Digital Certificates

The point of a digital certificate is to identify the owner of a document. This way the receiver knows that it is an authentic document. Companies can use these certificates in several different ways. They can be used as a replacement for user names and passwords. Each employee can be given these to access the documents that they need from wherever they are. These certificates also use encryption. They are a little more complicated than normal encryption, however. They actually used important information within the code. They do this in order to assure the authenticity of the documents as well as confidentiality and data integrity which always accompany encryption. Digital certificates are not commonly used because they are confusing for people to implement. There can be complications when using different browsers, which means they need to use multiple certificates. The process is being adjusted so that it is easier to use.

Digital Signatures

A final way to secure information online would be to use a digital signature. If a document has a digital signature on it, no one else is able to edit the information without being detected. That way if it is edited, it may be adjusted for reliability after the fact. In order to use a digital signature, one must use a combination of cryptography and a message digest. A message digest is used to give the document a unique value. That value is then encrypted with the sender's private key.

Check Your Progress

3. Explain the term data integrity.
4. Define the term online transaction.

14.4 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. A Real-Time Application (RTA) is an application program that functions within a time frame that the user senses as immediate or current. The use of RTAs is also called Real-Time Computing (RTC). Examples of RTAs include videoconference applications, VoIP (Voice over Internet Protocol), online gaming, community storage solutions, some e-commerce transactions, chatting IM (Instant Messaging), etc.
2. Real-time business system is a system that is put through real-time, i.e., the response is obtained within a specified timing constraint or system meets the specified deadline.
3. Data integrity answers the question, "Can the information be changed or corrupted in any way?" This leads to the assurance that the message received

NOTES

is identical to the message sent. A business needs to be confident that data is not changed in transit, whether deliberately or by accident. To help with data integrity, firewalls protect stored data against unauthorized access, while simply backing up data allows recovery should the data or equipment be damaged.

4. Online transaction is a payment method in which the transfer of fund or money happens online over electronic fund transfer. OnLine Transaction Processing (OLTP) is secure and password protected. Three steps involved in the online transaction are Registration, Placing an Order, and Payment.

14.5 SUMMARY

- A Real-Time Application (RTA) is an application program that functions within a time frame that the user senses as immediate or current. The latency must be less than a defined value, usually measured in seconds.
- Whether or not a given application qualifies as an RTA depends on the Worst-Case Execution Time (WCET), the maximum length of time a defined task or set of tasks requires on a given hardware platform.
- The use of RTAs is also called Real-Time Computing (RTC). Examples of RTAs include videoconference applications, VoIP (Voice over Internet Protocol), online gaming, community storage solutions, some e-commerce transactions, chatting IM (Instant Messaging), etc.
- The term 'Real-Time Business' refers to processes that permit the companies or business organisations to conduct a range of business activities instantaneously.
- The activities that make up real-time business include all aspects of business, including gathering and acting on business intelligence, developing promotional and marketing tactics, controlling and adjusting production processes, managing inventory, identifying and managing business risks, closing and fulfilling sales, and meeting customer needs.
- Implementing a real-time business approach can provide companies with numerous benefits at both the operational level and management level.
- At an operational level, the real-time business can be done by speeding up data capture and simplifying processes, executives can reduce inventories, minimize business risks, lower operational costs, accelerate speed to market, foster productivity and using improved methods to accomplish customer needs.
- At the management level, the real-time business can be done by accelerating decision-making and planning, executives can exploit market opportunities faster, identify competitive threats sooner, cope with market shifts more quickly and transform stagnating businesses.

- A business transaction, in the context of electronic commerce, is any monetary transaction that is made between consumers or businesses via the Internet.
- Business transactions free up time when conducted online since each party does not need to be physically present in order to make the transaction.
- 'Online Business' or 'e-business' is any kind of business or commercial transaction that includes sharing information across the Internet.
- E-business systems naturally have greater security risks than traditional business systems, therefore it is important for e-business systems to be fully protected against these risks.
- Online transaction is a payment method in which the transfer of fund or money happens online over electronic fund transfer. OnLine Transaction Processing (OLTP) is secure and password protected. Three steps involved in the online transaction are Registration, Placing an Order, and Payment
- The point of a digital certificate is to identify the owner of a document. This way the receiver knows that it is an authentic document.
- A final way to secure information online would be to use a digital signature. If a document has a digital signature on it, no one else is able to edit the information without being detected.

14.6 KEY WORDS

- **Real-time business system**: It is a system that is put through real-time, i.e., the response is obtained within a specified timing constraint or system meets the specified deadline.
- **Online transaction:** Online transaction is a payment method in which the transfer of fund or money happens online over electronic fund transfer.
- **Payment:** The buyer then has to select the payment option, he/she has various payment options.
- **Digital cash:** Digital Cash is a form of electronic currency that exists only in cyberspace and has no real physical properties.

14.7 SELF-ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. Define the term Real-Time Application (RTA).
2. What does the real-time business refer?
3. Explain the importance of security.
4. Write the stages of online transaction.

NOTES

Long-Answer Questions

1. Discuss briefly the concept of real time business applications giving its significance.
2. Briefly explain the online business transaction giving appropriate examples.
3. Discuss about the steps involved in online transaction with the help of examples.

14.8 FURTHER READINGS

Price, Wilson T. 1981. *Introduction to Computer Data Processing*, 3rd Edition. United States: Holt-Saunders International Editions.

Verzello, Robert J. and John Reutter. 1983. *Data Processing: Systems and Concepts*. New York: The McGraw-Hill Publishing Company, Inc.

Beech, Carol and Janice Burn. 1985. *Applications in Business Data Processing*. London (England): Pitman Publisher.

Fuori, William M. and Lawrence J. Aufiero. 1987. *Computers and Information Processing*. United States: Prentice Hall.

Sanders, Donald H. 1988. *Computers Today*. New York: McGraw-Hill.

Sinha, Pradeep K. and Priti Sinha. 2004. *Computer Fundamentals: Concepts, Systems & Applications*, 8th Edition. New Delhi: BPB Publications.

Rajaraman, V. 1999. *Fundamentals of Computers*. New Delhi: Prentice-Hall of India.

Lipschutz, Martin M. and S. Lipschultz. 1982. *Theory and Problems of Data Processing, Schaum's Outline Series*. New Delhi: Tata McGraw Hill Publishing Co. Ltd.

Stallings, William. 1996. *Data and Computer Communications*. New Jersey (USA): Prentice-Hall PTR.