UNIT 1 COMPUTER FOR DATA PROCESSING

Structure

1.0 INTRODUCTION

Educational research involves collection, collation and analysis of large amounts of data, which can be handled manually or by using electro-mechanical devices. Calculators and Mechanical Facit machines are the common calculating devices. Computers provide the best alternative for more than one reason. Besides its capability to process data, it can store data over a long period of time, its capacity is enormous, and it can house large amounts of data.

Research involves not only collection and storage of large amounts of data, it also involves complicated calculations for testing hypotheses and carrying out calculations. Imagine if you had to calculate long and complicated problems like multiple or stepwise regression, analysis of variance and covariance, or factor analysis with the help of a tiny calculator. Computers, with the help of relevant programs can carry' out these jobs for you in minutes.

Computers carry out such complicated calculations flawlessly and with mind boggling speed. This has also been possible due to sophistication in the software. Whereas earlier programmes had to be written for each incidence of data analysis, ready made packages

are now available for data analysis. The calculations that took months earlier, in the pre-computer age take now a few minutes and hours. Also, access to computers have increased substantially over the last few years all over the world including the third world countries.

In this unit, we will introduce you to computers for data processing - a brief introduction to hardware and software, data processing and how to use a computer for research.

1.1 OBJECTIVES

On the completion of this Unit, you should be able to

- Describe a computer configuration,
- Describe different types of software,
- Enumerate major steps in data processing, and
- Explain the methods of using computers in research.

1.2 DEFINITION OF COMPUTER

A computer is a high-speed electronic device capable of performing arithmetic and logical operations and of sorting and executing a set of instructions, which will enable it to perform a series of such operation without manual intervention. The ability to do different jobs on the same machine distinguishes computers from other machines. Computers can do a job exactly as per the instructions given. But they cannot do a job unless they are told how to do it.

1.3 COMPUTER HARDWARE

A typical computer terminal has five components:

- a computer - central processing unit
- a display unit - TV like box,
- a keyboard - similar to that of typewriter keyboard,'
- a mouse with two or three push down buttons, and
- a printer.

Of all the five, one that keeps blinking a light from one or two of the tiny windows is actually the computer. The rest are either input or output devices that connect the computer to the outside world.

A computer is defined in the Oxford dictionary as "an automatic electronic apparatus for making calculations or controlling operations that are expressible in numerical or logical terms". The definition clearly categorises computer as an electronic apparatus although the initial computers were mechanical and electromechanical. It also points to the two major areas of computer application, viz., data processing and computer assisted control/ operations and that the computer can perform only those operations/calculations which can be expressed in logical or numerical terms. The basic function performed by a

computer is the execution of a program. Program is a sequence of instructions, which operates on data to perform certain tasks.

There are several generations of computers and the current generation of computers has evolved over a period of time. The various generations of computers and their evolution have been "effectively described in the Certificate in Computer and Bachelor in Computer Applications course materials of IGNOU. For our limited purpose of introducing computers for educational research, we will examine the major categories of computers and their capabilities. But let us first examine what a computer consists of.

Basically, a computer comprises three components - a central processing unit, input devices and output devices.

1.3.1 Central Processing Unit

The Arithmetic Logic Unit (ALU) and the Control Unit (CD) together are termed as the Central Processing Unit (CPU). The CPU is the most important component of a computer's hardware. The ALU performs the arithmetic operations such as addition, subtraction, multiplication and division, and the logical operations such as: Is A = B? (Where A and B are both numeric or alphanumeric data), Is a given character equal to M (for male) or F (for female)? The control unit interprets instructions and produces the respective control signals. All the arithmetic and logical operations are performed in the CPU in special storage areas called registers. The size of the register is one of the important considerations in determining the processing capabilities of the CPU. Register size refers to the amount of information that can be held in a register at a time for processing. The larger the register size, the faster may be the speed of processing. A CPU's processing power is measured in Million Instructions Per Second (MIPS).

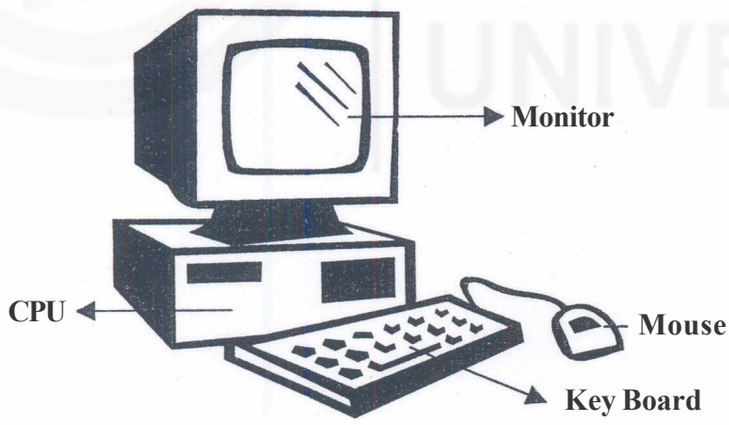


Fig. 1: A Computer

How can the instructions and data be put into the computer? The instructions and data need to be supplied from external environment; therefore, an input module is needed. The main responsibility of input module is to put the data in the form of signals that can be recognised by the system. Similarly, we need another component that will. report the results in proper format and form. This component is called the output module. These components are referred together as input/output (I/O) components.

Are these two components sufficient for a working computer? No, because input devices can bring instructions or data only sequentially and a program may not be executed sequentially as jump instructions are normally encountered in programming. In addition, more than one data elements may be required at a time. Therefore, a temporary storage area is needed in a computer to store temporarily the instructions and the data. This component is referred to as **memory**. The memory unit stores all the information in a group of memory cells, also called memory locations, as binary digits (bits).

Each memory location has a unique address and can be addressed independently. The contents of the desired memory locations are provided to the central processing unit by referring to the address of the memory location. The amount of information that can be held in the main memory is known as **memory capacity**. The capacity of the main memory is measured in Kilobytes (KB) or Megabytes (MB). One kilobyte stands for 2^{10} bytes, which is 1024 bytes (or approximately 1000 bytes). A megabyte stands for 2^{10} kilobytes, which is approximately little over one million bytes.

In addition, to transfer the information, the computer system needs internal interconnections. The most common interconnection structure is the bus structure. A bus is a set of wires (Lines) which you can visualise on the motherboard of computer. It is a shared media. A bus connecting the CPU, memory and I/O components is called a system bus. A system bus may consist of 50 to 100 separate lines.Figure 2. shows the basic structure of a conventional computer (Von Neumann machine).

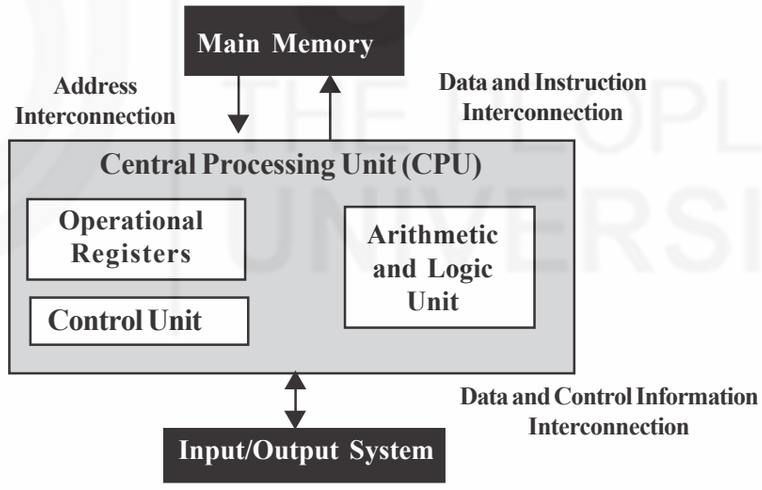


Fig. 2: Structure of a Computer

Memory System

Memory in a computer system is required for storage and subsequent retrieval of the instructions and data. A computer system uses a variety of devices for storing the instructions and data which are required for its operations. Let us review a few common terminologies on computer memory.

All modern computers use semiconductor memory as its main memory. Semiconductor

memory is known as Random Access Memory (RAM) since any part of the memory can be accessed for reading and writing. Another part of main memory is Read Only Memory (ROM). ROMs (Read Only Memory) are the memories on which it is not possible to write the data when they are on-line to the computer. They can only be read. The ROMs can be used in storing programs provided by the manufacturer of computer for basic operations. ROMs are non-volatile in nature and need not be loaded in a secondary storage device. ROMs are fabricated in large numbers in a way where there is no room for even a single error.

ROMs can be written only at the time of manufacture. Another similar memory is PROM. PROMs are also non-volatile and can be programmed only once by a special write device hence the name Programmable ROM (PROM). The writing process in PROM can be performed electrically by the supplier or the customer. Special equipment is needed to perform this writing operation. Therefore, PROMs are more flexible and convenient than ROMs.

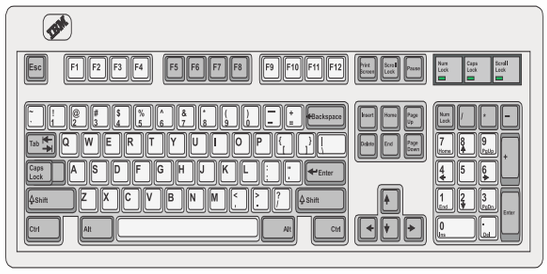
There is another kind of memory, which is increasingly being used in modern computers and this is called **Cache Memory**. It is logically positioned between the internal memory (registers) and the main memory. It stores or catches some of the content of the main memory, which is currently being used by the processor. Caches are fast and yet very extensive memories; thus are used in only small sizes: For example, Caches of sizes 64 K, 128K, 256 K, etc. are normally used in a typical PC-486 and Pentium based PCs while they can have 1 to 64 MB RAMs or even more. Thus, small Cache memories are intended to provide fast speed memory retrieval without sacrificing the size of the memory (because of the main memory size).

1.3.2 Input/Output Devices (Peripherals)

The computer will be of no use if it does not communicate with the external world. Thus, a computer must have a system to receive information from the outside world and able to communicate results to the external world. The devices that enable the computer to receive data and instruction and give output in the visual or print form are called input/ output devices.

Input Devices

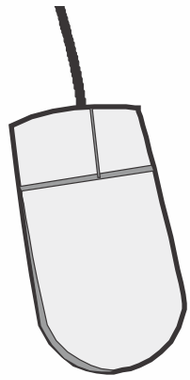
There are several input devices. The keyboard is one of the most common input devices for computers. The layout of the keyboard is like that of the traditional typewriter, although there are some extra commands and function keys provided for. Substantial development has taken place in the ergonomics of keyboard design to ensure that operator strain is minimal.



Key Board

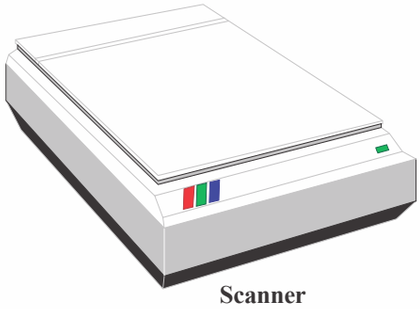
The mouse is another handy input device which can be moved on a smooth surface to simulate the movement of the cursor as desired on the display screen. A mouse could be optical, offering quiet and reliable operations, or mechanical, which is cheaper but noisier.

The user can move the mouse, stop it at a point where the pointer is to be located and, with the help of buttons, make selection of choices. Other input devices are pointing devices, light pen, speech/voice input, scanners, etc.



Mouse

Most recent trends in data input are towards source data automation. The equipment used for source data automation captures data as a by-product of a business activity thereby completely eliminating manual input of data. Some examples are Magnetic Ink Character Recognition (MICR). Optical Mark Recognition (OMR), and Optical Bar Code Reader (OBCR).

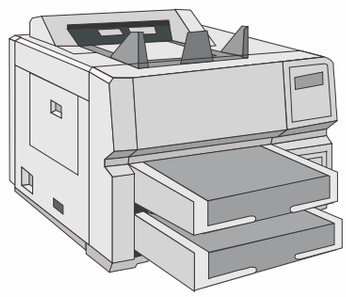


Scanner

Output Devices

The output can normally be produced in two ways, either on a display unit/device, or on paper. Other kinds of output such as speech output, or mechanical output is also being used in certain applications. The most common display output device is the graphic display device. Conventional computer display terminals knows as alphanumeric terminals, display characters (images) from a multi-dot array (normally 5 x 7 or 7 x 9). These are used to read text information displayed on the screen. However, there is an increasing demand for display of graphics, diagrams and pictures to make the visual presentation of information more effective for user interaction and decision making.

Graphic display is made up of a series of dots called 'pixels' (picture elements) whose pattern produces the image. Each, dot on the screen is defined as a separate unit, which can be directly addressed. Since each dot can be controlled individually there is a much greater flexibility in drawing pictures. Other kinds of displays, devices are CRT terminal, LCD, etc. In projection display, the personal size screen of the displays is replaced by a large screen upon which images are normally used for large group presentations. These systems can be connected to the computer and whatever appears on the computer terminal gets enlarged and projected on a large screen.



Printer

Printer - The printers are the output devices. They are classified by the method of printing. There are mainly three types of printers connected to personal computers.

1. Dot Matrix Printers
2. Daisy Wheel Printers
3. Laser Printers

1.3.3 Classification of Computers

Computers have been classified under three main classes, namely, microcomputers, minicomputers and mainframes.

Although with technological development the distinctions between these are becoming blurred, it is important to classify them as it is sometimes useful to differentiate the key elements and architecture of the different classes.

Supercomputer

The upper end of the state of the art mainframe machines comprises supercomputers. These are amongst the fastest machines in terms of processing speed and using multiprocessing techniques, where a number of processors are used to solve a problem. Lately, ranges of parallel computing products, which are multiprocessors sharing common buses, have been in use in combination with the mainframe supercomputers. The supercomputers are reaching speeds well over 25000 million arithmetic operations per second. India also has its indigenous supercomputer.

Supercomputers are mainly being used for weather forecasting, computational fluid dynamics, remote sensing, image processing, biomedical applications, etc. In India, we have one such mainframe supercomputer system- CRA Y XMP-14, which is at present, being used by the Meteorological Department.

Mainframes

Mainframe computers are generally 32-bit machines.. These are suitable for big organisations, to manage high volume applications. A few of the popular mainframe series are IBM, HP', etc. Mainframes are also used as central host computers in distributed systems. Libraries of application programs developed for mainframe computers are much larger than those of the micro or minicomputers because of their evolution over several decades as families of computing. These, and several other factors, make the mainframe computers indispensable in spite of the popularity of microcomputers.

Minicomputer

The term minicomputer originated in the 1960s when it was realised that many computing tasks do not require an expensive contemporary mainframe computer but could be done by a small, inexpensive computer. Initial minicomputers were 8-bit and 12-bit machines but by the 1970s almost all minicomputers were 16-bit machines. The 16-bit minicomputers have the advantage of large instruction sets and address fields, and efficient storage and handling of text, in comparison to lower bit machines. Thus, 16-bit minicomputer is a more powerful machine which can be used for a variety of applications and can support business applications along with scientific applications.

With the advances in technology, characteristics like the speed and memory size developed and the minicomputer was then used for various stand-alone or dedicated applications, and then as a multi-user system. Gradually, the architectural requirement of minicomputers grew and a 32-bit minicomputer, which was called supermini, was

introduced. The supermini has more peripheral devices, large memory and can support more users working simultaneously on the computer unlike in the case of the previous minicomputers.

Microcomputers

A microcomputer's CPU is a microprocessor. The first microcomputers were built around the 8-bit microprocessor chips. An improvement on the 8-bit chip technology was seen in the early 1980s, when a series of 16-bit chips, namely, 8086 and 8088 were introduced by Intel Corporation, each one with an advancement over the other. Both these chips can support a primary storage capacity of up to I-megabyte (MB).

Most of the popular microcomputers are developed around Intel's chips, while most of the minis and superminis are built around Motorola's 68000 series chips. With the advancement of display and VLSI technology now, a microcomputer of a very small size is available. Some of these are laptops, notebook computers, etc. Most of these are of the size of a small notebook but of capacity equivalent to an older mainframe.

Personal Computers

These are mini-computers as per the categorisation given above. The current generation Pentiums are 32-bit machines. Usually, the configuration of the current generation of personal computer is as follows:

- Pentium II or III Processors or Celeron Processor
- 64 mb RAM
- Cache Memory
- 4.3 GB Main Memory
- 33 Mhz Speed
- MB 3-114" Floppy Drive and
- CD ROM Drive

These computers are fully equipped to handle all the softwares that are normally required for data processing in research projects. They have sufficient memory to store huge amounts of data. The CD-ROM itself has a memory capacity of about 650-MB.

Check Your Progress 1

State true or false.

(a) In cases where graphical user interfaces are common, mouse should not be used.

True ☐ False ☐

(b) A keyboard is one of the most common input devices.

True ☐ False ☐

(c) Scanners are devices used for outputting pictures.

True ☐ False ☐

(d) Projection displays can be used for classroom teaching.

True ☐ False ☐

(e) A keyboard, a VDU and a printer are essential for computers.

True ☐ False ☐

1.4 COMPUTER SOFTWARE

Computer Software consists of set of instructions that mould the raw arithmetic and logical capabilities of the hardware units in to performance.

In order to communicate with each other, you use natural languages like Hindi, English, French, Spanish, Chinese, Japanese, Arabic, etc. Similarly, you use programming languages of one type or another in order to communicate instructions and commands to a computer for solving problems.

1.4.1 Program and Programming

A computer can neither think nor make any judgement on its own. It is equally impossible for any computer to analyse independently a given data and follow its own method of solving problems. It needs a program to instruct it as to what to do. A program is a set of instructions arranged in a sequence that guides the computer to solve a problem.

The process of writing a program is called *Programming.* Programming is a critical step in data processing. If the system is not correctly programmed, it delivers results that

cannot be used. There are two ways in which you can acquire a program. One is to purchase an existing program which is normally referred to as ***packaged software,*** and the other is to prepare a new program from scratch in which case it is called a ***customised software***.

A computer software can be broadly classified into two categories—***System Software*** and ***Application Software.***

Today, there are many languages available for developing programs. These languages are designed keeping in mind some specific areas of applications. Thus, some of the languages may be good for writing *system programs/software* while some others may be good for *application software.* Since a computer can be used for writing various types of application, there are different programming languages.

i) System Programming Languages

System programs are designed to make the computer easier to use. An example of system software is an operating system, which consists of many other programs for controlling input/output devices, memory, processor, etc. To write an operating system, the programmer needs instructions to control the computer's circuitry (hardware part); for example, instructions to move data from one location of storage to a register of the processor. C and C++ languages are widely used to develop system software.

ii) Application Programming Language

Application programs are designed for specific uses, such as, student information systems management, payroll processing, inventory control, etc. To write programs for payroll processing or other applications, the programmer does not need to control the basic circuitry of a computer. Instead, the programmer needs instructions that make it easy to feed data, produce output, make calculations and store and retrieve data. Programming languages that are suitable for such application programs support these instructions but not necessarily the types of instructions needed for development of system programs.

There are two main categories of application programs — business and scientific application programs. Most programming languages are designed to be good for one category of applications but not necessarily for the other, although there are some general purpose languages that support both types. Business applications are characterised by processing of large inputs and large outputs, high volume data storage and retrieval, but call for simple calculations. Languages, which are suitable for business program development must support high volume input, output and storage but need not support complex calculations. On the other hand, programming languages that are designed for writing scientific programs contain very powerful instructions for calculations but rather simple instructions for input, output, etc. Amongst traditionally used programming languages, COBOL (Commercial Business Oriented Programming Language) is more suitable for business applications whereas FORTRAN (Formula Translation - Language) more suitable for scientific applications. Let us briefly look at the categories of software, viz., system and application software.

1.4.2 System Software

Language Translator: A language translator is a system software which translates a computer program written by a user into a machine understandable form.

Operating System

An operating system (OS) is the most important system software used to operate a computer system. It manages a computer's resources very effectively, takes care of scheduling multiple jobs for execution and manages the flow of data and instructions between the input/output units and the main memory. Advances in the field of computer hardware have also helped in the development of more efficient operating systems.

Utilities

Utility programs are those which are often requested by many application programs. A few examples are:

SORT/MERGE utilities, which are used for sorting large volumes of data and merging them into a single sorted list, formatting, etc.

The positioning of the operating system, system software and application software in the overall computer system is shown in figure 3.

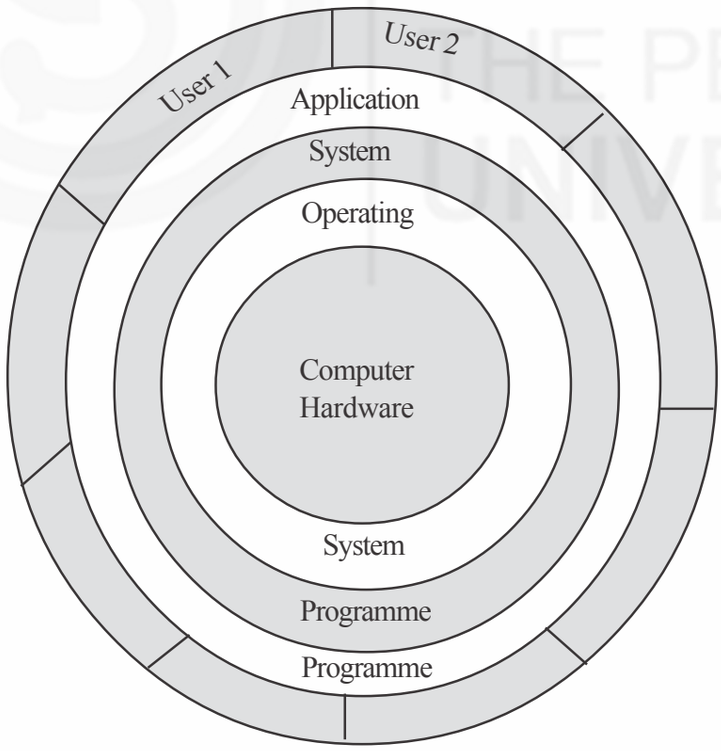


Fig. 3: Components of Computer Sysem

1.4.3 Application Software

Application software is written to enable the computer to solve a specific data processing task. A number of powerful application software packages which do not require significant programming knowledge, have been developed. They are easy to learn and use as compared to the more complex programming languages. Although these packages can perform many general and special functions, there are applications where these packages are not found adequate. In such cases, an application program is written to meet the exact requirements. A user application program may be written using one of these packages or a programming language. The most important categories of software packages available are:

- Data Base Management Software
- Spreadsheet Software
- Word Processing
- Desktop Publishing (DTP)
- Presentation Software
- Graphics Software
- Data Communication Software
- Statistical and Operational Research Software.

Data Base Management Software

Databases are very useful in creating and maintaining queries, the databases and generation of reports. Many of today's Database Management Systems are Relational Database Management Systems. Many RDBMS packages provide smart assistance for creating simple databases for distance education students - their socio-economic background, entry qualification, credits earned, final results, etc. Many database management systems are available in the market these days. You can select anyone of them based on your needs. Generally, it is good to have a few database packages like dBase, FoxPro, etc. If you require some additional features and moderate work then "load Lotus Approach and Microsoft Access also. However, if you are having high-end database requirements which requires multi-user environment and data security, access right, very good user interface, etc., then, you must go for professional RDBMS package like Ingress, Oracle, Integra etc.

Accounting Package

The accounting packages are one of the most important packages for an office. Some of the likely features which you may be looking for, are:

- Tax planner facility
- Facility for producing charts and graphs
- Finding accounts payable
- Simple inventory control facility
- Payroll functions
- On-line connection to stock quotes

- Creation of invoices Easy

One of the good packages in this connection is Quicken for windows.

Communication Package

Communication software includes software for fax. The fax-software is fast growing. An important fax software is Delrina's WinFax PRO 4.0. Some of the features such as Remote Retrieval and Fax Mailbox should be looked for in a fax software. These features ensure that irrespective of your location you will receive the fax message. Another important feature is fax broadcast. This allows you to send out large numbers of faxes all day long without tying your fax machine.

If you have to transfer files from your notebook computer to a desktop computer constantly, then you need a software program that coordinates and updates documents. On such software is Laplink for Windows. This software offers very convenient features. For example, a file can be transferred by simply dragging and dropping it. This software can work if you are connected to a serial cable or a Novell network or a modem.

Desktop Publishing Packages

Desktop Publishing Packages are very popular everywhere. Newer publishing packages also provide certain inbuilt formats such as brochures, newsletters, flyers, etc., which can be used directly. Already created texts and graphics placements can be put into these packages. Many DTP packages for English and other languages are available. Microsoft Publisher, Page Maker, Corel, Ventura are some of the popular names.

Desktop publishing packages, in general, are better equipped in Apple- Macintosh computers.

Information Providers

One of the very interesting information providers which has become popular all over the world, is Automap road atlas by Microsoft. This package provides city-to-city driving instructions and maps. It provides information about the best route and the time it will take.

Several Internet access programs also provide useful information. Today, Internet access packages come as a part of the operating system; however, many other packages can be used for accessing information on the World Wide Web. One very simple method of using the popular tool of browsing Internet is Netscape Navigator.

Organisers, Contact Managers, PIMs

Some of the significant tasks of-an office manager are:

- to track contacts
- to balance schedules
- to manage projects
- to prioritise tasks

These functions can be easily performed using organiser programs which have a phone book model for maintaining lists of contacts. They also have a calendar for entering appointments and to-dos. Some of these packages are Okna's DeskTop Set for Windows, Lotus Organiser, Microsoft Outlook, etc.

The Personal Information Manager (PIM) is a tool that stores virtually any information such as reference materials, project details, etc. The PIM document contains outlines, folders and links. Most of the data in the PIM is presented as an outline; for example the client may represent the top level followed by the date of appointment with him/her and the details of the meeting indented further below. This item can be further linked to all other appropriate places.

Suites

Suites are a set of packages sold as a group package mainly for the business user. A suite package includes programs for Word-Processing, Electronic Spreadsheet, Databases, and Presentation Graphics software and may also be a mail software. For example, Microsoft Office Professional for Windows includes programs like Microsoft Word, Microsoft Excel and Microsoft Access and a license for Microsoft Mail. The word-processing, spreadsheet, and presentation-graphics software interfaces in a suite are well-integrated allowing easy data transfer among these applications. Today there is a growing family of office-compatible products which can be included in various suites.

1.4.4 Software Languages

You can choose any language for writing a program as per your need, but a computer executes programs only after they are represented internally in a binary form (sequences of Is and Os). Programs written in any other language must be translated in to a binary representation of the instructions before a computer can execute them. Programs written for a computer may be in anyone of the following categories of languages:

Machine Language

It is a sequence of instructions written in the form of binary numbers consisting of Is and Os to which the computer responds directly. The machine language was initially referred to as a code, but the term code is now used more generally to refer to any program text.

An instruction prepared in any machine language will have at least two parts. The first part is the command or operation, which tells the computer what functions are to be performed. All computers have an operation code for each of their functions. The second part of the instruction is the operand i.e., it tells the computer where to find or store the data in question.
Just as hardware is classified into generations based on technology, computer languages also have a generation classification based on the level of interaction that can effect with the machine. A machine language is considered to be the first generation language.

The advantage of Machine Language that it is faster in execution since the computer directly starts executing it on the otherhand it is difficult to understand and develop a program using a machine language. Anybody going through this program for checking will

find it difficult to understand what is achieved by executing this program.

Assembly Language

A program that uses symbols (letters, digits or special characters) for the operation the address and other parts of the instruction code, is called an assembly language program. This is considered to be the second-generation language.

Machine and assembly languages are referred to as low level languages since the coding for a problem is at the individual instruction level. Each machine has got its own assembly language which is dependent upon the internal architecture of the processor. An assembler is a translator, which takes its input in the form of an assembly language program and produces machine language code as its output.

A machine cannot execute an assembly language program directly as it is not in a binary form. An assembler is needed to translate an assembly language program into the object code executable by the machine.

Writing a program in an assembly language is more convenient than in a machine language. Instead of following a binary sequence as is the case in a machine language, it is written in the form of symbolic instructions which increase its readability. An assembly language (program) is specific to a particular machine architecture. Assembly languages are designed for a specific makes and models of a microprocessor. It means that assembly language programs written for one processor will not work on a different processor if it is architecturally different. That is why an assembly language program is not portable. An assembly language program is not as fast as a machine language. It has to be first translated into a machine (binary) language code to get/be used.

High-Level Language

We have talked about programming languages such as COBOL, FORTRAN and BASIC. They are called high level programming languages. The time and cost of creating machine and assembly languages was quite high. And this was the prime motivation for the development of high level languages. A high level source program should be first translated into the form the machine can understand and this is done by a software called compiler which takes the source code as input and produces a machine language code as output.

During the process of translation, the compiler reads the source programs statement-wise and checks the syntax (grammatical) errors. If there is any error, the computer generates a printout of the errors it has detected. This action is known as diagnostics.

In educational research, you need primarily two or three types of software -software that can process data, software that can create graphic display of data and software that can produce research reports. Keeping in mind your software needs for research in general and data processing in particular, we have presented

- MS-Word in Unit 2
- Excel - Spreadsheet in Unit 3
- SPSS - for statistical analysis of data in Unit 4

Check Your Progress 2

1. What is computer software?
2. Explain the following terms in one or two sentences each.
   (a) Operating Systems
   (b) Database Management Software

Notes: (a) Space is given below for your answer.
(b) Compare the answer with the one given at the end of this unit.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

1.5 DATA PROCESSING

In computer literature, there are definitions and elaborate descriptions of various related words like data, information, etc. We will restrict our discussion to educational research. As you have seen in various units in previous blocks, research tools are used to collect data. There is a separate unit on data collection in Block IV. When you administer a structured questionnaire or a test, what you get is the data. Similarly, when you take interviews or use observation or participant observation techniques, you land up with a lot of information.

These data, however, are discrete and pertain to each individual respondent. As a researcher, you need to find a pattern — how does a variable behave in a group. For example, let us assume that you have collected information from 400 distance learners on their motivation for joining the distance education course. Instead of saying what 400

individuals say, you should be able to say on the basis of the 400 responses why people join distance education courses. To come to that kind of a conclusion, you need to' analyze and process data. There are several stages to be passed when you process data with the help of a computer, namely, data feeding, data checking, creating a data file, actual data processing, data output.

Fig. 4: Data Processing

Data collected through the use of research tools need to be entered into the computer. When you use EXCEL, the data can be entered directly on to the worksheets. The numbers in the columns can represent respondents and rows can represent variables. For others, specific data entry format has to be developed. The pattern, however, remains more or less the same - data on each variable against every respondent. Research students often find difficulty in entering data because they do not take sufficient care when the questionnaires and other tools are being designed. The research tools and response patterns can be structured keeping in mind the requirements of data entry in the computer.

Data entry can be done manually using the keyboard and electro--mechanically by using other types of input devices. For example, for large scale data on examinations or admissions, responses are sought on a pre--designed response sheet; these sheets are then used to transfer data to computer memory through OMR. The choice depends upon the volume of the data and the way data have been recorded. However, for the research projects that you will deal with in your project for MA(DE), data can be entered manually by using a computer keyboard since the data will be small in volume though on multiple variables, and these will be recorded on sheets or paper transferred from the research

tools scoring sheet. You can enter the data yourself if you are well versed with the computer and its keyboard. Alternatively, you can take help of a data entry operator. Whether you do it yourself or use the services of a data entry operator, caution must be taken to ensure correct entry of data. A computer will take the input and analyze whatever data are fed into it. Wrong and incorrect data will offer incorrect results and hence provide flawed conclusions. Therefore, it is necessary to check and verify data before it is analysed.

Data verification can be done in more than one way. There can be a sample check or a comprehensive check of the data. After the entry of the data, a data sheet should be printed. The printed data can be compared with the data contained in the original sheet. The second possibility is to check the data on the computer terminal itself by comparing the same on the display unit with the data recorded on paper. For large scale data and where complete accuracy is required, like examination results, they are fed parallel into two or more computers. Sample-check is done by using computer programs. What is important is to check data for accuracy in data entry. For our purpose of processing research data, it can be checked manually by comparing the entered data either iil the printed form or on the screen with the original data sheets. Following this, inaccuracies have to be corrected.

Research data entry and data checking are mechanical processes which require great attention. Unlike word processing, there are no automatic indications of mistakes and wrong entries. Therefore it is wiser to use the services of a competent data entry operator to reduce the chances of errors.

Coming to the issue of actual data processing, it is important to remember that as you enter data in the computer, you create a data file. Probably, the computer already houses a program file. Technically, data processing is interfacing the data file with the progran1 file. The moment you draw the mouse and block certain amount of data, say a particular column in a worksheet, the computer will immediately register the segment of data you wish to deal with. Next, when you call upon a particular symbol or a formula, the computer will activate the corresponding segment of the program and the subroutines required to carry out the operation. The. program draws the data from the data file according to its own specification. It carries out the necessary mathematical operations and produces the result. The entire process operates so fast that it appears to be automatic. Once, the processing of the data is over, the results can be seen on the display unit of the computer or can be printed. It may not be possible to use the printed results directly for the research reports. You may need to construct tables from the result sheets.

What is important for the researcher is to learn to enter data, verify data for accuracy and choose the right kind of statistics and programs for data analysis.

1.6 USING A COMPUTER FOR DATA PROCESSING

In using a computer for data processing, it is really not necessary to be an expert in computer applications or an expert programmer. You, as a researcher are the user and you should have the skills of the user. Personal knowledge of computers and software is an additional competence but is not crucial. What is important is to have a fair idea about

hardware and software.

The skill required of a researcher is his/her ability to understand the statistical tools for data analysis. The decision regarding the statistical tools depends upon the objectives and hypotheses, research designs, research tools, size of the sample, etc., the details of which you have read in Blocks 3 and 4. While you look at the use of a computer for "data processing" you need to be clear about the relevant software for data analysis. For example, if you as researcher warrants use of 'chi-square' or analysis of variance, you must choose programs that can do those functions. But, how do you know that?

As mentioned earlier, for data processing you can use two types of software - tailor-made or a package. For example, if you require to study the relationship between two variables, you can write or get a program written on correlation. Alternatively, you can look at a readily available package that has a program on correlation. For example, Statistical Package for Social Sciences (SPSS) has a wide range of statistical programs that are normally required by researchers in social sciences. Similarly, EXCEL can carry out certain operations including graphics. The details on EXCEL and SPSS are given in Units 3 and 4 of this Block.

One important factor in a research exercise is the process of consulting. Generally, you would consult your research guide. There are, however, certain specialized areas in research which need special consultations with experts, besides the research guides. Two such areas are research designs and data processing. If you or your research guide is not very conversant with statistical techniques and computer programs, it is worth consulting experts — specialists in research design and statistical methods, and then a computer software professional.

1.7 LET US SUM UP

In this Unit, our basic purpose was to introduce you to the use of computers for data processing. Although understanding of the architecture of a computer or the algorithm and rules of software programming are not necessary for research data processing, it is desirable for you to have a fair knowledge about computer hardware and software so that you know how it actually carries out the instructions to analyze your data. Accordingly, we have included in this Unit an introduction to hardware and software. We avoided details about the evolution of computers, the actual technology that goes into configuring a computer, nor have we dealt in great length with any machine language, assembly language, programming languages or the fourth generation languages among them.

1.8 CHECK YOUR PROGRESS: THE KEY

l. a). True b). False c). False d). True e). True

2.1. Computer software consists of set of instructions, which can be used for a variety of applications as it is not modelled for specific applications only. Von Neumann machines are general purpose machines since they can be programmed for any general application,

while a microprocessor based control systems are not general purpose machines as they are specifically modelled as control systems.

2.2 (a) Operating System Software manages a computers resources effectively. It also manages the flow of data and instructions between the input/output units and the main memory.

(b) It provides assistance for the creation of simple databases.