Unit 1: Introduction to Computer System

Information Technology:

IT is the techniques used in Information System. An information system collects, processes, stores, analyzes, and disseminates information for a specific purpose. Communication is the backbone of IT. IT can be defined as the implementation of Information Science where Information Science is a set of algorithms, principles and theories. Computer System, Internet, Telecommunication systems are the information technology we use.

Information systems are scattered throughout organizations, often in several locations and sometimes in two or more organizations. IT is implemented here to relay the information between these locations. Not only the organization but also the individual today is dependant on IT for career development and education. Education Institutions today provide distant education and online examinations. GIS has emerged due to the development in IT. Scientists now are able to acquire geographical information from the satellites.

Evolution of Information Processing :

Information Processing evolved due to the growth of national and international organizations. Since the early computing days, as costs declined and computer technologies improved, computers have been used to support managerial and other organizational activities of every sort.

Scopes/ Application field of Information System:

Telecommunication
Internet
GIS
Education
Security System
E-Governance
Medicine
Scientific Research

1.1. Introduction to Computers:

A computer is a programmable electronic machine. It takes raw facts as input, process these and gives the final output which is the result of processing. It responds to a specific set of instructions in a well-defined manner. It can execute a prerecorded list of instructions (a program).

1.2.Classification of Digital Computers (Types of Computers)

Computers can be classified by their size and power as follows:

1. Super Computers
2. Mainframe Computers
3. Mini Computers
4. Work Stations
5. Micro Computers/ Personal Computer

1. Super Computers: They are the computers with the most processing power.
   - The primary application of supercomputers has been in scientific and military work, but their use is growing in business as their prices decreases.
   - They are especially valuable for large simulation models of real world phenomena, where complex mathematical representations and calculations are required or for image creation and processing.
   - They are also used in weather prediction, design aircraft(Boeing 777), motion picture like star wars and Jurassic Park)
   - They operate generally at 4 to 10 times faster than the next most powerful computer class., the mainframe.
2. Mainframe Computers:
   - They are less powerful and generally less expensive than supercomputers.
   - Large corporate use mainframe computers for centralized data processing maintaining large databases.
   -
   - Application than run on a mainframe can be large and complex, allowing for data and information to be shared throughout the organization.
   - Examples: Airline Reservation System, Corporate Payroll, Student Information etc.
   - A mainframe system may have anywhere from 50megabytes to several gigabytes of primary storage.
   - Secondary storage may use high capacity magnetic and optical storage media with capacities in the gigabytes to terabyte range.
   - Typically, several hundreds or thousands or online computers can be linked to a mainframe.
   - Advance mainframe perform more than 1,000 MIPS and can handle up to one billion transactions per day.

Mini Computers

- Also called midrange computers, are relatively small, inexpensive and compact computers that perform the same functions as mainframe computers but to limited extent.
- They are designed specially for accomplishing specific tasks such as process control, scientific research, and engineering applications.
- IBM is the market leader in minicomputers with its AS/400 series of computers.
- Larger companies gain greater corporate flexibility by distributing data processing with minicomputers in organizations units instead of centralization at one place.
- They form the network .

Workstations

- Computers vendors originally developed desktop engineering workstations or to provide the high levels of performance demanded by engineers.
- They are based on RISC (Reduced Instruction Set Computing) architecture and provide both very high speed calculations and high resolution graphic displays.
- The distinction between workstation and personal computers is rapidly blurring.
- The latest PC s have the computing power of recent workstation.

Micro Computers

These are also called PCs and are the smallest and least expensive category of general purpose computers. Micro computer ranges from palmtops to Desktop computers. Laptops/ Notebooks are very popular now a days. Microprocessor is the core part of the Micro computers which forms the CPU.

1.3.Anatomy of Digital Computers:

Functions and Components of a Computer:

To function properly, the computer needs both hardware and software. Hardware consists of the mechanical and electronic devices, which we can see and touch. The different parts of the computer are Processor (CPU), Input devices, Output devices, Storage devices and Memory devices. The software consists of programs, the operating systems and the data that reside in the memory and storage devices.

A computer does mainly the following four functions:

- ❖ Receive input – Accept information from outside through various input devices like the keyboard, mouse, etc.
- ❖ Produce information – Perform arithmetic or logical operations on the information.
- ❖ Produce output – Communicate information to the outside word through output devices like monitor, printer, etc.
- ❖ Store information – Store the information in storage devices like hard disk, floppy disks, etc.

Computer hardware falls into two categories: processing hardware, which consists of the central processing unit (CPU), and the peripheral devices. The CPU, as its name implies, is where the data processing is done. Peripheral devices allow people to interact with the CPU. Together, they make it possible to use the computer for a variety of tasks.

Explain CPU and its different components: CU, ALU and MU.

How does CPU and memory works ?

- An instruction is fetched from primary storage by the Control Unit
- The Control Unit decodes the instruction
- The ALU receives the data and the instruction and performs the calculation or comparison
- The result is stored in primary storage which is sent to the proper output device.

1.4. Computer Architecture:

History: *Describe from Pascaline to the first generation computers:*

Pascaline :

In 1642, 19 year old French mathematician Blaise pascal invented a mechanical adding-machine called pascaline . The numbers were entered by means of add and subtract. It had 8 wheels and each wheel had 10 digits from 0 to 9.

Stepped Recknor

In 1671, German mathematician Gothfried von Leibnitsz invented an improved and strong computing machine called "stepped Recknor" which could multiply, divide and workout square roots apart from addition and subtraction. The machine also had handle and gear.

Analytical Engine

In 1833, Charles Babbage, (English mathematician) developed Analytical Engine which was a kind of a general purpose computer designed to solve any arithmetical problems. It was significant in a way that it had most of the elements present in today's digital computer systems, that's why he's called "Father of modern computer science." Lady Ada Augusta, disciple of charles Babbage, after his demise developed several programs for performing mathematical calculations on Analytical engine. She is considered as the first programmer in history and has to her credit a computer language called ADA named after her.

Dr. Herman Hollerith :
In 1890AD, Herman Hollerith designed a system to record census data. The information was stored as holes in punched cards, which were interpreted by machines with electrical sensors.

Mark-I
In 1944 Howard Aiken completed mark I. It was a electromecharical computer which was 51ft. long, 8 ft. high, 3 ft wide and consisted. Of 18000 vacuvm tubes. This consisted of 7 lakh 50 thousand parts and 500 miles long wire.

Electronic Numerical integration and calculator (ENIAC) – 1946 A.D.
ENIAC was the first electronic computer developed by John Mauchly and John presper Eckret in 1946. It could do 5000 additions per second. It was extremely huge, used 19,000 vacuum tubes, occupied an area of 150 sq. metres, weighed about 30 tons and required about 130 kw of power.

EDVAC:
In 1952, Electronic Discrete Variable Automatic Computer (EDVAC) was developed by John Mauchly and John Presper Eckert with the help of A. Burks and Neumann. This machine was used to store the data and information as well as the instructions.

- John Mauchly and Eckert founded their own company in 1946 and began to work on Universal Automatic Computer (UNIVAC): general purpose commercial computer in 1951. It was the first commercially used electronic computer in the world.
- In 1958, the first computer to use the transistor as a switching device, the IBM 7090, was introduced.
- In 1964, the first computer to use Integrated circuits (IC), the IBM 360 was announce.
- In 1975, the first microcomputer, the Altair, was introduced. In the same year, the first Supercomputer, the Cray-1 was announced.

RISC/ CISC

RISC (Reduced Instruction Set Computer) processors are designed for speeding up the processing power of the computer making the chip as simple as possible so that it uses less space and shorter design cycle. It is possible to use the technique of pipelining using RISC processors which gives the immense processing power.

The advantages of RISC processors are as follows:

1. Speed: Due to simplified instruction set RISC processors are 2 to 4 times faster.
2. Simpler hardware- Because of simpler instruction set the RISC processor uses much less chip space, as a result extra function are also placed in the same chip.
3. Shorter design cycle- Because of simple hardware and less instructions per task, the RISC processor uses very short machine cycles.

CISC (Complex Instruction Set Computer) processors use microcode, build rich instruction sets and build high-level instruction sets and these were the building blocks until the late 1980s and is still in major use today.
Some characteristics are:

1. Complex instruction-decoding logic, driven by the need for a single instruction to support multiple addressing modes.
2. A small number of general purpose registers.
3. Several special purpose registers.

Advantages of CISC:

1. Microprogramming is as easy as assembly language to implement, and much less expensive than hardwiring a control unit.
2. The ease of microcoding new instructions allowed the designers to make CISC machines upwardly compatible, i.e. a new computer could run the same programs as earlier computers.
3. As each instruction became more capable, fewer instructions could be used to implement a given task.
4. Because microprogram instruction sets can be written to match the constructs of high-level languages, the compiler does not have to be complicated.

The disadvantages of CISC:

1. The instruction set and chip hardware became more complex with each generation of computers.

2. The instructions set were lengthy and took more time to execute, slowing down the overall performance of the machine.
3. Many specialized instructions aren't used frequently enough to justify their existence.

1.5. Number System:
Refer to the exercise covered during the class.

Memory Units:
Memory units are the internal storages areas in a computer. These are in the form of chips. Usually we classify the computer's memory into two category: RAM and ROM.

RAM (Random Access Memory):
This is the main memory of the computer. This is also found in other devices like printers. This memory holds data as long as the electricity is supplied and therefore referred to as volatile memory. There are two basic types of RAM: (i) Dynamic RAM (DRAM) and (ii) Static RAM (SRAM).
Dynamic RAM needs to be refreshed thousands of times per second. Static RAM needs to be refreshed less often, which makes it faster; but it is more expensive than dynamic RAM.

ROM (Read Only Memory)
It is a non-volatile memory. The data is prerecorded in ROM. The program stored in ROM is known as Firmware and is programmed by the manufacturer. Once data has been written onto a ROM chip, it cannot be removed and can only be read. Most personal computers contain a small amount of ROM that stores critical programs such as the program that boots the computer.

1.7. Auxiliary Storage Units:
Hard Disk:

- A Hard disk is internal hardware which stores and provides access to large amounts of information.
- Hard disks have much greater data capacity and are much faster to use than floppy disks.
- Usually, it is a fixed disk, permanently sealed in the drive.
- Most new computers include an internal hard disk that contains several gigabytes or terabytes of storage capacity.
- The head of hard disk that reads the data floats over the hard disk's surface, while the head of the floppy disk touches the disk's surface while reading or writing data.
- Hard disk is a flat, circular, rigid plate with a magnetizable surface on one or both sides of which data can be stored.
- Hard disks are rigid aluminum or glass disks about 3.5" in diameter in a personal computer, and smaller in a laptop.
- Data is transferred magnetically by a read/write head.
- A hard disk is made of metallic disk coated with metallic oxide on both sides.
- To increase the storing capacity, several disks (platter) are packed together and mounted on a common drive to form a disk pack.
- A hard disk can have more than 1000 tracks per surface and contain 17 sectors per track.

Optical Disk

- An emerging technology that many expect will have a profound impact on mass storage strategies in the 1990s is the Optical Disk.
- With this technology becomes laser beams to wrote and read data at incredible densities.
- Thousand of times finer than the density of a typical magnetic disk.
- Data are placed onto optical disks with high-intensity laser beams that burn tiny holes into the disk's surface.
- Optical disk systems have recently started to become widely used on microcomputer systems.
- So, it is a storage medium from which data is read and to which it is written by lasers. Store much more data in portable magnetic media.
- There are three basic types of optical disks.
- CD-ROM (compact disk read only memory)
- WORM (write once read many)
- ERASABLE
- These three are not compatible with one another.
- WORM (Write-one Read-many)
- With a WORD disk, you can write data, but only once and then you can read number of times.
- ERASABLE Optical (rewritable & erasable)
    - Can be read to, written to and erased just like magnetic disk.
- CD-R (Compact Disk Recordable)
- CD-RD (Compact Disk Rewritable)
- DVD (Digital Versatile Disc)
    - Initial storage capacity of 4.7GB digital information on a single sided, single layer.

- Diameter & thickness is same as CD-ROM.

Magnetic Tape:

- Magnetic tape is a plastic tape with a magnetic surface for storing data as a series of magnetic spots.
- Magnetic tape has been one of the most prominent secondary storage alternatives.
- Magnetic tape is the most commonly used sequential access secondary storage medium.
- It is available in the form of cassettes, real & cartridges.
- Among these three, Reels are the most popular storage secondary media.
- Magnetic tape is a plastic ribbon coated on one side with iron oxide that can be magnetized

Floppy Disk:

Floppy disk is a soft magnetic disk. Floppy disks are being replaced by pen drives now a days. Foppy disks are slower to access than hard disks and have very less storage capacity.

1.8. Input Devices (covered in class discussion, Left as Assignment)
1.9. Output Devices (covered in class discussion, Left as Assignment)

Input Devices:

Keyboard, Mouse, Scanners, Digital cameras, Microphone, Digitizing Tablet (used in GIS application), Bar Code Reader, Stylus pen, Joystick, Track ball, OCR (Optical Character Reader), MICR (Magnetic Ink Character Reader etc.

Output Devices:

1. Monitor: CRT monitors, LCD monitors
2. Printers: Impact printers (Dot-matrix Printer, Daisy-wheel Printer), Non-Impact printer (Ink-jet printer, Laser Printer, LCD and LED Printers)
3. Plotter : Plotter is a device that draws pictures on paper based on commands from a computer. Plotters differ from printers in that they draw lines using a pen. In general, plotters are more expansive than printers.
4. Sound Cards and Speakers.

Unit 2: Computer Software and Software Development

2.1. Software:

- Software is a collection of set of programs, which are used to execute all kinds of specific instruction. It consists of a number of machine instructions, array in a specific order to perform a particular task. Software is used to describe all the programs and its associated documents which run on a computer. So, a computer needs both software and hardware for its proper functioning.
- Software means computer instructions or data. Anything that can be stored electronically is software.
- Firmware are software (programs or data) that has been permanently written onto read-only memory (ROM).

- All software falls into **two general types or categories**: System Software and Application Software. System software consists of low-level programs that interact with the computer at very basic level. This includes operating systems, compilers, and utilities for managing resources. On the other hand, application software includes database programs, word processors, and spreadsheets.

Operating Systems:

Operating systems are the most important programs that run on a computer. Every general-purpose computer must have an operating system to run other programs. Operating systems perform basic tasks, such as recognizing input from the keyboard, sending output to the display screen, keeping tracks of files and directories on the disk and controlling peripheral devices such as disk drives and printers. Most commonly used operating systems include Microsoft Windows, DOS, Xenix, Mac OS, OS/2, UNIX, MVS, etc.

Compilers and Interpreters:

Compiler is a program that translates source code into object code. The compiler takes the entire piece of source code and collects and recognizes the instructions. In contrast, the interpreter analyzes and executes each line of source code in succession, without looking at the entire program. The advantage of interpreters is that they can execute a program immediately but compilers requires some time before an executable program emerges. However, programs produced by compilers run much faster than the same programs executed by an interpreter.

Device Drivers:
Device drivers are the software to run the particular peripheral devices like printers, wireless mouse, modems, Ethernet cards e.t.c. Most of the peripheral devices attached with the CPU needs device drivers to carry out the functions. Mostly we get device drivers at the time of buying these devices. For example we get the printer driver when we buy the printer. Device driver lets the CPU know the type of hardware and the type of instruction for a particular device attached with it.

Application Software:

Word Processors:
A word processor is a program that makes us possible to perform word processing functions. We can create, edit, and print documents using word processors. We have many features that helps us to prepare a fine document. Some of the important feature of word processors are: editing, spelling checking, page setup, paragraph alignments, merging documents, typing in columns etc.
MS-Word is the word's most popular word-processor. Although every word processor provides almost the same features, MS-Word is most flexible to work with. It is used to write documents or letter. A file in Ms-Word is called a document. When a file is saved, MS-Word attaches the extension .doc to the file.

Spreadsheets:
A spreadsheet is a table of values arranged in rows and columns. Each value can have a predefined relationship to the other values. If one value is changed, others values need to be changed as well.
Spreadsheet applications are computer programs that let you create and manipulate spreadsheets electronically. In a spreadsheet application, each value sits in a cell. We can define what type of data is in each cell and how different cells depend on one another. The relationships between cells are called formulas, and the names of the cells are called labels. Once we have defined the cells and the formulas for linking them together, we can enter the data. We can then modify selected values to see how all the other values change accordingly. What-if analysis makes the complex decision-making a very easy process. MS-Excel is one of the most popular spreadsheet application.

Multimedia applications:
Multimedia applications makes us possible to run audio and video files. This application recognizes the digital signals and provides necessary signal to output devices and movie signals to the monitor and audio to the audio devices. Along with the video data we also get the text information about the file we are running. In windows operating system Windows Media Player is a good option to play the multimedia files.

Presentation Graphics:
Presentation Graphics enable users to create highly stylized images for slide shows and reports. The software includes functions for creating various types of charts and graphs and for inserting text in variety of fonts. Most systems enable us to import data from a spreadsheet application to create the charts and graphs. Presentation graphics is often called business graphics. Some of the popular presentation graphics software are Microsoft Powerpoint, Lotus Freelance Graphics, Harvard Presentation Graphics, etc.

(See: Image Processors: Paint Programs, Draw Programs and Image Editors in the book. Fund. of Info. Tech, Alexis Leon, Page 10.4,105)

Utility Software:
Utility is a program that performs a very specific task, usually related to managing system resources. Antivirus software, Disk Partition tools are the examples of utility software.

2.2. Operating System:

OS is an important system software package found in every computer systems. It is a set of programs that controls and supervises a computer system's hardware and it provides services to computer users. It permits the computer to supervise its own operations by automatically calling in application programs and managing data needed to produce the output desired by users. OS is an interface between the user and the computer. OS perform basic tasks, such as recognizing input from the keyboard, sending output to the display screen, keeping track of files and directories on the disk, and controlling peripheral devices such as disk drives and printers. OS, as a resource manager keeps track of who is using which resource, to grant resource requests, to account for usage and to mediate conflicting requests from different programs and users.

Functions of an Operating System:

Job Management:OS manages the jobs waiting to be processed. It recognizes the jobs, identifies their priorities, determines whether the appropriate main memory and secondary storage capability they require is available, and schedules and finally runs each job at the appropriate moment.

Batch Processing: Data are accumulated and processed in groups. The printing tasks in the printer are also done in groups. Most of the tasks of OS are grouped and performed one by one.

On-line Processing: In on-line processing, data are processed instantaneously. Most on-line operating systems have multi-user and multitasking capabilities. Now a days we can access the data from the remote sites using on-line processing.

Data Management : OS manages the storage and retrieval of data. As the system software handles many of the details associated with this process, such details are not a primary concern for users or programmers writing application programs.

Virtual Storage: Using this method the capacity of main memory increases without actually increasing its size. This is done by breaking a job into sequences of instructions, called pages or segments, and keeping only a few of these in main memory at a time; the remaining are kept on secondary storage devices. Thus, relatively large jobs can be processed by a CPU.

Input/ Output Management: OS manages the input to and output from a computer system. This applies to the flow of data among computers, terminals, and other devices such as printers. Application programs used the operating system extensively to handle input and output devices as needed.

Function of operating system can be further listed as follows:

- user interface implementation
- share hardware implementation
- allows users to share data.
- Prevent users from interfering with one another.
- Scheduling resources among users.
- Facilitate input/output
- Facilitate parallel operations
- Organize data for secure and rapid access.
- Handle network communications

Classification of Operating System:

Multi-user: These systems allow two or more users to run programs at the same time. Some OS permit hundreds or even thousands of concurrent users. The operating systems of mainframes and minicomputer are multi-user systems. Examples are MVS, UNIX, etc. Another term for multi-user is time-sharing.

Multiprocessing: It refers to a computer system's ability to support more than one process (program) at the same time. This system allows the simultaneous execution of programs on a computer that has several CPUs. Mainframe, Supercomputers have more than one CPU.

Multitasking: This system allows a single CPU to execute what appears to be more than one program at the same time when, in fact, only one program is being executed. In multitasking, only one CPU is involved, but it switches from one program to another so quickly that it gives the appearance of executing all of the programs at the same time.

Multithreading: Multithreading allows different parts of a single program to run concurrently. Multithreading is the ability of an OS to execute different parts of a program, called threads, simultaneously.

Real-time: These operating systems are system that respond to input immediately. It allows a computer to control or monitor tasks performed by other machines and people by responding to the input data in the required amount of time.

2.3. Programming Languages:

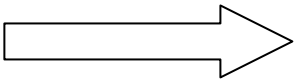
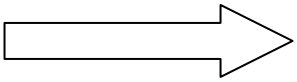
Programming Language is a set of rules that provides a way of instructing the computer to perform certain operations. Programming languages are said to be lower or higher, depending on whether they are closer to the language the computer itself uses or to the language that people uses.

We can study the programming languages under five levels (or generations) of language:

1. Machine Languages / First generation Languages
2. Assembly languages / 2nd Generation Languages
3. Procedural Languages/ Third-generation Languages
4. Problem-oriented Languages / Fourth generation Languages
5. Natural Languages / Fifth Generation Languages.

The characteristics & trends of these five languages are summarized in the following table:

	First	Second	Third	Fourth	Fifth

	Generation	Generation	Generation	Generation	Generation
Trend : Towards Conversational natural programming language 					
Software Trends	User-written programs	Packaged Programs	Operating Systems	Database Management Systems	Natural Languages
	Machine Language	Symbolic Language	High-Level Languages	Fourth Generation Languages Micro computer Pacages	Natural Languages Multipurpose graphic-interfaced packages.
Trend: Towards easy to use multi-purposed application packages 					

Computer programming languages are developed with the primary objectives without the need to know in detail the internal structure of the computer.

Machine Language:
A program written using the binary codes specified for a processor's operations and absolute binary addresses is known as machine language of the computer. The programs written in Machine languages are machine dependent. Programming using Machine Language is very tedious and time consuming. The programmer must keep track of a tremendous amount of detail and one must understand the technical operations of the computer. Programming in machine code has one advantage over programming at other language levels - its execution is very fast and efficient because the computer can accept the machine code as it is.

Assembly Language:
Assembly languages are also know as symbolic languages as they use abbreviations or mnemonic code which replace the 0s and 1s of machine language. An assembly language has a one-to-one correspondence with the machine language of a computer and is thus machine dependent. Assembly languages are more standardized and easier to used than machine languages. Though more standardized than machine languages, assembly languages are still machine dependent.

High Level Languages:
High level languages are the programming languages that use predefined common English words and clauses to write a program. They assisted programmers by reducing further the number of computer operations details they has to specify, so that they could concentrate more on the logic needed to solve the problem. We can classify the high level languages into three levels:

- Procedural-oriented or third generation
- Problem-oriented or fourth generation
- Natural or fifth generation

Procedural oriented Languages
General-purpose programming languages are called procedural languages or third generation languages. They are the languages such as Pascal, BASIC, COBOL, and FORTAN, C which are designed to express the logic, the procedure, of a problem. Because of their flexibility, procedural languages are able to solve a variety of problems. More focus is given to module while programming in Procedural-oriented languages.

Problem-oriented Languages:
Fourth-generation languages are also known as problem-oriented languages. These are the languages designed to solve specific problems or develop specific applications by enabling one to describe what he wants rather than step-by-step procedures for getting there. Thus the problem to considered most rather than procedures so as to get the solution to the particular problem. Visual Basic and PERL are Problem-oriented languages. These languages may be categorized into several kinds of application development tools:

- Personal computer application software
- Query languages and report generators
- Decision support system and financial planning languages
- Application generators

Natural Languages:
Natural languages are very near to the human conversational language. Natural languages are still in the development stages. These languages will be used most in the areas of artificial intelligence and expert systems. Natural languages have two characteristics:

- They are designed to make the connections that humans have with computers more natural – more humanlike.
- They are designed to allow the computer to become "smarter" – to actually simulate the learning process by remembering and improving upon earlier information.

Two popular natural languages are LISP and PROLOG.

Compilers and Interpreters:

Compiler is a program that translates source code into object code. The compiler takes the entire piece of source code and collects and recognizes the instructions. In contrast, the interpreter analyzes and executes each line of source code in succession, without looking at the entire program. The advantage of interpreters is that they can execute a program immediately but compilers requires some time before an executable program emerges. However, programs produced by compilers run much faster than the same programs executed by an interpreter.

2.4. General Software Features and Trends:

Introduction:

Now a days software projects are becoming more and more complex – in size, sophistication, and technologies used. Most software products are used by huge number of people, not only that, these software support different national languages and come in different sizes and shapes – desktop, standard, professional, Enterprise Resource Planning (ERP) packages and so on. Almost all application software products (like word processors, ERP packages) support more than on hardware and/or software platform. For example, we have web browsers for the PC and Mac; we have database management systems that run on MVS, UNIX, Windows NT, Linux and so on. The competition and the advancements in technology are driving software vendors to include additional functionality and new features to their products– just to stay in business.

The Information Technology is revolutionizing the way we live and work. The digital technology has given mankind the ability to treat information with mathematical precision, to transmit it at very high accuracy and to manipulate it at will.

Features:

The following are the trends that play a vital role in the development of software products:

Ease of use: The software systems are applications are becoming more and more easy to use. Software developers and system analysts are concerned for ensuring that the software they develop are user-friendly than their competitor's products. The user interfaces are more intuitive, the error messages are more descriptive, there is context sensitive help, there are wizards and templates to help the user when one encounters a problem.

Graphical User Interface (GUI): Today's software applications and products provide users with intuitive, graphical and easy-to-use interfaces. Now the users do not have to remember the cryptic system commands or shortcut keys that were a must in the character based era. Now almost any tasks can be accomplished by a mouse click. For example, in a DOS environment, to copy a file one needs to know the command for copying files, its exact syntax and so on, whereas in the Windows environment, you just have to drag the files you want to copy from the source to destination.

Requirement of more powerful hardware: Because software vendors are incorporating more and more features into their products these software need more and powerful machines to run. They need more main memory, more secondary storage, and faster and powerful processors. It is also not very difficult to go for new powerful computers as the price of computers are decreasing day by day.

Multi-platform capability: Today's software applications are not developed for just one platform. Most of the software applications supports multiple platforms– both hardware and software platforms. There are software applications that support hardware platforms ranging from mainframes to PCs and different software platforms like MVS, Solaris, AIX, UNIX, Windows and so on. Database like IBM's DB2 Universal is available for a variety of hardware and software platforms. Another important feature of today's software application is that they support multiple languages and multiple currencies. Many vendors are providing their application in many languages like English, Arabic, Japanese, Chinese and so on.

Network Capabilities: Network computers are becoming popular as they can work with minimal memory, disk storage and processor power. These computer are connected to a network especially the Internet. The idea behind network computers is that many users who are connected to a network don't need all the computer power they get from a typical personal computer. Instead, they can rely on the power of the network servers. As the popularity of network computers increase, the demand for software that can run these computers are increasing and the software applications of today and tomorrow will have that capability.

Compatibility with other software: Now a days most of the software products are compatible with each other. For example, we can import HTML documents and other text documents into a Microsoft Word document. Also as newer versions of software are released, most vendors maintain backward compatibility i.e. compatibility with earlier versions. These two features– backward compatibility and compatibility with other products make it easier for the users, as they can choose the application they want and still use the old files they created using other applications or using older versions of the same application.

Object Linking and Embedding: We have mechanism to interact with other software systems. One of the method to integrate external tolls into an application is using the Object Linking and Embedding (OLE) architecture to link or embed a component from another application running on the computer. OLE is a compound document standard developed by Microsoft Corporation. It enables us to create objects with one application and then link

or embed them into a second application. This embedding and linking of components enables applications to share components.

Groupwork capabilities: Technologies, which support collaboration, are in greater demand today than ever before. Therefore, vendors are integrating collaborating technologies into their products. Distributed workforces, information overload, and getting products to market as quickly as possible are just a few of the motivational aspects pushing collaboration technology development.

Mail Enabling: The mail enabling of an application is the process through which email is gradually replacing many of the single purpose applications now used on personal computers. In its simplest form, a mail-enabled application is a Windows program that has a 'Send' command in its 'File' menu. For example, the Microsoft Word 2000 has a 'Send To' menu item in the 'File' menu where we can choose to send the document to a mail recipient, fax recipient and son on.

Web Enabling: With the ever-increasing popularity of Internet and the amount of information that is available on the net, most software application are now web-enabled. Web enabling helps the user in many different ways. During installation, most of the applications will automatically connect the Internet and to the vendor's web site and will register their products (earlier one had to fill in a paper form and mail or fax it to the vendor).

Unit 3: Database Management System

3.1. Data Processing:

Data vs Information:

Data refers to a collection of natural phenomena descriptors, including the results of experience, observation or experiment, or a set of premises. This may consist of numbers, words, or images, particularly as measurements or observations of a set of variables.

Raw data are numbers, characters, images or other outputs from devices to convert physical quantities into symbols, in a very broad sense. Such data are typically further processed by a human or input into a computer, stored and processed there, or transmitted (output) to another human or computer. Raw data is a relative term; data processing commonly occurs by stages, and the "processed data" from one stage may be considered the "raw data" of the next. After processing of data we get information. Information is then useful for decision making. Information is nothing but refined data, data that have been put into meaningful and useful context and communicated to a recipient who uses it to make decisions.

File Processing:

1. Sequential file processing: It stores and access records in sequence. Such processing can be accomplished either by using tape storage or disk storage. To perform sequential file processing, records are sorted before they are processed. Sequential file processing is used in situations where data ca be processed in batches and where a substantial portion of the master file is changed with processing of each batch. Payroll processing is a classic example of sequential processing.
2. Direct-access file processing: There are many ways of organizing a file for direct-access. First, the file must be stored on a direct-access device like a disk, so that the records need not be processed in sequence. Second, some means must be developed for determining the location of a particular record. Indexes are on common means.

Database Processing:

A database is a self-describing collection of integrated records because it contains, a part of itself, a directory, or dictionary of its contents. The records are integrated because a database can contain multiple files (usually called tables in databases processing), and records within those tables are processed by their relationship to one another.

3.2. Database Management System:

Introduction:

A collection of programs that enables us to store, modify, and extract information from a database. There are many different types of DBMSs, ranging from small systems that run on personal computers to huge systems that run on mainframes. The following are examples of database applications:

- computerized library systems
- automated teller machines
- flight reservation systems
- computerized parts inventory systems

From a technical standpoint, DBMSs can differ widely. The terms *relational*, *network*, *flat*, and *hierarchical* all refer to the way a DBMS organizes information internally. The internal organization can affect how quickly and flexibly you can extract information.

Requests for information from a database are made in the form of a *query*, which is a stylized question. For example, the query

SELECT ALL WHERE NAME = "SMITH" AND AGE > 35

requests all records in which the NAME field is SMITH and the AGE field is greater than 35. The set of rules for constructing queries is known as a *query language*. Different DBMSs support different query languages, although there is a semi-standardized query language called *SQL (structured query language)*. Sophisticated languages for managing database systems are called *fourth-generation languages*, or *4GLs* for short.

The information from a database can be presented in a variety of formats. Most DBMSs include a *report writer program* that enables you to output data in the form of a report. Many DBMSs also include a graphics component that enables you to output information in the form of graphs and charts.

Quality of Information:

Accuracy: This means the data must be accurate. The data must be clear and accurately reflects the meaning of data on which it is based. It conveys an accurate picture to the recipient and may require a graphical presentation rather than a table full of numbers.

Timeliness: The recipients must the get the data within the needed time frame. For example, yesterday's newspaper today or stock quotes a day or two after are normally of little value.

Relevancy: This means the information for a particular person must be useful. Information relevant for one person may not be relevant for another.

Significance of DBMS :

- **Reduction in data redundancy** : Redundancy can be controlled using DBMS. Thus, space is efficiently used. The existing applications can share the data in the database.Reduces problem of inconsistencies in stored information, e.g. different addresses in different departments for the same customer
- **Maintenance of data integrity and quality** : Integrity means that the data in the database is accurate. Centralized control of the data helps in permitting the administrator to define integrity constraints to the data in the database.
- **Data are self-documented or self-descriptive:** Information on the meaning or interpretation of the data can be stored in the database, e.g. names of items, metadata
- **Avoidance of inconsistencies** : Reducing the redundancy also avoids the inconsistency of data. Data must follow prescribed models, rules, standards
- **Security restrictions :** With complete authority over the operational data, the database administrator can ensure that the only means of access to the database is through proper channels. He can define authorization checks to be carried out whenever access to sensitive data is attempted. Different checks can be established for each type of access (retrieve, modify, delete, etc). to each piece of information in the database.

Characteristics of Data in a database:

- **Shared** : Date in a database are shared among different users and applications.
- **Persistence**: Data in a database exist permanently in the sense, the data can live beyond the scope of the process that created it.
- **Validity/Integrity/Correctness**: Data should be correct with respect to the real world entity that they represent.
- **Security:** Data should be protected from unauthorized access.
- **Consistency:** Whenever more than one data element in a database represents related real-world values, the values should be consistent with respect to the relationship.
- **Non-redundancy**: No two data items in a database should represent the same real-world entity.
- **Independence:** The three levels in the schema (internal, conceptual and external) should be independent of each other so that the changes in the schema at one level should not affect the other levels.

Database Management System and its services:

- **Transaction Processing :** A transaction is a sequence of database operations that represent a logical unit of work. It accesses a database and transforms it from one state to another. A transaction can update a record, delete one, modify a set of records, etc. When the DBMS does a 'commit', the changes made by the transaction are made permanent. If you don't want to make the changes permanent you can roll back the transaction and the database will remain in its original state.
- **Concurrency Management:** It is the database management activity of coordinating the actions of database manipulation process that operate concurrently, access shared data and can potentially interface with each other. The goal of an idea concurrency management mechanism is to allow concurrency while maintaining the consistency of the shared data.
- **Recovery:** The objective of recovery in a database is to ensure that the aborted or failed transactions do not create any adverse effects on the database or other transactions. Recovery mechanisms in a DBMS make sure that the data is returned to a consistent state after a transaction fails or aborts. Recovery is very much related to concurrency in the sense that , the more the concurrency, the more is the chance of an aborted transaction can affecting many other transactions.
- **Security:** It refers to the protection of data against unauthorized access. Security mechanism of a DBMS make sure that only authorized users are given access to the data in the database. The level of access for each user and the operations that each user can perform on the data will be monitored and controlled by the DBMS depending on the access privileges of the user.
- **Language Interface:** The DBMS provides support languages used for the definition and manipulation of the data in the database. The data structures are created using the data definition language commands. The data

manipulation is done using the data manipulation commands. By providing language support for data definition and manipulation, the DBMS create an environment where the users can do their jobs without worrying about the physical implementation.

- **Data catalog:** Data catalog or Data Dictionary is a system of database that contains the description of data in the database (metadata). It contains information about data, relationships, constraints and the entire schema that organize these features into a unified database. The data catalog can be queried to get information about the structure of the database.
- **Storage Management:** The DBMS provides a mechanism for management of permanent storage of the data. The internal schema defines how the data should be stored by the storage management mechanism and the storage manager interfaces with the operating system to access the physical storage.

Types of Database Management System:

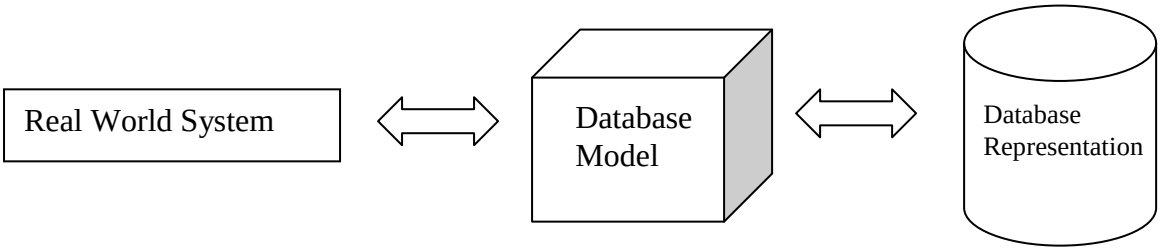
Hierarchical Model:
Hierarchical Database model is one of the oldest database models. The hierarchical model assumes that a tree structure is the most frequently occurring relationship.
In this model data follow the hierarchical model. Rather than one record type (flat file), a business has to deal with several types which are hierarchically related to each other , e.g. company has several departments, each with attributes: name of director, number of staff, address
Certain types of geographical data may fit the hierarchical model well , e.g. Census data organized by state, within state by city, within city by census tract. The database keeps track of the different record types, their attributes, and the hierarchical relationships between them. The attribute which assigns records to levels in the database structure is called the key (e.g. is record a department, part or supplier?)

Network Model:
The Network Model structures in a network connecting every nodes. The network model was evolved to specifically handle non-hierarchical relationships. The network model has greater flexibility than the hierarchical model for handling complex spatial relationships.

Relational Model:
In an ***RDBMS***, a database is considered to be a collection of interrelated data and programs. The data in a database has to be related. For example, in a College Management System, which takes care of maintaining students' records, storing data such as the salary details of the teachers would be inappropriate and considered unrelated. On the other hand the marks details, fee details and other personal details of students would be considered 'interrelated' data. The programs in a database perform the role of manipulating this data. A database that is designed on the concept of '***relational***' model is called a *'Relational Database Management System'*.
A relational database is made up of set of ***relations*** or ***tables***. These tables store user data as well as system data. Each of these relations is made up of ***attributes*** (fields) and ***tuples*** (records).
Structured Query Language (SQL) serves as a uniform interface for users providing a collection of standard expressions for storing and retrieving data.

Object-oriented Model
Object-oriented model represents an entity as a class. A class represents both object attributes as well as the behaviour of the entity. The objects that falls in the similar class have the similar behaviour as mentioned in the class and each object may have special attributes to distinguish itself from other object.

3.3 Database Design:
Database Design Process:



- To develop a good design, one has to understand the meaning of information and the intended use of the information's stored representation with the computer system. Once we develop the understanding and have identified the use of information in the application, we can determine how much and what kind of information we require.
- We could also determine in what format these information should be captured and represented in the computer system or database. During this phase, it will become clear that what data entities represent information redundancies and which entities are critical, which are useful and which are not related to the application.

- It is important to collect and analyze the static and dynamic information available about the real world application before starting the database design. For evolving a good database design, it is important that one uses a model or a database design model.

Data Normalization:
Normalization is the process of building database structures to store data. Normalization is a formal process of developing data structures in a manner that eliminates redundancy and promotes integrity. Data normalization is a corner stone of the relational theory.

Keys:
A key uniquely identifies a row in a table. There are two types of keys: intelligent keys and non-intelligent keys. (Covered in Class Discussion).

First Normal Form
Second normal Form
Third Normal Form.
(Already Discussed : Refer to the class note).

Data Warehouse:
To execute queries efficiently on diverse data, companies have built data warehouses. Data warehouse gather data from multiple sources under a unified schema, at a single site. Thus, they provide the user a single uniform interface to data.
A data warehouse is a large repository (or archive) of data which comes from operational sources and has four properties:
(1) Non volatile
(2) Time varying
(3) Subject oriented
(4) Integrated.
Once gathered, the data are stored for a long time, permitting access to historical data, making decision-support queries easier to write. Moreover, by accessing information for decision support from a data warehouse, the decision maker ensures that online transaction processing systems are not affected by the decision support workload.

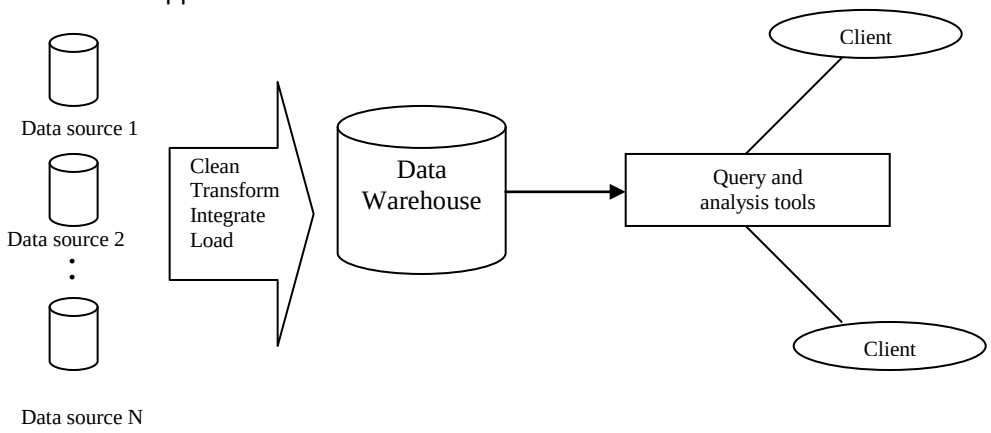


Fig: Typical architecture of a data warehouse.

Properties of Data warehouse:

- Subject-oriented: A data warehouse is organized around major subjects, such as customer, supplier, product, and sales. Data warehouses typically provide a simple and concise view around a particular subject issues by excluding data that are not useful in the decision support system.
- Integrated: A data warehouse is usually constructed by integrating multiple heterogeneous sources, such as relational databases, flat files, and on-line transaction records.
- Time-dependent: that is, containing information collected over time which implies there must always be a connection between the information in the warehouse and the time when it was entered.
- Nonvolatile: Data once loaded in the data warehouse is not going to change in the future. This means that a data warehouse will always be filled with historical data.

Data Mining:
It is a process of semi automatically finding of useful information from large repositories which was not previously found.
Simply stated, data mining refers to extracting or "mining" knowledge from large amounts of data.

A broad view of data mining functionality: Data mining is the process of discovering interesting knowledge from large amounts of data stored either in databases, data warehouses, or other information repositories.

Data mining involves an integrating of techniques from multiple disciplines such as database technology, statistics, machine learning, high-performance computing, pattern recognition, neural networks, data visualization, information retrieval, image and signal processing, and spatial data analysis.

By performing data mining, interesting knowledge, regularities or high-level information can be extracted from databases and viewed or browsed from different angles.

The discovered knowledge can be applied to decision making, process control, information management, and query processing.

Therefore, data mining is considered one of the most important frontiers in database systems and one of the most promising interdisciplinary developments in the information industry.

Data mining is a step in the Knowledge Discovery Process.

Other notes for reference:

Computer System

Computer system is the combination of input unit, central processing unit, Storage unit and the output unit. There are input devices such as keyboard and mouse. Monitor and printer are output devices. When these peripherals or devices are connected with the computer to achieve a common objective, which will make it a complete set, and referred as "Computer System".

Elements of Computer System:

(a) Input

Data and instructions are input from devices like keyboard, mouse etc. Other input devices are card reader, bar code reader, OCR, scanner, microphone, web camera, touch pad etc.

(b) Processor:

The processor is also known as CPU (Central Processing Unit). It consists of main memory, control unit and arithmetic and logic unit. It performs operations on data input and returns the result to the output devices. The control unit fetches instructions from main storage, interprets them, and issues the necessary signals to the components. It controls all the hardware operations. The ALU performs all the arithmetic and logic operations. The main memory holds the program instructions for the program to be executed, the input date to be processed and the intermediate results of any processing. Ram is an example of Main memory.

(c) Secondary Storage

This unit is the supplement to main memory. Some popular storage devices are Hard Disk, Floppy Disk, Optical disks etc.

(d) Output

Main memory gives the result to output devices. The output may be in the form of display or in printed form. Monitor and printer are common output devices.

Microprocessor:

The microprocessor is built onto a single piece of silicon, known as a wafer or chip. Its size is about 0.5 cm square. It is the developed form of integrated circuits manufactured using semiconductor technology. It serves as the CPU in our general computers. Microprocessors evolved rapidly due to

- Miniaturation of transistors
- Decreasing distance between transistors on the chip (decreasing line width)
- Improved conductivity (flow) of electricity
- Improved instruction sets programmed into the chip.

Functions of Processor (CPU):

The CPU is the heart of the entire computer system. It performs computations, executes instructions and transfer information to all the parts of a computer. The functions of the processor are listed below:

(a) It controls the main memory for storing intermediate data and instructions.

(b) It controls the sequence of operations.

(c) It give commands to all parts of the computer system and hence controls all the components.

(d) It carries out processing i.e. computations on data.

- Control Unit

- Access program instructions
- Decode (interpret) instructions
- Control flow of data throughout system
- Data flows through paths called buses

❖ Arithmetic-Logic Unit
- Perform computations on data
- Perform comparisons on data

❖ Registers
- High speed storage areas
- Hold data and instructions

❖ Primary Storage (Main Memory)
- Stores instructions from programs
- Stores data to be processed