adda 247 publications

# Ace IT OFFICER

## Professional Knowledge

**eBOOK**

Useful for IBPS SO IT, IBPS RRB SO IT, SBI SO IT Exams


**SECOND EDITION**

### THIS EDITION INCLUDES

- 1200+ Questions
- Covers Networking, DBMS, Operating System, Data Structure, Programming Languages and much more
- Practice Sets with detailed solution.

# CONTENTS

# 1 Software & Hardware

## 1. Introduction

Computer is an electronic machine which can perform arithmetic and logical operations according to the instructions given by the user.
Modern computer is based on analytical engine developed by Charles Babbage. He is considered as father of modern computer. Basic architecture of electronic computer is given by Johan Von Neumann. According to this architecture computer consists of following components.

1. Memory System which stores input data and result before it is output on the screen
2. Input and Output System like Monitor and keyboard
3. CPU (Central processing Unit) with ALU (Arithmetic Logic Unit) and CU(Control Unit)

Thus, we can say computer is a collection of different sub system. All the processing is done in CPU with the help of ALU and CU. Data in computer is represented by 0 and 1 known as binary form. Single 0 or 1 is known as bit and collection of eight bit is known as byte. Byte is a basic unit to represent data in computer
Computer cannot work by its own, it works according to the instruction given by the user. Collection of instruction in proper sequence to perform some task is known as **Software.** Therefore, we can say that computer is made of basic components -

1) Hardware
2) Software

There are different types of computer which has different speed, cost and application. like Micro Computer, mini Computer, Main Frame Computer and Super Computer.
Micro Computer is slowest, cheapest and used in simple application. e.g. PC, Home Computer laptop. Super Computer is fastest, costliest and use in special application like Weather fore casting, scientific research etc.
Computer can also be classified on the basis of their functions like 1) Analog Computer 2) Digital Computer Analog computer is generally used for measurement and Digital Computer is used for calculation. Most of the modern computers are Digital.

**Hardware :-** Hardware components of the computer consist of all the electronic components and electromechanical devices that comprise the physical devices of computer. The hardware of computer is divided into four major parts:-

1) Central Processing unit(CPU):- It contain an arithmetic and logic unit for manipulating data, a number of registers for storing data, and control unit for fetching and executing instructions.
2) Memory:- It contains data and instruction before execution and after execution. It is call ed random access memory because CPU can access any location randomly without affecting the other location.
3) Input Output Processor(IOP):- It contain electronic circuit for communication and controlling the transfer of information and the outside devices like input output devices
4) Input Output devices:- They are used to give information to the computer and display output like keyboard, printer, visual display unit(VDU).

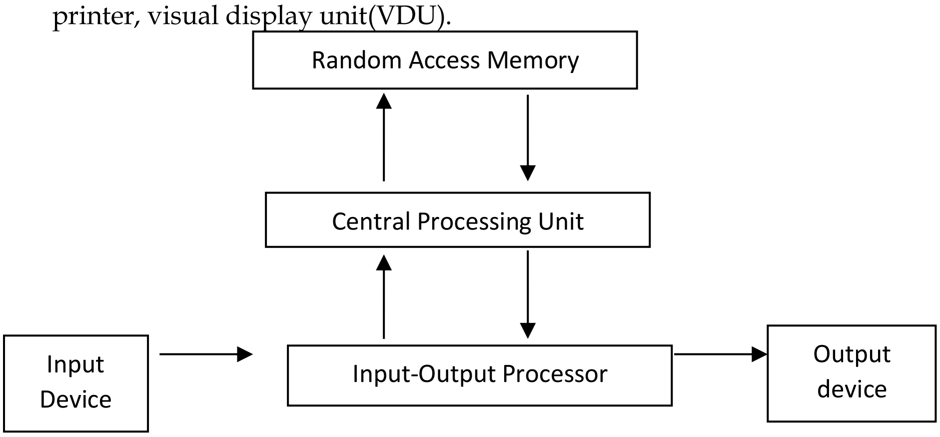



*Figure: Block Diagram Of Digital Computer*

**Logic Gate:** All the data in computer is present in binary form i.e. 0 and 1. All the computer components like RAM, ALU, CU and registered are made of Gates. There are different types of gates which perform different function e.g. AND gated, OR gate NOT Gate, etc. Therefore, we can say that logic gate are the main building blocks of digital computer.

**Memory System:** All the devices which are used to store information in computer is called memory system. Data and instruction is stored in memory before processing i.e. is before going to the CPU. In computer we use different type of memory because some memory are volatile but fast but some are nonvolatile but slow. To maintain the efficiency of computer we use different type of memory. Following is the hierarchy.

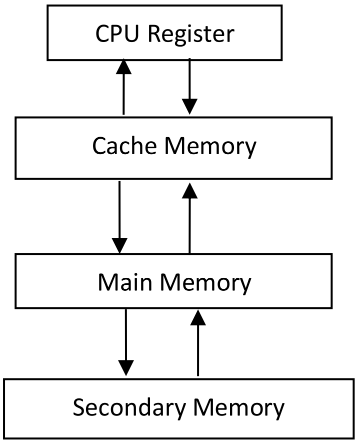


*Figure: Memory Hierarchy*

**CPU Register: -** They are small memory location which reside in CPU, they are fast and small. They are generally used to store data and instruction temporarily.
**Cache Memory: -** It is a fast-used memory between main memory and CPU. It's an electronic memory. It's size is smaller than main memory.
**Main Memory: -** It is also an electronic memory. All the programs before execution are loaded in main memory. Main memory consists of two parts
**a)** **RAM (**Random Access Memory**):-** It is read and write memory. It consists of flip-flop. It is volatile memory. There are different type of Ram 1) SRAM 2) DRAM
**b)** **ROM (**Read only Memory):- It is nonvolatile memory. It stores those programs which are essential to boot the program. It consists of combinational circuit. There are different types of ROM Like PROM,EPROM, EEPROM, Flash memory.
**Secondary Memory: -** Secondary Memory is a permanent storage device. It is nonvolatile memory, which can store data when there is no power.

**The following table highlights the points that differentiate a hardware from a software.**

| Hardware | Software |
|---|---|
| It is the physical component of a computer system. | It is the programming language that makes hardware functional. |
| It has the permanent shape and structure, which cannot be modified. | It can be modified and reused, as it has no permanent shape and structure. |
| The external agents such as dust, heat etc. can affect the hardware (as it is tangible). | The external agents such as dust, heat etc. cannot affect (as it is not tangible). |
| It works with binary code (i.e., 1's to 0's) . | It functions with the help of high level language like COBOL, BASIC, JAVA, etc. |
| It takes in only machine language, i.e., lower level language. | It takes in higher level language, easily readable by a human being. |
| It is not affected by the computer bug or virus. | It is affected by the computer bug or virus. |
| It cannot be transferred from one place to other electronically. | It can transfer from one place to other electronically. |
| Duplicate copy of hardware cannot be created. | A user can create copies of a software as many as he wishes. |

**Basics of CPU(Central Processing Unit)-**Central Processing Unit (CPU) performs all the arithmetic and logical calculations in a computer. The CPU is said to be the brain of the computer system. It reads and executes the program instructions, perform calculations and makes decisions. The CPU is responsible for storing and retrieving information on disks and other media.

**The CPU consists of Control Unit, Arithmetic and Logic Unit (ALU) and register set.**

- **Control Unit:** The control unit issue control signals to perform specific operation and it directs the entire computer system to carry out stored program instructions
- **Arithmetic and Logic Unit:** The ALU is the 'core' of any processor. It executes all arithmetic operations (addition, subtraction, multiplication and division), logical operations (compare numbers, letters, special characters etc.) and comparison operators (equal to, less than, greater than etc.).
- **Register Set:** Register set is used to store immediate data during the execution of instruction. This area of processor consists of various registers.

## 2. Software

As you know, the hardware devices need user instructions to function. A set of instructions which get a single outcome are called program. Many programs functioning together to do a task make a **software**.
For example, a word-processing software enables the user to create, edit and save documents. A web browser enables the user to view and share web pages and multimedia files. There are two categories of software −

- System Software
- Application Software

### 2.1 System Software

Software required to run the hardware parts of the computer and other application software are called **system software**. System software acts as **interface** between hardware and user applications.
System software includes −

- Operating System
- Language Processor
- Device Drivers

**Operating system (OS)** is the software program that manage software and hardware resources of computer. Operating system manages a computer's basic functions like storing data in memory, retrieving files from storage devices, scheduling tasks based on priority, etc.

**Language Processor**
An important function of system software is to convert all user instructions into machine understandable language. When we talk about human and machine interactions, languages are of three types –

**Machine language** − This language is nothing but a collection of 0s & 1s (binary digit) that the machines can understand. It is completely machine dependent.

**Assembly language** − This language introduces a layer of abstraction by defining **mnemonics**. **Mnemonics** are English like words or symbols used to denote a collection of 0s&1s. Assembly level language is **machine dependent**.

**High level language** − This language uses English like statements and is completely independent of machines and uses translator. Programs written in high level languages are easy to create, read and understand. It is also called **source code.**
**e.g.-**Java, C++, Fortran, Pascal etc. .

**Language Translator-** A language translator helps in converting programming language into machine language.
There are three types of language translator−

**Assembler** − Converts assembly level program into machine level program.

**Interpreter** − Converts high level programs into machine level program line by line.

**Compiler** − Converts high level programs into machine level programs at one go rather than line by line.

**Device Drivers**
System software that controls and monitors functioning of a specific device on computer is called **device driver**. Each device like printer, scanner, microphone, speaker, etc. that needs to be attached externally to the system has a specific driver associated with it. When you attach a new device, you need to install its driver so that the Operating System(OS) knows how it needs to be managed.

## 2.2 Application Software

A software that performs a single task and nothing else is called **application software**. Application software are very specialized in their function and approach to solving a problem. It is also called the end-user programs. These programs do the real work for users. Here are some commonly used application software −

- Word processing
- Spreadsheet
- Presentation
- Database management
- Multimedia tools

## 2.3 Utility Software

Utility software is system software designed to help analyze, configure, optimize or maintain a computer. Examples of utility software include −

- Antivirus software
- Disk management tools
- File management tools
- Compression tools
- Backup tools

## 2.4 Categorization of Software-

**Freeware**
Freeware is software that is distributed without demanding a fee for its usage. These programs are available either as fully functional software for an unlimited period.

Ownership of any freeware is retained by its developer. The developer can change future releases from freeware to a paid product (freeware) if he wishes so. Also, a freeware is typically distributed without its source code. This is done to prevent any sort of modification by its users. Plus, the license with which a free program is distributed may permit the software to be freely copied but not sold.

**Crippleware**
Some software are offered as freeware – but with very limited features – or with the major feature missing. These are referred to as Crippleware. The ones that provides fully functional version has all the functions enabled and is mostly available either as a commercial program or as a shareware. In most cases, the free programs promote a commercial offering.

**Donationware**
Sometimes, freeware is distributed to users with a regular reminder or request to make a donation to the author or to some third-party such as a charity. In such cases, freeware is referred as a Donationware.

**Free Software**
Many computer users aren't fully aware of this somewhat new and unrelated concept. Well, free software is software that gives a user freedom to run, copy, distribute, study, change and improve software. To be precise, free software is a matter of liberty, not price!
It essentially means a user can freely use, modify, and distribute a program stipulated to one condition: any redistributed version of the software must be distributed with the original terms of free use, modification, and distribution (known as copyleft). And unlike freeware, free software may be distributed for a fee.

**Open Source**
The term 'Open Source' is very close to 'free software' but not identical to it. We say this because, the source code of an open-source software is readily available to users but under a copyright, and one is freely allowed to re-distribute the software.

The concept of open-source program relies on the fact that a user can review a source-code for eliminating possible bugs in it. This is something that we do not observe in commercially developed and packaged programs. Programmers on the internet read and modify the source-code by eliminating the possible bugs. Thus, in this way programmers helps in providing more useful and bug-free product for everyone to use.

**Shareware**
Shareware is demonstration software that is distributed for free but for a specific evaluation period only, say, 15-30 days (Trialware). After the evaluation period the program gets expired and a user can no longer access the program. Only if you are interested in using the program further, the shareware provider may require you purchase a license for the software.

Apart from this there are some other terms related with computer softwares:

**Adware**
Adware, better known as advertising software is software that automatically renders advertisements. Most of these advertisements appear in the form of annoying pop-ups. However, one can disable the ads by purchasing a registration key. It can even change your home page, default search or install a tool bar. Like freeware, Adware too is available for computer users at no cost.

**Bundleware**
Bundleware gets its name from people 'bundling' different programs into one single installation program. The one installation for bundleware installs the main program that you want along with some other programs that you do not want.

**Spyware**
Spyware goes a few steps further and surreptitiously installs another software on your computer. The spyware may contain a code that sends information about the user's computer to the developer or to some other location whenever the user is connected to the Internet. This is done to display advertisements in the Web browser.

**Malware**
Typically referred as 'Malicious Software', Malware is any program with malafide intentions and which exploits data of a computer without its user's consent. Once on a computer hard drive, it can hijack your browser and track the websites you visit – and cause even worse damage.
In addition to this, it can hide itself deep within Windows and even reinstall itself after being removed completely, making it the most difficult program to be removed or cleaned. Virus, Trojans, etc. may all be considered as malware.

**Scareware**
Malware that is designed to trick users into downloading and buying non-functional or a dangerous software is referred to as Scareware or Rogue Software. How does it do this? Simple, it alarms scares a user by making him falsely believe that his computer is infected by potentially harmful viruses.
Once downloaded and installed, the program displays false virus alerts and instructs him to buy the 'full version' to remove the infections (fictional).

## 3. Hardware

**Hardware** refers to the physical elements of a computer. Examples of hardware in a computer are the keyboard, the monitor, the mouse and the central processing unit. A computer's hardware is comprised of many different parts, but perhaps the most important of these is the motherboard. The motherboard is made up of even more parts that power and control the computer. In contrast to software, hardware is a physical entity. Hardware and software are interconnected, without software, the hardware of a computer would have no function.

## 3.1 Introduction to Peripheral

A peripheral is a device that can be attached to the computer processor. Peripheral devices can be external, such as a mouse, keyboard, printer, monitor or scanner. Peripheral devices can also be internal, such as a CD-ROM drive, DVD-R drive or modem. Devices are usually classified as input and output devices.

**Input Devices**
A device that feeds data into a computer processor is called an input device.
Examples of common input devices are: keyboards, mouse, joystick, microphone and scanner

**Output Devices**
Output can also appear in a variety of forms - text, video, graphics, and so on. A device that shows data from a computer processor is an output device. Examples of common output devices are: monitors, laser printers, ink jet printers, speakers and headphones

**Input Devices** -
**A Keyboards** consists of a set of keys representing the alphabet and numbers. The keys are usually laid out in QWERTY style which originates from typewriters. The pressing of a key results in the generation, within the keyboard, of an 8-bit binary word, representing the character on the pressed key. This binary pattern is usually in ASCII code. The ASCII code for the chosen character is sent out from the keyboard using serial data transmission.

**A Scanners** is a digitizer as it converts graphics and text information into digital form. Modern scanners allow high resolution images to be scanned using high bit depths. This results in image files which are very large.

**The mouse**, sometimes called a pointer, is a hand-operated input device used to manipulate objects on a computer screen. Whether the mouse uses a laser or ball, or is wired or wireless, a movement detected from the mouse sends instructions to the computer to move the cursor on the screen in order to interact with files, windows, and other software elements. Even though the mouse is a peripheral device that sits outside the main computer housing,

**Output Device-**An output device is any peripheral that receives data from a computer, usually for display, projection, or physical reproduction. For example, the image shows an inkjet printer, an output device that can make a hard copy of any information shown on your monitor, which is another example of an output device. Following are some of the important output devices used in a computer.
- Monitors
- Graphic Plotter
- Printer

## 3.2 Monitors

Monitors, commonly called as **Visual Display Unit** (VDU), are the main output device of a computer. It forms images from tiny dots, called pixels that are arranged in a rectangular form. The sharpness of the image depends upon the number of pixels.

**A pixel** is the smallest unit of a digital image or graphic that can be displayed and represented on a digital display device. A pixel is represented by a dot or square on a computer monitor display screen. A pixel is the basic logical unit in digital graphics. Pixels are combined to form a complete image, video, text or any visible thing on a computer display.
There are two kinds of viewing screen used for monitors.
- Cathode-Ray Tube (CRT)
- Flat-Panel Display

**Cathode-Ray Tube (CRT) Monitor**
The CRT display is made up of small picture elements called pixels. The smaller the pixels, the better the image clarity or resolution. It takes more than one illuminated pixel to form a whole character, such as the letter 'e' in the word help.

**Flat-Panel Display Monitor**
The flat-panel display refers to a class of video devices that have reduced volume, weight and power requirement in comparison to the CRT. You can hang them on walls or wear them on your wrists. Current uses of flat-panel display include calculators, video games, monitors, laptop computer, and graphics display.

**The flat-panel display is divided into two categories -**
**Emissive Displays** − Emissive displays are devices that convert electrical energy into light. For example, plasma panel and LED (Light-Emitting Diodes).

**Non-Emissive Displays** − Non-emissive displays use optical effects to convert sunlight or light from some other source into graphics patterns. For example, LCD (Liquid-Crystal Device).

## 3.3 Printers

Printer is an output device, which is used to print information on paper.
There are two types of printers −
- Impact Printers
- Non-Impact Printers

**Impact Printers**
Impact printers print the characters by striking them on the ribbon, which is then pressed on the paper.
Characteristics of Impact Printers are the following −
- Very low consumable costs
- Very noisy
- Useful for bulk printing due to low cost
- There is physical contact with the paper to produce an image

These printers are of two types −
- Character printers
- Line printers

**Character Printers**
Character printers are the printers which print one character at a time.
These are further divided into two types:
- Dot Matrix Printer(DMP)
- Daisy Wheel

**Dot Matrix Printer**
In the market, one of the most popular printers is Dot Matrix Printer. These printers are popular because of their ease of printing and economical price. Dot matrix printers print one character at a time. It prints characters and images as a pattern of dots.

**Advantages**
- Inexpensive
- Widely Used
- Other language characters can be printed

**Daisy Wheel**
Head is lying on a wheel and pins corresponding to characters are like petals of Daisy (flower) which is why it is called Daisy Wheel Printer.

**Advantages**
- More reliable than DMP
- Better quality
- Fonts of character can be easily changed

**Line Printers**
Line printers are the printers which print one line at a time instead of one or more character at a time.

**Non-impact Printers**
Non-impact printers print the characters without using the ribbon. A non-impact printer uses electrostatic chemicals and ink-jet technologies. It can produce high quality of graphics.

These printers are of two types –
- Laser Printers
- Inkjet Printers

**Characteristics of Non-Impact Printers**
- Faster than impact printers
- They are not noisy
- High quality
- Supports many fonts and different character size

**Laser Printers**
These are non-impact page printers. Laser printer uses laser beam on to photo-sensitive surface for printing.

**Advantages**
- Very high speed
- Very high-quality output
- Good graphics quality
- Supports many fonts and different character size

**Inkjet Printers**
Inkjet printers are non-impact character printers based on a relatively new technology. They print characters by spraying small drops of ink onto paper. Inkjet printers produce high quality output with presentable features

## 4. Basics of Memory

**A memory** is just like a human brain. It is used to store data and instructions. Computer memory is the storage space in the computer, where data is to be processed and instructions required for processing are stored. Memory is primarily of three types –
- Cache Memory
- Primary Memory/Main Memory
- Secondary Memory

### 4.1 Cache Memory

Cache memory is a very high-speed semiconductor memory which can speed up the CPU. It acts as a buffer between the CPU and the main memory. It is used to hold those parts of data and program which are most frequently used by the CPU. The parts of data and programs are transferred from the disk to cache memory by the operating system, from where the CPU can access them.

**Advantages**
The advantages of cache memory are as follows –
- Cache memory is faster than main memory.
- It consumes less access time as compared to main memory.
- It stores the program that can be executed within a short period of time.
- It stores data for temporary use.

**Disadvantages**
The disadvantages of cache memory are as follows –
- Cache memory has limited capacity.
- It is very expensive.

### 4.2 Primary Memory (Main Memory)

Primary memory holds only those data and instructions on which the computer is currently working. It has a limited capacity and data is lost when power is switched off. It is generally made up of semiconductor device. These memories are not as fast as registers. The data and instruction required to be processed resides in the main memory. It is divided into two subcategories RAM and ROM.

**Characteristics of Main Memory**

- These are semiconductor memories.
- It is known as the main memory.
- Usually volatile memory.
- Data is lost in case power is switched off.
- It is the working memory of the computer.
- Faster than secondary memories.
- A computer cannot run without the primary memory.

### 4.3 Secondary Memory

This type of memory is also known as external memory or non-volatile. It is slower than the main memory. These are used for storing data/information permanently. CPU directly does not access these memories, instead they are accessed via input-output routines. The contents of secondary memories are first transferred to the main memory, and then the CPU can access it. For example, disk, CD-ROM, DVD, etc.

**Characteristics of Secondary Memory**

- These are magnetic and optical memories.
- It is known as the backup memory.
- It is a non-volatile memory.
- Data is permanently stored even if power is switched off.
- It is used for storage of data in a computer.
- Computer may run without the secondary memory.
- Slower than primary memories.

## 5. Ports

A connection point that acts as interface between the computer and external devices like mouse, printer, modem, etc. is called port. Ports are of two types −

**Internal port** − It connects the motherboard to internal devices like hard disk drive, CD drive, internal modem, etc.
**External port** − It connects the motherboard to external devices like modem, mouse, printer, flash drives, etc. Let us look at some of the most commonly used ports.

- **Serial ports** transmit data sequentially one bit at a time. So, they need only one wire to transmit 8 bits. However, it also makes them slower. Serial ports are usually 9-pin or 25-pin male connectors. They are also known as COM (communication) ports or RS323C ports.
- **Parallel ports** can send or receive 8 bits or 1 byte at a time. Parallel ports come in form of 25-pin female pins and are used to connect printer, scanner, external hard disk drive, etc.
- **USB Port (USB)** stands for Universal Serial Bus. It is the industry standard for short distance digital data connection. USB port is a standardized port to connect a variety of devices like printer, camera, keyboard, speaker, etc.
- **PS/2 Port (PS-2)** stands for Personal System/2. It is a female 6-pin port standard that connects to the male mini-DIN cable. PS/2 was introduced by IBM to connect mouse and keyboard to personal computers. This port is now mostly obsolete, though some systems compatible with IBM may have this port.
- **Infrared port** is a port that enables wireless exchange of data within a radius of 10m. Two devices that have infrared ports are placed facing each other so that beams of infrared lights can be used to share data.
- **Bluetooth Port- Bluetooth** is a telecommunication specification that facilitates wireless connection between phones, computers and other digital devices over short range wireless connection. Bluetooth port enables synchronization between Bluetooth-enabled devices. There are two types of Bluetooth ports −
  - Incoming − It is used to receive connection from Bluetooth devices.
  - Outgoing − It is used to request connection to other Bluetooth devices.

## 6. Types of Computer

Computers are generally are of 3 types on basis of electronic signal the process.
**Analog Computers:**

“An analog computer is a form of computer that uses the continuously-changeable aspects of physical fact such as electrical, mechanical, or hydraulic quantities to model the problem being solved.” They operate on inputs of varying voltage.
Analog computers were widely used in scientific and industrial applications. They are constructed to perform specific task and …
Example Tire pressure gauge, analog clock, thermometer

**Digital Computers**
A Digital Computer, as its name implies, works with digits to represent numerals, letters or other special symbols. Digital Computers operate on inputs which are ON-OFF type and its output is also in the form of ON-OFF signal. Normally, an ON is represented by a 1 and an OFF is represented by a 0 i.e. they process data in binary form.

The results of digital computers are more accurate than the results of analog computers. Analog computers are faster than digital. Analog computers lack memory whereas digital computers store information.

**Hybrid Computers**
They exhibit features of both analog and digital computers. It has the speed of analog computer and the memory and accuracy of digital computer. Traditionally, the analog components of the computer handle complex mathematical computations. The digital components take care of logical and numerical operations in addition to serving as the controller for the system.
Example is ECG used in hospitals to measure the heartbeat of the patient (ECG).
Now let us see another classification based on configuration and size

## 6.1. Based on configuration and size

**Super computers**
A supercomputer is a computer with great speed and memory. This kind of computer can do jobs faster than any other computer of its generation. They are huge and faster than other computers

**Mainframe computers**
They are also referred as **Big Iron**.
A mainframe computer is a very large computer capable of handling and processing very large amounts of data quickly. They are used by large institutions, such as government agencies and large corporations.
Mainframes are not good at number-crunching or don't do scientific calculations. A Mainframe is not a Super-computer. Mainframe computers don’t have a beautiful user-interface like the PC at your home.
Compared to a typical PC, mainframes commonly have hundreds to thousands of times as much data storage online, and can access it reasonably fast. Examples are: IBM 360, IBMzSeries, Unisys Dorado, Unisys Libra

**Mini Computers**
The term "minicomputer" developed in the 1960s to describe the smaller computers that became possible with the use of transistors and core memory technologies, minimal instructions sets and less expensive peripherals.
Minicomputers were also known as midrange computers. They grew to have relatively high processing power and capacity. The decline of the minis happened due to the lower cost of microprocessor-based hardware. The result was that minicomputers and computer terminals were replaced by networked workstations, file servers and PCs in some installations.

**Micro Computers**
A microcomputer is a standard desktop computer used at a home and in business. Microcomputer is a computer with a microprocessor as its CPU. They are cheap, compact and can be easily accommodated on a study table. They have evolved a lot over time. We are well familiar with uses of microcomputers as surfing on net, sending and receiving e-mails, database management and much more.
Examples: Laptops, Desktop Computer, Notebook, Tablet Computer, Smartphone, Palmtop

## 6.2. Based on functions

**Server-** A computer that is dedicated to provide services to its clients. For example, file server is for sharing file resources and likewise there is database server
**Workstations** serve usually one user only.

**Information appliances** are computers specially designed to perform a specific "user-friendly" function—such as playing music, photography. Typical examples are smartphones and personal digital assistants (PDAs).

**Embedded computers** - An embedded system is a computer system with a dedicated function within a larger system. Since the embedded system is dedicated to specific tasks, design engineers can optimize it to reduce the size and cost of the product and increase the reliability and performance. Examples of embedded systems are digital watches, MP3 players, Traffic lights, video game consoles.

### 6.3. Based on area of application

**Special Purpose Computers**
Special-purpose computer is the one that is designed to perform a specific task. Most of the times their job is to solve one particular problem. They are also known as dedicated computers, because they are dedicated to perform a single task over and over again.

**General Purpose Computers**
General-purpose computer is the one that can work on different types of programs input to it and thus be used in countless applications. General purpose computers are designed to perform a wide variety of functions and operations.

## PRACTICE SET

1. Which of the following is not a computer bus?
(a) data bus (b) timer bus
(c) control bus (d) address bus
(e) None of these
2. The technique of assigning a memory address to each I/O device in the SAM system is called:
(a) wired I/O (b) I/O mapping
(c) dedicated I/O (d) memory-mapped I/O
(e) None of these
3. How many bits are used in the data bus?
(a) 7 (b) 8
(c) 9 (d) 16
(e) None of these
4. A port can be:
(a) strictly for input (b) strictly for output
(c) bidirectional (d) all the above
(e) None of these
5. Which of the following is not a basic element within the microprocessor?
(a) microcontroller (b) arithmetic-logic unit (ALU)
(c) temporary register (d) accumulator
(e) None of these
6. When referring to instruction words, a mnemonic is:
(a) a short abbreviation for the operand address
(b) a short abbreviation for the operation to be performed
(c) a short abbreviation for the data word stored at the operand address
(d) short hand for machine language
(e) None of these
7. What is the difference between mnemonic codes and machine codes?
(a) There is no difference.
(b) Machine codes are in binary, mnemonic codes are in shorthand English.
(c) Machine codes are in shorthand English, mnemonic codes are in binary.
(d) Machine codes are in shorthand English, mnemonic codes are a high-level language.
(e) None of these
8. Which bus is bidirectional?
(a) data bus (b) control bus
(c) address bus (d) multiplexed bus
(e) None of these
9. Parity is:
(a) a byte stored in the FAT to indicated remaining slots
(b) the optimal transmission speed of data over a CAT 5 cable
(c) an extra bit stored with data in RAM that is used to check for errors when the data is read back
(d) the optimal transmission speed of data over a CAT 5 cable
(e) None of these
10. What does FDISK do?
(a) performs low-level formatting of the hard drive
(b) fixes bad sectors on the hard drive
(c) recovers lost clusters on the hard drive
(d) creates partitions on the hard drive
(e) None of these
11. A string of eight 0s and 1s is called a:
(a) megabyte. (b) byte.
(c) kilobyte. (d) gigabyte.
(e) None of these
12. The CPU and memory are located on the:
(a) expansion board. (b) motherboard.
(c) storage device. (d) output device.
(e) None of these

13. Word processing, spreadsheet, and photo-editing are examples of:
(a) application software.(b ) system software.
(c) operating system software.
(d) platform software. (e) None of these
14. System software is the set of programs that enables your computers hardware devices and ____________ software to work together.
(a) management (b) processing
(c) utility (d) application
(e) None of these
15. The PC (personal computer) and the Apple Macintosh are examples of two different:
(a) platforms. (b) applications.
(c) programs. (d) storage devices.
(e) None of these
16. ____________ are specially designed computers that perform complex calculations extremely rapidly.
(a) Servers (b) Supercomputers
(c) Laptops (d) Mainframes
(e) None of these
17. The operating system is the most common type of ____________ software.
(a) communication (b) application
(c) system (d) word-processing software
(e) None of these
18. The chip, used in computers, is made of
(a) chromium (b) iron oxide
(c) silica (d) silicon
(e) None of these
19. The following device allows the user to add external components to a computer system
(a) Storage devices (b) Keyboards
(c) Portal system boards (d) Diskettes
(e) None of these
20. All of the following are examples of storage devices except
(a) hard disk drives (b) printers
(c)floppy disk drives (d)CD drives
(e) None of these
21. From what location are the 1st computer instructions available on boot up?
(a) ROM BIOS (b)CPU
(c) boot.ini (d)CONFIG.SYS
(e)None of the above
22. Which one is the secondary memory device?
(a)CPU (b)ALU
(c)floppy disk (d)Mouse
(e)None of these
23. If a memory chip is volatile, it will ________________.
(a) Explode if exposed to high temperatures
(b) Lose it contents if current is turned off
(c)Be used for data storage only
(d)Be used to both read and write data
(e)None of these
24. What tool is used to test serial and parallel ports?
(a) high volt probe (b) cable scanner
(c) loop backs (wrap plugs)
(d) sniffer (e) None of the above
25. From what location are the 1st computer instructions available on boot up?
(a) ROM BIOS (b) CPU
(c) boot.ini (d) CONFIG.SYS
(e) None of the above
26. What could cause a fixed disk error?
(a) No-CD installed (b) bad ram
(c) slow processor (d) Incorrect CMOS settings
(e) None of the above
27. Missing slot covers on a computer can cause?
(a) over heat (b) power surges
(c) EMI. (d) incomplete path for ESD
(e) None of the above
28. A hard disk is divided into tracks which are further subdivided into:
(a) clusters (b) sectors
(c) vectors (d) heads
(e) None of the above
29. Which part of the laser printer should NOT be exposed to sunlight?
(a) Transfer corona assembly
(b) PC drum (c) Primary corona wire
(d) Toner cartridge (e) None of the above
30. Resistance is measured in?
(a) Volts (b) Amps
(c) Watts (d) Ohms
(e) None of the above
31. Program which is used to control system performance is classified as?
(a) experimental program
(b) system program (c) specialized program
(d) organized program (e) none of these
32. Examples of system programs includes
(a) operating system of computer
(b) trace program (c) compiler
(d) all of above (e) none of these
33. System programs which performs one simple task are classified as
(a) utility programs (b) function program
(c) compiling program (d) inquiry program
(e) None of these
34. In microcomputers, operating system is usually stored on
(a) random access memory
(b) read only memory (c) permanent memory
(d) temporary memory (e) none of these
35. Software which controls general operations of computer system is classified as
(a) dump programs (b) function system
(c) operating system (d) inquiry system
(e) none of these

36. Which provides the fastest access to large video files?
(a) Optical drives (b) IDE hard drives
(c) SCSI hard drives (d) EIDE hard drives
(e) None of the above
37. You install a second IDE hard drive, then boot to windows. But windows will not recognize it. What should you do next?
(a) Reinstall windows (b) Run CHKDISK
(c) Run defrag
(d) Change the jumper setting to slave
(e) None of these
38. Which of the following is true?
(a) DRAM is faster than SRAM
(b) DRAM and SRAM both are same with respect to speed
(c) SRAM is faster than DRAM
(d) All are true
(e) None of these
39. Serial port enables data flow in:
(a) One direction (b) Both directions
(c) Doesn't flow the data
(d) All of the above (e) None of these
40. Following cable provide immunity from electrical interference:
(a) UTP (b) Fiber Optic
(c) STP (d) Coaxial
(e) None of these
41. A 168-pin DIMM package has ________ pins on each side of the package
(a) 84 pins (b) 64 pins
(c) 32 pins (d)16 pins
(e) None of these
42. What is true regarding parallel port?
(a) A parallel port for connecting an external device.
(b) On PCs parallel port uses a 25-pin connector
(c) Parallel port uses parallel transmission of data
(d) All are true
(e) None of these
43. What Is LGA?
(a) An LGA socket is the connection point for a central processing unit (CPU) to fit into a motherboard.
(b) The LGA stands for Land Grid Array.
(c) The LGA stands for Land Graphic Array.
(d) Both (a) and (b)
(e) None of these
44. What is Pentium IV?
(a) It is a processor.
(b) In Pentium IV, the bus speed is 400 MHz
(c) Pentium IV processor start at 512KB
(d) All of the above
(e) None of these
45. What Are The Latest Processor Of Intel And Amd?
(a) For intel it is Intel Core i7 and AMD Opteron 6200 Series processor.
(b) For intel it is Intel Opteron 6200 Series and AMD Core i7 processor.
(c) For intel it is Intel Core i5 and AMD Opteron 6200 Series processor.
(d) For intel it is Intel Core i7 and AMD Opteron 5200 Series processor
(e) None of these
46. Which Type Of Socket Is Needed To Connect A Dual Core Processor Of Intel?
(a) Socket LPA 775 (b) Socket LGA 775
(c) Socket LGA 774 (d) All of the above
(e) None of these
47. What Is Cache Memory? What Is The Advantage If A Processor With More Cache Memory You Are Using?
(a) Cache memory is the memory area between RAM and SRAM. If cache memory decreases the speed of the system will also improved
(b) Cache memory is the memory area between RAM and Processor. If cache memory decreases the speed of the system will also improved
(c) Cache memory is the memory area between RAM and Processor. If cache memory increases the speed of the system will also improved.
(d) None of these
(e) All are true.
48. What is true about DRAM?
(a) Dynamic RAM stores data using a paired transistor and capacitor for each bit of data.
(b) In DRAM Capacitors constantly leak electricity, which requires the memory controller to refresh the DRAM several times a second to maintain the data.
(c) DDR-SDRAM is a type of DRAM
(d) All are true
(e) None of these
49. What is true about DDR2?
(a) DDR2 is the successor to DDR RAM.
(b) DDR2 incorporates several technological upgrades to computer system memory, as well as an enhanced data rate.
(c) DDR 2 is capable of achieving twice the data transfer rate of DDR-I memory because of its higher clock speed. It operates at a lower voltage than DDR-I as well: 1.8 volts instead of 2.5.
(d) All of the above
(e) None of these
50. What Is Full Name Of AMD?
(a) Advanced Micro Devices.
(b) Advanced Memory Devices
(c) Advanced Multipurpose Devices
(d) Advanced Multitasking Devices
(e) None of these
51. Winchester drive is also called:
(a) Hard Disk Drive (b) Floppy Disk Drive
(c) CD (d) DVD
(e) None of these

52. Which of the following is NOT a type of motherboard expansion slot?
(a) ISA (b) PCI
(c) AGP (d) ATX
(e) None of these

53. How much data will a high density (HD) floppy disk hold?
(a) 124 KB (b) 640 KB
(c) 1.44 MB (d) 2.88MB
(e) None of these

54. What does FDISK do?
(a) performs low-level formatting of the hard drive
(b) fixes bad sectors on the hard drive
(c) recovers lost clusters on the hard drive
(d) creates partitions on the hard drive
(e) None of these

55. What is different between AT and ATX power supplies?
(a) They are identical except for their shape.
(b) AT supplies use a single P1 power connector while ATX uses P8 and P9.
(c) AT supplies use P8 and P9 power connectors while ATX uses a single P1 connector.
(d) AT power supplies run on 120V AC current while ATX uses 220V AC
(e) None of these

56. A semiconductor memory which allows the eraser of the information stored in it so that new information can be stored in it is referred as
(a) EPROM (b) ROM
(c) RAM (d) None of these
(e) SDRAM

57. Index hole is related to?
(a) Scanner (b) Floppy disk
(c) Printer (d) CPU
(e) None of these

58. In MSDOS, the primary hard disk drives has the drive letter?
(a) a (b) b
(c) c (d) f
(e) None of these

59. A special type of Batch file that run automatically at startup is?
(a) Command.Com (b) Autoexec. Bat
(c) Config.Sys (d) Ansi.Sys
(e) None of these

60. Which command displays directories as well as subdirectories also in MSDOS?
(a) DIR/All (b) DIR/AN
(c) DIR/DS (d) DIR/S
(e) None of these

## SOLUTIONS

| | | | | | | |
|---|---|---|---|---|---|---|
| 1. (b) | 2. (d) | 3. (b) | 4. (c) | 5. (a) | 6. (b) | 7. (b) |
| 8. (a) | 9. (c) | 10. (d) | 11. (b) | 12. (b) | 13. (a) | 14. (d) |
| 15. (a) | 16. (b) | 17. (c) | 18. (d) | 19. (c) | 20. (b) | 21. (a) |
| 22. (c) | 23. (b) | 24. (c) | 25. (a) | 26. (d) | 27. (a) | 28. (b) |
| 29. (b) | 30. (d) | 31. (b) | 32. (d) | 33. (a) | 34. (b) | 35. (c) |
| 36. (c) | 37. (d) | 38. (c) | 39. (b) | 40. (b) | 41. (a) | 42. (d) |
| 43. (d) | 44. (d) | 45. (a) | 46. (b) | 47. (c) | 48. (d) | 49. (d) |
| 50. (a) | 51. (a) | 52. (d) | 53. (c) | 54. (d) | 55. (c) | 56. (a) |
| 57. (b) | 58. (c) | 59. (b) | 60. (d) | | | |

# 2 DBMS

## 1. Introduction

DBMS is one of the most important module for Specialist Officer (IT) Exam. As we've seen that the objective paper of Professional Knowledge (especially for Scale-I Officer) in IBPS Exam has many questions from Database and Networking Modules. Thus, aspirants should prepare DBMS thoroughly. The term DBMS stands for Data Base Management System. Now comes a question that *what is DBMS?*
**DBMS is the acronym of Data Base Management System. DBMS is a collection of interrelated data and a set of programs to access this data in a convenient and efficient way. It controls the organization, storage, retrieval, security and integrity of data in a database.**

A database management system (DBMS) is a computer software that manages databases, it may use any of a variety of database models, such as the Hierarchical DBMS, Network DBMS and Relational DBMS.

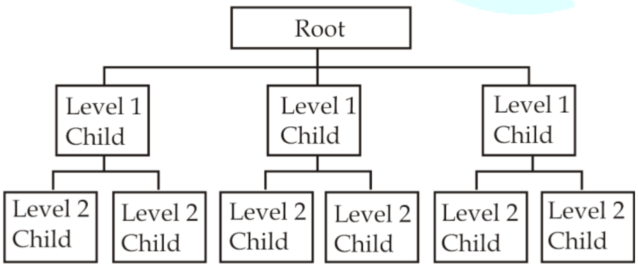
The emergence of the first type of DBMS was between 1960's-70's; that was the Hierarchical DBMS. IBM had the first model, developed on IBM 360 and their (DBMS) was called IMS, originally it was written for the Apollo program. This type of DBMS was based on binary trees, where the shape was like a tree and relations were only limited between parent and child records.

### 1.1 Database Models

A database model shows the logical structure of a database, including the relationships and constraints that determine how data can be stored and accessed.

**Hierarchical Database Model**

Hierarchical Database model



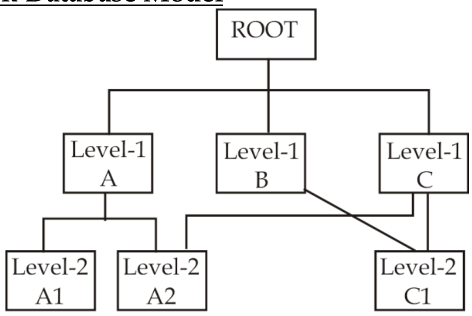

Hierarchical Database model is one of the oldest database models. In the hierarchical data model, records are linked with other superior records on which they are dependent and also on the records, which are dependent on them. A tree structure may establish one-to-many relationship. Parents can have many children exhibiting one to many relationships. The grandparents and children are the nodes or dependents of the root. In general, a root may have any number of dependents.
A tree-structure diagram is the schema for a hierarchical database. Such a diagram consists of two basic components:
1. Boxes, which correspond to **record types**
2. Lines, which correspond to **links**

**Pros:**
- ✓ The model allows easy addition and deletion of new information.
- ✓ Data at the top of the Hierarchy is very fast to access.
- ✓ It relates well to anything that works through a one to many relationships.

**Cons:**
- Realtime requirements are of more sophisticated relationships which this model fails to cater.

- The database can be very slow when searching for information on the lower entities.
- Many to many relationships are not supported.

**Network Database Model**



Network Database Model Diagram

The Network Database model can be viewed as an upside-down tree where each member information is the branch linked to the owner, which is the bottom of the tree. The network database model was a progression from the hierarchical database model and was designed to solve some of that model's problems, specifically the lack of flexibility. It addresses the need to model more complex relationships such as the many-to-many relationship which hierarchical model could not deal with.

The Network model replaces the hierarchical tree with a **graph** thus allowing more general connections among the nodes. The main difference of the network model from the hierarchical model, is its ability to handle many to many (N:N) relations. In other words, **it allows a record to have more than one parent.**

**Pros:**
- ✓ In the network database terminology, a relationship is a set. Each set comprises of two types of records.- an owner record and a member record,
- ✓ In a network model an application can access an owner record and all the member records within a set.
- ✓ Network Model supports data independence to some level as it draws a clear line of demarcation between programs and the complex physical storage details.

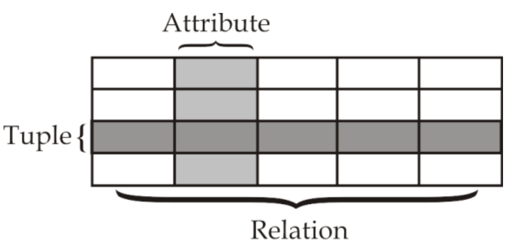
**Cons:**
- The insertion, deletion and updating operations of any record require large number of pointers adjustments.
- A change in structure demands a change in the application as well, which leads to lack of structural independence.

**Relational Database Model**

Relational data model is the primary data model, which is used widely around the world for data storage and processing. The relational database model was a huge leap forward from the network database model. Instead of relying on a parent-child or owner-member relationship, the relational model allows any file to be related to any other by means of a common field. Relational databases go hand-in-hand with the development of SQL. Structured Query Language is a standardized language for defining and manipulating data in a relational database.

**What are Tables in Relational Model?**

Relations are saved in the format of Tables. This format stores the relation among entities. A table has rows and columns, where rows represents records and columns represent the attributes. Following are some terms associated with relations:



1. **Tuple −** A single row of a table, which contains a single record for that relation is called a tuple.
2. **Relation instance −** A finite set of tuples in the relational database system represents relation instance. Relation instances do not have duplicate tuples.
3. **Relation schema −** A relation schema describes the relation name (table name), attributes, and their names.
4. **Relation key −** Each row has one or more attributes, known as relation key, which can identify the row in the relation (table) uniquely.
5. **Attribute domain** − Every attribute has some pre-defined value scope, known as attribute domain.

**Object-Oriented Database Model**

Object Oriented Database Model (also referred to as object-oriented database management system or OODBMS), is a database management system (DBMS) that supports the modelling and creation of data as objects. This includes some kind of support for classes of objects and the inheritance of class properties and methods by subclasses and their objects. ODBMS were originally thought of to replace RDBMS because of their better fit with object-oriented programming languages. However, high switching cost, the inclusion of object-oriented features in RDBMS to make them ORDBMS, and the emergence of object-relational mappers (ORMs) have made RDBMS successfully defend their dominance in the data center for server-side persistence.

Relational databases store data in tables that are two dimensional. The tables have rows and columns. Relational database tables are "normalized" so data is not repeated more often than necessary. With traditional databases, data manipulated by the application is transient and data in the database is persisted (Stored on a permanent storage device). In object databases, the application can manipulate both transient and persisted data.

**Entity-Relationship Database Model**

The Entity Relationship Data Model ensure that you get a precise understanding of the nature of the data and how it is used by the enterprise, you need to have a universal model for interaction that is non-technical and free of ambiguities and easy readable to both technical as well as non-technical members. ***This is implemented with use of the ER Diagrams.***

ER model is based on two concepts:

- Entities, defined as tables that hold specific information (data)
- Relationships, defined as the associations or interactions between entities

**What is Entity Relationship Diagram (ER-Diagram)?**
ER-Diagram is a pictorial representation of data that describes how data is communicated and related to each other. Any object, such as entities, attributes of an entity, sets of relationship and other attributes of relationship can be characterized with the help of the ER diagram.

## 1.2 Advantages of today's' DBMS over earlier File Management System

These are some important advantages of today's DBMS:

**Reduced Data Redundancy and Inconsistency:** This means with DBMS the chances of multiple file formats, duplication of information in different files got eliminated. Which means it reduced data duplication and with this the data could stay more consistent.

**Data Integrity:** data integrity" refers to the accuracy and consistency of data stored in a database
DBMS ensures data integrity by managing transactions through **ACID** test = atomicity, consistency, isolation, durability. While such integrity is absent in file management system.

**Sharing of Data:** In DBMS, data can be shared by authorized users of the organization. The database administrator manages the data and gives rights to users to access the data.

**Control Over Concurrency:** In a file-based system, if two users can access data simultaneously, it is possible that they will interfere with each other. For example, if both users attempt to perform update operation on the same record, then one may overwrite the values recorded by the other. Most database management systems have sub-systems to control the concurrency so that transactions are always recorded with accuracy.

**Backup and Recovery Procedures:** In a computer file-based system, the user creates the backup of data regularly to protect the valuable data from damage due to failures to the computer system or application program. It is very time consuming method, if amount of data is large. Most of the DBMSs provide the 'backup and recovery' sub-systems that automatically create the backup of data and restore data if required.

**Data Independence:** The separation of data structure of database from the application program that uses the data is called data independence. In DBMS, you can easily change the structure of database without modifying the application program.

## 2. Database Architecture

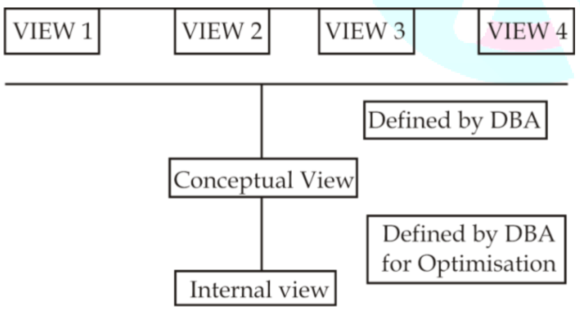
### 2.1 Data Abstraction – View Levels

The generalized architecture of DBMS is called ANSI/SPARC model. The architecture is divided into three levels:

**1. External view or User view/View Level**
It is the highest level of data abstraction. This includes only those portions of database of concern to a user or Application program. Each user has a different external view and it is described by means of a scheme called external schema.

**2. Conceptual view/Logical Level**
All the database entities and the relationship among them are included. One conceptual view represents the entire database called conceptual schema.

**3. Internal view/Physical Level**
It is the lowest level of abstraction, closest to the physical storage method. It describes how the data is stored, what is the structure of data storage and the method of accessing these data. It is represented by internal schema.
*View Level …Defined by User*




### 2.2 Instances and Schemas

Schema can be defined as the design of a database. The overall description of the database is called the database schema.
You can relate it as something like types and variables in programming languages. Thus, essentially **Schema** is the logical structure of the database. Just like the View Levels in Data Abstraction Schema is of 3 types:

1. **Physical Schema:**
The design of a database at physical level is called physical schema, how the data stored in blocks of storage is described at this level.

2. **Logical schema:**
Logical schema can be defined as the design of database at logical level. In this level, the programmers as well as the database administrator (DBA) work. At this level data can be described as certain types of data records which can be stored in the form of data structures. However, the internal details (such as implementation of data structure) will be remaining hidden at this level.

**3. View Schema**
View schema can be defined as the design of database at view level which generally describes end-user interaction with database systems.

- **Physical Data Independence**–the ability to modify the physical schema without changing the logical schema.
- Applications depend on the logical schema
- In general, the interfaces between the various levels and components should be well defined so that changes in some parts do not seriously influence others.

**What is an Instance?**
Databases change over time as information is inserted and deleted. The collection of information stored in the database at a particular moment is called an instance.

### 2.3 Database Languages

A database system provides a data-definition language to specify the database schema and a data-manipulation language to express database queries and updates.

**1. Data Definition Language:** DDL is used for specifying the database schema. It contains commands to create tables, alter the structure, delete tables or rename tables.

***Examples of DDL commands in SQL:***
To create the database instance – CREATE
To alter the structure of database – ALTER
To drop database instances – DROP
To delete tables in a database instance – TRUNCATE
To rename database instances – RENAME

**2. Data Manipulation Language:** As the name specifies itself DML is used for accessing and manipulating data in a database.

***Examples of DML commands in SQL:***
To read records from table(s) – SELECT
To insert records into the tables – INSERT
Update the data in tables– UPDATE
Delete all the records from the table – DELETE

**3. Data Control Language:** DCL is used for granting and revoking user access on a database –

***Examples of DCL commands in SQL:***
To grant access to user – GRANT
To revoke access from user – REVOKE

## 3. Entity-Relationship Model

**What is an Entity?**
In a database, we would be grouping only related data together and storing them under one group name called Entity / Table. This helps in identifying which data is stored where and under what name. It reduces the time to search for a particular data in a whole database.
Entities can be classified based on their strength. An entity is considered weak if its tables are existence dependent. Following are basic types of entities:

1. **Strong Entity**: Entities having its own attribute as primary keys are called strong entity. For example, EMPLOYE has EMPLOYE_ID as primary key. Hence it is a strong entity.
2. **Weak Entity:** Entities which cannot form their own attribute as primary key are known weak entities. These entities will derive their primary keys from the combination of its attribute and primary key from its mapping entity. The relationship between weak entity and strong entity set is called as ***Identifying Relationship***.
3. **Composite Entity:** Entities participating in the many to many relationships are called composite entity.

The relationship between weak entity and strong entity set is called as Identifying Relationship. The line connecting strong entity set with the relationship is single whereas the line connecting weak entity set with the identifying relationship is double. A member of a strong entity set is called dominant entity and member of weak entity set is called as subordinate entity. A weak entity set does not have a primary key, but we need a means of distinguishing among all those entries in the entity set that depend on one particular strong entity set. The discriminator of a weak entity set is a set of attributes that allows this distinction be made. A weak entity set is represented by doubly outlined box and corresponding identifying relation by a doubly outlined diamond. It is also called as the Partial key of the entity set.

**Weak Entity Sets**

Weak Entity Set: An entity set whose members owe their existence to some entity in a *strong entity set.*

- entities are not of independent existence.
- each weak entity is associated with some entity of the *owner* entity set through a special relationship.
- weak entity set may not have a key attribute.

**Weak Entity set- Example**

We depict a weak entity set by double rectangles. We underline the discriminator of a weak entity set with a dashed line.

*For example:* payment_number – discriminator of the *payment* entity set

Primary key for *payment* – (*loan_number, payment_number*)

***Note:*** the primary key of the strong entity set is not explicitly stored with the weak entity set, since it is implicit in the identifying relationship. If *loan_number* were explicitly stored, *payment* could be made a strong entity, but then the relationship between *payment* and *loan* would be duplicated by an implicit relationship defined by the attribute *loan_number* common to *payment* and *loan.*

## 3.1 Attributes

Each entity is described by a set of attributes/properties.

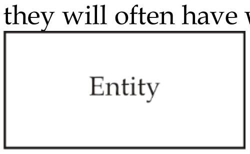
**Types of Attributes**

- Simple Attributes: having atomic or indivisible values. example: Dept – a string Phone Number - an eight-digit number
- Composite Attributes: having several components in the value. example: Qualification with components (Degree Name, Year, University Name)
- Derived Attributes: Attribute value is dependent on some other attribute. example: Age depends on Date of Birth. So, age is a derived attribute.
- Single-valued: having only one value rather than a set of values. for instance, Place Of Birth - single string value.
- Multi-valued: having a set of values rather than a single value. for instance, Courses Enrolled attribute for student Email Address attribute for student Previous Degree attribute for student.
- Attributes can be: simple single-valued, simple multi-valued, composite single-valued or composite multi-valued.

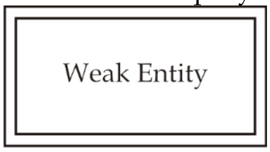
## 3.2 E-R Diagram

An ER diagram is a means of visualizing how the information a system produces is related. There are five main components of an ER Diagram:

1. **Connecting lines, solid lines that connect attributes to show the relationships of entities in the diagram.**
2. **Entities: Represented by Rectangle**
   - Strong Entity: These shapes are independent from other entities, and are often called parent entities, since they will often have weak entities that depend on them.

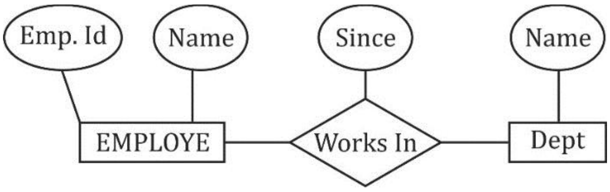


- Weak Entity: A weak entity is an entity that must defined by a foreign key relationship with another entity as it cannot be uniquely identified by its own attributes alone.




**3. Relationship: connects two or more entities into an association/relationship - Diamond**

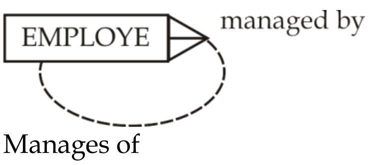
Here you can see: Employee Works in Department. EMPLOYE and Dept are Entity Types and WorksIn is the relationship represented with a diamond figure.




A **recursive relationship** is one in which the same entity participates more than once in the relationship. For Example: Every manager is also an employee. So, manager is not a new entity,
but just a subset of the instances of the entity EMPLOYEE.

**Recessive Relationship:**

| EMPLOYE | MANAGER |
|---|---|
| Nikhil | Anuj |
| Sumita | Nikhil |
| Anuj | |
| Rahul | |




This also a representation of - many cardinality.

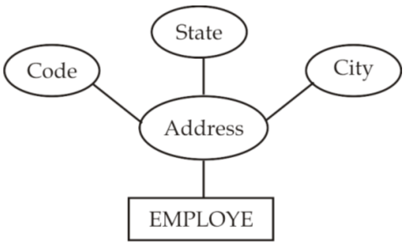
### 3.3 Attributes-Represented by Ovals

An Attribute describes a property or characteristic of an entity. For example, Name, ID, Age, Address etc can be attributes of an EMPLOYE.

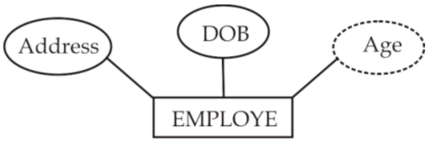
Key attribute represents the main characteristic of an Entity. It is used to represent Primary key. Ellipse with underlying lines represent Key Attribute. Here EmpId is the key attribute that is the primary key which will uniquely identify the EMPLOYE Records.

Double Ellipses is used to represent multivalued attributes.

An attribute can also have their own attributes. These attributes are known as **Composite attribute.**



**Derived Attribute** is calculated or otherwise derived from another attribute, such as age from a DOB (Date of Birth).

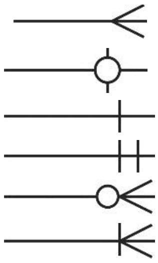



### 4. Cardinality

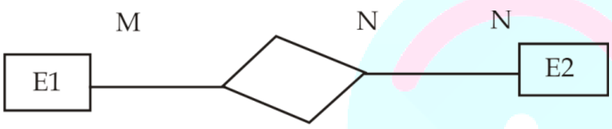
The cardinality of a relationship is the number of instances of entity B that can be associated with entity A. There is a minimum cardinality and a maximum cardinality for each relationship. Cardinality refers to the maximum number of times an instance in one entity can relate to instances of another entity. Ordinality, on the other hand, is the minimum number of times an instance in one entity can be associated with an instance in the related entity.

<u>**Cardinalit**</u>

Many
Zero or one
One
One (and only one)
Zero or many
One or many



**Binary Relationships and Cardinality Ratio**

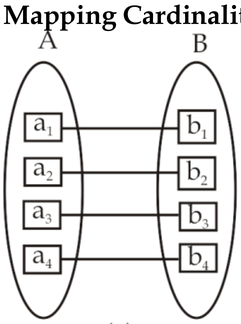



- The number of entities from *E2* that an entity from *E1* can possibly be associated thru *R* (and vice-versa) determines the *cardinality ratio* of *R.*
- Four possibilities are usually specified:
  1. *one-to-one (1:1)*
  2. *one-to-many (1:N)*
  3. *many-to-one (N:1)*
  4. *many-to-many (M:N)*

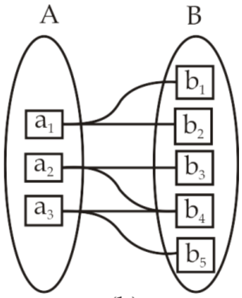
**Cardinality Ratios**

*One-to-one:* An *E1* entity may be associated with at most one *E2* entity and similarly an *E2* entity may be associated with at most one *E1* entity.
*One-to-many:* An *E1* entity may be associated with many *E2* entities whereas an *E2* entity may be associated with at most one *E*1 entity.
*Many-to-one:* … (similar to above)
*Many-to-many:* Many *E1* entities may be associated with a single *E2* entity and a single *E1* entity may be associated with many *E2* entities.

**Mapping Cardinalities**




(a)
One to one

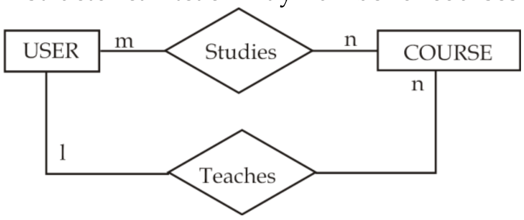



(b)
One to many

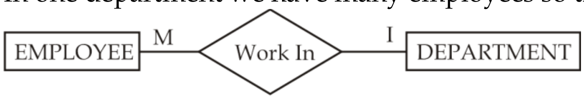
Note: Some elements in A and B may not be mapped to any elements in the other set.

*Many to one Many to many*
Note: Some elements in A and B may not be mapped to any elements in the other set.

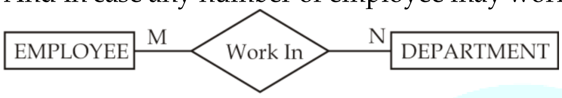
**Alternative Representation of Cardinality**
Many to Many Relationship b/w User and course meaning any number of users can study or enroll in any number of course and these is one to many relationship b/w a teacher (which is also a user) and course meaning only one instructor can teach may number of courses




In one department we have many employees so the following represents – one to many relationships




And in case any number of employee may work in any number of department (many to many)

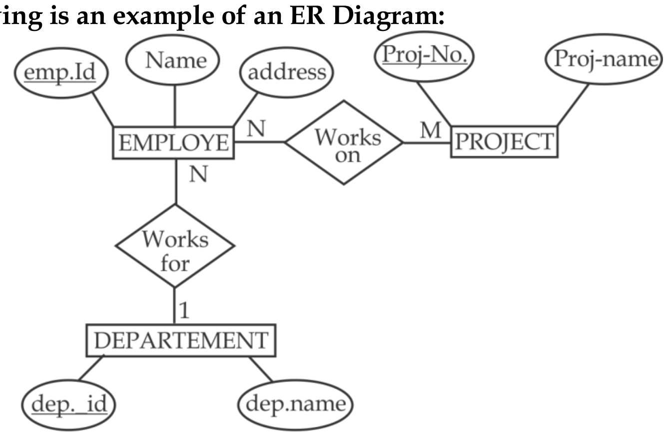



**5. Keys**
A ***super key*** of an entity set is a set of one or more attributes whose values uniquely determine each entity. A ***candidate key*** of an entity set is a minimal super key

For Example: *Customer-id* is candidate key of *customer*
*account-number* is candidate key of *account*
Although several candidate keys may exist, one of the candidate keys is selected to be the ***primary key***.

**Keys for Relationship Sets**
The combination of primary keys of the participating entity sets forms a super key of a relationship set.
(*customer-id, account-number*) is the super key of *depositor*
- *NOTE: this means a pair of entity sets can have at most one relationship in a particular relationship set.*
- E.g. if we wish to track all access-dates to each account by each customer, we cannot assume a relationship for each access. We can use a multivalued attribute though.
- Must consider the mapping cardinality of the relationship set when deciding the what are the candidate keys.
- Need to consider semantics of relationship set in selecting the *primary key* in case of more than one candidate key.

**Following is an example of an ER Diagram:**




**Entity- Relationship (E-R) Diagram**
The overall logical structure of a database can be expressed graphically by an E-R diagram. The diagram consists of the following major components.

- Rectangles: represent entity set.
- Ellipses: represent attributes.
- Diamonds: represents relationship sets.
- Lines: links attribute set to entity set and entity set to relationship set.
- Double ellipses: represent multi-valued attributes.
- Dashed ellipses: denote derived attributes.
- Double lines: represent total participation of an entity in a relationship set.
- Double rectangles: represent weak entity sets.

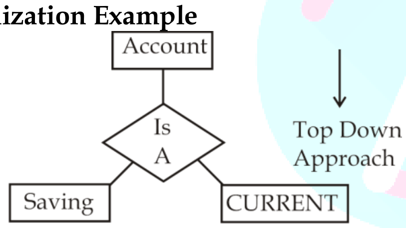
## 3.3 Specialization, Generalization and Aggregation

Generalization is a bottom-up approach in which two lower level entities combine to form a higher-level entity. In generalization, the higher-level entity can also combine with other lower level entity to make further higher-level entity. Specialization is opposite to Generalization. It is a top-down approach in which one higher level entity can be broken down into two lower level entities.

Top-down design process; we designate subgroupings within an entity set that are distinctive from other entities in the set.

- These subgroupings become lower-level entity sets that have attributes or participate in relationships that do not apply to the higher-level entity set.
- Depicted by a *triangle* component labelled ISA (E.g. *customer* "is a" *person*).
- **Attribute inheritance** – a lower-level entity set inherits all the attributes and relationship participation of the higher-level entity set to which it is linked.

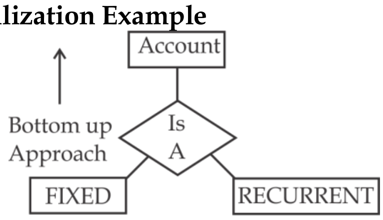
**Specialization Example**




**Generalization**

- A bottom-up design process – combine a number of entity sets that share the same features into a higher-level entity set.
- Specialization and generalization are simple inversions of each other; they are represented in an E-R diagram in the same way.
- The terms specialization and generalization are used interchangeably.

**Generalization Example**




**Specialization and Generalization**

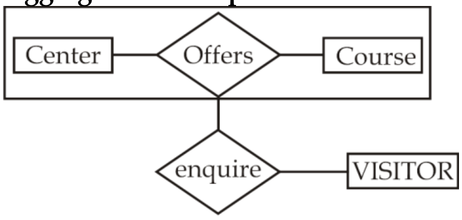
- Can have multiple specializations of an entity set based on different features.
- E.g. *permanent-employee* vs. *temporary-employee*, in addition to *officer* vs. *secretary* vs. *teller*
- Each particular employee would be a member of one of *permanent-employee* or *temporary-employee*, and also a member of one of *officer*, *secretary*, or *teller*
- The IS-A relationship also referred to as **superclass - subclass** relationship.

**Aggregation**

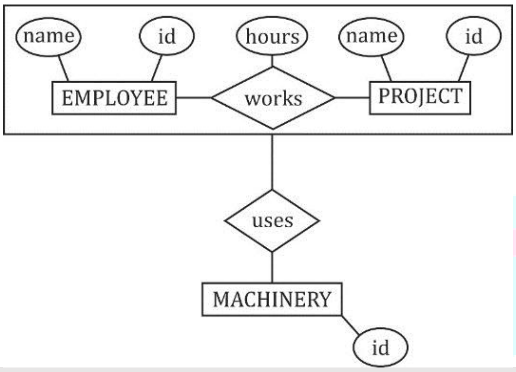
Aggregation is a process when relation between two entities is treated as a single entity. It is an abstraction that treats relationships as entities.

- Eliminate this redundancy via *aggregation*
- Treat relationship as an abstract entity
- Allows relationships between relationships
- Abstraction of relationship into new entity

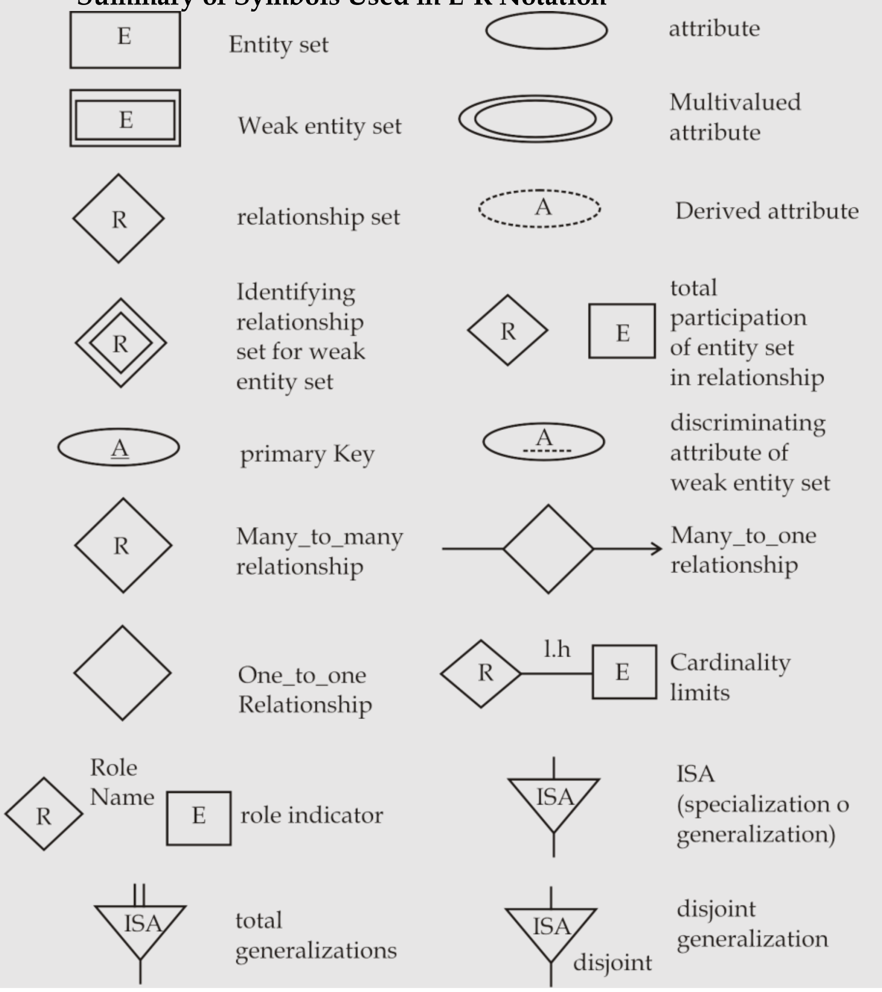
**Aggregation Example**




**ER Diagram with Aggregation**




**Summary of Symbols Used in E-R Notation**



# 4. Relational Database Management System

**The relational model** for database management is a data model based on predicate logic and set theory. It was invented by Edgar Codd. The fundamental assumption of the relational model is that all data are represented as mathematical n-ary **relations**, an n-ary relation being a subset of the Cartesian product of n sets.

**n-ary Relationship**
When there are n entities set participating in a relation, the relationship is called as n-ary relationship.

**1) Relation -** The fundamental organizational structure for data in the relational model is the relation. A relation is a two-dimensional table made up of rows and columns. Each relation also called a table, stores data about entities.
**2) Tuples -** The rows in a relation are called tuples. They represent specific occurrences (or records) of an entity. Each row consists of a sequence of values, one for each column in the table. In addition, each row (or record) in a table must be unique. A tuple variable is a variable that stand for a tuple.
**3) Attributes** – The column in a relation is called attribute. The attributes represent characteristics of an entity.
**4) Domain** – For each attribute there is a set of permitted values called domain of that attribute. For all relations 'r', the domain of all attributes of 'r' should be atomic. A domain is said to be **atomic** if elements of the domain are considered to be indivisible units.

**Database Schema** – Logical design of the database is termed as database schema.
**Database instance** – Database instance is a snapshot of the data in a database at a given instant of time.
**Relation schema** – The concept of relation schema corresponds to the programming notion of type definition. It can be considered as the definition of a domain of values. The database schema is the collection of relation schemas that define a database.
**Relation instance** – The concept of a relation instance corresponds to the programming language notion of a value of a variable. For relation instance, we actually mean the "relation" itself.

## 4.1 Database Keys

1. **Primary Key:** Which uniquely identifies a record in a table. Student_ID is the primary key in this STUDENT Table.
2. **Candidate Key:** A candidate key is a single field or the least combination of fields that uniquely identifies each record in the table. Every table must have at least one candidate key but at the same time can have several. For Example in the table STUDENT, Student_ID and Roll_No. Are Candidate keys.

| Roll_No. | Student_ID |
|---|---|
| 001 | 11093100 |
| 002 | 11093101 |
| 003 | 11093126 |
| 004 | 11093127 |

3. **Foreign Key:** A foreign key is generally a primary key from one table that appears as a field in another. For Example let us consider these two table STUDENT and LIBRARY_RECORD.

**STUDENT**

| Roll_No. | Student_ID | Student_Name | Student_Class |
|---|---|---|---|
| 001 | 11093100 | Ravi Kumar | 3 |
| 002 | 11093101 | Nihal Sharma | 4 |
| 003 | 11093126 | Astha Mathur | 3 |
| 004 | 11093127 | Nishi Arora | 5 |

**LIBRARY_RECORD**

| Lib_CardNo | Student_ID | Student_Name | Address |
|---|---|---|---|
| AX120 | 11093101 | Nihal Sharma | 12th Avenue Street, Delhi |
| AX121 | 11093126 | Astha Mathur | XYZ Lane,Delhi |
| BL101 | 11093127 | Nishi Arora | 5-D, Z Block, Delhi |

In the table LIBRARY_RECORD Lib_CardNo. Is the Primary key and Student_ID is the foreign key as it is the primary key of the table STUDENT.

4. **Alternate Key:** The candidate key other than primary key is called as alternate key.
5. **Super Key:** The set of attributes which can uniquely identify a tuple is known as Super Key. For Example, Student_Enroll_No, (Student_ID, Student_Name) etc.

Non-key attributes are attributes other than candidate key attributes in a table. And Non-prime Attributes are attributes other than Primary attribute.

## 4.2 Relational Query Languages

Relational query languages use relational algebra to break the user requests and instruct the DBMS to execute the requests. It is the language by which user communicates with the database. These relational query languages can be procedural or non-procedural.

- In a procedural language, the user instructs the system to perform a sequence of operations on the database to compute the desired result.

- In a nonprocedural language, the user describes the desired information without giving a specific procedure for obtaining that information.

## 4.3 Relational Algebra

Relational algebra is a procedural query language. It takes one or more relations / tables and performs the operation and produce the result. This result is also considered as a new table or relation. Specifically, since the result of a relational query is itself a relation, relational operations can be applied to the results of queries as well as to the given set of relations.

An operator can be either unary or binary. Following are some operations of relational algebra:

1. **Selection operator (σ):** Selection operator is used to select tuples from a relation based on some condition.
   **Syntax:** $\sigma_{(Cond)}$**(Relation Name)**

**Example**

Extract employees whose age is greater than 30 from EMPLOYEES relation

$\sigma_{(AGE>30)}(EMPLOYEES)$

2. **Project Operation (∏) :** It projects column(s) that satisfy a given predicate.
   **Syntax:** $\prod_{(Column\ 1, Column\ 2....Column\ n)}$**(Relation Name)**

**Example**

Extract EMP_ID and NAME from EMPLOYEE relation.

$\prod_{(EMP\_ID, NAME)}(EMPLOYEE)$

3. **Union Operation (∪):** It performs binary union between two given relations. Union on two relations R1 and R2 can only be computed if R1 and R2 are union compatible (These two relation should have same number of attributes and corresponding attributes in two relations have same domain) . Duplicate tuples are automatically eliminated in union operation.
   **Syntax: Relation1 U Relation2**
   **$r \cup s = \{ t \mid t \in r \text{ or } t \in s\}$**
   Note: r, and s must have the same number of attributes.

**Example**

Projects the names of th Employees who are Managers in IT_Dept or Managers in FUNC_Dept or Both

$\prod$ Managers (IT_Dept) ∪ $\prod$ Managers(FUNCT_Dept)

4. **Minus (-):** Minus on two relations R1 and R2 can only be computed if R1 and R2 are union compatible. Minus operator when applied on two relations as R1-R2 will give a relation with tuples which are in R1 but not in R2.
**Syntax: Relation1 - Relation2**

**Example**

Find person who are student but not employee, we can use minus operator like:

$\prod$ Name (STUDENTS) − $\prod$ Name (EMPLOYEE)

5. **Rename(ρ):** Rename operator is used to give another name to a relation.
**Syntax: ρ(Relation2, Relation1)**

**Example**

To rename STUDENT relation to STUDENT1, we can use rename operator like:

ρ(STUDENT1, STUDENT)

6. **Cartesian Product (X):** The cartesian product of two tables combines each row in one table with each row in the other table. It combines tuples from two relations, but unlike the join operation, its result contains all pairs of tuples from the two relations, regardless of whether their attribute values match.
Syntax: r X s
Where r and s are relations and their output will be defined as −
r X s = { q t | q ∈ r and t ∈ s}

**Tuple Relational Calculus**

Relational calculus is a non-procedural query language. It uses mathematical predicate calculus instead of algebra. It provides the description about the query to get the result whereas relational algebra gives the method to get the result.

A query in the tuple relational calculus is expressed as: {t | P(t)}
i.e. the set of tuples for which predicate is true.

{t | EMPLOYEE (t) and t.SALARY>20000} - implies that it selects the tuples from EMPLOYEE relation such that resulting employee tuples will have salary greater than 20000. It is example of selecting a range of values.

**Domain Relational Calculus**

In the tuple relational calculus, you have use variables that have series of tuples in a relation. In the domain relational calculus, you will also use variables but in this case the variables take their values from domains of attributes rather than tuples of relations. An domain relational calculus expression has the following general format –

{d1, d2, . . . , dn | F(d1, d2, . . . , dm)} m ≥ n
where d1, d2, . . . , dn, . . . , dm stand for domain variables and F(d1, d2, . . . , dm) stands for a formula

composed of atoms.

For example, select EMP_ID and EMP_NAME of employees who work for department ID 415

{<EMP_ID, EMP_NAME> | <EMP_ID, EMP_NAME> ? EMPLOYEE Λ DEPT_ID = 415}

## 5. Normalization

Normalization is a process of organizing the data in database to avoid data redundancy, insertion anomaly, update anomaly & deletion anomaly. Thus, database normalization is a database schema design technique, by which an existing schema is modified to minimize redundancy and dependency of data.

**What are Anomalies in Database Management?**

Anomalies are inconvenient or error-prone situation arising when we process the tables. There are three types of anomalies:

1. **Update Anomalies:** Incorrect data may have to be changed, which could involve many records having to be changed, leading to the possibility of some changes being made incorrectly.
2. **Delete Anomalies:** A record of data can legitimately be deleted from a database, and the deletion can result in the deletion of the only instance of other, required data, E.g. Deleting a book loan from a library member can remove all details of the particular book from the database such as the author, book title etc
3. **Insert Anomalies:** The nature of a database may be such that it is not possible to add a required piece of data unless another piece of unavailable data is also added. E.g. A library database that cannot store the details of a new member until that member has taken out a book.

### 5.1 Functional Dependency

We use functional dependencies to: test relations to see if they are legal under a given set of functional dependencies.
If a relation *r* is legal under a set *F* of functional dependencies, we say that *r* satisfies *F.*
We say that *F* holds on *R* if all legal relations on *R* satisfy the set of functional dependencies *F.*
Note: A specific instance of a relation schema may satisfy a functional dependency even if the functional dependency does not hold on all legal instances. For example, a specific instance of *loan* may, by chance, satisfy *amount →customer_name.*
*A* functional dependency is trivial if it is satisfied by all instances of a relation
Example:
*customer_name, loan_number →customer_name*
*customer_name→customer_name*

**Inference rules**
Armstrong's axioms - sound and complete i.e, enable the computation of any functional dependency. Functional dependencies are:
1. Reflexivity - if the B's are a subset of the A's then A → B
2. Augmentation - If A → B, then A, C → B, C.
3. Transitivity - If A → B and B → C then A → C.
*Additional inference rules*
4. Decomposition - If A → B, C then A → B
5. Union - If A → B and A → C then A → B, C
6. Pseudo transitive - If A → B and C, B → D then C, A → D
**Equivalence of sets of functional dependencies**
Two functional dependencies S & T are equivalent iff S→ T and T → S.
The dependency {A_1, ..., A_n} → {B_1, ..., B_m} is trivial if the B's are a subset of the A's is nontrivial if at least one of the B's is not among the A's is completely nontrivial if none of the B's is also one of the A's

**Closure ($F^+$)**
All dependencies that include F and that can be inferred from F using the above rules are called closure of F denoted by $F^+$.
After finding a set of functional dependencies that are hold on a relation, the next step is to find the Super key for that relation (table). The set of identified functional dependencies play a vital role in finding the key for the

relation. We can decide whether an attribute (or set of attributes) of any table is a key for that table or not by identifying the attribute or set of attributes' closure. If A is an attribute, (or set of attributes) then its attribute closure is denoted as A+.

**Algorithm to compute closure**

We have to find out whether $F \models X \rightarrow Y$. This is the case when $X \rightarrow Y \in F^+$

The better method is to generate $X^+$, closure of X under F and test $F \models X \rightarrow Y$ using the first two axioms augmentation and reflexive rules.

***Example:***

EMPLOYEE(empid, empname, dept, age, salary, experience)

Let the functional dependencies be as follows:

empid->empname
{age,experience}->salary
empid->{age,dept}
dept->experience

In above example, let us find the closure of the attribute empid, i.e, closure of {empid}

Since we are finding closure of empid. empid is an element of the closure set $C^+$. Now we go step by step.

- **Step 1:** Select each functional dependency and check whether the left side of functional dependency is a subset of closure. If yes, add the right side of that functional dependency to closure set. if not, check the next functional dependency
- **Step 2:** Keep on checking the functional dependencies until there is no more functional dependencies with its left side as a subset of closure set $C^+$.

***What is a subset?*** A set M is said to be a subset of another set N only if all elements of set M is present in set N. Set N can have more elements than M.

- So, in our example, empid is an element of the closure set $C^+$. So, initially, $C^+$={empid}.
- First functional dependency says empid functionally determines empname. Its left side ( {empid}) is subset of $C^+$. Therefore, its right side is added to $C^+$. Now $C^+$={empid, empname}.
- Second fd (functional dependency) says {age, experience}->salary. Here left side ( {age, experience} ) is not a subset of $C^+$. So we check the next fd.
- Third fd says, empid->{age, dept}. Here left side ( {empid} ) is subset of $C^+$. Therefore, its right side is added to $C^+$. Now, C becomes, $C^+$={empid, empname, age, dept}.
- Fourth fd says, dept->experience. Here left side ( {dept} ) is a subset of $C^+$. So we are adding its right side ( {experience} ) to Closure set. Now, $C^+$={empid, empname, age, dept, experience}.
- We are looking again for a functional dependency with its left side as a subset of closure set. Since the closure set $C^+$ is getting changed in some steps, there is more possibility to find another functional dependency with its left side as a subset of $C^+$. Again, we go through every functional dependency.
- Since sets do not allow duplication, we should do nothing if the right side of a functional dependency whose left side is subset of $C^+$, is already present in closure set $C^+$.
- Second fd has a left side that is subset of C+. {age, experience}->salary. Therefore, salary is added to $C^+$. Now, $C^+$={empid, empname, age, dept, experience, salary}.
- There isn't any more functional dependency whose left side is subset of C+ and give at least one new attribute to closure set. Therefore, we stop now.

Now closure of set C is C+={empid, empname, age, dept, experience, salary}.

**Minimal Cover of FD**

We say that a set of functional dependencies F covers another set of functional dependencies G, if every functional dependency in G can be inferred from F. More formally, F covers G if $G+ \subseteq F^+$. F is a minimal cover of G if F is the smallest set of functional dependencies that cover G.

We find the minimal cover by iteratively simplifying the set of functional dependencies. To do this, we will use three methods:

**Simplifying an FD by the Union Rule:** Let X, Y , and Z be sets of attributes. If $X \rightarrow Y$ and $X \rightarrow Z$, then $X \rightarrow Y Z$

**Simplifying an FD by simplifying the left-hand side:** Let X and Y be sets of attributes, and B be a single attribute not in X.

Let F be: $XB \to Y$
and H be $X \to Y$
If $F \Rightarrow X \to Y$ , then we can replace F by H. In other words, if $Y \subseteq X^+_F$, then we can replace F by H.
For example, say we have the following functional dependencies (F):
- $AB \to C$
- $A \to B$

And we want to know if we can simplify to the following (H):
- $A \to C$
- $A \to B$

Then $A^+_F = ABC$. Since, $Y \subseteq X^+_F$, we can replace F by H.

**Simplifying an FD by simplifying the right-hand side:** Let X and Y be sets of attributes, and C be a single attribute not in Y .
Let F be:
$X \to Y\ C$ and H be

$X \to Y$
If $H \Rightarrow X \to Y\ C$, then we can replace F by H. In other words, if $Y\ C \subseteq X^+_H$, then we can replace F by H.
For example, say we have the following functional dependencies (F):
- $A \to BC$
- $B \to C$

And we want to know if we can simplify to the following (H):
- $A \to B$
- $B \to C$

Then $A^+_H = ABC$. Since, $BC \subseteq X^+_H$, we can replace F by H.

**Finding the Minimal Cover**
Given a set of functional dependencies F :
1. Start with F
2. Remove all trivial functional dependencies
3. Repeatedly apply (in whatever order you like), until no changes are possible
- Union Simplification (it is better to do it as soon as possible, whenever possible)
- RHS Simplification
- LHS Simplification

4. Result is the minimal cover

Example: Applying to algorithm to EGS with
1. $E \to G$
2. $G \to S$
3. $E \to S$

*Using the union rule*, we combine 1 and 3 and get
1. $E \to GS$
2. $G \to S$

*Simplifying RHS of 1* (this is the only attribute we can remove), we get:
1. $E \to G$
2. $G \to S$

***Algorithm to find minimal cover for a set of FDs F***
Step 1: Let G be the set of FDs obtained from F by decomposing the
right hand sides of each FD to a single attribute.
Step 2: Remove all redundant attributes from the left hand sides of FDs in G.
Step 3: From the resulting set of FDs, remove all redundant FDs.
Output the resulting set of FDs.

**Example:** Consider R = ABCDEFGH and the following set of FDs, F:
$ABH \to C$

$A \rightarrow D$
$C \rightarrow E$
$BGH \rightarrow F$
$F \rightarrow AD$
$E \rightarrow F$
$BH \rightarrow E$
Converting right hand sides to single attributes, we get:
$ABH \rightarrow C$
$A \rightarrow D$
$C \rightarrow E$
$BGH \rightarrow F$
$F \rightarrow A$
$F \rightarrow D$
$E \rightarrow F$
$BH \rightarrow E$
Perform steps 2 and 3....

**Understanding Normalization**
In relational database theory, normalization is the process of restructuring the logical data model of a database to eliminate redundancy, organize data efficiently, and reduce repeating data and to reduce the potential for anomalies during data operations. Data normalization also may improve data consistency and simplify future extension of the logical data model. The formal classifications used for describing a relational database's level of normalization are called normal forms (**NF**).

A non-normalized database can suffer from data anomalies: A non-normalized database may store data representing a particular referent in multiple locations. An update to such data in some but not all of those locations results in an update anomaly, yielding inconsistent data. A normalized database prevents such an anomaly by storing such data (i.e. data other than primary keys) in only one location.

A non-normalized database may have inappropriate dependencies, i.e. relationships between data with no functional dependencies. Adding data to such a database may require first adding the unrelated dependency. A normalized database prevents such insertion anomalies by ensuring that database relations mirror functional dependencies.
Similarly, such dependencies in non-normalized databases can hinder deletion. That is, deleting data from such databases may require deleting data from the inappropriate dependency. A normalized database prevents such deletion anomalies by ensuring that all records are uniquely identifiable and contain no extraneous information.

**Normal forms**
Edgar F. Codd originally defined the first three normal forms. The first normal form requires that tables be made up of a primary key and a number of atomic fields, and the second and third deal with the relationship of non-key fields to the primary key. These have been summarized as requiring that all non-key fields be dependent on "the key, the whole key and nothing but the key". In practice, most applications in 3NF are fully normalized. However, research has identified potential update anomalies in 3NF databases. BCNF is a further refinement of 3NF that attempts to eliminate such anomalies. The fourth and fifth normal forms (4NF and 5NF) deal specifically with the representation of many-many and one-many relationships. Sixth normal form (6NF) only applies to temporal databases.

## 5.2 First Normal Form (1NF)

First normal form (1NF) lays the groundwork for an organized database design: **Ensure that each table has a primary key: minimal set of attributes which can uniquely identify a record.** It states that the domain of an attribute must include only atomic values and the value of any attribute in a tuple must be single value from the domain of that attribute. It doesn't allow nested relation. Data that is redundantly duplicated across multiple rows of a table is moved out to a separate table.

***Atomicity:*** *Each attribute must contain a single value, not a set of values.*
**Unnormalized form (UNF)**: A table that contains one or more repeating groups.
**First normal form (1NF):** A relation in which the intersection of each row and column contains one and only one value.

**UNF → 1NF: remove repeating groups:**

- Entering appropriate data in the empty columns of rows.
- Placing repeating data along with a copy of the original
- key attribute in a separate relation. Identifying a primary
- key for each of the new relations.

First normal form (1NF) lays the groundwork for an organized database design: Ensure that each table has a primary key: minimal set of attributes which can uniquely identify a record. It states that the domain of an attribute must include only atomic values and the value of any attribute in a tuple must be single value from the domain of that attribute. It doesn't allow nested relation. Data that is redundantly duplicated across multiple rows of a table is moved out to a separate table.

*For Example:*
*Consider a table STUDENT with fields Roll_No, Name, Course. Here a student may have opted for more than one courses thus the values in Course field will not be atomic:*

| Roll_No | Name | Course |
|---|---|---|
| 1 | Snehal | Polity, History, Economics |
| 2 | Kajal | DBMS, CD |
| 3 | Amit | Physics, Chemistry |

After converting it to First Normal Form (1NF)

| Roll_No | Name | Course |
|---|---|---|
| 1 | Snehal | Polity |
| 1 | Snehal | History |
| 1 | Snehal | Economics |
| 2 | Kajal | DBMS |
| 2 | Kajal | CD |
| 3 | Amit | Physics |
| 3 | Amit | Chemistry |

## 5.3 Second Normal Form (2NF)

General Definition: A relation schema R is in second normal form (2NF) if every nonprime attribute A in R is not partially dependent on any key of R.
Partial Dependency – If proper subset of candidate key determines non-prime attribute, it is called partial dependency.

- Create separate tables for sets of values that apply to multiple records.
- Relate the tables with a foreign key.
- Records should not depend on anything other than a table's primary key (a compound key, if necessary). For example, consider a customer's address in an accounting system. The address is needed by the Customers table, but also by the Orders, Shipping, Invoices, Accounts Receivable, and Collections tables. Instead of storing the customer's address as a separate entry in each of these tables, store it in one place, either in the Customers table or in a separate Addresses table.

A table is said to be in 2NF if both the following conditions hold:

- Table is in 1NF (First normal form)
- No non-prime attribute is dependent on the proper subset of any candidate key of table.

**1NF → 2NF: remove partial dependencies: the functionally dependent attributes are removed from the relation by placing them in a new relation along with a copy of their determinant.**

As per First Normal Form, no two Rows of data must contain repeating group of information i.e each set of column must have a unique value, such that multiple columns cannot be used to fetch the same row. The Primary key is usually a single column, but sometimes more than one column can be combined to create a single primary key which is actually called a Candidate Key. To identify or establish 2NF we must identify Candidate Key and Partial Dependencies.

***Example:*** Let us consider a Relation R with fields A,B,C,D :
R(A B C D)
Here AB-> D and B-> C.
A and B are the essential attributes here as from A and B you can find D and from B you can find C. Thus Candidate Key for this relation R is AB.
$(AB)^+ = ABCD$
AB-> Candidate Key
Hence, AB ∈ Prime Attribute - those who are a part of candidate key
CD ∈ Non-Prime Attribute
**Identify Partial Dependency:** Here C is only dependent on B and not the complete candidate key set AB, thus in this relation R there is a partial dependency.
Now for 2NF a relation must be in 1NF and there should not be any partial dependency. So we'll elimination partial dependency by creating another Relation $R_1$ and $R_2$:
$R_1$ (A B D)
AB->D
$R_2$ (B C)
B->C
In this scenario we have eliminated partial dependency as B is the only candidate key and C is dependent on B.

## 5.4 Third Normal Form (3NF)

For 3NF, first, the table must be in 2NF, plus, we want to make sure that the non-key fields are dependent upon ONLY the PK, and not on any other field in the table. This is very similar to 2NF, except that now you are comparing the non-key fields to OTHER non-key fields.

**3NF Rule:**
1. The database must meet all the requirements of the second normal form.
2. Any field which is dependent not only on the primary key but also on another field is moved out to a separate table. (No Transitive Dependencies)

***Example:***
STUDENT( Stu_ID, Stu_Name, City, Zip)

We find that in the above STUDENT relation, Stu_ID is the key and only prime key attribute.
We find that City can be identified by Stu_ID as well as Zip itself. Neither Zip is a superkey nor is City a prime attribute. Additionally, Stu_ID → Zip → City, so there exists transitive dependency.
To bring this relation into third normal form, we break the relation into two relations as follows:

Student_Detail (Stu_ID, Stu_Name, Zip)
ZipCode(Zip, City)

**General Definition:**
A relation schema R is in 3NF if, whenever a nontrivial functional dependency X→A holds in R,
Either a) X is a Super key Or b) Y is a prime attribute of R.
i.e. A relation schema R is in 3NF if every nonprime attribute of R meets both of the following terms:
1. It is fully functionally dependent on every key of R.
2. It is non-transitively dependent on every key of R.

## 5.5 Boyce-Codd Normal Form (BCNF)

A row is in BCNF if and only if every determinant is a candidate key. The second and third normal forms assume that all attributes not part of the candidate keys depend on the candidate keys but does not deal with dependencies within the keys.

BCNF deals with such dependencies.
*A relation R is said to be in BCNF if whenever **X -> A** holds in R, and A is not in X, then X is a candidate key for R.*

BCNF covers very specific situations where 3NF miss interdependencies between non-key attributes. It should be noted that most relations that are in 3NF are also in BCNF.

Infrequently, a 3NF relation is not in BCNF and this happens only if
(a) the candidate keys in the relation are composite keys (that is, they are not single attributes),
(b) there is more than one candidate key in the relation, and
(c) the keys are not disjoint, that is, some attributes in the keys are common.
**The BCNF differs from the 3NF only when there are more than one candidate keys and the keys are composite and overlapping**. Consider for example, the relationship *enrol (sno, sname, cno, cname, date-enrolled)*
Let us assume that the relation has the following candidate keys:
*(sno, cno)*
*(sno, cname)*
*(sname, cno)*
*(sname, cname)*
(we have assumed *sname* and *cname* are unique identifiers). The relation is in 3NF but not in BCNF because there are dependencies
*sno -> sname*
*cno -> cname*
where attributes that are part of a candidate key are dependent on part of another candidate key. Such dependencies indicate that although the relation is about some entity or association that is identified by the candidate keys.
e.g. *(sno, cno)*, there are attributes that are not about the whole thing that the keys identify. For example, the above relation is about an association (enrolment) between students and subjects and therefore the relation needs to include only one identifier to identify students and one identifier to identify subjects. Providing two identifiers about students *(sno, sname)* and two keys about subjects *(cno, cname)* means that some information about students and subjects that is not needed is being provided. This provision of information will result in repetition of information and the anomalies. If we wish to include further information about students and courses in the database, it should not be done by including the information in the present relation but by creating new relations that represent information about entities *student* and *subject*.
These difficulties may be overcome by decomposing the above relation in the following three relations:
*(sno, sname)*
*(cno, cname)*
*(sno, cno, date-of-enrolment)*
We now have a relation that only has information about students, another only about subjects and the third only about enrolments. All the anomalies and repetition of information have been removed.

## 5.6 Multivalued Dependency and Fourth Normal Form (4NF)

In a relational model, if all of the information about an entity is to be represented in one relation, it will be necessary to repeat all the information other than the multivalue attribute value to represent all the information that we wish to represent. This results in many tuples about the same instance of the entity in the relation and the relation having a composite key (the entity id and the multivalued attribute). Of course, the other option suggested was to represent this multivalued information in a separate relation. The situation of course becomes much worse if an entity has more than one multivalued attributes and these values are represented in one relation by a number of tuples for each entity instance. The multivalued dependency relates to this problem when more than one multivalued attributes exist. Consider the following relation that represents an entity employee that has one multivalued attribute *proj*:
*emp (e#, dept, salary, proj)*

We have so far considered normalization based on functional dependencies; dependencies that apply only to single-valued information. For example, *e# -> dept* implies only one *dept* value for each value of *e#*. Not all information in a database is single-valued, for example, *proj* in an employee relation may be the list of all projects that the employee is currently working on. Although *e#* determines the list of all projects that an employee is working on, *e# -> proj* is not a functional dependency.

We can more clearly analyze the multivalued dependency by the following example.
*programmer (emp_name, qualifications, languages)*
This relation includes two multivalued attributes of entity *programmer*; *qualifications* and *languages*. There are no functional dependencies.

The attributes qualifications and languages are assumed independent of each other. If we were to consider *qualifications* and *languages* as separate entities, we would have two relationships (one between *employees* and *qualifications* and the other between *employees* and programming *languages*). Both the above relationships are many-to-many i.e. one programmer could have several qualifications and may know several programming languages. Also one qualification may be obtained by several programmers and one programming language may be known to many programmers.
Functional dependency *A -> B* relates one value of *A* to one value of *B* while multivalued dependency *A ->> B* defines a relationship in which a set of values of attribute *B* are determined by a single value of *A*.

Now, more formally, *X ->> Y* is said to hold for *R(X, Y, Z)* if *t1* and *t2* are two tuples in *R* that have the same values for attributes *X* and therefore with *t1[x] = t2[x]* then *R* also contains tuples *t3* and *t4* (not necessarily distinct) such that
*t1[x] = t2[x] = t3[x] = t4[x]*
*t3[Y] = t1[Y]* and *t3[Z] = t2[Z]*
*t4[Y] = t2[Y]* and *t4[Z] = t1[Z]*
In other words if *t1* and *t2* are given by
*t1 = [X, Y1, Z1]*, and
*t2 = [X, Y2, Z2]*
then there must be tuples *t3* and *t4* such that
*t3 = [X, Y1, Z2]*, and
*t4 = [X, Y2, Z1]*

We are therefore insisting that every value of *Y* appears with every value of *Z* to keep the relation instances consistent. In other words, the above conditions insist that *X alone determines Y and Z* and there is no relationship between *Y* and *Z* since *Y* and *Z* appear in every possible pair and hence these pairings present no information and are of no significance.

**Fourth Normal Form**
Fourth normal form (or 4NF) requires that **there must be no non-trivial multivalued dependencies of attribute sets on something other than a superset of a candidate key.** A table is said to be in 4NF if and only if it is in the BCNF and multivalued dependencies are functional dependencies. The 4NF removes unwanted data structures: multivalued dependencies.

**Definition:** A relation schema R is in 4NF with respect to a set of dependencies F, if, for every non-trivial multivalued dependency X ->>Y in F+, X is a super key for R.

## 5.7 Properties of Relational Decompositions

If R doesn't satisfy a particular normal form, we decompose R into smaller schemas
**What's a decomposition?**
$R = (A_1, A_2, \ldots, A_n)$
$D = (R_1, R_2, \ldots, R_k)$ st $R_i \subseteq R$ and $R = R_1 \cup R_2 \cup \ldots \cup R_k$
($R_i$'s need not be disjoint)
Replacing R by $R_1, R_2, \ldots, R_k$ – process of decomposing R
Ex: gradeInfo (rollNo, studName, course, grade)
$R_1$: gradeInfo (rollNo, course, grade)
$R_2$: studInfo (rollNo, studName)

Decomposition Property: A relation must satisfy the following two properties during decomposition.
**i. Lossless**: A lossless-join dependency is a property of decomposition, which ensures that spurious rows are generated when relations are united through a natural join operation. i.e. The information in an instance r of R must be preserved in the instances r1, r2, r3, …..rk where ri = ΠRi (r)
Decomposition is lossless with respect to a set of functional dependencies F if, for every relation instance r on R satisfying F, $r = \pi_{R_1}(r) * \pi_{R_2}(r) * \ldots * \pi_{R_k}(r)$

**Lossless join property**
F – set of FDs that hold on R
R – decomposed into $R_1, R_2, \ldots, R_k$

Decomposition is lossless wrt F if for every relation instance r on R satisfying F,

| |
|---|
| Lossless joins are also called non-additive joins |

$r = \pi_{R_1}(r) * \pi_{R_2}(r) * . . . * \pi_{R_k}(r)$

R = (A, B, C); $R_1$ = (A, B); $R_2$ = (B, C)

Original info is distorted

r:

| A | B | C |
|---|---|---|
| $a_1$ | $b_1$ | $c_1$ |
| $a_2$ | $b_2$ | $c_2$ |
| $a_3$ | $b_1$ | $c_3$ |

Lossy join

$r_1$:

| A | B |
|---|---|
| $a_1$ | $b_1$ |
| $a_2$ | $b_2$ |
| $a_3$ | $b_1$ |

$r_2$:

| B | C |
|---|---|
| $b_1$ | $c_1$ |
| $b_2$ | $c_2$ |
| $b_1$ | $c_3$ |

$r_1 * r_2$:

| A | B | C |
|---|---|---|
| $a_1$ | $b_1$ | $c_1$ |
| $a_1$ | $b_1$ | $c_3$ |
| $a_2$ | $b_2$ | $c_2$ |
| $a_3$ | $b_1$ | $c_1$ |
| $a_3$ | $b_1$ | $c_3$ |

Spurious tuples

**ii. Dependency Preserving Property:** If a set of functional dependencies hold on R it should be possible to enforce F by enforcing appropriate dependencies on each $r_1$ .
Decomposition D= ($R_1$, $R_2$, $R_3$, ………, $R_k$) of schema R preserves a set of dependencies
F if,
($\pi_{R1}$ (F) U π R2 (F) U . . . . . . . . . . . π Rn (F)) +=F+
π Ri(F) is the projection of F onto Ri.
i.e Any FD that logically follows from F must also logically follows from the union of projection of F onto Ri 'S . Then D is called dependency preserving.

**Dependency Preserving Decompositions**
Decomposition D = ($R_1$, $R_2$, …, $R_k$) of schema R preserves a set of dependencies F if
$\left(\pi_{R_1}(F) \cup \pi_{R_2}(F) \cup ... \cup \pi_{R_k}(F)\right)^+ = F^+$
Here, $\pi_{R_i}(F) = \{ (X \to Y) \in F^+ \mid X \subseteq R_i, Y \subseteq R_i\}$
(called projection of F onto $R_i$)
Informally, any FD that logically follows from F must also logically follow from the union of projections of F onto $R_i$'s Then, D is called dependency preserving.

**Join Dependency**
Join dependency is the term used to indicate the property of a relation schema that cannot be decomposed losslesly into two relations schema, but can be decomposed losslesly into three or more simpler relation schema. It means that a table, after it has been decomposed into three or more smaller tables must be capable of being joined again on common keys to form the original table.

**Algorithm for BCNF decomposition**
R – given schema. F – given set of FDs
D = {R}  // initial decomposition
while there is a relation schema $R_i$ in D that is not in BCNF do
{let X → A be the FD in $R_i$ violating BCNF;
Replace $R_i$ by $R_{i1} = R_i - \{A\}$ and $R_{i2} = X \cup \{A\}$ in D;}

Decomposition of $R_i$ is lossless as
$R_{i1} \cap R_{i2} = X$, $R_{i2} - R_{i1} = A$ and X → A
Result: a lossless decomposition of R into BCNF relations

## 5.8 Fifth Normal Form

Fifth normal form (5NF also called PJ/NF) requires that there are no non-trivial join dependencies that do not follow from the key constraints. A table is said to be in the 5NF if and only if it is in 4NF and the candidate keys imply every join dependency in it.

**Goals of Normalization**
- Let *R* be a relation scheme with a set *F* of functional dependencies.
- Decide whether a relation scheme *R* is in "good" form.

- In the case that a relation scheme *R* is not in "good" form, decompose it into a set of relation scheme {*R*1, *R*2, ..., *R n*} such that each relation scheme is in good form the decomposition is a lossless-join decomposition Preferably, the decomposition should be dependency preserving.

**'A relation R is in fifth normal form (5NF) – also called projection-join normal form (PJ/NF) if and only if every join dependency in R is a consequence of the candidate keys of R.'**

For every normal form it is assumed that every occurrence of R can be uniquely identified by a primary key using one or more attributes in R.

**For R to be in 5th Normal Form it should follow the following rules:**
1. It should be in 4NF and
2. If Join Dependency does not exist
   else
   If Join Dependency exits then it should be Trivial in Nature
   else
   If all the Ri are Super Key; R-> ($R_1$, $R_2$, $R_3$ ... $R_i$)

- **Join dependencies** generalize multivalued dependencies lead to **project-join normal form (**PJNF) (also called **fifth normal form**)
- A class of even more general constraints, leads to a normal form called **domain key normal form**.
- Problem with these generalized constraints: are hard to reason with, and no set of sound and complete set of inference rules exists. Hence rarely used.

In short, Normalization of a Database is achieved by following a set of rules called 'forms' in creating the database.
These rules are 5 in number (with one extra one stuck in-between 3&4) and they are:

- **1st Normal Form or 1NF:** Each Column Type is Unique.
- **2nd Normal Form or 2NF:** The entity under consideration should already be in the 1NF and all attributes within the entity should depend solely on the entity's unique identifier.
- **3rd Normal Form or 3NF:** The entity should already be in the 2NF and no column entry should be dependent on any other entry (value) other than the key for the table. If such an entity exists, move it outside into a new table.
- Now if these 3NF are achieved, the database is considered normalized. But there are three more 'extended' NF for the elitist. These are:
- **BCNF (Boyce & Codd):** The database should be in 3NF and all tables can have only one primary key.
- **4NF:** Tables cannot have multi-valued dependencies on a Primary Key.
- **5NF:** There should be no cyclic dependencies in a composite key.

## 6. Structured Query Language

**What's SQL ?**
In 1971, IBM researchers created a simple non-procedural language called Structured English Query Language. or SEQUEL. This was based on Dr. Edgar F. (Ted) Codd's design of a relational model for data storage where he described a universal programming language for accessing databases.

In the late 80's ANSI and ISO (these are two organizations dealing with standards for a wide variety of things) came out with a standardized version called Structured Query Language or SQL. SQL is prounced as 'Sequel'. There have been several versions of SQL and the latest one is SQL-99. Though SQL-92 is the current universally adopted standard.
SQL is the language used to query all databases. It's simple to learn and appears to do very little but is the heart of a successful database application. Understanding SQL and using it efficiently is highly imperative in designing an efficient database application. The better your understanding of SQL the more versatile you'll be in getting information out of databases.

**Onto SQL**
There are four basic commands which are the workhorses for SQL and figure in almost all queries to a database.
INSERT - Insert Data
DELETE - Delete Data
SELECT - Pull Data
UPDATE - Change existing Data

As you can see SQL is like English. Let's build a real world example database using MySQL and perform some SQL operations on it. A database that practically anyone could use would be a Contacts database. In our example we are going to create a database with the following fields:

*First Name, Last Name, Birth Date, Street Address, City, State, Zip, Country, Telephone Home, Telephone Work, Email, Company Name, Designation*

**First, let's decide how we are going to store this data in the database. For illustration purposes, we are going to keep this data in multiple tables.**
This will let us exercise all the SQL commands pertaining to retrieving data from multiple tables. Also, we can separate different kinds of entities into different tables. So, let's say you have thousands of friends and need to send a mass email to all of them, a SELECT statement (covered later) will look at only one table.
Well, we can keep the FirstName, LastName and BirthDate in one table.
Address related data in another.
Company Details in another.
Emails can be separated into another.
Telephones can be separated into another.

**Let's build the database in MySQL**
While building a database - you need to understand the concept of data types. Data types allow the user to define how data is stored in fields or cells within a database. It's a way to define how your data will actually exist. Whether it's a Date or a string consisting of 20 characters, an integer etc. When we build tables within a database we also define the contents of each field in each row in the table using a data type. It's imperative that you use only the data type that fits your needs and don't use a data type that reserves more memory than the data in the field actually requires.
**Let's look at various Data Types under MySQL**

| Type | Size in bytes | Description |
|---|---|---|
| TINYINT (length) | 1 | Integer with unsigned range of 0-255 and a signed range from -128-127 |
| SMALLINT (length) | 2 | Integer with unsigned range of 0-65535 and a signed range from -32768-32767 |
| MEDIUMINT(length) | 3 | Integer with unsigned range of 0-16777215 and a signed range from -8388608-8388607 |
| INT(length) | 4 | Integer with unsigned range of 0-429467295 and a signed range from -2147483648-2147483647 |
| BIGINT(length) | 8 | Integer with unsigned range of 0-18446744 and a signed range from -9223372036854775808-9223372036854775807 |
| FLOAT(length, decimal) | 4 | Floating point number with max. value +/-3.402823466E38 and min.(non-zero) value +/11.175494351E-38 |
| DOUBLEPRECISION(length, decimal) | 8 | Floating point number with max. value +/- -1.7976931348623157E308 and min. (non-zero) value +/-2.2250738585072014E-308 |
| DECIMAL(length, decimal) | length | Floating point number with the range of the DOUBLE type that is stored as a CHAR field type. |
| TIMESTAMP(length) | 4 | YYYYMMDDHHMMSS or YYMMDDHHMMSS or YYYYMMDD, YYMMDD. A Timestamp value is updated each time the row changes value. A NULL value sets the field to the current time. |
| DATE | 3 | YYYY-MM-DD |
| TIME | 3 | HH:MM:DD |
| DATETIME | 8 | YYYY-MM-DD HH:MM:SS |
| YEAR | 1 | YYYY or YY |

| | | |
|---|---|---|
| CHAR(length) | **length** | A fixed length text string where fields shorter than the assigned length are filled with trailing spaces. |
| VARCHAR(length) | **length** | A fixed length text string (255 Character Max) where unused trailing spaces are removed before storing. |
| TINYTEXT | **length+1** | A text field with max. length of 255 characters. |
| TINYBLOB | **length+1** | A binary field with max. length of 255 characters. |
| TEXT | **length+1** | 64Kb of text |
| BLOB | **length+1** | 64Kb of data |
| MEDIUMTEXT | **length+3** | 16Mb of text |
| MEDIUMBLOB | **length+3** | 16 Mb of data |
| LONGTEXT | **length+4** | 4GB of text |
| LONGBLOB | **length+4** | 4GB of data |
| ENUM | **1,2** | This field can contain one of a possible 65535 number of options. Ex: ENUM('abc','def','ghi') |
| SET | **1-8** | This type of field can contain any number of a set of predefined possible values. |

The following examples will make things quite clear on declaring Data Types within SQL statements.
**Steps in Creating the Database using MySQL.**

From the shell prompt (either in DOS or UNIX):
**mysqladmin create contacts;**
This will create an empty database called "contacts".
Now run the command line tool "mysql" and from the mysql prompt do the following:
**mysql> use contacts;**
(You'll get the response "Database changed")
The following commands entered into the MySQL prompt will create the tables in the database.
CREATE TABLE names (contact_id SMALLINT NOT NULL AUTO_INCREMENT PRIMARY KEY, FirstName CHAR(20), LastName CHAR(20), BirthDate DATE);
CREATE TABLE address(contact_id SMALLINT NOT NULL PRIMARY KEY, StreetAddress CHAR(50), City CHAR(20), State CHAR(20), Zip CHAR(15), Country CHAR(20));
CREATE TABLE telephones (contact_id SMALLINT NOT NULL PRIMARY KEY, TelephoneHome CHAR(20), TelephoneWork(20));
CREATE TABLE email (contact_id SMALLINT NOT NULL PRIMARY KEY, Email CHAR(20));

**CREATE TABLE company_details (contact_id SMALLINT NOT NULL PRIMARY KEY, CompanyName CHAR(25), Designation CHAR(15));**

**Note:** *Here we assume that one person will have only one email address. Now if there were a situation where one person has multiple addresses, this design would be a problem. We'd need another field which would keep values that indicated to whom the email address belonged to. In this particular case email data ownership is indicated by the primary key. The same is true for telephones. We are assuming that one person has only one home telephone and one work telephone number. This need not be true. Similarly, one person could work for multiple companies at the same time holding two different designation. In all these cases an extra field will solve the issue. For now, however let's work with this small design.*

**KEYS:** The relationships between columns located in different tables are usually described through the use of keys.
As you can see we have a PRIMARY KEY in each table. The Primary key serves as a mechanism to refer to other fields within the same row. In this case, the Primary key is used to identify a relationship between a row under consideration and the person whose name is located inside the 'names' table. We use the AUTO_INCREMENT statement only for the 'names' table as we need to use the generated contact_id number in all the other tables for identification of the rows.
This type of table design where one table establishes a relationship with several other tables is known as a **'one to many'** relationship.

In a **'many to many'** relationship we could have several Auto Incremented Primary Keys in various tables with several inter-relationships.
**Foreign Key:** A foreign key is a field in a table which is also the Primary Key in another table. This is known commonly as 'referential integrity'.

Execute the following commands to see the newly created tables and their contents.
To see the tables inside the database:

```
mysql> SHOW TABLES;
+-----------------------+
| Tables in contacts |
+-----------------------+
| address |
| company_details |
| email |
| names |
| telephones |
+----------------------+
5 rows in set (0.00 sec)
```

To see the columns within a particular table:

```
mysql>SHOW COLUMNS FROM address;
+---------------+-------------+------+-----+---------+-------+---------------------------------+
| Field | Type | Null | Key | Default | Extra | Privileges
|
+---------------+-------------+------+-----+---------+-------+---------------------------------+
| contact_id | smallint(6) | | PRI | 0 | | select,insert,update,references |
| StreetAddress | char(50) | YES | | NULL | | select,insert,update,references |
| City | char(20) | YES | | NULL | | select,insert,update,references |
| State | char(20) | YES | | NULL | | select,insert,update,references |
| Zip | char(10) | YES | | NULL | | select,insert,update,references |
| Country | char(20) | YES | | NULL | | select,insert,update,references |
+---------------+-------------+------+-----+---------+-------+----------------- ---------------+
6 rows in set (0.00 sec)
```

So, we have the tables created and ready. Now we put in some data.
Let's start with the 'names' table as it uses a unique AUTO_INCREMENT field which in turn is used in the other tables.

**Inserting data, one row at a time:**

```
mysql> INSERT INTO names (FirstName, LastName, BirthDate) VALUES ('Yamila','Diaz ','1974-10-13');
Query OK, 1 row affected (0.00 sec)
```

**Inserting multiple rows at a time:**

```
mysql> INSERT INTO names (FirstName, LastName, BirthDate) VALUES
('Nikki','Taylor','1972-03-04'),('Tia','Carrera','1975-09-18');
Query OK, 2 rows affected (0.00 sec)
Records: 2 Duplicates: 0 Warnings: 0
```

Let's see what the data looks like inside the table. We use the SELECT command for this.

```
mysql> SELECT * from NAMES;
+------------+-----------+----------+------------+
| contact_id | FirstName | LastName | BirthDate |
+------------+-----------+----------+------------+
| 3 | Tia | Carrera | 1975-09-18 |
| 2 | Nikki | Taylor | 1972-03-04 |
| 1 | Yamila | Diaz | 1974-10-13 |
+------------+-----------+----------+------------+
3 rows in set (0.06 sec)
```

**Try another handy command called 'DESCRIBE'.**

```
mysql> DESCRIBE names;
+------------+-------------+------+-----+---------+----------------+---------------------------------+
| Field | Type | Null | Key | Default | Extra | Privileges
|
+------------+-------------+------+-----+---------+----------------+---------------------------------+
| contact_id | smallint(6) | | PRI | NULL | auto_increment | select,insert,update,references |
| FirstName | char(20) | YES | | NULL | | select,insert,update,references |
```

```
| LastName | char(20) | YES | | NULL | | select,insert,update,references |
| BirthDate | date | YES | | NULL | | select,insert,update,references |
+------------+-------------+------+-----+---------+----------------+---------------------------------+
4 rows in set (0.00 sec)
Now lets populate the other tables. Observer the syntax used.
mysql> INSERT INTO address(contact_id, StreetAddress, City, State, Zip, Country) VALUES ('1', '300 Yamila Ave.', 'Los Angeles', 'CA', '300012', 'USA'),('2','4000 Nikki St.','Boca Raton','FL','500034','USA'),('3','404 Tia Blvd.','New York','NY','10011','USA');
Query OK, 3 rows affected (0.05 sec)
Records: 3 Duplicates: 0 Warnings: 0
mysql> SELECT * FROM address;
+------------+-----------------+-------------+-------+--------+---------+
| contact_id | StreetAddress | City | State | Zip | Country |
+------------+-----------------+-------------+-------+--------+---------+
| 1 | 300 Yamila Ave. | Los Angeles | CA | 300012 | USA |
| 2 | 4000 Nikki St. | Boca Raton | FL | 500034 | USA |
| 3 | 404 Tia Blvd. | New York | NY | 10011 | USA |
+------------+-----------------+-------------+-------+--------+---------+
3 rows in set (0.00 sec)
mysql> INSERT INTO company_details (contact_id, CompanyName, Designation) VALUES ('1','Xerox','New Business Manager'), ('2','Cabletron','Customer Support Eng'),('3','Apple','Sales Manager');
Query OK, 3 rows affected (0.05 sec)
Records: 3 Duplicates: 0 Warnings: 0
mysql> SELECT * FROM company_details;
+------------+-------------+----------------------+
| contact_id | CompanyName | Designation |
+------------+-------------+----------------------+
| 1 | Xerox | New Business Manager |
| 2 | Cabletron | Customer Support Eng |
| 3 | Apple | Sales Manager |
+------------+-------------+----------------------+
3 rows in set (0.06 sec)
mysql> INSERT INTO email (contact_id, Email) VALUES ('1', 'yamila@yamila.com'),( '2', 'nikki@nikki.com'),('3','tia@tia.com');
Query OK, 3 rows affected (0.00 sec)
Records: 3 Duplicates: 0 Warnings: 0
mysql> SELECT * FROM email;
+------------+-------------------+
| contact_id | Email |
+------------+-------------------+
| 1 | yamila@yamila.com |
| 2 | nikki@nikki.com |
| 3 | tia@tia.com |
+------------+-------------------+
3 rows in set (0.06 sec)
mysql> INSERT INTO telephones (contact_id, TelephoneHome, TelephoneWork) VALUES ('1','333-50000','333-60000'),('2','444-70000','444-80000'),('3','555-30000','55 5-40000');
Query OK, 3 rows affected (0.00 sec)
Records: 3 Duplicates: 0 Warnings: 0
mysql> SELECT * FROM telephones;
+------------+---------------+---------------+
| contact_id | TelephoneHome | TelephoneWork |
+------------+---------------+---------------+
| 1 | 333-50000 | 333-60000 |
| 2 | 444-70000 | 444-80000 |
| 3 | 555-30000 | 555-40000 |
+------------+---------------+---------------+
3 rows in set (0.00 sec)
```

Okay, so we now have all our data ready for experimentation.

**Before we start experimenting with manipulating the data let's look at how MySQL stores the Data.**

To do this execute the following command from the shell prompt.

```
mysqldump contacts > contacts.sql
```

**Note: The reverse operation for this command is:**

```
mysql contacts < contacts.sql
```

The file generated is a text file that contains all the data and SQL instruction needed to recreate the same database. As you can see, the SQL here is slightly different than what was typed in. Don't worry about this. It's all good ! It would also be obvious that this is a good way to backup your stuff.

```
# MySQL dump 8.2
#
# Host: localhost Database: contacts
#--------------------------------------------------------
# Server version 3.22.34-shareware-debug
#
# Table structure for table 'address'
#
CREATE TABLE address (
contact_id smallint(6) DEFAULT '0' NOT NULL,
StreetAddress char(50),
City char(20),
State char(20),
Zip char(10),
Country char(20),
PRIMARY KEY (contact_id)
);
#
# Dumping data for table 'address'
#
INSERT INTO address VALUES (1,'300 Yamila Ave.','Los Angeles','CA','300012','USA');
INSERT INTO address VALUES (2,'4000 Nikki St.','Boca Raton','FL','500034','USA');
INSERT INTO address VALUES (3,'404 Tia Blvd.','New York','NY','10011','USA');
#
# Table structure for table 'company_details'
#
CREATE TABLE company_details (
contact_id smallint(6) DEFAULT '0' NOT NULL,
CompanyName char(25),
Designation char(20),
PRIMARY KEY (contact_id)
);
#
# Dumping data for table 'company_details'
#
INSERT INTO company_details VALUES (1,'Xerox','New Business Manager');
INSERT INTO company_details VALUES (2,'Cabletron','Customer Support Eng');
INSERT INTO company_details VALUES (3,'Apple','Sales Manager');
#
# Table structure for table 'email'
#
CREATE TABLE email (
contact_id smallint(6) DEFAULT '0' NOT NULL,
Email char(20),
PRIMARY KEY (contact_id)
);
#
# Dumping data for table 'email'
#
```

```
INSERT INTO email VALUES (1,'yamila@yamila.com');
INSERT INTO email VALUES (2,'nikki@nikki.com');
INSERT INTO email VALUES (3,'tia@tia.com');
#
# Table structure for table 'names'
#
CREATE TABLE names (
contact_id smallint(6) DEFAULT '0' NOT NULL auto_increment,
FirstName char(20),
LastName char(20),
BirthDate date,
PRIMARY KEY (contact_id)
);
#
# Dumping data for table 'names'
#
INSERT INTO names VALUES (3,'Tia','Carrera','1975-09-18');
INSERT INTO names VALUES (2,'Nikki','Taylor','1972-03-04');
INSERT INTO names VALUES (1,'Yamila','Diaz','1974-10-13');
#
# Table structure for table 'telephones'
#
CREATE TABLE telephones (
contact_id smallint(6) DEFAULT '0' NOT NULL,
TelephoneHome char(20),
TelephoneWork char(20),
PRIMARY KEY (contact_id)
);
#
# Dumping data for table 'telephones'
#
INSERT INTO telephones VALUES (1,'333-50000','333-60000');
INSERT INTO telephones VALUES (2,'444-70000','444-80000');
INSERT INTO telephones VALUES (3,'555-30000','555-40000');
```

Let's try some SELECT statement variations:

To select all names whose corresponding contact_id is greater than 1.

```
mysql> SELECT * FROM names WHERE contact_id > 1;
+------------+-----------+----------+------------+
| contact_id | FirstName | LastName | BirthDate |
+------------+-----------+----------+------------+
| 3 | Tia | Carrera | 1975-09-18 |
| 2 | Nikki | Taylor | 1972-03-04 |
+------------+-----------+----------+------------+
2 rows in set (0.00 sec)
```

As a condition we can also use NOT NULL. This statement will return all names where there exists a contact_id.

```
mysql> SELECT * FROM names WHERE contact_id IS NOT NULL;
+------------+-----------+----------+------------+
| contact_id | FirstName | LastName | BirthDate |
+------------+-----------+----------+------------+
| 3 | Tia | Carrera | 1975-09-18 |
| 2 | Nikki | Taylor | 1972-03-04 |
| 1 | Yamila | Diaz | 1974-10-13 |
+------------+-----------+----------+------------+
3 rows in set (0.06 sec)
```

Results can be arranged in a particular way using the statement ORDER BY.

```
mysql> SELECT * FROM names WHERE contact_id IS NOT NULL ORDER BY LastName;
+------------+-----------+----------+------------+
```

```
| contact_id | FirstName | LastName | BirthDate |
+------------+-----------+----------+------------+
| 3 | Tia | Carrera | 1975-09-18 |
| 1 | Yamila | Diaz | 1974-10-13 |
| 2 | Nikki | Taylor | 1972-03-04 |
+------------+-----------+----------+------------+
3 rows in set (0.06 sec)
```

'asc' and 'desc' stand for ascending and descending respectively and can be used to arrange the results.

```
mysql> SELECT * FROM names WHERE contact_id IS NOT NULL ORDER BY LastName desc;
+------------+-----------+----------+------------+
| contact_id | FirstName | LastName | BirthDate |
+------------+-----------+----------+------------+
| 2 | Nikki | Taylor | 1972-03-04 |
| 1 | Yamila | Diaz | 1974-10-13 |
| 3 | Tia | Carrera | 1975-09-18 |
+------------+-----------+----------+------------+
3 rows in set (0.04 sec)
```

You can also place date types into conditional statements.

```
mysql> SELECT * FROM names WHERE BirthDate > '1973-03-06';
+------------+-----------+----------+------------+
| contact_id | FirstName | LastName | BirthDate |
+------------+-----------+----------+------------+
| 3 | Tia | Carrera | 1975-09-18 |
| 1 | Yamila | Diaz | 1974-10-13 |
+------------+-----------+----------+------------+
2 rows in set (0.00 sec)
```

LIKE is a statement to match field values using wildcards. The % sign is used for denoting wildcards and can represent multiple characters.

```
mysql> SELECT FirstName, LastName FROM names WHERE LastName LIKE 'C%';
+-----------+----------+
| FirstName | LastName |
+-----------+----------+
| Tia | Carrera |
+-----------+----------+
1 row in set (0.06 sec)
```

'_' is used to represent a single wildcard.

```
mysql> SELECT FirstName, LastName FROM names WHERE LastName LIKE '_iaz';
+-----------+----------+
| FirstName | LastName |
+-----------+----------+
| Yamila | Diaz |
+-----------+----------+
1 row in set (0.00 sec)
```

**SQL Logical Operations (operates from Left to Right)**

1. NOT or !
2. AND or &&
3. OR or ||
4. = : Equal
5. <> or != : Not Equal
6. <=
7. >=
8. <,>

Here are some more variations with Logical Operators and using the 'IN' statement.

```
mysql> SELECT FirstName FROM names WHERE contact_id < 3 AND LastName LIKE 'D%';
+-----------+
| FirstName |
+-----------+
| Yamila |
```

```
+-----------+
1 row in set (0.00 sec)
mysql> SELECT contact_id FROM names WHERE LastName IN ('Diaz','Carrera');
+------------+
| contact_id |
+------------+
| 3 |
| 1 |
+------------+
2 rows in set (0.02 sec)
```

To return the number of rows in a table

```
mysql> SELECT count(*) FROM names;
+----------+
| count(*) |
+----------+
| 3 |
+----------+
1 row in set (0.02 sec)
mysql> SELECT count(FirstName) FROM names;
+------------------+
| count(FirstName) |
+------------------+
| 3 |
+------------------+
1 row in set (0.00 sec)
```

To do some basic arithmetic aggregate functions.

```
mysql> SELECT SUM(contact_id) FROM names;
+-----------------+
| SUM(contact_id) |
+-----------------+
| 6 |
+-----------------+
1 row in set (0.00 sec)
```

To select a largest value from a row. Substitute 'MIN' and see what happens next.

```
mysql> SELECT MAX(contact_id) FROM names;
+-----------------+
| MAX(contact_id) |
+-----------------+
| 3 |
+-----------------+
1 row in set (0.00 sec)
```

**HAVING**

Take a look at the first query using the statement WHERE and the second statement using the statement HAVING.

```
mysql> SELECT * FROM names WHERE contact_id >=1;
+------------+-----------+----------+------------+
| contact_id | FirstName | LastName | BirthDate |
+------------+-----------+----------+------------+
| 1 | Yamila | Diaz | 1974-10-13 |
| 2 | Nikki | Taylor | 1972-03-04 |
| 3 | Tia | Carrera | 1975-09-18 |
+------------+-----------+----------+------------+
3 rows in set (0.03 sec)
mysql> SELECT * FROM names HAVING contact_id >=1;
+------------+-----------+----------+------------+
| contact_id | FirstName | LastName | BirthDate |
+------------+-----------+----------+------------+
| 3 | Tia | Carrera | 1975-09-18 |
```

```
| 2 | Nikki | Taylor | 1972-03-04 |
| 1 | Yamila | Diaz | 1974-10-13 |
+------------+-----------+----------+------------+
3 rows in set (0.00 sec)
```

Now lets work with multiple tables and see how information can be pulled out of the data.

```
mysql> SELECT names.contact_id, FirstName, LastName, Email FROM names, email WHERE
names.contact_id = email.contact_id;
+------------+-----------+----------+------------------+
| contact_id | FirstName | LastName | Email |
+------------+-----------+----------+------------------+
| 1 | Yamila | Diaz | yamila@yamila.com |
| 2 | Nikki | Taylor | nikki@nikki.com |
| 3 | Tia | Carrera | tia@tia.com |
+------------+-----------+----------+------------------+
3 rows in set (0.11 sec)
mysql> SELECT DISTINCT names.contact_id, FirstName, Email, TelephoneWork FROM names, email,
telephones WHERE names.contact_id=email.contact_id=telephones.contact_id;
+------------+-----------+------------------+--------------+
| contact_id | FirstName | Email | TelephoneWork |
+------------+-----------+------------------+--------------+
| 1 | Yamila | yamila@yamila.com | 333-60000 |
| 2 | Nikki | nikki@nikki.com | 333-60000 |
| 3 | Tia | tia@tia.com | 333-60000 |
+------------+-----------+------------------+--------------+
3 rows in set (0.05 sec)
```

**So what's a JOIN?**

JOIN is the action performed on multiple tables that returns a result as a table. It's what makes a database 'relational'.

There are several types of joins. Let's look at LEFT JOIN (OUTER JOIN) and RIGHT JOIN

Let's first check out the contents of the tables we're going to use

```
mysql> SELECT * FROM names;
+------------+-----------+----------+------------+
| contact_id | FirstName | LastName | BirthDate |
+------------+-----------+----------+------------+
| 3 | Tia | Carrera | 1975-09-18 |
| 2 | Nikki | Taylor | 1972-03-04 |
| 1 | Yamila | Diaz | 1974-10-13 |
+------------+-----------+----------+------------+
3 rows in set (0.00 sec)
mysql> SELECT * FROM email;
+------------+------------------+
| contact_id | Email |
+------------+------------------+
| 1 | yamila@yamila.com |
| 2 | nikki@nikki.com |
| 3 | tia@tia.com |
+------------+------------------+
3 rows in set (0.00 sec)
```

**A LEFT JOIN First:**

```
mysql> SELECT * FROM names LEFT JOIN email USING (contact_id);
+------------+-----------+----------+------------+------------+-----------------+
| contact_id | FirstName | LastName | BirthDate | contact_id | Email|
+------------+-----------+----------+------------+------------+------------------+
| 3 | Tia | Carrera | 1975-09-18 | 3 | tia@tia.com |
| 2 | Nikki | Taylor | 1972-03-04 | 2 | nikki@nikki.com |
| 1 | Yamila | Diaz | 1974-10-13 | 1 | yamila@yamila.com |
```

```
+------------+-----------+----------+------------+------------+-------------------+
3 rows in set (0.16 sec)
```

To find the people who have a home phone number.

```
mysql> SELECT names.FirstName FROM names LEFT JOIN telephones ON names.contact_id =
telephones.contact_id WHERE TelephoneHome IS NOT NULL;
+-----------+
| FirstName |
+-----------+
| Tia |
| Nikki |
| Yamila |
+-----------+
3 rows in set (0.02 sec)
```

These same query leaving out 'names' (from names.FirstName) is still the same and will generate the same result.

```
mysql> SELECT FirstName FROM names LEFT JOIN telephones ON names.contact_id =
telephones.contact_id WHERE TelephoneHome IS NOT NULL;
+-----------+
| FirstName |
+-----------+
| Tia |
| Nikki |
| Yamila |
+-----------+
3 rows in set (0.00 sec)
```

**And now a RIGHT JOIN:**

```
mysql> SELECT * FROM names RIGHT JOIN email USING(contact_id);
+------------+-----------+----------+------------+------------+----------------- --+
| contact_id | FirstName | LastName | BirthDate | contact_id | Email |
+------------+-----------+----------+------------+------------+-------------------+
| 1 | Yamila | Diaz | 1974-10-13 | 1 | yamila@yamila.com |
| 2 | Nikki | Taylor | 1972-03-04 | 2 | nikki@nikki.com
|
| 3 | Tia | Carrera | 1975-09-18 | 3 | tia@tia.com
|
+------------+-----------+----------+------------+------------+-----------------
--+
3 rows in set (0.03 sec)
```

**BETWEEN**

This conditional statement is used to select data where a certain related contraint falls between a certain range of values. The following example illustrates it's use.

```
mysql> SELECT * FROM names;
+------------+-----------+----------+------------+
| contact_id | FirstName | LastName | BirthDate |
+------------+-----------+----------+------------+
| 3 | Tia | Carrera | 1975-09-18 |
| 2 | Nikki | Taylor | 1972-03-04 |
| 1 | Yamila | Diaz | 1974-10-13 |
+------------+-----------+----------+------------+
3 rows in set (0.06 sec)
mysql> SELECT FirstName, LastName FROM names WHERE contact_id BETWEEN 2 AND 3;
+-----------+----------+
| FirstName | LastName |
+-----------+----------+
| Tia | Carrera |
```

```
| Nikki | Taylor |
+-----------+----------+
2 rows in set (0.00 sec)
```

**ALTER**

The ALTER statement is used to add a new column to an existing table or to make changes to it.

```
mysql> ALTER TABLE names ADD Age SMALLINT;
Query OK, 3 rows affected (0.11 sec)
Records: 3 Duplicates: 0 Warnings: 0
```

Now let's take a look at the 'ALTER'ed Table.

```
mysql> SHOW COLUMNS FROM names;
+------------+-------------+------+-----+---------+----------------+
| Field | Type | Null | Key | Default | Extra |
+------------+-------------+------+-----+---------+----------------+
| contact_id | smallint(6) | | PRI | 0 | auto_increment |
| FirstName | char(20) | YES | | NULL | |
| LastName | char(20) | YES | | NULL | |
| BirthDate | date | YES | | NULL | |
| Age | smallint(6) | YES | | NULL | |
+------------+-------------+------+-----+---------+----------------+
5 rows in set (0.06 sec)
```

But we don't require Age to be a SMALLINT type when a TINYINT would suffice. So we use another ALTER statement.

```
mysql> ALTER TABLE names CHANGE COLUMN Age Age TINYINT;
Query OK, 3 rows affected (0.02 sec)
Records: 3 Duplicates: 0 Warnings: 0
mysql> SHOW COLUMNS FROM names;
+------------+-------------+------+-----+---------+--------+----------------+
| Field | Type | Null | Key | Default | Extra |
+------------+-------------+------+-----+---------+--------+----------------+
| contact_id | smallint(6) | | PRI | NULL |
| FirstName | char(20) | YES | | NULL | |
| LastName | char(20) | YES | | NULL | |
| BirthDate | date | YES | | NULL | |
| Age | tinyint(4) | YES | | NULL | |
+------------+-------------+------+-----+---------+--------+----------------+
5 rows in set (0.00 sec)
```

**MODIFY**

You can also use the MODIFY statement to change column data types.

```
mysql> ALTER TABLE names MODIFY COLUMN Age SMALLINT;
Query OK, 3 rows affected (0.03 sec)
Records: 3 Duplicates: 0 Warnings: 0
mysql> SHOW COLUMNS FROM names;
+------------+-------------+------+-----+---------+----------------+---------------+
| Field | Type | Null | Key | Default | Extra |
+------------+-------------+------+-----+---------+----------------+---------------+
| contact_id | smallint(6) | | PRI | NULL | auto_increment |
| FirstName | char(20) | YES | | NULL | |
| LastName | char(20) | YES | | NULL | |
| BirthDate | date | YES | | NULL | |
| Age | smallint(6) | YES | | NULL | |
+------------+-------------+------+-----+---------+----------------+---------------+
5 rows in set (0.00 sec)
```

**To Rename a Table:**

```
mysql> ALTER TABLE names RENAME AS mynames;
Query OK, 0 rows affected (0.00 sec)
mysql> SHOW TABLES;
+--------------------+
```

```
| Tables_in_contacts |
+--------------------+
| address |
| company_details |
| email |
| mynames |
| telephones |
+--------------------+
5 rows in set (0.00 sec)
```

We rename it back to the original name.

```
mysql> ALTER TABLE mynames RENAME AS names;
Query OK, 0 rows affected (0.01 sec)
```

**UPDATE**

The UPDATE command is used to add a value to a field in a table.

```
mysql> UPDATE names SET Age ='23' WHERE FirstName='Tia';
Query OK, 1 row affected (0.06 sec)
Rows matched: 1 Changed: 1 Warnings: 0
```

The Original Table:

```
mysql> SELECT * FROM names;
+------------+-----------+----------+------------+------+
| contact_id | FirstName | LastName | BirthDate | Age |
+------------+-----------+----------+------------+------+
| 3 | Tia | Carrera | 1975-09-18 | 23 |
| 2 | Nikki | Taylor | 1972-03-04 | NULL |
| 1 | Yamila | Diaz | 1974-10-13 | NULL |
+------------+-----------+----------+------------+------+
3 rows in set (0.05 sec)
```

The Modified Table:

```
mysql> SELECT * FROM names;
+------------+-----------+----------+------------+------+
| contact_id | FirstName | LastName | BirthDate | Age |
+------------+-----------+----------+------------+------+
| 3 | Tia | Carrera | 1975-09-18 | 24 |
| 2 | Nikki | Taylor | 1972-03-04 | NULL |
| 1 | Yamila | Diaz | 1974-10-13 | NULL |
+------------+-----------+----------+------------+------+
3 rows in set (0.00 sec)
```

**DELETE**

```
mysql> DELETE FROM names WHERE Age=23;
Query OK, 1 row affected (0.00 sec)
mysql> SELECT * FROM names;
+------------+-----------+----------+------------+------+
| contact_id | FirstName | LastName | BirthDate | Age |
+------------+-----------+----------+------------+------+
| 2 | Nikki | Taylor | 1972-03-04 | NULL |
| 1 | Yamila | Diaz | 1974-10-13 | NULL |
+------------+-----------+----------+------------+------+
2 rows in set (0.00 sec)
```

**A DEADLY MISTAKE**

```
mysql> DELETE FROM names;
Query OK, 0 rows affected (0.00 sec)
mysql> SELECT * FROM names;
Empty set (0.00 sec)
```

One more destructive tool...

**DROP TABLE**

```
mysql> DROP TABLE names;
Query OK, 0 rows affected (0.00 sec)
mysql> SHOW TABLES;
+--------------------+
| Tables in contacts |
+--------------------+
| address |
| company_details |
| email |
| telephones |
+--------------------+
4 rows in set (0.05 sec)
mysql> DROP TABLE address ,company_details, telephones;
Query OK, 0 rows affected (0.06 sec)
mysql> SHOW TABLES;
Empty set (0.00 sec)
```

As you can see, the table 'names' no longer exists. MySQL does not give a warning so be careful.

**Full Text Indexing and Searching**
Since version 3.23.23, Full Text Indexing and Searching has been introduced into MySQL. FULLTEXT indexes can be created from VARCHAR and TEXT columns. FULLTEXT searches are performed with the MATCH function. The MATCH function matches a natural language query on a text collection and from each row in a table it returns relevance. The resultant rows are organized in order of relevance.
Full Text searches are a very powerful way to search through text. But is not ideal for small tables of text and may produce inconsistent results. Ideally it works with large collections of textual data.

**Optimizing your Database:** Well, databases do tend to get large at some or the other. And here arises the issue of database optimization. Queries are going to take longer and longer as the database grows, and certain things can be done to speed things up.

**Clustering:** The easiest method is that of 'clustering'. Suppose you do a certain kind of query often, it would be faster if the database contents were arranged in the same way data was requested. To keep the tables in a sorted order you need a clustering index. Some databases keep stuff sorted automatically.

**Ordered Indices:** These are a kind of 'lookup' tables of sorts. For each column that may be of interest to you, you can create an ordered index. It needs to be noted that again these kinds of optimization techniques produce a system load in terms of creating a new index each time the data is re-arranged.
There are additional methods such as B-Trees, hashing which you may like to read up about but will not be discussed here.

**Replication:** Replication is the term given to the process where databases synchronize with each other. In this process one database updates its own data with respect to another or with reference to certain criteria for updates specified by the programmer. Replication can be used under various circumstances. Examples may be: safety and backup, to provide a closer location to the database for certain users.

**What are Transactions?**
In an RDBMS, when several people access the same data or if a server dies in the middle of an update, there must be a mechanism to protect the integrity of the data. Such a mechanism is called a Transaction. A transaction groups a set of database actions into a single instantaneous event. This event can either succeed or fail. i.e. either get the job done or fail.
The definition of a transaction can be provided by an Acronym called '**ACID**'.

**(A) Atomicity:** If an action consists of multiple steps - it's still considered as one operation.

**(C) Consistency:** The database exists in a valid and accurate operating state before and after a transaction.

**(I) Isolation:** Processes within one transaction are independent and cannot interfere with that in others.

**(D) Durability:** Changes affected by a transaction are permanent.

To enable transactions a mechanism called 'Logging' needs to be introduced. Logging involves a DBMS writing details on the tables, columns and results of a particular transaction, both before and after, onto a log file. This log file is used in the process of recovery. Now to protect a certain database resource (ex. a table) from being used and written onto simultaneously several techniques are used. One of them is 'Locking' another is to put a 'time stamp' onto an action. In the case of Locking, to complete an action, the DBMS would need to acquire locks on all resources needed to complete the action. The locks are released only when the transaction is completed.

Now if there were say a large number of tables involved in a particular action, say 50, all 50 tables would be locked till a transaction is completed. To improve things a bit, there is another technique used called 2 Phase Locking or 2PL. In this method of locking, locks are acquired only when needed but are released only when the transaction is completed. This is done to make sure that that altered data can be safely restored if the transaction fails for any reason. This technique can also result in problems such as "**deadlocks".**
In this case - processes requiring the same resources lock each other up by preventing the other to complete an action. Options here are to abort one, or let the programmer handle it. MySQL implements transactions by implementing the Berkeley DB libraries into its own code. So it's the source version you'd want here for MySQL installation. Read the MySQL manual on implementing this.

## 7. Beyond MySQL

**What are Views?**
A view allows you to assign the result of a query to a new private table. This table is given the name used in your VIEW query. Although MySQL does not support views yet a sample SQL VIEW construct statement would look like:

**CREATE VIEW TESTVIEW AS SELECT * FROM names;**
**What are Triggers?**
A trigger is a pre-programmed notification that performs a set of actions that may be commonly required. Triggers can be programmed to execute certain actions before or after an event occurs. Triggers are very useful as they they increase efficiency and accuracy in performing operations on databases and also are increase productivity by reducing the time for application development. Triggers however do carry a price in terms of processing overhead.

**What are Procedures?**
Like triggers, Procedures or 'Stored' Procedures are productivity enhancers. Suppose you needed to perform an action using a programming interface to the database in say PERL and ASP. If a programmed action could be stored at the database level, it's obvious that it has to be written only once and can be called by any programming language interacting with the database. Procedures are executed using triggers.

**<u>Beyond RDBMS</u>**
**Distributed Databases (DDB)**
A distributed database is a collection of several, logically interrelated database located at multiple locations of a computer network. A distributed database management system permits the management of such a database and makes the operation transparent to the user. Good examples of distributed databases would be those utilized by banks, multinational firms with several office locations where each distributed data system works only with the data that is relevant to it's operations. DDBs have full functionality of any DBMS. It's also important to know that the distributed databases are considered to be actually one database rather than discrete files and data within distributed databases are logically interrelated.

**Object Database Management Systems or ODBMS**
When the capabilities of a database are integrated with object programming language capabilities, the resulting product is an ODBMS. Database objects appear as programming objects in an **ODBMS.** Using an ODBMS offers several advantages. The ones that can be most readily appreciated are:
**1. Efficiency:** When you use an ODBMS, you're using data the way you store it. You will use less code as you're not dependent on an intermediary like SQL or ODBC. When this happens you can create highly complex data structures through your programming language.
**2. Speed:** When data is stored the way you'd like it to be stored (i.e. natively) there is a massive performance increase as no to-and-fro translation is required.

**A Quick Tutorial on Database Normalization**

Let's start off by taking some data represented in a Table.

**Table Name: College Table**

| StudentName | CourseID1 | CourseTitle1 | CourseProfessor1 | CourseID2 | CourseTitle2 | CourseProfessor2 | StudentAdvisor | StudentID |
|---|---|---|---|---|---|---|---|---|
| Tia Carrera | CS123 | Perl Regular Expressions | Don Corleone | CS003 | Object Oriented Programming 1 | Daffy Duck | Fred Flintstone | 400 |
| John Wayne | CS456 | Socket Programming | DJ Tiesto | CS004 | Algorithms | Homer Simpson | Barney Rubble | 401 |
| Lara Croft | CS789 | OpenGL | Bill Clinton | CS001 | Data Structures | Papa Smurf | Seven of Nine | 402 |

**(text size has been shrunk to aid printability on one page)**

**The First Normal Form: (Each Column Type is Unique and there are no repeating groups [types] of data)**

This essentially means that you indentify data that can exist as a separate table and therefore reduce repetition and will reduce the width of the original table.

We can see that for every student, Course Information is repeated for each course. So if a student has three course, you'll need to add another set of columns for Course Title, Course Professor and CourseID. So Student information and Course Information can be considered to be two broad groups.

| Table Name: Student Information |
|---|
| StudentID (Primary Key)<br>StudentName<br>AdvisorName |

| Table Name: Course Information |
|---|
| CourseID (Primary Key)<br>CourseTitle<br>CourseDescription<br>CourseProfessor |

It's obvious that we have here a Many to Many relationship between Students and Courses.

**Note:** In a Many to Many relationship we need something called a relating table which basically contains information exclusively on which relatioships exist between two tables. In a One to Many relationship we use a foreign key.

So in this case we need another little table called: **Students and Courses**

| Table Name: Students and Courses |
|---|
| SnCStudentID<br>SnCCourseID |

**The Second Normal Form: (All attributes within the entity should depend solely on the entity's unique identifier)**

The AdvisorName under Student Information does not depend on the StudentID. Therefore it can be moved to it's own table.

| Table Name: Student Information |
|---|
| StudentID (Primary Key)<br>StudentName |

| Table Name: Advisor Information |
|---|
| AdvisorID<br>AdvisorName |

| Table Name: Course Information |
|---|
| CourseID (Primary Key)<br>CourseTitle<br>CourseDescription<br>CourseProfessor |

| Table Name: Students and Courses |
|---|
| SnCStudentID<br>SnCCourseID |

**Note: Relating Tables can be created as required.**
**The Third Normal Form:** (no column entry should be dependent on any other entry (value) other than the key for the table)
In simple terms - a table should contain information about only one thing.
In Course Information, we can pull CourseProfessor information out and store it in another table.

| Table Name: Student Information |
|---|
| StudentID (Primary Key)<br>StudentName |

| Table Name: Advisor Information |
|---|
| AdvisorID<br>AdvisorName |

| Table Name: Course Information |
|---|
| CourseID (Primary Key)<br>CourseTitle<br>CourseDescription |

| Table Name: Professor Information |
|---|
| ProfessorID<br>CourseProfessor |

| Table Name: Students and Courses |
|---|
| SnCStudentID<br>SnCCourseID |

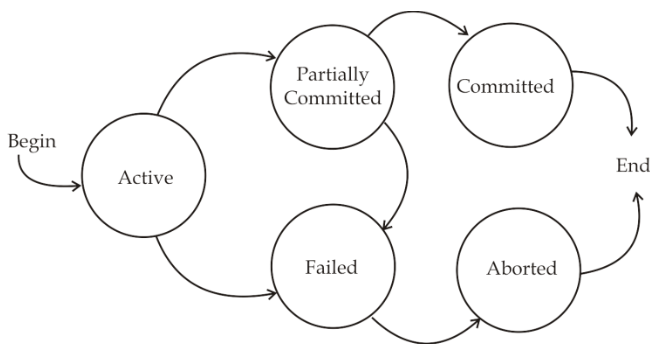
**Note: Relating Tables can be created as required.**
Well that's it. Ones you are done with 3NF the database is considered Normalized.
Now let's consider some cases where normalization would have to avoided for practical purposes. Suppose we needed to store a student's home address along with State and Zip Code information. Would you create a separate table for every zip code in your country along with one for cities and one for states? It actually depends on you. I would prefer just using a non-normalized address table and stick everything in there. So exceptions crop up often and it's up to your better judgement

## 8. Transaction Control

Transaction: A collection of actions that transforms the DB from one consistent state into another consistent state; during the execution, the DB might be inconsistent. A transaction can be defined as a logical unit of work on the database. This may be an entire program, a piece of a program or a single command (like the SQL commands such as INSERT or UPDATE) and it may engage in any number of operations on the database.

*Consider an example of steps involved in a simple transaction of deducting 100 Rupees from Amit's Account.*
Open_Acc (Amit)
OldBal = Amit.bal
NewBal = OldBal - 5000
Ram.bal = NewBal
CloseAccount(Amit)

## 8.1 States of Transaction




- **Active:** Initial state and during the execution
- **Partially committed:** After the final statement has been executed
- **Committed:** After successful completion
- **Failed:** After the discovery that normal execution can no longer proceed
- **Aborted:** After the transaction has been rolled back and the DB restored to its state prior to the start of the transaction. Restart it again or kill it.

A transaction enters the failed state after the system determines that the transaction can no longer proceed with its normal execution (for example, because of hardware or logical errors). Such a transaction must be rolled back. Then, it enters the aborted state. At this point, the system has two options:

- It can restart the transaction, but only if the transaction was aborted as a result of some hardware or software error that was not created through the internal logic of the transaction. A restarted transaction is considered to be a new transaction.
- It can kill the transaction. It usually does so because of some internal logical error that can be corrected only by rewriting the application program, or because the input was bad, or because the desired data were not found in the database.

## 8.2 Concurrency Control

**Problems in Concurrency Control**

*Lost update*

• Occurs in two concurrent transactions when:
• Same data element is updated
• One of the updates is lost

*Uncommitted data*

• Occurs when:
• Two transactions are executed concurrently
• First transaction is rolled back after the second transaction has already accessed uncommitted data

*Inconsistent retrievals*

• Occurs when a transaction accesses data before and after one or more other transactions finish working with such data

## 8.3 Schedules

Schedule – a sequences of instructions that specify the chronological order in which instructions of concurrent transactions are executed. A schedule for a set of transactions must consist of all instructions of those transactions. Must preserve the order in which the instructions appear in each individual transaction.

A transaction that successfully completes its execution will have a commit instruction as the last statement (will be omitted if it is obvious). A transaction that fails to successfully complete its
execution will have an abort instruction as the last statement (will be omitted if it is obvious).

The execution sequences just described are called schedules. They represent the chronological order in which instructions are executed in the system. Clearly, a schedule for a set of transactions must consist of all instructions of those transactions, and must preserve the order in which the instructions appear in each individual transaction. Following is an example of a serial schedule:

| $T_1$ | $T_2$ |
|---|---|
| read($A$) | |
| $A := A - 50$ | |
| write($A$) | |
| read($B$) | |
| $B := B + 50$ | |
| write($B$) | |
| commit | |
| | read($A$) |
| | $temp := A * 0.1$ |
| | $A := A - temp$ |
| | write($A$) |
| | read($B$) |
| | $B := B + temp$ |
| | write($B$) |
| | commit |

## 8.4 Serializability

If a schedule S can be transformed into a schedule S´ by a series of swaps of non-conflicting instructions, we say that S and S´ are conflict equivalent. We say that a schedule S is serializable if it is conflict equivalent to a serial schedule.

Schedule S1 can be transformed into S2, a serial schedule where T2 follows T1, by series of swaps of non-conflicting instructions
Therefore, Schedule S1 is serializable

**Schedule S1:**

| T1 | T2 |
|---|---|
| read (A)<br>write (A) | |
| | read (A)<br>write (A) |
| read (B)<br>write (B) | |
| | read (B)<br>write (B) |

**Schedule S2:**

| T1 | T2 |
|---|---|
| read (A)<br>write (A)<br>read (B)<br>write (B) | |
| | read (A)<br>write (A)<br>read (B)<br>write (B) |

Schedule S3 is not serializable: We are unable to swap instructions in the schedule to obtain either the serial schedule <T3, T4>, or the serial schedule <T4,T3>

Schedule S3:

| T3 | T4 |
|---|---|
| read (P) | |
| | write (P) |
| write (P) | |

## 8.5 Recoverability

Recoverable schedule: For each pair of transactions Ti and Tj, where Tj reads data items written by Ti, Ti must commit before Tj commits.

The following schedule is not recoverable, if T6 commits immediately after the read

| T5 | T6 |
|---|---|
| read (A)<br>write (A) | |
| | read (A) |
| read (B) | |

If T5 should abort, T6 would have read (and possibly shown to the user) an inconsistent database state. Hence, database must ensure that schedules are recoverable.

## 8.6 Cascadeless schedules

| T7 | T8 | T9 |
|---|---|---|
| read (A)<br>read (B)<br>write (A) | | |
| | read (A)<br>write (A) | |
| | | read (A) |

Transaction T7 writes a value of A that is read by Transaction T8. Transaction T8 writes a value of A that is read by Transaction T9. Suppose at this point T7 fails. T7 must be rolled back, since T8 is dependent on T7, T8 must be rolled back, T9 is dependent on T8, T9 must be rolled back.

**This phenomenon, in which a single transaction failure leads to a series of transaction rollbacks is called Cascading rollback.**

- Cascading rollback is undesirable, since it leads to the undoing of a significant amount of work.
- It is desirable to restrict the schedules to those where cascading rollbacks cannot occur, Such schedules are called Cascadeless Schedules.
- Formally, a cascadeless schedule is one where for each pair of transaction Ti and Tj such that Tj reads data item, previously written by Ti the commit operation of Ti appears before the read operation of Tj .
- Every Cascadeless schedule is also recoverable schedule.

**Cascadeless Schedule**

| T10 | T11 |
|---|---|
| read (A)<br>write (A) | |
| | read (B) |
| commit | |
| | read (A) |

## 8.7 Implementation of Isolation Levels

The goal of concurrency-control policies is to provide a high degree of concurrency, while ensuring that all schedules that can be generated are conflict or view serializable, recoverable, and cascadeless.

**Locking**

Instead of locking the entire database, a transaction could, instead, lock only those data items that it accesses. Under such a policy, the transaction must hold locks long enough to ensure serializability, but for a period short enough not to harm performance excessively.

Further improvements to locking result if we have two kinds of locks: shared and exclusive. Shared locks are used for data that the transaction reads and exclusive locks are used for those it writes. Many transactions can hold shared locks on the same data item at the same time, but a transaction is allowed an exclusive lock on a data item only if no other transaction holds any lock (regardless
of whether shared or exclusive) on the data item. This use of two modes of locks along with two-phase locking allows concurrent reading of data while still ensuring serializability.

**Timestamps**
Another category of techniques for the implementation of isolation assigns each transaction a timestamp, typically when it begins. For each data item, the system keeps two timestamps. The read timestamp of a data item holds the largest (that is, the most recent) timestamp of those transactions that read the data item. The write timestamp of a data item holds the timestamp of the transaction that Transactions wrote the current value of the data item.

**Snapshot Isolation**
By maintaining more than one version of a data item, it is possible to allow a transaction to read an old version of a data item rather than a newer version written by an uncommitted transaction or by a transaction that should come later in the serialization order. There are a variety of multi-version concurrency control techniques. One in particular, called snapshot isolation, is widely used in practice.

Snapshot isolation ensures that attempts to read data never need to wait (unlike locking). Read-only transactions cannot be aborted; only those that modify data run a slight risk of aborting. Since each transaction reads its own version or snapshot of the database, reading data does not cause subsequent update attempts by other transactions to wait (unlike locking). Since most transactions are read-only (and most others read more data than they update), this is often a major source of performance improvement as compared to locking.

## 8.9 Lock Based Protocols

A lock is a mechanism to control concurrent access to a data item! Data items can be locked in two modes:
**1. exclusive (X) mode.** Data item can be both read as well as written. X-lock is requested using lock-X instruction.
**2. shared (S) mode.** Data item can only be read. S-lock is requested using lock-S instruction.
Lock requests are made to concurrency-control manager. Transaction can proceed only after request is granted.

A transaction may be granted a lock on an item if the requested lock is compatible with locks already held on the item by other transactions. Any number of transactions can hold shared locks on an item, but if any transaction holds an exclusive on the item no other transaction may hold any lock on the item. If a lock cannot be granted, the requesting transaction is made to wait till all incompatible locks held by other transactions have been released. The lock is then granted.

**Drawbacks**
The potential for deadlock exists in most locking protocols. Deadlocks are a necessary evil.
Starvation is also possible if concurrency control manager is badly designed.
For example:

- A transaction may be waiting for an X-lock on an item, while a sequence of other transactions request and are granted an S-lock on the same item.
- The same transaction is repeatedly rolled back due to deadlocks.

Concurrency control manager can be designed to prevent starvation.

**The Two-Phase Locking Protocol**

One protocol that ensures serializability is the two-phase locking protocol. This protocol requires that each transaction issue lock and unlock requests in two phases:
1. Growing phase. A transaction may obtain locks, but may not release any lock.
2. Shrinking phase. A transaction may release locks, but may not obtain any new locks.

Initially, a transaction is in the growing phase. The transaction acquires locks as needed. Once the transaction releases a lock, it enters the shrinking phase, and it can issue no more lock requests.

***Two-phase locking does not ensure freedom from deadlocks***

- Cascading roll-back is possible under two-phase locking. To avoid this, follow a modified protocol called strict two-phase locking. Here a transaction must hold all its exclusive locks till it commits/aborts.
- Rigorous two-phase locking is even stricter: here all locks are held till commit/abort. In this protocol transactions can be serialized in the order in which they commit.

## 9. Deadlock Handling

In a database, a deadlock is a situation in which two or more transactions are waiting for one another to give up locks. For example, Transaction A might hold a lock on some rows in the Accounts table and needs to update some rows in the Orders table to finish. Transaction B holds locks on those very rows in the Orders table but needs to update the rows in the Accounts table held by Transaction A. Transaction A cannot complete its transaction because of the lock on Orders. Transaction B cannot complete its transaction because of the lock on Accounts. All activity comes to a halt and remains at a standstill forever unless the DBMS detects the deadlock and aborts one of the transactions.

**Deadlock Prevention**

To prevent any deadlock situation in the system, the DBMS aggressively inspects all the operations which transactions are about to execute. DBMS inspects operations and analyze if they can create a deadlock situation. If it finds that a deadlock situation might occur then that transaction is never allowed to be executed.
There are deadlock prevention schemes, which uses time-stamp ordering mechanism of transactions in order to predict a deadlock situation.

***WAIT-DIE SCHEME:***

In this scheme, if a transaction request to lock a resource (data item), which is already held with conflicting lock by some other transaction, one of the two possibilities may occur:

- If TS(Ti) < TS(Tj), that is Ti, which is requesting a conflicting lock, is older than Tj, Ti is allowed to wait until the dataitem is available.
- If TS(Ti) > TS(tj), that is Ti is younger than Tj, Ti dies. Ti is restarted later with random delay but with same timestamp.
- This scheme allows the older transaction to wait but kills the younger one.

***WOUND-WAIT SCHEME:***

In this scheme, if a transaction request to lock a resource (data item), which is already held with conflicting lock by some other transaction, one of the two possibilities may occur:

- If TS(Ti) < TS(Tj), that is Ti, which is requesting a conflicting lock, is older than Tj, Ti forces Tj to be rolled back, that is Ti wounds Tj. Tj is restarted later with random delay but with same timestamp.
- If TS(Ti) > TS(Tj), that is Ti is younger than Tj, Ti is forced to wait until the resource is available.

This scheme, allows the younger transaction to wait but when an older transaction request an item held by younger one, the older transaction forces the younger one to abort and release the item.
In both cases, transaction, which enters late in the system, is aborted.

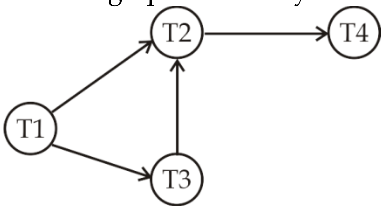
**Deadlock Detection**

Deadlocks can be described precisely in terms of a directed graph called a wait for graph. This graph consists of a pair G = (V, E), where V is a set of vertices and E is a set of edges. The set of vertices consists of all the transactions in the system. Each element in the set E of edges is an ordered pair Ti → Tj. If Ti → Tj is in E, then there is a directed edge from transaction Ti to Tj , implying that transaction Ti is waiting for transaction Tj to release a data item that it needs.

When transaction Ti requests a data item currently being held by transaction Tj, then the edge Ti → Tj is inserted in the wait-for graph. This edge is removed only when transaction Tj is no longer holding a data item needed by transaction Ti.

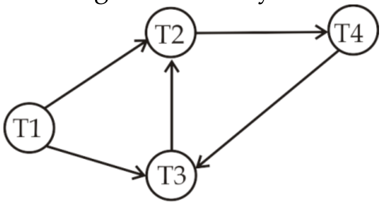
A deadlock exists in the system if and only if the wait-for graph contains a cycle. Each transaction involved in the cycle is said to be deadlocked. To detect deadlocks, the system needs to maintain the wait-for graph, and periodically to invoke an algorithm that searches for a cycle in the graph.

Wait for graph with no cycle




Since the graph has no cycle, the system is not in a deadlock state.

Suppose now that transaction T4 is requesting an item held by T3. The edge T4 → T3 is added to the wait-for graph, resulting in the new system state. This time, the graph contains the cycle:




T2 →T4 →T3 →T2
implying that transactions T2, T3, and T4 are all deadlocked.

**Deadlock Recovery**

When a detection algorithm determines that a deadlock exists, the system must recover from the deadlock. The most common solution is to roll back one or more transactions to break the deadlock. Choosing which transaction to abort is known as Victim Selection.

**Selection of a victim:** In the above discussed wait-for graph transactions T3, T2 and T4 are deadlocked. In order to remove deadlock one of the transaction out of these three transactions must be roll backed.

We should roll back those transactions that will incur the minimum cost. When a deadlock is detected, the choice of which transaction to abort can be made using following criteria:

- The transaction which have the fewest locks
- The transaction that has done the least work
- The transaction that is farthest from completion

**Rollback:** Once we have decided that a particular transaction must be rolled back, we must determine how far this transaction should be rolled back.
The simplest solution is a total rollback: Abort the transaction and then restart it. However, it is more effective to roll back the transaction only as far as necessary to break the deadlock. Such partial rollback requires the system to maintain additional information about the state of all the running transactions. Specifically, the sequence of lock requests/grants and updates performed by the transaction needs to be recorded. The deadlock detection mechanism should decide which locks the selected transaction needs to release in order to break the deadlock. The selected transaction must be rolled back to the point where it obtained the first of these locks, undoing all actions it took after that point. The recovery mechanism must be capable of performing such partial rollbacks. Furthermore, the transactions must be capable of resuming execution after a partial rollback.

**Problem of Starvation:** In a system where the selection of victims is based primarily on cost factors, it may happen that the same transaction is always picked as a victim. As a result this transaction never completes can be picked as a victim only a (small) finite number of times. The most common solution is to include the number of rollbacks in the cost factor.

## PRACTICE SET

1. Mechanism developed to enforce users to enter data in required format is?
(a) Data validation (b) Input mask
(c) Criteria (d) Data verification
(e) None of these
2. What is the size of Data & Time field type?
(a) 1 (b) 8
(c) 255 (d) 50
(e) None of these
3. The options like Save, Open Database, Print are available in
(a) Home tab (b) Backstage View tab
(c) None of these (d) Database Tools tab
(e) File menu
4. Which of the following method can be used to add more tables in a database?
(a) Design View (b) Table Wizard
(c) By Entering Data (d) All of above
(e) None of these
5. The feature that database allows to access only certain records in database is?
(a) Forms (b) Reports

(c) Queries (d) Tables
(e) None of these

6. You can find Sort & Filter group of commands in?
(a) Create ribbon (b) Home ribbon
(c) Database tools ribbon
(d) Fields ribbon (e) None of these

7. Arrange according to the size
(a) Record, field, byte, bit (b) Bit, field, byte, record
(c) Field, byte, record, bit (d) Byte, bit, record, field
(e) None of these

8. What is the maximum allowed field size for Boolean (Yes/No) fields?
(a) 1 (b) 8
(c) 50 (d) 255 (e) None of these

9. What is relational database?
(a) A database structured to recognize relations between stored items of information. It is based on the *relational* model of data
(b) A database that is not related to other databases
(c) A database to store human relations
(d) Both (a) and (b)
(e) None of these

10. What is the best data type for a field that stores mobile numbers?
(a) Memo (b) Number
(c) Date/Time (d) Text
(e) None of these

11. Which filter method lets you filter records based on criterion you specify?
(a) Filter by form (b) Filter by selection
(c) Auto filter (d) Advanced filter
(e) None of these

12. Which of the following field type is used to store photograph of employees?
(a) Memo (b) None of these
(c) OLE (d) Photo
(e) Picture

13. Which of the following in not a function of DBA?
(a) Network Maintenance
(b) Routine Maintenance
(c) Schema Definition
(d) Authorization for data access
(e) None of these

14. Which of the following is a Data Model?
(a) Entity-Relationship model
(b) Relational data model
(c) Object-Based data model
(d) Network model (e) All of the above

15. Which of the following is not Modification of the Database?
(a) Deletion (b) Insertion
(c) Sorting (d) Updating
(e) None of these

16. Which of the following represents a relationship among a set of values?
(a) A row (b) A table
(c) A field (d) A column
(e) None of these

17. Column header is referring as?
(a) Table (b) Relation
(c) Attributes (d) Domain
(e) None of these

18. In mathematical term table is referred as?
(a) Relation (b) Attribute
(c) Tuple (d) Domain
(e) None of these

19. Minimal Super keys are called?
(a) Schema keys (b) Candidate keys
(c) Domain keys (d) Attribute keys
(e) None of these

20. The Primary key must be?
(a) Non Null (b) Unique
(c) Either (e) or (b) (d) Both (a) and (b)
(e) Null

21. By Grouped Report you understand-
(a) Type of report generated by the Report Wizard
(b) Type of report that present records sorted in ascending or descending order as you specify
(c) Type of report that displays data grouped by fields you specified
(d) Both (a) and (b) (e) None of these

22. Which of the following is not a level of data abstraction?
(a) Physical Level (b) Critical Level
(c) Logical Level (d) View Level
(e) None of these

23. Data Manipulation Language enables users to
(a) Retrieval of information stored in database
(b) Only (a) and (b)
(c) Insertion of new information into the database
(d) Deletion of information from the database
(e) All of the above

24. Which of the following is not a Storage Manager Component?
(a) Transaction Manager (b) Logical Manager
(c) Buffer Manager (d) None of these
(e) File Manager

25. To display associated record from related table in datasheet view, you can?
(a) Double click the record
(b) Apply filter by form command
(c) Single click on expand indicator (+) next to the record
(d) Double click on expand indicator (+) next to the record
(e) None of these

26. A Relation is a
(a) Subset of a Cartesian product of a list of attributes
(b) Subset of a Cartesian product of a list of domains
(c) Subset of a Cartesian product of a list of tuple
(d) Subset of a Cartesian product of a list of relations
(e) None of these

27. Who proposed the relational model?
(a) Bill Gates (b) E.F. Codd
(c) Herman Hollerith (d) Charles Babbage

(e) None of these

28. In an Entity-Relationship Diagram "Ellipses" represents
(a) Attributes (b) Weak entity set
(c) Relationship sets (d) Multi-valued attributes
(e) None of these

29. What type of relationship exists between a Teacher table and Class table?
(a) One to many (b) Many to many
(c) One to one (d) Two to two
(e) None of these

30. Data Manipulation Language (DML) is not to
(a) Create information table in the Database
(b) Insertion of new information into the Database
(c) Deletion of information in the Database
(d) Modification of information in the Database
(e) None of these

31. Which of the following in true regarding Referential Integrity?
(a) Every primary-key value must match a primary-key value in an associated table
(b) Every primary-key value must match a foreign-key value in an associated table
(c) Every foreign-key value must match a primary-key value in an associated table
(d) Every foreign-key value must match a foreign-key value in an associated table
(e) None of these

32. Group names in ribbon can be helpful to
(a) Group the commands so that when you move one, you can move all of them together
(b) Give a name for buttons on ribbon
(c) Find the required option by inspecting through them
(d) All of above
(e) None of these

33. Which of the following is an unary operation?
(a) Selection operation (b) Generalized selection
(c) Primitive operation (d) Projection operation
(e) None of these

34. Which of the following NF is based on Multi Valued Dependency?
(a) First (b) Second
(c) Third (d) Fourth
(e) None of these

35. Which of the following in not Outer join?
(a) Left outer join (b) Right outer join
(c) Full outer join (d) All of the above
(e) None of these

36. A relation that has no partial dependencies is in which normal form?
(a) First (b) Second
(c) Third (d) BCNF
(e) None of these

37. A transaction completes its execution is said to be
(a) Saved (b) Loaded
(c) Rolled (d) Committed
(e) None of these

38. An entity type whose existence depends on another entity type is called a _____ entity.
(a) Strong (b) Weak
(c) Codependent (d) Variant
(e) Independent.

39. The advantage of computerized database over manual database is?
(a) We can get the information quickly
(b) We can put in information quick
(c) Solve the repeated information and consistency problem
(d) All of above
(e) None of these

40. The property of transaction which ensures that either all operations of the transaction are reflected properly in the database or none, is called
(a) Atomicity (b) Durability
(c) Isolation (d) Consistency
(e) Deadlock.

41. In a super type/subtype hierarchy, each subtype has?
(a) Only one super type (b) Many super types
(c) At most two super types
(d) At least one subtype (e) Not at all.

42. A property or characteristic of an entity type that is of interest to the organization is called an
(a) Attribute (b) Coexisting entity
(c) Relationship (d) Cross-function
(e) Weak entity.

43. In the context of a database table, the statement "A determines B" indicates that
(a) Knowing the value of attribute A you cannot look up the value of attribute B
(b) You do not need to know the value of attribute A in order to look up the value of attribute B
(c) Knowing the value of attribute B you can look up the value of attribute A
(d) Knowing the value of attribute A you can look up the value of attribute B
(e) None of the above.

44. A method that speeds query processing by running a query at the same time against several partitions of a table using multi processors is called
(a) Multiple partition query
(b) Perpendicular query processing
(c) Parallel query processing
(d) Query optimization (e) Query Execution.

45. A database management software (DBMS) includes
(a) Automated tools (CASE) used to design databases and application programs
(b) A software application that is used to define, create, maintain and provide controlled access to user databases
(c) Application programs that are not used to provide information to users
(d) Database that contains only occurrences of logically organised data or information

(e) Repository of meta data, which is a central storehouse for all data definitions, data relationships, screen and report formats and other system components.

46. Which of the following statements concerning the primary key is true?
(a) All primary key entries are unique
(b) The primary key may be null
(c) The primary key is not required for all tables
(d) The primary key data does not have to be unique
(e) None of these

47. An index record appears for every search key value in the file is
(a) Secondary index (b) Dense index
(c) Sparse index (d) Multi level index
(e) B+ tree.

48. What does the following SQL statement do?
Select * From Customer Where Cust_Type = "Best";
(a) Selects all the fields from the Customer table for each row with a customer labeled "best"
(b) Selects the "*" field from the Customer table for each row with a customer labeled "best"
(c) Selects fields with a "*" in them from the Customer table
(d) Selects all the fields from the Customer table for each row with a customer labeled "*"
(e) Counts all records and displays the value.

49. If K is a foreign key in a relation R1, then
(a) Every tuple of R1 has a distinct value for K
(b) K cannot have a null value for tuples in R1
(c) K is a key for some other relation
(d) K is a Primary key for R1
(e) K is a Composite key for R1.

50. Select the correct statement from the following on proper naming of schema constructs:
(a) Entity type name applies to all the entities belonging to that entity type and therefore a plural name is selected for entity type
(b) In the narrative description of the database requirements, verbs tend to indicate the names of relationship types
(c) The nouns arising from a database requirement description can be considered as names of attributes
(d) Additional nouns which are appearing in the narrative description of the database requirements represent the weak entity type names
(e) Adjectives written in the database requirement description help to identify the partial relationships among entities.

51. Embedded SQL means
(a) Using the EMBED key word in a SQL statement
(b) Writing a SQL statement to retrieve data from more than one relation
(c) Writing SQL statements within codes written in a general programming language
(d) Specifying a condition and action to be taken in case the given condition is satisfied in a trigger
(e) Using SQL language constructs like revoke and grant respectively for revoking and granting privileges to users.

52. State the unit of storage that can store one or more records in a hash file organization
(a) Buckets (b) Disk pages
(c) Blocks (d) Nodes (e) Baskets.

53. Which of the following questions is answered by the SQL statement?
Select Count (Product_Description) from Product_T;
(a) How many products are in the Product Table?
(b) How many different product descriptions are in the Product Table?
(c) How many characters are in the field name "Product_Description"?
(d) How many different columns named "Product Description" is there in table Product_T?
(e) How many total records in a table?

54. Consider the following table obtained using Student and Instructor relations.

| **Fname:** | **Lname:** |
|---|---|
| Ajith | Gamage |
| Sujith | Hewage |
| Kasun | Peiris |

Which relational algebra operation could have been applied on the pair of relations Student and Instructor to obtain the above data?
(a) Student n Instructor (b) Instructor ÷ Student
(c) Student – Instructor (d) Student ? Instructor
(e) Instructor – Student.

55. Which of the following type of index is automatically created when we do not specify?
(a) Bitmap (b) Balanced Tree Index
(c) Binary Tree Index (d) Hashed
(e) Sparse Index.

56. Which of the following is a procedure for acquiring the necessary locks for a transaction where all necessary locks are acquired before any are released?
(a) Record controller (b) Exclusive lock
(c) Authorization rule (d) Two phase lock
(e) Three Phase lock.

57. In the relational modes, cardinality is termed as
(a) Number of tuples (b) Number of attributes
(c) Number of tables (d) Number of constraints
(e) None of these

58. Out of the following activities, which is the one that normally performed by DBMS, without the interference of the DBA?
(a) Integrity (b) Retention
(c) Security (d) Granting the Privileges
(e) Recovery.

59. Which of the following Relational Algebra operations require that both tables (or virtual tables) involved have the exact same attributes/data types?
(a) Join, Projection, Restriction
(b) Multiplication and Division
(c) Union, Intersection, Minus
(d) Minus, Multiplication, Intersection

(e) Projection, Selection, Rename.

60. Which of the following is a component of the relational data model included to specify business rules to maintain the integrity of data when they are manipulated?
(a) Business rule constraint (b) Data integrity
(c) Business integrity (d) Data structure
(e) Entity Integrity.

## SOLUTIONS

1. (b); 'Input mask'
2. (b); '8'
3. (e); 'File menu'
4. (d); 'All of above'
5. (c); 'Queries'
6. (b); 'Home ribbon'
7. (a); 'Record, field, byte, bit'
8. (a); '1'
9. (a)
10. (d); 'Text'
11. (a); 'Filter by form'
12. (c); 'OLE'
13. (a); 'Network Maintenance'
14. (e); 'All of the above'
15. (c); 'Sorting'
16. (a); 'A row'
17. (c); 'Attributes'
18. (a); 'Relation'
19. (b); 'Candidate key'
20. (d); 'Both (a) and (b)'
21. (b);
22. (b)
23. (e)
24. (b)
25. (c)
26. (b)
27. (b)
28. (a)
29. (b)
30. (a)
31. (c)
32. (c)
33. (b)
34. (d)
35. (d)
36. (b)
37. (d)
38. (b)
39. (d)
40. (a)
41. (a); Reason: In a super type/sub-type hierarchy, each sub-type has only one super type
42. (a); Reason: A property or characteristic of an entity type that is of interest to the organization is called attribute
43. (d); Reason: Knowing the value of attribute A you can look up the value of attribute B.
44. (c); Reason: A method that speeds query processing by running a query at the same time against several partitions of a table using multi processors is called parallel query processing.
45. (b); Reason: A software application that is used to define, create, maintain and provide controlled access to user databases.
46. (a)
47. (b); Reason: Dense Index record appears for every search key valued in the file.
48. (a);
49. (c); Reason : If k is a foreign key in a relation R1, then K is a key for some other relation.
50. (b); Reason: In the narrative description of the database requirements, verbs tend to indicate the names of relationship types.
51. (c); Embedded SQL refers to writing SQL statements within codes written in a general programming language.
52. (a); Buckets are used to store one or more records in a hash file organization.
53. (b); Reason: How many different product descriptions are in the Product Table?
54. (e); Instructor – Student is the relational algebra operation that could be applied on the pair of relations Student and Instructor to obtain the above data.
55. (b); Balanced Tree Index is automatically created when we do not specify.
56. (d); Two-phase lock is a procedure for acquiring the necessary locks for a transaction where all necessary locks are acquired before any are released
57. (a);
58. (e); Recovery is the one that normally is performed by DBMS, without the interference of the DBA
59. (c); n relational algebra Union, Intersection, Minus operations require that both tables (or virtual tables) involved have the exact same attributes/data types.
60. (b); Data integrity is a component of the relational data model included to specify business rules to maintain the integrity of data when they are manipulated

# 3 DATA WAREHOUSING & DATAMINING

## 1. Data warehousing:

### 1.1 Introduction

Data warehousing is combining data from multiple sources into one comprehensive and easily manipulated database. The primary aim for data warehousing is to provide businesses with analytic results from data mining, OLAP, Score carding and reporting.

A data warehouse is a subject-oriented, integrated, time-variant and non-volatile collection of data in support of management's decision making process.

**Subject Oriented**: Data that gives information about a particular subject instead of about a company's ongoing operations.

**Integrated**: Data that is gathered into the data warehouse from a variety of source and merged into a coherent whole.

**Time variant**: All data in the data warehouse is identified with a particular time period.

**Non-volatile**: Data is stable in a data warehouse. More data is added but data is never removed.

### 1.2. Benefits of data warehousing:

- Enhance Business Intelligence
- Increased Query and System performance
- Business Intelligence from Multiple Source
- Timely Access to Data
- Enhanced Data quality and Consistency
- Historical intelligence
- High return on investment

### 1.3. Operational and Informational Data

**Operational Data:**

- Focusing on transactional function such as bank card withdrawals and deposits
- Detailed
- Updateable
- Reflects current data

**Informational Data:**

- Focusing on providing answers to problems posed by decision makers
- summarized
- Non updateable

**Important Terminology:**

**Enterprise Data Warehouse-** It collects all information about subjects (Customers, products, sale assets, personnel) that span the entire organization.

**Data Mart-** Departmental subsets that focus on selected subjects. A data mart is a segment of a data warehouse that can provide data for reporting and analysis on a section, unit, department or operation in the company, e.g. sales, payroll, production. Data marts are sometimes complete individual data warehouses which are usually smaller than the corporate data warehouse.

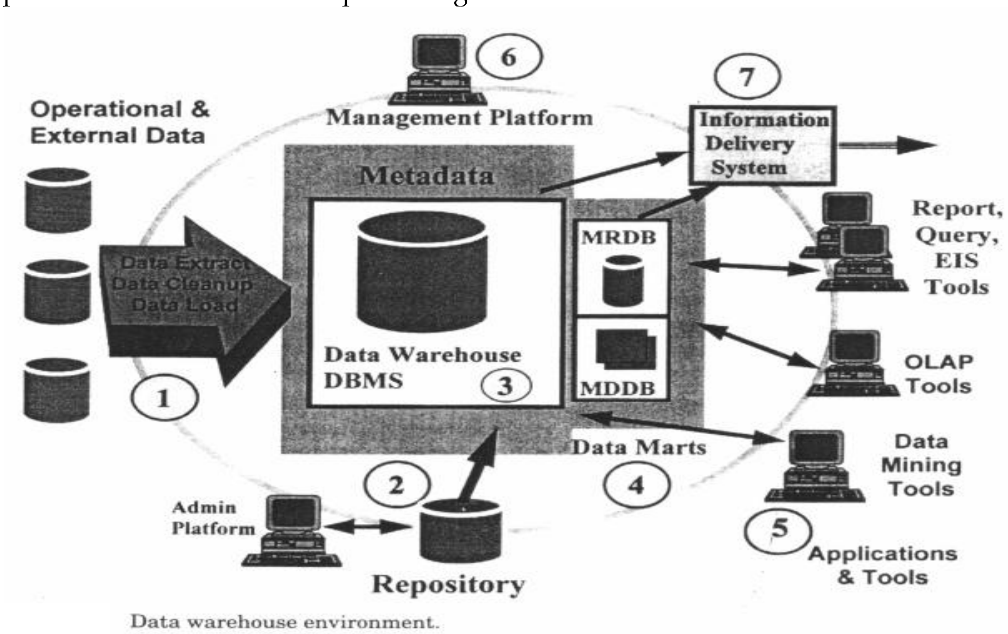
**Decision Support System(DSS)-** Information technology to help the knowledge worker(executive,manager and analyst) makes faster and better decision.

**Drill-down-** Traversing the summarization levels from highly summarized data to the underlying current or old detail.

**Meta data-** Data about data. Containing location and description of warehouse system components: names, definition, structure etc.

## 1.4. Data Warehouse Architecture:

Data warehouse architecture is based on a relational database management system server that functions as the central repository for informational data. In the data warehouse architecture, operational data and processing is completely separate from data warehouse processing.



Data warehouse environment.

Components of Data Warehouse Architecture:
1. Data sourcing, cleanup, transformation and migration tools.
2. Metadata repository
3. Warehouse database technology
4. Data marts
5. Data query, reporting, analysis and mining tools
6. Data warehouse administration and management
7. Information delivery system

NOTE on Operational Data Store: Operational Data store (ODS) is an architecture concept to support day-to-day operational decision support and contains current value data propagated from operational applications
- ODS is subject-oriented, similar to a classic definition of a Data warehouse
- ODS is integrated, in the same sense as a data warehouse

However
- ODS is volatile, while a data warehouse is nonvolatile
- ODS contains very current data, while a data warehouse contains both current and historical data.
- ODS contains detailed data only, and not precalculated summaries and aggregates, as is typical for a data warehouse.

## 1.5. Data Warehouse Design Approaches:

Designing or Building of a Data Warehouse can be done following either one of the approaches. These approaches are notably known as:

- **Top-Down Approach-**In the top down approach suggested by Bill Inmon, we build a centralized repository to house corporate wide business data. This repository is called Enterprise Data Warehouse (EDW). The data in the EDW is stored in a normalized form in order to avoid redundancy. The central repository for corporate wide data helps us maintain one version of truth of the data. The data in the EDW is stored at the most detail level. The reason to build the EDW on the most detail level is to leverage- Flexibility to be used by multiple departments and Flexibility to cater for future requirements.
- **Bottom-Up Approach-**The bottom up approach suggested by Ralph Kimball is an incremental approach to build a data warehouse. Here we build the data marts separately at different points of time as and when the specific subject area requirements are clear. The data marts are integrated or combined together to form a data warehouse. Separate data marts are combined through the use of conformed dimensions and conformed facts. A conformed dimension and a conformed fact is one that can be shared across data marts.

A Conformed dimension has consistent dimension keys, consistent attribute names and consistent values across separate data marts. The conformed dimension means exact same thing with every fact table it is joined.
A Conformed fact has the same definition of measures, same dimensions joined to it and at the same granularity across data marts.
The bottom up approach helps us incrementally build the warehouse by developing and integrating data marts as and when the requirements are clear. We don't have to wait for knowing the overall requirements of the warehouse

## 1.6. Meta Data:

It is data about data. It is used for maintaining, managing and using the data warehouse. It is classified into two:
***Technical Meta data*:** It contains information about data warehouse data used by warehouse designer, administrator to carry out development and management tasks. It includes,

- Info about data stores
- Transformation descriptions. That is mapping methods from operational db to warehouse db
- Warehouse Object and data structure definitions for target data
- The rules used to perform clean up, and data enhancement
- Data mapping operations
- Access authorization, backup history, archive history, info delivery history, data acquisition history, data access etc.,

***Business Meta data:*** It contains info that gives info stored in data warehouse to users. It includes,

- Subject areas, and info object type including queries, reports, images, video, audio clips etc.
- Internet home pages
- Info related to info delivery system
- Data warehouse operational info such as ownerships, audit trails etc.,

Meta data helps the users to understand content and find the data. Meta data are stored in a separate data stores which is known as informational directory or Meta data repository which helps to integrate, maintain and view the contents of the data warehouse.

A metadata repository is a database of data about data (metadata). The purpose of the metadata repository is to provide a consistent and reliable means of access to data. The repository itself may be stored in a physical location or may be a virtual database, in which metadata is drawn from separate sources.

## 1.7. Access Tools In Data warehouse

Its purpose is to provide info to business users for decision making. There are five categories:

1. **Data query and reporting tools-** Query and reporting tools are used to generate query and report.

2. **Application development tools-** It used to generate SQL query. It uses Meta layer software in between users and databases which offers a point-and-click creation of SQL statement.
3. **Executive info system tools (EIS)-** An Executive Information System (EIS) as a management information system is generally designed to be emphasized with graphical display and very easy to use and appealing interfaces as this is assumed to be used for supporting and facilitating the information and decision making needs of senior executives.
4. **OLAP tools-** These tools are based on concepts of multidimensional database and allow a sophisticated user to analyse the data using elaborate, multidimensional and complex views. Typical business applications for these tools include product performance and profitability, effectiveness of a sales program or a marketing campaign, sales forecasting and capacity planning. These tools assume that the data is organised in a multidimensional model, which is supported by a special multidimensional database or by a Relational database designed to enable multidimensional properties.
5. **Data mining tools-** Data mining tools are used to discover knowledge from the data warehouse data also can be used for data visualization and data correction purposes.

## 1.8. Data marts:

Data mart is a subset of a data warehouse that support the requirements of particular department or business function.
The characteristic that differentiate data marts and data warehouse is, a data mart focuses only the requirements of users associated with one department or business function.

## 1.9. OLAP (Online Analytical Processing):

OLAP is an approach to answering multi dimentional analytical queries. OLAP is part of the broader category of business intelligence, which also encompasses relational database, report writing and data mining. OLAP tools enable users to analyze multidimentional data interactively from multiple perspectives.
The OLAP databases are highly de-normalized, which makes the files redundant and helps to improve analytic performance. The processing speed of the system is very slow and can take up to many hours depending on the data involved.
**Types of OLAP:**
- Relational OLAP
- Multidimentional OLAP
- Hybrid OLAP

## 1.10. OLTP (Online Transaction Processing):

It is a class of systems that facilitate and manage transaction-oriented applications, typically for data entry and retrieval transaction processing. It manages current data and stores all of the given data. It is characterized by a large number of short online transactions and their quick real time response to the users.
The main purpose of the OLTP system is to control or run the fundamental business tasks.

## 1.11. Difference between OLAP and OLTP:

| OLAP | OLTP |
|---|---|
| OLAP technology used to perform complex analysis of the data in a data warehouse. | OLTP technology used to preform updates on operational or transactional system (e.g. Point of sale system). |
| It holds historical and only relevant data | It holds current and all type of data |
| It has few concurrent users. | It has many concurrent users. |
| OLAP systems are used by knowledge workers such as executives, managers and analysts. | OLTP systems are used by clerks, DBAs, or database professionals. |
| Provides summarized and multidimensional view of data. | Provides detailed and flat relational view of data. |
| Based on Star Schema, Snowflake, Schema and Fact Constellation Schema. | It is highly normalized with many tables. |

## 2. Data Mining:

Data mining is a process of extracting previously unknown, valid and actionable information from large set of data and then using the information to make crucial business decision.
Data mining is concerned with the analysis of data and the use of software techniques for finding hidden and unexpected patterns and relationships in sets of data. The focus of data mining is to find the information that is hidden and unexpected.

### 2.1. Data Mining Techniques:

- Association
- Classification
- Clustering
- prediction
- Sequential Patterns
- Decision Trees
- Combinations

### 2.2. Data Mining Applications:

Various fields uses data mining technologies because of fast access of data and valuable information from vast amount of data. Data mining technologies have been applied successfully in many areas:

- **Financial Data Analysis:**
  The financial data in banking and financial industry is generally reliable and of high quality which facilitates the systematic data analysis and data mining. Here are the few typical cases: Design and construction of data warehouses for multidimensional data analysis and data mining. Loan payment prediction and customer credit policy analysis. Classification and clustering of customers for targeted marketing. Detection of money laundering and other financial crimes

- **Retail Industry:**
  Data Mining has its great application in Retail Industry because it collects large amount data from on sales, customer purchasing history, goods transportation, consumption and services. It is natural that the quantity of data collected will continue to expand rapidly because of increasing ease, availability and popularity of web. The Data Mining in Retail Industry helps in identifying customer buying patterns and trends. That leads to improved quality of customer service and good customer retention and satisfaction.

- **Telecommunication Industry:**
  Today the Telecommunication industry is one of the most emerging industries providing various services such as fax, pager, cellular phone, Internet messenger, images, email, web data transmission etc. Due to the development of new computer and communication technologies, the telecommunication industry is rapidly expanding. This is the reason why data mining is become very important to help and understand the business. Data Mining in Telecommunication industry helps in identifying the telecommunication patterns, catch fraudulent activities, make better use of resource, and improve quality of service.

- **Biological Data Analysis:**
  Now a days we see that there is vast growth in field of biology such as genomics, proteomics, functional Genomics and biomedical research. Biological data mining is very important part of Bioinformatics. Following are the aspects in which Data mining contribute for biological data analysis:
  - Semantic integration of heterogeneous, distributed genomic and proteomic databases.
  - Alignment, indexing, similarity search and comparative analysis multiple nucleotide sequences.
  - Discovery of structural patterns and analysis of genetic networks and protein pathways.

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

## 2.3. Difference Between Data warehousing and Data Mining:

Data mining is a method for comparing large amounts of data for the purpose of finding patterns. Data mining is normally used for models and forecasting. Data mining is the process of correlations, patterns by shifting through large data repositories using pattern recognition techniques.

Data warehousing is the central repository for the data of several business systems in an enterprise. Data from various resources extracted and organized in the data warehouse selectively for analysis and accessibility.

# PRACTICE SET

1. The data is stored, retrieved and updated in……..?
   (a) OLAP (b) OLTP
   (c) SMTP (d) FTP
   (e) None of these
2. …………………… is a good alternative to the star schema.
   (a) Star schema (b) Snowflake schema
   (c) Fact constellation (d) Star-snowflake schema
   (e) None of these
3. ……. is an essential process where intelligent methods are applied to extract data patterns.
   (a) Data warehousing (b) Data mining
   (c) Text mining (d) Data selection
   (e) None of these
4. Strategic value of data mining is …………………..?
   (a) cost-sensitive (b) work-sensitive
   (c) time-sensitive (d) technical-sensitive
   (e) None of these
5. The type of relationship in star schema is ………?
   (a) many to many (b) one to one
   (c) one to many (d) many to one
   (e) None of these
6. The ……………… allows the selection of the relevant information necessary for the data warehouse.
   (a) top-down view (b) data warehouse view
   (c) data source view (d) business query view
   (e) None of these
7. Which of the following is not a component of a data warehouse?
   (a) Meta data (b) Current detail data
   (c) Lightly summarized data (d) Component Key
   (e) None of these
8. Which of the following is not an ETL tool?
   (a) Informatica
   (b) Oracle warehouse builder(c) Datastage
   (d) Visual studio (e) DT/studio.
9. Which of the following is/are the Data mining tasks?
   (a) Regression (b) Classification
   (c) Clustering
   (d) Inference of associative rules
   (e) All (a), (b), (c), and (d) above.
10. Which of the following is not a kind of data warehouse application?
    (a) Information processing
    (b) Analytical processing
    (c) Data mining
    (d) Transaction processing
    (e) None of these
11. The output of KDD is …………..?
    (a) Data (b) Information
    (c) Query (d) Useful information
    (e) None of these
12. The full form of OLAP is?
    (a) Online Analytical Processing
    (b) Online Advanced Processing
    (c) Online Advanced Preparation
    (d) Online Analytical Performance
    (e) None of these
13. ……….. is a summarization of the general characteristics or features of a target class of data.
    (a) Data Characterization (b) Data Classification
    (c) Data discrimination (d) Data selection
    (e) None of these
14. Which of the following should not be considered for each dimension attribute?
    (a) Attribute name
    (b) Rapid changing dimension policy
    (c) Attribute definition
    (d) Sample data (e) Cardinality.
15. The full form of KDD is ………………
    (a) Knowledge Database
    (b) Knowledge Discovery Database
    (c) Knowledge Data House
    (d) Knowledge Data Definition
    (e) None of these
16. Which of the following is not a data mining functionality?
    (a) Characterization and Discrimination
    (b) Classification and regression
    (c) Selection and interpretation
    (d) Clustering and Analysis
    (e) None of these
17. ……….. is the process of finding a model that describes and distinguishes data classes or concepts.
    (a) Data Characterization (b) Data Classification
    (c) Data discrimination (d) Data selection
    (e) None of these
18. The process of removing the deficiencies and loopholes in the data is called as?
    (a) Aggregation of data (b) Extracting of data
    (c) Cleaning up of data. (d) Loading of data
    (e) Compression of data.

19. Concept description is the basic form of the?
(a) Predictive data mining
(b) Descriptive data mining
(c) Data warehouse (d) Relational data base
(e) Proactive data mining.
20. An OLAP tool provides for
(a) Only transaction control
(b) Only Roll-up and drill-down
(c) Slicing and dicing (d) Only Rotation
(e) Setting up only relations.
21. Which one manages both current and historic transactions?
(a) OLTP (b) OLAP
(c) Spread sheet (d) XML
(e) All (a), (b), (c) and (d) above.
22. Which of the following is the collection of data objects that are similar to one another within the same group?
(a) Partitioning (b) Grid
(c) Cluster (d) Table
(e) Data source.
23. Which of the following employees data mining techniques to analyze the intent of a user query, provided additional generalized or associated information relevant to the query?
(a) Iceberg query method (b) Data analyzer
(c) Intelligent query answering
(d) DBA (e) Query parser.
24. Which of the following process includes data cleaning, data integration, data selection, data transformation, data mining, pattern evolution and knowledge presentation?
(a) KDD process (b) ETL process
(c) KTL process (d) MDX process
(e) None of the above.
25. At which level we can create dimensional models?
(a) Business requirements level
(b) Architecture models level
(c) Detailed models level
(d) Implementation level (e) Testing level.
26. Which of the following is not related to dimension table attributes?
(a) Verbose (b) Descriptive
(c) Equally unavailable (d) Complete
(e) Indexed.
27. Data warehouse bus matrix is a combination of
(a) Dimensions and data marts
(b) Dimensions and facts
(c) Facts and data marts
(d) Dimensions and detailed facts
(e) All (a), (b), (c) and (d) above.
28. Which of the following is not the managing issue in the modeling process?
(a) Content of primary units column
(b) Document each candidate data source
(c) Do regions report to zones
(d) Walk through business scenarios
(e) Ensure that the transaction edit flat is used for analysis.
29. Data modeling technique used for data marts is
(a) Dimensional modeling (b) ER – model
(c) Extended ER – model (d) Physical model
(e) Logical model.
30. A warehouse architect is trying to determine what data must be included in the warehouse. A meeting has been arranged with a business analyst to understand the data requirements, which of the following should be included in the agenda?
(a) Number of users (b) Corporate objectives
(c) Database design (d) Routine reporting
(e) Budget.
31. A Business Intelligence system requires data from:
(a) Data warehouse
(b) Operational systems
(c) All possible sources within the organization and possibly from external sources
(d) Web servers
(e) Database servers.
32. Data mining application domains are
(a) Biomedical
(b) DNA data analysis
(c) Financial data analysis
(d) Retail industry and telecommunication industry
(e) All (a), (b), (c), and (d) above.
33. Which of the following project is a building a data mart for a business process/department that is very critical for your organization?
(a) High risk high reward
(b) High risk low reward
(c) Low risk low reward
(d) Low risk high reward
(e) Involves high risks.
34. Which of the following tools a business intelligence system will have?
(a) OLAP tool (b) Data mining tool
(c) Reporting tool (d) Both (a) and (b) above
(e) All (a), (b) and (c) above.
35. The Synonym for data mining is
(a) Data warehouse
(b) Knowledge discovery in database
(c) ETL
(d) Business intelligence
(e) OLAP.
36. The …… exposes the information being captured, stored, and managed by operational systems.
(a) top-down view (b) data warehouse view
(c) data source view (d) business query view
(e) None of these
37. An ……… system is market-oriented and is used for data analysis by knowledge workers, including managers, executives, and analysts.
(a) OLAP (b) OLTP
(c) Both of the above (d) Text Mining
(e) None of these

38.……….. is a comparison of the general features of the target class data objects against the general features of objects from one or multiple contrasting classes.
(a) Data Characterization (b) Data Classification
(c) Data discrimination (d) Data selection
(e) None of these

39.……. is a subject-oriented, integrated, time-variant, nonvolatile collection or data in support of management decisions.
(a) Data Mining (b) Data Warehousing
(c) Document Mining (d) Text Mining
(e) None of these

40. Most common kind of queries in a data warehouse
(a) Inside-out queries (b) Outside-in queries
(c) Browse queries (d) Range queries
(e) All (a), (b), (c) and (d) above.

41. In a data warehouse, if D1 and D2 are two conformed dimensions, then
(a) D1 may be an exact replica of D2
(b) D1 may be at a rolled up level of granularity compared to D2
(c) Columns of D1 may be a subset of D2 and vice versa
(d) Rows of D1 may be a subset of D2 and vice versa
(e) All (a), (b), (c) and (d) above.

42. The generalization of multidimensional attributes of a complex object class can be performed by examining each attribute, generalizing each attribute to simple-value data and constructing a multidimensional data cube is called as
(a) Object cube (b) Relational cube
(c) Transactional cube (d) Tuple
(e) Attribute.

43. Which of the following statements is true?
(a) A fact table describes the transactions stored in a DWH
(b) A fact table describes the granularity of data held in a DWH
(c) The fact table of a data warehouse is the main store of descriptions of the transactions stored in a DWH
(d) The fact table of a data warehouse is the main store of all of the recorded transactions over time
(e) A fact table maintains the old records of the database.

44. What is/are the different types of Meta data?
I. Administrative. II. Business.
III. Operational.
(a) Only (I) above
(b) Both (II) and (III) above
(c) Both (I) and (II) above
(d) Both (I) and (III) above
(e) All (I), (II) and (III) above.

45. Multiple Regression means
(a) Data are modeled using a straight line
(b) Data are modeled using a curve line
(c) Extension of linear regression involving only one predicator value
(d) Extension of linear regression involving more than one predicator value
(e) All (a), (b), (c) and (d) above.

46. Biotope are-
(a) This takes only two values. In general, these values will be 0 and 1 and they can be coded as one bit.
(b) The natural environment of a certain species
(c) Systems that can be used without knowledge of internal operations
(d) None of these
(e) Both (a) and (c)

47. Naive prediction is-
(a) A class of learning algorithms that try to derive a Prolog program from examples
(b) A table with n independent attributes can be seen as an n-dimensional space.
(c) A prediction made using an extremely simple method, such as always predicting the same output.
(d) None of these
(e) Both (a) and (b)

48. The apriori property means
(a) If a set cannot pass a test, all of its supersets will fail the same test as well
(b) To improve the efficiency the level-wise generation of frequent item sets
(c) If a set can pass a test, all of its supersets will fail the same test as well
(d) To decrease the efficiency the level-wise generation of frequent item sets
(e) All (a), (b), (c) and (d) above.

49. Which of following form the set of data created to support a specific short lived business situation?
(a) Personal data marts (b) Application models
(c) Downstream systems
(d) Disposable data marts (e) Data mining models.

50. Which of the following is the most important when deciding on the data structure of a data mart?
(a) XML data exchange standards
(b) Data access tools to be used
(c) Metadata naming conventions
(d) Extract, Transform, and Load (ETL) tool to be used
(e) All (a), (b), (c) and (d) above.

51. The various aspects of data mining methodologies is/are ……………….?
i) Mining various and new kinds of knowledge
ii) Mining knowledge in multidimensional space
iii) Pattern evaluation and pattern or constraint-guided mining.
iv) Handling uncertainty, noise, or incompleteness of data
(a) i, ii and iv only (b) ii, iii and iv only
(c) i, ii and iii only (d) All i, ii, iii and iv
(e) None of these

52. Data mining can also applied to other forms such as …………….?
i) Data streams ii) Sequence data

iii) Networked data    iv) Text data
v) Spatial data
(a) i, ii, iii and v only    (b) ii, iii, iv and v only
(c) i, iii, iv and v only    (d) All i, ii, iii, iv and v
(e) None of these

53. Classification accuracy is?
(a) A subdivision of a set of examples into a number of classes
(b) Measure of the accuracy, of the classification of a concept that is given by a certain theory
(c) The task of assigning a classification to a set of examples
(d) None of these
(e) Both (a) and (b)

54. Learning algorithm referrers to?
(a) An algorithm that can learn
(b) A sub-discipline of computer science that deals with the design and implementation of learning algorithms
(c) A machine-learning approach that abstracts from the actual strategy of an individual algorithm and can therefore be applied to any other form of machine learning.
(d) None of these
(e) All (a), (b) and (c)

55. Knowledge is referred to?
(a) Non-trivial extraction of implicit previously unknown and potentially useful information from data
(b) Set of columns in a database table that can be used to identify each record within this table uniquely
(c) collection of interesting and useful patterns in a database
(d) None of these
(e) Sets of rows in a database table

56. Machine learning is?
(a) An algorithm that can learn
(b) A sub-discipline of computer science that deals with the design and implementation of learning algorithms
(c) An approach that abstracts from the actual strategy of an individual algorithm and can therefore be applied to any other form of machine learning.
(d) None of these
(e) Programming language

57. Query tools are?
(a) A reference to the speed of an algorithm, which is quadratically dependent on the size of the data
(b) Attributes of a database table that can take only numerical values.
(c) Tools designed to query a database.
(d) None of these
(e) Both (a) and (b)

58. Operational database is?
(a) A measure of the desired maximal complexity of data mining algorithms
(b) A database containing volatile data used for the daily operation of an organization
(c) Relational database management system.
(d) None of these
(e) Both (a) and (c)

59. Projection pursuit is?
(a) The result of the application of a theory or a rule in a specific case
(b) One of several possible enters within a database table that is chosen by the designer as the primary means of accessing the data in the table.
(c) Discipline in statistics that studies ways to find the most interesting projections of multi-dimensional spaces
(d) None of these
(e) The result of the algorithm.

60. Case-based learning is?
(a) A class of learning algorithm that tries to find an optimum classification of a set of examples using the probabilistic theory.
(b) Any mechanism employed by a learning system to constrain the search space of a hypothesis
(c) An approach to the design of learning algorithms that is inspired by the fact that when people encounter new situations, they often explain them by reference to familiar experiences, adapting the explanations to fit the new situation.
(d)Both (a) and (c)
(e) None of these

## SOLUTIONS

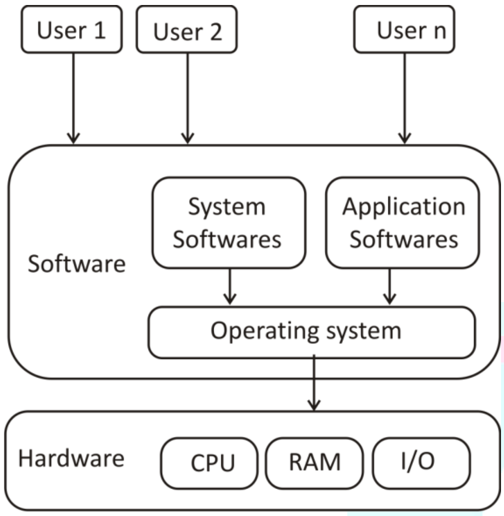
1. (b); OLTP
2. (c); Fact constellation
3. (b); Data mining
4. (c); time-sensitive
5. (c); one to many
6. (a); top-down view
7. (d); Component Key
8. (d); Visual Studio is not an ETL tool.
9. (e); Regression, Classification and Clustering are the data mining tasks.
10. (d); Transaction processing
11. (d); Useful information
12. (a); Online Analytical Processing
13. (a); Data Characterization
14. (b); Rapid changing dimension policy should not be considered for each dimension attribute.
15. (b); Knowledge Discovery Database
16. (c); Selection and interpretation
17. (b); Data Classification

18. (c); The process of removing the deficiencies and loopholes in the data is called as cleaning up of data.
19. (b); Concept description is the basic form of the descriptive data mining.
20. (c);
21. (b); Online Analytical Processing (OLAP) manages both current and historic transactions.
22. (c); Cluster is the collection of data objects that are similar to one another within the same group.
23. (c); Intelligent Query Answering employee's data mining techniques to analyze the intent of a user query provided additional generalized or associated information relevant to the query.
24. (a); KDD Process includes data cleaning, data integration, data selection, data transformation, data mining, pattern evolution, and knowledge presentation.
25. (b); Dimensional models can be created at Architecture models level.
26. (c); Equally unavailable is not related to dimension table attributes.
27. (a); Data warehouse bus matrix is a combination of Dimensions and data marts.
28. (e); Ensure that the transaction edit flat is used for analysis is not the managing issue in the modeling process.
29. (a); Data modeling technique used for data marts is Dimensional modeling.
30. (d); Routine reporting should be included in the agenda.
31. (a); A business Intelligence system requires data from Data warehouse
32. (e); Data mining application domains are Biomedical, DNA data analysis, financial data analysis and Retail industry and telecommunication industry
33. (a); High risk high reward project is a building a data mart for a business process/department that is very critical for your organization
34. (a); Business intelligence system will have OLAP, Data mining and reporting tolls.
35. (b); The synonym for data mining is Knowledge discovery in Database.
36. (c); data source view
37. (a); OLAP
38. (c); Data discrimination
39. (b); Data Warehousing
40. (a); The Most common kind of queries in a data warehouse is Inside-out queries.
41. (b)
42. (a)
43. (a)
44. (c)
45. (a)
46. (d)
47. (a)
48. (b)
49. (d)
50. (c)
51. (b)
52. (c)
53. (a)
54. (b)
55. (c)
56. (d)
57. (d)
58. (d)
59. (a)
60. (c)

# 4 OPERATING SYSTEM

## 1. Introduction

**Operating System** is software that works as an interface between a user and the computer hardware. The primary objective of an operating system is to make computer system convenient to use and to utilize computer hardware in an efficient manner. The operating system performs the basic tasks such as receiving input from the keyboard, processing instructions and sending output to the screen.




### 1.1. FUNCTIONS OF OPERATING SYSTEM:

- Memory Management
- Processor Management
- Device Management
- File Management
- Security
- Control over system performance
- Job accounting
- Error detecting aids
- Coordination between other software and users

### 1.2. TYPES OF OPERATING SYSTEM:

**(a) Batch operating system**

The users of a batch operating system do not interact with the computer directly. Each user prepares his job on an off-line device like punch cards and submits it to the computer operator. To speed up processing, jobs with similar needs are batched together and run as a group. The programmers leave their programs with the operator and the operator then sorts the programs with similar requirements into batches.

The problems with Batch Systems are as follows −

- Lack of interaction between the user and the job.
- CPU is often idle, because the speed of the mechanical I/O devices is slower than the CPU.
- Difficult to provide the desired priority.

**(b) Time-sharing operating systems**

Time-sharing is a technique which enables many people, located at various terminals, to use a particular computer system at the same time. Time-sharing or multitasking is a logical extension of multiprogramming. Processor's time which is shared among multiple users simultaneously is termed as time-sharing.

The main difference between Multiprogrammed Batch Systems and Time-Sharing Systems is that in case of Multiprogrammed batch systems, the objective is to maximize processor use, whereas in Time-Sharing Systems, the objective is to minimize response time.

Multiple jobs are executed by the CPU by switching between them, but the switches occur so frequently. Thus, the user can receive an immediate response. For example, in a transaction processing, the processor executes each user program in a short burst or quantum of computation. That is, if **n** users are present, then each user can get a time quantum. When the user submits the command, the response time is in few seconds at most.
Computer systems that were designed primarily as batch systems have been modified to time-sharing systems.
Advantages of Timesharing operating systems are as follows −

- Provides the advantage of quick response.
- Avoids duplication of software.
- Reduces CPU idle time.

Disadvantages of Time-sharing operating systems are as follows −

- Problem of reliability.
- Question of security and integrity of user programs and data.
- Problem of data communication.

**(c) Multiprocessor system**
A multiprocessor system consists of several processors that share a common physical memory. Multiprocessor system provides higher computing power and speed. In multiprocessor system all processors operate under single operating system. Multiplicity of the processors and how they do act together are transparent to the others.
Following are some advantages of this type of system.

- Enhanced performance
- Execution of several tasks by different processors concurrently, increases the system's throughput without speeding up the execution of a single task.
- If possible, system divides task into many subtasks and then these subtasks can be executed in parallel in different processors. Thereby speeding up the execution of single tasks.

**(d)Distributed operating System**
Distributed systems use multiple central processors to serve multiple real-time applications and multiple users. Data processing jobs are distributed among the processors accordingly.
Processors in a distributed system may vary in size and function. These processors are referred as sites, nodes, computers, and so on.

> The motivation behind developing distributed operating system is the availability of power and inexpensive microprocessor and advances in communication technology.

The advantages of distributed systems are as follows −

- With resource sharing facility, a user at one site may be able to use the resources available at another.
- Speedup the exchange of data with one another via electronic mail.
- If one site fails in a distributed system, the remaining sites can potentially continue operating.
- Better service to the customers.
- Reduction of the load on the host computer.
- Reduction of delays in data processing.

**(e) Real Time operating System**
A real-time system is defined as a data processing system in which the time interval required to process and respond to inputs is so small that it controls the environment. The time taken by the system to respond to an input and display of required updated information is termed as the **response time**. So, in this method, the response time is very less as compared to online processing.

Real-time systems are used when there are rigid time requirements on the operation of a processor or the flow of data and real-time systems can be used as a control device in a dedicated application. A real-time operating system must have well-defined, fixed time constraints, otherwise the system will fail. For example, Scientific experiments, medical imaging systems, industrial control systems, weapon systems, robots, air traffic control systems, etc.
There are two types of real-time operating systems.

**Hard real-time systems**
Hard real-time systems guarantee that critical tasks complete on time. In hard real-time systems, secondary storage is limited or missing, and the data is stored in ROM. In these systems, virtual memory is almost never found.

**Soft real-time systems**
Soft real-time systems are less restrictive. A critical real-time task gets priority over other tasks and retains the priority until it completes. Soft real-time systems have limited utility than hard real-time systems. For example, multimedia, virtual reality, Advanced Scientific Projects like undersea exploration and planetary rovers, etc.

**(f)Embedded Operating System:**
These are embedded in a device, which is located in ROM. These are applicable and developed only for the needed resources and accordingly developed. These OS are less resource intensive. Mainly, applicable in appliances like microwaves, washing machines, traffic control systems etc.

## 1.3. Tasks of OS:

**1. Process Management:**
The operating system manages many kinds of activities ranging from user programs to system programs.
Each of these activities is encapsulated in a process. There are many processes can be running the same program. The five major activities of an operating system in regard to process management are
- Creation and deletion of user and system processes.
- Suspension and resumption of processes.
- A mechanism for process synchronization.
- A mechanism for process communication.
- A mechanism for deadlock handling.

**2.Memory Management:**
- Keeping track of which parts of memory are currently being used and by whom.
- Deciding which processes (or parts thereof) and data to move into and out of memory.
- Allocating and reallocating memory space as needed.

**3. Device Management:**
The assembling-disassembling of I/O peripherals devices are managed by the OS.

**4. Storage Management:**
The three major activities of an operating system in regard to secondary storage management are:
- Managing the free space available on the secondary-storage device.
- Allocation of storage space when new files have to be written.
- Scheduling the requests for memory access.

## 1.4. Operating System - Services

An Operating System provides services to both the users and to the programs.
- It provides programs an environment to execute.
- It provides users the services to execute the programs in a convenient manner.

Following are a few common services provided by an operating system −
- Program execution
- I/O operations
- File System manipulation
- Communication
- Error Detection
- Resource Allocation
- Protection

**Program execution**
Operating systems handle many kinds of activities from user programs to system programs like printer spooler, name servers, file server, etc. Each of these activities is encapsulated as a process.
A process includes the complete execution context (code to execute, data to manipulate, registers, OS resources in use).
Following are the major activities of an operating system with respect to program management −

- Loads a program into memory.
- Executes the program.
- Handles program's execution.
- Provides a mechanism for process synchronization.
- Provides a mechanism for process communication.
- Provides a mechanism for deadlock handling.

**I/O Operation**
An I/O subsystem comprises of I/O devices and their corresponding driver software. Drivers hide the peculiarities of specific hardware devices from the users.
An Operating System manages the communication between user and device drivers.
- I/O operation means read or write operation with any file or any specific I/O device.
- Operating system provides the access to the required I/O device when required.

**File system manipulation**
A file represents a collection of related information. Computers can store files on the disk (secondary storage), for long-term storage purpose. Examples of storage media include magnetic tape, magnetic disk and optical disk drives like CD, DVD. Each of these media has its own properties like speed, capacity, data transfer rate and data access methods.
A file system is normally organized into directories for easy navigation and usage. These directories may contain files and other directions. Following are the major activities of an operating system with respect to file management −
- Program needs to read a file or write a file.
- The operating system gives the permission to the program for operation on file.
- Permission varies from read-only, read-write, denied and so on.
- Operating System provides an interface to the user to create/delete files.
- Operating System provides an interface to the user to create/delete directories.
- Operating System provides an interface to create the backup of file system.

**Communication**
In case of distributed systems which are a collection of processors that do not share memory, peripheral devices, or a clock, the operating system manages communications between all the processes. Multiple processes communicate with one another through communication lines in the network.
The OS handles routing and connection strategies, and the problems of contention and security. Following are the major activities of an operating system with respect to communication −
- Two processes often require data to be transferred between them
- Both the processes can be on one computer or on different computers, but are connected through a computer network.
- Communication may be implemented by two methods, either by Shared Memory or by Message Passing.

**Error handling**
Errors can occur anytime and anywhere. An error may occur in CPU, in I/O devices or in the memory hardware. Following are the major activities of an operating system with respect to error handling −
- The OS constantly checks for possible errors.
- The OS takes an appropriate action to ensure correct and consistent computing.
-

**Resource Management**
In case of multi-user or multi-tasking environment, resources such as main memory, CPU cycles and files storage are to be allocated to each user or job. Following are the major activities of an operating system with respect to resource management −
- The OS manages all kinds of resources using schedulers.
- CPU scheduling algorithms are used for better utilization of CPU.

**Protection**
Considering a computer system having multiple users and concurrent execution of multiple processes, the various processes must be protected from each other's activities.
Protection refers to a mechanism or a way to control the access of programs, processes, or users to the resources defined by a computer system. Following are the major activities of an operating system with respect to protection −
- The OS ensures that all access to system resources is controlled.

- The OS ensures that external I/O devices are protected from invalid access attempts.
- The OS provides authentication features for each user by means of passwords.

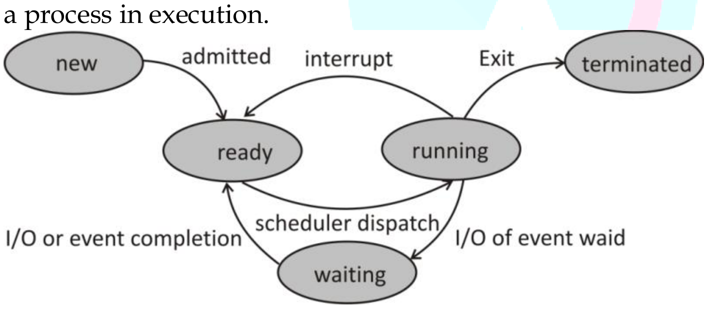
## 2. Process Management:

A program in execution is a process. A process is executed sequentially, one instruction at a time. A program is a passive entity. For example, a file on the disk. A process on the other hand is an active entity. In addition to program code, it includes the values of the program counter, the contents of the CPU registers, the global variables in the data section and the contents of the stack that is used for subroutine calls. A program becomes a process when the executable file of a program is loaded in memory. There can be a possibility in large programs that more than one process is needed to completely run the program, but there are nevertheless considered two separate execution sequences.
Process after being loaded in memory, the above diagram shows the process memory state when it is in execution. The stack region is used to store temporary data being used by the process. The Heap is the dynamic memory region i.e. i.e. on run-time allocated to the process. It consumes the free space in memory (Random access memory). Data region holds the computed results and values, and the global variables being used by the process. The text region holds the code or the executable instructions for the process.

### 2.1. PROCESS STATE

A process being an active entity changes state as execution proceeds. A process can be any one of the following states:
- New: Process being created.
- Running: Instructions being executed.
- Waiting: Process waiting for an event to occur.
- Ready: Process waiting for CPU.
- Terminated: Process has finished execution.

A state diagram is used to diagrammatically represent the states and also the events that trigger the change of state of a process in execution.




### 2.2 PROCESS CONTROL BLOCK

Every process has a number and a process control block (PCB) represents a process in an operating system. The PCB contains information that makes the process an active entity.
The PCB consists of the number of the process, its state, value of the program counter, contents of the CPU registers and much more as given below. The PCB serves as a repository of information about a process and varies from process to process.

Information associated with each process:
1. Process State – The state, may be any of the 5 states shown in process state diagram such as New, ready, running, waiting and terminated.
2. Program Counter – It indicates the address of the next instruction to be executed for this process.
3. CPU registers - the registers highly vary based on number and type and also depending on the computer architecture. They include accumulators, index registers, stack pointers and general-purpose registers. Along with the program counter, this state information must be saved when an interrupt occurs to allow the process to be continued correctly after it recovers from the interrupt.
4. CPU scheduling information – it includes information such as process priority, pointers to scheduling queues and other scheduling parameters.
5. Memory-management information – this information contains value of base register and limit register, page tables, segment tables. These tables are used for referring the right locations of memory to be used by the process.

6. Accounting information – this includes amount of processor time and real time used, time limits, account numbers etc.
7. I/O status information – This includes list of I/O devices allocated to the process, list of open files, etc.

## 2.3. Inter-process Communication

Processes executing concurrently in the operating system may be either independent process or cooperating process. A process is called cooperating if it is affected by the other processes executing in the system, whereas it is called Independent when its execution is not dependent on execution of other processes. Cooperating processes require Inter-process communication (IPC) mechanisms that enable them to send and receive data. Since processes frequently need to communicate with other processes therefore, there is a need for a well-structured communication, without using interrupts, among processes. Inter-process communication has two models:
(a)Message passing system model
(b)Shared memory system model

Message passing is useful for exchanging smaller amounts of data, because no conflicts need be avoided. Message passing is also easier to implement than is shared memory for intercomputer communication. Shared memory allows maximum speed and convenience of communication. Shared memory is faster than message passing, as message-passing systems are typically implemented using system calls and thus require the more time-consuming task of kernel intervention.

**NOTE:** If two processes want to communicate, a communication link must exist between them. This link can be implements in a various ways. But we are concerned here not with the link's physical implementation, rather with its logical implementation. Here are several methods for logically implementing a link and the send/ receive operation.

1. **Direct or Indirect Communication**
Under Direct communication, Process A should explicitly name the recipient of the message i.e. Process B. In this the send( ) and receive( ) are defined as
SEND ( A , MESSAGE ) – Sends a message to process A.
RECEIVE( B, MESSAGE) – Receive a message from process B.
Three characteristics of Direct communication
a) A link is automatically established between every pair of processes that want to communicate. The process only needs to know about other process identity.
b) A link is established with exactly two of the processes.
c) Between each pair of process, there exists only a single link.

Under Indirect communication, messages are sent and received via mailboxes also known as ports. Mailboxes can be simply understood as an object in which messages are temporarily stored by first process and removed by the other process In Indirect communication, the send( ) and receive( ) are defined as
SEND (X, MESSAGE ) – Send/store message to mailbox X
RECEIVE(X, MESSAGE) – Receive/remove message from mailbox X
Three characteristics of Indirect communication
a) A link is established between every pair of processes only if both processes share a common mailbox.
b) A link can be established between two or more processes.
c) Between each pair of process, there may exists more than a single link

**2. Synchronous or Asynchronous Communication**
Communication between processes takes places through calls to send( ) and receive( ) calls. There are different design options for implementing each function. Message passing may be either synchronous and asynchronous or blocking and nonblocking.
Blocking Send – sending is blocking until the message. Is received by receiving process
Non-Blocking Send – sending process sends message and resumes operation
Blocking Receive – receiver blocks until a message is available
Non-Blocking Receive – receiver retrieves either a valid message or a NULL.

**3. Automatic or Explicit Buffering**

Whether communication is direct or indirect, messages exchanged by communicating processes reside in a temporary queue. Depending on how the queue is implemented, the method of execution of message-passing model varies:

(a). Zero Capacity – Queue has a maximum length zero, hence link can't have any messages waiting in it. So sender must wait until recipient retrieves the message.

(b). Bounded Capacity – Finite size queue. Till queue is not full, sender can send messages. Receiver can retrieve messages until queue is empty.

(c). Unbounded Capacity – Queue length is assumed to be infinite. Thus any number of messages can wait in it. The sender is never delayed.

## 2.4. Client-Server Communication

In section above, we explained how communication was done between 2 processes via shared memory or message passing model. This can be used for client – server communications as well, but we have special techniques available, that is especially suited for client – server communication viz Sockets and Remote procedure calls.

### 2.4.1 Sockets

- A socket is defined as an endpoint for communication
- It is a concatenation of IP address and port
- The socket 192.168.10.1:808 refers to port 808 on host 192.168.10.1
- Communication takes place between a pair of sockets

When a client process initiates a request for communication, it is assigned a port by the host computer. This port is a random number and is greater than 1024. The scenario can be something as shown in diagram above. The packets travelling between the hosts are delivered to the appropriate process based on the destination port number. All connection must be unique. Therefore, if another process also on host X wishes to establish another communication with the same web server, it would be assigned a port number greater than 1024, but not equal to 1625. This thus ensures all connections consist of a unique pair of sockets.

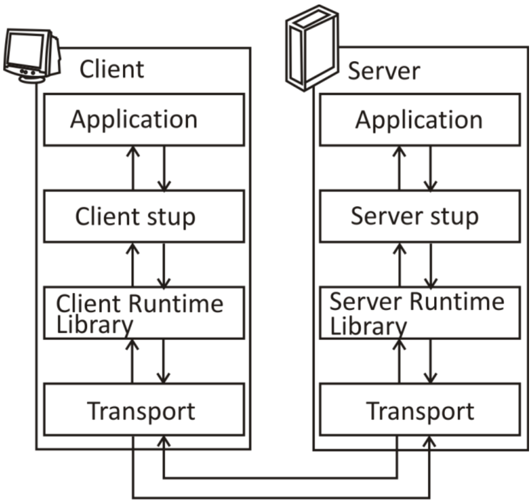
### 2.4.2 Remote Procedure Calls

Remote procedure calls (RPC) is a powerful technique for constructing client-server based applications. It is based on extending the notion of conventional, or local procedure calling, so that the called procedure need not exist in the same address space as the calling procedure.

The two processes may be on the same system, or they may be on different systems with a network connecting them. By using RPC, programmers of distributed applications avoid the details of the interface with the network. The transport independence of RPC isolates the application from the physical and logical elements of the data communications mechanism and allows the application to use a variety of transports.
RPC is similar in many respects to IPC mechanism, and is often built on top of them. But usually RPC is used when the processes are executing on two different systems, client and server. The RPC allows a user to do the communication as if the processes are in same address space, and the complexity of the network, the packing-unpacking of data packets from client-server-client is handled in background by the background stubs.

**How RPC Works**

An RPC is analogous to a function call. Like a function call, when an RPC is made, the calling arguments are passed to the remote procedure and the caller waits for a response to be returned from the remote procedure. The client makes a procedure call that sends a request to the server and waits. When the request arrives, the server calls a dispatch route that performs the requested service, and sends the reply to the client. After the RPC call is completed, the client program continues. RPC specifically supports network applications.
The diagram clearly explains how the remote procedure call works.

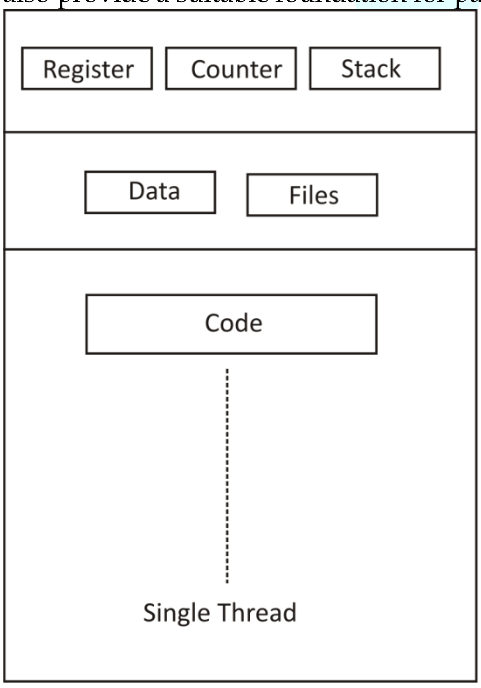



## 2.5. THREAD

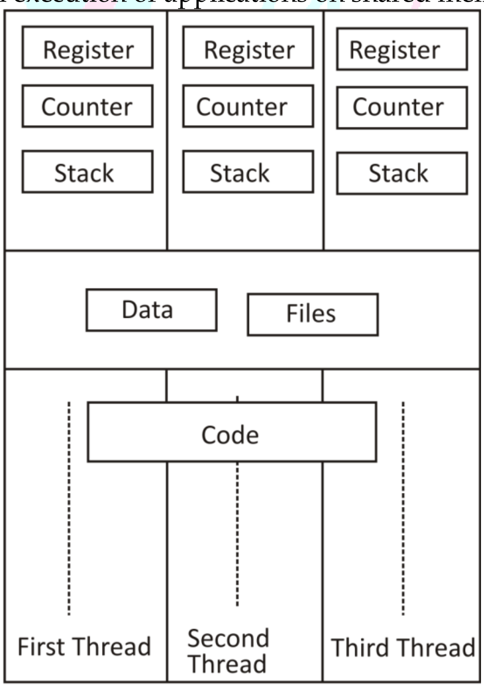
A thread is a flow of execution through the process code, with its own program counter that keeps track of which instruction to execute next, system registers which hold its current working variables, and a stack which contains the execution history. A thread shares with its peer threads few information like code segment, data segment and open files. When one thread alters a code segment memory item, all other threads see that.

A thread is also called a **lightweight process**. Threads provide a way to improve application performance through parallelism. Threads represent a software approach to improving performance of operating system by reducing the overhead thread is equivalent to a classical process.

Each thread belongs to exactly one process and no thread can exist outside a process. Each thread represents a separate flow of control. Threads have been successfully used in implementing network servers and web server. They also provide a suitable foundation for parallel execution of applications on shared memory multiprocessors.




Single Process P with single thread




Single Process P with three threads

### 2.5.1. Benefits

1. Responsiveness. Multithreading an interactive application may allow a program to continue running even if part of it is blocked or is performing a lengthy operation, thereby increasing responsiveness to the user. For instance, a multithreaded web browser could still allow user interaction in one thread while an image was being loaded in another thread.

2. Resource sharing. By default, threads share the memory and the resources of the process to which they belong. The benefit of sharing code and data is that it allows an application to have several different threads of activity within the same address space.
3. Economy. Allocating memory and resources for process creation is costly. Because threads share resources of the process to which they belong, it is more economical to create and context-switch threads. Empirically gauging the difference in overhead can be difficult, but in general it is much more time consuming to create and manage processes than threads. In Solaris, for example, creating a process is about thirty times slower than is creating a thread, and context switching is about five times slower.
4. Utilization of multiprocessor architectures. The benefits of multithreading can be greatly increased in a multiprocessor architecture, where threads may be running in parallel on different processors. A single threaded process can only run on one CPU, no matter how many are available. Multithreading on a multi-CPU machine increases concurrency.

## 2.5.2. Types of Thread

**1. User-Level Threads**
User-level threads implement in user-level libraries, rather than via systems calls, so thread switching does not need to call operating system and to cause interrupt to the kernel. In fact, the kernel knows nothing about user-level threads and manages them as if they were single-threaded processes. Three primary thread libraries: POSIX Pthreads, Win32 threads, Java threads.

**Advantages**
- Thread switching does not require Kernel mode privileges.
- User level thread can run on any operating system.
- Scheduling can be application specific in the user level thread.
- User level threads are fast to create and manage.

**Disadvantages**
- In a typical operating system, most system calls are blocking.
- Multithreaded application cannot take advantage of multiprocessing.

**2. Kernel-Level Threads**
In this method, the kernel knows about and manages the threads. No runtime system is needed in this case. Instead of thread table in each process, the kernel has a thread table that keeps track of all threads in the system. In addition, the kernel also maintains the traditional process table to keep track of processes. Operating Systems kernel provides system call to create and manage threads.
Examples: Windows XP/2000, Solaris, Linux, Tru64 UNIX, Mac OS X

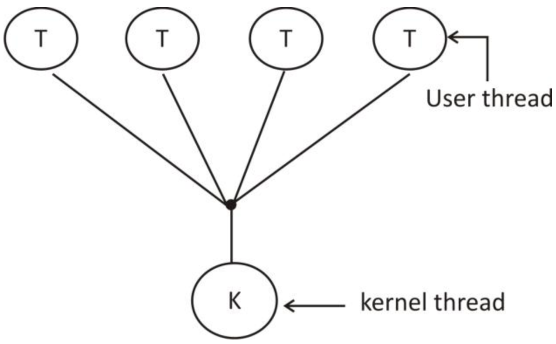
*Advantages:*
- Because kernel has full knowledge of all threads, Scheduler may decide to give more time to a process having large number of threads than process having small number of threads.
- Kernel-level threads are especially good for applications that frequently block.

*Disadvantages:*
- The kernel-level threads are slow and inefficient. For instance, threads operations are hundreds of times slower than that of user-level threads.
- Since kernel must manage and schedule threads as well as processes. It requires a full thread control block (TCB) for each thread to maintain information about threads. As a result, there is significant overhead and increased in kernel complexity.

## 2.5.3. Multithread Model

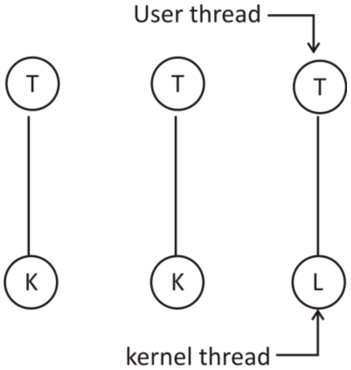
As seen above, there may be user threads, and kernel level threads. The way they can relate to each other in multi-thread process are as follows,

**1. Many-to-One Model:**
- In the many-to-one model, many user-level threads are all mapped onto a single kernel thread.
- Thread management is handled by the thread library in user space, which is efficient in nature.



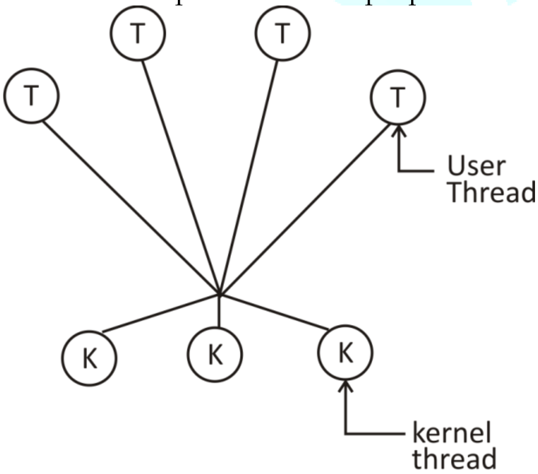

**2. One to One Model:**

- The one-to-one model creates a separate kernel thread to handle each and every user thread.
- Most implementations of this model place a limit on how many threads can be created.
- Linux and Windows from 95 to XP implement the one-to-one model for threads.




**3. Many to Many Model :**

- The many-to-many model multiplexes any number of user threads onto an equal or smaller number of kernel threads, combining the best features of the one-to-one and many-to-one models.
- Users can create any number of the threads.
- Blocking the kernel system calls does not block the entire process.
- Processes can be split across multiple processors.




**Difference between User-Level & Kernel-Level Thread:**

| User- Level Thread | Kernel -level Thread |
|---|---|
| User thread are implemented by user | Kernel thread implemented by operating system |
| User-level thread are faster to create and manage | Kernel-level threads are slower to create and manage |
| User-level thread is generic and can run on any operating system | Kernel –level thread is specific to the operating system |
| Multi-threaded application cannot take advantage of multiprocessing | Kernel routines themselves can be multithreaded |
| Context switch required no hardware support | Hardware support is needed |

### 2.5.4. Difference between Process and Thread:

| Process | Thread |
|---|---|
| Process is heavy weight or resource intensive. | Thread is light weight, taking lesser resources than a process. |
| Process switching needs interaction with operating system. | Thread switching does not need to interact with operating system. |
| In multiple processing environments, each process executes the same code but has its own memory and file resources. | All threads can share same set of open files, child processes. |
| Multiple processes without using threads use more resources. | Multiple threaded processes use fewer resources. |
| If one process is blocked, then no other process can execute until the first process is unblocked. | While one thread is blocked and waiting, a second thread in the same task can run. |
| In multiple processes each process operates independently of the others. | One thread can read, write or change another thread's data. |

### 2.6. Schedulers

Schedulers are special system software which handles process scheduling in various ways. Their main task is to select the jobs to be submitted into the system and to decide which process to run. Schedulers are of three types −

- **Long-Term Scheduler:** It is also called a **job scheduler**. A long-term scheduler determines which programs are admitted to the system for processing. It selects processes from the queue and loads them into memory for execution. Process loads into the memory for CPU scheduling. The primary objective of the job scheduler is to provide a balanced mix of jobs, such as I/O bound and processor bound. It also controls the degree of multiprogramming. If the degree of multiprogramming is stable, then the average rate of process creation must be equal to the average departure rate of processes leaving the system.
- **Short-Term Scheduler:** It is also called as **CPU scheduler**. Its main objective is to increase system performance in accordance with the chosen set of criteria. It is the change of ready state to running state of the process. CPU scheduler selects a process among the processes that are ready to execute and allocates CPU to one of them. Short-term schedulers, also known as dispatchers, make the decision of which process to execute next. Short-term schedulers are faster than long-term schedulers.
- **Medium-Term Scheduler:** This scheduler removes the processes from memory (and from active contention for the CPU), and thus reduces the degree of multiprogramming. At some later time, the process can be reintroduced into memory and its execution van be continued where it left off. This scheme is called **swapping**. The process is swapped out, and is later swapped in, by the medium-term scheduler. Swapping may be necessary to improve the process mix, or because a change in memory requirements has overcommitted available memory, requiring memory to be freed up. This complete process is descripted in the below diagram:

### 2.7. Dispatcher:

Another component involved in the CPU-scheduling function is the dispatcher. The dispatcher is the module that gives control of the CPU to the process selected by the short-term scheduler. This function involves the following:

- Switching context
- Switching to user mode
- Jumping to the proper location in the user program to restart that program

The dispatcher should be as fast as possible, since it is invoked during every process switch. The time it takes for the dispatcher to stop one process and start another running is known as the dispatch latency.

### 2.8. Scheduling Criteria

Many criteria exist for comparing and deciding on CPU scheduling algorithms. Which characteristics are used for comparison can make a substantial difference in which algorithm is judged to be best. The criteria include the following:

1. **CPU utilization:** We want to keep the CPU as busy as possible. Conceptually, CPU utilization can range from 0 to 100 percent. In a real system, it should range from 40 percent (for a lightly loaded system) to 90 percent (for a heavily used system).

2. **Throughput**: If the CPU is busy executing processes, then work is being done. One measure of work is the number of processes that are completed per time unit, called throughput. For long processes, this rate may be one process per hour; for short transactions, it may be 10 processes per second.

3. **Turnaround time**: From the point of view of a particular process, the important criterion is how long it takes to execute that process. The interval from the time of submission of a process to the time of completion is the turnaround time. Turnaround time is the sum of the periods spent waiting to get into memory, waiting in the ready queue, executing on the CPU, and doing I/O.

5. **Waiting time:** The CPU scheduling algorithm does not affect the amount of time during which a process executes or does I/O; it affects only the amount of time that a process spends waiting in the ready queue. Waiting time is the sum of the periods spent waiting in the ready queue.

6. **Response time:** In an interactive system, turnaround time may not be the best criterion. Often, a process can produce some output fairly early and can continue computing new results while previous results are being output to the user. Thus, another measure is the time from the submission of a request until the first response is produced. This measure, called response time, is the time it takes to start responding, not the time it takes to output the response. The turnaround time is generally limited by the speed of the output device.

It is needed to maximize the CPU utilization and the throughput and also to minimize turnaround time, waiting time, and response time. In most cases, we optimize the average measure.

## 2.9. Scheduling Algorithms:

**1. First Come First Serve Scheduling**
The simplest of all scheduling algorithm is the first come first serve algorithm also known as FCFS. It implies that, the process that request for CPU firsts, gets the CPU allocated to it firsts. It can be easily implemented using a Queue (FIFO).
Average waiting time is usually high in case of FCFS.

**Example 1:** Let processes come in order P1, P2, P3 and request for CPU

| process | Burst time |
|---|---|
| P1 | 24 |
| P2 | 3 |
| P3 | 3 |

If process arrive in the order P1, P2, P3 and are server in FCFS order, the result is:

| P1 | P2 | P3 |
|---|---|---|

0 24 27 30

The waiting time is 0 milliseconds for process P1, 24 milliseconds for process P2 and 27 milliseconds for process P3.
Thus, the average waiting time is - (0+24+27)/3= 17 milliseconds
If the processes arrive in the order P2, P3, P1, however, the result will be shown:

| P2 | P3 | P1 |
|---|---|---|

0 3 6 30

The average waiting time is now (6+0+3)/3=3 milliseconds
This reduction is substantial. Thus, the average waiting times under FCFS policy is generally not minimal and varies substantially if the process's CPU burst times vary greatly.

The FCFS scheduling algorithm is nonpreemptive. Once the CPU has been allocated for a process that process keeps the CPU until it releases the CPU either by terminating or by requesting I/O. The FCFS algorithm is thus particularly troublesome for time-sharing systems, where it is important that each user get a share of the CPU at regular intervals. It would be disastrous to allow one process to keep the CPU for an extended period.

**2. Shortest Job First Scheduling**

This algorithm associates with each process the length of the process's next CPU burst. When the CPU is available, it is assigned to the process that has the smallest next CPU burst. If the next CPU bursts of two processes are the same, FCFS scheduling is used to break the tie. Note that a more appropriate term for this scheduling method would be the shortest-next-CPU-burst algorithm, because scheduling depends on the length of the next CPU burst of a process, rather than its total length.

> Shortest Job First Scheduling algorithm is the best approach to minimize waiting time.

But, the real difficulty with the SJF algorithm knows the length of the next CPU request. Although the SJF algorithm is optimal, it cannot be implemented at the level of short-term CPU scheduling. There is no way to know the length of the next CPU burst, but we can predict it; i.e. we may not know the length of the next CPU burst, but we may be able to predict its value. We expect that the next CPU burst will be similar in length to the previous ones. Thus, by computing an approximation of the length of the next CPU burst, we can pick the process with the shortest predicted CPU burst.
The SJF algorithm may be preemptive or non-preemptive depending on the kernel, or CPU requirement.

**Example:** consider the following set of processes, with the length of the CPU burst given in milliseconds:

| Process | Burst time |
|---|---|
| P1 | 6 |
| P2 | 8 |
| P3 | 7 |
| P4 | 3 |

Using SJF scheduling, we would schedule these process:

| P4 | P1 | P3 | P2 | |
|---|---|---|---|---|
| 0 | 3 | 9 | 16 | 24 |

The waiting time is 3 milliseconds for process P1, 16 milliseconds for process P2, 9 milliseconds for process P3, and 0 milliseconds for process P4.
Thus the average waiting time is- (3+16+9+0)/4= 7 milliseconds
The average waiting time of SJF is less in comparison to FCFS.

**3.Priority Scheduling**
In this CPU scheduling algorithm, a priority is associated with each process, and the CPU is allocated to the process with the highest priority. Equal-priority processes are scheduled in FCFS order. We can think of priority algorithm as simply a SJF algorithm where the priority (p) is the inverse of the (predicted) next CPU burst. The larger the CPU burst, the lower the priority, and vice versa. Thus only difference is we decide on scheduling in terms of high priority and low priority. Priorities are generally indicated by some fixed range of numbers, such as 0 to 7 or 0 to 4095.

**Example:** consider the set of process P1, P2, P3, P4, P5.

| Process | Burst Time | Priority |
|---|---|---|
| P1 | 10 | 3 |
| P2 | 1 | 1 |
| P3 | 2 | 4 |
| P4 | 1 | 5 |
| P5 | 5 | 2 |

Using priority scheduling we would schedule these process:

| P2 | P5 | P1 | P3 | P4 | |
|---|---|---|---|---|---|
| 0 | 1 | 6 | 16 | 18 | 19 |

The average waiting time is-(6+0+16+18+1)/5= 8.2 milliseconds

Priority scheduling can be either preemptive or non-preemptive. When a process arrives at the ready queue, its priority is compared with the priority of the currently running process. A preemptive priority scheduling algorithm will preempt the CPU if the priority of the newly arrived process is higher than the priority of the current running process.

A non preemptive priority scheduling algorithm will simply put the new process at the head of the ready queue. A major problem with priority scheduling algorithms is of starvation. A process that is ready to run but waiting for the CPU can be considered blocked. A priority scheduling algorithm can leave some low-priority processes waiting indefinitely. In a heavily loaded computer system, a steady stream of higher-priority processes can prevent a low-priority process from ever getting the CPU. This is highly undesirable, and this problem can be solved by using the concept of Aging of a process, which means that as the process waits in ready queue, it's priority gradually keeps on increasing. This ensures that even low priority process get CPU allocated to it, after some waiting time.

**4.Round Robin Scheduling**

The round-robin CPU scheduling algorithm is basically a preemptive scheduling algorithm designed for time-sharing systems. One unit of time is called a time slice. Duration of a time slice may range between 10 milliseconds to about 100 milliseconds. The CPU scheduler allocates to each process in the ready queue one time slice at a time in a round-robin fashion. Hence the name round-robin.

The ready queue in this case is a FIFO queue with new processes joining the tail of the queue. The CPU scheduler picks processes from the head of the queue for allocating the CPU. The first process at the head of the queue gets to execute on the CPU at the start of the current time slice and is deleted from the ready queue. The process allocated the CPU may have the current CPU burst either equal to the time slice or smaller than the time slice or greater than the time slice. In the first two cases, the current process will release the CPU on its own and there by the next process in the ready queue will be allocated the CPU for the next time slice. In the third case, the current process is preempted, stops executing, goes back and joins the ready queue at the tail there by making way for the next process.

**Example:**

| Process | Burst Time |
|---|---|
| P1 | 24 |
| P2 | 3 |
| P3 | 3 |

If we use a time quantum of 4 milliseconds, then process P1 gets the first 4 milliseconds. since It require another 20 milliseconds, it is preempted after the first time quantum, the CPU is given to the next process in the queue, process P2. Process p2 does not need 4 milliseconds, so it quite before its time quantum expires. The CPU is given to the next process, Process P3. Once each process has received 1 time quantum, the CPU is returned to process P1 for an additional time quantum.

| P1 | P2 | P3 | P1 | P1 | P1 | P1 | P1 |
|---|---|---|---|---|---|---|---|

0 4 7 10 14 18 22 26 30

The average waiting time for the above schedule. P1 waits for 6 milliseconds (10-4), P2 waits for 4 milliseconds, and P3 waits for 7 milliseconds.

Thus the average waiting time is - (6+4+7)/3 = 5.66 milliseconds

If there are n. processes in the ready queue and the time quantum is q, then each process gets 1/n of the CPU time in chunks of at most q time units. Each process must wait no longer than (n - 1) x q time units until its next time quantum. The performance of the RR algorithm is very much dependent on the length of the time slice. If the duration of the time slice is indefinitely large then the RR algorithm is the same as FCFS algorithm. If the time slice is too small, then the performance of the algorithm deteriorates because of the effect of frequent context switching.

**5. MULTILEVEL QUEUE SCHEDULING:**

Processes can be classified into groups. For example, interactive processes, system processes, batch processes, student processes and so on. Processes belonging to a group have a specified priority. This algorithm partitions the ready queue into as many separate queues as there are groups. Hence the name multilevel queue. Based on certain properties is process is assigned to one of the ready queues. Each queue can have its own scheduling algorithm like FCFS or RR. For example, Queue for interactive processes could be scheduled using RR algorithm where queue for batch processes may use FCFS algorithm. An illustration of multilevel queues is shown below

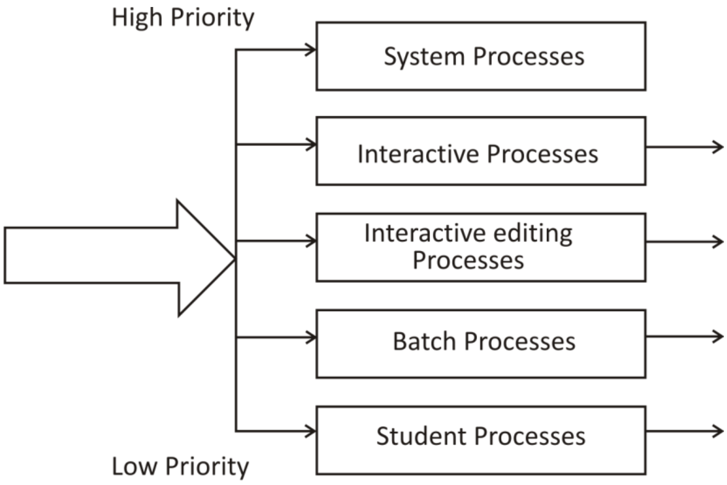



Figure-Multilevel Priority Queue Scheduling

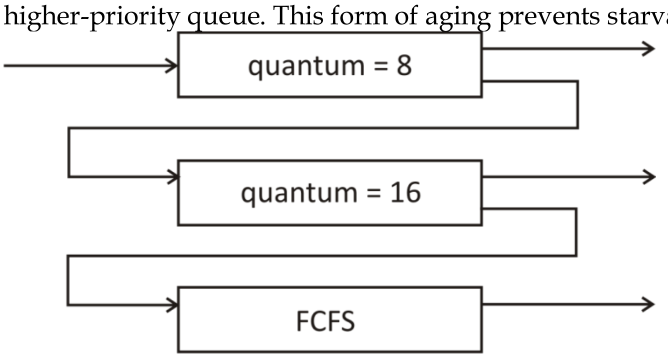
Queues themselves have priorities. Each queue has absolute priority over low priority queues that is a process in a queue with lower priority will not be executed until all processes in a queue with higher priority have finished executing.

This priority on the queues themselves may lead to starvation. To overcome this problem, time slices may be assigned to queues when each queue gets some amount of CPU time. The duration of the time slices may be different for queues depending on the priority of the queues.

**6. MULTILEVEL FEEDBACK QUEUE SCHEDULING**

In a multilevel queue-scheduling algorithm, processes are permanently assigned to a queue on entry to the system. Processes do not move between queues. This setup has the advantage of low scheduling overhead, but the disadvantage of being inflexible.

Multilevel feedback queue scheduling, however, allows a process to move between queues. The idea is to separate processes with different CPU-burst characteristics. If a process uses too much CPU time, it will be moved to a lower-priority queue. Similarly, a process that waits too long in a lower-priority queue may be moved to a higher-priority queue. This form of aging prevents starvation.




**An example of a multilevel feedback queue can be seen in the below figure.**

In general, a multilevel feedback queue scheduler is defined by the following parameters:

- The number of queues.
- The scheduling algorithm for each queue.
- The method used to determine when to upgrade a process to a higher-priority queue.
- The method used to determine when to demote a process to a lower-priority queue.
- The method used to determine which queue a process will enter when that process needs service.

The definition of a multilevel feedback queue scheduler makes it the most general CPU-scheduling algorithm. It can be configured to match a specific system under design. Unfortunately, it also requires some means of selecting values for all the parameters to define the best scheduler. Although a multilevel feedback queue is the **most general scheme**, it is also the **most complex**.

## 2.9.Process synchronization:

Process Synchronization means sharing system resources by processes in a such a way that, Concurrent access to shared data is handled thereby minimizing the chance of inconsistent data. Maintaining data consistency demands mechanisms to ensure synchronized execution of cooperating processes.

**Race Condition**

When several processes access and manipulate the same data concurrently, the outcome of the execution depends on the particular order in which the access takes place, is called race condition.

Example – Let's take a common variable counter which is shared by 2 processes P1 and P2. At any instance, value of counter is 7.

Also counter++ may be implemented in machine language as

register1 = counter
register1 = register1 + 1
counter= register1.

counter-- can be implemented in the same way.

Now P1 executes counter++ and at the same time P2 executes counter--.

The concurrent execution of both statements may be sequentially implemented as

register1 = counter
register1 = register1 + 1
register2 = counter
register2= register2 – 1
counter = register1
counter = register2

Thus depending on which statement gets executed first the value of counter, after the execution of counter++ and counter-- simultaneously, is 6, 7 or 8 i.e. unpredictable. Thus processes race to modify the shared variable and this condition is called Race Condition.

Process Synchronization was introduced to handle problems that arose while multiple process executions. Some of the problems are discussed below.

**1. Critical Section**

A critical section is a piece of code that accesses a shared resource (data structure or device), writing a file or updating a table that must not be concurrently accessed by more than one thread of execution. Thus, execution of critical sections by the processes is mutually exclusive in time.

General structure of a process is:

```
do
{
Entry section
Critical section
Exit section
Remainder section
}while(TRUE);
```

Each process must request permission to enter its critical section, which is done in entry section. The critical section problem is to design a protocol that the processes can use to co-operate. A solution to the critical section problem must satisfy the three requirements:

1) Mutual Exclusion: If a process Pi is executing in its critical section, then no other process can enter it.
2) Progress: If no process is executing in its critical section and some processes wish to enter, then only those process which are not in remainder section can enter and the selection cannot be postponed indefinitely.
3) Bounded Waiting: After a process makes a request to enter critical section, then there is a bound on the number of times other processes are allowed to enter critical section before this process.

**Software solution for critical section problem:**

**1. Two – process solution**

The processes share two variables:

```
booleanflag[2];
int turn;
Initially flag[0]=flag[1]= false and turn can be 0 or 1;
do
{
flag[i] = true;
turn = j;
while (flag[j] && turn == j);
```

```
Critical section
flag[i] = false;
Remainder section
}while(TRUE);
```

**2. Multiple process solution**

This algorithm is also called Bakery algorithm. The common data structures for n processes is

```
boolean choosing: array [n];
int: array [n];
```

Initially, these data structures are initialized to false and 0, respectively. For convenience, the following notation is used in the solution:

(a, b) < (c, d) if a < c or if a== c and b < d.

max(a0,a1..., an-1) is a number, k, such that k >= a ; for i = 0,..., n — 1

```
do
{
choosing[i] := true;
number[i] := max(number[Q], number^}, ..,, number[n — 1]) + 1;
choosing[i] := false;
for(j=0; j<n; j++)
{
while ( choosing [j] );
while ( (number[j]!=0) && (number[j], j)< (number[i], i) );
}
critical section
number[i] :=0;
remainder section
}while(1);
```

**Hardware solution for critical section problem**

Features of the hardware can make the programming task easier and improve system efficiency. The critical-section problem could be solved simply in a uniprocessor environment if we could disallow interrupts to occur while a shared variable is being modified. Unfortunately, this solution is not feasible in a multiprocessor environment.

Many machines provide special hardware instructions that allow us either to test and modify the content of a word, or to swap the contents of two words, atomically.

The most common instructions provided by hardware are TestAndSet instruction and the Swap instruction.

The *Test-and-Set* instruction can be defined as

```
booleanTestAndSet (boolean&target) {
booleanrv = target;
target = true;
return rv;
}
```

The solution which satisfies all requirements of critical section problem is the following algorithm. Following data structures are common with processes.

```
boolean waiting[n];
boolean lock;
```

In addition each process has a local boolean variable key.

Initially these data structures have value set to false.

```
do {
waiting[i] = true;
key = true;
while (waiting [i] && key)
key = TestAndSet (lock);
waiting[i] = false;
Critical section
j = (i+1) % n;
while ((j!=i) && !waiting[j])
j = (j+1) % n;
if ( j == i)
```

```
lock = false;
else
waiting[j] = false;
Remainder section
} while(1);
```

**Semaphores**

Semaphores are used for synchronization mechanisms. A semaphore is an integer variable that, apart from initialization, is accessed only through two standard atomic operations: wait and signal.

Definition of wait and signal operation. Let S be a semaphore.

```
wait(S) {
While (S <= 0); // no-operation
S-- ;
}
signal(S) {
S++;
}
```

Using semaphores, implementing mutual exclusion becomes very easier.

```
do {
wait (mutex)
Critical section
signal (mutex)
Remainder section
} while (1);
```

In this solution, if a process is in critical section, then any other process trying to get into CS must loop continuously in the entry code. This continuous looping is called Busy Waiting and it wastes CPU cycles. This type of semaphore is called Spinlock because the process spins while waiting for the lock.

Spinlocks are useful in multiprocessor systems where context switch is highly expensive and if locks are to be held for smaller time than spinning is beneficial.

**Binary Semaphores**

A binary semaphore is a semaphore with an integer value that can range only between 0 and 1. A binary semaphore can be simpler to implement rather than a counting semaphore, depending on the underlying hardware.

**Bounded buffer problem**

Bounded buffer problem is a classical synchronization problem for co-operating sequential processes. A perfect example of co-operating sequential processes is producer and consumer processes. Producer produces an item and the consumer consumes. In bounded buffer environment producer cannot produces items more than the size of buffer and consumer cannot consume items more than buffer size.

A shared memory solution to this problem exist which makes use of a shared variable counter initialized to 0. Producer can produce at max Buffer-1 items.

Code for Producer process

```
while (1) {
/* produce a newItem*/
while ( counter == Buffer_Size);
buffer [in] = newItem;
in = (in +1) % Buffer_Size;
counter ++;
}
```

Code for Consumer Process

```
while (1) {
while ( counter == 0);
item = buffer [out];
out = (out +1) % Buffer_Size;
counter --;
/* consume item */
}
```

Here buffer and counter variable is shared between the 2 processes and in and out are local variables for producer and consumer respectively.
Although both producer and consumer codes are correct separately but on concurrent execution may not produce correct result due to race condition as no proper synchronization mechanism exists.
Another solution to bounded buffer problem where there is no race condition exists. This solution makes use of semaphores. In this problem it is assumed that there exist n buffers each of size 1 instead of a single buffer of size n. Also, three semaphores are used. One is **mutex** used for providing mutual exclusion in critical section and it is initialized to 1.
The **empty** and **full** semaphores count the number of empty and full buffers, and are initialized to n and 0 respectively.
The structure of producer process is

```
do {
produce an item
wait(empty);
wait(mutex);
add item to buffer
signal(mutex);
signaltfull);
}while(1);
```

The structure of consumer process is

```
do {
wait(full);
wait(mutex);
remove item from buffer
signal(mutex);
signal(empty);
consume the item
}while(1);
```

## 2.10.Deadlock

In a multiprogramming environment, several processes may compete for a finite number of resources. A process requests resources; if the resources are not available at that time, the process enters a wait state. It may happen that waiting processes will never again change state, because the resources they have requested are held by other waiting processes. This situation is called a deadlock.
A system consists of finite number of resources to be distributed among a number of competing resources. The resources are partitioned into types each of which may have several instances.
A process must request a resource before using it and must release it after using it. A process can request as many resources for carrying out its task; obviously this number has to be less than total available resources. A system table records whether each resource is free or allocated, and, if a resource is allocated, to which process. If a process requests a resource that is currently allocated to another process, it can be added to a queue of processes waiting for this resource.
A set of processes is in a deadlock state when every process in the set is waiting for an event that can be caused by only another process in the set. The events with which we are mainly concerned here are resource acquisition and release. For example, consider a system with three tape drives. Suppose that there are three processes, each holding one of these tape drives. If each process now requests another tape drive, the three processes will be in a deadlock state. Each is waiting for the event "tape drive is released," which can be caused only by one of the other waiting processes. This example illustrates a deadlock involving processes competing for the same resource type. Deadlocks may also involve different resource types.

### *2.10.1. Necessary condition for deadlocks*

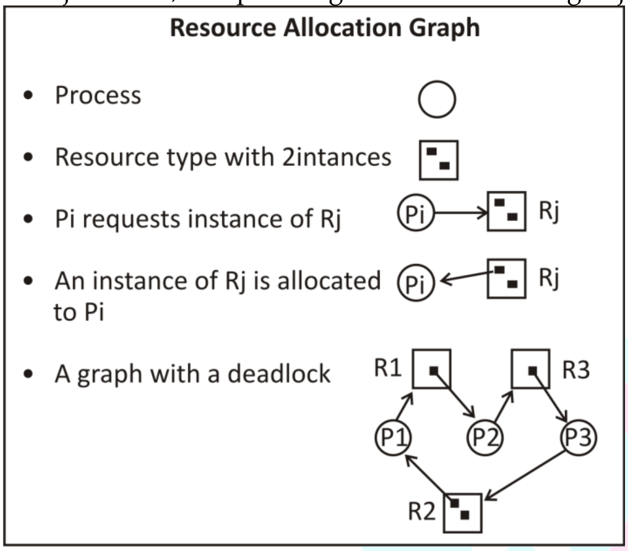
A deadlock situation can arise if the following four conditions hold simultaneously in a system.

1. **Mutual exclusion:** At least one resource must be held in a non-sharable mode; that is, only one process at a time can use the resource. If another process requests that resource, the requesting process must be delayed until the resource has been released.
2. **Hold and wait:** A process must be holding at least one resource and waiting to acquire additional resources that are currently being held by other processes.

3. **No preemption:** Resources cannot be preempted; that is, a resource can be released only voluntarily by the process holding it, after that process has completed its task.
4. **Circular wait:** There must exist a set {P1, P2, ..., Pn} of waiting processes such that P0 is waiting for a resource that is held by P1, P1 is waiting for a resource that is held by P2,..., Pn-1 is waiting for a resource that is held by Pn, and Pn is waiting for a resource that is held by P0.

### *2.10.2. Resource Allocation Graph*

A deadlock can be described more precisely in terms of a directed graph called a system resource allocation graph. In this graph the vertices represent two different types of nodes, one for active processes and other for resources. A directed edge from process Pi to resource type Rj denoted by Pi —>Rj signifies that process Pi requested an instance of resource type Rj and is currently waiting for that resource. A directed edge from resource type Rj to process Pi denoted by Rj —> Pi signifies that an instance of resource type Rj has been allocated to process Pi. A directed edge Pi —>Rj is called, a request edge and a directed edge Rj —> Pi is called an assignment edge.




When process P, requests an instance of resource type R, a request edge is inserted in the resource-allocation graph. When this request can be fulfilled, the request edge is instantaneously transformed to an assignment edge. When the process no longer needs access to the resource it releases the resource, and as a result the assignment edge is deleted.

If the graph contains no cycles, then no process in the system is deadlocked. If, on the other hand, the graph contains a cycle, then a deadlock may exist. If the cycle involves only a set of resource types, each of which has only a single instance, then a deadlock has occurred. Each process involved in the cycle is deadlocked. In this case, a cycle in the graph is both a necessary and a sufficient condition for existence of deadlock.

If each resource type has several instances, then a cycle does not necessarily imply that a deadlock occurred. In this case, a cycle in the graph is a necessary but not a sufficient condition for the existence of deadlock.
If a resource-allocation graph does not have a cycle, then the system is not in a deadlock state. On the other hand, if there is a cycle, then the system may or may not be in a deadlock state.

### *2.10.3. Methods for handling deadlocks*

Principally, there are three different methods for dealing with the deadlock problem:
- We can use a protocol to ensure that the system will never enter a deadlock state. To ensure this, the system can use either deadlock prevention or deadlock avoidance scheme. Deadlock prevention is a set of methods for ensuring that at least one of the necessary conditions of deadlock cannot hold. Deadlock avoidance, on the other hand, requires that the operating system be given in advance additional information concerning which resources a process will request and use during its lifetime. With this additional knowledge, we can decide for each request whether or not the process should wait.
- We can allow the system to enter a deadlock state and then recover. In this environment, the system can provide an algorithm that examines the state of the system to determine whether a deadlock has occurred, and an algorithm to recover (if a deadlock has indeed occurred) from the deadlock.
- We can ignore the deadlock problem all together, and pretend that they never occur in the system. This solution is the one used by most operating systems, including UNIX. Although this method does not seem to be a viable

approach to the deadlock problem, it is nevertheless used in some operating systems. In many systems, deadlocks occur rarely (say, once per year); thus, it is cheaper to use this method instead of the costly deadlock prevention, deadlock avoidance, or deadlock detection and recovery methods that must be used constantly.

### *2.10.4. Deadlock Prevention*

By ensuring that at least one of the necessary conditions cannot hold, deadlock can be prevented.
**1) Mutual Exclusion –** This condition must hold for non-sharable resources like printer, which cannot be used simultaneously by several processes. Thus, deadlocks cannot be prevented b denying mutual exclusion condition. Some resources are intrinsically non-sharable.

**2) Hold and Wait –** To ensure that the hold-and-wait condition never occurs in the system, we must guarantee that, whenever a process requests a resource, it does not hold any other resources. This can be done in 2 ways. First, a process should ask for all the resources it needs before it starts executing. Second, if it asks for a resource in between, it must release all the resources previously held by it and then ask again for all the resources it needs.
These protocols have 2 main disadvantages –

- Resource utilization is low, since many resources may be held for long time, even when required for less time.
- Starvation is possible. A process needing many resources may not get a chance to execute as getting many resources may not be possible.

**3) No Preemption –** To ensure this condition does not hold, following protocol is followed. If a process requests some resources, check whether they are available. If they are, allocate them. If they are not available, check whether they are allocated to some other process that is waiting for additional resources. If so, preempt the desired resources from the waiting process and allocate them to the requesting process. If the resources are not either available or held by a waiting process, the requesting process must wait. While it is waiting, some of its resources may be preempted, but only if another process requests them. This protocol cannot be applied to resources like printers and tape drives.

**4) Circular Wait –** To ensure that the circular-wait condition never holds, impose a total ordering of all resource types, and ensure that each process requests resources in an increasing order of enumeration. Example - assign to each resource type a unique integer number, which allows to compare two resources and to determine whether one precedes another in our ordering. Then the following protocol can be used to prevent deadlocks: Each process can request resources only in an increasing order of enumeration. That is, a process can initially request any number of instances of a resource type, say Ri. After that, the process can request instances of resource type Rj if and only if F(Rj) > F(Ri). If several instances of the same resource type are needed, a single request for all of them must be issued.
Let the set of processes involved in the circular wait be {P0, P1, … ,Pn}, where Pi is waiting for a resource Rj, which is held by process Pi+1. (Modulo arithmetic is used on the indexes, so that Pn is waiting for a resource Rn held by P0.) Then, since process Pi+i is holding resource Ri, while requesting resource Ri+1, we must have F(Ri) < F(Ri+1), for all i. But this condition means that F(R0) < F(R1) < ... < F(Rn) < F(R0). By transitivity, F(R0) < F(R0), which is impossible. Therefore, there can be no circular wait.

### *2.10.5. Deadlock Avoidance*

A deadlock-avoidance algorithm dynamically examines the resource-allocation state to ensure that there can never be a circular-wait condition. The resource-allocation state is defined by the number of available and allocated resources, and the maximum demands of the processes.

**Safe state -** A state is safe if the system can allocate resources to each process (up to its maximum) in some order and still avoid a deadlock. A sequence of processes <P1, P2, ..., Pn> is a safe sequence for the current allocation state if, for each Pi, the resources that Pi can still request can be satisfied by the currently available resources plus the resources held by all the Pj with j<i.

A safe state is not a deadlock state. Conversely, a deadlock state is an unsafe state. However, not all unsafe states are deadlocks. An unsafe state may lead to a deadlock. As long as the state is safe, the operating system can avoid unsafe (and deadlock) state. In an unsafe state, the operating system cannot prevent process from requesting resources in such a way that a deadlock occurs. The behavior of the process controls unsafe states.

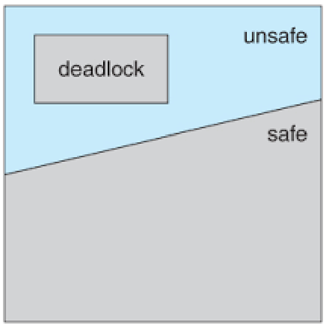



**Figure: safe, unsafe, and deadlocked state spaces**

**Banker's Algorithm**
The resource-allocation graph algorithm is not applicable to a resource allocation system with multiple instances of each resource type. The deadlock avoidance algorithm that we describe next is applicable to such a system but is less efficient than the resource allocation graph scheme. This algorithm is commonly known as the banker's algorithm.
When a new process enters the system, it must declare the maximum number of instances of each resource type that it may need. This number may not exceed the total number of resources in the system. When a user requests a set of resources the system must determine the allocation of these resources will leave the system in a safe state. If it will, the resources are allocated; otherwise the process must wait until some other process releases enough resources.
Several data structures must be maintained to implement the banker's algorithm. These data structures encode the state of the resource-allocation system. Let $n$ be the number of processes in the system and $m$ be the number of resource types. We need the following data structures:

**Available**: A vector of length m indicates the number of available resources of each type.
If Available[j] = k, there are k instances of resource type Rj available.
**Max**: An n x m matrix defines the maximum demand of each process. If Max [i, j] = k, then Pi may request at most k instances of resource type Rj.
**Allocation**: An n x m matrix defines the number of resources of each type currently allocated to each process. If Allocation [i, j] = k, then process Pi is currently allocated k instances of resource type Rj.
**Need**: An n x m matrix indicates the remaining resource need of each process. If Need[i,j] = k, then Pi may need k more instances of resource type Rj to complete its task. Note that Need [i, j] = Max [i, j] – Allocation [i, j].

### *2.10.6 Deadlock Detection*

These algorithms are needed when the system does not have mechanisms to prevent deadlock or avoid it. Thus, chances of deadlocks occurring in the system are increased and system ten employs algorithms to detect the deadlock and then recover from it. This detection and recovery scheme requires overhead that includes not only the run-time costs of maintaining the necessary information and executing algorithm but also potential losses inherent in recovering from the deadlock.

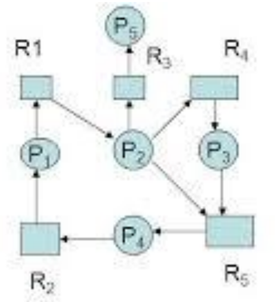
**For single instance of resources**
When all the resources existing in the system have single instance, then a variant of resource allocation graph called wait-for graph can be used to detect deadlocks. A wait for graph is obtained by removing the nodes of type resource and collapsing the appropriate edges from the resource-allocation graph. More precisely, an edge from Pi to Pj in a wait-for graph implies that process Pi is waiting for process Pj to release a resource that Pi needs. An edge Pi —>Pj exists in a wait-for graph if and only if the corresponding resource allocation graph contains two edges Pi —>Rq and Rq —>Pj for some resource Rq.

For several instances of resource, an algorithm similar to banker's algorithm is used to detect deadlock.
**Available**: A vector of length m indicates the number of available resources of each type.

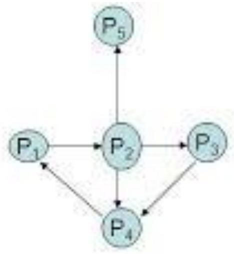
**Allocation**: An n x m matrix defines the number of resources of each type currently allocated to each process.

**Request**: An n x m matrix indicates the current request of each process. If Request [i, j]=k then process Pi is requesting k more instances of resource of type Rj.




Resource allocation graph




Corresponding wait for Graph

The algorithm works as follows:
1. Let Work and Finish be vectors of length m and n, respectively. Initialize
Work=Available. For i = l, 2… n, if $Allocation_i \neq 0$, then Finish[i] = false; otherwise,
Finish[i] = true.
2. Find an index i such that both
a. Finish[i]= = false.
b. $Request_i \leq$ Work.
If no such i exists, go to step 4.
3. Work = Work + $Allocation_i$
Finish[i] = true
go to step 2.
4. If Finish[i] == false, for some i, $0 \leq i < n$, then the system is in a deadlock state.
Moreover, if Finish[i]= =false, then process Pi is deadlocked.
It also requires m x $n^2$ operations to detect whether the system is in a deadlocked state.

### *2.10.7 Recovery from deadlock*

When a detection algorithm determines that a deadlock exist the system must recover from the deadlock. One solution is simply to abort one or more processes to break the circular wait. The second option is to preempt some resources from one or more of the deadlocked processes.

**(a) Process Termination**
To eliminate deadlocks by aborting a process, we use one of two methods. In both methods, the system reclaims all resources allocated to the terminated processes.

- **Abort all deadlocked processes**: This method clearly will break the deadlock cycle, but at a great expense, since these processes may have computed for a long time, and the results of these partial computations must be discarded, and probably must be recomputed later.
- **Abort one process at a time until the deadlock cycle is eliminated**: This method incurs considerable overhead, since, after each process is aborted, a deadlock-detection algorithm must be invoked to determine whether any processes are still deadlocked. Aborting a process may not be easy. If the process was in the midst of updating a file, terminating it in the middle will leave that file in an incorrect state. Many factors may determine which process is chosen for abortion, including:

1. What the priority of the process is?
2. How long the process has computed, and how much longer the process will compute before completing its designated task?
3. How many and what type of resources the process has used (for example, whether the resources are simple to preempt)?
4. How many more resources the process needs in order to complete?
5. How many processes will need to be terminated?
6. Whether the process is interactive or batch?

**(b) Resource Preemption**
To eliminate deadlocks using resource preemption, we successively preempt some resources from processes and give these resources to other processes until the deadlock cycle is broken.
If preemption is required to deal with deadlocks, then three issues need to be addressed:

1. **Selecting a victim**: As in process termination, we must determine the order of preemption to minimize cost. Cost factors may include such parameters as the number of resources a deadlock process is holding, and the amount of time a deadlocked process has thus far consumed during its execution.

2. **Rollback:** If we preempt a resource from a process, what should be done with that process? Clearly, it cannot continue with its normal execution; it is missing some needed resource. We must roll back the process to some safe state, and restart it from that state.
3. **Starvation**: How do we ensure that starvation will not occur? That is, how can we guarantee that resources will not always be preempted from the same process? In a system where victim selection is based primarily on cost factors, it may happen that the same process is always picked as a victim. As a result, this process never completes its designated task, a starvation situation that needs to be dealt with in any practical system. Clearly, we must ensure that a process can be picked as a victim only a (small) finite number of times.

## 3. Memory Management:

Main Memory refers to a physical memory that is the internal memory to the computer. The word main is used to distinguish it from external mass storage devices such as disk drives. Main memory is also known as RAM. The computer is able to change only data that is in main memory. Therefore, every program we execute and every file we access must be copied from a storage device into main memory.

### 3.1. Memory Organization

The main purpose of a computer system is to execute programs. These programs, together with the data they access, must be in main memory (at least partially) during execution. To improve both the utilization of CPU and the speed of its response to its users, the computer must keep several processes in memory. There are many different memory-management schemes. Since main memory is usually too small to accommodate all the data and programs permanently, the computer system must provide secondary storage to back-up main memory.

**Address mapping and translation**
Most systems allow a user process to reside in any part of the physical memory. Although the address space of the computer starts at 00000, the first address of the user process does not need to be 00000. In most cases, a user program goes through several steps before being executed. Addresses may be represented in different ways during these steps. Addresses in the source program are generally symbolic. A compiler typically binds these symbolic addresses to relocatable addresses. The linkage editor or loader will in turn bind these relocatable addresses to absolute addresses. Each binding is a mapping from one address space to another. The addresses are translated using the relocation registers.
As in the diagram, the binding of instructions and data to memory can be done at any step along the way:

- Compile time: If it is known at compile time where the process will reside in memory, then absolute code can be generated. For example, if it is known in advance that a user process resides starting at location R, then the generated compiler code will start at that location and extend up from there. If, at some later time, the starting location changes, then it will be necessary to recompile this code.
- Load time: If it is not known at compile time where the process will reside in memory, then the compiler must generate relocatable code. In this case, final binding is delayed until load time. If the starting address changes, we need only to reload the user code to incorporate this changed value.
- Execution time: If the process can be moved during its execution from one memory segment to another, then binding must be delayed until run rime.

**Logical and physical addresses**
An address generated by the CPU is commonly referred to as a logical address, whereas an address seen by the memory unit, i.e., the one loaded into the memory address register of the memory, is called physical address. The logical address space may also be referred to as virtual address space. Now, since as user we only see the logical address, but the system works with physical addresses, some mapping must be done to allow correct working. This mapping is done by the memory-management unit (MMU).

**Base and Limit Registers**
Base and limit registers are special registers in the system which are used by MMU to convert logical addresses to physical addresses. Base register keep the starting address of physical memory where the program resides where as the limit register keeps the length of the program. These registers allow protecting the OS code and data as well as a simpler model of executable (user) code. The purpose of having base-limit register pairs is to address specific sub regions of the memory unit.

## 3.2.Memory management unit

A memory management unit (MMU) is a computer hardware component responsible for handling accesses to memory requested by the central processing unit (CPU). Its functions include translation of virtual addresses to physical addresses, memory protection etc.

The MMU uses a base register also called relocation register. The value in this register is added to the logical address, to get the physical address. That means, depending on the value of the relocation register, the logical address is relocated to its actual physical location.

## 3.3. Memory Partitioning

The main memory must accommodate both the operating system and the various user processes. The memory is usually divided into two partitions, one for the resident operating system, and one for the user processes. If single partition scheme is used, then we need to protect the operating-system code and data from changes by the user processes and also protect the user processes from one another. This can be done by loading the relocation and limit registers with the correct values as part of the context switch.

For multiple partitions, one of the simplest methods is to divide memory into several fixed-size partitions. Each partition may contain exactly one process. Thus, the degree of multiprogramming is bound by the number of partitions. When a partition is free, a process is selected from the input queue and is loaded into the free partition. When the process terminates, the partition becomes available for another process. Other method is to consider all memory as one large block, a hole. When a process arrives and needs memory, we search for a hole large enough for this process. If we find one, we allocate only as much memory as is needed, keeping the rest available to satisfy future requests. Thus after some time, there is at any time a set of holes, of various sizes, scattered throughout memory. The operating system keeps a table indicating which parts of memory are available and which are occupied.

The set of holes is searched to determine which hole is best to allocate. First-fit, best-fit, and worst-fit are the most common strategies used to select a free hole from the set of available holes.

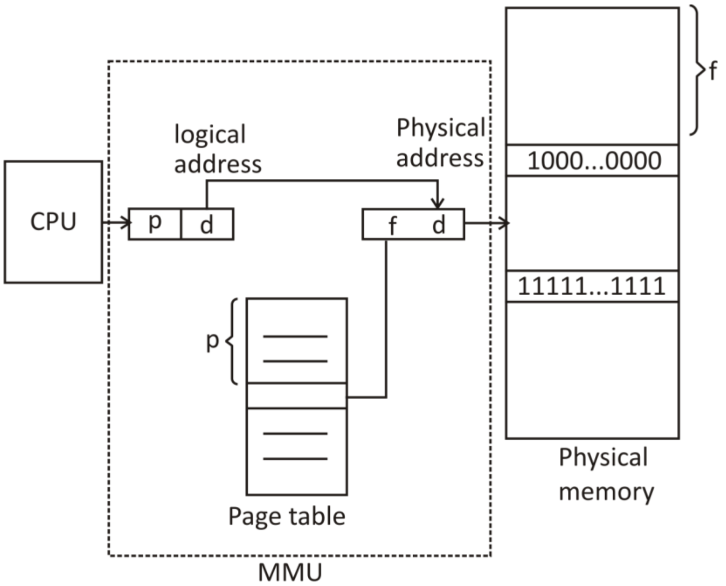
**Fragmentation**
By partitioning the memory and allocating them using various strategies, the memory is fragmented. Fragmentation can be internal and external. When fixed size partitions are used, the process may or may not be the same size as the partition. Thus, some amount of memory is left unused leading to internal fragmentation.
When variable size partitions are used, problem of external fragmentation is created as a lot of memory may still be available but since it is not contiguous, it cannot be allotted to a process. The solution to the external fragmentation problem is **compaction**. The goal is to shuffle the memory contents to place all free memory together in one large block.

## 3.4. Memory management techniques

Various memory management techniques exist such as partitioning memory into various fixed sized partitions or dynamic memory allocation. Also, overlays and swapping is also used as part of memory management. But the two most important and widely used strategies are paging and segmentation.
**Paging**

Paging is a memory management scheme that permits the physical address space of a process to be non-contiguous. The physical memory is broken into fixed-sized blocks called frames. Logical memory is also broken into blocks of the same size called pages. When a process is to be executed, its pages are loaded into any available memory frames from the backing store. The backing store, swap space on disk, is divided into fixed-sized blocks that are of the same size as the memory frames.

Every address generated by the CPU is divided into two parts: a page number (p) and a page offset (d). The page number is used as an index into a page table. The page table contains the base address of each page in physical memory. This base address is combined with the page offset to define the physical memory address that is sent to the memory unit.




The page size (like the frame size) is defined by the hardware. The size of a page is typically a power of 2. The selection of a power of 2 as a page size makes the translation of a logical address into a page number and page offset particularly easy. If the size of logical address space is $2^m$ and a page size is $2^n$ addressing units (bytes or words), then the high-order m - n bits of a logical address designate the page number, and the n low order bits designate the page offset. Thus, the logical address is as follows:

| Page number | Page Offset |
|---|---|
| p | d |
| m-n | n |

Where p is an index into the page table and d is the displacement within the page.

Paging is to deal with external fragmentation problem. This is to allow the logical address space of a process to be noncontiguous, which makes the process to be allocated physical memory.

**Translation Lookaside Buffer**

Each operating system has its own methods for storing page tables. Most allocate a page table for each process and store a pointer to it in the process control block. The hardware implementation of this page table can be done in several ways. In the simplest case, the page table is implemented as a set of dedicated registers. These registers should be built with very high-speed logic to make the paging-address translation efficient. The use of registers for the page table is satisfactory if the page table is reasonably small (for example, 256 entries). Most contemporary computers, however, allow the page table to be very large (for example, 1 million entries). For these machines, the use of fast registers to implement the page table is not feasible. Rather, the page table is kept in main memory, and a page table base register (PTBR) points to the page table. Changing page tables requires changing only this one register, substantially reducing context-switch time.

The problem with this approach is the time required to access a user memory location.

The standard solution to this problem is to use a special, small, fast lookup hardware cache, called a translation look aside buffer (TLB). The TLB is associative, high-speed memory. Each entry in the TLB consists of two parts: a key (or tag) and a value. When the associative memory is presented with an item, the item is compared with all keys simultaneously. If the item is found, the corresponding value field is returned. The search is fast; the hardware, however, is expensive. Typically, the number of entries in a TLB is small, often numbering between 64 and 1,024.

**Segmentation**

**Segmentation** is a memory-management scheme that supports this user view of memory. In this scheme, logical address space is a collection of segments. Each segment has a name and a length. The addresses specify both the segment name and the offset within the segment.

For simplicity of implementation, segments are numbered and are referred to by a segment number, rather than by a segment name. Thus, a logical address consists of a two tuple:

<Segment-number, offset>.

Although the user can now refer to objects in the program by a two-dimensional address, the actual physical memory is still, of course, a one-dimensional sequence of bytes. Thus, we must define an implementation to map two dimensional user-defined addresses into one-dimensional physical addresses. This mapping is affected by a **segment table**. Each entry of the segment table has a segment base and a segment limit. The segment base contains the starting physical address where the segment resides in memory, and the segment limit specifies the length of the segment.

### 3.5. Virtual memory

Virtual memory is a technique that allows the execution of processes that may not be completely in memory i.e., programs can be larger than physical memory. Further, it abstracts main memory into an extremely large, uniform array of storage, separating logical memory as viewed by the user from physical memory. This technique frees programmers from concern over memory storage limitations. Virtual memory also allows processes to share files easily and to implement shared memory. In addition, it provides an efficient mechanism for process creation.

Virtual memory is commonly implemented by **demand paging**. A demand-paging system is similar to a paging system with swapping. Processes reside on secondary memory. When we want (on demand) to execute a process, we swap it into memory. Rather than swapping the entire process into memory, however, we use a lazy swapper. A lazy swapper never swaps a page into memory unless that page will be needed.

**Page Fault:**
A page fault is a type of interrupt, raised by the hardware when a running program accesses a memory page that is mapped into the virtual address space, but not loaded in physical memory.

**Page Replacement**
Whenever a page fault occurs, a new page has to be loaded into memory. But since other process are also running, chances that a free frame does not exist increases. Thus page replacement is a possibility to continue with the running of the program. When the page being replaced is needed by the program, another page is replaced to bring in that page again. This series continues.
Some of the common page replacement algorithms used for this purpose are:

**1. FIFO** – this is the simplest page replacement algorithm where with each page the time when that page was brought into memory is associated. For the replacement of the page, the oldest page is selected. Thus, the name of this algorithm. The page that is paged in first is also the one which is paged out first. This algorithm suffers from Belady's Anomaly, which is for some page replacement algorithms, the page-fault rate increases with increase in the number of available frames. Normally with increase in number of frames, page-fault rate decreases.

**2. Optimal Page Replacement** – this algorithm replaces the page that will not be used for the longest period of time. This is called optimal because it guarantees the lowest possible page-fault rate for a fixed number of frames. But, this algorithm is not practical as it requires the knowledge of usage of pages in advance, which is not always possible.
**3. LRU Page Replacement –** this algorithm like its name replaces the least recently used page. For this it keeps track of the time of last usage of each page. According to simulation results, this algorithm performs better than FIFO but has higher page fault rate than optimal page replacement algorithm.

### 3.6. Thrashing

To make the effective utilization of CPU and virtual memory, we increase the degree of multiprogramming by starting more process and reducing the number of frames for each process. This can be done as if we need a new page for a process, it can be page in using any of the page replacement algorithm. Thus, if it so happens, that the degree of multiprogramming becomes so high that more of the CPU time is consumed in swapping in and out of pages than execution of the programs, the condition is called thrashing.

## 4. File System:

File system provides the mechanism for on-line storage of and access to both data and programs of the operating system and all the users of the computer system. The file system consists of two distinct parts: a collection of files, each storing related data, and a directory structure, which organizes and provides information about all the files in the system.
Computers store information on several different storage media, such as magnetic disks, magnetic tapes, and optical disks. The operating system provides a uniform logical view of information storage by abstracting the physical properties of the storage devices to define a logical storage unit, the file. A file is a named collection of related information that is recorded on secondary storage.
Data cannot be written to secondary storage unless they are within a file. Commonly, files represent programs and data. Data files may be numeric, alphabetic, alphanumeric, or binary. In general, a file is a sequence of bits, bytes, lines, or records whose meaning is defined by the file's creator and user. The information in a file is defined by its creator. Many different types of information may be stored in a file: source programs, object programs, executable programs,

numeric data, text, payroll records, graphic images, sound recordings, and so on. A file has a certain defined structure according to its type.

## 4.1.File Attributes

- Name - Every file has a name which helps users to differentiate between files.
- Identifier - This unique tag (number) identifies the file within the file system.
- Location - This information is a pointer to a device and to the location of the file on that device.
- Size - The current size of the file (in bytes, words or blocks), and possibly the maximum allowed size are included in this attribute.
- Protection - Access-control information controls, who can do reading, writing, executing, and so on.
- Time, date, and user identification - This information is kept for creation, last modification, and last use. These data can be useful for protection, security, and usage monitoring.

## 4.2.File Operations

A file is an abstract data type. To define a file properly, we need to consider the operations that can be performed on files. The operating system provides system calls to create, write, read, reposition, delete, and truncate files.

- Creating a file - To create a file, space in the file system must be found for the file and an entry for the new file must be made in the directory. The directory entry records the name of the file and the location in the file system.
- Writing a file - To write a file, we make a system call specifying both the name of the file and the information to be written to the file. The system must keep a *write* pointer to the location in the file where the next write is to take place.
- Reading a file - To read from a file, we use a system call that specifies the name of the file and where (in memory) the next block of the file should be put.
- Repositioning within a file - The directory is searched for the appropriate entry, and the current-file-position is set to a given value. This file operation is also known as file seek.
- Deleting a file - To delete a file, we search the directory for the named file. Having found the associated directory entry, we release all file space and erase the directory entry.
- Truncating a file - This function allows all attributes to remain unchanged (except for file length) but for the file to be reset to length zero.

## 4.3.File Types

Files have various types; most common are text and binary. Depending on the type of file certain operations are limited to types. Example - a common mistake occurs when a user tries to print the binary-object form of a program. This attempt normally produces garbage, but can be prevented if the operating system has been told that the file is a binary-object program.

A common technique for implementing file types is to include the type as part of the file name. The name is split into two parts — a name and an extension, usually separated by a period character. The system uses the extension to indicate the type of the file and the type of operations that can be done on that file. For instance, only a file with a ".com", ".exe", or ".bat" extension can be executed. Some of the common file types are:

File types Usual Extension Function:

- executable exe, com, bin Run machine language program
- object obj, o Compiled, machine language, not linked
- Source code c, cc, java, asm Source code in various languages
- Batch bat, sh Commands to the command interpreter
- text txt, doc Textual data, documents
- Word processor wp, tex, doc Various word processor formats
- archive arc, zip, tar, rar Related files group into one file,
- sometimes after compressing
- Multimedia mpeg, mov, rm Binary file containg audio or a/v information.

## 4.4. Access Methods

Files store information. When it is used, this information must be accessed and read into computer memory. The information in the file can be accessed in several ways.

**Sequential Access** - Information in the file is processed in order, one record after the other. A read operation reads the next portion of the file and automatically advances a file pointer, which tracks the I/O location. Similarly, a write appends to the end of the file and advances to the end of the newly written material.
**Direct Access** – This method is based on disk model and allows random access to any file block. Direct-access files are of great use for immediate access to large amounts of information. Databases are often of this type. When a query concerning a particular subject arrives, we compute which block contains the answer, and then read that block directly to provide the desired information.
**Other Methods** – The additional methods generally involve the construction of an index for the file. The index, like an index in the back of a book, contains pointers to the various blocks. To find a record in the file, we first search the index, and then use the pointer to access the file directly and to find the desired record.

## 5. INPUT- OUTPUT system

The two main jobs of a computer are I/O and processing. In many cases, the main job is I/O, and the processing is merely incidental. For instance, when we browse a web page or edit a file, our immediate interest is to read or type in some information; it is not to compute an answer. The role of the operating system in computer I/O is to manage and control I/O operations and I/O devices. The operating system provides interfaces to interact with the underlying hardware. These interfaces abstract away the detailed differences in I/O devices and provide a uniform way for interaction. The actual differences are encapsulated in kernel modules called device drivers that internally are custom tailored to each device. The purpose of the device-driver layer is to hide the differences among device controllers from the I/O subsystem of the kernel.

**Devices vary in many aspects. Some of them are:**

- Character-stream or block - A character-stream device transfers bytes one by one, whereas a block device transfers a block of bytes as a unit.
- Sequential or random-access - A sequential device transfers data in a fixed order that is determined by the device, whereas the user of a random-access device can instruct the device to seek to any of the available data storage locations.
- Synchronous or asynchronous - A synchronous device is one that performs data transfers with predictable response times while an asynchronous device exhibits irregular or unpredictable response times.
- Sharable or dedicated - A sharable device can be used concurrently by several processes or threads but a dedicated one cannot.
- Speed of operation - Device speed range from a few bytes per second to a few gigabytes per second.
- Read-write, read only, or write only - Some devices perform both input and output, but others support only one data direction.

Kernels provide many services related to I/O. Several services - I/O scheduling, buffering, caching, spooling, and error handling - are provided by the kernel's I/O subsystem, and build on the hardware and device-driver infrastructure.

**I/O scheduling** - To schedule a set of I/O requests means to determine a good order in which to execute them. Scheduling can improve overall system performance, can share device access fairly among processes, and can reduce the average waiting time for I/O to complete. Operating-system developers implement scheduling by maintaining a queue of requests for each device. When an application issues a blocking I/O system call, the request is placed on the queue for that device. The I/O scheduler rearranges the order of the queue to improve the overall system efficiency and the average response time experienced by applications.

**Buffering** - A buffer is a memory area that stores data while they are transferred between two devices or between a device and an application. Buffering is done for three reasons. One reason is to cope with a speed mismatch between the producer and consumer of a data stream. Secondly to adapt between devices that have different data-transfer sizes. Third, to support copy semantics for application I/O.

**Caching** - A cache is region of fast memory that holds copies of data. Access to the cached copy is more efficient than access to the original.

**Spooling** - A spool is a buffer that holds output for a device, such as a printer, that cannot accept interleaved data stream. Although a printer can serve only one job at a time, several applications may wish to print their output concurrently, without having their output mixed together. The operating system solves this problem by intercepting all output to the printer. Each application's output is spooled to a separate disk file. When an application finishes printing, the spooling system queues the corresponding spool file for output to the printer. The spooling system copies the queued spool files to the printer one at a time.

**Error handling** - An operating system that uses protected memory can guard against many kinds of hardware and application-errors, so that a complete system failure is not the usual result of each minor mechanical glitch. As a general rule, an I/O system call will return 1 bit of information about the status of the call, signifying either success or failure.

## 6. Directory system/structures:

The most common directory structure used by multiuser systems are:
**(a) Single level directory-** In a single level; directory system, all the files are placed in one directory. This is very common on single user operating system. It has significant limitations when the number of files or when there is more than one user. Since all the files are in same folder, they must have unique name.

**(b) Two level directory-** In the two level directory system, the system maintains a master block that has one entry for each user. this master block contains the address of directories of the users. the problem with this structure is that it effectively isolates one user from another.

**(c) Tree structure directories-** In this structure, the directory are files. This lead to the possibility of having sub-directories that can contain files and sub-subdirectories.

**(d) Acyclic graph Directories-** It is an extension of the tree structured directory structure. In the tree structure directory, files and directories starting from some fixed directory are owned by one particular user. In the acyclic structure, this prohibition is taken out and thus a directory or file under directory can be owned by several users

## 7. Short Introduction of UNIX operating system:

### 7.1. What Is UNIX?

UNIX is an operating system which was first developed in the 1960s, and has been under constant development ever since. By operating system, we mean the suite of programs which make the computer work. It is a stable, multi-user, multi-tasking system for servers, desktops and laptops.

UNIX systems also have a graphical user interface (GUI) like Microsoft Windows which provides an easy to use environment. However, knowledge of UNIX is required for operations which aren't covered by a graphical program, or for when there is no windows interface available, for example, in a telnet session.

### 7.2. Types of UNIX

There are many different versions of UNIX, although they share similarities. The most popular varieties of UNIX are Sun Solaris, GNU/Linux, and MacOS X.

### 7.3. The UNIX Operating System

The UNIX operating system is made up of three parts; the kernel, the shell and the programs.

**(a)The kernel:**
The kernel of UNIX is the hub of the operating system: it allocates time and memory to programs and handles the file store and system calls.

**(b)The shell:**
The shell acts as an interface between the user and the kernel. When a user logs in, the login program checks the username and password, and then starts another program called the shell. The shell is a command line interpreter (CLI). It interprets the commands the user types in and executes them. The commands are themselves programs. Once programs terminate, control is returned to the shell and the user receives another prompt ($ on our systems), indicating that another command may be entered.

The shell keeps a list of the commands you have typed in. If you need to repeat a command, use the cursor keys to scroll up and down the list or type history for a list of previous commands.

(c) **The Programs:**

Programs are not part of the operating system as such, but they are logical sequences of commands, developed for implementing specific tasks. They usually include application software running at the user end.

### 7.4. Some Important UNIX Command:

| Command | Example | Description |
|---|---|---|
| Is | ls<br>ls -alF | Lists files in current directory<br>List in long format |
| cd | cd tempdir<br>cd ..<br>cd ~dhyatt/web-docs | Change directory to tempdir<br>Move back one directory<br>Move into dhyatt's web-docs directory |
| mkdir | mkdir graphics | Make a directory called graphics |
| rmdir | rmdiremptydir | Remove directory (must be empty) |
| cp | cp file1 web-docs<br>cp file1 file1.bak | Copy file into directory<br>Make backup of file1 |
| rm | rm file1.bak<br>rm *.tmp | Remove or delete file<br>Remove all file |
| mv | mv old.html new.html | Move or rename files |
| more | more index.html | Look at file, one page at a time |
| lpr | lpr index.html | Send file to printer |
| man | man ls | Online manual (help) about command |

## PRACTICE SET

1. What is true about an operating system?
(a) An operating system is a program that acts as an intermediary between the user and the computer hardware.
(b) The purpose of an OS is to provide a convenient environment in which user can execute programs in a convenient and efficient manner.
(c) Both (a) and (b)
(d) None of these
(e) An operating system or OS is a hardware program that enables the computer software.
2. Which of the following is not a type of operating systems?
(a) Batched operating systems
(b) Multi-programmed operating systems
(c) Timesharing operating systems
(d) Distributed operating systems
(e) All of the above are types of Operating System
3. What are the basic functions of an operating system?
(a) Operating system controls and coordinates the use of the hardware among the various applications programs for various uses.
(b) Operating system acts as resource allocator and manager.
(c) Operating system is control program which controls the user programs to prevent errors and improper use of the computer.
(d) Operating system doesn't concern with the operation and control of I/O devices.
(e) All except (d)
4. What do you understand by the kernel in operating system?
(a) Kernel is the core and essential part of computer operating system that provides basic services for all parts of OS.
(b)Mostly, it is one of the first programs loaded on start-up (after the boot loader).
(c) None of these
(d) the 'kernel' is the corner component of computer operating systems
(e) Both (a) and (b)
5. Page stealing is a…..
(a)It is a sign of an efficient system
(b) It is taking page frames from other working sets
(c) It should be the tuning goal
(d) It is taking larger disk spaces for pages paged out
(e) None of the above
6. How will you define a dead lock in operating system?
(a) Deadlock is a situation where the two processes are waiting for each other to complete so that they can start. This result both the processes to hang.
(b) A deadlock occurs when a thread enters a waiting state because a requested resource is held by another waiting process, which in turn is waiting for another resource held by another waiting process.
(c) Both (a) and (b)
(d) None of these
(e) A deadlock is a situation in which two computer programs dividing the same resource.

7. Which of the following is not a state of a process?
(a) New (b) Running
(c) Waiting (d) Testing
(e) Terminated
8. What is a process?
(a) A program in execution is called a process.
(b) Process contains the program code and its current activity.
(c) None of these
(d) Both (a) and (b)
(e) Process contains the operating system.
9. What is the difference between Complier and Interpreter?
(a) An interpreter translates entire instructions at a time. But a compiler translates a single instruction at a time.
(b) An interpreter reads one instruction at a time and carries out the actions implied by that instruction. It does not perform any translation. But a compiler translates the entire instructions.
(c) None of these
(d) The overall execution time is slower in interpreter but in compiler overall execution time is comparatively faster.
(e) Both (b) and (d)
10. What is a thread?
(a) A thread is a program line under execution.
(b) Thread sometimes called a light-weight process, is a basic unit of CPU utilization
(c) Thread comprises a thread id, a program counter, a register set, and a stack
(d) All of the above (e) None of these
11. What is Fork ?
(a) The dispatching of a task
(b) The creation of a new job
(c) The creation of a new process
(d) Increasing the priority of a task
(e) None of the above
12. What is Thrashing?
(a) It is a natural consequence of virtual memory systems
(b) It can always be avoided by swapping
(c) It always occurs on large computers
(d) It can be caused by poor paging algorithms
(e) None of the above
13. A system program that sets up an executable program in main memory ready for execution is
(a) Assembler (b) Linker
(c) Loader (d) Compiler
(e) None of the above
14. The register or main memory location which contains the address of the operand is known as.
(a) Pointer (b) Indexed register
(c) Special location (d) scratch pad
(e) None of the above
15. Resolution of externally defined symbols is performed by
(a) Linker (b) Loader
(c) Compiler (d) Assembler
(e) None of the above
16. What do you understand by the logical address space?
(a) Logical address space is generated from CPU
(b) It bound to a separate physical address space is central to proper memory management.
(c) None of these
(d) Both (a) and (b)
(e) Either (a) or (b)
17. Which of the following is not true about Caching?
(a) A cache is a temporary storage area.
(b) It is used to served data faster
(c) Caching is for improving computer system performance
(d) All of the above
(e) None of these
18. Which of the following is true regarding FCFS in operating system?
(a) FCFS stands for First Come, First Served.
(b) It is a type of scheduling algorithm.
(c) In FCFS, if a process requests the CPU first, it is allocated to the CPU first. Its implementation is managed by a FIFO queue.
(d) All (a),(b), and (c) are true
(e) None of these
19. What is true about Fragmentation of the file system?
(a) It occurs only if the file system is used improperly
(b) It increases the access time
(c) Fragmentation causes slow access time because read/write head accessing the data must find all fragments of a file before it can be opened or executed. It is temporary removed by the compaction.
(d) It is a characteristic of all file systems
(e) None of the above
20. Which of the following are(is) Language Processor(s)
(a) Assembles (b) Compilers
(d) Interpreters (d) All of the above
(e) None of the above
21. What do you understand by micro kernel?
(a) Micro kernel is a kernel which run services those are minimal for operating system performance.
(b) It provides the minimal number of mechanisms, just enough to run the most basic functions of a system, in order to maximize the implementation flexibility
(c) In it all operating system code is in single executable image.
(d) Both (a) and (b)
(e) None of these
22. Which of the following is true with respect to Belady's Anomaly?
(a) It is also called FIFO anomaly.

(b) Usually, on increasing the number of frames allocated to a process virtual memory, the process execution is faster, because fewer page faults occur.
(c) Both (a) and (b)
(d) None of these
(e) It is also called GIGO anomaly

23. What is a binary semaphore?
(a) A binary semaphore is one, which takes only 0 and 1 as values.
(b) A binary semaphores are used to implement mutual exclusion and synchronize concurrent processes.
(c) A binary semaphore is one, which takes only 0 values.
(d) None of these
(e) Both (a) and (b)

24. Which of the following is not a Coffman's conditions that lead to a deadlock.
(a) Mutual Exclusion (b) Hold & Wait
(c) No Pre-emption (d) Circular Wait
(e) All are conditions of Coffman's conditions that lead to a deadlock.

25. A system program that combines the separately compiled modules of a program into a form suitable for execution
(a) Assembler (b) Linking loader
(c) Cross compiler (d) Load and go
(e) None of the above

26. The strategy of allowing processes that are logically runnable to be temporarily suspended is called
(a) Preemptive scheduling
(b) Non preemptive scheduling
(c) Shortest job first
(d) First come first served
(e) None of the above

27. The FIFO algorithm
(a) Executes first the job that last entered the queue
(b) Executes first the job that first entered the queue
(c) Execute first the job that has been in the queue the longest
(d) Executes first the job with the least processor needs
(e) None of the above

28. In analyzing the compilation of PL/I program, the term "Lexical analysis" is associated with
(a) Recognition of basic syntactic constructs through reductions.
(b) Recognition of basic elements and creation of uniform symbols
(c) Creation of more optional matrix.
(d) Use of macro processor to produce more optimal assembly code
(e) None of the above

29. A non-relocatable program is one which
(a) cannot be made to execute in any area of storage other than the one designated for it at the time of its coding or translation.
(b) consists of a program and relevant information for its relocation.
(c) can itself performs the relocation of its address-sensitive portions.
(d) all of the above
(e) None of the above

30. Which of the following is true about Banker's algorithm?
(a) Banker's algorithm is used to avoid deadlock.
(b) It tests for safety by simulating the allocation of predetermined maximum possible amounts of all resources
(c) Sometimes referred to as the detection algorithm.
(d) All (a), (b), and (c)
(e) None of these

31. What is the difference between turnaround time and response time?
(a) Turnaround time is the interval between the submission of a job and its completion. Response time is the interval between submission of a request, and the first response to that request.
(b) Response time is the interval between the submission of a job and its completion. Turnaround time is the interval between submission of a request, and the first response to that request.
(c) None of these
(d) Response time is the interval between the submission of a job and its completion. Turnaround time is the interval between submission of a request, and the last response to that request.
(e) Both (a) and (b)

32. Which of the following is not an element of a process image?
(a) User data (b) User program
(c) System Stack (d) Process control Block
(e) All are element of process image

33. What do you understand by batched operating systems?
(a) In it users gives their jobs to the operator who sorts the programs according to their requirements and executes them.
(b) It is the execution of a series of jobs in a program on a computer without manual intervention.
(c) Either (a) or (b)
(d) Both (a) and (b)
(e) None of these

34. What do you understand by throughput?
(a) Number of processes that complete their execution per time unit.
(b) Amount of time to execute a particular process
(c) Amount of time a process has been waiting in the ready queue.
(d) Amount of time it takes from when a request was submitted until the first response is produced
(e) None of these

35. What is a job queue in operating system?
(a) When a process enters the system it is placed in the job queue.

(b) When a process completes the job then it is placed in job queue
(c) When a process waiting for resources then it is placed in job queue
(d) None of these
(e) Both (b) and (c)

36. Which of the following is true regarding ready queue?
(a) The ready queue is a queue of all processes that are waiting to be scheduled on a core/CPU.
(b) The ready queue is a queue of a single process that is waiting to be scheduled on a core/CPU.
(c) None of these
(d) Both (a) and (e)
(e) Every ready queue is the job queue

37. Which one of the following is the address generated by CPU?
(a) Physical address
(b) Absolute address (c) Logical address
(d) None of these (e) Both (a) and (b)

38. The address of a page table in memory is pointed by
(a) Stack pointer (b) Page table base register
(c) Page register (d) Program counter
(e) None of these

39. Program always deals with
(a) Logical address (b) Absolute address
(c) Physical address (d) Relative address
(e) None of these

40. Operating System maintains the page table for
(a) Each process (b) Each thread
(c) Each instruction (d) Each address
(e) None of these

41. What is compaction in operating system?
(a) It is a technique for overcoming internal fragmentation
(b) It is a paging technique
(c) It is a technique for overcoming external fragmentation
(d) It is a technique for overcoming fatal error
(e) None of these

42. Which module gives control of the CPU to the process selected by the short-term scheduler?
(a) Dispatcher (b) Interrupt
(c) Scheduler (d) None of these
(e) Compiler

43. Which scheduling algorithm allocates the CPU first to the process that requests the CPU first?
(a) First-come, first-served scheduling
(b) Shortest job scheduling
(c) Priority scheduling
(d) None of these
(e) Real time scheduling

44. In priority scheduling algorithm, when a process arrives at the ready queue, its priority is compared with the priority of
(a) All process (b) Currently running process
(c) Parent process (d) Init process
(e) None of these

45. Time quantum is defined in which of the following scheduling algorithm in operating system?
(a) Shortest job scheduling algorithm
(b) Round robin scheduling algorithm
(c) Priority scheduling algorithm
(d) Multilevel queue scheduling algorithm
(e) None of these

46. In multilevel feedback scheduling algorithm
(a) A process can move to a different classified ready queue
(b) Classification of ready queue is permanent
(c) Processes are not classified into groups
(d) None of these
(e) Classification of job queue is permanent

47. Which of the following is/are the reasons for the execution of process in two state process model.
(i) A process is created when a user at a terminal logs on to the system.
(ii) A process is created in response to the submission of a job
(iii) A process is created to perform a function on behalf of a user program
(a) i and ii only (b) ii and iii only
(c) i and iii only (d) All i, ii and iii
(e) i only

48. Which of the following are the characteristics of suspended process?
(i) The process is not immediately available for execution.
(ii) The process may or may not be waiting on an event
(iii) The process will remove from this state whether the agent explicitly orders the removal or not.
(a) i and ii only (b) ii and iii only
(c) i and iii only (d) All i, ii and iii
(e) None of these

49. A process in the ……………….. state is moved to the ………….. state if there are no ready processes, then at least one blocked process is swapped out to make room for another process that is not blocked.
(a) blocked, blocked/suspend
(b) ready, ready/suspend
(c) blocked/suspend, ready/suspend
(d) ready/suspend, ready
(e) ready/ suspend

50. When the operating system has performed a …………….. operation, it has two choices for selecting a process either admitting a newly created process or bring in a previously suspended process.
(a) Swapping-in (b) Swapping-out
(c) Blocked-in (d) Blocked-out

51. In …………. there is not necessary to load all of the segments of a process and non-resident segments that are needed are brought in later automatically.
(a) Fixed partitioning (b) Simple Paging
(c) Virtual memory segmentation
(d) Simple segmentation (e) Virtual Paging

52. 14. Which of the following is/are the strengths of virtual memory segmentation techniques used in memory management?
(i) No internal fragmentation
(ii) Higher degree of multi programming
(iii) More efficient to use of main memory
(iv) Large virtual address space
(v) Protection and sharing support
(a) i, ii, iii and iv only (b) i, ii, iii and v only
(c) i, ii, iv and v only (d) ii, iii, iv and v only
(e) i and ii only

53. ……………… when a process leaves a critical section and more than one process is waiting, the selection of a waiting process is arbitrary.
(a) Busy waiting is employed
(b) Starvation is possible (c) Deadlock is possible
(d) All of the above (e) None of these

54. In ……………….. messages are not send directly from sender to receiver but rather are sent to a shared data structure consisting queues that can temporarily hold messages.
(a) Direct addressing (b) Indirect addressing
(c) One-to-one-addressing (d) One-to-many addressing
(e)None of these

55. For …………. page to be replaced is chosen from all available frames in main memory.
(a) Fixed allocation in Local Scope
(b) Fixed allocation in Global Scope
(c) Variable allocation in Local Scope
(d) Variable allocation in Global Scope
(e) None of these

56. What is Mutex in operating system?
(a) Mutex is a program object that allows multiple program threads to share the same resource, such as file access, but not simultaneously.
(b) The Mutex is set to unlock when the data is no longer needed or the routine is finished.
(c) Both (a) and (b)
(d) None of these
(e) Some operating system does not support using mutex between process.

57. The ………………… module deals with the device as a logical resource and is not concerned with the details of actually controlling the device.
(a) Directory Management (b) Logical I/O
(c) Device I/O (d) Scheduling and control
(e) Linear buffer

58. When a user process issues an I/O request, the operating system assigns a buffer in the system portion of main memory to the operation is called ………….
(a) Double buffer (b) Single buffer
(c) Linear buffer (d) Circular buffer
(e) None of these

59. In …………………….. file organization, a fixed format is used for records where all records are of the same length, consisting of the same number of fixed length fields in a particular order.
(a) Pile (b) Sequential
(c) Indexed sequential (d) Indexed
(e) Buffer

60. An interactive user or a process has associated with pathname is a current directory which is often referred to as the …………………….
(a) Update directory (b) List directory
(c) Working directory (d) Create directory
(e) Name directory

## SOLUTIONS

| | | | | | | |
|---|---|---|---|---|---|---|
| 1. (c) | 2. (e) | 3. (e) | 4. (e) | 5. (b) | 6. (c) | 7. (d) |
| 8. (d) | 9. (e) | 10. (d) | 11. (c) | 12. (d) | 13. (c) | 14. (a) |
| 15. (a) | 16. (d) | 17. (d) | 18. (d) | 19. (c) | 20. (d) | 21. (d) |
| 22. (c) | 23. (e) | 24. (e) | 25. (b) | 26. (a) | 27. (b) | 28. (b) |
| 29. (a) | 30. (d) | 31. (a) | 32. (e) | 33. (d) | 34. (a) | 35. (a) |
| 36. (d) | 37. (c) | 38. (b) | 39. (a) | 40. (a) | 41. (c) | 42. (a) |
| 43. (a) | 44. (b) | 45. (b) | 46. (a) | 47. (d) | 48. (a) | 49. (a) |
| 50. (b) | 51. (c) | 52. (c) | 53. (b) | 54. (b) | 55. (d) | 56. (c) |
| 57. (b) | 58. (b) | 59. (b) | 60. (c) | | | |

# 5 Networking

## 1. Networking

Networking is a collection of interconnected computers and devices that communicate with each other and share resources and information.

**Advantages of Networking**

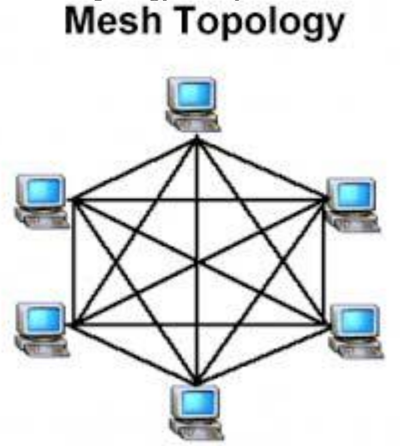
- **Sharing of resources –** Both hardware (printer, hard disk etc.) and software resources can be shared on a network.
- **Reliability-** Since we have more than one copies of a file on the network, reliability increases in case of system failure.
- **Cost Reduction-** Sharing of resources enables considerable cost reduction of resources.
- **Efficient Communication medium-** E-mail, Video Conferencing etc are technologies that have provided better communication and increased productivity.
- **File sharing –** Files can be transferred on a network faster than any other me

A network is a set of devices(often referred to as nodes) connected by communication links.
A node cab be a computer, printer or any other device capable of sending and/or receiving data generated by other nodes on the network.
A network must be able to meet a certain number of criteria. The most important of these are:

- Performance
- Reliability
- Security

**1.1.Network Topology:**
Topology refers to the way in which the network of computers is connected.
The Topology of a network is a geometric representation of the relationship of all the links and linking devices(usually called nodes) to one another.

**1.1.1.Mesh Topology:**
In mesh topology every device has a dedicated point-to-point link to every other device.




Two nodes are connected by dedicated point-point links between them. So the total number of links to connect n nodes = n(n-1)/2; which is proportional to $n^2$ . Media used for the connection (links) can be twisted pair, co-axial cable or optical fiber

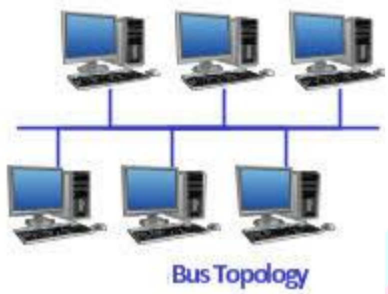
Mesh topology is not flexible and has a poor expandability as to add a new node n links have to be laid because that new node has to be connected to each of the existing nodes via dedicated link.
So the cost of cabling will be very high for a large area, due to these reasons this topology is rarely used in practice.

**Advantages of a mesh topology**
- Can handle high amounts of traffic, because multiple devices can transmit data simultaneously.
- A failure of one device does not cause a break in the network or transmission of data.
- Adding additional devices does not disrupt data transmission between other devices.

**Disadvantages of a mesh topology**
- The cost to implement is higher than other network topologies, making it a less desirable option.
- Building and maintaining the topology is difficult and time consuming.
- The chance of redundant connections is high, which adds to the high costs and potential for reduced efficiency.

**1.1.2.Bus Topology:**
Bus topology uses one main cable to which all nodes are directly connected. The main cable acts as a backbone for the network. One of the computers in the network typically acts as the computer server. A bus topology is multipoint topology. One long cable act as a back bone to link all the devices in the network.
Nodes connected to the bus cable by drop lines and taps. Drop line is a connection line which is used to connect each computer with the main cable. Tap is a connector where drop line is connected with main cable.




**Advantages of bus topology**
- It works well when you have a small network.
- Easiest network topology for connecting computers or peripherals in a linear fashion.
- Requires less cable length than a star topology.

**Disadvantages of bus topology**
- Difficult to identify the problems if the whole network goes down.
- It can be hard to troubleshoot individual device issues.
- Terminators are required for both ends of the main cable.
- Additional devices slow the network down.
- If a main cable is damaged, the network fails or splits into two.

**1.1.3.Ring Topology:**
In Ring Topology , each device has a dedicated point-to-point connection with only the two devices on either side of it. A signal is passed along the ring in one direction, from device to device, until it reaches its destination.Each device in the ring incorporates a repeater. When a device receives a signal intended for another device, its repeater regenerates the bits and passes them along.

A ring topology is relatively easy to install and reconfigure. Each device is linked to only its immediate neighbours (either physically or logically). To add or delete a device requires changing only two connections.


**Advantages of ring topology**
- All data flows in one direction, reducing the chance of packet collisions.

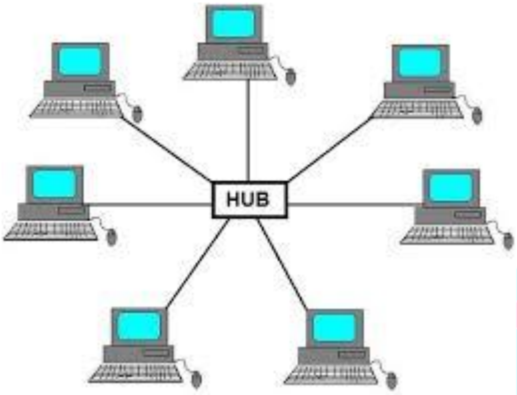
- Data can transfer between workstations at high speeds.
- Additional workstations can be added without impacting performance of the network.

**Disadvantages of ring topology**

- All data being transferred over the network must pass through each workstation on the network, which can make it slower.
- The entire network will be impacted if one workstation shuts down.
- The hardware needed to connect each workstation to the network is more expensive.

**1.1.4.Star Topology:**

In star topology , each device has a dedicated point-to-point link only to a central controller, usually called a hub. The device are not directly linked to one another. The controller act as an exchange: If one device wants to send data to another, it sends the data to the controller, which then relays the data to the other connected device.




**Advantages of star topology**

- Centralized management of the network, through the use of the central computer, hub, or switch.
- Easy to add another computer to the network.
- If one computer on the network fails, the rest of the network continues to function normally.

**Disadvantages of star topology**

- Can have a higher cost to implement, especially when using a switch or router as the central network device.
- The central network device determines the performance and number of nodes the network can handle.
- If the central computer, hub, or switch fails, the entire network goes down and all computers are disconnected from the network.

**1.1.5.Tree Topology:**

Tree Topology integrates the characteristics of Star and Bus Topology.
In Tree Topology, the number of Star networks are connected using Bus. This main cable seems like a main stem of a tree, and other star networks as the branches. It is also called Expanded Star Topology.

**Advantages of Tree Topology**

- It is an extension of Star and bus Topologies, so in networks where these topologies can't be implemented individually for reasons related to scalability, tree topology is the best alternative.
- Expansion of Network is possible and easy.
- Here, we divide the whole network into segments (star networks), which can be easily managed and maintained.
- Error detection and correction is easy.
- Each segment is provided with dedicated point-to-point wiring to the central hub.
- If one segment is damaged, other segments are not affected.

**Disadvantages of Tree Topology**

- Because of its basic structure, tree topology, relies heavily on the main bus cable, if it breaks whole network is crippled.
- As more and more nodes and segments are added, the maintenance becomes difficult.
- Scalability of the network depends on the type of cable used.

**1.1.6.Hybrid Topology:**
Hybrid topology is an integration of two or more different topologies to forma resultant topology which has many advantages of all the constituent basic topologies rather than having characteristics of one specific topology.

**Advantages of Hybrid Topology:**
- Unlike other networks, fault detection and troubleshooting is easy in this topology.
- It is easy to increase the size of network by adding new components, without disturbing existing architecture.
- Hybrid network is flexible, it can be designed according to the requirements of the organization and by optimizing the available resources.
- This type of topology is very effective because it is the combination of two or more topologies,so we can design it in such a way that strengths of constituent topologies are maximized while there weaknesses are neutralized. Forexample we saw Ring Topology has good data reliability and Star topology has high tolerance capability so these two can be used effectively in hybrid star-ring topology.

**Disadvantages of Hybrid Topology:**
- One of the biggest drawbacks of hybrid topology is its complex design. Configuration and installation process is very difficult.
- The hubs used to connect two distinct networks are very expensive.
- As hybrid architectures are usually larger in scale, they require a lot of cables; cooling systems and sophisticate network devices.

**Categories of network:**
- **LAN(Local Area Network):** A LAN connects network devices over a relatively short distance. A networked office building, school, or home usually contains a single LAN, though sometimes one building will contain a few small LANs (perhaps one per room), and occasionally a LAN will span a group of nearby buildings.A LAN is very useful for sharing resources, such as data storage and printers. LANs can be built with relatively inexpensive hardware, such as hubs, network adapters and Ethernet cables.
- **MAN(Metropolitan Area Network):** MAN consists of a computer network across an entire city, college campus or small region. A MAN is larger than a LAN, which is typically limited to a single building or site. Depending on the configuration, this type of network can cover an area from several miles to tens of miles. A MAN is often used to connect several LANs together to form a bigger network.
- **WAN(Wide Area Network):**A WAN provides long distance transmission of data, image, audio,and video information over large geographic areas that may comprise a country, a continent,or even the whole world.
- **SAN(Storage Area Network):** SAN is a high-speed network of storage devices that also connects those storage devices with servers.
- **CAN(Campus Area Network):** A campus area networks (CANs) is a computer network interconnecting a few local area networks (LANs) within a university campus or corporate campus Network.Campus area network may link a variety of campus buildings.
- **PAN(Personal Area Network):**A personal area network is a computer network organized around an individual person. Personal area networks typically involve a mobile computer,Personal area networks can be constructed with cables or wireless.

## 2. Data Communication:

Data communication are the exchange of data between two devices via some form of transmission medium such as a wire cable.
The communicating devices must be part of a communication system made up of a combination of hardware and software. The effectiveness of data communication depends on the four basic characteristics: delivery, accuracy, timeliness, and jitter.

**Components of Data Communication:**
- **Message:**Message is the information(data) to be communicated. e.g. text, picture, audio, video etc.
- **Sender:**The sender is the device that sends the data message. It can be computer, workstation, telephone handset and so on.

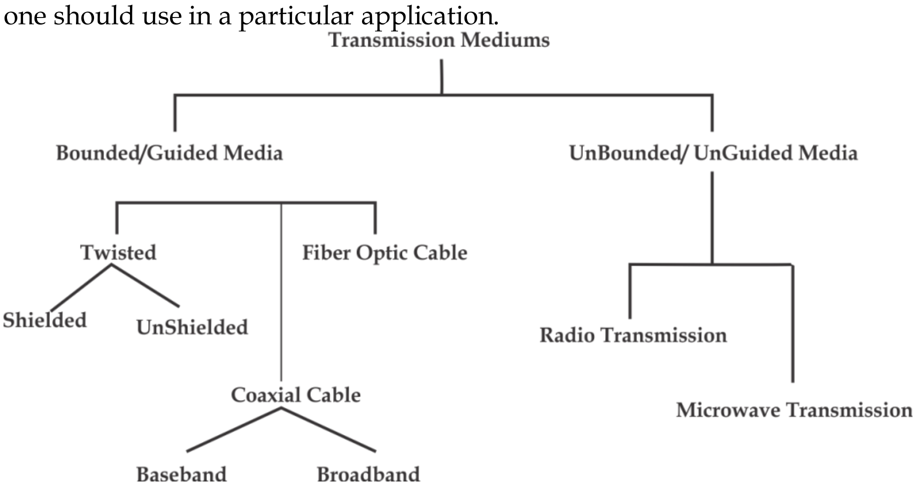
- **Receiver:**The receiver is the device that receive the message. It can be computer, workstation, telephone handset and so on.
- **Transmission Medium:**Transmission medium is the physical path by which a message travels from sender to receiver.

**Data flow:** Communication between two device can be:
**Simplex:** Simplex mode communication is unidirectional or a one-way street. Only one of the two devices on a link can transmit; the other can only receive. Keyboards and traditional monitors are example of simplex devices.
**Half-Duplex:** In half duplex, each station can both transmit and receive, but not at the same time. When one device is sending, the other can only receive and vice versa. Walkie-talkies is the example of half duplex.
**Full-Duplex:** In full duplex, both stations can transmit and receive simultaneously. The full duplex mode is like a two-way street with traffic flowing in both directions at the same time. one common example of full duplex communication is the telephone network. When two people are communicating by a telephone line, both can talk and listen at the same time.

**Protocol:**
A protocol is a set of rules that govern data communication.A protocol defines what is communicated,how it is communicated and when it is communicated.
The key element of protocol is:
**Syntax-** syntax refers to the structure or format of the data,meaning the order in which they are presented.
**Semantics-**Semantics refer to the meaning of each section of bits. How is a particular pattern to be interpreted, and what action is to be taken based on that interpretation?
**Timing-** Timing refers two characteristics: When data should be sent and how fast they can be sent.

**2.1. Transmission medium:**
Transmission media can be defined as physical path between transmitter and receiver in a data transmission system. And it may be classified into two types as:
- Guided: Transmission capacity depends critically on the medium, the length, and whether the medium is point-to-point or multipoint (e.g. LAN). Examples are coaxial cable, twisted pair, and optical fiber.
- Unguided: provides a means for transmitting electro-magnetic signals but do not guide them. Example wireless transmission.

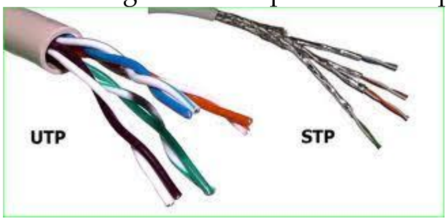
Characteristics and quality of data transmission are determined by medium and signal characteristics. For guided media, the medium is more important in determining the limitations of transmission. While in case of unguided media, the bandwidth of the signal produced by the transmitting antenna and the size of the antenna is more important than the medium. Signals at lower frequencies are omni-directional (propagate in all directions). For higher frequencies, focusing the signals into a directional beam is possible. These properties determine what kind of media one should use in a particular application.




**Guided media:**
Guided media, which are those that provide a conduit from one device to another, include Twisted-Pair Cable, Coaxial Cable, and Fibre-Optic Cable.
A signal travelling along any of these media is directed and contained by the physical limits of the medium.

**Twisted Pair Cable**
This cable is the most commonly used and is cheaper than others. It is lightweight, cheap, can be installed easily, and they support many different types of network. Its frequency range is 0 to 3.5 kHz.
A twisted pair consists of two conductors (normally copper), each with its own plastic insulation, twisted together.
Twisted Pair are two types:

- **Unshielded Twisted Pair Cable**
  It is the most common type of telecommunication when compared with Shielded Twisted Pair Cable which consists of two conductors usually copper, each with its own colour plastic insulator. Identification is the reason behind coloured plastic insulation.
- **Shielded Twisted Pair Cable**
  This cable has a metal foil or braided-mesh covering which encases each pair of insulated conductors. Electromagnetic noise penetration is prevented by metal casing.

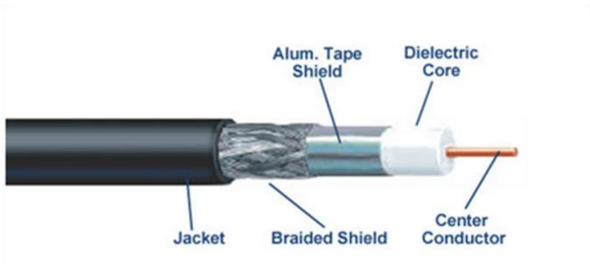



**Coaxial Cable:**
Coaxial is called by this name because it contains two conductors that are parallel to each other. Copper is used in this as centre conductor which can be a solid wire or a standard one. It is surrounded by PVC installation, a sheath which is encased in an outer conductor of metal foil, braid or both.

Outer metallic wrapping is used as a shield against noise and as the second conductor which completes the circuit. The outer conductor is also encased in an insulating sheath. The outermost part is the plastic cover which protects the whole cable.

Co-axial cable has superior frequency characteristics compared to twisted-pair and can be used for both analog and digital signaling.

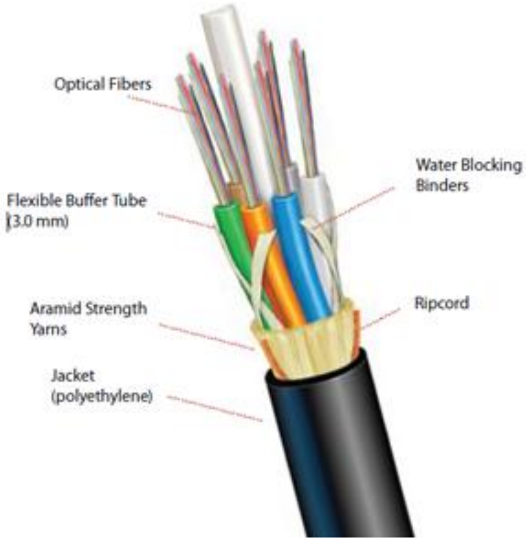
One of the most popular use of co-axial cable is in cable TV (CATV) for the distribution of TV signals. Another importance use of co-axial cable is in LAN.




Foil/Braid Shield Coaxial Cable

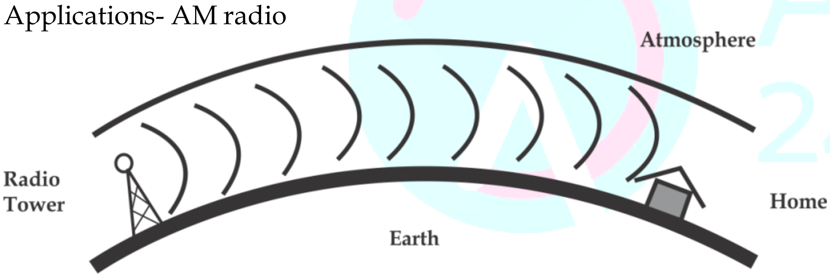
**Fiber Optics:**
It transmits signals in the form of light and is made up of an inner core of glass or plastic. The core is surrounded by a cladding that reflects light back into the core. Each fiber is surrounded by a plastic casing. It is very efficient medium because it provides maximum bandwidth, lower attenuation and is immune to Electromagnetic Interference.

Fiber optics cable have Very high data rate, low error rate. 1000 Mbps (1 Gbps) over distances of kilometers common. Error rates are so low they are almost negligible. Because of greater bandwidth (2Gbps), smaller diameter, lighter weight, low attenuation, immunity to electromagnetic interference and longer repeater spacing, optical fiber cables are finding widespread use in long-distance telecommunications. Especially, the single mode fiber is suitable for this purpose. Fiber optic cables are also used in high-speed LAN applications. Multi-mode fiber is commonly used in LAN.

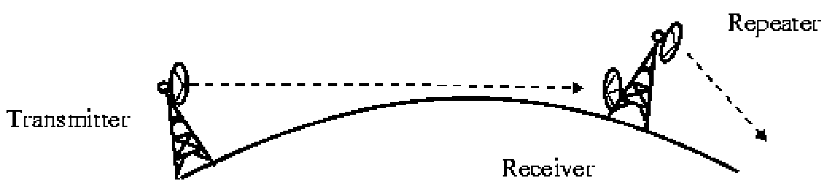



**Unguided Transmission:**
Unguided transmission is used when running a physical cable (either fiber or copper) between two end points is not possible. For example, running wires between buildings is probably not legal if the building is separated by a public street. Infrared signals typically used for short distances (across the street or within same room), Microwave signals commonly used for longer distances (10's of km). Sender and receiver use some sort of dish antenna.

**Radiowaves:** These are the electromagnetic waves having a frequency range of 3KHz-1GHz. These are omnidirectional i.e. the senders and receivers do not have to be in line of sight with each other. These can penetrate walls and are prone to interference.
Applications- AM radio




**Microwaves**
These are the electromagnetic waves having frequencies ranging from 1 to 300GHz. They are unidirectional and incorporates two antennas(sending & receiving) which should be aligned or in line of sight with each other. They provide higher data rate but Very high frequency microwaves are unable to penetrate the walls. Applications- cellular phones, satellite networks, wireless L




**Infrared**
Infrared waves use infrared light for signal transmission. These have frequency range from 300 GHz to 400 THz and are used for short-range communication. They also incorporate line of sight propagation. It also provides high bandwidth and high data rate.
Applications- for communication between PCs, mobile phones etc.

**2.2 Multiplexing:**
Multiplexing is the set of techniques that allows the simultaneous transmission of multiple signals across a single data link. There are three basic multiplexing techniques:
**Frequency-division multiplexing:**

Frequency Division Multiplexing is a technique which uses various frequencies to combine many streams of data for sending signals over a medium for communication purpose. It carries frequency to each data stream and later combines various modulated frequencies to transmission. Television Transmitters are the best example for FDM, which uses FDM to broad cast many channels at a time.

**Time-division Multiplexing:**

It is also called synchronous TDM, which is commonly used for multiplexing digitized voice stream. The users take turns using the entire channel for short burst of time.

**Wavelength-division Multiplexing:**

Wavelength Division Multiplexing (WDM) is analog multiplexing technique and it modulates many data streams on light spectrum. This multiplexing is used in optical fiber. Various signals in WDM are optical signal that will be light and were transmitted through optical fiber.WDM similar to FDM as it mixes many signals of different frequencies into single signal and transfer on one link. Wavelength of wave is reciprocal to its frequency, if wavelength increase then frequency decreases. Several light waves from many sources are united to get light signal which will be transmitted across channel to receiver.

## 3. Networking Devices:

These are used to connect different devices in the network or two connect two or more different networks.
Following devices are used for interconnection:

- Modem
- Hub
- Switch
- Repeater
- Router
- Bridge
- Gateway

**Modem:** Modem stands for Modulator-Demodulator. It is used to connect computers for communication via telephone lines.

**Hub:** It works at the Physical layer. It just acts like a connector of several computers i.e it simply connects all the devices on its ports together .It broadcasts all the data packets arriving at it with no filtering capacity.

**Switch:** Switch is data link layer device. A network switch also connects computers to each other, like a hub. Where the switch differs from a hub is in the way it handles packets of data. When a switch receives a packet of data, it determines what computer or device the packet is intended for and sends it to that computer only. It does not broadcast the packet to all computers as a hub does which means bandwidth is not shared and makes the network much more efficient.

**Repeater:** It operates at the physical layer. It is used to amplify a signal that has lost its original strength so as to enable them to travel long distances. It can only join the networks that transmit similar data packets. It does not have filtering capacity i.e. all data including noise is amplified and passed on in the network so don't help in reducing network traffic.

**Router:** It works at the network layer and is used to connect different networks that have different architectures and protocols. It sends the data packets to desired destination by choosing the best path available thus reducing network traffic. It routes the data packets using the routing table that contains all the Information regarding all known network addresses, possible paths and cost of transmission over them. Availability of path and cost of transmission decide sending of data over that path. It is of 2 types: static (manual configuration of routing table is needed) and dynamic (automatically discovers paths).

**Gateway:** It operates in all the layers of the network architecture. It can be used to connect two different networks having different architectures, environment and even models. It converts the data packets in form that is suitable to the destination application. The two different networks may differ in types of communication protocols they use, language, data formats etc.

**Bridge:** They are used two connect two LANs with the same standard but using different types of cables. It provides an intelligent connection by allowing only desired messages to cross the bridge thus improving performance. It uses physical addresses of the packets for this decision. It works on data link layer of the OSI model. A bridge uses Spanning tree Algorithm for data transmission so as to avoid loops in the network.

**Brouter:** A brouter is a device that functions as both a bridge and a router. It can forward data between networks (serving as a bridge), but can also route data to individual systems within a network (serving as a router).
The brouters functions at the network and data link layer of the OSI model.

## 4. Networking Switching:

A network switch is a computer networking device that connects devices together on a computer network by using packet switching to receive, process, and forward data to the destination device.

There are basically three types of switching methods are made available. Out of three methods, circuit switching and packet switching are commonly used but the message switching has been opposed out in the general communication procedure but is still used in the networking application.

- **Circuit Switching:** Circuit switched network consists of a set of switches connected by physical links. In circuit switched network, two nodes communicate with each other over a dedicated communication path. There is a need of pre-specified route from which data will travel and no other data is permitted. Before starting communication, the nodes must make a reservation for the resources to be used during the communication. In this type of switching, once a connection is established, a dedicated path exists between both ends until the connection is terminated.
- **Packet Switching:** In packet switching, messages are divided into packets of fixed or variable size. The size of packet is decided by the network and the governing protocol. Resource allocation for a packet is not done in packet switching. Resources are allocated on demand. The resource allocation is done on first-come, first-served basis. Each switching node has a small amount of buffer space to hold packets temporarily. If the outgoing line is busy, the packet stays in queue until the line becomes available.

  Packet switching method uses two routing methods:

  1. Datagram Packet Switching
  2. Virtual Circuit Packet Switching
- **Message Switching:**In message switching, it is not necessary to establish a dedicated path between transmitter and receiver.In this, each message is routed independently through the network. Each message carries a header that contains the full information about the destination. Each intermediate device receives the whole message and buffers it until there are resources available to transfer it to the next hop. If the next hop does not have enough resources to accommodate large size message, the message is stored and switch waits. For this reason a message switching is sometimes called as Store and Forward Switching. Message switching is very slow because of store-and-forward technique. Message switching is not recommended for real time applications like voice and video.

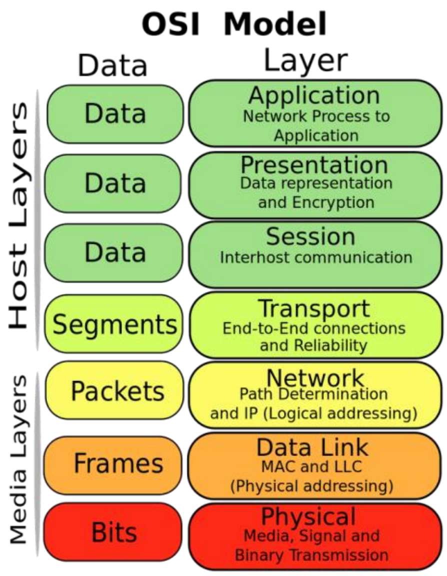
## 5. Network Models:

### 5.1. OSI Model:

OSI (Open System Interconnection) Model was developed by International Standards Organisation (ISO) to standardize the network architecture internationally.

The purpose of the OSI Model is to show how facilitate communication between different system without requiring changes to the logic of the underlying hardware and software. OSI model is not a protocol; it is a model for understanding and designing a network architecture that is flexible, robust and interoperable.

It is a layered framework having seven layers. The layers communicate with each other in a hierarchical manner where control is passed from one layer to another in the hierarchy beginning from the application layer at one computer, then to the bottom layer of that computer. From here the control passes to the bottom layer of the next computer and then back up in the hierarchy.




**Layers in the OSI Model:**

**Layer 1-Physical layer:**

Physical layer coordinates the function required to carry a bit stream over a physical medium.

It defines the mechanical, electrical & physical specifications of the interface & the transmission medium used for communication. It determines how a cable is attached with LAN card & is responsible for transmitting bit stream from one computer to another. Fast Ethernet, ATM etc are some of the protocols that exist here.

| The Physical Layer is responsible for movements of individual bits from one hop(node) to the next. |
|---|

**Functions of Physical Layer:**

- Physical Characteristics of interface and medium: It defines the characteristics of the interface between the device and the transmission medium.
- Representation of bits: Data in this layer consists of stream of bits. The bits must be encoded into signals for transmission. It defines the type of encoding i.e. how 0's and 1's are changed to signal.
- Data Rate:This layer defines the rate of transmission which is the number of bits per second.
- Synchronization of bits: It deals with the synchronization of the transmitter and receiver. The sender and receiver are synchronized at bit level.
- Line Configuration:The Physical layer is concerned with the connection of devices to the media. It connect the device in Point-to-Point and Multipoint configuration.
- Physical Topology:The Physical topology defines how devices are connected to make a network.
- Transmission Mode: The physical layer also defines the direction of transmission between two device: simplex, half duplex or full duplex.

**Layer 2-Data Link Layer:**

Data link layer is most reliable node to node delivery of data. It forms frames from the packets that are received from network layer and gives it to physical layer. It also synchronizes the information which is to be transmitted over the data. Error controlling is easily done. The encoded data are then passed to physical.

Error detection bits are used by the data link layer. It also corrects the errors. Outgoing messages are assembled into frames. Then the system waits for the acknowledgement to be received after the transmission. It is reliable to send message.

**Data link layer has two sub-layers:**

Logical Link Control: It deals with protocols, flow-control, and error control

Media Access Control: It deals with actual control of media

**Responsibility of the data link layer:**

- **Framing:** The datalink layer divides the stream of bits receive from the network layer into manageable data units called frames.
- **Physical addressing:**The Data Link layer adds a header to the frame in order to define physical address of the sender or receiver of the frame, if the frames are to be distributed to different systems on the network.

- **Flow control:** If the rate at which the data are absorbed by the receiver is less than the rate at which data are produced in the sender, the data link layer imposes a flow control mechanism to avoid overwhelming the receiver.
- **Error control:** The data link layer adds reliability to the physical layer by adding mechanisms to detect and retransmit damaged or lost frames.It also uses a mechanism to recognize duplicate frames.
- **Access control:**When two or more devices are connected to the same link, data link layer protocols are necessary to determine which device has control over the link at any given time.

**Layer 3-Network Layer:**
The network layer is responsible for the source-to-destination delivery of a packet, possibly across multiple networks(links). The function of network layer called routing.

**Responsibility of network layer:**
- **Logical addressing**: It translates logical network address into physical address. Concerned with circuit, message or packet switching.
- **Routing**: When independent networks or links are connected to create internetworks (networks of networks) or a large networks, the connecting device(called routers or switches) routes or switch the packets to their final destinations.

**Layer 4-Transport Layer:**

The transport layer is responsible for process-to-process delivery of the entire message. A process is an application program running on a host. whereas the network layer oversees source-to-destination delivery of individual packets, it does not recognize any relationship between those packets.
The transport layer on the other hand, ensure that the whole message arrives intact and in order, overseeing both error control and flow control at the source-to-destination level.

**Responsibility of the transport layer:**
- **service-point addressing:**Transport Layer header includes service point address which is port address. This layer gets the message to the correct process on the computer unlike Network Layer, which gets each packet to the correct computer.
- **Segmentation and reassembly:**A message is divided into segments; each segment contains sequence number, which enables this layer in reassembling the message. Message is reassembled correctly upon arrival at the destination and replaces packets which were lost in transmission.
- **Connection Control:** The transport layer can be either connectionless or connection-oriented. A connectionless transport layer treats each segments as an independent packet and delivers it to the transport layer at the destination mechine. A connection-oriented transport layer makes a connection with the transport layer at the destination machnine first before delivering the packets. After all the data are transfer, the connection is terminated.
- **Flow control:**In this layer, flow control is performed end to end.
- **Error control**:Error Control is performed end to end in this layer to ensure that the complete message arrives at the receiving transport layer without any error. Error Correction is done through retransmission.

**Layer 5-Session Layer:**
The session layer is the network dialog controller. It establishes, maintains and synchronizes the interaction among communicating system.
**Responsibilities of the session layer:**
- **Dialog Control:** The session layer allows two systems to enter into dialog. It allows the communication between two process to take place in either half-dulex or full duplex mode.
- **Synchronization:** The session layer allows a process to add checkpoints, or synchronization points, to a stream of data.

**Layer 6-Presentation Layer:**
The presentation layer is concerned with the syntax and semantics of the information exchange between two systems.
**Responsibilities of Presentation layer:**
- **Translation:** Before being transmitted, information in the form of characters and numbers should be changed to bit streams. The presentation layer is responsible for interoperability between encoding methods as

different computers use different encoding methods. It translates data between the formats the network requires and the format the computer.

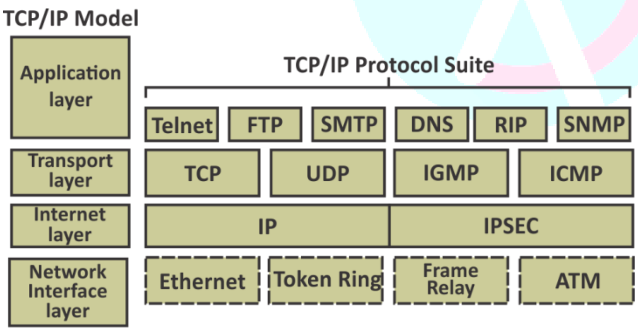
- **Encryption:** To carry sensitive information, a system must be able to ensure privacy. Encryption means that the sender transforms the original information to another form and sends the resulting message out over the network. Decryption reverse the original process to transform the message back to its original form.
- **Compression:** Data compression reduce the number of bits contained in the information. Data compression becomes particularly important in the transmission of multimedia such as text, audio and video.

**Layer 7-Application Layer:**
The application layer enables the user, whether human or software, to access the network. It provides user interface and support for services such as electronic mail, remote file access and transfer, a shared database management,and other type of distributed information services.
**Services provided by application layer:**

- **Network Virtual terminal:**It allows a user to log on to a remote host. The application creates software emulation of a terminal at the remote host. User's computer talks to the software terminal which in turn talks to the host and vice versa. Then the remote host believes it is communicating with one of its own terminals and allows user to log on.
- **File transfer, access, and management:**It is a standard mechanism to access files and manages it. Users can access files in a remote computer and manage it. They can also retrieve files from a remote computer.
- **Mail Services:** This layer provides the basis for E-mail forwarding and storage.

**Directory Services :** This layer provides access for global information about various services.

**5.2.TCP/IP Model:**
TCP/IP means Transmission Control Protocol and Internet Protocol. The TCP/IP protocol suite was developed prior to the OSI model.
The layers of the TCP/IP protocol suite contain relatively independent protocol that can be mixed and matched depending on the needs of the system.
TCP/IP is a layered framework having four layers:-




**Layer 1- Network Interface Layer:**
It is responsible for breaking down the data packets from Internet layer into frames which are then converted into bits for transmission across the physical media. Here, Ethernet, FDDI, Token ring etc. Some of the standards that are defined for data transmission. The Network Interface layer encompasses the Data Link and Physical layers of the OSI model.

**Institute of Electrical and Electronics Engineers (IEEE)**
The Institute of Electrical and Electronics Engineers is an organization which was formed in 1963 in USA. The Institute of Electrical and Electronics Engineers (IEEE) is the world's largest association for Electrical and Electronics Engineers. Institute of Electrical and Electronics Engineers (IEEE) was formed by the merger of two other technical organizations, American Institute of Electrical Engineers and Institute of Radio Engineers in 1st January, 1963. Today, IEEE has about 500,000 members, from different countries in the world.
The IEEE is best known for developing standards for the computer and electronics industry. In particular, the IEEE 802 standards for local-area networks are widely followed.

**LIST OF SOME IMPORTANT IEEE 802 standards:**

| IEEE 802 Standard | |
|---|---|
| **802.1** | Bridging & Management |

| | |
|---|---|
| **802.2** | Logical Link Control |
| **802.3** | Ethernet – CSMA/CD Access Method |
| **802.4** | Token Passing Bus Access Method |
| **802.5** | Token Ring Access Method |
| **802.6** | Distribute Queue Dual Bus Access Method |
| **802.7** | Broadband LAN |
| **802.8** | Fiber Optic |
| **802.9** | Integrated Services LAN |
| **802.10** | Security |
| **802.11** | Wireless LAN |
| **802.12** | Demand Priority Access |
| **802.14** | Medium Access Control |
| **802.15** | Wireless Personal Area Network |
| **802.16** | Broadband Wireless Metro Area Network |
| **802.17** | Resilient Packet Ring |

**Ethernet**

Ethernet is the most popular physical layer LAN technology in use today. Other LAN types include Token Ring, Fast Ethernet, Fiber Distributed Data Interface (FDDI), Asynchronous Transfer Mode (ATM) and LocalTalk. Ethernet is popular because it strikes a good balance between speed, cost and ease of installation. These benefits, combined with wide acceptance in the computer marketplace and the ability to support virtually all popular network protocols, make Ethernet an ideal networking technology for most computer users today. The Institute for Electrical and Electronic Engineers (IEEE) defines the Ethernet standard as IEEE Standard 802.3. This standard defines rules for configuring an Ethernet network as well as specifying how elements in an Ethernet network interact with one another. By adhering to the IEEE standard, network equipment and network protocols can communicate efficiently.

**Fast Ethernet**

For Ethernet networks that need higher transmission speeds, the Fast Ethernet standard (IEEE 802.3u) has been established. This standard raises the Ethernet speed limit from 10 Megabits per second (Mbps) to 100 Mbps with only minimal changes to the existing cable structure. There are three types of Fast Ethernet: 100BASE-TX for use with level 5 UTP cable, 100BASE-FX for use with fiber-optic cable, and 100BASE-T4 which utilizes an extra two wires for use with level 3 UTP cable. The 100BASE-TX standard has become the most popular due to its close compatibility with the 10BASE-T Ethernet standard. For the network manager, the incorporation of Fast Ethernet into an existing configuration presents a host of decisions. Managers must determine the number of users in each site on the network that need the higher throughput, decide which segments of the backbone need to be reconfigured specifically for 100BASE-T and then choose the necessary hardware to connect the 100BASE-T segments with existing 10BASE-T segments. Gigabit Ethernet is a future technology that promises a migration path beyond Fast Ethernet so the next generation of networks will support even higher data transfer speeds.

| Speed (Mbit/s) | Distance (m) | Name | Standard/ Year | Description |
|---|---|---|---|---|
| 10 | 100 (nominally) | 10BASE-T | 802.3i 1990 | Runs over four wires (two twisted pairs) on a Category 3 or Category 5 cable. Star topology with an active hub or switch sits in the middle and has a port for each node. This is also the configuration used for 100BASE-T and gigabit Ethernet. Manchester coded signaling. |
| 100 | 100 | 100BASE-TX | 802.3u 1995 | 4B5B MLT-3 coded signaling, Category 5 cable copper cabling with two twisted pairs. |

| | | | | |
|---|---|---|---|---|
| 1000 | 100 | 1000BASE-T | 802.3ab 1999 | PAM-5 coded signaling. At least Category 5 cable with four twisted pairs copper cabling. Category 5 cable has since been deprecated and new installations use Category 5e. Each pair is used in both directions simultaneously. |
| | 100 | 10GBASE-T | 802.3an 2006 | THP PAM-16 coding. Uses category 6a cable. |
| | ≥30 | 40GBASE-T | 802.3bq | Under development, uses encoding from 10GBASE-T on proposed Cat 8.1.2 shielded cable. |

**Token Ring**

Token Ring is another form of network configuration which differs from Ethernet in that all messages are transferred in a unidirectional manner along the ring at all times. Data is transmitted in tokens, which are passed along the ring and viewed by each device. When a device sees a message addressed to it, that device copies the message and then marks that message as being read. As the message makes its way along the ring, it eventually gets back to the sender who now notes that the message was received by the intended device. The sender can then remove the message and free that token for use by others.

**FDDI**

FDDI (Fiber-Distributed Data Interface) is a standard for data transmission on fiber optic lines in a local area network that can extend in range up to 200 km (124 miles). The FDDI protocol is based on the token ring protocol. In addition to being large geographically, an FDDI local area network can support thousands of users.

**Layer 2-Internet Layer:**

The **Internet layer** is responsible for addressing, packaging, and routing functions. The core protocols of the **Internet layer** are **IP**, ARP, ICMP, and IGMP. The **Internet** Protocol (**IP**) is a routable protocol responsible for **IP** addressing, routing, and the fragmentation and reassembly of packets.

- The Internet Protocol (IP) is a routable protocol responsible for IP addressing, routing, and the fragmentation and reassembly of packets.
- The Address Resolution Protocol (ARP) is responsible for the resolution of the Internet layer address to the Network Interface layer address such as a hardware address.
- The Internet Control Message Protocol (ICMP) is responsible for providing diagnostic functions and reporting errors due to the unsuccessful delivery of IP packets.
- The Internet Group Management Protocol (IGMP) is responsible for the management of IP multicast groups.

**Addressing**

To send a packet from a source node to a destination node correctly through a network,the packet must contain enough information about the destination address. It is also common to include the source address, so that retransmission can be done, if necessary.The addressing scheme used for this purpose has considerable effect on routing.

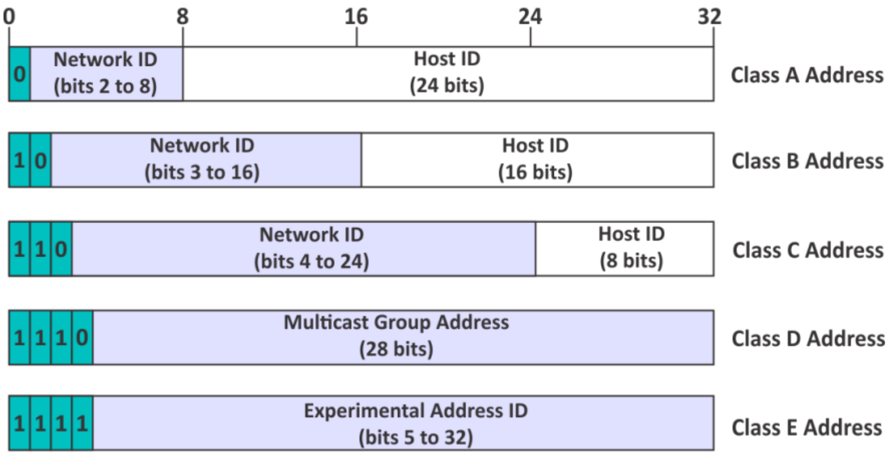
**IP Addressing**

Every host and router on the internet is provided with a unique standard form of network address, which encodes its network number and host number. The combination is unique; no two nodes have the same IP addresses.

**IPv4 Addressing:**

The IPv4 addresses are 32-bit long. The main address formats are assigned with network addresses (net id) and host address (host id) fields of different sizes. The class A format allows up to 126 networks with 16 million hosts each. Class B allows up to 16,382 networks with up to 64 K hosts each. Class C allows 2 million networks with up to 254 hosts each. The Class D is used for multicasting in which a datagram is directed to multiple hosts.

| |
|---|
| there are two prevalent notations to show an IPv4 address: binary notation and dotted decimal notation. |
| IPv4 addressing, at its inception, used the concept of classes. This architecture is called classful addressing. In classful addressing, the address space is divided into five classes: A,B,C,D,E. |




The netid determines the network address while the hostid determines the host connected to that network.

**Range of Host Addresses:**

| | | | |
|---|---|---|---|
| Class A | 1.0.0.0 | to | 126.255.255.255 |
| Class B | 128.0.0.0 | to | 191.255.255.255 |
| Class C | 192.0.0.0 | to | 223.255.255.255 |
| Class D | 224.0.0.0 | to | 239.255.255.255 |
| Class E | 240.0.0.0 | to | 254.255.255.255 |

Class D address reserved for multicast groups and Class E address reserved for future use, or Research and Development process.

**Subnetting:**
Subnetting is the strategy used to partition a single physical network into more than one smaller logical sub-networks (subnets). An IP address includes a network segment and a host segment. Subnets are designed by accepting bits from the IP address's host part and using these bits to assign a number of smaller sub-networks inside the original network.

**Subnet mask:** A Subnet mask is a 32-bit number that masks an IP address, and divides the IP address into network address and host address.

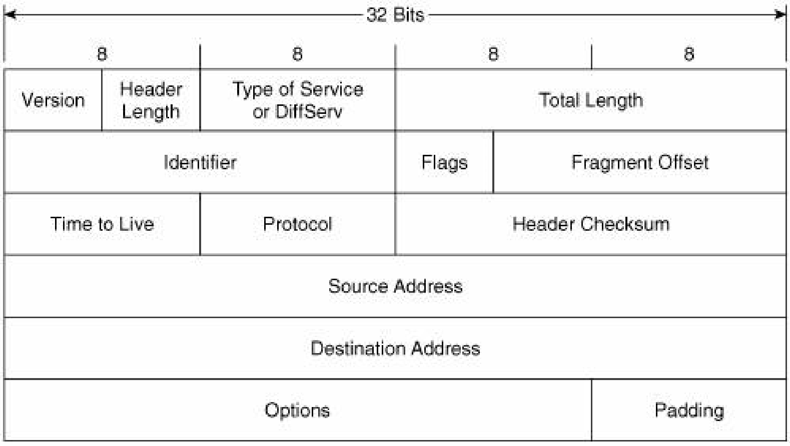
**Supernetting:** It is the process of combining several IP networks with a common network prefix. Supernetting was introduced as a solution to the problem of increasing size in routing tables. Supernetting also simplifies the routing process. For example, the subnetworks 192.60.2.0/24 and 192.60.3.0/24 can be combined in to the supernetwork denoted by 192.60.2.0/23.
In the supernet, the first 23 bits are the network part of the address and the other 9 bits are used as the host identifier. So, one address will represent several small networks and this would reduce the number of entries that should be included in the routing table. Typically, supernetting is used for class C IP addresses (addresses beginning with 192 to 223 in decimal), and most of the routing protocols support supernetting.

**Difference between Subnetting and Supernetting:**
Subnetting is the process of dividing an IP network in to sub divisions called subnets whereas, Supernetting is the process of combining several IP networks with a common network prefix.
Supernetting will reduce the number of entries in a routing table and also will simplify the routing process. In subnetting, host ID bits (for IP addresses from a single network ID) are borrowed to be used as a subnet ID, while in supernetting, bits from the network ID are borrowed to be used as the host ID.

**Classless Inter-Domain Routing:**
Short for Classless Inter-Domain Routing, an IP addressing scheme that replaces the older system based on classes A, B, and C. With CIDR, a single IP address can be used to designate many unique IP addresses.
A CIDR IP address looks like a normal IP address except that it ends with a slash followed by a number, called the IP network prefix.

*For example:* 172.200.0.0/16
CIDR addresses reduce the size of routing tables and make more IP addresses available within organizations. CIDR is also called supernetting

**IPv4 header:**




The fields in the IP header and their descriptions are

**Version—**A 4-bit field that identifies the IP version being used. The current version is 4, and this version is referred to as IPv4.

**Length—**A 4-bit field containing the length of the IP header in 32-bit increments. The minimum length of an IP header is 20 bytes, or five 32-bit increments. The maximum length of an IP header is 24 bytes, or six 32-bit increments. Therefore, the header length field should contain either 5 or 6.

**Type of Service (ToS)—**The 8-bit ToS uses 3 bits for IP Precedence, 4 bits for ToS with the last bit not being used. The 4-bit ToS field, although defined, has never been used.

**IP Precedence—** A 3-bit field used to identify the level of service a packet receives in the network.

**Differentiated Services Code Point (DSCP)—**A 6-bit field used to identify the level of service a packet receives in the network. DSCP is a 3-bit expansion of IP precedence with the elimination of the ToS bits.

**Total Length—**Specifies the length of the IP packet that includes the IP header and the user data. The length field is 2 bytes, so the maximum size of an IP packet is 216 – 1 or 65,535 bytes.

**Identifier, Flags, and Fragment Offset—**As an IP packet moves through the Internet, it might need to cross a route that cannot handle the size of the packet. The packet will be divided, or fragmented, into smaller packets and reassembled later. These fields are used to fragment and reassemble packets.

**Time to Live (TTL)—**It is possible for an IP packet to roam aimlessly around the Internet. If there is a routing problem or a routing loop, then you don't want packets to be forwarded forever. A routing loop is when a packet is continually routed through the same routers over and over. The TTL field is initially set to a number and decremented by every router that is passed through. When TTL reaches 0 the packet is discarded.

**Protocol—**In the layered protocol model, the layer that determines which application the data is from or which application the data is for is indicated using the Protocol field. This field does not identify the application, but identifies a protocol that sits above the IP layer that is used for application identification.
**Header Checksum—**A value calculated based on the contents of the IP header. Used to determine if any errors have been introduced during transmission.

**Source IP Address—**32-bit IP address of the sender.

**Destination IP Address—**32-bit IP address of the intended recipient.

**Options and Padding**—A field that varies in length from 0 to a multiple of 32-bits. If the option values are not a multiple of 32-bits, 0s are added or padded to ensure this field contains a multiple of 32 bits.

**IPv6:**
IPv6 is of 128 bits represented in 8 combinations of 4 hexa decimal numbers each, separated by a colon.
An example of an IPv6 address is: 2001:0db8:85a3:0000:0000:8a2e:0370:7334.

**Categories of IPv6 address:**

- **Unicast:** Unicast represents a single interface. A packet sent to a unicast address is delivered to the interface identified by that address.
- **Multicast:** Multicast represents a group of interfaces. A packet sent to a multicast address is delivered to all interfaces identified by that address.
- **Anycast:** Anycast identifies one or more interface. A packet sent to an anycast address is delivered to the closest member of a group, according to the routing protocols' measure of distance.

**Routing:**
Routing is the act of moving information across an inter-network from a source to a destination. Along the way, at least one intermediate node typically is encountered. It's also referred to as the process of choosing a path over which to send the packets. Routing is often contrasted with bridging, which might seem to accomplish precisely the same thing to the casual observer. The primary difference between the two is that bridging occurs at Layer 2 (the data link layer) of the OSI reference model, whereas routing occurs at Layer 3 (the network layer).
Routing protocols use metrics to evaluate what path will be the best for a packet to travel. A metric is a standard of measurement; such as path bandwidth, reliability, delay, current load on that path etc.; that is used by routing algorithms to determine the optimal path to a destination.

Routing algorithms can be classified based on the following criteria:

- Static versus Adaptive
- Single-path versus multi-path
- Intra-domain versus inter-domain
- Flat versus hierarchical
- Link-state versus distance vector
- Host-intelligent versus router-intelligent

**IPsec:**
IPsec short for IP security, a set of protocol developed by the Internet Engineering Task Force(IETF) to support secure exchange of packets at the IP layer. IPsec has been deployed widely to implement Virtual Private Networks(VPNs).
IPsec support two encryption modes: Transport and Tunnel. Transport mode encrypts only the data portion(payload) of each packets, but leaves the header untouched. The more secure Tunnel mode encrypts both the header and the payload.

**Layer 3 – Transport Layer**
It encapsulates raw data received from application layer into data segments and performs error control and flow control. It is represented by the two protocols i.e. TCP & UDP.
**TCP**(Transmission Control Protocol)- It is a connection oriented protocol. First a connection is established between the sender and the receiver and then data is sent across the network. It gives the data segments proper sequence numbers for reordering at the destination side and also the acknowledgment nos. Are given for the data packets received. So it is a reliable protocol.
**UDP**(User Datagram Protocol)It is an unreliable, connectionless protocol i.e. no reliable connection is established between sender & receiver before data transmission. It is used for client- server type requests where prompt delivery of requests-replies is more important than accurate delivery.

**Application Layer**
It enables network access to the user. Following are some of the protocols defined here:-

**File Transfer Protocol (FTP)**
File Transfer Protocol (FTP) is a TCP/IP client-server application for transfer filesbetween two remote machines through internet. A TCP connection is set up before file transfer and it persists throughout the session. It is possible to

send more than one file before disconnecting the link. A control connection is established first with a remote host before any file can be transferred.

**HTTP (Hyper Text Transfer Protocol)**
It permits the user to upload and upload webpages through browser. It is a connection less protocol.

**Telnet**
Telnet is a simple remote terminal protocol that provides a remote log-on capability,which enables a user to log on to a remote computer and behaves as if it is directlyconnected to it. The following three basic services are offered by TELNET:o It defines a network virtual terminal that provides a standard interface toremote systemso It includes a mechanism that allows the client and server to negotiate optionsfrom a standard seto It treats both ends symmetrically

**Simple Network Management Protocol (SNMP)**
Network managers use network management software that help them to locate, diagnose and rectify problems. Simple Network Management Protocol (SMTP) provides a systematic way for managing network resources. It uses transport layer protocol for communication. It allows them to monitor switches, routers and hosts. There are four components of the protocol:

- Management of systems
- Management of nodes; hosts, routers, switches
- Management of Information Base; specifies data items a host or a router must keep and the operations allowed on each (eight categories)
- Management of Protocol; specifies communication between network managementclient program a manager invokes and a network management server running on ahost or router

## 6. Internet

The Internet is generally defined as a global network connecting millions of computers. Many countries are linked into exchanges of data, news, opinions etc.
The Internet contains billions of web pages created by people and companies from around the world, making it a limitless place to locate information and entertainment. The Internet also has thousands of services that help make life more convenient. For example, many financial institutions offer online banking that enables a user to manage and view their account online.

**History of internet:**
The internet was developed in the United States by the "United States Department of Defense Advanced Research Projects Agency" (DARPA). It was first connected in October, 1969, and was called ARPANET. The World Wide Web was created at CERN in Switzerland in 1990 by a British (UK) man named Tim Berners-Lee.

**IMPORTANT ORGANIZATION:**
**Internet service provider:**
An Internet service provider (ISP) is a company that provides customers with Internet access. Data may be transmitted using several technologies, including dial-up, DSL, cable modem, wireless or dedicated high-speed interconnects.

**W3C:**
Short for World Wide Web Consortium, W3C is an organization founded by Tim Berners-Lee in 1994 to help with the development of common protocols for the unified evolution of the Web.

**Internet Architecture Board (IAB):**
Internet Architecture Board defines the architecture for the Internet. The Internet Architecture Board (IAB) purpose is to provide oversight of the architecture for the protocols and other procedures used by the Internet.

**Internet Society (ISOC):**
The Internet Society (ISOC) is mainly involved in policy, governance, technology, education & training and development of internet.

**Internet Corporation for Assigned Names and Numbers (ICANN) & Internet Assigned Numbers Authority (IANA):**

The Internet Corporation for Assigned Names and Numbers is an international non-profit corporation which is in charge of Internet Protocol (IP) address allocation (IPv4 and IPv6), Domain Names allocation (examples, omnisecu.com, msn.com, google.com) Global public Domain Name System management, DNS Root Server maintenance, Port Number allocation etc.

**Institute of Electrical and Electronics Engineers (IEEE):**
The Institute of Electrical and Electronics Engineers (IEEE) develop and maintain standards in every technology field related with electricity. The Institute of Electrical and Electronics Engineers (IEEE) develop and maintain Local Area Network (LAN) networking standards including Ethernet (IEEE 802.3 family standards) and Wireless LAN (IEEE 802.11 family standards).

**Internet Research Task Force (IRTF) & Internet Engineering Task Force (IETF):**
The Internet Research Task Force is a technology research organization which is working on focused long-term research on technical topics related to standard Internet protocols, applications, architecture and technology.
Internet Engineering Task Force is working to develop the short-term issues of network engineering protocols and standards.
Internet Engineering Task Force (IETF) develop the maintain high quality relevant technical standards, mainly network protocols. The network protocol standards are developed under a platform, called as Request for Comments (RFCs).
A Request for Comments (RFC) is a technical publication of the Internet Engineering Task Force (IETF) and the Internet Society. Request for Comments (RFCs) are mainly used to develop a network protocol, a function of a network protocol or any feature which is related with network communication. All the standard network protocols (like, HTTP, FTP, SMTP, TCP, UDP, IP etc) are defined as RFSs.

**VPN (Virtual Private Network):**

There are different technologies available for Wide Area Network (WAN) connectivity. But the main drawback of many Wide Area Network (WAN) connectivity solutions is "Cost". Think about an organization which has 100 offices all over the world. Providing Wide Area Network (WAN) connectivity using Leased Lines, for all these offices will be too costly.

If broadband internet access is available at all these 100 offices, linking all these offices using broadband internet is the most budget friendly Wide Area Network (WAN) connectivity solution. But we have a very serious problem related with security if we use public internet to connect all our 100 offices using broadband internet. Security!
Internet is a public network consisting of thousands of service providers and your organization's private Data is not much secure in a public network. We need protection for our private data against eavesdropping, tampering and we must make sure we are sending the data to exact recipient (mutual authentication).

A Virtual Private Network (VPN) is a Network Security Technology, which is used to secure private network traffic over a public network such as the Internet. A VPN ensures Data Confidentiality (privacy) and Data Integrity for network data in its journey from the source device to destination device using network security protocols like IPSec (Internet Protocol Security). IPSec (Internet Protocol Security) VPN provide Data Confidentiality by encrypting the data at the sending device and decrypting the data at receiving end. IPSec (Internet Protocol Security) VPN also provides Data Integrity (making sure that the Data is not changed while its journey) by using Hashing Algorithms like MD5 (Message Digest) and SHA (Secure Hashing Algorithm).

## 7. Some Important Networking Protocol

**User Datagram protocol (UDP)**
UDP is responsible for differentiating among multiple source and destination processes within one host. Multiplexing and demultiplexing operations are performed using the port mechanism.
UDP Datagram:
A brief description of different fields of the datagram are given below:
- Source port (16 bits): It defines the port number of the application program in the host of the sender
- Destination port (16 bits): It defines the port number of the application program in the host of the receiver
- Length: It provides a count of octets in the UDP datagram, minimum length = 8
- Checksum: It is optional, 0 in case it is not in use Characteristics of the UDP Key characteristics of UDP are given below:
- UDP provides an unreliable connectionless delivery service using IP to transport messages between two processes

- UDP messages can be lost, duplicated, delayed and can be delivered out of order
- UDP is a thin protocol, which does not add significantly to the functionality of IP
- It cannot provide reliable stream transport service

**Transmission Control Protocol (TCP)**

TCP provides a connection-oriented, full-duplex, reliable, streamed delivery service using IP to transport messages between two processes. Reliability is ensured by:

- Connection-oriented service
- Flow control using sliding window protocol
- Error detection using checksum
- Error control using go-back-N ARQ technique
- Congestion avoidance algorithms; multiplicative decrease and slow-start

**TCP Datagram**

A brief explanation of the functions of different fields are given below:

- Source port (16 bits): It defines the port number of the application program in the host of the sender
- Destination port (16 bits): It defines the port number of the application program in the host of the receiver
- Sequence number (32 bits): It conveys the receiving host which octet in this sequence comprises the first byte in the segment
- Acknowledgement number (32 bits): This specifies the sequence number of the next octet that receiver expects to receive
- HLEN (4 bits): This field specifies the number of 32-bit words present in the TCP header
- Control flag bits (6 bits): URG: Urgent pointer
- ACK: Indicates whether acknowledge field is valid
- PSH: Push the data without buffering
- RST: Resent the connection
- SYN: Synchronize sequence numbers during connection establishment
- FIN: Terminate the connection
- Window (16 bits): Specifies the size of window
- Checksum (16 bits): Checksum used for error detection.
- User pointer (16 bits): Used only when URG flag is valid
- Options: Optional 40 bytes of information

**Domain Name System**

Although IP addresses are convenient and compact way for identifying machines and are fundamental in TCP/IP, it is unsuitable for human user. Meaningful high-level symbolic names are more convenient for hum Application software permits users to use symbolic names, but the underlying network protocols require addresses. This requires the use of names with proper syntax with efficient translation mechanism. A concept known as Domain Name System (DNS) was invented for this purpose. DNS is a naming scheme that uses a hierarchical, domain-based naming scheme on a distributed database system. The basic approach is to divide the internet into several hundred top-level domains, which come in two flavors - generic and countries.

**HTTP protocol:**

The Hypertext Transfer Protocol (HTTP) is an application-level protocol for distributed, collaborative, hypermedia information systems. This is the foundation for data communication for the World Wide Web (i.e. internet) since
Standards port number of http connection is port 80.

**HTTPS**

Short for Hypertext Transfer Protocol Secure, HTTPS is a protocol which uses HTTP on a connection encrypted by transport-layer security. HTTPS is used to protect transmitted data from eavesdropping. It is the default protocol for conducting financial transactions on the web, and can protect a website's users from censorship by a government or an ISP. HTTPS port use port 443.

**Voice over IP (VoIP)**

VoIP technology allows telephone calls to be made over digital computer networks including the Internet. VoIP converts analog voice signals into digital data packets and supports real-time, two-way transmission of conversations using Internet Protocol (IP).

**Open Shortest Path First (OSPF)**

It is Interior Gateway Protocol. It is a routing protocol developed for Internet Protocol (IP) networks by the Interior Gateway Protocol (IGP) working group of the Internet Engineering Task Force (IETF). The working group was formed in 1988 to design an IGP based on the Shortest Path First (SPF) algorithm for use in the Internet.

**Routing Information Protocol (RIP)**

It is one of the most commonly used Interior Gateway Protocol on internal networks which helps a router dynamically adapt to changes of network connections by communicating information about which networks each router can reach and how far away those networks are. Although RIP is still actively used, it is generally considered to have been obsolete by Link-state routing protocol such as OSPF.

**Border Gateway Protocol (BGP)**

BGP is used to exchange routing information for the Internet and is the protocol used between Internet service providers (ISP). One of the most important characteristics of BGP is its flexibility. The protocol can connect together any internetwork of autonomous systems using an arbitrary topology

**ARP:**

Address Resolution Protocol (ARP) is one of the major protocol in the TCP/IP suit and the purpose of Address Resolution Protocol (ARP) is to resolve an IPv4 address (32 bit Logical Address) to the physical address (48 bit MAC Address). Network Applications at the Application Layer use IPv4 Address to communicate with another device. But at the Datalink layer, the addressing is MAC address (48 bit Physical Address), and this address is burned into the network card permanently. You can view your network card's hardware address by typing the command "ipconfig /all" at the command prompt (Without double quotes using Windows Operating Systems).

The purpose of Address Resolution Protocol (ARP) is to find out the MAC address of a device in your Local Area Network (LAN), for the corresponding IPv4 address, which network application is trying to communicate.

**RARP:**

The Reverse Address Resolution Protocol (RARP) is the earliest and simplest protocol designed to allow a device to obtain an IP address for use on a TCP/IP network. It is based directly on ARP and works in basically the same way, but in reverse: a device sends a request containing its hardware address and a device set up as an RARP server responds back with the device's assigned IP address.

**SIP:**

Session Initiation Protocol (SIP) is one of the most common protocols used in VoIP technology. It is an application layer protocol that works in conjunction with other application layer protocols to control multimedia communication sessions over the Internet.

**DHCP:**

Dynamic Host Configuration Protocol (DHCP) is used to dynamically (automatically) assign TCP/IP configuration parameters to network devices (IP address, Subnet Mask, Default Gateway, DNS server etc). Dynamic Host Configuration Protocol (DHCP) is described in RFC 1531. Other RFCs related with Dynamic Host Configuration Protocol (DHCP) are RFC 1534, RFC 1541, RFC 2131, and RFC 2132. DHCP is an IETF standard based on the BOOTP protocol. A computer that gets its configuration information by using Dynamic Host Configuration Protocol (DHCP) is known as a Dynamic Host Configuration Protocol (DHCP) client. DHCP clients communicate with a DHCP server to obtain IP addresses and related TCP/IP configuration information. DHCP server should be configured properly by the DHCP administrator.

Using Dynamic Host Configuration Protocol (DHCP), DHCP Clients can be configured with TCP/IP configuration values like IP Address, Subnet Mask, Default Gateway, DNS Server, DNS suffix etc.

**Stop and Wait protocol:**

This is the simplest form of flow control where a sender transmits a data frame. After receiving the frame, the receiver indicates its willingness to accept another frame by sending back an ACK frame acknowledging the frame just received. The sender must wait until it receives the ACK frame before sending the next data frame. This is sometimes referred to as ping-pong behavior, request/reply is simple to understand and easy to implement, but not very efficient. In LAN environment with fast links, this isn't much of a concern, but WAN links will spend most of their time idle, especially if several hops are required. Major drawback of Stop-and-Wait Flow Control is that only one frame can be in transmission at a time, this leads to inefficiency if propagation delay is much longer than the transmission delay. Some protocols pretty much require stop-and-wait behavior. For example, Internet's Remote Procedure Call (RPC) Protocol is used to implement subroutine calls from a program on one machine to library routines on another machine. Since most programs are single threaded, the sender has little choice but to wait for a reply before continuing the program and possibly sending another request. Error correction in Stop-and-Wait ARQ is done by keeping a copy of the sent frame and retransmitting of the frame when the timer expires.

**Sliding Window Protocol:**

In sliding window method, multiple frames are sent by sender at a time before needing an acknowledgment. Multiple frames sent by source are acknowledged by receiver using a single ACK frame. In sliding window protocols, the sender's data link layer maintains a 'sending window' which consists of a set of sequence numbers corresponding to the frames it is permitted to send. Similarly, the receiver maintains a 'receiving window' corresponding to the set of frames it is permitted to accept. The window size is dependent on the retransmission policy and it may differ in values for the receiver's and the sender's window. The sequence numbers within the sender's window represent the frames sent but as yet not acknowledged. Whenever a new packet arrives from the network layer, the upper edge of the window is advanced by one. When an acknowledgement arrives from the receiver the lower edge is advanced by one. The receiver's window corresponds to the frames that the receiver's data link layer may accept. When a frame with sequence number equal to the lower edge of the window is received, it is passed to the network layer, an acknowledgement is generated and the window is rotated by one. If however, a frame falling outside the window is received, the receiver's data link layer has two options. It may either discard this frame and all subsequent frames until the desired frame is received or it may accept these frames or buffer them until the appropriate frame is received and then pass the frames to the network layer in sequence.

**GO-back-N ARQ:**

The most popular ARQ protocol is the go-back-N ARQ, where the sender sends the frames continuously without waiting for acknowledgement. That is why it is also called as continuous ARQ. As the receiver receives the frames, it keeps on sending ACKs or a NACK, in case a frame is incorrectly received. When the sender receives a NACK, it retransmits the frame in error plus all the succeeding frames. Hence, the name of the protocol is go-back-N ARQ. If a frame is lost, the receiver sends NAK after receiving the next frame. In case there is long delay before sending the NAK, the sender will resend the lost frame after its timer times out. If the ACK frame sent by the receiver is lost, the sender resends the frames after its timer times out.

**Piggybacking:**

In practice, the link between receiver and transmitter is full duplex and usually both transmitter and receiver stations send data to each over. So, instead of sending separate acknowledgement packets, a portion (few bits) of the data frames can be used for acknowledgement. This phenomenon is known as piggybacking. The piggybacking helps in better channel utilization. Further, multi-frame acknowledgement can be done.

## PRACTICE SET

1. UDP stands for?
(a) Universal data Protocol
(b) User Datagram Protocol
(c) Universal Datagram Process
(d) None of the above
(e) User Division Protocol
2. Radio Broadcasting is an example of which type of data communication ?
(a) Simplex (b) Half Duplex
(c) Full Duplex (d) Both (b) and (c)
(e) None of the above
3. IEEE subdivided the datalink layer to provide for environments that need connectionless or connection-oriented services. One of the two layers is called?
(a) Physical (b) MAC
(c) Session (d) IP
(e)None of these
4. _____________ is used as transmission media in cable networks?
(a) Coaxial Cable (b) Optical fibre
(c) STP (d) UTP
(e)None of these
5. Which layer of the OSI model specifies the path that the data should take?
(a) Network layer (b) Transport layer
(c) Physical layer (d) Data link layer
(e)None of these
6. In TDM, slots are further divided into?
(a) Seconds (b) Frames
(c) Packets (d) None of the mentioned
(e) Routers
7. Coaxial cable is a type of?
(a) Unguided media (b) guided media
(c) connection less (d) Window media
(e) None of the above
8. Which of the following gives Sequence and Acknowledgment nos. to the data packets?
(a) TCP (b) UDP
(c) IP (d) HTTP
(e)None of these
9. A layer of glass surrounding the central fiber of glass inside a fiber optic cable is called as?
(a) Core (b) Cladding
(c) Refractive layer (d) None of the above
(e) Surrounded layer
10. Which of the following are transport layer protocols?
(a) TCP and UDP (b) ATM
(c) CISC (d) HTTP and FTP
(e)None of these
11. Which of the following protocols uses both TCP and UDP?
(a) FTP (b) SMTP
(c) Telnet (d) DNS
(e)None of these
12. HDLC is an acronym for _______.
(a) High-duplex line communication
(b) High-level data link control
(c) Half-duplex digital link combination
(d) Host double-level circuit
(e)None of these
13. Which of the following is a data link layer function?
(a) Line discipline (b) Flow control
(c) Error control (d) All of the above
(e) None of these
14. In ________ protocols, we use ________.
(a) character-oriented; byte stuffing
(b) character-oriented; bit stuffing
(c) bit-oriented; character stuffing
(d) None of these
(e) Both (b) and (c)
15. In OSPF database descriptor packet, if more database descriptor packet flows, 'M' field is set to
(a) 1 (b) 0
(c) more (d) none
(e) 2
16. When a secondary device is ready to send data it must wait for ________ frame.
(a) An ACK (b) A Poll
(c) A SEL (d) An ENQ
(e)None of these
17. Which of the following is false about a hub?
(a) It simply acts as a connector between machines in a network.
(b) It works at the data link layer
(c) It broadcasts all the data packets coming to it
(d) Hubs are devices commonly used to connect segments of a LAN
(e) None of the above
18. UDP is a ___________ protocol.
(a) Connectionless (b) Connection oriented
(c) layer 2 (d) Wireless
(e) None of the above
19. There are different ways in which a network can be connected. What is the term which best describes the physical means of connection on a network?
(a) Network Services (b) Protocols
(c) Rules (d) Transmission Media

(e)None of these

20. Which of the following is not a type of OSPF packet?
(a) Hello (b) Link-state request
(c) Link-state response (d) Link-state ACK
(e)None of these

21. Which layer translates between physical (MAC) and logical addresses?
(a) network (b) data link
(c) transport (d) presentation
(e) None of these

22. The purpose of the OSI model is:
(a) Standardize communication and facilitate vendor interoperability
(b) Allow the Department of Defence to create the DARPANET
(c) Allow IBM compatible computers to interoperate with Macintosh PCs
(d) Allow rapid development of Windows 3.11
(e) None of these

23. Which NetWare protocol works on layer 3–network layer—of the OSI model?
(a) IPX (b) NCP
(c) SPX (d) NetBIOS
(e) None of these

24. Which type of Ethernet framing is used for TCP/IP and DEC net?
(a) Ethernet 802.3 (b) Ethernet 802.2
(c) Ethernet II (d) Ethernet SNAP
(e) None of these

25. Novell's implementation of RIP updates routing tables every _________ seconds.
(a) 60 (b) 90
(c) 10 (d) 30
(e) None of these

26. The state when dedicated signals are idle are called
(a) Death period (b) Poison period
(c) Silent period (d) None of these
(e) Both (a) and (b)

27. Multiplexing can provide
(a) Efficiency (b) Privacy
(c) Anti jamming (d) Both (a) and (b)
(e) None of these

28. Correct order of the operations of OSPF
1. Hello packets
2. Propagation of link-state information and building of routing tables
3. Establishing adjacencies and synchronisation database
(a) 1-2-3 (b) 1-3-2
(c) 3-2-1 (d) 2-1-3
(e) None of these

29. Wireless networks -
(a) Can only have one type of operating system on them
(b) Have a limited range that can be furthered reduced by physical obstacles
(c) Only work with laptops
(d) Are quite difficult to secure
(e) None of these

30. Match the following.

| | |
|---|---|
| P. SMTP | 1. Application layer. |
| Q. IP | 2. Transport layer. |
| R. TCP | 3. Data Link Layer. |
| S. Token ring | 4. Network Layer. |

5.Physical layer.
(a) P-1,Q-4,R-2,S-5 (b) P-1,Q-4,R-2,S-3
(c) P-3,Q-5,R-2,S-3 (d) P-5,Q-4,R-2,S-5
(e) None of these

31. The Reverse Address Resolution Protocol (RARP) is used for
(a) Finding the MAC address from the vendor
(b) Finding the IP address from the gateway
(c) Finding the MAC address that corresponds to an IP address
(d) Finding the IP address that corresponds to a MAC address
(e) None of these

32. Ping can
(a) Measure round-trip time (b) Report packet loss
(c) Report latency (d)All of the mentioned
(e) None of these

33. This allows to check if a domain is available for registration.
(a) Domain Check (b) Domain Dossier
(c) Domain Lookup (d) None of these
(d) All of above

34. In the layer hierarchy as the data packet moves from the upper to the lower layers, headers are
(a) Added (b) Removed
(c) Rearranged (d) Modified
(e) None of these

35. Which organization has authority over interstate and international commerce in the communications field?
(a) ITU-T (b) IEEE
(c) FCC (d) ISOC
(e) None of these

36. A set of rules that governs data communication
(a) Protocols (b) Standards
(c) RFCs (d) Both (b) and (c)
(e) None of these

37. Which protocol does ping use?
(a) ICMP (b) IP
(c) HTTP (d) IGMP

(e) None of these

38. Which of the following best characterizes TCP versus UDP in most cases?
(a) TCP is less reliable and quicker
(b) TCP is slower, more reliable and requires more overhead
(c) TCP is faster, more reliable and more streamlined
(d) TCP is less reliable and connection-oriented
(e) None of these

39. The speed of a fibre optic cable is?
(a) 300BPS-10MBPS (b) 56KBPS-200MBPS
(c) 56KBPS-10GBPS (d) None of these.
(e) All of above

40. Three or more devices share a link in ________ connection
(a) Unipoint (b) Multipoint
(c) Point to point (d) All of above
(e) None of these

41. State the error control scheme used in Bluetooth.
(a) Idle ARQ. (b) Go-back-N.
(c) Selective repeat. (d) None of these
(e) Both (b) and (c)

42. Which of the following is a form of DoSattack ?
(a) Vulnerability attack (b) Bandwidth flooding
(c) Connection flooding
(d) All of the mentioned (e) None of these

43. In a network, If P is the only packet being transmitted and there was no earlier transmission, which of the following delays could be zero
(a) Propogation delay (b) Queuing delay
(c) Transmission delay (d) Processing delay
(e) None of these

44. The most common protocol for point-to-point access is the Point-to-Point Protocol (PPP), which is a____ protocol.
(a) bit-oriented (b) byte-oriented
(c) character-oriented (d)None of the above
(e) Megabyte-oriented

45. Stop-and-Wait ARQ is a special case of Go-Back-N ARQ in which the size of the send window is ____.
(a) 2 (b) 1
(c) 8 (d) 4
(e)None of the above

46. Firewalls are often configured to block
(a) UDP traffic (b) TCP traffic
(c) Both of the mentioned
(d) None of the mentioned (e) TGIP traffic

47. Which of the following is not a transition strategies?
(a) Dual stack (b) Tunnelling
(c) Conversion (d) Header translation
(e) None of these

48. The strategy used when two computers using IPv6 want to communicate with each other and the packet must pass through a region that uses IPv4 is?
(a) Dual stack (b) Header translation
(c) Conversion (d) Tunnelling
(e) None of these

49. Poll /Select line discipline requires __________ to identify the packet recipient.
(a) A timer (b) A buffer
(c) An address (d) A dedicated line
(e) None of these

50. HTTP client requests by establishing a __________ connection to aparticular port on the server.
(a) User datagram protocol
(b) Transmission control protocol
(c) Broader gateway protocol
(d) None of the mentioned
(e) Transfer protocol

51. In sliding window flow control, the frames to the left of the receiver window are frames_______.
(a) Received but not acknowledged
(b) Received and acknowledged
(c) Not received
(d) Not sent (e) None of these

52. Which of the following is NOT true with respect to a transparent bridge and a router?
(a) Both bridge and router selectively forward data packets.
(b) A bridge uses IP addresses while a router uses MAC addresses
(c) A bridge builds up its routing table by inspecting incoming packets
(d) A router can connect between a LAN and a WAN
(e) None of these

53. The function of DSLAM is
(a) Convert analog signals into digital signals
(b) Convert digital signals into analog signals
(c) Amplify digital signals
(d) Both (b) and (c)
(e) None of the mentioned

54. The data link layer
(a) Provides a well defined service interface to the network layer
(b) Deals with the transmission errors
(c) Regulates the flow of data so that receivers are not swamped by fast senders
(d)All of the above
(e) None of these

55. Choose the statement which is not applicable for cable internet access
(a) It is a shared broadcast medium
(b) It includes HFCs

(c) Cable modem connects home PC to Ethernet port
(d) Analog signal is converted to digital signal in it
(e) None of these

56. If you have to send multimedia data over SMTP it has to be encoded into?
(a) Binary
(b) Signal
(c) ASCII
(d) None of the mentioned
(e) Machine

57. What does a router do?
(a) It determines the entire route for an IP packet from source to destination.
(b) It uses ARP to route all packets.
(c) It attempts to get an IP packet one hop closer to the destination.
(d) It uses DNS to route all packets
(e) None of these

58. Which of the following is FALSE about Microwaves
(a) They have a frequency range of 1 to 300GHz
(b) They are unidirectional.
(c) The sender and receiver need not be in line of sight with each other.
(d) They are unable to penetrate the walls.
(e) None of these

59. Mostly ________ is used in wireless LAN.
(a) time division multiplexing
(b) orthogonal frequency division multiplexing
(c) space division multiplexing
(d) Frequently division multiplexing
(e)None of the mentioned

60. What is WPA?
(a) wi-fi protected access
(b) wired protected access
(c) wired process access
(d) wi-fi process access
(e) None of these

## SOLUTIONS

| | | | | | | |
|---|---|---|---|---|---|---|
| 1. (b) | 2. (a) | 3. (b) | 4. (a) | 5. (a) | 6. (b) | 7. (b) |
| 8. (a) | 9. (b) | 10. (a) | 11. (d) | 12. (b) | 13. (d) | 14. (a) |
| 15. (a) | 16. (a) | 17. (b) | 18. (a) | 19. (d) | 20. (c) | 21. (a) |
| 22. (a) | 23. (a) | 24. (c) | 25. (a) | 26. (c) | 27. (d) | 28. (b) |
| 29. (b) | 30. (a) | 31. (d) | 32. (d) | 33. (a) | 34. (a) | 35. (c) |
| 36. (a) | 37. (a) | 38. (b) | 39. (c) | 40. (b) | 41. (a) | 42. (d) |
| 43. (b) | 44. (b) | 45. (b) | 46. (a) | 47. (c) | 48. (d) | 49. (c) |
| 50. (b) | 51. (b) | 52. (b) | 53. (a) | 54. (d) | 55. (d) | 56. (c) |
| 57. (c) | 58. (c) | 59. (b) | 60. (a) | | | |

# 6. Information Security

## 1. Introduction

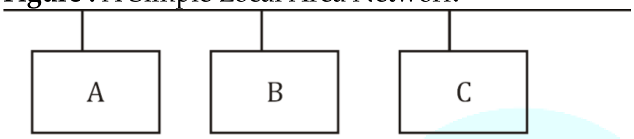
Internet: This is a word, now everyone is familiar with. Movies, books, newspapers, magazines, television programs, and practically every other sort of media imaginable has dealt with the Internet recently.

**WHAT IS THE INTERNET?**

The Internet is the world's largest network *of NETWORKS*. When you want to access the resources offered by the Internet, you don't really connect to *the* Internet; you connect to a network that is eventually connected to the *Internet BACKBONE*, a network of extremely fast (and incredibly overloaded!) network components. This is an important point: The Internet is a network of *NETWORKS* -- not a network of hosts.

A simple network can be constructed using the same protocols and such that the Internet uses without CONNECTING it to anything else. Such a basic network is shown in Figure.

**Figure :** A Simple Local Area Network

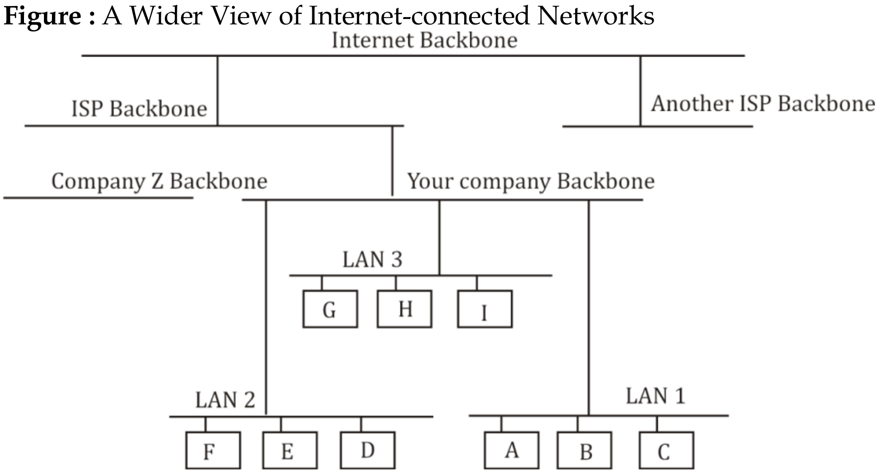



I might be allowed to put one of my hosts on one of my employer's networks. We have a number of networks, which are all connected together on a *backbone* , that is a network of our networks. Our backbone is then connected to other networks, one of which is to an *Internet Service Provider* (ISP) whose backbone is connected to other networks, one of which is the Internet backbone.

If you have a connection ``to the Internet'' through a local ISP, you are actually connecting your computer to one of their networks, which is connected to another, and so on. To use a service from my host, such as a web server, you would tell your web browser to connect to my host. Underlying services and protocols would send *packets* (small datagrams) with your query to your ISP's network, and then a network they're connected to, and so on, until it found a path to my employer's backbone, and to the exact network my host is on. My host would then respond appropriately, and the same would happen in reverse: packets would traverse all of the connections until they found their way back to your computer, and you were looking at my web page.

The following figure shows how the hosts on that network are provided connectivity to other hosts on the same LAN, within the same company, outside of the company, but in the same ISP *cloud* , and then from another ISP somewhere on the Internet.

**Figure :** A Wider View of Internet-connected Networks




The Internet is made up of a wide variety of hosts, from supercomputers to personal computers, including every imaginable type of hardware and software. How do these computers understand each other and work together?

**TCP/IP: THE LANGUAGE OF THE INTERNET**

*TCP/IP* (Transport Control Protocol/Internet Protocol) is the ``language'' of the Internet. Anything that can learn to ``speak TCP/IP'' can play on the Internet. This is functionality that occurs at the Network (IP) and Transport (TCP) layers in the ISO/OSI Reference Model. Consequently, a host that has TCP/IP functionality (such as Unix, OS/2, MacOS, or Windows NT) can easily support applications (such as Netscape's Navigator) that uses the network.

**OPEN DESIGN**

One of the most important features of TCP/IP isn't a technological one: The protocol is an ``open'' protocol, and anyone who wishes to implement it may do so freely. Engineers and scientists from all over the world participate in the *IETF* (Internet Engineering Task Force) working groups that design the protocols that make the Internet work. Their time is typically donated by their companies, and the result is work that benefits everyone.

**IP**

As noted, IP is a ``network layer'' protocol. This is the layer that allows the hosts to ``talk'' to each other. Such things as carrying datagrams, mapping the Internet address (such as 10.2.3.4) to a physical network address (such as 08:00:69:0a:ca:8f), and routing, which takes care of making sure that all the devices that have Internet connectivity can find the way to each other.

**UNDERSTANDING IP**

IP has many very important features which make it an extremely robust and flexible protocol. For our purposes, though, we're going to focus on the security of IP, or more specifically, the lack thereof.

**ATTACKS AGAINST IP**

Many attacks against IP are possible. Typically, these exploit the fact that IP does not perform a robust mechanism for *authentication*, which is proving that a packet came from where it claims it did. A packet simply claims to originate from a given address, and there isn't a way to be sure that the host that sent the packet is telling the truth. This isn't necessarily a weakness, but it is an important point, because it means that the facility of host authentication has to be provided at a higher layer on the ISO/OSI Reference Model. Today, applications that require strong host authentication (such as cryptographic applications) do this at the application layer.

**IP SPOOFING**

This is where one host claims to have the IP address of another. Since many systems (such as router access control lists) define which packets may and which packets may not pass based on the sender's IP address, this is a useful technique to an attacker: he can send packets to a host, perhaps causing it to take some sort of action.

**IP SESSION HIJACKING**

This is a relatively sophisticated attack, first described by Steve Bellovin. This is very dangerous, however, because there are now toolkits available in the underground community that allow otherwise unskilled bad-guy-wannabes to perpetrate this attack. IP Session Hijacking is an attack whereby a user's session is taken over, being in the control of the attacker. If the user was in the middle of email, the attacker is looking at the email, and then can execute any commands he wishes as the attacked user. The attacked user simply sees his session dropped, and may simply login again, perhaps not even notice that the attacker is still logged in and doing things.

In this attack, a user on host A is carrying on a session with host G. Perhaps this is a telnet session, where the user is reading his email, or using a Unix shell account from home. Somewhere in the network between A and G sits host H which is run by a naughty person. The naughty person on host H watches the traffic between A and G, and runs a tool which starts to impersonate A to G, and at the same time tells A to shut up, perhaps trying to convince it that G is no longer on the net (which might happen in the event of a crash, or major network outage). After a few seconds of this, if the attack is successful, naughty person has ``hijacked'' the session of our user. Anything that the user can do legitimately can now be done by the attacker, illegitimately. As far as G knows, nothing has happened.

This can be solved by replacing standard telnet-type applications with encrypted versions of the same thing. In this case, the attacker can still take over the session, but he'll see only ``gibberish'' because the session is encrypted. The attacker will not have the needed cryptographic key(s) to decrypt the data stream from G, and will, therefore, be unable to do anything with the session.

**TCP**

TCP is a transport-layer protocol. It needs to sit on top of a network-layer protocol, and was designed to ride atop IP. (Just as IP was designed to carry, among other things, TCP packets.) Because TCP and IP were designed together and wherever you have one, you typically have the other, the entire suite of Internet protocols are known collectively as ``TCP/IP.'' TCP itself has several important features.

**GUARANTEED PACKET DELIVERY**

Probably the most important is guaranteed packet delivery. Host A sending packets to host B expects to get acknowledgments back for each packet. If B does not send an acknowledgment within a specified amount of time, A will resend the packet.

Applications on host B will expect a data stream from a TCP session to be complete, and in order. As noted, if a packet is missing, it will be resent by A, and if packets arrive out of order, B will arrange them in proper order before passing the data to the requesting application. This is suited well toward a number of applications, such as a telnet session. A user wants to be sure every keystroke is received by the remote host, and that it gets every packet sent back, even if this means occasional slight delays in responsiveness while a lost packet is resent, or while out-of-order packets are rearranged.

It is not suited well toward other applications, such as streaming audio or video, however. In these, it doesn't really matter if a packet is lost (a lost packet in a stream of 100 won't be distinguishable) but it *does* matter if they arrive late (i.e., because of a host resending a packet presumed lost), since the data stream will be paused while the lost packet is being resent. Once the lost packet is received, it will be put in the proper slot in the data stream, and then passed up to the application.

**UDP**

*UDP* (User Datagram Protocol) is a simple transport-layer protocol. It does not provide the same features as TCP, and is thus considered ``unreliable.'' Again, although this is unsuitable for some applications, it does have much more applicability in other applications than the more reliable and robust TCP.

**LOWER OVERHEAD THAN TCP**

One of the things that makes UDP nice is its simplicity. Because it doesn't need to keep track of the sequence of packets, whether they ever made it to their destination, etc., it has lower overhead than TCP. This is another reason why it's more suited to streaming-data applications: there's less screwing around that needs to be done with making sure all the packets are there, in the right order, and that sort of thing.

## 2. Security Threats and Malwares

***What Is Network Security?***

Network security is any activity designed to protect the usability and integrity of your network and data. It includes both hardware and software technologies. Effective network security manages access to the network. It targets a variety of threats and stops them from entering or spreading on your network.

The primary goal of network security are Confidentiality, Integrity, and Availability. These
three pillars of Network Security are often represented as CIA triangle.

1. **Confidentiality.** The function of confidentiality is to protect precious business data from unauthorized persons. Confidentiality part of network security makes sure that the data is available only to the intended and authorized persons.
2. **Integrity.** This goal means maintaining and assuring the accuracy and consistency of data. The function of integrity is to make sure that the data is reliable and is not changed by unauthorized persons.
3. **Availability.** The function of availability in Network Security is to make sure that the data, network resources/services are continuously available to the legitimate users, whenever they require it.

**Security Attacks:** A useful means of classifying security attacks, used both in X.800 and RFC 2828, is in terms of passive attacks and active attacks. A passive attack attempts to learn or make use of information from the system but does not affect system resources. An active attack attempts to alter system resources or affect their operation.

**Passive Attacks:** Passive attacks are in the nature of eavesdropping on, or monitoring of, transmissions. The goal of the opponent is to obtain information that is being transmitted. Two types of passive attacks are release of message contents and traffic analysis. Passive attacks are very difficult to detect because they do not involve any alteration of the data. Typically, the message traffic is sent and received in an apparently normal fashion and neither the sender nor receiver is aware that a third party has read the messages or observed the traffic pattern. However, it is feasible to prevent the success of these attacks, usually by means of encryption. Thus, the emphasis in dealing with passive attacks is on prevention rather than detection.

**Active attacks:** They involve some modification of the data stream or the creation of a false stream and can be subdivided into four categories: masquerade, replay, modification of messages, and denial of service.

Modification of messages simply means that some portion of a legitimate message is altered, or that messages are delayed or reordered, to produce an unauthorized effect

***What is a Malware?***

Malware is a general term for a piece of software inserted into an information system to cause harm to that system or other systems, or to subvert them for use other than that intended by their owners. Malware can gain remote access to an information system, record and send data from that system to a third party without the use's permission or knowledge, conceal that the information system has been compromised, disable security measures, damage the information system, or otherwise affect the data and system integrity. Different types of malware are commonly described as viruses, worms, trojan horses, backdoors, keystroke loggers, rootkits or spyware. These terms correspond to the functionality and behaviour of the malware (e.g. a virus is self-propagating, a worm is self-replicating).

***How does malware work?***

Malware can compromise information systems due to a combination of factors that include insecure operating system design and related software vulnerabilities. Malware works by running or installing itself on an information system manually or automatically. Software may contain vulnerabilities, or "holes" in its fabric caused by faulty coding. Software may also be improperly configured, have functionality turned off, be used in a manner not compatible with suggested uses or improperly configured with other software. These are potential vulnerabilities and vectors for attack.

Once these vulnerabilities are discovered, malware can be developed to exploit them for malicious purposes before the security community has developed a “fix”, known as a patch. Malware can also compromise information systems due to non-technological factors such as poor user practices and inadequate security policies and procedures.

Malware is a program that must be triggered or somehow executed before it can infect your computer system and spread to others. Here are some examples on how malware is distributed:
a) Social network
b) Pirated software
c) Removable media
d) Emails
e) Websites

1. **Viruses:** A program or piece of code that is loaded onto your computer without your knowledge and runs against your wishes. Viruses can also replicate themselves. Viruses copy themselves to other disks to spread to other computers. They can be merely annoying or they can be vastly destructive to your files.

- **File infector viruses**
  File infector viruses infect program files. These viruses normally infect executable code, such as .com and .exe files. The can infect other files when an infected program is run from floppy, hard drive, or from the network.
- **Boot sector viruses**
  Boot sector viruses infect the system area of a disk--that is, the boot record on floppy disks and hard disks. All floppy disks and hard disks (including disks containing only data) contain a small program in the boot record that is run when the computer starts up. Boot sector viruses attach themselves to this part of the disk and activate when the user attempts to start up from the infected disk. These viruses are always memory resident in nature. Most were written for DOS, but, all PCs, regardless of the operating system, are potential targets of this type of virus.
- **Master boot record viruses**
  Master boot record viruses are memory resident viruses that infect disks in the same manner as boot sector viruses. The difference between these two virus types is where the viral code is located. Master boot record infectors normally save a legitimate copy of the master boot record in an different location. Windows NT

computers that become infected by either boot sector viruses or master boot sector viruses will not boot. This is due to the difference in how the operating system accesses its boot information, as compared to Windows 95/98

- **Multipartite viruses**
  Multipartite (also known as polypartite) viruses infect both boot records and program files. These are particularly difficult to repair. If the boot area is cleaned, but the files are not, the boot area will be reinfected. The same holds true for cleaning infected files. If the virus is not removed from the boot area, any files that you have cleaned will be reinfected. Examples of multipartite viruses include One_Half, Emperor and Anthrax.

- **Macro viruses**
  These types of viruses infect data files. They are the most common and have cost corporations the most money and time trying to repair. With the advent of Visual Basic in Microsoft's Office 97, a macro virus can be written that not only infects data files, but also can infect other files as well. Macro viruses infect Microsoft Office Word, Excel, PowerPoint and Access files.

2. **Trojan Horses:** A Trojan Horse program has the appearance of having a useful and desired function. A Trojan Horse neither replicates nor copies itself, but causes damage or compromises the security of the computer. A Trojan Horse must be sent by someone or carried by another program and may arrive in the form of a joke program or software of some sort. These are often used to capture your logins and passwords.
   Example of Trojan Horses
   - Remote access Trojans (RATs)
   - Backdoor Trojans (backdoors)
   - IRC Trojans (IRCbots)
   - Keylogging Troj

3. **Worms:** A computer worm is a self-replicating computer program. It uses a network to send copies of itself to other nodes (computers on the network) and it may do so without any user intervention. It does not need to attach itself to an existing program.
4. **Spyware:** Spyware is a type of malware installed on computers that collects information about users without their knowledge. The presence of spyware is typically hidden from the user and can be difficult to detect. Spyware programs lurk on your computer to steal important information, like your passwords and logins and other personal identification information and then send it off to someone else.
5. **Phishing:** Phishing (pronounced like the word 'fishing') is a message that tries to trick you into providing information like your social security number or bank account information or logIn and password for a web site. The message may claim that if you do not click on the link in the message and log onto a financial web site that your account will be blocked, or some other disaster.
6. **Spam:** Spam is email that you did not request and do not want. Spam is a common way to spread viruses, trojans, and the like.
7. **Adware:** Adware (short for advertising-supported software) is a type of malware that automatically delivers advertisements. Common examples of adware include pop-up ads on websites and advertisements that are displayed by software. Often times software and applications offer “free” versions that come bundled with adware.
8. **Ransomware:** Ransomware is a form of malware that essentially holds a computer system captive while demanding a ransom. The malware restricts user access to the computer either by encrypting files on the hard drive or locking down the system and displaying messages that are intended to force the user to pay the malware creator to remove the restrictions and regain access to their computer.

## 3. Botnets

A now prevalent form of malware, botnets are key tools attackers use to conduct a variety of malicious activity and cybercrime. A botnet is a group of malware infected computers also called “zombies” or bots that can be used remotely to carry out attacks against other computer systems.

Bots are generally created by finding vulnerabilities in computer systems, exploiting these vulnerabilities with malware, and inserting malware into those systems, inter alia. Botnets are maintained by malicious actors commonly referred to as “bot herders” or “bot masters” that can control the botnet remotely. The bots are then programmed and instructed by the bot herder to perform a variety of cyber-attacks, including attacks involving the further distribution and installation of malware on other information systems. Malware, when used in conjunction with botnets, allows attackers to create a self-sustaining renewable supply of Internet-connected computing resources to facilitate their crimes.

**Spam and botnets**
There is a correlation between botnets and spam due to changes in spamming techniques over the last few years. Spam commonly refers to bulk, unsolicited, unwanted and potentially harmful electronic messages. Attackers have found convenience in co-operating with spammers by using their e–mail lists to send mass quantities of spam – which often contain other malware as an e–mail attachment through botnets. It is important to note that not all spam contains malware and it is often difficult to determine how much spam directly contains malware.

## 4. Authentication and Authorization

**Authorization** is a process by which a server determines if the client has permission to use a resource or access a file and Authentication is used by a client when the client needs to know that the server is system it claims to be.
In **authentication**, the user or computer must prove its identity to the server or client. So, if we provide you with the credentials of a secured network we will be giving you authorization to access our network and while accessing the network each time you'll have authenticate your credentials to prove that you are the authorized user. Simply Authentication is the process of verifying the identity of a user by obtaining some sort of credentials and using those credentials to verify the user's identity. If the credentials are valid, the authorization process starts. Authentication process always proceeds to Authorization process.
Whereas Authorization is the process of allowing authenticated users to access the resources by checking whether the user has access rights to the system. Authorization helps you to control access rights by granting or denying specific permissions to an authenticated user.

## 5. Cryptography

Cryptography is the science of providing security for information. It has been used historically as a means of providing secure communication between individuals, government agencies, and military forces. Today, cryptography is a cornerstone of the modern security technologies used to protect information and resources on both open and closed networks. In Cryptography, **encryption** is the process of obscuring information to make it unreadable without special knowledge. This is usually done for secrecy, and typically for confidential communications. Encryption can also be used for authentication, digital signatures, digital cash e.t.c.

Cryptography is most often associated with the confidentiality of information that it provides. However, cryptography can offer the following four basic functions:

**Confidentiality**: Assurance that only authorized users can read or use confidential information. Without confidentiality, anyone with network access can use readily available tools to eavesdrop on network traffic and intercept valuable proprietary information. Intruders who gain illicit network rights and permissions can steal proprietary information that is transmitted or stored as plaintext. For example, unauthorized users might be able to intercept information, but the information is transmitted and stored as ciphertext and is useless without a decoding key that is known only to authorized users.

**Authentication:** Verification of the identity of the entities that communicate over the network. Without authentication, anyone with network access can use readily available tools to forge originating Internet Protocol (IP) addresses and impersonate others. Therefore, cryptosystems use various techniques and mechanisms to authenticate both the originators and recipients of information. For example, online entities can choose to trust communications with other online entities based on the other entities ownership of valid digital authentication credentials.

**Integrity:** Verification that the original contents of information have not been altered or corrupted. Without integrity, someone might alter information or information might become corrupted, and the alteration could be undetected. Therefore, many cryptosystems use techniques and mechanisms to verify the integrity of information.

**Nonrepudiation**: Assurance that a party in a communication cannot falsely deny that a part of the actual communication occurred. Without nonrepudiation, someone can communicate and then later either falsely deny the communications entirely or claim that it occurred at a different time. For example, without nonrepudiation, an originator of information might falsely deny being the originator of that information. Likewise, without nonrepudiation, the recipient of a communication might falsely deny having received the communication.

## 5.1 What is Encryption?

Encryption is the process of making information unreadable by unauthorized persons. The process may be manual, mechanical, or electronic. It consists of a sender and a receiver, a message (called the "plain text"), the encrypted message (called the "cipher text"), and an item called a "key." The encryption process, which transforms the plain text into the cipher text, may be thought of as a "black box." It takes inputs (the plain text and key) and produces output (the cipher text). Encryption methods can be divided into symmetric key algorithm and asymmetric key algorithm.

- **Cryptography:** process of making and using codes to secure transmission of information
- **Encryption:** converting original message into a form unreadable by unauthorized individuals
- **Cryptanalysis:** process of obtaining original message from encrypted message without knowing algorithms
- **Cryptology:** science of encryption; combines cryptography and cryptanalysis

**Ciphers**

A cipher is an algorithm for performing encryption (and the reverse, decryption) — a series of well-defined steps that can be followed as a procedure. An alternative term is encipherment. The original information is known as plaintext, and the encrypted form as ciphertext. The ciphertext message contains all the information of the plaintext message, but is not in a format readable by a human or computer without the proper mechanism to decrypt it; it should resemble random gibberish to those not intended to read it.

A symmetric-key algorithm is an algorithm for cryptography that uses the same cryptographic key to encrypt and decrypt the message. Actually, it is sufficient for it to be easy to compute the decryption key from the encryption key and vice versa. In cryptography, an asymmetric key algorithm uses a pair of different, though related, cryptographic keys to encrypt and decrypt. The two keys are related mathematically; a message encrypted by the algorithm using one key can be decrypted by the same algorithm (e.g., RSA), there are two separate keys: a public key is published and enables any sender to perform encryption, while a private key is kept secret by the receiver and enables him to perform decryption. Common asymmetric encryption algorithms available today are all based on the Diffie-Hellman key agreement algorithm.

Symmetric key ciphers can be distinguished into two types, depending on whether they work on blocks of symbols usually of a fixed size (block ciphers), or on a continuous stream of symbols (stream ciphers).

All information in a computer is reduced to a representation as 1s and 0s (or the "on" and "off" state of an electronic switch). All of the operations within the computer can be reduced to logical OR, EXCLUSIVE OR, and AND. Arithmetic in the computer (called binary arithmetic) obeys the following rules (represented by "addition" and "multiplication" tables):

| $\oplus$ | 0 | 1 |
|---|---|---|
| 0 | 0 | 1 |
| 1 | 1 | 0 |

| $\otimes$ | 0 | 1 |
|---|---|---|
| 0 | 0 | 0 |
| 1 | 0 | 1 |

The symbol $\oplus$ is called modulo 2 addition and $\otimes$ is called modulo 2 multiplication. If '1' represents a truth value of TRUE and '0' represents FALSE, then $\oplus$ is equivalent to exclusive OR in logic (XOR) and $\otimes$ is equivalent to AND. For example, A XOR B is true only if A or B is TRUE but not both. Likewise, A AND B is true only when both A and B are TRUE.

All messages, both plain text and cipher text, may be represented by strings of 1s and 0s. Because the actual method used to digitize the message is not relevant to an understanding of cryptography, the details will not be discussed here.

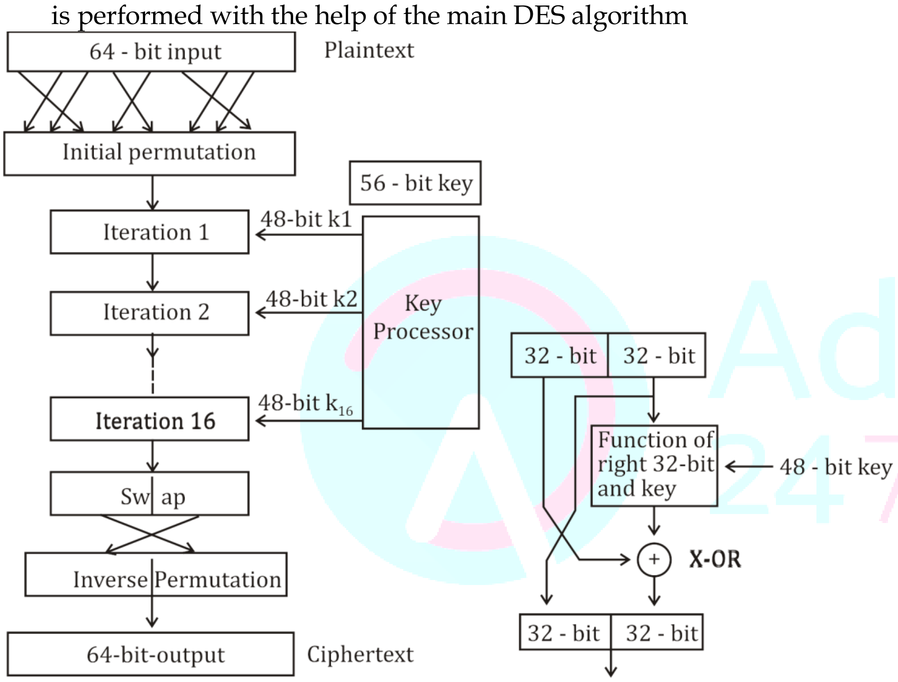
There are two main classes of algorithms:

- Stream Ciphers — operate on essentially continuous streams of plain text represented as 1s and 0s.
- Block Ciphers — operate on blocks of plain text fixed size

These two divisions overlap because a block cipher may be operated as a stream cipher. Generally speaking, stream ciphers tend to be implemented more in hardware devices while block ciphers are more suited to implementation in software to execute on a general purpose computer. Again, these guidelines are not absolute and there are a variety of reasons for choosing one method over another.

**Data Encryption Standard (DES)**
One example of the block cipher is the Data Encryption Standard (DES). Basic features of the DES algorithm are given below:

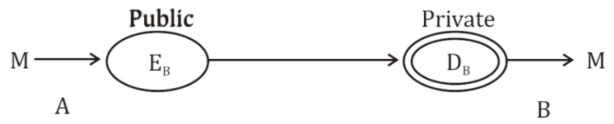
- A monoalphabetic substitution cipher using a 64-bit character
- It has 19 distinct stages
- Although the input key for DES is 64 bits long, the actual key used by DES is only 56 bits in length.
- The decryption can be done with the same password; the stages must then be carried out in reverse order.
- DES has 16 rounds, meaning the main algorithm is repeated 16 times to produce the ciphertext.
- As the number of rounds increases, the security of the algorithm increases exponentially.
- Once the key scheduling and plaintext preparation have been completed, the actual encryption or decryption is performed with the help of the main DES algorithm




64-bit Date Encryption Standard (DES)

## 5.2 Public and Private Key Cryptography

The concept of two-key cryptosystems was introduced by Diffie and Hellman in 1976. They proposed a new method of encryption called public-key encryption, wherein each user has both a public and private key, and two users can communicate knowing only each other's public keys. In a public-key system, each user A has a public enciphering transformation $E_A$, which may be registered with a public directory, and a private deciphering transformation $D_A$, which is known only to that user. The private transformation $D_A$ is described by a private key, and the public transformation $E_A$ by a public key derived from the private key by a one-way transformation. It must be computationally infeasible to determine $D_A$ from $E_A$ (or even to find a transformation equivalent to $D_A$).




**Public Key System**
In a public-key system, secrecy and authenticity are provided by the separate transformations. Suppose user A wishes to send a message M to another user B. If A knows B's public transformation $E_B$, A can transmit M to B in secrecy by sending the ciphertext $C = E_B(M)$. On receipt, B deciphers C using B's private transformation $D_B$, getting

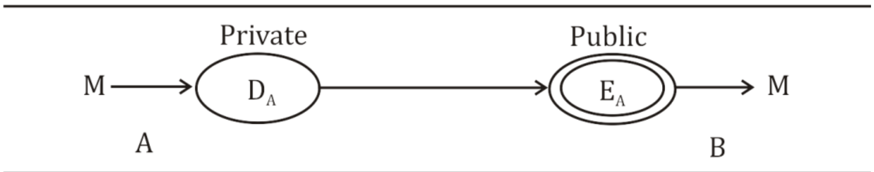
$$D_B(C) = DB(E_B(M)) = M$$

The preceding scheme does not provide authenticity because any user with access to B's public transformation could substitute another message M' for M by replacing C with C' = Es(M' ).
For authenticity, M must be transformed by A's own private transformation $D_A$. Ignoring secrecy for the moment, A sends C = $D_A$(M ) to B. On receipt, B uses A's public transformation $E_A$ to compute $E_A$(C) = $E_A$(DA(M)) = M.

Authenticity is provided because only A can apply the transformation $D_A$. Secrecy is not provided because any user with access to A's public transformation can recover M. Now, we had previously defined a transformation $D_A$ as a function from the ciphertext space Ç to the message space ɱ. To apply $D_A$ to plaintext messages, $D_A$ must instead map ɱ to Ç. Furthermore, to restore the original message, E a must be the inverse of $D_A$; that is, $E_A$ must be a function from Ç to ɱ such that $E_A(D_A(M)) = M$.

To use a public-key system for both secrecy and authenticity, the ciphertext space Ç must be equivalent to the plaintext space ɱ so that any pair of transformations $E_A$ and $D_A$ can operate on both plaintext and ciphertext messages. Furthermore, both $E_A$ and $D_A$ must be mutual inverses so that $E_A(D_A(M)) = D_A(E_A(M)) = M$

Authenticity public-key system

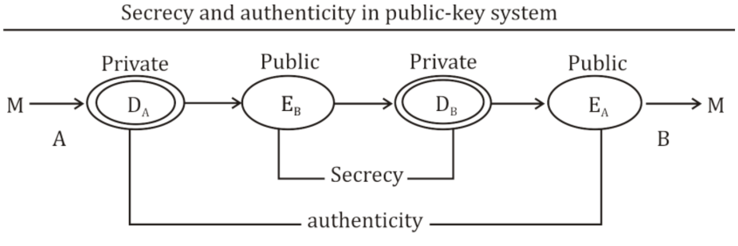



| Secrecy | Authenticity | Both |
|---|---|---|
| $E_A$: ɱ -> Ç | $D_A$ : ɱ ->Ç | $E_A$ : ɱ ->ɱ |
| $D_A$ : Ç-> ɱ | $E_A$ : Ç-> ɱ | $D_A$ : ɱ ->ɱ |
| $D_A(E_A(M)) = M$ | $E_A(D_A(M)) = M$ | $D_A(E_A(M)) = M$<br>$E_A(D_A(M)) = M$ |

To achieve both secrecy and authenticity, the sender and receiver must each apply two sets of transformations. Suppose A wishes to send a message M to B. First A's private transformation $D_A$ is applied. Then A enciphers the result using B's public enciphering transformation $E_B$, and transmits the doubly transformed message $C = E_B(D_A(M))$ to B. B recovers M by first applying B's own private deciphering transformation $D_B$, and then applying A's public transformation $E_A$ to validate its authenticity, getting

$$\begin{aligned} E_A(D_B(C)) &= E_A(D_B(E_B(D_A(M))) \\ &= E_A(D_A(M)) \\ &= M. \end{aligned}$$

Secrecy and authenticity in public-key system



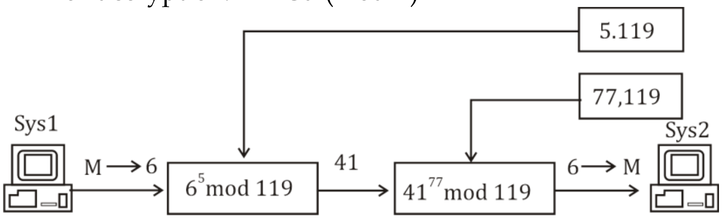

**Advantages:**
The pair of keys can be used with any other entity
The number of keys required is small

**Disadvantages:**
It is not efficient for long messages
Association between an entity and its public key must be verified

**RSA**
The most popular public-key algorithm is the RSA (named after their inventors Rivest, Shamir and Adleman) Key features of the RSA algorithm are given below:

- Public key algorithm that performs encryption as well as decryption based on number theory

- Variable key length; long for enhanced security and short for efficiency (typical 512 bytes)
- Variable block size, smaller than the key length
- The private key is a pair of numbers (d, n) and the public key is also a pair of numbers (e, n)
- Choose two large primes p and q (typically around 256 bits)
- Compute n=p×q and z= (p-1) × (q-1)
- Choose a number d relatively prime to z
- Find e such that e × d mod (p-1) × (q-1) = 1
- For encryption: C = Pe (mod n)
  For decryption: P = Cd (mod n)




The RSA public key encryption technique

**Private-Key (aka Symmetric) Encryption**

- The setting for private-key encryption is the following: two parties share a secret key and want to exchange messages privately over "insecure channel". For now, we will not worry about how they came to share the secret key.
- Kerchkhoff's Principle: Assume encryption/decryption algorithms are known to adversary. Only thing secret is the key.
- For now, "insecure channel" means that adversary can listen to all messages sent, but cannot inject/alter messages, i.e. passive rather than active.
- **Definition** A (private-key) encryption scheme consists of three algorithms (G, E, D), as follows:

– The key generation algorithm G is a randomized algorithm that returns a key $k \in \mathcal{K}$; we write $k \overset{R}{\leftarrow} G$.
– The encryption algorithm E is a randomized algorithm that takes a key $k \in \mathcal{K}$ and a plaintext (aka message) $m \in \mathcal{P}$ and outputs a ciphertext $c \in C$; we write $c \overset{R}{\leftarrow} E_k(m)$.
– The decryption algorithm D is a deterministic algorithm that takes a key $k \in \mathcal{K}$ and a cyphertext $c \in C$ and returns a plaintext $m \in P$.

The message space $\mathcal{P}$is often the set of strings of a given length. The ciphertext space $C$ does not have to equal the plaintext space. We require $D_k(E_k(m)) = m$ for all $m \in \mathcal{P}$.

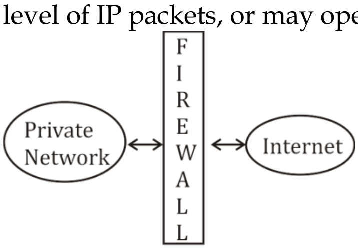
The definition describes the functionalities of the encryption scheme but does not take security into account yet.

### 5.3 Decryption

The decryption process involves converting the encrypted data back to its original form for the receiver's understanding. The same process is performed at the beginning of the encryption and decryption process (connection established) as described in the encryption part at the sender side to generate the same private position at the receiver side to eliminate the key from the cipher text.

## 6. Firewall

A firewall forms a barrier through which the traffic going in each direction must pass. A firewall security policy dictates which traffic is authorized to pass in each direction. A firewall may be designed to operate as a filter at the level of IP packets, or may operate at a higher protocol layer.




Schematic diagram of a firewall

There is no need for a firewall if each and every host of a private network is properly secured. Unfortunately, in practice the situation is different. A private network may consist of different platforms with diverse OS and applications running on them. Many of the applications were designed and developed for an ideal environment, without considering the possibility of the existence of bad guys. Moreover, most of the corporate networks are not designed for security. Therefore, it is essential to deploy a firewall to protect the vulnerable infrastructure of an enterprise.

## 6.1 Access Control Policies

**Access control policies** play an important role in the operation of a firewall. The policies can be broadly categorized in to the following four types:

**Service Control:**

• Determines the types of internet services to be accessed
• Filters traffic based on IP addresses and TCP port numbers
• Provides Proxy servers that receives and interprets service requests before it is passed on

**Direction Control:**

Determines the direction in which a particular service request may be initiated and allowed to flow through the firewall

**User Control:**

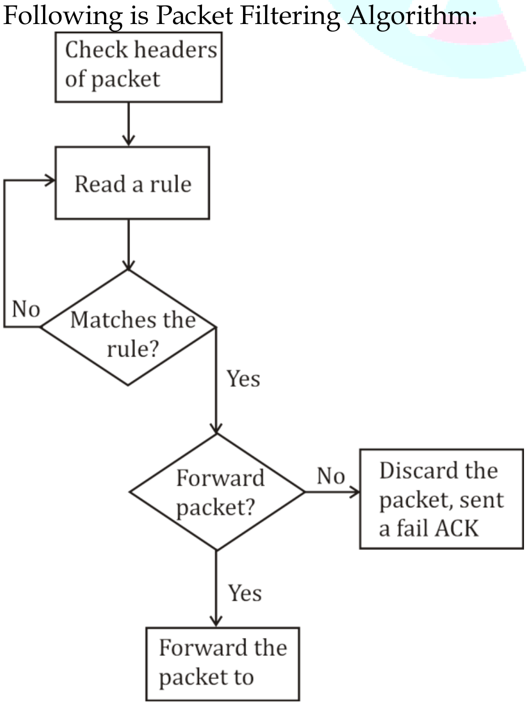
• Controls access to a service according to which user is attempting to access it
• Typically applied to the users inside the firewall perimeter
• Can be applied to the external users too by using secure authentication technique

**Behavioural Control:**

• Controls how a particular service is used
• For example, a firewall may filter email to eliminate spam
• Firewall may allow only a portion of the information on a local web server to an external user

## 6.2 Types of Firewalls

1. **Packet Filters:** Packet filtering router applies a set of rules to each incoming IP packet and then forwards or discards it. Packet filter is typically set up as a list of rules based on matches of fields in the IP or TCP header. Following is Packet Filtering Algorithm:




2. **Application-level Gateway:** Application level gateway, also called a Proxy Server acts as a relay of application level traffic. Users contact gateways using an application and the request is successful after authentication. The application gateway is service specific such as FTP, TELNET, SMTP or HTTP.

3. **Circuit Level Gateway:** Circuit-level gateway can be a standalone or a specialized system. It does not allow end-to-end TCP connection; the gateway sets up two TCP connections. Once the TCP connections are established,

the gateway relays TCP segments from one connection to the other without examining the contents. The security function determines which connections will be allowed and which are to be disallowed.

**Limitations of a Firewall**

Main limitations of a firewall system are given below:

- A firewall cannot protect against any attacks that bypass the firewall. Many organizations buy expensive firewalls but neglect numerous other back-doors into their network
- A firewall does not protect against the internal threats from traitors. An attacker may be able to break into network by completely bypassing the firewall, if he can find a "helpful" insider who can be fooled into giving access to a modem pool
- Firewalls can't protect against tunneling over most application protocols. For example, firewall cannot protect against the transfer of virus-infected programs or files

## 7. Proxies

A proxy is a program that receives traffic destined for another computer. Proxies sometimes require user authentication; they can verify that the user is allowed to connect to the destination, and then connect to the destination service on behalf of the user. When a proxy is used, the connection to the remote machine comes from the machine running the proxy instead of the original machine making the request. Because the proxy generates the connection to the remote machine, it has no problems determining which connections are real and which are spoofed; this is in contrast to stateless packet filtering firewalls.

Proxy server is an intermediary server between client and the internet. Proxy servers offers the following basic functionalities:

- Firewall and network data filtering.
- Network connection sharing
- Data caching

Proxies appear in firewalls primarily at the Transport and Application ISO network levels. In the Internet, the transport level consists of only two protocols, TCP and UDP. This small number of protocols makes writing a proxy easy—one proxy suffices for all protocols that use TCP. Contrast this with the application-level proxies (covered below), where a separate proxy is required for each service, e.g., Telnet, FTP, HTTP, SMTP, etc

Transport-level proxies have the advantage that a machine outside of the firewall cannot send packets through the firewall which claim to be a part of an established connection. Because the state of the TCP connection is known by the firewall, only packets that are a legitimate part of a communication are allowed inside the firewall.

Proxies at the application level provide the benefits of transport-level proxies, and additionally they can enforce the proper application-level protocol and prevent the abuses of the protocol by either client or server. The result is excellent security and auditing. Unfortunately, application proxies are not without their drawbacks:

- The proxy must be designed for a specific protocol. New protocols are developed frequently, requiring new proxies; if there is no proxy, there is no access.
- To use an application proxy, the client program must be changed to accommodate the proxy. The client needs to understand the proxy's authentication method and it must communicate the actual packet destination to the proxy. Because source code is not publicly available for some applications, in these cases the required changes can be made only by the application's vendor, a significant bottleneck.
- Each packet requires two trips through the complete network protocol stack which adversely affects performance. This is in contrast to packet filtering, which handles packets at the network layer.

## 8. Antivirus Software

Anti-virus software is a program or set of programs that are designed to prevent, search for, detect, and remove software viruses, and other malicious software like worms, trojans, adware, and more. If a virus infects a computer without an antivirus program, it may delete files, prevent access to files, send spam, spy on you, or perform other malicious actions. In some situations, a computer may not meet the requirements of a virus and the computer is only used to help spread the virus to other computers that may meet the requirements. Thus an antivirus software can be an effective tool for information and data security.

Regardless of the scope of coverage of various antivirus software, the underlying mechanisms of an antivirus package remain mostly the same. It actively scans files that are introduced to a system, relying on a method to identify potentially hazardous files. This is called **signature detection.**

Basically, antivirus applications maintain a database of known viruses and compare the scanned files to that database in order to find out whether the characteristics match. If they do, the file is quarantined, which is to say that it is moved to a new, safe location and renamed, so that it does not affect other files on the system.

In addition to signature detection, antivirus programs also attempt to identify suspicious behaviour on a system. This ranges from making suspicious registry entries, or adding items to a list that executes automatically upon system startup. This approach is what helps protect against encrypted viruses, or viruses that are yet to be identified.

## 9. Intrusion Detection System (IDS)

An Intrusion Detection System is used to detect all types of malicious network traffic and computer usage that can't be detected by a conventional firewall. This includes network attacks against vulnerable services, data driven attacks on applications, host based attacks such as privilege escalation, unauthorized logins and access to sensitive files, and malware (viruses, trojan horses, and worms).

An IDS is composed of the following three components:

- Sensors: - which sense the network traffic or system activity and generate events.
- Console: - to monitor events and alerts and control the sensors,
- Detection Engine: - that records events logged by the sensors in a database and uses a system of rules to generate alerts from the received security events.

There are several ways to categorize an IDS depending on the type and location of the sensors and the methodology used by the engine to generate alerts. In many simple IDS implementations, all three components are combined in a single device or appliance.

**Types of Intrusion-Detection systems**

1. **Network Intrusion Detection System: -** identifies intrusions by examining network traffic and monitors multiple hosts. Network Intrusion Detection Systems gain access to network traffic by connecting to a hub, network switch configured for port mirroring, or network tap. An example of a NIDS is Snort.

2. **Host-based Intrusion Detection System: -** consists of an agent on a host which identifies intrusions by analysing system calls, application logs, file-system modifications (binaries, password files, databases) and other host activities and state.

3. **Hybrid Intrusion Detection System: -** combines one or more approaches. Host agent data is combined with network information to form a comprehensive view of the network. An example of a Hybrid IDS is Prelude.

## 10. Vulnerability Scanners

A vulnerability scanner differs from an intrusion detection system, as vulnerability looks for static configurations and the IDS looks for transient misuse or abnormalities. A vulnerability scanner may look for a known NFS vulnerability by examining the available services and configuration on a remote system. An IDS, handling the same vulnerability, would only report the existence of the vulnerability when an attacker attempted to exploit it. Vulnerability scanners, whether network or host scanners, give the organization the opportunity to fix problems before they arise, rather than reacting to an intrusion or misuse that is already in progress. An intrusion detection system detects intrusions in progress, while a vulnerability scanner allows the organization to prevent the intrusion in the first place. Vulnerability scanners may be helpful in organizations without a good incident response capability.

**Network Vulnerability Scanner:** A network vulnerability scanner operates remotely by examining the network interface on a remote system. It will look for vulnerable services running on that remote machine, and report on a possible vulnerability. Since a network vulnerability scanner can be run from a single machine on the network, it can be installed without impacting the configuration management of other machines. Frequently, these scanners are used by auditors and security groups because they can provide an "outsider's view" of security holes in a computer or network.

**Host Vulnerability Scanner:** A host vulnerability scanner differs from a network vulnerability scanner in that it is confined entirely to the local operating system. A network vulnerability scanner requires the target machine be accessible from the network in order for it to operate; a host vulnerability scanner does not. Host vulnerability scanners are software packages that are installed on particular operating systems. Once the software is installed it can be configured to run at any time of the day or night.

## PRACTICE SET

1. Which of the following is true about Data encryption?
   (a) Data encryption ensures data safety and very important for confidential or critical data.
   (b) It protects data from being read, altered or forged while transmission.
   (c) It is used to secure the data while transmission of data over network.
   (d) None of these
   (e) All (a), (b), and (c)
2. Which of the following is an antivirus program?
   (a) Norton (b) K7
   (c) Quick Heal (d) All of the above
   (e) None of these
3. Which of the following is a default port number for Apache and most web servers?
   (a) 20 (b) 27
   (c) 80 (d) 87
   (e) None of these
4. A hash function guarantees integrity of a message. It guarantees that message has not be?
   (a) Replaced (b) Over view
   (c) Changed (d) Violated
   (e) None of these
5. MAC stands for
   (a) Message authentication code
   (b) Message arbitrary connection
   (c) Message authentication control
   (d) Message authentication cipher
   (e) Media authentication connection
6. What is the use of Digital Signatures?
   (a) It is an attachment to an electronic message used for security purpose. It is used to verify the authenticity of the sender.
   (b) It is not used to verify the authenticity of the sender.
   (c) None of these
   (d) Both (a) and (b)
   (e) It is not used for security purpose
7. A hacker contacts you my phone or email and attempts to acquire your password.
   (a) Spoofing (b) Phishing
   (c) Spamming (d) Bugging
   (e) None of these
8. Authentication is.
   (a) Verification of user identification
   (b) Verification of the data
   (c) Both (a) and (b) (d) None of the above
   (e) Verified the computer information
9. IPsec is designed to provide the security at the?
   (a) Transport layer (b) Network layer
   (c) Application layer (d) Session layer
   (e) None of these
10. WPA2 is used for security in?
    (a) Ethernet (b) Bluetooth
    (c) Wi-Fi (d) None of these
    (e) Both (a) and (b)
11. Message must be encrypted at sender site and decrypted at the?
    (a) Sender Site (b) Site
    (c) Receiver site (d) Conferencing
    (e) None of these
12. In tunnel mode IPsec protects the
    (a) Entire IP packet (b) IP header
    (c) IP payload (d) None of these
    (e) Both (a) and (c)
13. Which of the following are possible sizes of MACs?
    (i) 12 Bytes (ii) 16 Bytes
    (iii) 20 Bytes (iv) 24 Bytes
    (a) i and iii (b) ii only
    (c) ii and iii (d) ii iii and iv
    (e) i only
14. Frequency band definition and Wireless signal encoding are functions of which layer?
    (a) Physical Layer
    (b) Logic Link Control Layer
    (c) Medium Access Layer
    (d) None of these

(e) Both (a) and (b)

15. ___________ services are used to control IEEE 302.11 LAN access and confidentiality.
(a) 4 (b) 5
(c) 2 (d) 3
(e) 8

16. …………… is to protect data and passwords.
(a) Encryption (b) Authentication
(c) Authorization (d) Non-repudiation
(e) None of these

17. What do you understand by spyware?
(a) It is software that is installed without your permission.
(b) It affect a computer's performance by installing additional software, changing computer settings, changing the homepage or even completely disrupting network connection ability.
(c) Both (a) and (b)
(d) None of these
(e) Spyware is antivirus

18. Which of the following is true about SID?
(a) SID stands for security identifier.
(b) It is used to uniquely identify a user or a group
(c) SID stands for symmetric identifier
(d) Both (a) and (b)
(e) None of these

19. What is Trojan horse?
(a) It is an Antivirus
(b) It is any malicious computer program which misleads users of its true intent.
(c) Trojans may allow an attacker to access users' personal information such as banking information, passwords, or personal identity (IP address).
(d) Both (c) and (b)
(e) None of these

20. Define confidentiality and authentication Confidentiality:
(a) It means how to maintain the secrecy of message. It ensures that the information in a computer system and transmitted information are accessible only for reading by authorized person.
(b) It doesn't maintain the secrecy of the message.
(c) It doesn't ensure that the information in a computer system and transmitted information are accessible only for reading by authorized person.
(d) Both (a) and (c)
(e) None of these

21. A proxy firewall filters at the…
(a) Physical layer (b) Application layer
(c) Data link layer (d) Network layer
(e) None of these

22. A sender must not be able to deny sending a message that was sent, is known as
(a) Message Non-repudiation
(b) Message Integrity (c) Message Confidentiality
(d) Message Sending (e) None of these

23. In tunnel mode IPsec protects the
(a) Entire IP packet (b) IP header
(c) IP payload (d) None of these
(e) Both (a) and (b)

24. Network layer firewall works as a
(a) Frame filter (b) Packet filter
(c) Both (a) and (b) (d) None of these
(e) Bit filter

25. Network layer firewall has two sub-categories as
(a) Stateful firewall and stateless firewall
(b) Bit oriented firewall and byte oriented firewall
(c) Frame firewall and packet firewall
(d) None of these (e) Both (a) and (c)

26. Frame relay has error detection at the
(a) Physical layer (b) Data link layer
(c) Network layer (d) Transport layer
(e) None of these

27. PGP encrypts data by using a block cipher called
(a) International data encryption algorithm
(b) Private data encryption algorithm
(c) Internet data encryption algorithm
(d) None of these
(e) Intranet data encryption algorithm

28. When a DNS server accepts and uses incorrect information from a host that has no authority giving that information, then it is called
(a) DNS lookup (b) DNS hijacking
(c) DNS spoofing (d) None of these
(e) Both (a) and (c)

29. Which protocol consists of only 1 bit?
(a) Alert Protocol (b) Handshake Protocol
(c) Upper-Layer Protocol
(d) Change Cipher Spec Protocol
(e) None of these

30. The correct order of the of the MAC header is-
(a) MAC Control, Destination MAC Address, Source MAC Address
(b) Destination MAC Address, Source MAC Address, MAC Control
(c) Source MAC Address, Destination MAC Address, MAC Control
(d) None of these
(e) Source MAC Address, MAC Control, Destination MAC Address.

31. A hash function guarantees integrity of a message. It guarantees that message has not be
(a) Replaced (b) Over view
(c) Changed (d) Violated
(e) None of these

32. In ………………. Mode, the authentication header is inserted immediately after the IP header.
(a) Tunnel (b) Transport
(c) Authentication (d) Both A and B
(e) None of these

33. Which of the following is true regarding boot sector virus?
(a) A boot sector virus usually infects the computer by altering the boot sector program.
(b) A boot sector virus is able to infect a computer only if the virus is used to boot up the computer.
(c) A boot sector virus is malware that infects the computer storage sector where startup files are found.
(d) None of these
(e) All (a),(b),and (c)
34. One of protocols to provide security at application layer is
(a) Pretty Good Privacy (b) Handshake Protocol
(c) Alert Protocol (d) Record Protocol
(e) None of these
35. MAC address also known as.
(a) Hardware address (b) Physical address
(c) Both (a) and (b) (d) IP address
(e) None of these
36. What do you understand by VPN?
(a) VPN means Virtual Private Network, a technology that allows a secure tunnel to be created across a network such as the Internet.
(b) VPN means Virtual Public Network.
(c) It is a technology that allows a non- secure tunnel to be created across a network such as the Internet.
(d) None of these
(e) Both (a) and (b)
37. What do you understand by proxy servers and how do they protect computer networks?
(a) Proxy servers primarily prevent external users who identifying the IP addresses of an internal network. Without knowledge of the correct IP address, even the physical location of the network cannot be identified.
(b) Proxy servers can make a network virtually invisible to external users.
(c) Either (a) and (b)
(d) Both (a) and (b)
(e) None of these
38. Which of the following is a function of a network administrator?
(a) Installation of a network,
(b) Configuration of network settings,
(c) Maintenance/troubleshooting of networks.
(d) None of these
(e) All (a), (b), and (c) are functions of network administrator
39. Which of the following in not true about DHCP?
(a) DHCP is short for Dynamic Host Configuration Protocol.
(b) DHCP doesn't assign automatically an IP address to devices across the network.
(c) DHCP first checks for the next available address not yet taken by any device, then assigns this to a network device.
(d) Both (a) and (b)
(e) None of these
40. What do you understand by Malware?
(a) It is, short for malicious software, is an umbrella term used to refer to a variety of forms of hostile or intrusive software, including computer viruses, worms, etc.
(b) It is, short for malicious software, is an umbrella term used to refer to a variety of forms of hostile or intrusive software, but doesn't include computer viruses, worms, etc.
(c) It is not malicious software, is an umbrella term used to refer to a variety of forms of hostile or intrusive software, but doesn't include computer viruses, worms, etc.
(d) None of these
(e) Both (a) and (c)
41. Which one of the following is not a higher –layer SSL protocol?
(a) Alert Protocol (b) Handshake Protocol
(c) Alarm Protocol
(d) Change Cipher Spec Protocol
(e) None of these
42. Which one of the following is not a session state parameter?
(a) Master Secret (b) Cipher Spec
(c) Peer Certificate (d) Server Write Key
(e) None of these
43. Which protocol is used to convey SSL related alerts to the peer entity?
(a) Alert Protocol (b) Handshake Protocol
(c) Upper-Layer Protocol
(d) Change Cipher Spec Protocol
(e) None of these
44. What is the error (if any) in the following representation – 111.56.045.78?
(a) There should be no leading zeros
(b) We cannot have more than 4 bytes in an IPv4 address
(c) Each byte should be less than or equal to 255
(d) No error
(e) None of these
45. What is the error (if any) in the following representation – 221.34.7.8.20?
(a) There should be no leading zeros
(b) Each byte should be less than or equal to 255
(c) We cannot have more than 4 bytes in an IPv4 address
(d) No error
(e) None of these
46. The components of IP security includes …………………
(a) Authentication Header (AH)

(b) Encapsulating Security Payload (ESP)
(c) Internet key Exchange (IKE)
(d) All of the above
(e) None of these

47. State true or false.
(i) Socks are a standard for circuit level gateways.
(ii) The NAT is used for small number of the hosts in a private network.
(a) True, False (b) False, True
(c) True, True (d) False, False
(e) None of these

48. A ………………. is an extension of an enterprise's private intranet across a public Network such as the Internet across a public Network such as the Internet, creating a secure private connection.
(a) VNP (b) VPN
(c) VSN (d) VSPN
(e) None of these

49. Which of the following is true with respect to CryptoAPI?
(a) It is a set of encryption APIs.
(b) It allows developer to develop applications that work securely over non secure networks.
(c) None of these
(d) The Cryptography API has not a number of significant uses within the Enterprise Computing Model.
(e) Both (a) and (b)

50. Which Feature on a network switch can be used to prevent rogue DHCP Servers?
(a) DHCP Snooping (b) DHCP Spoofing
(c) DHCP fishing (d) None of these
(e) Both (a) and (b)

51. Which of the following is true when describing a multicast address?
(a) Packets addressed to a unicast address are delivered to a single interface.
(b) Packets are delivered to all interfaces identified by the address. This is also called a one-to-many address.
(c) Identifies multiple interfaces and is only delivered to one address. This address can also be called one-to-one-of-many.
(d) These addresses are meant for non-routing purposes, but they are almost globally unique so it is unlikely they will have an address overlap.
(e) None of these

52. Which statement(s) about IPv6 addresses are true?
(1) Leading zeros are required.
(2) Two colons are used to represent successive hexadecimal fields of zeros.
(3) Two colons are used to separate fields.
(4) A single interface will have multiple IPv6 addresses of different types.
(a) 1 and 3 (b) 2 and 4
(c) 1, 3 and 4 (d) All of the above
(e) None of these

53. How Can You Prevent A Brute Force Attack On A Windows Login Page?
(a) Setup an account lockout for specific number of attempts, so that the user account would be locked up automatically after the specified number.
(b) There is no need to prevent a Brute Force Attack on a window login page.
(c) Both (a) and (b)
(d) None of these
(e) By using only firewall

54. What is true regarding SRM?
(a) The SRM is abbreviation Security Reference Monitor
(b) It is the kernel mode component that does the actual access validation, as well as audit generation.
(c) Either (a) or (b)
(d) None of these
(e) Both (a) and (b)

55. ……………. Work To Protect The Entire Network And All Devices That Are Connected To It?
(a) NIPS (b) NAPS
(c) NICS (d) NIC
(e) None of these

56. An Attacker could alter the Mac Address in the ARP Cache so that the corresponding IP address would point to a different computer, which is known as ____?
(a) ARP poisoning. (b) ARP addressing
(c) DHCP spoofing (d) None of these
(e) Both (a) and (b)

57. In order to avoid detection some Viruses can alter how they appear. These are known as ____ Viruses?
(a) Trojan (b) Metamorphic
(c) Antivirus (d) None of these
(e) Both (a) and (c)

58. How do you remove Network Security keys?
(a) Go to your router options on your computer and it should say remove.
(b) Go to control panel and uninstall it.
(c) Either (a) or (b)
(d) None of these
(e) Go to computer properties and remove it.

59. What do you understand by IP Grabber?
(a) An IP grabber is a program that is used to assign IP address of computer.
(b) An IP grabber is a program that will find the IP address of another computer. Often used by hackers.
(c) Both (a) and (b)
(d) None of these
(e) It is used to assign MAC address

60. A ____ Virus can interrupt almost any function executed by the computer Operating System and alter it for its own malicious purposes?
(a) Worm
(b) Boot sector
(c) Resident
(d) None of these
(e) Trojan horse

## SOLUTIONS

1. (e)
2. (d)
3. (c)
4. (c)
5. (a)
6. (a)
7. (b)
8. (a)
9. (b)
10. (c)
11. (c)
12. (a)
13. (c); MACs can be 0, 16 or 20 Bytes.
14. (a)
15. (d)
16. (a)
17. (c)
18. (d)
19. (d)
20. (a)
21. (b)
22. (a)
23. (a)
24. (b)
25. (a)
26. (b)
27. (a)
28. (c)
29. (d)
30. (a); The correct order of arrangement is MAC Control, Destination MAC Address, Source MAC Address.
31. (c)
32. (a)
33. (e)
34. (a)
35. (c)
36. (a)
37. (d)
38. (e)
39. (b)
40. (a)
41. (c); Three higher –layer protocols are defined as part of SSL: The Handshake Protocol, The Change Cipher Spec Protocol and The Alert Protocol.
42. (d); Session state is defined by the following parameters – Session identifier, Peer certificate, Compression method, Cipher spec, Master secret, Is resumable. Server Write Key falls under Connection State.
43. (a)
44. (a)
45. (c); We cannot have more than 4 bytes in an IPv4 address.
46. (d)
47. (c)
48. (b)
49. (e)
50. (a)
51. (b)
52. (b)
53. (a)
54. (e)
55. (a); NIPS stands for network intrusion protection system.
56. (a)
57. (b)
58. (a)
59. (b)
60. (c)

# 7 Web Technology

## 1. Introduction

The Internet started as early as in the 1950s where the US defense organization ARPA, it stands for Advanced Research Projects Agency. This started to network a number of computers that are funded by the in a very small way. So, a few computers which are located in different paths of the country where provided with some sort of connectivity. So that they can communicate among themselves, now subsequently while it continued for some time like this in 1970s and beyond this ARPA. ARPA became to know as ARPA network advanced research project agency network. So, ARPANET started to create a standard which is basically the predecessor to the TCP standard that we have today. So, it was a preliminary protocol which through subsequent refinements and modification became finally the TCP as we see today. In 1971 the universities were added to the network, the main purpose was that many of the defense funded research used to take place in the universities and ARPANET felt universities should be part of the network. And some basic internet services like telnet and FTP were made available. Now these you will be studying later in more detail. Now using telnet you can start remote session on a different computer sitting on your own computer. And using FTP File Transfer Protocol. You can transfer a file or a group of files between two machines. These were the basic facilities which were provided at that time for communicating between machines. In 1972 the first version of electronic mail came into you can say being coming to be first email message was sent during that time.

**What is Internet?**

Internet can be defined as the network which is formed by the co-operative interconnection of a large number of computer networks. Now since internet is formed by the interconnection of a number of computer networks, sometimes it is also known as **a network of networks.** Now here there are a few interesting things that work with internet. The first and foremost is that there is no single owner of the internet. Now just unlike a network that you can see in your organization, may be your organization owns the network. In contrast internet, you cannot identify a single owner who owns or who administrates or manages the whole network. Suppose you are a member of the internet group which means you are computer is also connected to the internet it is as simple as dial up telephone line from your residence. Now in that case you are also a part of the internet.

## 2. Scripting and Markup Languages

A markup language is a formal way of annotating a document or collection of digital data using embedded encoding tags to indicate the structure of the document or data file, and the contents of its data elements. This markup provides a computer with information about how to process and display marked-up documents. Example: SGML, XML, HTML.

Scripting is the action of writing scripts using a scripting language, distinguishing neatly between programs, which are written in conventional programming language such as C, C++, java, and scripts, which are written using a different kind of language. Scripting is used as a new style of programming which allows applications to be developed much faster than traditional methods allow, and makes it possible for applications to evolve rapidly to meet changing user requirements. This style of programming frequently uses a scripting language to interconnect 'off the shelf 'components that are themselves written in conventional language. Applications built in this way are called 'glue applications', and the language is called a 'glue language'.

## 3. HTML

**What is HTML?**

Web Pages are written in HTML. HTML is short for Hypertext Markup Language.

- Hypertext is simply a piece of text that works as a link.
- Markup Language is a way of writing layout information within documents.

Basically, an HTML document is a plain text file that contains text and nothing else. When a browser opens an HTML file, the browser will look for HTML codes in the text and use them to change the layout, insert images, or create links to other pages. Since HTML documents are just text files they can be written in even the simplest text editor.
A more popular choice is to use a special HTML editor - maybe even one that puts focus on the visual result rather than the codes - a so-called **WYSIWYG** editor ("What You See Is What You Get").
Some of the most popular HTML editors, such as FrontPage or Dream weaver will let you create pages more or less as you write documents in Word or whatever text editor you're using.

## 3.1 Understanding TAGS

Basically, a computer sees an "A" as simply an "A" - whether it is bold, italic, big or small.
To tell the browser that an "A" should be bold we need to put a markup in front of the A.
Such a markup is called a Tag.
All HTML tags are enclosed in < and >.
Example: a piece of text as it appears on the screen.
This is an example of bold text.
HTML: the HTML for the above example:
This is an example of <b>bold</b> text.
As you can see, the start tag <b> indicates that whatever follows should be written in bold. The corresponding end tag </b> indicates that the browser should stop writing text in bold.

## 3.2 PAGE STRUCTURE

All normal webpages consist of a head and a body.
* Head (The head is used for text and tags that do not show directly on the page.)
* Body (The body is used for text and tags that are shown directly on the page.)
Finally, all webpages have an <html> tag at the beginning and the end, telling the browser where the document starts and where it stops.

The most basic code - the code you will use for any page you make, is shown below:

```
<html>
<head>
<!-- This section is for the title and technical info of the page. -->
</head>
<body>
<!-- This section is for all that you want to show on the page. -->
</body>
</html>
```

**HEAD SECTION**
The head section of the webpage includes all the stuff that does not show directly on the resulting page.
The <title> and </title> tags encapsulate the title of your page. The title is what shows in the top of your browser window when the page is loaded.

Right now it should say something like "Basics - Html Tutorial" on top of the window containing this text. Another thing you will often see in the head section is metatags. Metatags are used for, among other things, to improve the rankings in search engines. Quite often the head section contains javascript which is a programming language for more complex HTML pages. Finally, more and more pages contain codes for cascading style sheets (CSS). CSS is a rather new technique for optimizing the layout of major websites. Since these aspects are way out of reach at this stage we will proceed with explaining the body section.

**BODY SECTION**
The body of the document contains all that can be seen when the user loads the page.
* Text
- Formatting
- Resizing
- Layout

- Listing

* Links
    - To local pages
    - To pages at other sites
    - To bookmarks
* Images
    - Inserting images (GIF and jpg)
    - Adding a link to an image
* Backgrounds
    - Colors
    - Images
    - Fixed Image
* Tables
* Frames
* Forms
* Metatags
* Hexadecimal Colors

## 3.3 HTML Links

**How to create links in an HTML document.**
**Example:**

```
<html>
<body>
<p><a href="newpage.html">This text</a> is a link to a page on this Web site.</p>
</body>
</html>
```

**Open a link in a new browser window** (how to link to another page by opening a new window.)
**Example:**

```
<html>
<body>
<p><a href="newpage.html" target="_blank">Last Page</a> This text is a link to a page on this Web site.</p>
</body>
</html>
```

**An image as a link** ( how to use an image as a link. )
**Example:**

```
<html>
<body>
<p><a href="page.htm"><img border="0" src="buttonnext.gif" width="65" height="38">
</a> This text is a link to a page on this Web site.</p>
</body>
</html>
```

**Link to a location on the same page** (how to use a link to jump to another part of a document.)
**Example:**

```
<html>
<body>
<p><a href="#C4">See also Chapter 4.</a></p>
<h2>Chapter 1</h2>
<p>This chapter explains ba bla bla</p>
<h2>Chapter 2</h2>
<p>This chapter explains ba bla bla</p>
<h2>Chapter 3</h2>
<p>This chapter explains ba bla bla</p>
<h2><a name="C4">Chapter 4</a></h2>
<p>This chapter explains ba bla bla</p>
<h2>Chapter 5</h2>
```

```html
<p>This chapter explains ba bla bla</p>
<h2>Chapter 6</h2>
<p>This chapter explains ba bla bla</p>
<h2>Chapter 7</h2>
<p>This chapter explains ba bla bla</p>
<h2>Chapter 8</h2>
<p>This chapter explains ba bla bla</p>
<h2>Chapter 9</h2>
<p>This chapter explains ba bla bla</p>
<h2>Chapter 10</h2>
<p>This chapter explains ba bla bla</p>
<h2>Chapter 11</h2>
<p>This chapter explains ba bla bla</p>
<h2>Chapter 12</h2>
<p>This chapter explains ba bla bla</p>
<h2>Chapter 13</h2>
<p>This chapter explains ba bla bla</p>
<h2>Chapter 14</h2>
<p>This chapter explains ba bla bla</p>
<h2>Chapter 15</h2>
<p>This chapter explains ba bla bla</p>
<h2>Chapter 16</h2>
<p>This chapter explains ba bla bla</p>
<h2>Chapter 17</h2>
<p>This chapter explains ba bla bla</p>
</body>
</html>
```

**Create a mailto link** (how to link to a mail)
**Example:**

```html
<html>
<body>
<p>This is a mail link:
<a href="mailto:someone@microsoft.com?subject=Hello%20again">
Send Mail</a></p>
</body>
</html>
```

## 3.4 Tables

Tables are defined with the <table> tag. A table is divided into rows (with the <tr> tag), and each row is divided into data cells (with the <td> tag). The letters td stands for "table data," which is the content of a data cell. A data cell can contain text, images, lists, paragraphs, forms, horizontal rules, tables, etc.

```html
<table border="1">
<tr>
<td>row 1, cell 1</td>
<td>row 1, cell 2</td>
</tr>
<tr>
<td>row 2, cell 1</td>
<td>row 2, cell 2</td>
</tr>
</table>
```

**Image Tag (**and the Src Attribute)
In HTML, images are defined with the <img> tag. The <img> tag is empty, which means that it contains attributes only and it has no closing tag. To display an image on a page, you need to use the src attribute. Src stands for "source". The value of the src attribute is the URL of the image you want to display on your page.

The syntax of defining an image:
```
<img src="url" />
```

## 3.5 HTML Lists

1) Unordered list 2) Ordered list

**Unordered Lists**

An unordered list is a list of items. The list items are marked with bullets (typically small black circles).
An unordered list starts with the <ul> tag. Each list item starts with the <li> tag.

```
<ul>
<li>Coffee</li>
<li>Milk</li>
</ul>
```

Here is how it looks in a browser:

* Coffee
* Milk

Inside a list item you can put paragraphs, line breaks, images, links, other lists, etc.

**Ordered Lists**

An ordered list is also a list of items. The list items are marked with numbers.
An ordered list starts with the <ol> tag. Each list item starts with the <li> tag.

```
<ol>
<li>Coffee</li>
<li>Milk</li>
</ol>
```

Here is how it looks in a browser:

1. Coffee
2. Milk

Inside a list item you can put paragraphs, line breaks, images, links, other lists, etc.

## 3.6 Frames

With frames, you can display more than one HTML document in the same browser window. Each HTML document is called a frame, and each frame is independent of the others.
The disadvantages of using frames are:

- The web developer must keep track of more HTML documents
- It is difficult to print the entire page

**The Frameset Tag**

The <frameset> tag defines how to divide the window into frames**.** Each frameset defines a set of rows or columns. The values of the rows/columns indicate the amount of screen area each row/column will occupy

**Vertical frameset** (This example demonstrates how to make a vertical frameset with three different documents.)

```
<html>
<frameset cols="25%,50%,25%">
<frame src="frame_a.htm">
<frame src="frame_b.htm">
<frame src="frame_c.htm">
</frameset>
</html>
```

**Horizontal frameset** (This example demonstrates how to make a horizontal frameset with three different documents.)

```
<html>
<frameset rows="25%,50%,25%">
<frame src="frame_a.htm">
<frame src="frame_b.html">
<frame src="frame_c.html">
</frameset>
</html>
```

## 3.7 Other HTML Tags

<!DOCTYPE> Defines the document type

<html> Defines an html document
<body> Defines the body element
<h1> to <h6> Defines header 1 to header 6
<p> Defines a paragraph
<br> Inserts a single line break
<hr> Defines a horizontal rule
<!--...--> Defines a comment

**Text Formatting Tags**
<b> Defines bold text
<b> Defines bold text
<font> Deprecated. Defines text font, size, and color
<i> Defines italic text
<em> Defines emphasized text
<big> Defines big text
<strong> Defines strong text
<small> Defines small text
<sup> Defines superscripted text
<sub> Defines subscripted text
<bdo> Defines the direction of text display
<u> Deprecated. Defines underlined text

**Output**
<pre> Defines preformatted text
<code> Defines computer code text
<tt> Defines teletype text
<kbd> Defines keyboard text
<var> Defines a variable
<dfn> Defines a definition term
<samp> Defines sample computer code
<xmp> Deprecated. Defines preformatted



**Blocks**
<acronym> Defines an acronym
<abbr> Defines an abbreviation
<address> Defines an address element
<blockquote> Defines a long quotation
<center> Deprecated. Defines centered text
<q> Defines a short quotation
<cite> Defines a citation
<ins> Defines inserted text
<del> Defines deleted text
<s> Deprecated. Defines strikethrough text
<strike> Deprecated. Defines strikethrough text

**Links**
<a> Defines an anchor
<link> Defines a resource reference

**Frames**
<frame> Defines a sub window (a frame)
<frameset> Defines a set of frames
<noframes> Defines a noframe section
<iframe> Defines an inline sub window
(frame)
**Input**
<form> Defines a form
<input> Defines an input field
<textarea> Defines a text area

<button> Defines a push button
<select> Defines a selectable list
<optgroup> Defines an option group
<option> Defines an item in a list box
<label> Defines a label for a form control
<fieldset> Defines a fieldset
<legend> Defines a title in a fieldset
<isindex> Deprecated. Defines a single-line input field

**Lists**
<ul> Defines an unordered list
<ol> Defines an ordered list
<li> Defines a list item
<dir> Deprecated. Defines a directory list
<dl> Defines a definition list
<dt> Defines a definition term
<dd> Defines a definition description
<menu> Deprecated. Defines a menu list

**Images**
<img> Defines an image
<map> Defines an image map
<area> Defines an area inside an image map

**Tables**
<table> Defines a table
<caption> Defines a table caption
<th> Defines a table header
<tr> Defines a table row
<td> Defines a table cell
<thead> Defines a table header
<tbody> Defines a table body
<tfoot> Defines a table footer
<col> Defines attributes for table columns
<colgroup> Defines groups of table columns

**Styles**
<style> Defines a style definition
<div> Defines a section in a document
<span> Defines a section in a document

**Meta Info**
<head> Defines information about the document
<title> Defines the document title
<meta> Defines meta information
<base> Defines a base URL for all the links in a page
<basefont> Deprecated. Defines a base font

**Programming**
<script> Defines a script
<noscript> Defines a noscript section
<applet> Deprecated. Defines an applet
<object> Defines an embedded object
<param> Defines a parameter for an object

## 3.8 Forms

Usually the information supplied by the QUERY_STRING variable should come from the user pressing buttons and entering text in the HTML document. It is this information we would like to package up and send to the CGI script. Each group of buttons and text boxes is called a *form*, and forms are enclosed between the HTML tags <form> ...

</form>. You also have to tell it the URL to send the information to, and how the information is sent. The result is something like this:

```
<form action="http://www.comp.leeds.ac.uk/sam-cgi/answerme"
method="GET">
Some text in here.
It can anything except another form.
</form>
```

The action tag is the URL of the CGI script. The method GET tells it to use the QUERY_STRING method of sending information. As indicated, almost anything can go between the form tags, including text and various types of input devices. In particular we can have...

**Submit buttons**

A submit button is the input device that actually calls the URL. It has a value which is the message that appears on the button. Here is the code for a form with just a submit button in it. When you click on the submit button the URL specified in the form's action is called.

```
<form action="http://www.comp.leeds.ac.uk/cgi-bin/Perl/environment-example"
method="GET">
<input type="submit" value="Click me">
</form>
```

The result is a form which looks like this.

Click me

If you click the submit button then the URL will be called. However the QUERY_STRING variable will be null because no information was specified. The answer is to use...

**Checkboxes**

A checkbox is a simple on/off button. A checkbox has a name (its key) and a value that this key has when the box is checked. As an example, here is the HTML code for a form with a checkbox and a submit button in it.

```
<form action="http://www.comp.leeds.ac.uk/cgi-bin/Perl/environment-example"
method="GET">
<input type="checkbox" name="lights" value="on">
<input type="submit" value="Do it">
</form>
```

The result of this code is the following form

Do it

Now if the submit button is clicked when the box is checked then the information lights=on is packaged into QUERY_STRING. However if the box is not checked then no information is packaged into QUERY_STRING and it remains empty. Notice also that the checkbox does not appear with a message. This is something you have to add yourself as ordinary HTML text.

Here is example HTML code for a form with two checkboxes and a message for each.

```
<form action="http://www.comp.leeds.ac.uk/cgi-bin/Perl/environment-example"
method="GET">
<input type="checkbox" name="lights" value="on"> Lights
<input type="checkbox" name="camera" value="on"> Camera
<input type="submit" value="Do it">
</form>
```

The result of this code is the following form

Lights Camera

Do it

Click the submit button with various combinations of checked boxes and watch how the QUERY_STRING environment variable changes. If both boxes are checked then the names are separated by an & sign, as we saw earlier.

**Radio buttons**

Radio buttons are just like checkboxes except they are grouped together and only one button in the group may be selected at a time. All the buttons in a group must have the same name and each one should have a different value. You can also specify which buttons (if any) are checked initially. When the submit button is clicked the name and the value of the selected button are packaged up for QUERY_STRING.

Here is some example code for five such buttons. They are all of type radio, and are in the group named cert. The 15 button is checked initially.

```
<form action="http://www.comp.leeds.ac.uk/cgi-bin/Perl/environment-example"
method="GET">
<input type="radio" name="cert" value="u"> U
<input type="radio" name="cert" value="pg"> PG
<input type="radio" name="cert" value="12"> 12
<input type="radio" name="cert" value="15" checked> 15
<input type="radio" name="cert" value="18"> 18
<input type="submit" value="Certify">
</form>
```

The result of this HTML code is the following form.

U PG 12 15 18

Certify

Again, try this out for yourself and watch QUERY_STRING change. Notice that the value of cert for the U and PG buttons are in lowercase because this is what we specified with the value tag.

**Text boxes**

Finally we deal with text input devices. These are simply boxes into which the user can enter some text which is then packaged up under a particular name. Here is some example code for two text boxes and a submit button. The <br> tag causes a line break.

```
<form action="http://www.comp.leeds.ac.uk/cgi-bin/Perl/environment-example"
method="GET">
Director: <input type="text" name="dir"> <br>
Producer: <input type="text" name="prod">
<input type="submit" value="Fire">
</form>
```

The result is the following form

Director:

Producer:

Fire

Recall that spaces are encoded as + signs and some other characters are encoded in their hexadecimal form. Try entering signs like &, + and % in particular.

**Text areas and the POST method**

As well as allowing single-line text boxes, forms also allow multiline text areas. A text area does not use the input tag; it is a pair <textarea> ... </textarea> in its own right, with the default contents going between the two tags. A text area must still have a name, but we can also specify how many rows and columns it has.

Here is the HTML code for a 40 by 4 text box with some initial default text.

```
<form action="http://www.comp.leeds.ac.uk/cgi-bin/Perl/environment-example"
method="GET">
<textarea name="review" cols=40 rows=4>I urge you to see it.
</textarea>
<input type="submit" value="Publish">
</form>
```

This looks like

I urge you to see it.

Publish

When the submit button is clicked the contents of the entire text box is packaged up and sent as the query string. It is at this point that things can start to go wrong. The information to the query string is sent as part of the URL, but the URL can often only be so many characters long (about two hundred) before the HTTP server chokes on information overload. This isn't very likely with the examples we've seen before now, but text areas can contain potentially unlimited amounts of text and so it starts to be a danger.

The solution is to use another method to send the data.

**The POST method**

Until now we've been using the GET method to send information to the HTTP server. The GET method is the method of packaging the information into the URL and then passing it to the CGI script as the QUERY_STRING environment variable. A generally more reliable method is the POST method. This packages the information in exactly the same

way, but instead of sending it as a text string after a ? in the URL it sends it as a separate message. This message comes into the CGI script in the form of the standard input.
Once again these details needn't bother us, though, because the & read input subroutine is designed to cope with this. All we need to worry about is setting the form's method to POST, and then everything else stays the same. The resultant HTML code should look like this

```
<form action="http://www.comp.leeds.ac.uk/cgi-bin/Perl/environment-example"
method="POST">
<textarea name="review" cols=40 rows=4>I urge you to see it.
</textarea>
<input type="submit" value="Publish">
</form>
```

Notice that only the method has changed from GET to POST. Everything else remains the same. Now when the URL is accessed the query string should be empty because the information is no longer sent that way. Try it:
I urge you to see it.
Publish
The way to access the information is to write a CGI script which uses the &read_input subroutine. The script in webm/WWW/Perl/Source/ta-example is one such. Here is the HTML code for the form which accesses it, followed by the form itself.

```
<form action="http://www.comp.leeds.ac.uk/cgi-bin/Perl/ta-example"
method="POST">
<textarea name="review" cols=40 rows=4>I urge you to see it.
</textarea>
<input type="submit" value="Publish">
</form>
```

I urge you to see it.
Publish
It's interesting to note that since this text is fed straight into the HTML document being generated it's actually interpreted as HTML. For a graphic illustration of this try pasting the above HTML code into the text area and submitting that.

**What is DHTML?**

Dynamic HTML, called DHTML for short, is the name given to a set of Web development techniques that are mostly used in Web pages that have non-trivial user-input features. DHTML means manipulating the Document Object Model of an HTML document, fiddling with CSS directives in style information, and using client-side JavaScript scripting to tie everything together. DHTML is the combination of several built-in browser features in fourth-generation browsers that enable a Web page to be more dynamic. Dynamic HTML has also been described as a set of commands mixed with text that describe how a Web page should appear.

**Advantages of DHTML**

(1) DHTML makes documents dynamic. Dynamic documents:
- Allow the designer to control how the HTML displays Web pages' content.
- React and change with the actions of the visitor.
- Can exactly position any element in the window, and change that position after the document has loaded.
- Can hide and show content as needed.

(2) DHTML allows any HTML element (any object on the screen that can be controlled independently using JavaScript) in Internet Explorer to be manipulated at any time, turning plain HTML into dynamic HTML.
(3) With DHTML, changes occur entirely on the client-side ( on the user's browser).
(4) Using DHTML gives the author more control over how the page is formatted and how content is positioned on the page.

The use of "dynamic" HTML pages is only possible on the latest web browsers and sometimes the interpretation of HTML code can be different from one browser to another.

**CSS and DHTML**

Cascading Style Sheets (CSS) is a technique that allows you to describe the presentation of your HTML. In essence, it allows you to state how you want each element on your page to look. An element is a piece of HTML that represents one thing: one paragraph, one heading, one image, one list. Elements usually correspond to a particular tag and its content. When CSS styles are used, DHTML pages can work on the appearance and the content of the page independently.

## 4. XML

XML is a markup language. The mighty ones who created this acronym cheated a little, as XML stands for extensible Markup Language. XML was released in the late 90's and has since received a great amount of hype. The XML standard was created by W3C to provide an easy to use and standardized way to store self-describing data (self-describing data is data that describes both its content and its structure).

XML is nothing by itself. XML is more of a "common ground" standard. The main benefit of XML is that you can use it to take data from a program like MSSQL (Microsoft SQL), convert it into XML, and then share that XML with a slough of other programs and platforms. Each of these receiving platforms can then convert the XML into a structure the platform uses normally, and presto! You have just communicated between two platforms which are potentially very different! What makes XML truly powerful is the international acceptance it has received. Many individuals and corporations have put forth their hard work to make XML interfaces for databases, programming, office application, mobile phones and more. It is because of this hard work that the tools exist to do this conversion from whatever platform into standardized XML data or convert XML into a format used by that platform.
In the past, attempts at creating a standardized format for data that could be interpreted by many different platforms (or applications) failed miserably. XML has largely succeeded in doing this.

**The main difference between XML and HTML**
XML is not a replacement for HTML.
XML and HTML were designed with different goals:
XML was designed to describe data and to focus on what data is.
HTML was designed to display data and to focus on how data looks.
HTML is about displaying information, XML is about describing information.

**XML is extensible**
The tags used to markup HTML documents and the structure of HTML documents are predefined. The author of HTML documents can only use tags that are defined in the HTML standard.
XML allows the author to define his own tags and his own document structure.

**XML is a complement to HTML**
It is important to understand that XML is not a replacement for HTML. In the future development of the Web it is most likely that XML will be used to structure and describe the Web data, while HTML will be used to format and display the same data.

**XML in future Web development**
We have been participating in XML development since its creation. It has been amazing to see how quickly the XML standard has been developed, and how quickly a large number of software vendors have adopted the standard.

We strongly believe that XML will be as important to the future of the Web as HTML has been to the foundation of the Web. XML is the future for all data transmission and data manipulation over the Web

### 4.1 XML Syntax Rules

**All XML Elements Must Have a Closing Tag**
Example: <p>This is a paragraph</p>
**XML Tags are Case Sensitive**
Example: Opening and closing tags must be written with the same case:
<message>This is correct</message>

**XML Elements Must be Properly Nested**
Example: In XML, all elements must be properly nested within each other:
<b><i>This text is bold and italic</i></b>
**XML Documents Must Have a Root Element**
XML documents must contain one element that is the parent of all other elements. This element is called the root element.
Example:-
<root>
<child>
<subchild>.....</subchild>
</child>
</root>

**XML Attribute Values Must be Quoted**
XML elements can have attributes in name/value pairs just like in HTML.
In XML the attribute value must always be quoted. Study the two XML documents below. The first one is incorrect, the second is correct:
<note date=12/11/2007>
<to>Tove</to>
<from>Jani</from>
</note>
<note date="12/11/2007">
<to>Tove</to>
<from>Jani</from>
</note>
The error in the first document is that the date attribute in the note element is not quoted.

**Entity References**
Some characters have a special meaning in XML.
If you place a character like "<" inside an XML element, it will generate an error because the parser interprets it as the start of a new element.
This will generate an XML error:
<message>if salary < 1000 then</message>
To avoid this error, replace the "<" character with an **entity reference**:
<message>if salary < 1000 then</message>
There are 5 predefined entity references in XML:
< < less than
> > greater than
& & ampersand
' ' apostrophe
" " quotation mark
**Note:** Only the characters "<" and "&" are strictly illegal in XML. The greater than character is legal, but it is a good habit to replace it.

**Comments in XML**
The syntax for writing comments in XML is similar to that of HTML.
<!-- This is a comment -->

**White-space is Preserved in XML**
HTML truncates multiple white-space characters to one single white-space:
HTML: Hello my name is Tove
Output: Hello my name is Tove.
With XML, the white-space in a document is not truncated.

**XML Stores New Line as LF**
In Windows applications, a new line is normally stored as a pair of characters: carriage return (CR) and line feed (LF). The character pair bears some resemblance to the typewriter actions of setting a new line. In Unix applications, a new line is normally stored as a LF character. Macintosh applications use only a CR character to store a new line.

## 4.2 What is an XML Element?

An XML element is everything from (including) the element's start tag to (including) the element's end tag.
An element can contain other elements, simple text or a mixture of both. Elements can also have attributes.

```
<bookstore>
<book category="CHILDREN">
<title>Harry Potter</title>
<author>J K. Rowling</author>
<year>2005</year>
<price>29.99</price>
</book>
<book category="WEB">
<title>Learning XML</title>
<author>Erik T. Ray</author>
<year>2003</year>
<price>39.95</price>
</book>
</bookstore>
```

In the example above, <bookstore> and <book> have element contents, because they contain other elements. <author> has text content because it contains text.
In the example above only <book> has an attribute (category="CHILDREN").

## 4.3 XML Attributes

XML attributes are normally used to describe XML elements, or to provide additional information about elements. From HTML you can remember this construct: <IMG SRC="computer.gif">. In this HTML example SRC is an attribute to the IMG element.
The SRC attribute provides additional information about the element.
Attributes are always contained within the start tag of an element. Here are some examples:
HTML examples:

```
<img src="computer.gif">
<a href="demo.asp">
```

XML examples:

```
<file type="gif">
<person id="3344">
```

Usually, or most common, attributes are used to provide information that is not a part of the content of the XML document. Did you understand that? Here is another way to express that: Often attribute data is more important to the XML parser than to the reader.
Did you understand it now? Anyway, in the example above, the person id is a counter value that is irrelevant to the reader, but important to software that wants to manipulate the person element.

## 4.4 XML Validation

The following is a "Well Formed" XML document:

```
<?xml version="1.0"?>
<note>
<to>Tove</to>
<from>Jani</from>
<heading>Reminder</heading>
<body>Don't forget me this weekend!</body>
</note>
```

"Valid" XML documents
A "Valid" XML document is a "Well Formed" XML document which conforms to the rules of a Document Type Definition (DTD).
The following is the same document as above but with an added reference to a DTD:

```
<?xml version="1.0"?>
<!DOCTYPE note SYSTEM "InternalNote.dtd">
<note>
```

```
<to>Tove</to>
<from>Jani</from>
<heading>Reminder</heading>
<body>Don't forget me this weekend!</body>
</note>
```

A "Well Formed" XML document is a document that conforms to the XML syntax rules that we described in the previous chapter.

The following is a "Well Formed" XML document:

```
<?xml version="1.0"?>
<note>
<to>Tove</to>
<from>Jani</from>
<heading>Reminder</heading>
<body>Don't forget me this weekend!</body>
</note>
```

"Valid" XML documents

A "Valid" XML document is a "Well Formed" XML document which conforms to the rules of a Document Type Definition (DTD).

The following is the same document as above but with an added reference to a DTD:

```
<?xml version="1.0"?>
<!DOCTYPE note SYSTEM "InternalNote.dtd">
<note>
<to>Tove</to>
<from>Jani</from>
<heading>Reminder</heading>
<body>Don't forget me this weekend!</body>
</note>
```

## 5. Proxy

### 5.1 What is Proxy server?

Proxy server is the software installed on some network server. Proxy server sends all the requests made by the client and in turn replies without any modifications. Proxy server acts as a firewall server between the local client and the global server by providing network security, network control and caching requests. If your browser is configured to work through the proxy server then all your network traffic will go through that proxy server.

Schematic representation of a proxy server, where the computer in the middle acts as the proxy server between the other two.

### 5.2 The main purposes of proxy servers

A proxy server acts as a go-between for web browsers and the websites they access. When a web browser is configured to use a proxy server, all traffic from the web browser goes to the proxy server, which forwards traffic to the final destination. Both users and administrators have many reasons to use a proxy server.

**Privacy:** An anonymous proxy server hides the originating IP of the web browser from the website. The website will see the IP of the proxy server instead of the user.

**Caching:** Proxy servers can cache data and images locally, thus allowing the web browser to fetch the data and images faster from slow websites.

**Bypass Content Blocks:** A web browser using a proxy server on the Internet can get around websites blocked by the local network administrator. The local network thinks the browsing is going to the proxy server instead of a prohibited site.

**Web Browsing Access:** At some installations, a locally installed proxy server protected by a username and password is required for permission to web browse.

**Usage Monitoring:** A proxy server can monitor the web browsing activity of its users since all traffic is logged on the proxy server.

### 5.3 How to setup your browser for proxy access:

**SETTING UP PROXING IN FIREFOX**

1. Click **Options** from the *Tools* menu.
2. Click the *Advanced* icon and then open the *Network* tab.
3. In the *Connection* area of the tab, click the **Settings** button to bring up the *Connection Settings* dialog box.
4. Click the *Automatic proxy configuration URL* radio button to make it the selected choice.
5. Enter this URL in the box beneath the radio button:
6. It's best to copy the URL from these instructions. Make sure **not to include any spaces** before or after the URL.
7. Click **Reload** and then click **OK** to close each of the dialog boxes you've opened.
   • Select Tools => Internet Options.
   • Select the Connections tab.
   • If you are using LAN, click the LAN Settings button. If you are using Dial-up or Virtual

Private Network connection, select necessary connection and click the Settings button.

- Make sure the "automatically detect proxy settings" and "use a proxy automatic configuration script" options are not checked.
- In the "Proxy Server" area, click the check box next to *Use a proxy server ....*
- If nessesary, enable "bypass proxy server for local addresses".
- Click the "Advanced" button and set Proxy Server address (proxy IP), proxy server port.
- Click OK

**There are 3 levels of HTTP proxies:**

1. Transparent - these proxies are not anonymous. They don't hide IP-address of a client and let a web server know that you are surfing through a proxy server. The task of such proxies, as a rule, is information caching and/or support of Internet access for several computers via single connection. Such proxies are not applicable for security and privacy while surfing on net. You can use them only for network speed improvement.
2. Anonymous - these proxies don't show your real IP but change the request fields so it is very easy to detect that you are using proxy.
3. High Anonymous - these proxy servers do not pass an IP-address of a client and don't send any variables indicating that you are using proxy server to host and look like real browser.

### 5.4 What is a proxy chain?

A proxy chain is a connection of 2 or more proxy servers. Proxy way allows to work with any Internet service through a chain of SOCKS or HTTP(S) proxies. Using the program you can create proxy services with any number of proxy servers in the chain (supports HTTP, HTTPS, SOCKS4 and SOCKS5) to provide IP address security and tunnel Internet activity through proxy servers.

### 5.5 What is SOCKS?

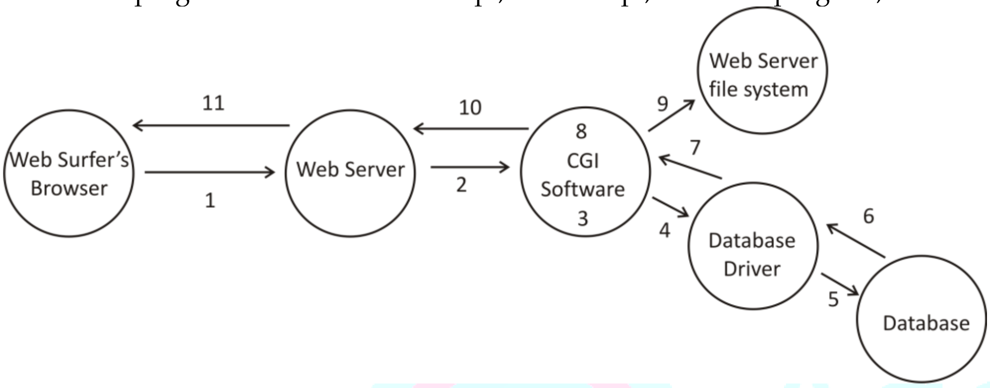
SOCKS is a networking proxy protocol that enables hosts on one side of a SOCKS server to gain full access to hosts on the other side of the SOCKS server without requiring direct IP-reachability. SOCKS is often used as a network firewall, redirecting connection requests from hosts on opposite sides of a SOCKS server. The SOCKS server authenticates and authorizes requests, establishes a proxy connection, and relays data between hosts.

## 6. Common Gateway Interface (CGI)

CGI is a standard interface by which the web server passes the client's request to a program and receives the response from that program. It is a standard for interfacing executable files with Web servers. It allows for the interactive, dynamic, flexible features that have become standard on many Web sites, such as guestbooks, counters, bulletin boards, chats, mailing lists, searches, shopping carts, surveys, and quizzes. Several newer, faster means for accomplishing these same kinds of tasks have been developed, but CGI is more flexible in a number of ways. CGI is commonly used whenever one needs a Web server to run a program in real-time, take some kind of action, and then send the results back to a user’s browser. Scripts can be written in any language that allows a file to be executed, but the most common language for CGI scripts is Perl.

## 6.1 How CGI Scripts Work

To understand the concept of CGI, lets see what happens when we click a hyper link available on a web page to browse a particular web page or URL. Your browser contacts web server using HTTP protocol and demands for the URL, i.e., web page filename. Web Server will check the URL and will look for the filename requested. If web server finds that file then it sends the file back to the browser without any further execution otherwise sends an error message indicating that you have requested a wrong file. Web browser takes response from web server and displays either the received file content or an error message in case file is not found. However, it is possible to set up HTTP server in such a way so that whenever a file in a certain directory is requested that file is not sent back; instead it is executed as a program, and whatever that program outputs as a result, that is sent back for your browser to display. This can be done by using a special functionality available in the web server and it is called Common Gateway Interface or CGI and such programs which are executed by the server to produce final result, are called CGI scripts. These CGI programs can be a PERL Script, Shell Script, C or C++ program, etc.




1. The Web surfer fills out a form and clicks, "Submit." The information in the form is sent over the Internet to the Web server.
2. The Web server "grabs" the information from the form and passes it to the CGI software.
3. The CGI software performs whatever validation of this information that is required. For instance, it might check to see if an e-mail address is valid. If this is a database program, the CGI software prepares a database statement to either add, edit, or delete information from the database.
4. The CGI software then executes the prepared database statement, which is passed to the database driver.
5. The database driver acts as a middleman and performs the requested action on the database itself.
6. The results of the database action are then passed back to the database driver.
7. The database driver sends the information from the database to the CGI software.
8. The CGI software takes the information from the database and manipulates it into the format that is desired.
9. If any static HTML pages need to be created, the CGI program accesses the Web server computer's file system and reads, writes, and/or edits files.
10. The CGI software then sends the result it wants the Web surfer's browser to see back to the Web server.
11. The Web server sends the result it got from the CGI software back to the Web surfer's browser.

**CGI Implementation**

Many applications can be implemented using other means as well (e.g., server-side JavaScript, PHP, ACGI, VRML, DHTML), but many of these other means developed after CGI. CGI, then, has become a standard, and many programmers prefer simply to "tweak" their old CGI scripts for new purposes, instead of starting from scratch with the newer languages. Also, CGI is more versatile in many ways. A traditional CGI application using Perl, for instance, can be run on a large number of platforms with a wide variety of Web servers. A programmer using server-side JavaScript, however, would be limited to Netscape Enterprise Server. CGI has its disadvantages though. Many of the newer languages developed in response to CGI being slow, so they are significantly faster. An informal study of the performance of CGI versus server-side JavaScript. Also, there are significant security issues with CGI. Since a file that uses CGI is executable, it is equivalent to letting anyone in the world run a program on your machine. Obviously, this is not the safest thing to do. For this reason, many Web hosts do not allow users to run CGI scripts. In this case, though, you can have your CGI applications hosted for you remotely.

**First CGI Program**

Here is a simple link which is linked to a CGI script called hello.cgi. This file has been kept in /cgibin/directory and it has the following content.

```
#!/usr/bin/perl
print "Content-type:text/html\r\n\r\n";
print '<html>';
print '<head>';
print '<title>Hello Word - First CGI Program</title>';
print '</head>';
print '<body>';
print '<h2>Hello Word! This is my first CGI program</h2>';
print '</body>';
print '</html>';
1;
```

Now if you click hello.cgi link then request goes to web server who search for hello.cgi in /cgi-bin directory, execute it and whatever result got generated, web server sends that result back to the web browser, which is as follows −

```
Hello Word! This is my first CGI program
```

This hello.cgi script is a simple Perl script which is writing its output on STDOUT file, i.e., screen. There is one important and extra feature available which is first line to be printed Contenttype:text/html\r\n\r\n. This line is sent back to the browser and specifies the content type to be displayed on the browser screen.

The very first line Content-type:text/html\r\n\r\n is a part of HTTP header, which is sent to the browser so that browser can understand the incoming content from server side. All the HTTP header will be in the following form –

```
HTTP Field Name: Field Content
For Example −
Content-type:text/html\r\n\r\n
```

| HEADER | DESCRIPTION |
|---|---|
| HTTP_ACCEPT | The list of MIME types a client can accept, for example, "image/gif, image/xxbitmap, image/jpeg, image/pjpeg, image/png, */*". |
| HTTP_ACCEPT_CHARSET | A list of character sets that the client can accept, for example, "iso88591,*,utf8". |
| HTTP_ACCEPT_ENCODING | A list of character coding types the client can accept, for example, "gzip". |
| HTTP_ACCEPT_LANGUAGE | The languages which the client can accept, for example, "en". |
| HTTP_AUTHORIZATION | The authorization data of an HTTP authentication, if any. See AUTH_TYPE REMOTE_USER above. |
| HTTP_CACHE_CONTROL | Set if a request can be cached by the server. |
| HTTP_CONNECTION | The connection type, for example, "Keep-alive" if the connection is desired to be persistent. |
| HTTP_COOKIE | The cookie or cookies transmitted by the client. The third-party CGI::Cookie module is useful for processing this variable to extract the cookies it contains. |
| HTTP_HOST | The name of the server requested by the client (this can be useful on a system with many virtual hosts in operation). |
| HTTP_REFERER | The URL of the page from which this page was accessed. |
| HTTP_USER_AGENT | The user agent (client or browser) that send the request, for example, "Mozilla/4.72 [en] (X11; I; Linux 2.2.9 i686)". Note that user agents often pretend to be other agents to work with web sites that treat particular agents differently. |
| HTTP_VIA | Information about which proxy cache or caches were used for making this request. |

A client is free to send any headers it likes (including no headers at all), and further revisions of the HTTP protocol may add more variables to this list. The server may also set its own variables, especially if additional functionality has been enabled.

### 6.2 CGI and Perl

You will often see the term "Perl" used with the term "CGI". The two are NOT the same. CGI programs, or scripts, can be written in a variety of computer languages, including C. CGI is the process by which scripts are run. Perl is the most common language used for writing CGI scripts, and for very good reason. (See the Messin' around with Perl section below).

CGI allows for the interactive, dynamic, flexible features that have become standard on many Web sites, such as guestbooks, counters, bulletin boards, chats, mailing lists, searches, shopping carts, surveys, and quizzes. Several newer, faster means for accomplishing these same kinds of tasks have been developed, but CGI is more flexible in a number of ways. CGI is commonly used whenever one needs a Web server to run a program in real-time, take some kind of action, and then send the results back to a user's browser. Scripts can be written in any language that allows a file to be executed, but the most common language for CGI scripts is Perl.

Because Perl has its' roots in UNIX, many people think that Perl CGI scripts cannot be used on Windows NT Web servers. Not true! Perl CGI scripts can not only run on UNIX and NT servers, but with a little tweaking for "AppleScript", many can run on Macintosh servers as well.

Perl is by far the most widely used language for CGI programming! It contains many powerful features, and is very easy for the novice programmer to learn. The advantages of Perl include:

- It is highly portable and readily available.
- It contains extremely powerful string manipulation operators, as well as functions to deal with binary data.
- It contains very simple and concise constructs.
- It makes calling shell commands very easy, and provides some useful equivalents of certain UNIX system functions.
- There are numerous extensions built on top of Perl for specialized functions.

## PRACTICE SET

1. There are how many Heading Tags?
   (a) 3 (b) 4
   (c) 5 (d) 6
   (e) 7
2. Which of the following are attributes of Font Tag?
   (a) Font (b) color
   (c) size (d) All (a),(b) and (c)
   (e) None
3. HTML is a subset of ______
   (a) SGMT (b) SGML
   (c) SGMD (d) STML
   (e) NONE
4. HTML tags are surrounded by ______
   (a) Square Brackets (b) Curly Brackets
   (c) Angle Brackets (d) None of the above
   (e) None
5. Which of the following tag is used to mark a beginning of paragraph?
   (a) <TD> (b) <BR>
   (c) <P> (d) <TR>
   (e) <PS>
6. Correct HTML tag for the largest heading is
   (a) <head> (b) <h6>
   (c) <heading> (d) <h1>
   (e) <h2>
7. From which tag descriptive list starts?
   (a) <LL> (b) <DD>
   (c) <DL> (d) <DS>
   (e) <TG>
8. HTML stands for?
   (a) Hyper Text Markup Language
   (b) High Text Markup Language
   (c) Hyper Tabular Markup Language
   (d) Hyper Text Mapeup Language
   (e) None of the above
9. www is based on which model?
   (a) Local server (b) client server
   (c) 3-tier (d) 2-tier system
   (e) none
10. Web pages starts with which of the following tag?
    (a) <Body> (b) <Title>
    (c) <HTML> (d) <Forum>

(e) None of the above

11. The ______ method requests data from a specified source.
(a)GET (b)POST
(c)Both (1) & (2) (d)None of the above
(e)All of these

12. <Base> tag is designed to appear only between
(a) <HEAD> (b) <TITLE>
(c) <BODY> (d) <FORM>
(e) <CAPTION>

13. How can you open a link in a new browser window?
(a) < a href = "url" target = "new">
(b) <a href = "url" target= "-blank">
(c) <a href = "url".new>
(d) <a href = "url" target ="open">
(e) None of the above

14. A much better approach to establish the base URL is to use
(a) BASE element (b) HEAD element
(c) both (a) and (b) (d) BODY Element
(e) None of these

15. The tag used to create a new list item and also include a hyperlink is
(a) <LI> (b) <DL>
(c) <DD> (d) <UL>
(e) <Lh>

16. Which of the tag is used to creates a number list?
(a) <LI> (b) <OL>
(c) <LI> and <OL> (d) <IL>
(e) None of the above

17. <INPUT> is
(a) Format tag (b) Empty tag
(c) Input tag (d) Both (a) and (b)
(e) None of these

18. The latest HTML standard is
(a) XML (b) SGML
(c) HTML 4.0 (d) STML
(e) HTML 5.0

19. The tag used to create a hypertext relationship between current document and another URL is
(a) <A> (b) <Link>
(c) <Isindex> (d) <Tag>
(e) None of these

20. Markup tags tell the web browser
(a)How to organise the page
(b)How to display the page
(c) How to display message box on page
(d)None of these
(e) Both (a) and (b)

21. The body tag usually used after
(a) Title tag (b) HEAD tag
(c) EM tag (d) FORM tag
(e) None of the above

22. Main container for <TR>, <TD> and <TH> is
(a) <TABLE> (b) <GROUP>
(c) <DATA> (d) <Attribute>
(e) None of these

23. Symbol which is not used at the beginning of the HREF text is
(a) All (b),(c), and (d) (b) $
(c) & (d) ^
(e) None of these

24. How can you make an e-mail link?
(a) <mail href +"xxx@y.com">
(b) <a href ="mail to: xxx@y.com">
(c) <a href = "xxx@y.com">
(d) Both (b) and (c)
(e) None of these

25. Which tag creates a number/order list?
(a) <UL> (b) <OL>
(c) <OT> (d) <TO>
(e) None of these

26. Which of the following is a Valid Name?
(a) <_person> (b) <123 person>
(c) Both (a) and (b) (d) <person 123>
(e) None of the these

27. The text inside the <TEXT AREA> tag works like
(a) <P> formatted text (b) <T> formatted text
(c) <PRE> formatted text
(d) <DI>formatted text (e) None of these

28. Symbol used at the beginning of the HREF text is
(a) # (b) $
(c) & (d) ^
(e) None of these

29. The web standard allows programmers on many different computer platforms to dispersed format and display the information server. These programs are called
(a) Web Browsers (b) HTML
(c) Internet Explorer (d) SDML
(e) None of these

30. URL stands for
(a) unique reference label
(b) uniform reference label
(c) uniform resource locator
(d) unique resource locator
(e) None of the above

31. Dynamic web page
(a)is same every time whenever it displays
(b) generates on demand by a program or a request from browser
(c) both (a) and (b)
(d) generate pre demand by a program or a request from browser
(e) None of these

32. Common gateway interface is used to
(a) generate executable files from web content by web server
(b) generate web pages (c) stream videos

(d) generate images (e) None of these

33. Which one of the following is not used to generate dynamic web pages?
(a) PHP (b) ASP.NET
(c) JSP (d) java script
(e) none of these

34. ________ programs are automatically loaded and operates as a part of browser.
(a) Utilities (b) Plug-ins
(c) Widget (d) Add-ons
(e) None of these

35. ___________ programs automatically connects to web sites and download documents and save them to local drive.
(a) web servers
(b) web Downloading utilites
(c) offline browser (d) online browser
(e) None of the above

36. Which of the following is an advantage of putting presentation information in a separate CSS file rather than in HTML itself?
(a) The content becomes easy to manage
(b) Becomes easy to make site for different devices like mobile by making separate CSS files
(c) CSS Files are generally cached and therefore decrease server load and network traffic.
(d) All of the above
(e) None of these

37. Consider the three commands : PROMPT, HEAD and RCPT. Which of the following options indicate a correct association of these commands with protocols where these are used?
(a) HTTP, SMTP, FTP (b) FTP, HTTP, SMTP
(c) HTTP, FTP, SMTP (d) SMTP,HTTP,FTP
(e) None of these

38. The MIME text file is saved with
(a) HMT extension (b) HTML extension
(c) THM extension (d) XML extension
(e) None of these

39. White spaces in XML includes
(a) Things like space characters, new lines and tabs
(b) Only spaces
(c) Spaces between two double quotes
(d) no space
(e) None of these

40. The attribute, which define the relationship between current document and HREF'ed URL is
(a) REL (b) URL
(c) REV (d) RPV
(e) All of the above

41. The tags elements in XML are
(a) case-insensitive (b) case-sensitive
(c) browser dependent (d) all the above
(e) none of these

42. A much better approach to establish the base URL is to use
(a) Base element (b) Head element
(c) Body element (d) all the above
(e) None of these

43. MSXML is
(a) a Microsoft language
(b) XML parser that ships with IE5.0
(c) used for interacting with XML
(d) Both (b) & (c)
(e) None of the above

44. A document have how many DTD's?
(a) 2 (b) None
(c) 4 (d) 5
(e) 3

45. Attributes in XML are
(a) elements in XML
(b) child nodes
(c) a way of attaching characteristics or properties to elements of a document
(d) all the above (e) None of these

46. The DOM specification describes how strings are to be manipulated by the DOM by defining the datatype _______. It is encoded using _______ encoding scheme.
(a) DOMString, UTF-8
(b) DOMString, Unicode
(c) UNICODE String, Unicode
(d) String, Unicode (e) None of these

47. Which of the following is an XML-based service IDL that defines the service interface and its implementation characteristics?
(a) UDDI (b) WSDL
(c) SOAP (d) Path
(e) None of these

48. Which of the following is (are) a valid XML name(s)?
(a) 1-2-4_6 (b) :3:-3:5:-7
(c) ;123456 (d) 3:4;-7
(e)None of the above

49. What is the correct HTML for creating a hyperlink?
(a) <a url="http://www.w3schools.com">W3Schools.com </a>
(b) <a href="http://www.w3schools.com">W3Schools</a >
(c) <a>http://www.w3schools.com</a>
(d) <a name="http://www.w3schools.com">W3Schools.co m</a>
(e)None of the above

50. Which of the following are CGI Proxy Server software?
(a) JAP (b) Cherokee HTTP Server
(c) Habitat (d) Proxy

(e)None of these

51. For the XML parser to ignore a certain section of your XML document, which syntax is correct?
(a) <![CDATA[ Text to be ignored ]]>
(b) <xml:CDATA[ Text to be ignored ]>
(c) <PCDATA> Text to be ignored </PCDATA>
(d) <CDATA> Text to be ignored </CDATA>
(e) None of these

52. Which of the following are all generic fonts?
(a) Fantasy, Cursive, Serif, Monospace, and Times New Roman
(b) Cursive, Fantasy, Serif, Monospace, and Sans Serif
(c) Arial, Cursive, Fantasy, Serif, and Monospace
(d) Cursive, Fantasy, Serif, Monospace, and Garamond
(e)None of the above

53. How DOM differs from SAX?
(a) DOM is not event driven and builds up the whole memory
(b) SAX is event driven and requires less space
(c) DOM and SAX are packages
(d) All the above
(e) None of these

54. Which of the following creates push button?
(a) Chech box (b) Radio
(c) Reset (d) Input
(e) None

55. The map definition file is generally stored in
(a) CGI-BIN (b) RECYCLE-BIN
(c) BIN (d) RECYCLE
(e) None of these

56. What does the URL need to access documents?
I. Path name II. Host name
III. DNS IV. Retrieval method
V. Server port number
(a) I, II, III (b) I, III, V
(c) I, II, IV (d) III, IV, V
(e) None of these

57. Unlike Ipv4, Ipv6 does not include the following field in the base header
(a) Next Header field
(b) Field for Fragmentation information
(c) Flow Label
(d) Kind field
(e) None of these

58. XML DSOs has the property for the number of pages of data the recordset contains
(a) Count (b) Number
(c) PageCount (d) PageNumber
(e) None of these

59. Microsoft XML schema data types for hexadecimal digits representing octates
(a) UID (b) UXID
(c) UUID (d) XXID
(e) None of these

60. CIDR stands for
(a) Classified Internet Domain Routing
(b) Classless Inter Domain Routing
(c) Classless Internet Domain Routing
(d) Classified Inter Domain Routing
(e) None of these

## SOLUTIONS

| | | | | | | |
|---|---|---|---|---|---|---|
| 1. (d) | 2. (d) | 3. (b) | 4. (c) | 5. (c) | 6. (d) | 7. (c) |
| 8. (a) | 9. (b) | 10. (c) | 11. (a) | 12. (a) | 13. (b) | 14. (a) |
| 15. (a) | 16. (c) | 17. (b) | 18. (e) | 19. (b) | 20. (b) | 21. (b) |
| 22. (a) | 23. (a) | 24. (b) | 25. (b) | 26. (a) | 27. (c) | 28. (a) |
| 29. (a) | 30. (c) | 31. (b) | 32. (a) | 33. (e) | 34. (b) | 35. (b) |
| 36. (d) | 37. (b) | 38. (b) | 39. (c) | 40. (c) | 41. (b) | 42. (a) |
| 43. (d) | 44. (b) | 45. (c) | 46. (b) | 47. (b) | 48. (b) | 49. (b) |
| 50. (a) | 51. (a) | 52. (b) | 53. (b) | 54. (c) | 55. (a) | 56. (b) |
| 57. (b) | 58. (c) | 59. (c) | 60. (b) | | | |

# 8 Computer Organization & Microprocessor

## 1. Number System

Number systems provide the basis for all operations in information processing systems. In a number system the information is divided into a group of symbols; for example, 26 English letters, 10 decimal digits etc. In conventional arithmetic, a number system based upon ten units (0 to 9) is used. However, arithmetic and logic circuits used in computers and other digital systems operate with only 0's and 1's because it is very difficult to design circuits that require ten distinct states. The number system with the basic symbols 0 and 1 is called binary. ie. A binary system uses just two discrete values. The binary digit (either 0 or 1) is called a bit.

A group of bits which is used to represent the discrete elements of information is a symbol. The mapping of symbols to a binary value is known a binary code. This mapping must be unique. For example, the decimal digits 0 through 9 are represented in a digital system with a code of four bits. Thus, a digital system is a system that manipulates discrete elements of information that is represented internally in binary form.

**Decimal and Binary Number System**

When we write decimal (base 10) numbers, we use a positional notation system. Each digit is multiplied by an appropriate power of 10 depending on its position in the number:

The Decimal Representation of Number $843_{10}$ Is:

$843 = 8 \times 10^2 + 4 \times 10^1 + 3 \times 10^0$
$= 8 \times 100 + 4 \times 10 + 3 \times 1$
$= 800 + 40 + 3$

For whole numbers, the rightmost digit position is the one's position ($10^0 = 1$). The numeral in that position indicates how many ones are present in the number. The next position to the left is ten's, then hundred's, thousand's, and so on. Each digit position has a weight that is ten times the weight of the position to its right.

In the decimal number system, there are ten possible values that can appear in each digit position, and so there are ten numerals required to represent the quantity in each digit position. The decimal numerals are the familiar zero through nine (0, 1, 2, 3, 4, 5, 6, 7, 8, 9).

In a positional notation system, the number base is called the radix. Thus, the base ten system that we normally use has a radix of 10. The term radix and base can be used interchangeably. When writing numbers in a radix other than ten, or where the radix isn't clear from the context, it is customary to specify the radix using a subscript.

The binary number system is also a positional notation numbering system, but in this case, the base is not ten, but is instead 2. Each digit position in a binary number represents a power of two. So, when we write a binary number, each binary digit is multiplied by an appropriate power of 2 based on the position in the number.

The Decimal Representation of Binary Number $101101_2$ Is:

$101101 = 1 \times 2^5 + 0 \times 2^4 + 1 \times 2^3 + 1 \times 2^2 + 0 \times 2^1 + 1 \times 2^0$
$= 1 \times 32 + 0 \times 16 + 1 \times 8 + 1 \times 4 + 0 \times 2 + 1 \times 1$
$= 32 + 8 + 4 + 1$

**Octal Number System**

Digital systems operate only on binary numbers. Since binary numbers are often very long, two shorthand notations, octal and hexadecimal, are used for representing large binary numbers. Octal systems use a base or radix of 8. Thus it has digits from 0 to 7 (r-1). As in the decimal and binary systems, the positional valued of each digit in a sequence of numbers is fixed. Each position in an octal number is a power of 8, and each position is 8 times more significant than the previous position.

The Decimal Representation of Binary Number $15.2_8$ Is:

$15.2_8 = 1 \times 8^1 + 5 \times 8^0 + 2 \times 8^{-1}$

| Decimal | Binary | Octal | Hexadecimal |
|---|---|---|---|
| 0 | 0000 | 0 | 0 |
| 1 | 0001 | 1 | 1 |
| 2 | 0010 | 2 | 2 |
| 3 | 0011 | 3 | 3 |
| 4 | 0100 | 4 | 4 |
| 5 | 0101 | 5 | 5 |
| 6 | 0110 | 6 | 6 |
| 7 | 0111 | 7 | 7 |
| 8 | 1000 | 10 | 8 |
| 9 | 1001 | 11 | 9 |
| 10 | 1010 | 12 | A |
| 11 | 1011 | 13 | B |
| 12 | 1100 | 14 | C |
| 13 | 1101 | 15 | D |
| 14 | 1110 | 16 | E |
| 15 | 1111 | 17 | F |

**Hexadecimal Number System:**

The hexadecimal number system contains 16 unique numerals. This system is most often referred to in spoken word as "hex" for short. Since we only have 10 Arabic numerals in our familiar decimal system, we need to use other symbols to represent the remaining 6 numerals. We use the alphabetic characters A–F in order to expand the system to 16 numerals. The 16 numerals in the hexadecimal

system are 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, and F. The relative magnitudes of the symbols are 0 < 1 < 2 < 3 < 4 < 5 < 6 < 7 < 8 < 9 < A < B < C < D < E < F. We use the generic term digit to describe the numerals within a hexadecimal number.

## 1.1 Base Conversion

**Converting to Decimal**

The value of each digit within a number is based on the individual digit value and the digit's position.

Each position in the number contains a different weight based on its relative location to the radix point. The weight of each position is based on the radix of the number system that is being used.

In order to find the decimal value of each of the numerals in the number, its individual numeral value is multiplied by its positional weight. In order to find the value of the entire number, each value of the individual numeral-weight products is summed.

**Binary to Decimal**

Let's convert $101.11_2$ to decimal. The same process is followed with the exception that the base in

the summation is changed to 2. Converting from binary to decimal can be accomplished quickly in your head due to the fact that the bit values in the products are either 1 or 0. That means any bit that is a 0 has no impact on the outcome and any bit that is a 1 simply yields the weight of its position.

**Example:** convert $101.11_2$ to Decimal:

**Example:** convert $101.11_2$ to Decimal:

$$\begin{array}{lccccccc}
 & & 1 & 0 & 1 & .1 & 1_2 \\
 & & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\
\text{Position} & \rightarrow & 2 & 1 & 0 & -1 & -2 \\
 & & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\
\text{Weight} & \rightarrow & (2)^2 & (2)^1 & (2)^0 & (2)^{-1} & (2)^{-2} \\
 & & & \downarrow & & &
\end{array}$$

$$\text{Value} = \sum_{i=-2}^{2} d_1 \cdot 2^i$$

$$\downarrow$$

$$\text{value} = 1.2^2 + 0.2^1 + 1.2^0 + 1.2^{-1} + 1.2^{-2}$$

$$\downarrow$$

$$\text{Value} = 1.(4) + 0.(2) + 1.(1) + 1.(1/2) + 1.(1/4)$$

$$\downarrow$$

$$\text{Value} = 4 + 0 + 1 + 0.5 + 0.25$$

$$\downarrow$$

$$\text{Value} = 5.75_{10}$$

**Octal to Decimal**

When converting from octal to decimal, the same process is followed with the exception that the base in the weight is changed to 8. Below is an example:

**Example:** Convert $17.17_8$ to Decimal

$$\begin{array}{rcccc}
 & 1 & 7 & .1 & 7_8 \\
 & \downarrow & \downarrow & \downarrow & \downarrow \\
\text{Position (p)} \rightarrow & 1 & 0 & -1 & -2 \\
 & \downarrow & \downarrow & \downarrow & \downarrow \\
\text{Weight} \rightarrow & (8)^1 & (8)^0 & (8)^{-1} & (8)^{-2}
\end{array}$$

$$\downarrow$$

$$\text{Value} = \sum_{i=-2}^{2} d_1 \cdot 2^i$$

$$\downarrow$$

$$\text{Value} = 1.8^1 + 7.8^0 + 1.8^{-1} + 7.8^{-2}$$

$$\downarrow$$

$$\text{Value} = 1.(8) + 7.(1) + 1.(1/8) + 7.(1/64)$$

$$\downarrow$$

$$\text{Value} = 8 + 7 + 0.125 + 0.109375$$

$$\downarrow$$

$$\text{Value} = 15.234375_{10}$$

**Hexadecimal to Decimal**

Let's convert $1AB.EF_{16}$ to decimal. The same process is followed with the exception that the base is changed to 16. When performing the conversion, the decimal equivalents of the numerals A–F need to be used.

Example: Convert $1AB.EF_{16}$ to Decimal

$$\begin{array}{rccccc}
 & 1 & A & B & .E & F_{16} \\
 & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\
\text{Position (p)} \rightarrow & 2 & 1 & 0 & -1 & -2 \\
 & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\
\text{weight} \rightarrow & (16)^2 & (16)^1 & (16)^0 & (16)^{-1} & (16)^{-2}
\end{array}$$

$$\downarrow$$

$$\text{Value} = \sum_{i=-2}^{2} d_1 \cdot 16^i$$

$$\downarrow$$

$$\text{Value} = 1.16^2 + A.16^1 + B.16^0 + E.16^1 + F.16^{-2}$$

$$\downarrow$$

$$\text{Value} = 1.(256) + 10.(16) + 11.(1) + 14.(1/16) + 15.(1/256)$$

$$\downarrow$$

$$\text{Value} = 256 + 160 + 11 + 0.875 + 0.05859375$$

$$\downarrow$$

$$\text{Value} = 427.93359375_{10}$$

**Converting from Decimal**

The process of converting from decimal to another base consists of two separate algorithms. There is one algorithm for converting the whole number portion of the number and another algorithm for converting the fractional portion of the number. The process for converting the whole number portion is to divide the decimal number by the base of the system you wish to convert to. The division will result in a quotient and a whole number remainder. The remainder is recorded as the least significant numeral in the converted number. The resulting quotient is then divided again by the base, which results in a new quotient and new remainder. The remainder is recorded as the next higher order numeral in the new number. This process is repeated until a quotient of 0 is achieved. At that point the conversion is complete. The remainders will always be within the numeral set of the base being converted to.

The process for converting the fractional portion is to multiply just the fractional component of the number by the base. This will result in a product that contains a whole number and a fraction. The whole number is recorded as the most significant digit of the new converted number. The new fractional portion is then multiplied again by the base with the whole number portion being recorded as the next lower order numeral. This process is repeated until the product yields a fractional component equal to zero or the desired level of accuracy has been achieved. The level of accuracy is specified by the number of numerals in the new converted number. For example, the conversion would be stated as "convert this

decimal number to binary with a fractional accuracy of 4 bits." This means the algorithm would stop once 4-bits of fraction had been achieved in the conversion.

**Decimal to Binary**

Let's convert $11.375_{10}$ to binary. Example shows the step-by-step process converting a decimal number to binary.

**Example:** Convert $11.375_{10}$ to Binary:

$$11.375_{10}$$

Part 1: Converting the whole number protion:

| | | Quotient | Remainder | |
|---|---|---|---|---|
| Step 1: | $2\overline{)11}$ | 5 | 1 | USB ↓ |
| Step 2: | $2\overline{)5}$ | 2 | 1 | Next highest order bit ↓ |
| Step 3: | $2\overline{)2}$ | 1 | 0 | Next highest order bit ↓ |
| Step 4: | $2\overline{)1}$ | 0 ↑ Done | 1 ↓ | MSB |

Converted whole Number = $1011_2$

Part 2: Converting the tractional number portion

| | | Product | Whole Number | |
|---|---|---|---|---|
| Step 1: | 2.(0.375) | 0.75 | 0 | MSB ↓ |
| Step 2: | 2.(0.75) | 1.50 | 1 | Next lower order bit ↓ |
| Step 3: | 2.(0.5) | 1.00 ↑ Done | 1 ↓ | LSB |

Converted Fractional Number = $.011_2$

Part 3: Combine the two components to form the new number:

$$1011.011_2$$

**Decimal to Octal**

Let's convert $10.4_{10}$ to octal with an accuracy of four fractional digits. When converting the fractional component of the number, the algorithm is continued until four digits worth of fractional numerals have been achieved. Once the accuracy has been achieved, the conversion is finished even though a product with a zero fractional value has not been obtained.

**Example :** Convert $10.4_{10}$ to Octal with an Accuracy of 4 fractional digits:

$10.4_{10}$

Part 1: Converting the whole number portion:

| | | Quotient | Remainder | |
|---|---|---|---|---|
| Step 1: | $8 \overline{)10}$ | 1 | 2 | Least significant digit ↓ |
| Step 2: | $8 \overline{)1}$ | 0 ↑ Done | 1 ↓ | Most significant digit |

Converted whole Number = $12_8$

Part 2: Converting the fractional number portion:

| | | Product | Whole Number | |
|---|---|---|---|---|
| Step 1: | 8 . (0.4) | <u>3</u>.2 | 3 | Most significant digit ↓ |
| Step 2: | 8 . (0.2) | <u>1</u>.6 | 1 | Next lower order digit ↓ |
| Step 3: | 8 . (0.6) | <u>4</u>.8 | 4 | Next lower order digit ↓ |
| Step 4: | 8 . (0.8) | <u>6</u>.4 | 6 | Least significant digit |

Converted Fractional Number = $.3146_8$

Done because we have achieved the desired accuracy

Part 3: Combine the two components to from the new number:

$12.3146_8$

**Decimal to Hexadecimal**

Let's convert $254.655_{10}$ to hexadecimal with an accuracy of three fractional digits. When doing this conversion, all of the divisions and multiplications are done using decimal. If the results end up between $10_{10}$ and $15_{10}$, then the decimal numbers are substituted with their hex symbol equivalent (i.e., A to F).

Example: Convert 254.65510 to Hexadecimal with an Accuracy of 3 fractional digits:

**254.65510**

Part 1: Converting the whole number portion:

| | | Quotient | Remainder | |
|---|---|---|---|---|
| Step 1: | $16 \overline{)254}$ | 15($F_{16)}$ | 14($F_{16)}$ | Least significant digit ↓ |
| Step 2: | $16 \overline{)15}$ | 0 ↑ Done | 15(F1$_{6)}$ ↓ | Most significant digit |

Converted whole Number = $FE_{16}$

Part 2: Converting the fractional number portion:

| | | Product | Whole Number | |
|---|---|---|---|---|
| Step 1: | 16.(0.655) | <u>10</u>.48 | 10($A_{16}$) | Most significant digit ↓ |
| Step 2: | 16.(0.48) | <u>7</u>.68 | 7 | Next lower order digit ↓ |
| Step 3: | 16.(0.68) | <u>10</u>.88 | 10($A_{16}$) | Least significant digit |

Converted Fractional Number = $.A7A_{16}$

Done because we have achieved the desired accuracy

Part 3: Combine the two components to form the new number:

$FE.A7A_{16}$

**Binary to Octal**

Example: Convert $10111.01_2$ to Octal

$10111\ .\ 01_2$

Part 1: Form groups of 3 bits representing octal symbols.

$(0\ 1\ 0)\ (1\ 1\ 1)\ .\ (0\ 1\ 0)_2$

Whole number groupings start at the radix point and work left. Leading 0's are added as necessary

Fractional number groupings start at the radix point and work right. Trailing 0's are added as necessary

Part: Perform a direct substitution of the bit groupings with the equivalent octal symbol.

$(0\ 1\ 0)\ (1\ 1\ 1)\ .\ (0\ 1\ 0)_2$

$2\ 7\ .\ 2_8$

**Binary to Hexadecimal**

Example: Convert $111011.11111_2$ to Hexadecimal

$111011\ .\ 11111_2$

Part 1: Form groups of 4 bits representing hex symbols.

$(0\ 0\ 1\ 1)\ (1\ 0\ 1\ 1)\ .\ (1\ 1\ 1\ 1)\ (1\ 0\ 0\ 0)_2$

Whole number groupings start at the radix point and work left. Leading 0's are added as necessary.

Fractional number groupings start at the radix point and work right Trailing 0's are added as necessary.

Part 2: Perform a direct substitution of the bit groupings with the equivalent hex symbol.

$(0\ 0\ 1\ 1)\ (1\ 0\ 1\ 1)\ .\ (1\ 1\ 1\ 1)\ (1\ 0\ 0\ 0)_2$

$3\ B\ .\ F\ 8_{16}$

**Octal to Binary**

An octal symbol will be replaced with 3 binary bits while a hexadecimal symbol will be replaced with 4 binary bits. Any leading or trailing 0s can be removed from the converted number once complete.

Example: Convert $347.12_8$ to Binary:

$347\ .\ 12_8$

Part 1: Each of the octal symbols is replaced with its 3 bit binary equivalent.

$3\ 4\ 7\ .\ 1\ 2_8$

$(0\ 1\ 1)\ (1\ 0\ 0)\ (1\ 1\ 1)\ .\ (0\ 0\ 1)\ (0\ 1\ 0)_2$

Leading and Trailing 0's can be removed

$11100111\ .\ 00101_2$

**Hexadecimal to Binary**
Example: Convert $1B.A_{16}$ to Binary:

Part 1: Each of the hex symbols is replaced with its 4 bit binary equivalents

$1\ B\ .\ A_{16}$

$(0\ 0\ 0\ 1)\ (1\ 0\ 1\ 1)\ .\ (1\ 0\ 1\ 0)_2$

Part 2: Leading and trailing Zeros can be removed.

$11011\ .\ 101_2$

**Octal to Hexadecimal**
When converting between $2_n$ bases (excluding binary) the number is first converted into binary and then converted from binary into the final $2_n$ base using the algorithms described before.

Example convert $71.5_8$ to Hexadecimal:
Part 1: Convert the octal number into binary. Each octal symbol is represented with 3 bits.

$7\ 1\ .\ 5_8$

$(1\ 1\ 1)\ (0\ 0\ 1)\ .\ (1\ 0\ 1)_2$

$111001.101_2$

Part 2: Convert the binary number into hexadecimal. Form groups of 4 bits representing hex symbols.

Step 1: $(0\ 0\ 1\ 1)\ (1\ 0\ 0\ 1)\ .\ (1\ 0\ 1\ 0)_2$

Whole number groupings start at the radix point and work left. Leading 0's are added as necessary.

Fractional number groupings start at the radix point and work right. Trailing 0's are added as necessary.

Step 2: $(0\ 0\ 1\ 1)\ (1\ 0\ 0\ 1)\ .\ (1\ 0\ 1\ 0)_2$

$3\ 9\ .\ A_{16}$

**Hexadecimal to Octal**
Example: Convert $AB.C_{16}$ to Octal:

$A\ B\ .\ C_{16}$

Part 1: Convert the hex number into binary. Each hex symbol represented with 4 bits.

$AB\ .\ C_{16}$

$(1\ 0\ 1\ 0)\ (1\ 0\ 1\ 1)\ .\ (1\ 1\ 0\ 0)_2$

$10101011\ .\ 11_2$

Part 2: Convert the binary number into octal. From groups of 3 bits representing octal symbols.

Step 1: $(0\ 1\ 0\ )\ (1\ 0\ 1)\ (0\ 1\ 1)\ .\ (1\ 1\ 0)_2$

Step 2: $2\ 5\ 3\ .\ 6_8$

## 1.2 Binary Arithmetic

Binary addition is a straightforward process that mirrors the approach we have learned for longhand decimal addition. The two numbers (or terms) to be added are aligned at the radix point and addition begins at the least significant bit. If the sum of the least significant position yields a value with two bits (e.g., $10_2$), then the least significant bit is recorded and the most significant bit is carried to the next higher position. The sum of the next higher position is then performed including the potential carry bit from the prior addition. This process continues from the least significant position to the most significant position.

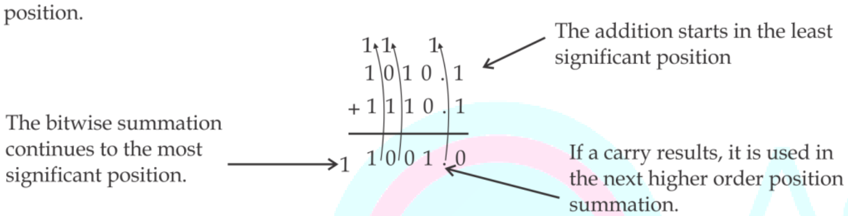
**Example:** Single Bit Binary Addition
There are four Possible results when adding two bits.

$$\begin{array}{r} 0 \\ +\ 0 \\ \hline 0 \end{array} \qquad \begin{array}{r} 0 \\ +\ 1 \\ \hline 1 \end{array} \qquad \begin{array}{r} 1 \\ +\ 0 \\ \hline 1 \end{array} \qquad \begin{array}{r} 1 \\ +\ 1 \\ \hline \text{Carry} \rightarrow \underline{1}0 \end{array}$$

When performing binary addition, the width of the inputs and output is fixed (i.e., n-bits). Carries that exist within the n-bits are treated in the normal fashion of including them in the next higher position sum; however, if the highest position summation produces a carry, this is a uniquely named event. This event is called a carry out or the sum is said to generate a carry. The reason this type of event is given special terminology is because in real circuitry, the number of bits of the inputs and output is fixed in hardware and the carry out is typically handled by a separate circuit.

Example: What is the sum of $1010.1_2$ and $1110.1_2$? Did this addition generate a carry?

The two numbers are aligned at the radix point and addition begins at the least significant position. Carries are recorded at each position and used in the addition of the next higher position.




The sum of these two numbers is $11001.0_2$. Since the inputs each had n=5 but the sum required n=6, we say that this addition "generated a carry". Another way of stating the result is "$1001_2$ with a carry".

$$\begin{array}{r} 0 \\ -\ 0 \\ \hline 0 \end{array} \qquad \begin{array}{r} \text{Borrow} \rightarrow 10 \\ \text{Required}\ \ \cancel{0} \\ -\ 1 \\ \hline 1 \end{array} \qquad \begin{array}{r} 1 \\ -\ 0 \\ \hline 1 \end{array} \qquad \begin{array}{r} 1 \\ -\ 1 \\ \hline 0 \end{array} \begin{array}{l} \leftarrow \text{Minuend} \\ \leftarrow \text{Subtrahend} \\ \\ \end{array}$$

**Subtraction (Borrows)**
Binary subtraction also mirrors longhand decimal subtraction. In subtraction, the formal terms for the two numbers being operated on are minuend and subtrahend. The subtrahend is subtracted from the minuend to find the difference. In longhand subtraction, the minuend is the top number and the subtrahend is the bottom number. For a given position if the minuend is less than the subtrahend, it needs to borrow from the next higher order position to produce a difference that is positive. If the next higher position does not have a value that can be borrowed from (i.e., 0), then it in turn needs to borrow from the next higher position, and so forth.

$$\begin{array}{r} 0 \\ -\ 0 \\ \hline 0 \end{array} \qquad \begin{array}{r} \text{Borrow} \rightarrow 10 \\ \text{Required}\ \ \cancel{0} \\ -\ 1 \\ \hline 1 \end{array} \qquad \begin{array}{r} 1 \\ -\ 0 \\ \hline 1 \end{array} \qquad \begin{array}{r} 1 \\ -\ 1 \\ \hline 0 \end{array} \begin{array}{l} \leftarrow \text{Minuend} \\ \leftarrow \text{Subtrahend} \\ \\ \end{array}$$

**Example:** Single Bit Binary Subtraction
There are four possible results when subtracting two bits.

As with binary addition, binary subtraction is accomplished on fixed widths of inputs and output (i.e., n-bits). The minuend and subtrahend are aligned at the radix point and subtraction begins at the least significant bit position. Borrows are used as necessary as the subtractions move from the least significant position to the most significant position. If the most significant position requires a borrow, this is a uniquely named event. This event is called a borrow in or the subtraction is said to require a borrow. Again, the reason this event is uniquely named is because in real circuitry, the number of bits of the input and output is fixed in hardware and the borrow in is typically handled by a separate circuit.

**Example:** What is the difference between $1011.0_2$ and $0100.1_2$? Did this subtraction require a borrow in?

The way this question is phrased indicates that $1011.0_2$ is the minuend and $0100.1_2$ is the subtrahend. The two numbers are aligned at the radix point and subtraction begins at the least significant position. Borrows are taken as needed from the next higher order position.

Borrow Required     Borrow Required

$$\begin{array}{r} 0\ 10\ \ 0\ \ 10 \\ \cancel{1}\ \cancel{0}\ 1\ \cancel{1}\ .\ \cancel{0} \\ -\ 0\ 1\ 0\ 0\ .\ 1 \\ \hline 0\ 1\ 1\ 0\ .\ 1 \end{array}$$

← The subtraction starts in the least significant position

The difference of these two numbers is $0110.1_2$ and it did not require a borrow in. To double-check if this subtraction worked, we can look at the decimal equivalents of the numbers: $1011.0_2$ $(11_{10})$ - $0100.1_2(4.5_{10})$ = $0110.1_2(6.5_{10})$, which verifies the subtraction was correct.

## 1.3 Unsigned and Signed Numbers

All of the number systems presented in the prior sections were positive. We need to also have a mechanism to indicate negative numbers. When looking at negative numbers, we only focus on the mapping between decimal and binary since octal and hexadecimal are used as just another representation of a binary number. In decimal, we are able to use the negative sign in front of a number to indicate that it is negative (e.g., $-34_{10}$). In binary, this notation works fine for writing numbers on paper (e.g., $-1010_2$), but we need a mechanism that can be implemented using real circuitry. In a real digital circuit, the circuits can only deal with 0s and 1s. There is no "-" in a digital circuit. Since we only have 0s and 1s in the hardware, we use a bit to represent whether a number is positive or negative. This is referred to as the sign bit. If a binary number is not going to have any negative values, then it is called an unsigned number and it can only represent positive numbers. If a binary number is going to allow negative numbers, it is called a signed number. It is important to always keep track of the type of number we are using as the same bit values can represent very different numbers depending on the coding mechanism that is being used.

### Unsigned Numbers

An unsigned number is one that does not allow negative numbers. When talking about this type of code, the number of bits is fixed and stated up front. We use the variable n to represent the number of bits in the number. For example, if we had an 8-bit number, we would say, "This is an 8-bit, unsigned number." The number of unique codes in an unsigned number is given by $2^n$. For example, if we had an 8-bit number, we would have $2^8$ or 256 unique codes (e.g., $0000\ 0000_2$ to $1111\ 1111_2$).

The range of an unsigned number refers to the decimal values that binary code can represent. If we use the notation $N_{unsigned}$ to represent any possible that n-bit- unsigned number can take on, the range would be defined as $0 < N_{unsigned} < (2^n - 1)$:

Range of an UNSIGNED Number $\Rightarrow 0 \le N_{unsigned} \le (2^n - 1)$

### Signed Numbers

Signed numbers are able to represent both positive and negative numbers. The most significant bit of these numbers is always the sign bit, which represents whether the number is positive or negative. The sign bit is defined to be a 0 if the number is positive and 1 if the number is negative. When using signed numbers, the number of bits is fixed so that the sign bit is always in the same position. There are a variety of ways to encode negative numbers using a sign bit. The encoding method used exclusively in modern computers is called two's complement. There are two other encoding techniques called signed magnitude and one's complement that are rarely used but are studied to motivate the power of two's complement. When talking about a signed number, the number of bits and the type of encoding are always stated.

### Signed Magnitude

Signed magnitude is the simplest way to encode a negative number. In this approach, the most significant bit (i.e., leftmost bit) of the binary number is considered the sign bit (0 = positive, 1 = negative). The rest of the bits to the right of the sign bit represent the magnitude or absolute value of the number. As an example of this approach, let's look at the decimal values that a 4-bit, signed magnitude number can take on.

Example: What decimal values can a
4-bit "Signed Magnitude" code represent?

| Decimal | 4-bit Signed Magnitude |
|---|---|
| –7 | 1111 |
| –6 | 1110 |
| –5 | 1101 |
| –4 | 1100 |
| –3 | 1011 |
| –2 | 1010 |
| –1 | 1001 |
| –0 | 1000 |
| 0 | 0000 |
| 1 | 0001 |
| 2 | 0010 |
| 3 | 0011 |
| 4 | 0100 |
| 5 | 0101 |
| 6 | 0110 |
| 7 | 0111 |

↑ Sign bit

There are drawbacks of signed magnitude encoding that are apparent from this example. First, the value of $0_{10}$ has two signed magnitude codes ($0000_2$ and $1000_2$). This is an inefficient use of the available codes and leads to complexity when building arithmetic circuitry since it must account for two codes representing the same number.

The second drawback is that addition using the negative numbers does not directly map to how decimal addition works. For example, in decimal if we added (-5) + (1), the result would be -4. In signed magnitude, adding these numbers using a traditional adder would produce (-5) + (1) = (-6). This is because the traditional addition would take place on the magnitude portion of the number. A $5_{10}$ is represented with $10_{12}$. Adding 1 to this number would result in the next higher binary code $110_2$ or $6_{10}$.

Since the sign portion is separate, the addition is performed on |5|, thus yielding 6. Once the sign bit is included, the resulting number is -6. It is certainly possible to build an addition circuit that works on signed magnitude numbers, but it is more complex than a traditional adder because it must perform a different addition operation for the negative numbers versus the positive numbers. It is advantageous to have a single adder that works across the entire set of numbers.

Due to the duplicate codes for 0, the range of decimal numbers that signed magnitude can represent is reduced by 1 compared to unsigned encoding. For an n-bit number, there are $2^n$ unique binary codes available but only $2^{n-1}$ can be used to represent unique decimal number. If we use the notation $N_{SM}$ to represent any possible value that that an n-bit, signed magnitude can take on, the range would be defined as

Range of a SIGNED MAGNITUDE number $\Rightarrow -(2^{n-1} - 1) \leq N_{SM} \leq +(2^{n-1} - 1)$

### 1.4 One's Complement

One's complement is another simple way to encode negative numbers. In this approach, the negative number is obtained by taking its positive equivalent and flipping all of the 1s to 0s and 0s to 1s. This procedure of flipping the bits is called a complement. In this way, the most significant bit of the number is still the sign bit (0=positive, 1= negative). The rest of the bits represent the value of the number, but in this encoding scheme the negative number values are less intuitive. As an example of this approach, let's look at the decimal values that a 4-bit, one's complement number can take on.

Example: What decimal values can a 4-bit 'One's Complement' code represent?

| Decimal | 4-bit One's Complement |
|---|---|
| – 7 | 1000 |
| –6 | 1001 |
| –5 | 1010 |
| –4 | 1011 |
| –3 | 1100 |
| –2 | 1101 |
| –1 | 1110 |
| –0 | 1111 |
| 0 | 0000 |
| 1 | 0001 |
| 2 | 0010 |
| 3 | 0011 |
| 4 | 0100 |
| 5 | 0101 |
| 6 | 0110 |
| 7 | 0111 |

↑Sign bit

Again, we notice that there are two different codes for $0_{10}$ ($0000_2$ and $1111_2$). This is a drawback of one's complement because it reduces the possible range of numbers that can be represented from $2^n$ to ($2^n$ - 1) and requires arithmetic operations that consider the gap in the number system. There are advantages of one's complement, however. First, the numbers are ordered such that traditional addition works on both positive and negative numbers (excluding the double 0 gap). Taking the example of (-5) + (1) again, in one's complement the result yields -4, just as in a traditional decimal system.

Notice that in one's complement, $-5_{10}$ is represented with $1010_2$. Adding 1 to this entire binary code would result in the next higher binary code $1011_2$ or $-4_{10}$ from the above table. This makes addition circuitry less complicated, but still not as simple as if the double 0 gap was eliminated. Another advantage of one's complement is that as the numbers are incremented beyond the largest value in the set, they roll over and start counting at the lowest number. For example, if you increment the number $0111_2$ ($7_{10}$), it goes to the next higher binary code $1000_2$, which is $-7_{10}$. The ability to have the numbers roll over is a useful feature for computer systems.

The process of finding the decimal value of a one's complement number involves first identifying whether the number is positive or negative by looking at the sign bit. If the number is positive (i.e., the sign bit is 0), then the number is treated as an unsigned code and is converted to decimal using the standard conversion procedure described in prior sections. If the number is negative (i.e., the sign bit is 1), then the number sign is recorded separately and the code is complemented in order to convert it to its positive magnitude equivalent. This new positive number is then converted to decimal using the standard conversion procedure. As the final step, the sign is applied.

Example: What is the decimal value of the 5 bit. One's complement code $11010_2$ ?

The most significant bit of this 5-bit number is a 1, which indicates that the number is negative.

Sign Bit ⟶ $\underline{1}1010$

To find the magnitude of the number, we first perform a complement on the entire number to find its positive equivalent

$$1\ 1\ 0\ 1\ 0_2 \downarrow 0\ 0\ 1\ 0\ 1_2$$

A complement operation turns all 1's to 0's and all 0's to 1's

The number can now be converted into decimal to find its magnitude.

$$|\text{Value}| = \sum_{I=0}^{4} d_i . 2_i$$

$$\downarrow$$

$$|\text{Value}| = 0.2^4 + 0.2^3 + 1.2^2 + 0.2^1 + 1.2^0$$

$$\downarrow$$

$$|\text{Value}| = 0.(16) + 0.(8) + 1.(4) + 0.(2) + 1.(1)$$

$$\downarrow$$

$$|\text{Value}| = 0 + 0 + 4 + 0 + 1 = 5_{10}$$

The negative sign is then added back to the converted number given a decimal value of $-5_{10}$

## 1.5 Two's Complement

Two's complement is an encoding scheme that addresses the double 0 issue in signed magnitude and 1's complement representations. In this approach, the negative number is obtained by subtracting its positive equivalent from $2^n$. This is identical to performing a complement on the positive equivalent and then adding one. If a carry is generated, it is discarded. This procedure is called "taking the two's complement of a number." The procedure of complementing each bit and adding one is the most common technique to perform a two's complement. In this way, the most significant bit of the number is still the sign bit (0 = positive, 1 = negative) but all of the negative numbers are in essence shifted up so that the double 0 gap is eliminated. Taking the two's complement of a positive number will give its negative counterpart and vice versa. Let's look at the decimal values that a 4-bit, two's complement number can take on.

Example: What decimal values can a 4-bit "Two's Complement" code represent?

| Decimal | 4-bit Two's complement |
|---|---|
| –8 | 1000 |
| –7 | 1001 |
| –6 | 1010 |
| –5 | 1011 |
| –4 | 1100 |
| –3 | 1101 |
| –2 | 1110 |
| –1 | 1111 |
| 0 | 0000 |
| 1 | 0001 |
| 2 | 0010 |
| 3 | 0011 |
| 4 | 0100 |
| 5 | 0101 |
| 6 | 0110 |
| 7 | 0111 |

↑ Sign bit

There are many advantages of two's complement encoding. First, there is no double 0 gap, which means that all possible $2^n$ unique codes that can exist in an n-bit number are used. This gives the largest possible range of numbers that can be represented. Another advantage of two's complement is that addition with negative numbers works exactly the same as decimal. In our example of (-5) + (1), the result is (-4). Arithmetic circuitry can be built to mimic the way our decimal arithmetic works without the need to consider the double 0 gap. Finally, the rollover characteristic is preserved from one's complement. Incrementing +7 by +1 will result in -8.

The process of finding the decimal value of a two's complement number involves first identifying whether the number is positive or negative by looking at the sign bit. If the number is positive (i.e., the sign bit is 0), then the number is treated as an unsigned code and is converted to decimal using the standard conversion procedure described in prior sections. If the number is negative (i.e., the sign bit is 1), then the number sign is recorded separately and a two's complement is performed on the code in order to convert it to its positive magnitude equivalent. This new positive number is then converted to decimal using the standard conversion procedure. The final step is to apply the sign.

Example: What is the decimal value of the 5- bit, 2's complement code $11010_2$?

The most significant bit of this 5-bit number is a 1, which indicates that the number is negative.

Sign Bit → $\underline{1}1010$

To find the magnitude of the number, we take the 2's complement of the entire number to find its positive equivalent.

Step 1 - Complement the number $11010_2$ ↓ $00101_2$

Step 2- Add 1, ignore carry Out if any

$$\begin{array}{r} \overset{1}{0\,0\,1\,0\,1} \\ +\quad 1 \\ \hline 0\,0\,1\,1\,0_2 \end{array}$$

The number can now be converted into decimal to find its magnitude (i.e., $00110_2 = 6_{10}$)

The negative sign is then added giving a final decimal value of $-6_{10}$

To convert a decimal number into its two's complement code, the range is first checked to determine whether the number can be represented with the allocated number of bits. The next step is to convert the decimal number into unsigned binary. The final step is to apply the sign bit. If the original decimal number was positive, then the conversion is complete. If the original decimal number was negative, then the two's complement is taken on the unsigned binary code to find its negative equivalent.

Example: What is the 8-bit. 2's complement code for $-99_{10}$?

Step 1- Determine if $-99_{10}$ can be represented within the 2's complement number range

An 8-bit, 2's complement number has a range of

$$-(2^{n-1}) \le N_{2comp} \le +(2^{n-1}-1)$$
$$\downarrow$$
$$-(2^{8-1}) \le N_{2comp} \le +(2^{8-1}-1)$$
$$\downarrow$$
$$-128 \le N_{2comp} \le +127$$

Yes, the number $-99_{10}$ falls within the range that an 8 bit, 2's complement number.

Step 2 - Find the positive binary code for $-99_{10}$

| | Quotient | Remainder | |
|---|---|---|---|
| 2 ⟌ 99 | 49 | 1 | LSB |
| 2 ⟌ 49 | 24 | 1 | |
| 2 ⟌ 24 | 12 | 0 | |
| 2 ⟌ 12 | 6 | 0 | |
| 2 ⟌ 6 | 3 | 0 | |
| 2 ⟌ 3 | 1 | 1 | |
| 2 ⟌ 1 | 0 | 1 | MSB |

Done

The converted 8-bit number is $0110\ 0011_2$

Step 3- Perform 2's Complement on the positive equivalent of $99_{10}$

First, complement the number $0110\ 0011_2$
$$\downarrow$$
$$1001\ 1100_2$$

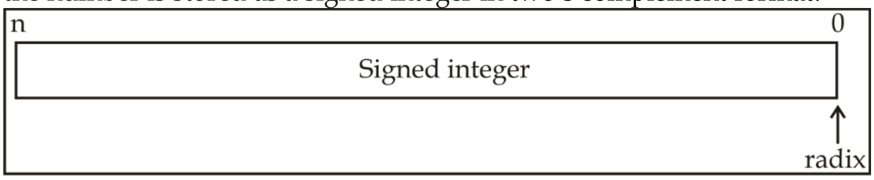
Second, add 1. ignore carry out if any

$$\begin{array}{r} 1001\ 1100 \\ +\qquad\quad 1 \\ \hline 10011101_2 \end{array}$$

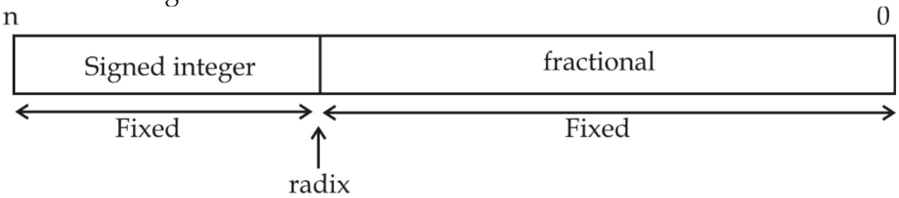
The 8-bit, 2's complement code for $-99_{10}$ is $1001\ 1101_2$

## 1.6. Fixed Point Notation

**Fixed Point Notation** is a representation of our fractional number as it is stored in memory. In Fixed Point Notation, the number is stored as a signed integer in two's complement format.




On top of this, we apply a notional split, locating the radix point (the separator between integer and fractional parts) a fixed number of bits to the left of its notational starting position to the right of the least significant bit. We've illustrated this in the diagram below.




When we interpret the bits of the signed integer stored in memory we reposition the radix point by multiplying the stored integer by a fixed scaling factor. The scaling factor in binary is always 2 raised to a fixed exponent. As the scaling factor is a power of 2 it relocates the radix point some number of places to the left or right of its starting position.

During this conversion, there are three directions that the radix point can be moved:

- **The radix point is moved to the right:** This is represented by a scaling factor whose exponent is 1 or more. In this case additional zeros are appended to the right of the least-significant bit and means that the actual number being represented is larger than the binary integer that was stored.
- **The radix point remains where it is:** This is represented by a scaling factor whose exponent is 0 and means that the integer value stored is exactly the same as the integer value being represented.
- **The radix point is moved to the left:** This is represented by a scaling factor whose exponent is negative. This means that the number being represented is smaller than the integer number that was stored and means that the number being represented has a fractional component.

Let's take a look at a couple of examples.

**Examples of Fixed Point Numbers**

Lets assume we have an 8-bit signed binary number $00011011_2$ that is stored in memory using 8-bits of storage (hence the leading zeros).

In our first scenario, lets also assume this number was stored as a signed fixed-point representation with a scale factor of $2^2$.

As our scale factor is greater than 1, when we translated the bits stored in memory into the number we are representing, we move the radix point two places to the right. This gives us the number: $1101100_2$ (Note the additional zeros that are appended to the right of the least significant bit).

In our second scenario, let us assume that we start off with the same binary number in memory but this time we'll assume that it is stored as a signed fixed-point representation with a scale factor of $2^{-3}$. As the exponent is negative we move the radix point three places to the left. This gives us the number $00011.011_2$

**Advantages and Disadvantages of Fixed Point Representation**

The major advantage of using a fixed-point representation is performance. As the value stored in memory is an integer the CPU can take advantage of many of the optimizations that modern computers have to perform integer arithmetic without having to rely on additional hardware or software logic. This in turn can lead to increases in performance and when writing your apps, can therefore lead to an improved experience for your users.

However, there is a downside! Fixed Point Representations have a relatively limited range of values that they can represent.

So how do we work out the maximum and minimum numbers that can be stored in a fixed-point representation and determine whether it is suitable for our needs? All we do is take the largest and smallest integer values that can be stored in the given number of bits and multiply that by the scale factor associated with our fixed-point representation. For a given signed binary number using *b* bits of storage with a scale factor of *f* the maximum and minimum values that can be stored are:

Minimum: $-2^{b-1}/2^f$

Maximum: $(2^{b-1}-1)/2^f$

If the number you want to represent fits into this range then things are great. If it doesn't though, you have to look for an alternative! This is where Floating Point Notation comes in.

## 1.7 Floating Point Notation

**Floating Point Notation** is an alternative to the Fixed-Point notation and is the representation that most modern computers use when storing fractional numbers in memory. Floating Point Notation is a way to represent very large or very small numbers precisely using scientific notation in binary. In doing so, Floating Point Representation provides a varying degree of precision depending on the scale of the numbers that you are using.

For example, the level of precision we need when we are talking about the distance between atoms ($10^{-10}$ m) is very different from the precision we need when we're talking about the distance between the earth and the sun ($10^{11}$ m). This is a major benefit and allows a much wider range of numbers to be represented than is possible in Fixed Point Notation. Floating Point Representation is based on **Scientific Notation**. You may have used Scientific Notation in school. When we use Scientific Notation in decimal (the form you're probably most familiar with), we write numbers in the following form:

+/- mantissa x $10^{\text{exponent}}$

In this form, there is an optional sign indicating whether the overall number is positive or negative, followed by a mantissa (also known as a significand) which is a real (fractional) number which in turn is multiplied by a number base (or radix) raised by an exponent. As we know, in decimal this number base is 10.
Floating Point Representation is essentially Scientific Notation applied to binary numbers. In binary, the only real difference is that the number base is 2 instead of 10. We would therefore write Floating Point Numbers in the following form:
+/- mantissa x $2^{exponent}$

Now, you may not have realised it but when we write numbers in scientific notation (whether they be binary or decimal) we can write them in a number of different ways.

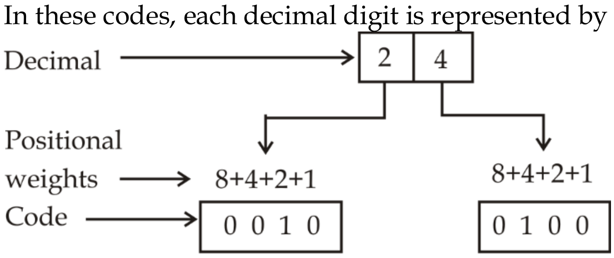
In decimal, we could write $1.5 \times 10^2$, $15 \times 10^1$ and $150 \times 10^0$ and yet all these numbers have exactly the same value.
This provides flexibility but with this flexibility also comes confusion. To try and address this confusion a common set of rules known as normalized scientific notation are used to define how numbers in scientific notation are normally written.

In the coding, when numbers, letters or words are represented by a specific group of symbols, it is said that the number, letter or word is being encoded. The group of symbols is called as a code. The digital data is represented, stored and transmitted as group of binary bits. This group is also called as binary code. The binary code is represented by the number as well as alphanumeric letter.

Following is the list of advantages that binary code offers. Binary codes are suitable for the computer applications.
- Binary codes are suitable for the digital communications.
- Binary codes make the analysis and designing of digital circuits if we use the binary codes.
- Since only 0 & 1 are being used, implementation becomes easy.

## 1.8 Classification of Binary Codes
The codes are broadly categorized into following four categories.
- Weighted Codes
- Non-Weighted Codes
- Binary Coded Decimal Code
- Alphanumeric Codes
- Error Detecting Codes
- Error Correcting Codes

**Weighted Codes**
Weighted binary codes are those binary codes which obey the positional weight principle. Each position of the number represents a specific weight. Several systems of the codes are used to express the decimal digits 0 through 9. In these codes, each decimal digit is represented by a group of four bits.




**Non-Weighted Codes**
In this type of binary codes, the positional weights are not assigned. The examples of non-weighted codes are Excess-3 code and Gray code.

**Excess-3 code**
The Excess-3 code is also called as XS-3 code. It is non-weighted code used to express decimal numbers. The Excess-3 code words are derived from the 8421 BCD code words adding W112 or 3 10 to each code word in 8421. The excess-3 codes are obtained as follows –

$$\text{Decimal Number} \longrightarrow \text{8421 BCD} \xrightarrow[0011]{\text{Add}} \text{Excess} - 3$$

**Example**

| Decimal | BCD | | | | Excess-3 | | | |
|---|---|---|---|---|---|---|---|---|
| | 8 | 4 | 2 | 1 | BCD+0011 | | | |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 |
| 2 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 |
| 3 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 0 |
| 4 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 |
| 5 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 |
| 6 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 1 |
| 7 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 0 |
| 8 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 1 |
| 9 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 0 |

**Gray Code**

It is the non-weighted code and it is not arithmetic codes. That means there are no specific weight assigned to the bit position. It has a very special feature that only one bit will change each time the decimal number is incremented as shown in fig. As only one bit changes at a time, the gray code is called as a unit distance code. The gray code is a cyclic code. Gray code cannot be used for arithmetic operation.

| Decimal | BCD | | | | Gray | | | |
|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| 2 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 |
| 3 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 |
| 4 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 |
| 5 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| 6 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 1 |
| 7 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 0 |
| 8 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 |
| 9 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 1 |

**Application of Gray code**

- Gray code is popularly used in the shaft position encoders.
- A shaft position encoder produces a code word which represents the angular position of the shaft.

**Binary Coded Decimal BCD code**

In this code, each decimal digit Is represented by a 4-bit binary number. BCD is a way to express each of the decimal digits with a binary code. In the BCD, with four bits we can represent sixteen numbers 0000 to 1111. But in BCD code only first ten of these are used 00000 to 1001. The remaining six code combinations i.e. 1010 to 1111 are invalid in BCD.

| **Decimal** | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| **BCD** | 0000 | 0001 | 0010 | 0011 | 0100 | 0101 | 0110 | 0111 | 1000 | 1001 |

**Advantages of BCD Codes**

- It is very similar to decimal system.
- We need to remember binary equivalent of decimal numbers 0 to 9 only.

**Disadvantages of BCD Codes**

- The addition and subtraction of BCD have different rules.
- The BCD arithmetic is little more complicated.
- BCD needs more number of bits than binary to represent the decimal number. So, BCD is less efficient than binary.

## 2. General Register Organization

Memory access is very time-consuming, especially on today's computers, which typically have many wait-states. Compounding the technological limitations on memory speed is the fact than RAM memory banks are getting bigger.

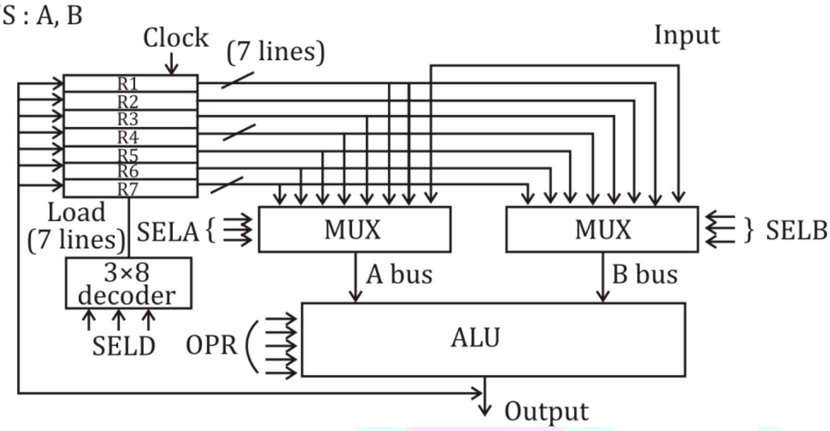
Bigger memory units mean longer propagation delays in decoding the address. Memory access is also expensive in terms of the instruction code size necessary to accommodate 32 or 64-bit addresses.

If many CPU registers are available for heavily used variables and intermediate results, we can avoid memory references much of the time, thus vastly increasing program execution speed, and reducing program size.

In a general-register organization such as this one, any two registers can be inputs to the ALU, and the results from the ALU can be stored in any register. When all registers are interchangeable, the architecture is orthogonal, or symmetric. An orthogonal processor is ideal for programmers, because they can use any register for any purpose. This is the other end of the spectrum from an accumulator-based architecture.

General register Organization

BUS : A, B




**The Control Word**
Is the binary word formed by the selections in the general register organization

It has 14 bits, 3 bits for SELA, 3 bits for SELB, 3 bits for SELD, and 5 bits for OPR

| Number of bits | 3 | 3 | 3 | 5 |
|---|---|---|---|---|
| **Field** | **SELA** | **SELB** | **SELD** | **OPR** |

**Control Word (CW)**
Control word is defined as a word whose individual bits represent the various control signal. Therefore, each of the control steps in the control sequence of an instruction defines a unique combination of 0s and 1s in the CW.
A sequence of control words (CWs) corresponding to the control sequence of a machine instruction constitutes the micro program for that instruction.

The individual control words in this micro-program are referred to as micro instructions.
The micro programs corresponding to the instruction set of a computer are stored in a special memory which will be referred to as the micro-program memory. The control words related to an instruction are stored in micro-program memory.

## 3. Micro programmed Control

In hardwired control, we saw how all the control signals required inside the CPU can be generated using a state counter and a PLA circuit.
There is an alternative approach by which the control signals required inside the CPU can be generated. This alternative approach is known as micro programmed control unit. In micro programmed control unit, the logic of the control unit is specified by a micro program.
A microprogram consists of a sequence of instructions in a microprogramming language. These are instructions that specify microoperations.

A microprogrammed control unit is a relatively simple logic circuit that is capable of (1) sequencing through micro instructions and (2) generating control signals to execute each microinstruction.
The concept of micro program is similar to computer program. In computer program the complete instructions of the program is stored in main memory and during execution it fetches the instructions from main memory one after another. The sequence of instruction fetch is controlled by program counter (PC).

Microprogram are stored in microprogram memory and the execution is controlled by microprogram counter (PC).

## 3.1 Basic Concepts of Microprogramming

- **Control word (CW):**
  A word with each bit for one of the control signals. Each step of the instruction execution is represented by a control word with all of the bits corresponding to the control signals needed for the step set to one.
- **Microinstruction:**
  Each step in a sequence of steps in the execution of a certain machine instruction is considered as a *microinstruction*, and it is represented by a control word. All of the bits corresponding to the control signals that need to be asserted in this step are set to 1, and all others are set to 0 (*horizontal organization*).
- **Microprogram:**
  Composed of a sequence of microinstructions corresponding to the sequence of steps in the execution of a given machine instruction.
- **Microprogramming:**
  The method of generating the control signals by properly setting the individual bits in a control word of a step.

**Micro-programmed Control Unit –**

- The control signals associated with operations are stored in special memory units inaccessible by the programmer as Control Words.
- Control signals are generated by a program are similar to machine language programs.
- Micro-programmed control unit is slower in speed because of the time it takes to fetch micro instructions from the control memory.

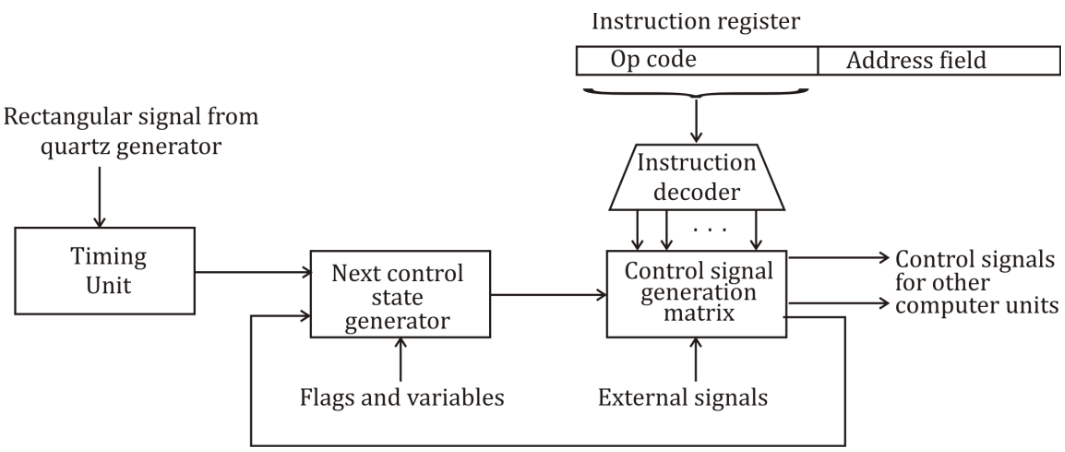
**Some Important Terms –**

1. **Control Word :** A control word is a word whose individual bits represent various control signals.
2. **Micro-routine:** A sequence of control words corresponding to the control sequence of a machine instruction constitutes the micro-routine for that instruction.
3. **Micro-instruction:** Individual control words in this micro-routine are referred to as micro instructions.
4. **Micro-program:** A sequence of micro-instructions is called a micro-program, which is stored in a ROM or RAM called a Control Memory (CM).
5. **Control Store:** the micro-routines for all instructions in the instruction set of a computer are stored in a special memory called the Control Store.

## 3.2 Hardwired Control Unit

The control hardware can be viewed as a state machine that changes from one state to another in every clock cycle, depending on the contents of the instruction register, the condition codes and the external inputs. The outputs of the state machine are the control signals. The sequence of the operation carried out by this machine is determined by the wiring of the logic elements and hence named as "hardwired".

- Fixed logic circuits that correspond directly to the Boolean expressions are used to generate the control signals.
- Hardwired control is faster than micro-programmed control.
- A controller that uses this approach can operate at high speed.



### 3.3 Control address register

Control memory address register specifies the address of the micro-instruction, and the control data register holds the micro-instruction read from memory, The micro-instruction contains a control word that specifies one or more micro operations for the data processor. Once these operations are executed, the control must determine the next address. The location of the next micro-instruction may be the one next in sequence, or it may be located somewhere else in the control memory for this reason it is necessary to use some bits of the present micro-instruction to control the generation of the address of the next micro-instruction. The next address may also be a function of the external input condition.

## 4. Instruction Pipeline

**Instruction pipelining** is a technique used in the design of modern microprocessors, microcontrollers and CPUs to increase their instruction throughput (the number of instructions that can be executed in a unit of time).

Most modern CPUs are driven by a clock. The CPU consists internally of logic and memory (flip flops). When the clock signal arrives, the flip flops store their new value then the logic requires a period of time to decode the flip flops new values. Then the next clock pulse arrives and the flip flops store another values, and so on. By breaking the logic into smaller pieces and inserting flip flops between pieces of logic, the time required by the logic (to decode values till generating valid outputs depending on these values) is reduced. In this way the clock period can be reduced.
For example, the RISC pipeline is broken into five stages with a set of flip flops between each stage as follow:

1. Instruction fetch
2. Instruction decode and register fetch
3. Execute
4. Memory access
5. Register write back

Pipelining can occur in the data stream and the instruction stream.
In the most general case, an instruction cycle can be broken into the following stages:

1. Fetch instruction from memory
2. Decode instruction
3. Calculate effective address(es)
4. Fetch operand(s)
5. Execute instruction
6. Store result

Each stage can be performed more or less independently from the rest.
A queue made of up a small number of registers in the CPU can be filled by the instruction fetch stage while other stages are working on previously fetched instructions. The instruction fetch hardware can prefetch instructions whenever there is space in the queue and memory is not being used for other purposes (fetching operands or storing results).

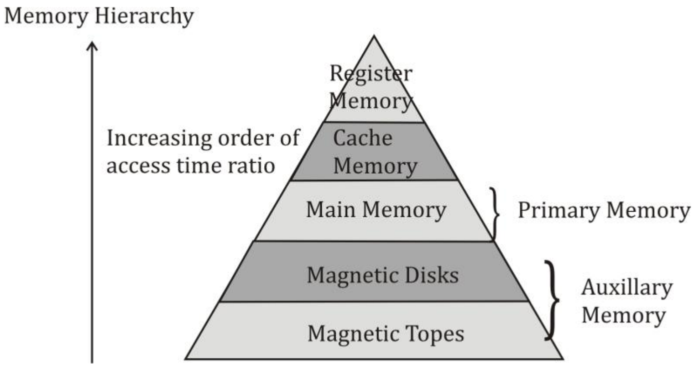
## 5. Memory Organization

A memory unit is the collection of storage units or devices together. The memory unit stores the binary information in the form of bits. Generally, memory/storage is classified into 2 categories:

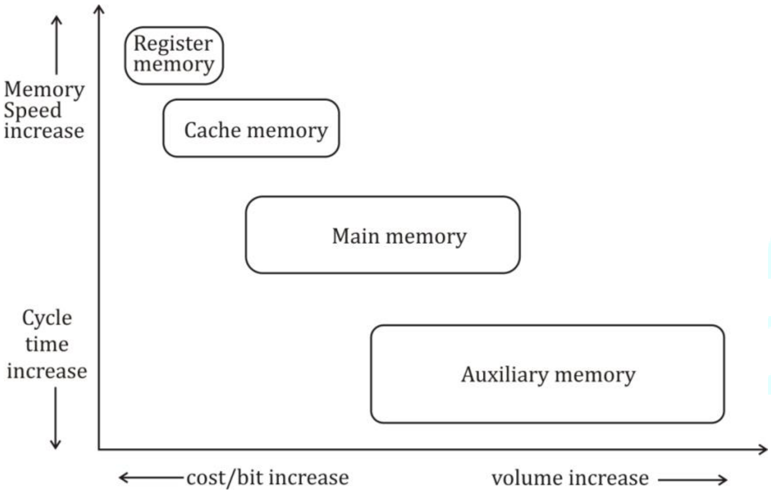
- Volatile Memory: This loses its data, when power is switched off.
- Non-Volatile Memory: This is a permanent storage and does not lose any data when power is switched off.

### 5.1 Memory Hierarchy

The total memory capacity of a computer can be visualized by hierarchy of components. The memory hierarchy system consists of all storage devices contained in a computer system from the slow Auxiliary Memory to fast Main Memory and to smaller Cache memory.




Memory devices in computers can be grouped depending on their access time and their "distance" from the processor (more precisely from the arithmetical-logical unit and control unit), which means the number of elementary transfers when data (instructions) are fetched. The groups have similar parameters such as access time, cycle time, volume, information storage cost per bit. The so defined groups can be set in the order which corresponds to direct mutual access between neighbouring elements, in a Cartesian coordinate system, where the vertical axis corresponds to the memory speed and access time and the horizontal axis corresponds to memory volume and the data storing cost per bit.




## 5.2 Primary and Auxiliary Memory

Main memory unit is the storage unit, there are several locations for storing information in the main memory module. The main memory stores instructions and data of the currently executed programs. Usually it is a random-access memory RAM with reads and writes available. Sometimes, its part can be implemented as the fixed memory or read-only memory ROM. The capacity of a memory module is specified by the number of memory location and the information stored in each location.

A memory module of capacity 16 X 4 indicates that, there are 16 locations in the memory module and in each location, we can store 4 bits of information.
We need two operations to work with memory.

- **READ Operation:** This operation is to retrieve the data from memory and bring it to CPU register
- **WRITE Operation:** This operation is to store the data to a memory location from CPU register

A main memory can be built of a single or many memory modules. A main memory module is built of an address decoder and a set of memory locations. The locations store words of bits of data assigned to consecutive addresses. The word can contain any, but fixed for a given computer, number of bits. There can be several word formats available in the same computer. Usually, the words are so defined as to contain an integer number of bytes. To store one bit of information, a bit cell is used in main memory.

Organization structures of main memories can be divided, according to the circuit that selects memory locations, into the following types:

- Main memory with linear selection (with a single address decoder)
- Main memory with two-dimensional selection (with two address decoders)

- Main memory with linear selection of multiple words (with a single address decoder and a selector)

**Random-Access Memory**

Random-access memory (RAM) comes in two varieties—static and dynamic. Static RAM (SRAM) is faster and significantly more expensive than Dynamic RAM (DRAM). SRAM is used for cache memories, both on and off the CPU chip. DRAM is used for the main memory plus the frame buffer of a graphics system. Typically, a desktop system will have no more than a few megabytes of SRAM, but hundreds or thousands of megabytes of DRAM.

**Static RAM**

SRAM stores each bit in a bi-stable memory cell. Each cell is implemented with a six-transistor circuit. This circuit has the property that it can stay indefinitely in either of two different voltage configurations, or states. Any other state will be unstable—starting from there, the circuit will quickly move toward one of the stable states.
Due to its bi-stable nature, an SRAM memory cell will retain its value indefinitely, as long as it is kept powered. Even when a disturbance, such as electrical noise, perturbs the voltages, the circuit will return to the stable value when the disturbance is removed.

**Dynamic RAM**

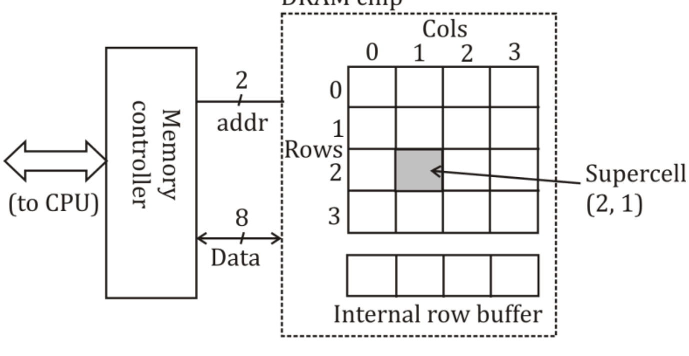
DRAM stores each bit as charge on a capacitor. This capacitor is very small. DRAM storage can be made very dense—each cell consists of a capacitor and a single access-transistor. Unlike SRAM, however, a DRAM memory cell is very sensitive to any disturbance. When the capacitor voltage is disturbed, it will never recover. Exposure to light rays will cause the capacitor voltages to change. In fact, the sensors in digital cameras and camcorders are essentially arrays of DRAM cells. Various sources of leakage current cause a DRAM cell to lose its charge within a time period of around 10 to 100 milliseconds. Fortunately, for computers operating with clock cycles times measured in nanoseconds, this retention time is quite long. The memory system must periodically refresh every bit of memory by reading it out and then rewriting it. Some systems also use error-correcting codes, where the computer words are encoded a few more bits (e.g., a 32-bit word might be encoded using 38 bits), such that circuitry can detect and correct any single erroneous bit within a word.)

SRAM is persistent as long as power is applied to them. Unlike DRAM, no refresh is necessary. SRAM can be accessed faster than DRAM. SRAM is not sensitive to disturbances such as light and electrical noise. The trade-off is that SRAM cells use more transistors than DRAM cells, and thus have lower densities, are more expensive, and consume more power.

**Conventional DRAMs**

The cells (bits) in a DRAM chip are partitioned into *d supercells*, each consisting of *w DRAM cells*. A $d \times w$ DRAM stores a total of $dw$ bits of information. The supercells are organized as a rectangular array with $r$ rows and $c$ columns, where $rc = d$. Each supercell has an address of the form $(i,j)$, where $i$ denotes the row, and $j$ denotes the column.

For example, consider the organization of a 16 × 8 DRAM chip with d = 16 supercells, w = 8bits per supercell, r = 4 rows, and c = 4 columns. The shaded box denotes the supercell at address (2, 1). Information flows in and out of the chip via external connectors called pins. Each pin carries a 1-bit signal.




High level view of a 128-bit 16 × 8 DRAM chip.

**RAS and CAS Request:**

Each DRAM chip is connected to some circuitry, known as the memory controller, that can transfer w bits at a time to and from each DRAM chip. To read the contents of supercell (i,j), the memory controller sends the row address i to the DRAM, followed by the column address j. The DRAM responds by sending the contents of supercell (i,j) back to the controller. The row address i is called **a RAS (Row Access Strobe)** request. The column address j is called a **CAS (Column Access Strobe)** request. Notice that the RAS and CAS requests share the same DRAM address pins.

**Memory Modules:** DRAM chips are packaged in memory modules that plug into expansion slots on the main system board (motherboard). Common packages include the 168-pin dual inline memory module (DIMM), which transfers data to and from the memory controller in 64-bit chunks, and the 72-pin single inline memory module (SIMM), which transfers data in 32-bit chunks.

**Enhanced DRAMs**

There are many kinds of DRAM memories, and new kinds appear on the market with regularity as manufacturers attempt to keep up with rapidly increasing processor speeds.

- **Fast page mode DRAM (FPM DRAM):** A conventional DRAM copies an entire row of supercells into its internal row buffer, uses one, and then discards the rest. FPM DRAM improves on this by allowing consecutive accesses to the same row to be served directly from the row buffer.
- **Extended data out DRAM (EDO DRAM):** An enhanced form of FPM DRAM that allows the individual CAS signals to be spaced closer together in time.
- **Synchronous DRAM (SDRAM):** Conventional, FPM, and EDO DRAMs are asynchronous in the sense that they communicate with the memory controller using a set of explicit control signals. SDRAM replaces many of these control signals with the rising edges of the same external clock signal that drives the memory controller. Without going into detail, the net effect is that an SDRAM can output he contents of its supercells at a faster rate than its asynchronous counterparts.
- **Double Data-Rate Synchronous DRAM (DDR SDRAM):** DDR SDRAM is an enhancement of SDRAM that doubles the speed of the DRAM by using both clock edges as control signals. Different types of DDR SDRAMs are characterized by the size of a small prefetch buffer that increases the effective bandwidth: DDR (2 bits), DDR2 (4 bits), and DDR3 (8 bits).
- **Rambus DRAM (RDRAM):** This is an alternative proprietary technology with a higher maximum bandwidth than DDR SDRAM.
- **Video RAM (VRAM):** Used in the frame buffers of graphics systems. VRAM is similar in spirit to FPM DRAM. Two major differences are that (1) VRAM output is produced by shifting the entire contents of the internal buffer in sequence, and (2) VRAM allows concurrent reads and writes to the memory. Thus, the system can be painting the screen with the pixels in the frame buffer (reads) while concurrently writing new values for the next update (writes).

*Non-volatile Memory*

This category of memory retains the information even when they are powered off. There are a variety of non-volatile memories. For historical reasons, they are referred to collectively as **read-only memories (ROMs)**, even though some types of ROMs can be written to as well as read. ROMs are distinguished by the number of times they can be reprogrammed (written to) and by the mechanism for reprogramming them.

A programmable ROM (PROM) can be programmed exactly once. PROMs include a sort of fuse with each memory cell that can be blown once by zapping it with a high current. An erasable programmable ROM (**EPROM**) has a transparent quartz window that permits light to reach the storage cells. The EPROM cells are cleared to zeros by shining ultraviolet light through the window. Programming an EPROM is done by using a special device to write ones into the EPROM. An EPROM can be erased and reprogrammed on the order of 1000 times.

An electrically erasable PROM **(EEPROM)** is akin to an EPROM, but does not require a physically separate programming device, and thus can be reprogrammed in-place on printed circuit cards. An EEPROM can be reprogrammed on the order of 105 times before it wears out.

Flash memory is a type of non-volatile memory, based on EEPROMs, that has become an important storage technology. Flash memories are everywhere, providing fast and durable non-volatile storage for a slew of electronic devices, including digital cameras, cell phones, music players, PDAs, and laptop, desktop, and server computer systems.

Programs stored in ROM devices are often referred to as firmware. When a computer system is powered up, it runs firmware stored in a ROM. Some systems provide a small set of primitive input and output functions in firmware, for example, a PC's BIOS (basic input/output system) routines. Complicated devices such as graphics cards and disk drive controllers also rely on firmware to translate I/O (input/output) requests from the CPU.

**Secondary Memory/Auxiliary Memory**
Also termed as 'auxiliary' or 'backup' storage, it is typically used as a supplement to main storage. It is much cheaper than the main storage and stores large amount of data and instructions permanently. Hardware devices like magnetic tapes and disks fall under this category.

**Magnetic Disk**
The Magnetic Disk is Flat, circular platter with metallic coating that is rotated beneath read/write heads. It is a Random access device; read/write head can be moved to any location on the platter.

**Floppy Disk**
These were small removable disks that are plastic coated with magnetic recording material. Floppy disks were typically 3.5″ in size (diameter) and could hold 1.44 MB of data (their use has been discontinued). This portable storage device is a rewritable media and can be reused a number of times.

**HARD DISK**
Another form of auxiliary storage is a hard disk. A hard disk consists of one or more rigid metal plates coated with a metal oxide material that allows data to be magnetically recorded on the surface of the platters. The hard disk platters spin at a high rate of speed, typically 5400 to 7200 revolutions per minute (RPM). Storage capacities of hard disks for personal computers range from 10 GB to 120 GB (one billion bytes are called a gigabyte).

**Optical Disks: CD-R Drive, CD-RW disks, DVD, Blue ray Discs**
Optical Mass Storage Devices Store bit values as variations in light reflection. They have higher area density & longer data life than magnetic storage. They are also Standardized and relatively inexpensive. Their Uses: read-only storage with low performance requirements, applications with high capacity requirements & where portability in a standardized format is needed.

**Their Types:**
– CD-ROM (read only)
– CD-R: (record) to a CD
– CD-RW: can write and erase CD to reuse it (re-writable)
– DVD (Digital Video Disk)

**Blu-ray Technology**
The name is derived from the blue-violet laser used to read and write data. It was developed by the Blu-ray Disc Association with more than 180 members. Some companies with the technology are Dell, Sony, LG. The Data capacity is very large because Blu-ray uses a blue laser (405 nanometres) instead of a red laser(650 nanometres) this allows the data tracks on the disc to be very compact. This allows for more than twice as small pits as on a DVD. Because of the greatly compact data Blu-ray can hold almost 5 times more data than a single layer DVD.

The Variations in the formats are as follows:
• BD-ROM (read-only) - for pre-recorded content
• BD-R (recordable) - for PC data storage
• BD-RW (rewritable) - for PC data storage
• BD-RE (rewritable) - for HDTV recording

**Pen Drive**: A flash drive consists of a small printed circuit board carrying the circuit elements and a USB connector, insulated electrically and protected inside a plastic, metal, or rubberized case which can be carried in a pocket or on a key chain, for example; the USB connector may be protected by a removable cap or by retracting into the body of the drive, although it is not likely to be damaged if unprotected. Most flash drives use a standard type-A USB connection allowing plugging into a port on a personal computer, but drives for other interfaces also exist.
USB flash drives draw power from the computer via the USB connection. Some devices combine the functionality of a digital audio player with USB flash storage; they require a battery only when used to play music.

## 5.3 Cache Memory Organization

Cache Memory is a fast random access memory where the computer hardware stores copies of information currently used by programs (data and instructions), loaded from the main memory. The cache has a significantly shorter access time than the main memory due to the applied faster but more expensive implementation technology. The ***cache has a*** *limited volume* that also results from the properties of the applied technology. If information fetched to the cache memory is used again, the access time to it will be much shorter than in the case if this information were stored in the main memory and the program will execute faster.

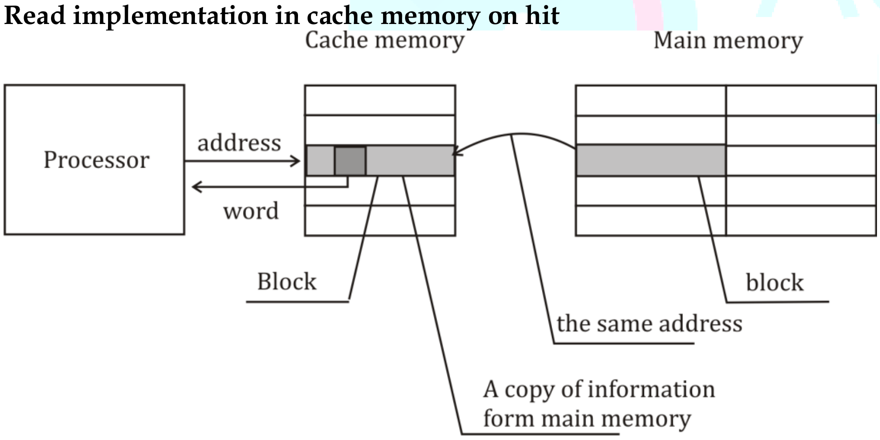
Time efficiency of using cache memories results from the locality of access to data that is observed during program execution. We observe here time and space locality:

- **Time locality** consists in a tendency to use many times the same instructions and data in programs during neighbouring time intervals,
- **Space locality** is a tendency to store instructions and data used in a program in short distances of time under neighbouring addresses in the main memory.

Due to these localities, the information loaded to the cache memory is used several times and the execution time of programs is much reduced. Cache can be implemented as a multi-level memory. Contemporary computers usually have two levels of caches. In older computer models, a cache memory was installed outside a processor (in separate integrated circuits than the processor itself). The access to it was organized over the processor external system bus. In today's computers, the first level of the cache memory is installed in the same integrated circuit as the processor. It significantly speeds up processor's co-operation with the cache. Some microprocessors have the second level of cache memory placed also in the processor's integrated circuit. The volume of the first level cache memory is from several thousands to several tens of thousands of bytes. The second level cache memory has volume of several hundred thousand bytes. A cache memory is maintained by a special processor subsystem called cache controller.

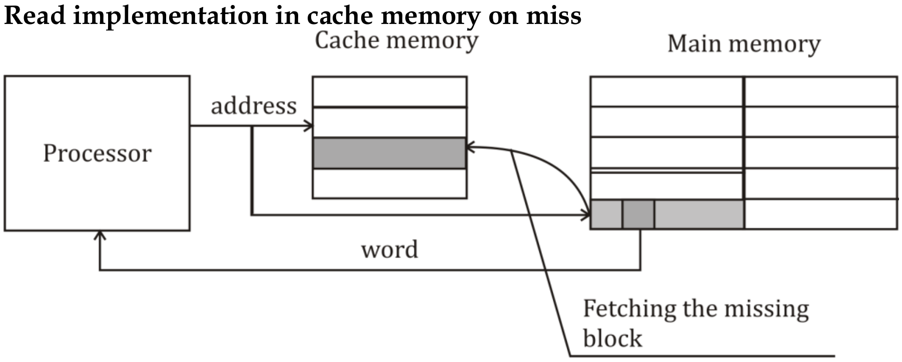
If there is a cache memory in a computer system, then at each access to a main memory address in order to fetch data or instructions, processor hardware sends the address first to the cache memory. The cache control unit checks if the requested information resides in the cache. If so, we have a "hit" and the requested information is fetched from the cache.

**Read implementation in cache memory on hit**




If the requested information does not reside in the cache, we have a "miss" and the necessary information is fetched from the main memory to the cache and to the requesting processor unit. The information is not copied in the cache as single words but as a larger block of a fixed volume. Together with information block, a part of the address of the beginning of the block is always copied into the cache. This part of the address is next used at readout during identification of the proper information block.

**Read implementation in cache memory on miss**



If there are two cache levels, then on "miss" at the first level, the address is transferred in a hardwired way to the cache at the second level. If at this level a "hit" happens, the block that contains the requested word is fetched from the second level cache to the first level cache. If a "miss" occurs also at the second cache level, the blocks containing the requested word are fetched to the cache memories at both levels. The size of the cache block at the first level is from 8 to several tens of bytes (a number must be a power of 2). The size of the block in the second level cache is many times larger than the size of the block at the first level.

## 5.4 Cache Mapping

Cache Mapping is necessary as there are far fewer number of available cache addresses than the memory. Cache mapping used to assign main memory address to cache address and determine hit or miss. Caches are partitioned into indivisible blocks or lines of adjacent memory addresses usually 4 or 8 addresses per line.
There are three basic methods used for mapping of information fetched from the main memory to the cache memory:

- associative mapping
- direct mapping
- set-associative mapping

**Cache memory with direct mapping**
The simplest way of associating main memory blocks with cache block is the direct mapping technique. In this technique, block k of main memory maps into block k modulo m of the cache, where m is the total number of blocks in cache. In this example, the value of m is 128. In direct mapping technique, one particular block of main memory can be transferred to a particular block of cache which is derived by the modulo function.
Since more than one main memory block is mapped onto a given cache block position, contention may arise for that position. This situation may occur even when the cache is not full. Contention is resolved by allowing the new block to overwrite the currently resident block. So, the replacement algorithm is trivial.
The detail operation of direct mapping technique is as follows:

- The main memory address is divided into three fields. The field size depends on the memory capacity and the block size of cache. In this example, the lower 5 bits of address is used to identify a word within a block. Next 7 bits are used to select a block out of 128 blocks (which is the capacity of the cache). The remaining 4 bits are used as a TAG to identify the proper block of main memory that is mapped to cache.
- When a new block is first brought into the cache, the high order 4 bits of the main memory address are stored in four TAG bits associated with its location in the cache. When the CPU generates a memory request, the 7-bit block address determines the corresponding cache block. The TAG field of that block is compared to the TAG field of the address. If they match, the desired word specified by the low-order 5 bits of the address is in that block of the cache.
- If there is no match, the required word must be accessed from the main memory, that is, the contents of that block of the cache is replaced by the new block that is specified by the new address generated by the CPU and correspondingly the TAG bit will also be changed by the high order 4 bits of the address.

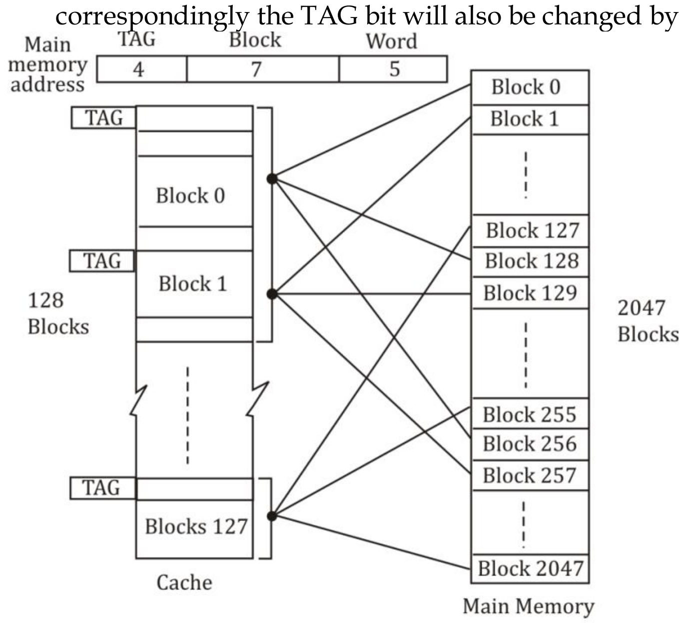
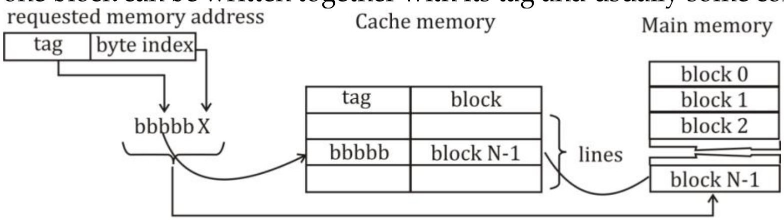



**Figure:** Direct-mapping cache

**Cache memory with associative mapping**

With the associative mapping of the contents of cache memory, the address of a word in the main memory is divided into two parts: the tag and the byte index (offset). Information is fetched into the cache in blocks. The byte index determines the location of the byte in the block whose address is generated from the tag bits, which are extended by zeros in the index part (it corresponds to the address of the first byte in the block. In the number of bits in the byte index is n then the size of the block is a power of 2 with the exponent n. The cache is divided into lines. In each line one block can be written together with its tag and usually some control bits.




Information organization in cache with associative mapping

The principle of the read operation in cache memory is shown below. The requested address contains the tag (bbbbb) and the byte index in the block (X). The tag is compared in parallel with all tags written down in all lines. If a tag match is found in a line, we have a hit and the line contains the requested information block. Then, based on the byte index, the requested byte is selected in the block and read out into the processor. If none of the lines contains the requested tag, the requested block does not reside in the cache. The missing block is next fetched from the main memory or an upper level cache memory.

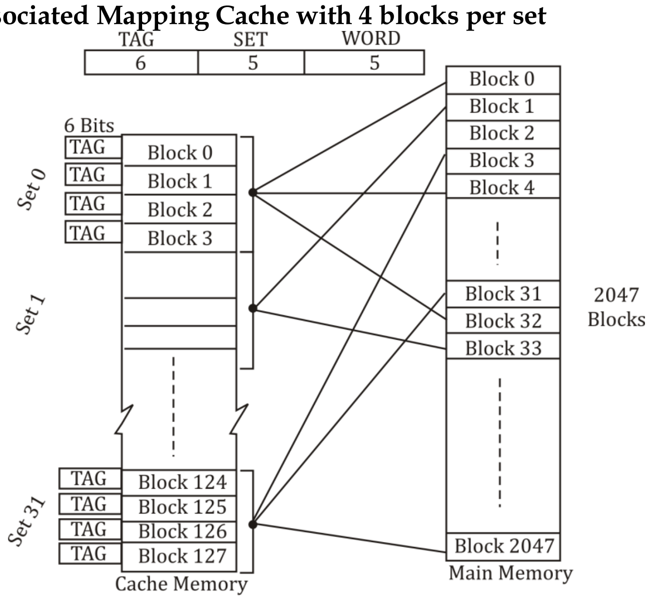
**Cache memory with set associative mapping**

This mapping technique is intermediate to the previous two techniques. Blocks of the cache are grouped into sets, and the mapping allows a block of main memory to reside in any block of a specific set. Therefore, the flexibity of associative mapping is reduced from full freedom to a set of specific blocks. This also reduces the searching overhead, because the search is restricted to number of sets, instead of number of blocks. Also, the contention problem of the direct mapping is eased by having a few choices for block replacement.

Consider the same cache memory and main memory organization of the previous example. Organize the cache with 4 blocks in each set. The TAG field of associative mapping technique is divided into two groups, one is termed as SET bit and the second one is termed as TAG bit. Each set contains 4 blocks, total number of set is 32. The main memory address is grouped into three parts: low-order 5 bits are used to identifies a word within a block. Since there are total 32 sets present, next 5 bits are used to identify the set. High-order 6 bits are used as TAG bits.

The 5-bit set field of the address determines which set of the cache might contain the desired block. This is similar to direct mapping technique, in case of direct mapping, it looks for block, but in case of block-set-associative mapping, it looks for set. The TAG field of the address must then be compared with the TAGs of the four blocks of that set. If a match occurs, then the block is present in the cache; otherwise the block containing the addressed word must be brought to the cache. This block will potentially come to the corresponding set only. Since, there are four blocks in the set, we have to choose appropriately which block to be replaced if all the blocks are occupied. Since the search is restricted to four block only, so the searching complexity is reduced.

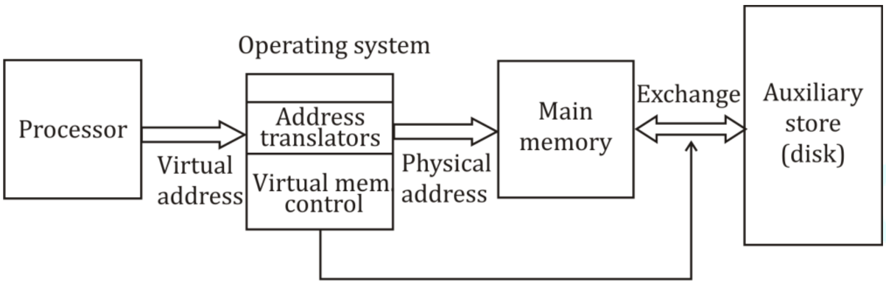
**Set Associated Mapping Cache with 4 blocks per set**



## 5.5. Virtual Memory

In early computers, freedom of programming was seriously restricted by a limited volume of main memory comparing program sizes. Small main memory volume was making large programs execution very troublesome and did not enable flexible maintenance of memory space in the case of many co-existing programs. It was very uncomfortable, since programmers were forced to spend much time on designing a correct scheme for data and code distribution among the main memory and auxiliary store. The solution to this problem was supplied by introduction of the virtual memory concept. This concept was introduced at the beginning of years 1970 under the name of one-level storage in the British computer called Atlas. Only much later, together with application of this idea in computers of the IBM Series 370, the term virtual memory was introduced.

Virtual memory provides a computer programmer with an addressing space many times larger than the physically available addressing space of the main memory. Data and instructions are placed in this space with the use of virtual addresses, which can be treated as artificial in some way. In the reality, data and instructions are stored both in the main memory and in the auxiliary memory (usually disk memory). It is done under supervision of the virtual memory control system that governs real current placement of data determined by virtual addresses. This system automatically (i.e. without any programmer's actions) fetches to the main memory data and instructions requested by currently executed programs.

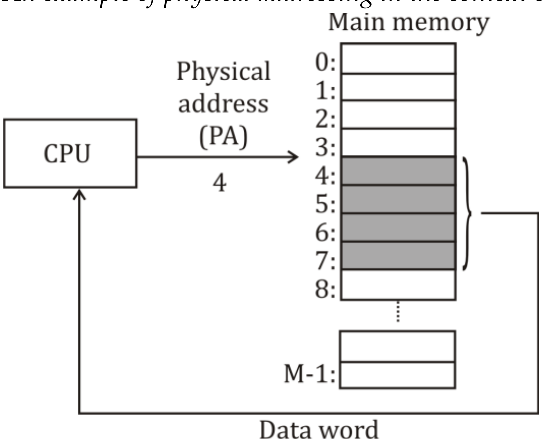



General scheme of the virtual memory

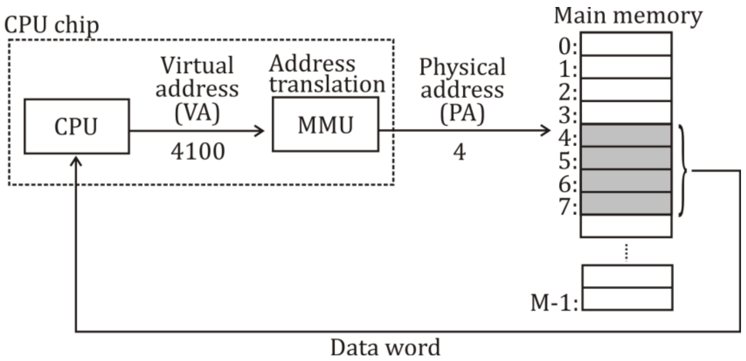
**Physical and Virtual Addressing**

The main memory of a computer system is organized as an array of M contiguous byte-sized cells. Each byte has a unique physical address (PA). The first byte has an address of 0, the next byte an address of 1, the next byte an address of 2, and so on. Given this simple organization, the most natural way for a CPU to access memory would be to use physical addresses. We call this approach physical addressing.

*An example of physical addressing in the context of a load instruction that reads the word starting at physical address 4:*




When the CPU executes the load instruction, it generates an effective physical address and passes it to main memory over the memory bus. The main memory fetches the 4-byte word starting at physical address 4 and returns it to the CPU, which stores it in a register. Early PCs used physical addressing, and systems such as digital signal processors, embedded microcontrollers, and Cray supercomputers continue to do so. However, modern processors use a form of addressing known as **virtual addressing.**

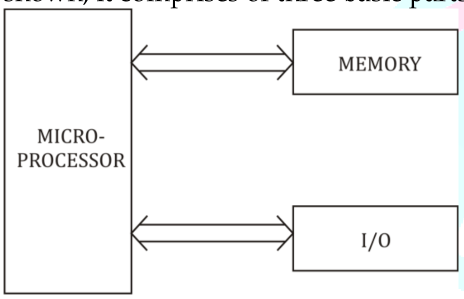



With virtual addressing, the CPU accesses main memory by generating a virtual address (VA), which is converted to the appropriate physical address before being sent to the memory. The task of converting a virtual address to a physical one is known as address translation. Like exception handling, address translation requires close cooperation between the CPU hardware and the operating system. Dedicated hardware on the CPU chip called the memory management unit (MMU) translates virtual addresses on the fly, using a look-up table stored in main memory whose contents are managed by the operating system.

## 6. Microprocessor Architecture

**MICROPROCESSOR ARCHITECHTURE-**

A computer, large or small, can be represented functionally (in a simplified form) by the block diagram in Figure. As shown, it comprises of three basic parts or sub-systems:




**(a) Central Processing Unit (CPU)**

It performs the necessary arithmetic and logic operations and controls the timing and general operation of the complete system.

**(b) Input/ Output (I/O) Devices**

Input devices are used for feeding data into the CPU, examples of these devices are toggle switches, analog-to-digital converters, paper tape readers, card readers, keyboards, disk etc.

The output devices are used for delivering the results of computations to the outside world; examples are light emitting diodes, cathode ray tube (CRT) displays, digital-to-analog converters, card and paper-tape punches, character printers, plotters, communication lines etc. The input output subsystem thus allows the computer to usefully communicate with the outside world. Input-output devices are also called as peripherals.

**(c) Memory**

It stores both the instructions to be executed (i.e., the program) and the data involved. It usually consists of both RAMs (random-access memories) and ROMS (read-only memories). A microprocessor is an integrated circuit designed to function as the CPU of a microcomputer.

**Internal Architecture of a MICROPROCESSOR**

The microprocessor or CPU reads each instruction from the memory, decodes it and executes it. It processes the data as required in the instructions. The processing is in the form of arithmetic and logical operations. The data is retrieved from memory or taken from an input device and the result of processing is stored in the memory or delivered to an appropriate output device, all as per the instructions.

To perform all these functions, the µP (microprocessor) incorporates various functional units in an appropriate manner. Such an internal structure or organizational structure of µP, which determines how it operates, is known as its architecture.

A typical microprocessor architecture is shown in Figure. The various functional units are as follows:

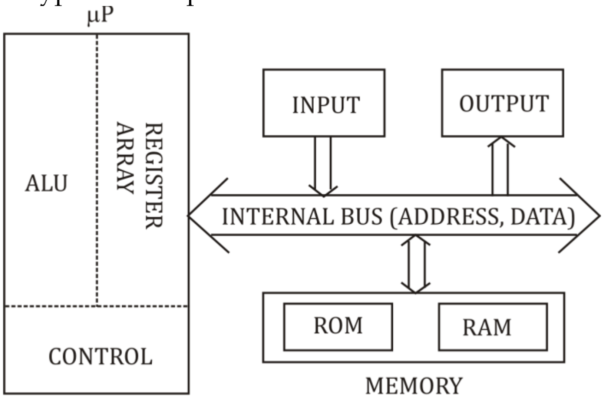


Figure : Architecture of Microprocessor

**Busses-**
µC (microcomputer), like all computers, manipulates binary information. The binary information is represented by binary digits, called bits. µC operates on a group of bits which are referred to as a word. The number of bits making-µP a word varies with the µP. Common word sizes are 4, 8, 12 and 16 bits (µPs with 32 bit-word have also of late entered the market). Another binary terms that will be of interest in subsequent discussions are the byte and the nibble, which represent a set of 8 bits and 4 bits, respectively.
Figure shows busses interconnecting various blocks. These busses allow exchange of words between the blocks. A bus has a wire or line for each bit and thus allows exchange of all bits of a word in parallel. The processing of bits in the µP is also in parallel. The busses can thus be viewed as data highways. The width of a bus is the number of signal lines that constitute the bus.
The figure shows for simplicity three busses for distinct functions. Over the address bus, the µP transmits the address of that I/O device or memory locations which it desires to access. This address is received by all the devices connected to the processor, but only the device which has been addressed responds. The data bus is used by the µP to send and receive data to and from different devices (I/O and memory) including instructions stored in memory. Obviously the address bus is unidirectional and the data bus is bi-directional. The control bus is used for transmitting and receiving control signals between the µP and various devices in the system.

**Arithmetic-Logic Unit (ALU)**
The arithmetic-logic unit is a combinational network that performs arithmetic and logical operations on the data.

**Internal Registers**
A number of registers are normally included in the microprocessor. These are used for temporary storage of data, instructions and addresses during execution of a program. Those in the Intel µP.
**8085 microprocessor are typical and are described below:**
**(i) Accumulator (Acc) or Result Register**
This is an 8-bit register used in various arithmetic and logical operations. Out of the two operands to be operated upon, one comes from accumulator (Acc), whilst the other one may be in another internal register or may be brought in by the data bus from the main memory. Upon completion of the arithmetic/logical operation, the result is placed in the accumulator (replacing the earlier operand). Because of the later function, this register is also called as result register.

**(ii) General Purpose Registers or Scratch Pad Memory**
There are six general purpose 8-bit registers that can be used by the programmer for a variety of purposes. These registers, labelled as B, C, D, E, H and L, can be used individually (e.g., when operation on 8-bit data is desired) or in pairs (e.g., when a 16-bit address is to be stored). Only B-C, D-E and H-L pairs are allowed.

**(iii) Instruction Register (IR)**
This 8-bit register stores the next instruction to be executed. At the proper time this stored word (instruction) is fed to an instruction decoder which decodes it and supplied appropriate signals to the control unit. When the execution has been accomplished the new word in the instruction register is processed.

### (iv) Program Counter (PC)

This is a 16-bit register which holds the address of the next instruction that has to be fetched from the main memory and loaded into the instruction register. The program controlling the operation is stored in the main memory and instructions are retrieved from this memory normally in order. Therefore, normally the address contained in the PC is incremented after each instruction is fetched. However, certain classes of instruction can modify the PC so that the programmer can provide for branching away from the normal program flow. Examples are instructions in the "jump" and 'call subroutine' groups.

### (v) Stack Pointer (SP)

This is also a 16-bit register and is used by the programmer to maintain a stack in the memory while using subroutines.

### (vi) Status Register or Condition Flags

A status register consisting of a few flip-flops, called as condition flags (in 8085 the number of flags is five) is used to provide indication of certain conditions that arise during arithmetic and logical operations.
These are:
**'zero'** Flag is set if result of instruction is 0.
**'sign'** Set if MSB of result is 1.
'**parity'** Set if result has even parity.
'**carry'** Set if carry or borrow resulted.
'**auxiliary carry'** Set if instruction caused a carry out of bit 3 and into bit 4 of the resulting value.

### (vii) Dedicated Registers

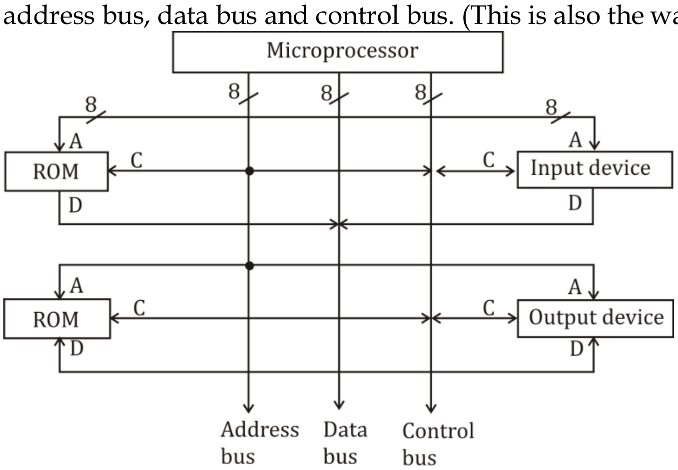
Several other registers are incorporated in the μP for its internal operation. They cannot be accessed by the programmer and hence do not concern much a μP user.

### Instruction Decoder and Control Unit

It decodes each instruction and under the supervision of a clock controls the external and internal units ensuring correct logical operation of the system.

## SEMICONDUCTOR MEMORIES

As mentioned earlier, semiconductor memories are required in a microcomputer for storing information which may comprise of (a) the data to be used for computation, (b) instructions and (c) computational results. A program starts as a set of instructions on a paper, then this is transferred to a set of cards with the instructions punched in code on them. These instructions also can be transferred to magnetic tape, paper tape or directly into semiconductor memory which is the eventual storage space for a program. The semiconductor memory chips are connected to the μP through the address bus, data bus and control bus. (This is also the way that I/O devices are connected to the μP).




## PERIPHERAL INTERFACING

### Functions

When one or more I/O devices (peripherals) are to be connected to a μP, an interface network for each device, called peripheral interface, is required.

**The interface incorporate commonly the following four functions:**
**(a) Buffering** : Which is necessary to take care of incompatibility between the μP and the peripheral.

**(b) Address Decoding** : Which is required to select one of the several peripherals connected in the system.
**(c) Command Decoding** : Which is required for peripherals that perform actions other than data transfers.
**(d) Timing and Control** : All the above functions require timing and control.

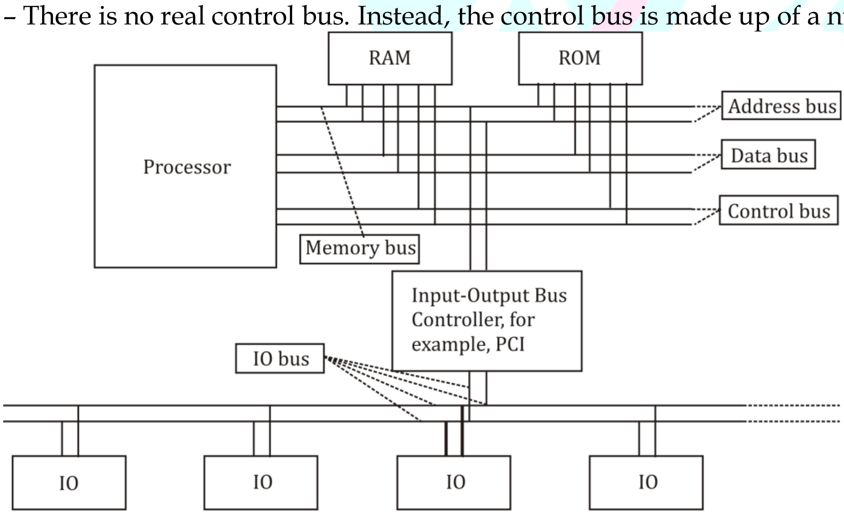
## 7. Microprocessor Bus Organisation

The 8085 uses three separate busses to perform its operations
– The address bus.
– The data bus.
– The control bus.

**The Address Bus**
– 16 bits wide (A0 A1…A15)
• Therefore, the 8085 can access locations with numbers from 0 to 65,536. Or, the 8085 can access a total of 64K addresses.
– "Unidirectional".
• Information flows out of the microprocessor and into the memory or peripherals.
– When the 8085 wants to access a peripheral or a memory location, it places the 16-bit address on the address bus and then sends the appropriate control signals.

**The Data Bus**
– 8 bits wide (D0 D1…D7)
– "Bi-directional".
• Information flows both ways between the microprocessor and memory or I/O.
– The 8085 uses the data bus to transfer the binary information.
– Since the data bus has 8-bits only, then the 8085 can manipulate data 8 bits at-a-time only.

**The Control Bus**
– There is no real control bus. Instead, the control bus is made up of a number of single bit control signals.

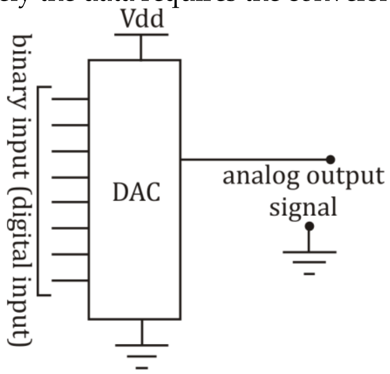



**Devices on I/O Bus Devices on I/O Bus-**
Devices can be designed to interface with the bus, allowing them to be compatible with any computer that uses the same type of I/O bus.

## 8. Digital to Analog Converters

A D/A Converter is used when the binary output from a digital system is to be converted into its equivalent analog voltage or current. The binary output will be a sequence of 1's and 0's. Thus they may be difficult to follow. But, a D/A converter help the user to interpret easily.
Digital to Analog Converter (DAC) is a device that transforms digital data into an analog signal. According to the Nyquist-Shannon sampling theorem, any sampled data can be reconstructed perfectly with bandwidth and Nyquist criteria.

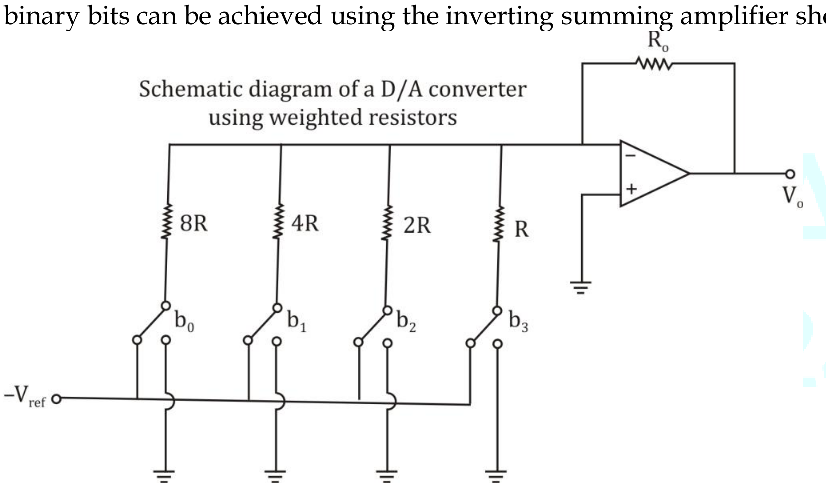
A DAC can reconstruct sampled data into an analog signal with precision. The digital data may be produced from a microprocessor, Application Specific Integrated Circuit (ASIC), or Field Programmable Gate Array (FPGA), but ultimately the data requires the conversion to an analog signal in order to interact with the real world.




Basic Digital to Analog Converter

**DAC using Weighted Resistors method**

The basic operation of DAC is the ability to add inputs that will ultimately correspond to the contributions of the various bits of the digital input. In the voltage domain, that is if the input signals are voltages, the addition of the binary bits can be achieved using the inverting summing amplifier shown in the below figure.

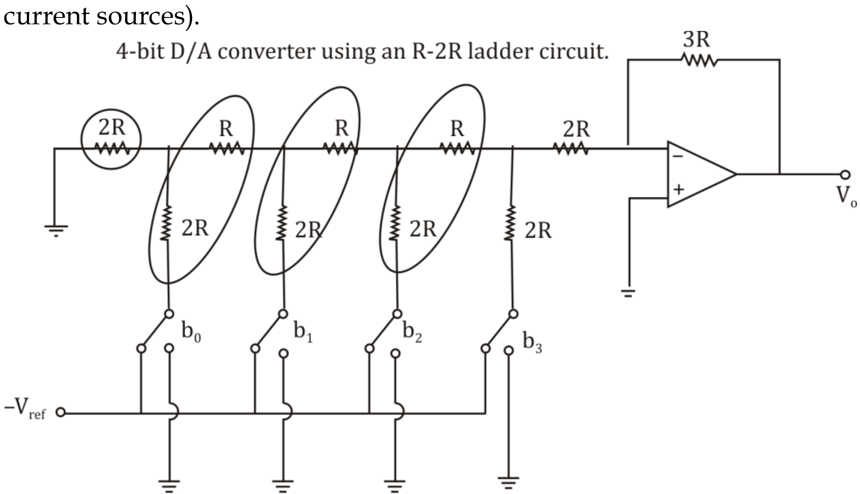



The input resistors of the op-amp have their resistance values weighted in a binary format. When the receiving binary 1 the switch connects the resistor to the reference voltage. When the logic circuit receives binary 0, the switch connects the resistor to ground. All the digital input bits are simultaneously applied to the DAC.

As the number of bits is increasing in the digital input voltage, the range of the resistor values becomes large and accordingly, the accuracy becomes poor.

**R-2R Ladder Digital to Analog Converter (DAC)**

The R-2R ladder DAC constructed as a binary-weighted DAC that uses a repeating cascaded structure of resistor values R and 2R. This improves the precision due to the relative ease of producing equal valued-matched resistors (or current sources).



## 9. Boolean Algebra and Logic Gates

The flow of digital signals is controlled by transistors in various configurations depending on the logic family. For most purposes we can imagine that the logic gates are composed of ideal switches with just two states: OPEN and CLOSED. The state of a switch is controlled by a digital signal. The switch remains closed so long as a logical (1) signal is applied. A logical (0) control signal keeps it open.

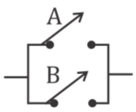
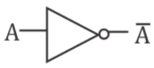
Logic signals interact by means of gates. The three fundamental gates AND, OR, and NOT, are named after the three fundamental operations of logic that they carry out. The AND and OR gates each have two inputs and one output. The output state is determined by the states of the two inputs.

The function of each gate is defined by a truth table, which specifies the output state for each possible combination of input states. The output values of the truth tables can be understood in terms of two switches. If the switches are in series, you get the AND function. Parallel switches perform the OR operation. A bubble after a gate or at an input indicates NOT.
The three compound gates NAND, NOR and XOR can be made from AND, OR, and NOT. NAND means an AND gate followed by a NOT, while NOR means an OR gate followed by a NOT. The EXCLUSIVE-OR (XOR) is similar to OR but it has a low output if both inputs are High.

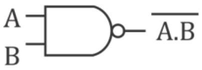
**Types of Gate-**
1- NOT or INVERTER Gate

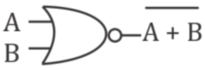
2- AND Gate

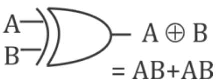
3- OR Gate

| Operation | Switches | Condition that Circuit is closed | Boolean Notation | symbol | Truth Table |
|---|---|---|---|---|---|
| AND |  Series | (A and B are close) | A • B or AB |  | A B \| A·B<br>0 0 \| 0<br>0 1 \| 0<br>1 0 \| 0<br>1 1 \| 1 |
| OR |  Paralle | (A OR B is closed) | A + B |  | A B \| A + B<br>0 0 \| 0<br>0 1 \| 1<br>1 0 \| 1<br>1 1 \| 1 |
| NOT (same as invert) | Different Kind of switch | 1 means open<br>0 means closed | NOT A ≡ $\overline{A}$ |  | A \| $\overline{A}$<br>0 \| 1<br>1 \| 0 |

**Compound Gates-**
There are three types of gates-
1- NAND
2- NOR
3- XOR

NAND 


NOR 


XOR 


### 9.1 BOOLEAN ALGEBRA

We imagine a logical variable, A, that takes on the values 0 or 1. If A = 0 then A = 1 and if A = 1 then A = 0. Here are some obvious identities using the AND, OR and NOT operations.

Looking at these identities you can see why the 'plus' symbol was chosen for OR and 'times' was chosen for AND.

| OR | AND | NOT |
|---|---|---|
| $A+0=A$ | $A\bullet 0=0$ | $A+\bar{A}=1$ |
| $A+1=1$ | $A\bullet 1=A$ | $A\bullet \bar{A}=0$ |
| $A+A=A$ | $A\bullet A=A$ | $\bar{\bar{A}}=A$ |
| $A+\bar{A}=1$ | $A\bullet \bar{A}=0$ | |

**The basic Laws related to Boolean expressions-**

**1. Equality**

Two Boolean expressions are equal if and only if their truth tables are identical.

**2. Associative Laws**

(A + B)+ C = A+ (B + C)

(AB)C = A(BC)

**3. Distributive Laws**

A(B + C)= AB + AC

**4. DeMorgan's Theorems**

A• B •K = A + B +K

A + B +K = A •B • K

**Example of Proof**

Each of the above equalities is a theorem that can be proved. Let's do an example by directly comparing the truth tables for the left and right sides. We take on DeMorgan's first theorem for two variables,

AB = A + B :

| A | B | AB | $\overline{AB}$ |
|---|---|---|---|
| 0 | 0 | 0 | 1 |
| 0 | 1 | 0 | 1 |
| 1 | 0 | 0 | 1 |
| 1 | 1 | 1 | 0 |

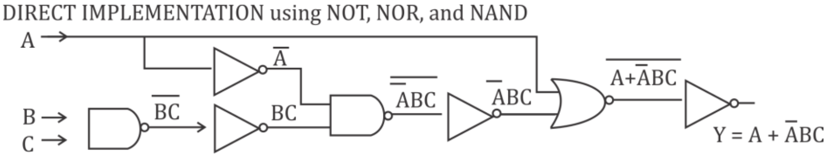
| $\bar{A}$ | $\bar{B}$ | $\bar{A}$ | $\bar{B}$ | $\bar{A}+\bar{B}$ |
|---|---|---|---|---|
| 0 | 0 | 1 | 1 | 1 |
| 0 | 1 | 1 | 0 | 1 |
| 1 | 0 | 0 | 1 | 1 |
| 1 | 1 | 0 | 0 | 0 |

**Example of simplification**

Boolean algebra can be used to simplify logical expressions and reduce the number of gates required in a circuit. In figure, we show two ways to implement the expression,

Y = A + A BC.

(A) DIRECT IMPLEMENTATION using NOT, NOR, and NAND



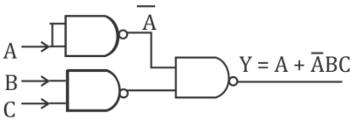

(B) SIMPLIFIED CIRCUIT

$Y = A + \bar{A}BC$

$= A+BC$

$= \overline{\overline{A+BC}}$ (by property of NOT)

$= \overline{\bar{A}(\overline{BC})}$ (by De Morgan's Law)




## 9.2 THE INVERTER (NOT GATE)

The inverter (NOT circuit) performs the operation called inversion or complementation. The inverter changes one logic level to the opposite level. In terms of bits, it changes a 1 to a 0 and a 0 to a 1.

Standard logic symbols for the inverter are shown in Fig, shows the distinctive shape symbols.

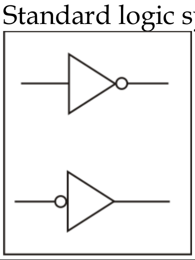
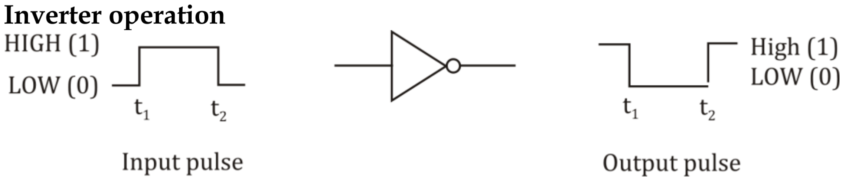


Fig- Logic symbol for the inverter.

When a HIGH level is applied to an inverter input, a LOW level will appear on its output. When a LOW level is applied to its input, a HIGH will appear on its output as shown in Fig. This operation is summarized in Table, which shows the output for each possible input in terms of levels and corresponding bits. A table such as this is called a truth table, which gives the output state for all possible input combinations.

| Input | OUTPUT |
|---|---|
| LOW (0) | HIGH (1) |
| HIGH (1) | LOW (0) |

**Inverter operation**



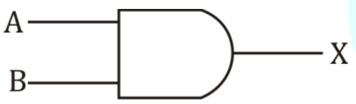

The operation of an inverter (NOT circuit) can be expressed as follows: If the input variable is called A and the output variable is called X, then

X = A

This expression states that the output is the complement of the input, so if A = 0, then X = 1, and if A = 1, then X = 0.

## 9.3 The AND Gate

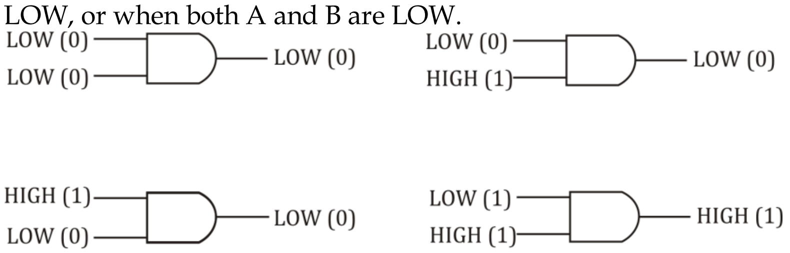
The term gate is used to describe a circuit that performs a basic logic operation. The AND gate is composed of two or more inputs and a single output, as indicated by the standard logic symbols shown in Fig. Inputs are on the left, and the output is on the right in each symbol. Gates with two inputs are shown, however, an AND gate can have any number of inputs greater than one.




**Operation of an AND Gate**

An AND gate produces a HIGH output only when all of the inputs are HIGH. When any of the inputs is LOW, the output is LOW. Therefore, the basic purpose of an AND gate is to determine when certain conditions are simultaneously true, as indicated by HIGH levels on all of its inputs, and to produce a HIGH on its output to indicate that all these conditions are true. The inputs of the 2-input AND gate in Fig. are labelled A and B, and the output is labelled X. The gate operation can be stated as follows:

For a 2-input AND gate, output X is HIGH only when inputs A and B are HIGH; X is LOW when either A or B is LOW, or when both A and B are LOW.




The logical operation of a gate can be expressed with a truth table that lists all input combinations with the corresponding outputs, as illustrated in Table for a 2-input AND gate. The truth table can be expanded to any number of inputs. For any AND gate, regardless of the number of inputs, the output is HIGH only when all inputs are HIGH.

| Input | | OUTPUT |
|---|---|---|
| A | B | X |
| 0 | 0 | 0 |
| 0 | 1 | 0 |
| 1 | 0 | 0 |
| 1 | 1 | 1 |
| 1 = HIGH, 0 =LOW | | |

The total number of possible combinations of binary inputs to a gate is determined by the following formula:

**$N=2^n$**

**Logic Expressions for an AND Gate**

The logical AND function of two variables is represented mathematically either by placing a dot between the two variables, as A . B, or by simply writing the adjacent letters without the dot, as AB. We will normally use the latter notation because it is easier to write. Boolean multiplication follows the same basic rules governing binary multiplication:

0 . 0 = 0

0 . 1 = 0

1 . 0 = 0

1 . 1 = 1

**Boolean multiplication is the same as the AND function.**

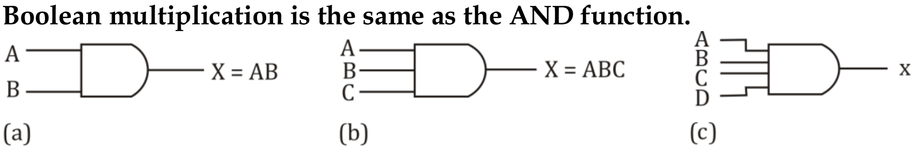



Fig. Boolean expressions for AND gates with two, three, and four inputs.

## 9.4 The OR Gate

An OR gate can have more than two inputs. The OR gate is another of the basic gates from which all logic functions are constructed. An OR gate can have two or more inputs and performs what is known as logical addition.

An OR gate has two or more inputs and one output, as indicated by the standard logic symbol in Fig., where OR gates with two inputs are illustrated. An OR gate can have any number of inputs greater than one.

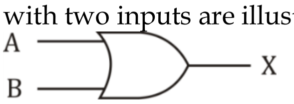



Fig. Standard logic symbol for the OR gate.

**Operation of an OR Gate**

- An OR gate produces a HIGH on the output when any of the inputs is HIGH. The output is LOW only when all of the inputs are LOW.
- The inputs of the 2-input OR gate in Fig. are labelled A and B. and the output is labelled X. The operation of the gate can be stated as follows:
- For a 2-input OR gate, output X is HIGH when either input A or input B is HIGH, or when both A and B are HIGH; X is LOW only when both A and B are LOW.
- The HIGH level is the active or asserted output level for the OR gate. Fig. illustrates the operation for a 2-input OR gate for all four possible input combinations.

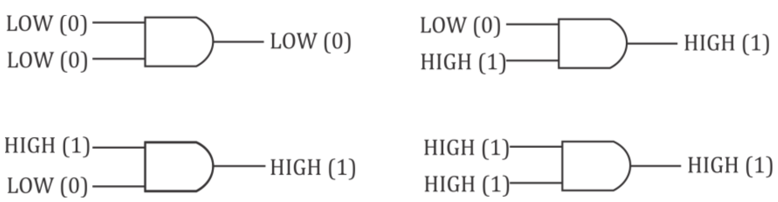



**Fig. All possible logic levels for a 2-input OR gate.**

**OR Gate Truth Table**

The operation of a 2-input OR gate is described in Table. This truth table can be expanded for any number of inputs; but regardless of the number of inputs. the output is HIGH when one or more of the inputs are HIGH.

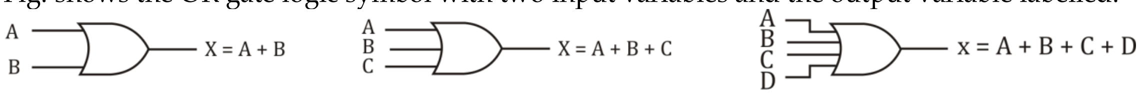
**Table- The truth table for a 2-input OR gate.**

| Input | | OUTPUT |
|---|---|---|
| 0 | 0 | 0 |
| 0 | 1 | 1 |
| 1 | 0 | 1 |
| 1 | 1 | 1 |
| 1 = HIGH, 0 =LOW | | |

**Logic Expressions for an OR Gate**
The logical OR function of two variables is represented mathematically by a + between the two variables, for example, A + B.
Addition in Boolean algebra involves variables whose values are either binary 1 or binary 0. The basic rules for Boolean addition are as follows:
$0 + 0 = 0$
$0 + 1 = 1$
$1 + 0 = 1$
$1 + 1 = 1$

**Boolean addition is the same as the OR function.**
Notice that Boolean addition differs from binary addition in the case where two 1 s are added. There is no carry in Boolean addition.

**The operation of a 2-input OR gate can be expressed as follows**: If one input variable is A, if the other input variable is B, and if the output variable is X, then the Boolean expression is
$X=A+B$
Fig. shows the OR gate logic symbol with two input variables and the output variable labelled.



(a) (b) (c)

## 9.5 The Nand Gate

The NAND gate is a popular logic element because it can be used as a universal gate: that is, NAND gates can be used in combination to perform the AND, OR, and inverter operations.

The term NAND is a contraction of NOT-AND and implies an AND function with a complemented (inverted) output. The standard logic symbol for a 2-input NAND gate and its equivalency to an AND gate followed by an inverter are shown in Fig., where the symbol ≡ means equivalent to.

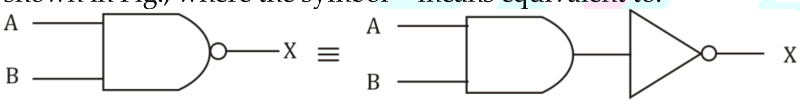


**Fig. Standard NAND gate logic symbols**

**Operation of a NAND Gate**
A NAND gate produces a LOW output only when all the inputs are HIGH. When any of the inputs is LOW, the output will be HIGH. For the specific case of a 2-input NAND gate, as shown in Fig with the inputs labelled A and B and the output labelled X, the operation can be stated as follows:
For a 2-input NAND gate, output X is LOW only when inputs A and B are HIGH; X is HIGH when either A or B is LOW, or when both A and B are LOW.
Note that this operation is opposite that of the AND in terms of the output level. In a NAND gate, the LOW level (0) is the active or asserted output level, as indicated by the bubble on the output. Fig illustrates the operation of a 2-input NAND
gate for all four input combinations, and Table, is the truth table summarizing the logical operation of the 2-input NAND gate.
**Note- The NAND is the same as the AND except the output is inverted.**

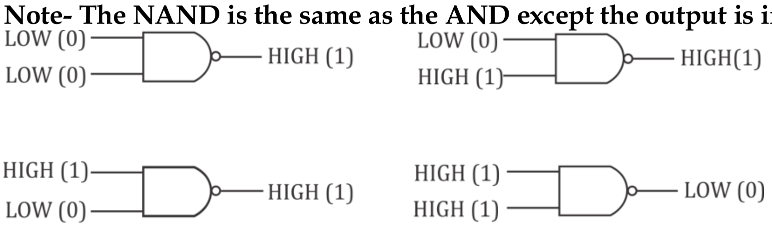


**Fig. Operation of a 2-input NAND gate.**

**Table- Truth table for a 2-input NAND gate.**

| Input | | OUTPUT |
|---|---|---|
| 0 | 0 | 1 |
| 0 | 1 | 1 |

| | | |
|---|---|---|
| 1 | 0 | 1 |
| 1 | 1 | 0 |
| 1 = HIGH, 0 =LOW | | |

**Logic Expressions for a NAND Gate**

The Boolean expression for the output of a 2-input NAND gate is

X = AB

This expression says that the two input variables, A and B, are first ANDed and then complemented, as indicated by the bar over the AND expression. This is a description in equation form of the operation of a NAND gate with two inputs. Evaluating this expression for all possible values of the two input variables, you get the results shown in Table.

| A | B | $\overline{AB} = X$ |
|---|---|---|
| 0 | 0 | $\overline{0.0} = \overline{0} = 1$ |
| 0 | 1 | $\overline{0.1} = \overline{0} = 1$ |
| 1 | 0 | $\overline{1.0} = \overline{0} = 1$ |
| 1 | 1 | $\overline{1.1} = \overline{1} = 0$ |

## 9.6 The NOR Gate

The NOR gate, like the NAND gate, is a useful logic element because it can also be used as a universal gate; that is, NOR gates can be used in combination to perform the AND, OR, and inverter operations.

The term NOR is a contraction of NOT-OR and implies an OR function with an inverted (complemented) output. The standard logic symbol for a 2-input NOR gate and its equivalent OR gate followed by an inverter are shown in Fig.

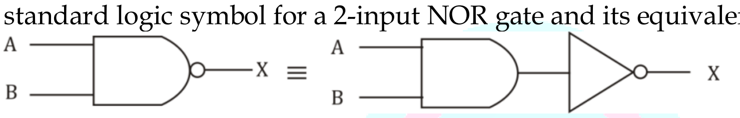



**Fig. Standard NOR gate logic symbols.**

**Operation of a NOR Gate**

A NOR gate produces a LOW output when any of its inputs is HIGH. Only when all of its inputs are LOW is the output HIGH. For the specific case of a 2-input NOR gate, as shown in Fig. with the inputs labelled A and B and the output labelled X, the operation can be stated as follows:

For a 2-input NOR gate, output X is LOW when either input A or input B is HIGH, or when both A and B are HIGH; X is HIGH only when both A and B are LOW.

This operation results in an output level opposite that of the OR gate. In a NOR gate, the LOW output is the active or asserted output level as indicated by the bubble on the output.

Fig. illustrates the operation of a 2-input NOR gate for all four possible input combinations, and Table, is the truth table for a 2-input NOR gate.

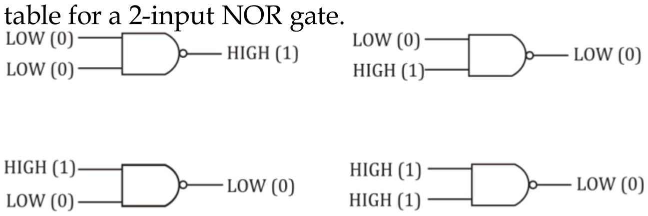



**Fig.** Operation of a 2-input NOR gate.

| **Input** | | **OUTPUT** |
|---|---|---|
| A | B | X |
| 0 | 0 | 1 |
| 0 | 1 | 0 |
| 1 | 0 | 0 |
| 1 | 1 | 0 |
| 1 = HIGH, 0 =LOW | | |

**Table- Truth table for a 2-input NOR gate.**

**Logic Expressions for a NOR Gate**

The Boolean expression for the output of a 2-input NOR gate can be written as

X = $\overline{A + B}$

This equation says that the two input variables are first ORed and then complemented, as indicated by the bar over the OR expression. Evaluating this expression, you get the results shown in Table. The NOR expression can be extended to more than two input variables by including additional letters to represent the other variables.

| A | B | $\overline{A+B}=X$ |
|---|---|---|
| 0 | 0 | $\overline{0+0}=\overline{0}=1$ |
| 0 | 1 | $\overline{0+1}=\overline{1}=0$ |
| 1 | 0 | $\overline{1+0}=\overline{1}=0$ |
| 1 | 1 | $\overline{1+1}=\overline{1}=0$ |

## 9.7 The Exclusive-Or and Exclusive-Nor Gates

Exclusive-OR and exclusive-NOR gates are formed by a combination of other gates already discussed. However, because of their fundamental importance in many applications, these gates are often treated as basic logic elements with their own unique symbols.

### 9.7 (a) The Exclusive-OR Gate

Standard symbol for an exclusive-OR (XOR for short) gate is shown in Fig. The XOR gate has only two inputs.

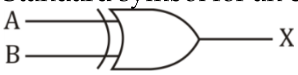



**Fig. Standard symbol for an exclusive-OR.**

**For an exclusive-OR gate, output X is HIGH when the two inputs are different.**

The four possible input combinations and the resulting outputs for an XOR gate are illustrated in Fig. The HIGH level is the active or asserted output level and occurs only when the inputs are at opposite levels. The operation of an XOR gate is summarized in the table shown in Table.

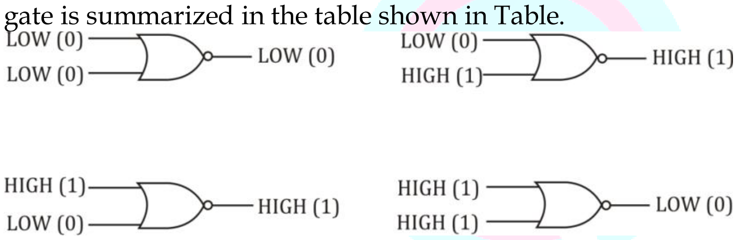



**Fig. All possible logic levels for an exclusive-OR gate.**

**Table- Truth table for an exclusive-OR gate.**

| Input | | OUTPUT |
|---|---|---|
| A | B | X |
| 0 | 0 | 0 |
| 0 | 1 | 1 |
| 1 | 0 | 1 |
| 1 | 1 | 0 |

### 9.7 (b) The Exclusive-NOR Gate

Standard symbols for an exclusive-NOR (XNOR) gate are shown in Fig. Like the XOR gate, an XNOR has only two inputs. The bubble on the output of the XNOR symbol indicates that its output is opposite that of the XOR gate. When the two input logic levels are opposite, the output of the exclusive-NOR gate is LOW. The operation can be stated as follows (A and B are inputs, X is the output):

**For an exclusive-NOR gate, X is HIGH only when A and B are both HIGH or both LOW.**

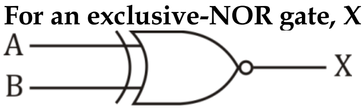



**Fig. Standard logic symbols for the exclusive-NOR gate.**

The four possible input combinations and the resulting outputs for an XNOR gate are shown in Fig. The operation of an XNOR gate is summarized in Table. Notice that the output is HIGH when the same level is on both inputs.

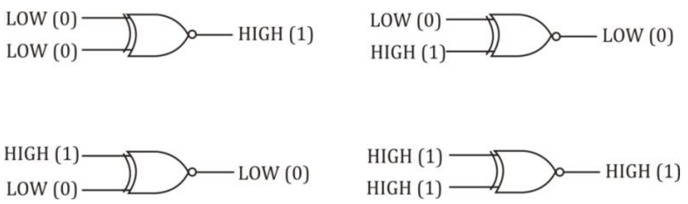


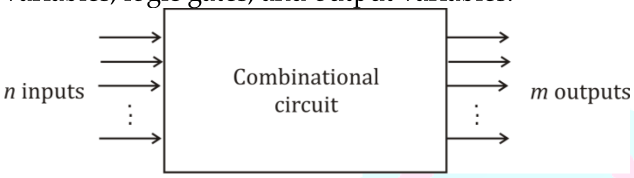
**Fig. All possible logic levels for an exclusive-NOR gate.**

**Table- Truth table for an exclusive-NOR gate.**

| Input | | OUTPUT |
|---|---|---|
| A | B | X |
| 0 | 0 | 1 |
| 0 | 1 | 0 |
| 1 | 0 | 0 |
| 1 | 1 | 1 |

## 10. Combinational Logic Design

Logic circuits for digital systems may be combinational or sequential. A combinational circuit consists of input variables, logic gates, and output variables.




### 10.1 Half Adder and Full Adder Circuit

An adder is a digital circuit that performs addition of numbers. The half adder adds two binary digits called as augend and addend and produces two outputs as sum and carry; XOR is applied to both inputs to produce sum and AND gate is applied to both inputs to produce carry. The full adder adds 3 one bit numbers, where two can be referred to as operands and one can be referred to as bit carried in. And produces 2-bit output, and these can be referred to as output carry and sum.

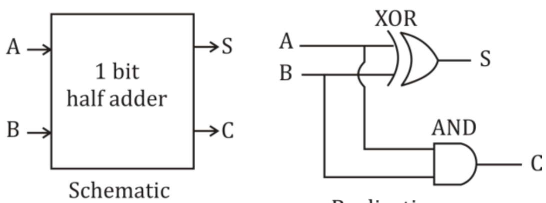
**Half adder Circuit**

To understand what is a half adder you need to know what is an adder first. Adder circuit is a combinational digital circuit that is used for adding two numbers. A typical adder circuit produces a sum bit (denoted by S) and a carry bit (denoted by C) as the output. Typically, adders are realized for adding binary numbers but they can be also realized for adding other formats like BCD (binary coded decimal, XS-3 etc. Besides addition, adder circuits can be used for a lot of other applications in digital electronics like address decoding, table index calculation etc.

Half adder is a combinational arithmetic circuit that adds two numbers and produces a sum bit (S) and carry bit (C) as the output. If A and B are the input bits, then sum bit (S) is the X-OR of A and B and the carry bit (C) will be the AND of A and B. From this it is clear that, a half adder circuit can be easily constructed using one X-OR gate and one AND gate. Half adder is the simplest of all adder circuit, but it has a major disadvantage. The half adder can add only two input bits (A and B) and has nothing to do with the carry if there is any in the input. So, if the input to a half adder have a carry, then it will be neglected it and adds only the A and B bits. That means the binary addition process is not complete and that's why it is called a half adder. The truth table, schematic representation and XOR//AND realization of a half adder are shown in the figure below.

| Inputs | | Outputs | |
|---|---|---|---|
| A | B | S | C |
| 0 | 0 | 0 | 0 |
| 1 | 0 | 1 | 0 |
| 0 | 1 | 1 | 0 |
| 1 | 1 | 0 | 1 |

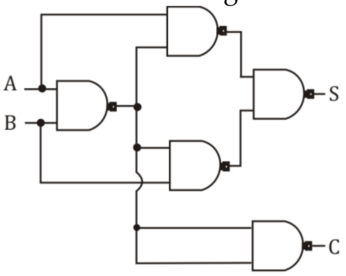
Truth table



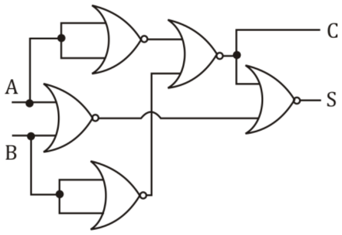
Schematic Realization

**Truth table, schematic and realization of half adder**

NAND gates or NOR gates can be used for realizing the half adder in universal logic and the relevant circuit diagrams are shown in the figure below.



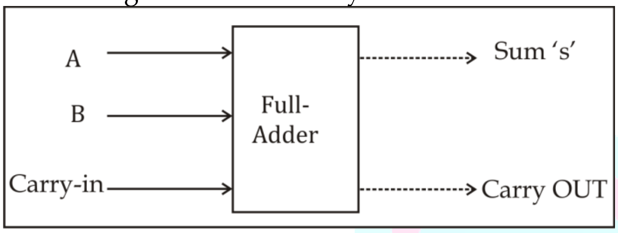
Half adder using NAND logic



Half adder using NOR logic

**Half adder using NAND & NOR logic**

**Full Adder**
This adder is difficult to implement than a half-adder. The difference between a half-adder and a full-adder is that the full-adder has three inputs and two outputs, whereas half adder has only two inputs and two outputs. The first two inputs are A and B and the third input is an input carry as C-IN. When a full-adder logic is designed, you string eight of them together to create a byte-wide adder and cascade the carry bit from one adder to the next.

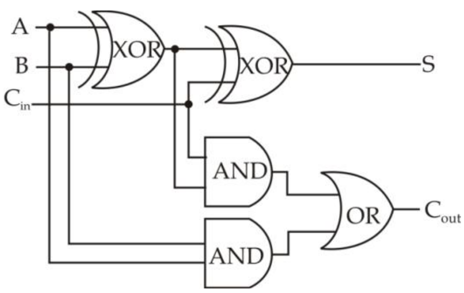



The output carry is designated as C-OUT and the normal output is designated as S.
Full Adder Truth Table:

| INPUTS | | | OUTPUT | |
|---|---|---|---|---|
| A | B | C-IN | C-OUT | S |
| 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 1 | 0 | 1 |
| 0 | 1 | 0 | 0 | 1 |
| 0 | 1 | 1 | 1 | 0 |
| 1 | 0 | 0 | 0 | 1 |
| 1 | 0 | 1 | 1 | 0 |
| 1 | 1 | 0 | 1 | 0 |
| 1 | 1 | 1 | 1 | 1 |

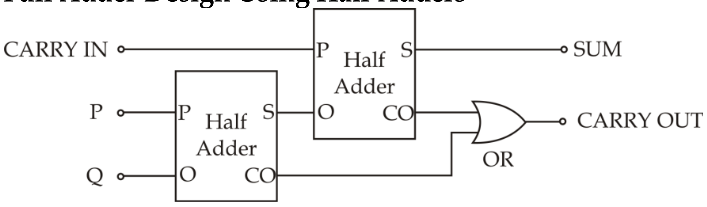
With the truth-table, the full adder logic can be implemented. You can see that the output S is an XOR between the input A and the half-adder, SUM output with B and C-IN inputs. We take C-OUT will only be true if any of the two inputs out of the three are HIGH.
So, we can implement a full adder circuit with the help of two half adder circuits. At first, half adder will be used to add A and B to produce a partial Sum and a second half adder logic can be used to add C-IN to the Sum produced by the first half adder to get the final S output.

**Full Adder Logic Circuit**



If any of the half adder logic produces a carry, there will be an output carry. So, COUT will be an OR function of the half-adder Carry outputs.

**Full Adder Design Using Half Adders**




With this type of symbol, we can add two bits together, taking a carry from the next lower order of magnitude, and sending a carry to the next higher order of magnitude. In a computer, for a multi-bit operation, each bit must be represented by a full adder and must be added simultaneously. Thus, to add two 8-bit numbers, you will need 8 full adders which can be formed by cascading two of the 4-bit blocks.

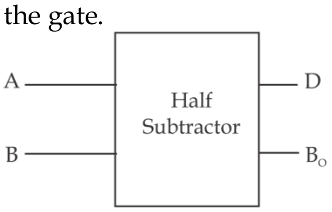
## 10.2 Binary Subtraction Circuits

Subtraction is a mathematical operation in which one integer number is deducted from another to obtain the equivalent quantity. The number from which other number is to be deducted is called as minuend and the number subtracted from the minuend is called subtrahend.

**Half Subtractors**

A half subtractor is a multiple output combinational logic network that does the subtraction of two bits of binary data. It has input variables and two output variables. Two inputs are corresponding to two input bits and two output variables corresponds to the difference bit and borrow bit.

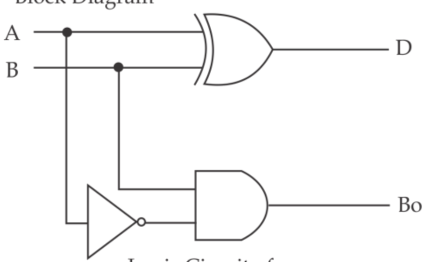
The binary subtraction is also performed by the Ex-OR gate with additional circuitry to perform the borrow operation. Thus, a half subtractor is designed by an Ex-OR gate including AND gate with A input complemented before fed to the gate.




Block Diagram

| A | B | D | $B_0$ |
|---|---|---|---|
| 0 | 0 | 0 | 0 |
| 0 | 1 | 1 | 1 |
| 1 | 0 | 1 | 0 |
| 1 | 1 | 0 | 0 |

Truth Table

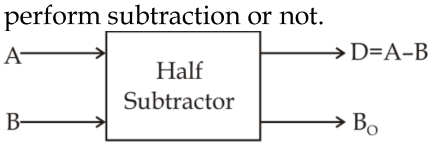



Logic Circuit of Half Subtractor

**Full Subtractor**

A combinational logic circuit that performs a subtraction between the two binary bits by considering borrow of the lower significant stage is called as the full subtractor. In this, subtraction of the two digits is performed by taking into consideration whether a 1 has already borrowed by the previous adjacent lower minuend bit or not.

Half-subtractor is a combinational circuit capable of subtracting a binary number from another binary number. It produces two outputs: difference and borrow. The borrow output specifies whether a binary number 1 is borrowed to perform subtraction or not.

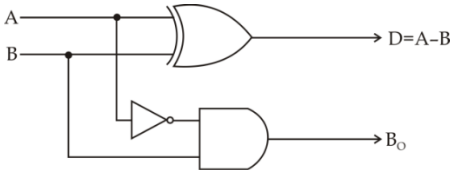



Graphic symbol of half-subtractor

The inputs are A and B. The output D is the result of subtraction of B from A and output $B_0$ is the borrow output.

| A | B | D | $B_0$ |
|---|---|---|---|
| 0 | 0 | 0 | 0 |
| 0 | 1 | 1 | 1 |
| 1 | 0 | 1 | 0 |
| 1 | 1 | 0 | 0 |

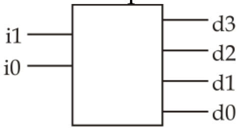
**Table 1: Truth Table of half-subtractor**




## 10.3 Decoders and Multiplexers

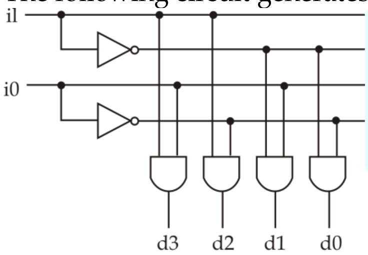
**Decoders**

A decoder is a circuit which has n inputs and $2^n$ outputs, and outputs 1 on the wire corresponding to the binary number represented by the inputs. For example, a 2-4 decoder might be drawn like this:

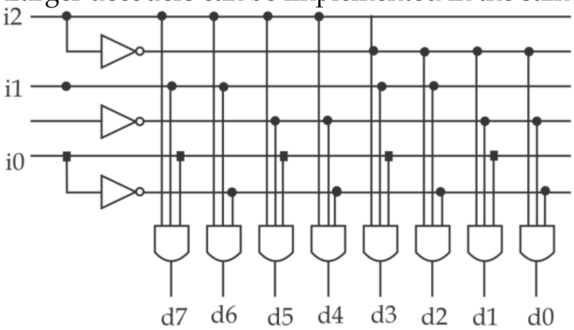



**The Decoder Circuit**

The following circuit generates all four minterms from two inputs, and implements the 2-4 decoder.

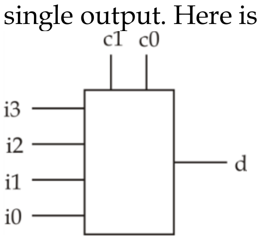



Larger decoders can be implemented in the same way. Here is a 3-8 decoder.




**Multiplexers**

Multiplexing is the property of combining one or more signals and transmitting on a single channel. This is achieved by the device multiplexer. A multiplexer is the most frequently used combinational circuits and important building block in many in digital systems. A multiplexer is a device which allows one of a number of inputs to be routed to a single output. Here is a 4-1 multiplexer.




Multiplexers are useful in many situations. For example, in a CPU, data being written to memory might come from one of a number of sources - from a register, from the result of a calculation, etc - so a multiplexer would be used to select data from the appropriate source.

## 10.4 Parity Generator and Checker

A parity generator is a combinational logic circuit that generates the parity bit in the transmitter. On the other hand, a circuit that checks the parity in the receiver is called parity checker. A combined circuit or devices of parity generators and parity checkers are commonly used in digital systems to detect the single bit errors in the transmitted data word.
The sum of the data bits and parity bits can be even or odd. In even parity, the added parity bit will make the total number of 1s an even amount whereas in odd parity the added parity bit will make the total number of 1s odd amount.

**Parity Generator**
It is combinational circuit that accepts an n-1 bit stream data and generates the additional bit that is to be transmitted with the bit stream. This additional or extra bit is termed as a parity bit.

In **even parity** bit scheme, the parity bit is '**0**' if there are **even number of 1s** in the data stream and the parity bit is '**1**' if there are **odd number of 1s** in the data stream.

In **odd parity** bit scheme, the parity bit is '**1**' if there are **even number of 1**s in the data stream and the parity bit is '**0**' if there are **odd number of 1s** in the data stream. Let us discuss both even and odd parity generators.

**Even Parity Generator**
Let us assume that a 3-bit message is to be transmitted with an even parity bit. Let the three inputs A, B and C are applied to the circuits and output bit is the parity bit P. The total number of 1s must be even, to generate the even parity bit P.

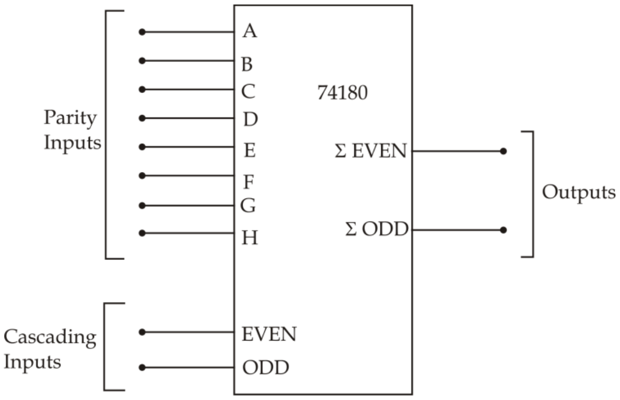
**Odd Parity Generator**
Let us consider that the 3-bit data is to be transmitted with an odd parity bit. The three inputs are A, B and C and P is the output parity bit. The total number of bits must be odd in order to generate the odd parity bit.

**Parity Check**
It is a logic circuit that checks for possible errors in the transmission. This circuit can be an even parity checker or odd parity checker depending on the type of parity generated at the transmission end. When this circuit is used as even parity checker, the number of input bits must always be even.
When a parity error occurs, the 'sum even' output goes low and 'sum odd' output goes high. If this logic circuit is used as an odd parity checker, the number of input bits should be odd, but if an error occurs the 'sum odd' output goes low and 'sum even' output goes high.

**Even Parity Checker**
Consider that three input message along with even parity bit is generated at the transmitting end. These 4 bits are applied as input to the parity checker circuit which checks the possibility of error on the data. Since the data is transmitted with even parity, four bits received at circuit must have an even number of 1s.
If any error occurs, the received message consists of odd number of 1s. The output of the parity checker is denoted by PEC (parity error check).

**Odd Parity Checker**
Consider that a three-bit message along with odd parity bit is transmitted at the transmitting end. Odd parity checker circuit receives these 4 bits and checks whether any error are present in the data.
If the total number of 1s in the data is odd, then it indicates no error, whereas if the total number of 1s is even then it indicates the error since the data is transmitted with odd parity at transmitting end.

**Parity Generator/Checker ICs**
There are different types of parity generator /checker ICs are available with different input configurations such as 5-bit, 4-bit, 9-bit, 12-bit, etc. A most commonly used and standard type of parity generator/checker IC is 74180.
It is a 9-bit parity generator or checker used to detect errors in high speed data transmission or data retrieval systems.
This IC can be used to generate a 9-bit odd or even parity code or it can be used to check for odd or even parity in a 9-bit code (8 data bits and one parity bit).




This IC consists of eight parity inputs from A through H and two cascading inputs. There are two outputs even sum and odd sum. In implementing generator or checker circuits, unused parity bits must be tied to logic zero and the cascading inputs must not be equal.

## 11. Flip Flop

Flip flops are actually an application of logic gates. With the help of Boolean logic, you can create memory with them. Flip flops can also be considered as the most basic idea of a Random-Access Memory [RAM]. When a certain input value is given to them, they will be remembered and executed, if the logic gates are designed correctly. A higher application of flip flops is helpful in designing better electronic circuits.

The most commonly used application of flip flops is in the implementation of a feedback circuit. As a memory relies on the feedback concept, flip flops can be used to design it.

A digital computer needs devices which can store information. A flip flop is a binary storage device. It can store binary bit either 0 or 1. It has two stable states HIGH and LOW i.e. 1 and 0. It has the property to remain in one state indefinitely until it is directed by an input signal to switch over to the other state. It is also called bistable multivibrator.

There are mainly four types of flip flops that are used in electronic circuits. They are-

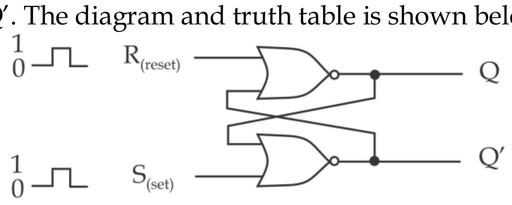
1. The basic Flip Flop or S-R Flip Flop
2. Delay Flip Flop [D Flip Flop]
3. J-K Flip Flop
4. T Flip Flop

### 11.1 S-R Flip Flop

The SET-RESET flip flop is designed with the help of two NOR gates and also two NAND gates. These flip flops are also called S-R Latch.

**S-R Flip Flop using NOR Gate**

The design of such a flip flop includes two inputs, called the SET [S] and RESET [R]. There are also two outputs, Q and Q′. The diagram and truth table is shown below.




(a) Logic diagram

| S R | Q Q′ | |
|---|---|---|
| 1 0 | 1 0 | |
| 0 0 | 1 0 | (after S=1, R = 0) |
| 0 1 | 0 1 | |
| 0 0 | 0 0 | (after S=0, R = 1) |
| 1 1 | 0 0 | |

(b) Truth table

Basic flip-flop circuit with NOR gates
S-R Flip Flop using NOR Gate

From the diagram it is evident that the flip flop has mainly four states. They are
S=1, R=0 – Q=1, Q′=0

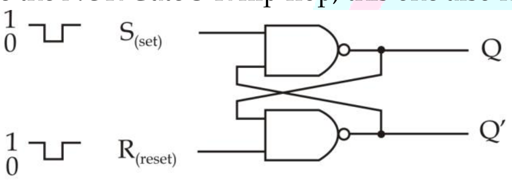
This state is also called the **SET state.**
S=0, R=1 – Q=0, Q′=1

This state is known as the **RESET state.**
In both the states you can see that the outputs are just compliments of each other and that the value of Q follows the value of S.
S=0, R=0 – Q & Q′ = Remember

If both the values of S and R are switched to 0, then the circuit remembers the value of S and R in their previous state.
S=1, R=1 – Q=0, Q′=0 [Invalid]

This is an invalid state because the values of both Q and Q′ are 0. They are supposed to be compliments of each other. Normally, this state must be avoided.

**S-R Flip Flop using NAND Gate**
The circuit of the S-R flip flop using NAND Gate and its truth table is shown below.
Like the NOR Gate S-R flip flop, this one also has four states. They are




(a) Logic diagram

| S R | Q Q′ | |
|---|---|---|
| 1 0 | 0 1 | |
| 1 1 | 0 1 | (after S=1, R=0) |
| 0 1 | 1 0 | |
| 1 1 | 1 0 | (after S=0, R=1) |
| 0 0 | 1 1 | |

(b) Truth table
Basic flip-flop circuit with NAND gates
S-R Flip Flop using NAD gate
S=1, R=0 – Q=0, Q′=1
This state is also called the SET state.
S=0, R=1 – Q=1, Q′=0
This state is known as the RESET state.

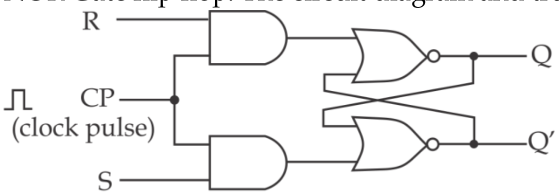
In both the states you can see that the outputs are just compliments of each other and that the value of Q follows the compliment value of S.
S=0, R=0 – Q=1, & Q′ =1 [Invalid]
If both the values of S and R are switched to 0 it is an invalid state because the values of both Q and Q′ are 1. They are supposed to be compliments of each other. Normally, this state must be avoided.
S=1, R=1 – Q & Q′= Remember

If both the values of S and R are switched to 1, then the circuit remembers the value of S and R in their previous state.

## 11.2 Clocked S-R Flip Flop

It is also called a Gated S-R flip flop.
The problems with S-R flip flops using NOR and NAND gate is the invalid state. This problem can be overcome by using a bi stable SR flip-flop that can change outputs when certain invalid states are met, regardless of the condition of either the Set or the Reset inputs. For this, a clocked S-R flip flop is designed by adding two AND gates to a basic NOR Gate flip flop. The circuit diagram and truth table is shown below.




(a) Logic diagram

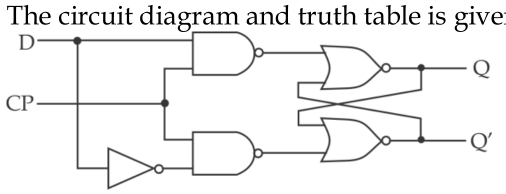
| Q S R | Q(t+1) |
|---|---|
| 0 0 0 | 0 |
| 0 0 1 | 0 |
| 0 1 0 | 1 |
| 0 1 1 | indeterminate |
| 1 0 0 | 1 |
| 1 0 1 | 0 |
| 1 1 0 | 1 |
| 1 1 1 | indeterminate |

(b) Truth table

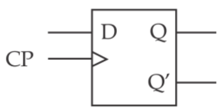
**Cocked S-R Flip Flop**
A clock pulse [CP] is given to the inputs of the AND Gate. When the value of the clock pulse is ‘0’, the outputs of both the AND Gates remain ‘0’. As soon as a pulse is given the value of CP turns ‘1’. This makes the values at S and R to pass through the NOR Gate flip flop. But when the values of both S and R values turn ‘1’, the HIGH value of CP causes both of them to turn to ‘0’ for a short moment. As soon as the pulse is removed, the flip flop state becomes intermediate. Thus either of the two states may be caused, and it depends on whether the set or reset input of the flip-flop remains a ‘1’ longer than the transition to ‘0’ at the end of the pulse. Thus the invalid states can be eliminated.

## 11.3 D Flip Flop

The circuit diagram and truth table is given below.




(a) Logic diagram with NAND GATES




(b) Graphical symbol

| Q D | Q(t+1) |
|---|---|
| 0 0 | 0 |
| 0 1 | 1 |
| 1 0 | 0 |
| 1 1 | 1 |

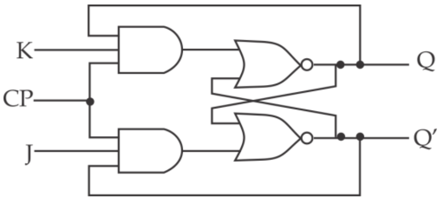
( c ) Transition table

Clocked D flip-flop

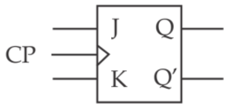
D flip flop is actually a slight modification of the above explained clocked SR flip-flop. From the figure you can see that the D input is connected to the S input and the complement of the D input is connected to the R input. The D input is passed on to the flip flop when the value of CP is '1'. When CP is HIGH, the flip flop moves to the SET state. If it is '0', the flip flop switches to the CLEAR state.

### 11.4 J-K Flip Flop

The circuit diagram and truth-table of a J-K flip flop is shown below.




(a) Logic diagram




(b) Graphical symbol

| Q J K | Q(t+1) |
|---|---|
| 0 0 0 | 0 |
| 0 0 1 | 0 |
| 0 1 0 | 1 |
| 0 1 1 | 1 |
| 1 0 0 | 1 |
| 1 0 1 | 0 |
| 1 1 0 | 1 |
| 1 1 1 | 0 |

(c) Transition table

Clocked JK flip-flop

A J-K flip flop can also be defined as a modification of the S-R flip flop. The only difference is that the intermediate state is more refined and precise than that of a S-R flip flop.

The behaviour of inputs J and K is same as the S and R inputs of the S-R flip flop. The letter J stands for SET and the letter K stands for CLEAR.

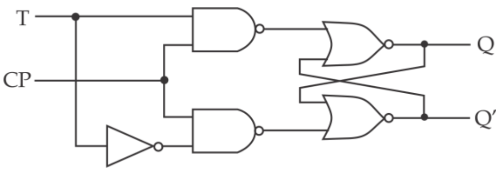
When both the inputs J and K have a HIGH state, the flip-flop switch to the complement state. So, for a value of Q = 1, it switches to Q=0 and for a value of Q = 0, it switches to Q=1.

The circuit includes two 3-input AND gates. The output Q of the flip flop is returned back as a feedback to the input of the AND along with other inputs like K and clock pulse [CP]. So, if the value of CP is '1', the flip flop gets a CLEAR signal and with the condition that the value of Q was earlier 1. Similarly output Q′ of the flip flop is given as a feedback to the input of the AND along with other inputs like J and clock pulse [CP]. So, the output becomes SET when the value of CP is 1 only if the value of Q′ was earlier 1.

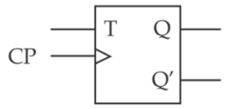
The output may be repeated in transitions once they have been complimented for J=K=1 because of the feedback connection in the JK flip-flop. This can be avoided by setting a time duration lesser than the propagation delay through the flip-flop. The restriction on the pulse width can be eliminated with a master-slave or edge-triggered construction.

### 11.5 T Flip Flop

This is a much simpler version of the J-K flip flop. Both the J and K inputs are connected together and thus are also called a single input J-K flip flop. When clock pulse is given to the flip flop, the output begins to toggle. Here also the restriction on the pulse width can be eliminated with a master-slave or edge-triggered construction. Take a look at the circuit and truth table below.




(a) Logic diagram with NAND GATES




(b) Graphical symbol

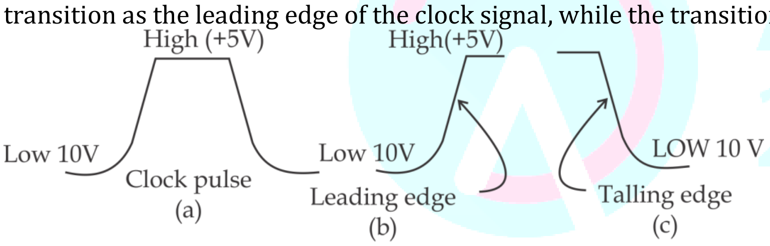
| Q T | Q(t+1) |
|---|---|
| 0 0 | 0 |
| 0 1 | 1 |
| 1 0 | 0 |
| 1 1 | 0 |

( c ) Transition table

Clocked T flip-flop

## 11.6 Edge-triggered Flip-Flop

The pulse goes from a low level 0 volt, the positive logical 0 condition, to a high level (+5 volts) , the positive logic logical 1 condition going between the two logic levels at a fixed frequency rate. A clock signal as seen in Figure, has two transitions, one from low to high level the other from high to low level. For positive logic operation, we define the low to high transition as the leading edge of the clock signal, while the transition from high to low is called the clock trailing edge.



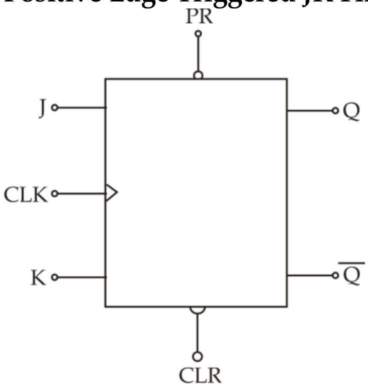

Clock Waveform (a) Full clock pulse (b) Leading edge (c) Trailing edge

Some flip flop circuits are triggered by the clock leading edge while other units are triggered on the clock trailing edge. The particular flip flop specifications will provide this information as we shall see. Some flip flop are other logic units are triggered when the clock reaches prescribed voltage levels or goes from one voltage level to another usually without regard to voltage rise or fall time. A circuit clocked by the leading edge, is referred to as being positive edge triggered while another circuit triggering on the trailing edge, is negative edge triggered.

**Positive Edge Triggered Flip Flop**

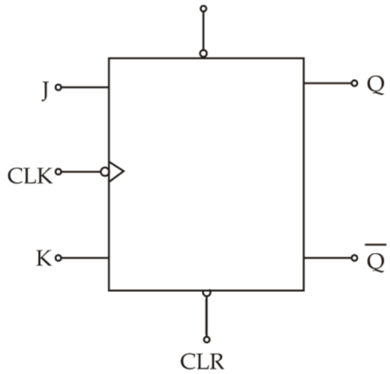
In positive edge triggered flip flops the clock samples the input line at the positive edge (rising edge or leading edge) of the clock pulse. The state of the output of the flip flop is set or reset depending upon the state of the input at positive edge of the clock. This state of the output remains for one clock cycle and the clock again samples the input line on the next positive edge of the clock. The arrow head at clock terminal indicates positive edge triggering. The arrow head symbol is termed as dynamic signal indicator.

**Positive Edge Triggered JK Flip Flop**



**Negative Edge Triggered Flipflop-**

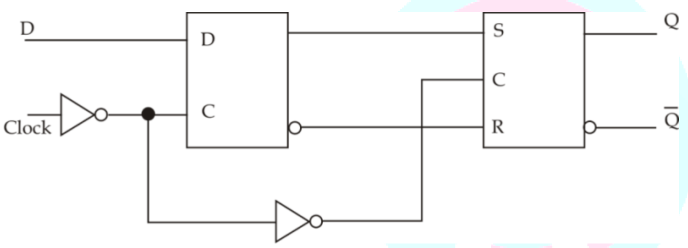
In negative edge triggered flip flops the clock samples the input lines at the negative edge (falling edge or trailing edge) of the clock pulse. The output of the flip flop is set or reset at the negative edge of the clock pulse. A symbolic representation of negative edge triggering has been shown in Figure. A small circle is put before the arrow head to indicate negative edge triggering.




Negative Edge Triggered Flip Flop

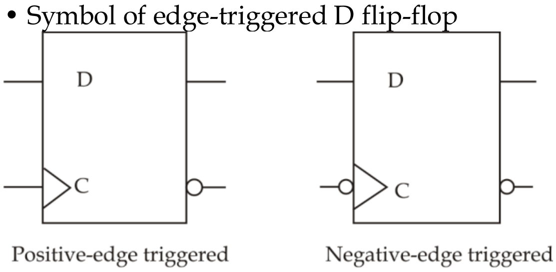
• **Example: Positive Edge-Triggered D Flip-Flop**

on the positive edge (while the clock is going from 0 to 1), the input D is read, and almost immediately propagated to the output Q. Only the value of D at the positive.




**Symbol**

• Symbol of edge-triggered D flip-flop




**Flip-Flop Timing**

• Set-up time: $t_s$

Input needs to be stable before trigger

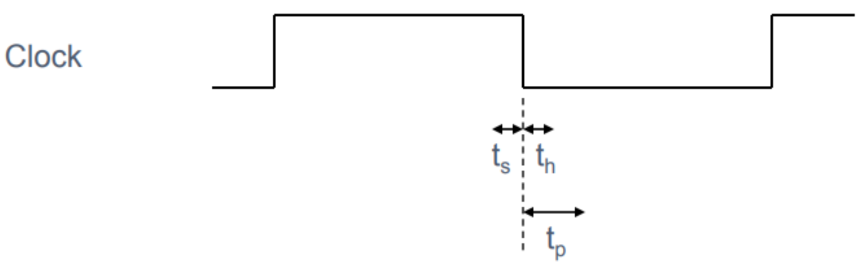
• Hold time: $t_h$

Input needs to be stable after trigger

• Propagation delay: $t_p$

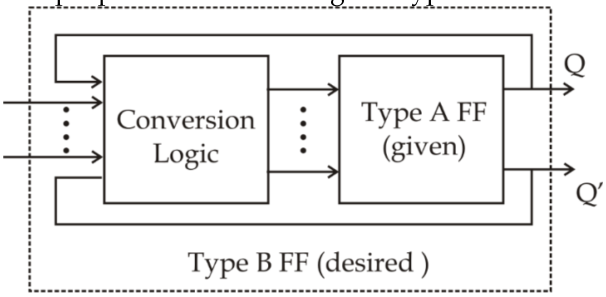
Some delay from trigger to output change

• Example: Negative edge triggered flip-flip



## 11.7 Flip-flop Conversions

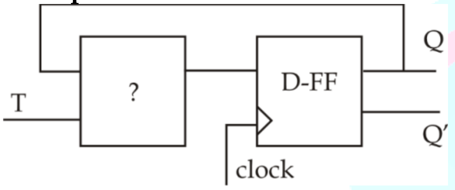
The purpose is to convert a given type A FF to a desired type B FF using some conversion logic.




The key here is to use the excitation table, which shows the necessary triggering signal (S,R, J,K, D and T) for a desired flip-flop state transition :

| $Q_t$ | $Q_{t+1}$ | S | R | J | L | D | T |
|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | X | 0 | X | 0 | 0 |
| 0 | 1 | 1 | 0 | 1 | X | 1 | 1 |
| 1 | 0 | 0 | 1 | X | 1 | 0 | 1 |
| 1 | 1 | X | 0 | X | 0 | 1 | 0 |

**Example: Convert a D-FF to a T-FF:**

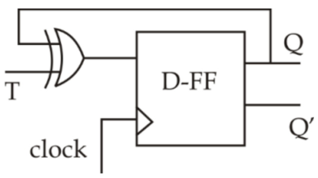



We need to design the circuit to generate the triggering signal D as a function of T and Q:
D = f(T, Q)

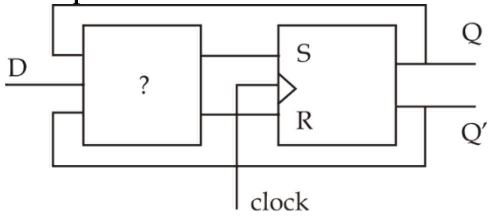
Consider the excitation table:

| $Q_t$ | $Q_{t+1}$ | T | D |
|---|---|---|---|
| 0 | 0 | 0 | 0 |
| 0 | 1 | 1 | 1 |
| 1 | 0 | 1 | 0 |
| 1 | 1 | 0 | 1 |

Treating D as a function of T and current FF state Q ($Q_T$), we have
$D = T'Q + TQ' = T \oplus Q$



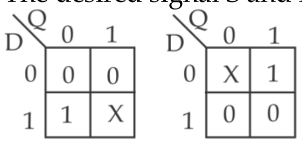

**Example 2: Convert a RS-FF to a D-FF:**




We need to design the circuit to generate the triggering signals S and R as functions of D and Q Consider the excitation table:

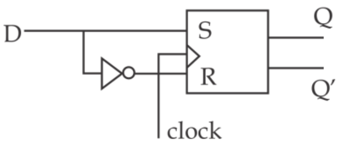
| $Q_t$ | $Q_{t+1}$ | D | S | R |
|---|---|---|---|---|
| 0 | 0 | 0 | 0 | X |
| 0 | 1 | 1 | 1 | 0 |
| 1 | 0 | 0 | 0 | 1 |
| 1 | 1 | 1 | X | 0 |

The desired signal S and R can be obtained as functions of T and current FF state Q from the Karnaugh maps:

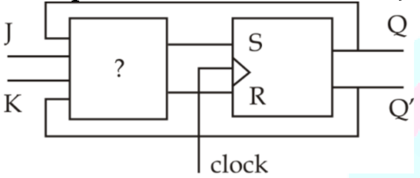


S=D R=D′

S = D, R =D′



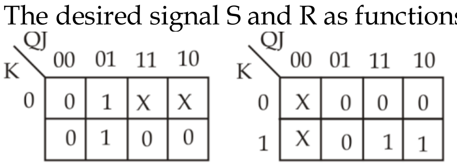

**Example 3: Convert a RS-FF to a JK-FF**




We need to design the circuit to generate the triggering signals S and R as functions of J, K and Q. Consider the excitation table:

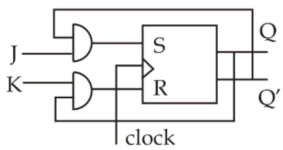
| $Q_t$ | $Q_{t+1}$ | J | K | S | R |
|---|---|---|---|---|---|
| 0 | 0 | 0 | X | 0 | X |
| 0 | 1 | 1 | X | 1 | 0 |
| 1 | 0 | X | 1 | 0 | 1 |
| 1 | 1 | X | 0 | X | 0 |

The desired signal S and R as functions of J, K and current FF state Q can be obtained from the Karnaugh maps:



S=Q′J R=QK

S = Q′J, R = QK



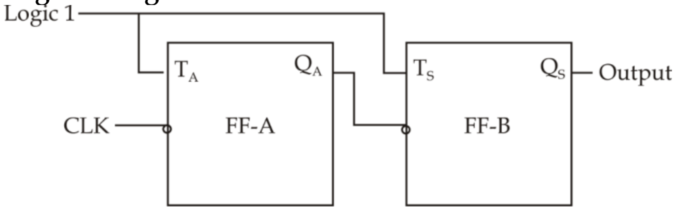

## 11.8 Counters

Counter is a sequential circuit. A digital circuit which is used for a counting pulses is known counter. Counter is the widest application of flip-flops. It is a group of flip-flops with a clock signal applied. Counters are of two types.

**Asynchronous or ripple counters**

The logic diagram of a 2-bit ripple up counter is shown in figure. The toggle (T) flip-flop are being used. But we can use the JK flip-flop also with J and K connected permanently to logic 1. External clock is applied to the clock input of flip-flop A and QA output is applied to the clock input of the next flip-flop i.e. FF-B.

**Logical Diagram-**




**Truth Table**

| Clock | Counter output | | State Number | Decimal Counter Output |
|---|---|---|---|---|
| | $Q_S$ | $Q_A$ | | |
| Initially | 0 | 0 | – | 0 |
| 1st | 0 | 1 | 1 | 1 |
| 2nd | 1 | 0 | 2 | 2 |
| 3rd | 1 | 1 | 3 | 3 |
| 4th | 0 | 0 | 4 | 0 |

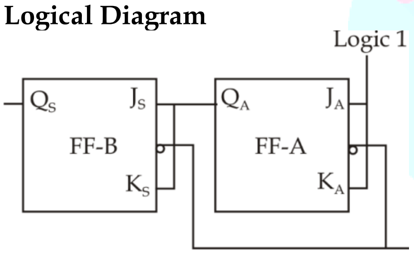
**Synchronous counters**

If the "clock" pulses are applied to all the flip-flops in a counter simultaneously, then such a counter is called as synchronous counter.

**2-bit Synchronous up counter**

The $J_A$ and $K_A$ inputs of FF-A are tied to logic 1. So FF-A will work as a toggle flip-flop. The $J_B$ and $K_B$ inputs are connected to $Q_A$.

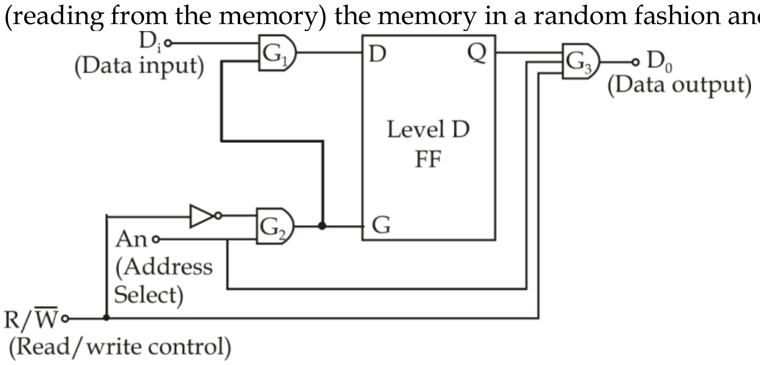
**Logical Diagram**




## 11.9 RAM-

In computers, digital control systems, information processing systems, etc. it is necessary to store digital data and retrieve the data as desired. For this purpose, earlier only magnetic memory devices were possible, whereas these days it has become possible to make memory devices using semiconductor devices. Semiconductor memories have become very popular because of their small size (available in ICs) and convenience to use.

FLIP-FLOPS can be used for making memories in which data can be stored for any desired length of time and then read out whenever required. In such a memory, data can be put into (writing into the Memory) or retrieved from (reading from the memory) the memory in a random fashion and is known as random-access memory.




l-bit read/write memory.

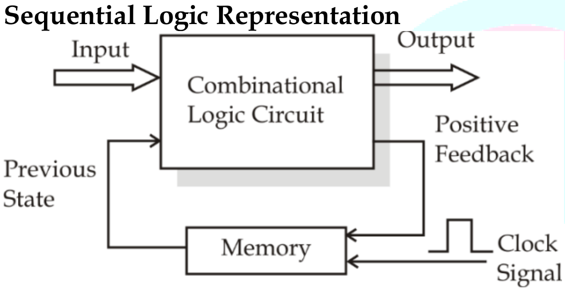
A 1-bit read/write memory is shown in Fig. which is the basic memory element and memory ICs are built around a system of basic 1-bit cell.

In this memory cell, a level D FLIP-FLOP is used which has Q output that follows the D input as long as G terminal is at logic 1. The moment the G input changes to logic 0, the Q output does not change, and it retains the D input level that existed just before the transition from 1 to 0 at input G. This input is used to select the memory cell. In the 1-bit cell shown there are three inputs–$D_1$ (data input), A (address select) and R/W (read/write control) and one output $D_0$ (data output). A= 0, all input and output activities are blocked, and the cell is in the hold mode where its stored output is protected.
The complete function of this cell can be Understood from the function table of Table. The read operation is nondestructive, that is, the stored bit can be read out any number of times without disturbing it. The stored bit will be protected as long as power is on. Therefore, this type-of memory is known **as volatile memory**.
As far as writing into the cell is concerned, it is not required to be cleared before entering the new bit. Whenever a new bit is entered the earlier one gets destroyed automatically.

## 12. Sequential Logic Design

Unlike Combinational Logic circuits that change state depending upon the actual signals being applied to their inputs at that time, ***Sequential Logic*** circuits have some form of inherent "Memory" built in. This means that sequential logic circuits are able to take into account their previous input state as well as those actually present, a sort of "before" and "after" effect is involved with sequential circuits. In other words, the output state of a "sequential logic circuit" is a function of the following three states, the "present input", the "past input" and/or the "past output". Sequential Logic circuits remember these conditions and stay fixed in their current state until the next clock signal changes one of the states, giving sequential logic circuits "Memory".

**Sequential Logic Representation**

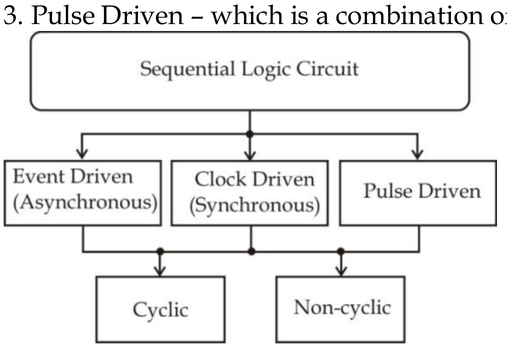



The word "Sequential" means that things happen in a "sequence", one after another and in Sequential Logic circuits, the actual clock signal determines when things will happen next. Simple sequential logic circuits can be constructed from standard Bistable circuits such as: Flip-flops, Latches and Counters and which themselves can be made by simply connecting together universal NAND Gates and/or NOR Gates in a particular combinational way to produce the required sequential circuit.

### 12.1 Classification of Sequential Logic

As standard logic gates are the building blocks of combinational circuits, bistable latches and flip-flops are the basic building blocks of sequential logic circuits. Sequential logic circuits can be constructed to produce either simple edge-triggered flip-flops or more complex sequential circuits such as storage registers, shift registers, memory devices or counters. Either way sequential logic circuits can be divided into the following three main categories:

1. Event Driven – asynchronous circuits that change state immediately when enabled.
2. Clock Driven – synchronous circuits that are synchronized to a specific clock signal.
3. Pulse Driven – which is a combination of the two that responds to triggering pulses.




**Latches:** Latches and flip flops are the basic elements and these are used to store information. One flip flop and latch can store one bit of data. The main difference between the latches and flip flops is that, a latch checks input continuously and changes the output whenever there is a change in input. But, flip flop is a combination of latch and clock that continuously checks input and changes the output time adjusted by the clock. In this article, we are going to look at the operations of the numerous latches and flip-flops.

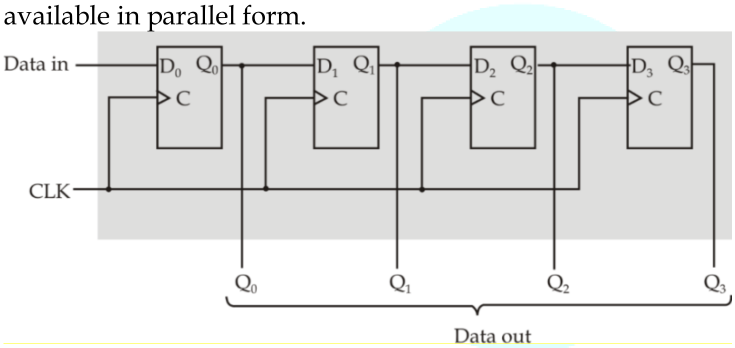
### 12.2 Shift Registers

The Shift Register is another type of sequential logic circuit that is used for the storage or transfer of data in the form of binary numbers and then "shifts" the data out once every clock cycle, hence the name shift register. It basically consists of several single bit "D-Type Data Latches", one for each bit (0 or 1) connected together in a serial or daisy-chain arrangement so that the output from one data latch becomes the input of the next latch and so on. The data bits may be fed in or out of the register serially, i.e. one after the other from either the left or the right direction, or in parallel, i.e. all together. The number of individual data latches required to make up a single Shift Register is determined by the number of bits to be stored with the most common being 8-bits wide, i.e. eight individual data latches.

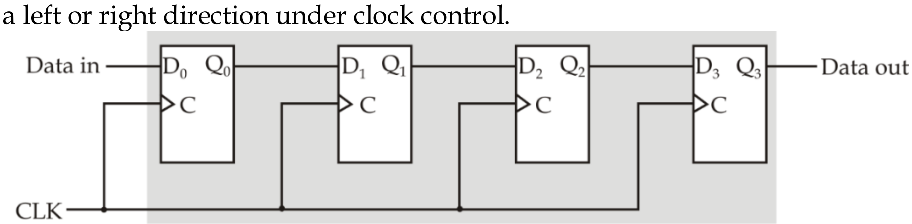
Shift Registers are used for data storage or data movement and are used in calculators or computers to store data such as two binary numbers before they are added together, or to convert the data from either a serial to parallel or parallel to serial format. The individual data latches that make up a single shift register are all driven by a common clock (Clk) signal making them synchronous devices. Shift register IC's are generally provided with a clear or reset connection so that they can be "SET" or "RESET" as required.

Generally, shift registers operate in one of four different modes with the basic movement of data through a shift register being:

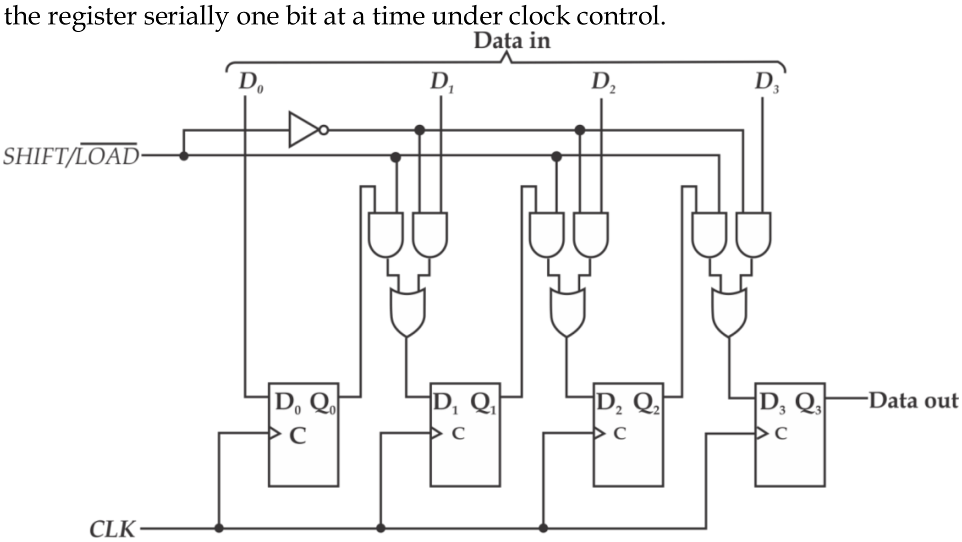
• **Serial-in to Parallel-out (SIPO)** - The register is loaded with serial data, one bit at a time, with the stored data being available in parallel form.




• **Serial-in to Serial-out (SISO)** - The data is shifted serially "IN" and "OUT" of the register, one bit at a time in either a left or right direction under clock control.




• **Parallel-in to Serial-out (PISO)** - The parallel data is loaded into the register simultaneously and is shifted out of the register serially one bit at a time under clock control.



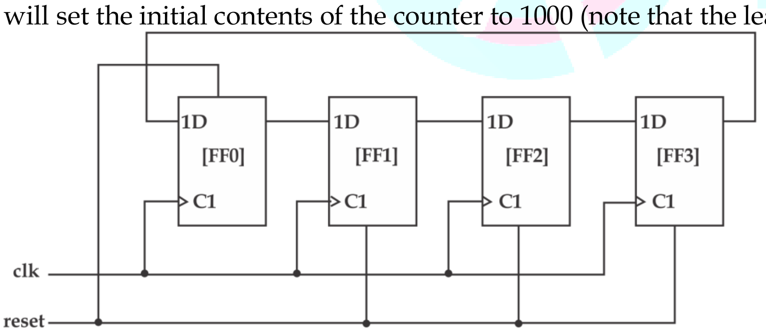
• **Parallel-in to Parallel-out (PIPO)** - The parallel data is loaded simultaneously into the register, and transferred together to their respective outputs by the same clock pulse.

**4-Bit Bidirectional Universal Shift Register**
This bidirectional shift register is designed to incorporate virtually all of the features a system designer may want in a shift register; they feature parallel inputs, parallel outputs, right-shift and left-shift serial inputs, operating-mode-control inputs, and a direct overriding clear line. The register has four distinct modes of operation, namely:
A) Parallel
B) load Shift right
C) Shift left
D) Inhibit clock

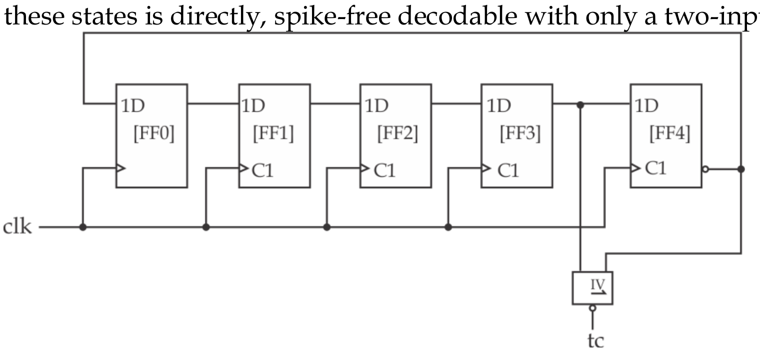
## 13. Ring Counter

Ring counters are a type of counter created using shift registers. A shift register is constructed using D-type flip-flops where the output of one flip-flop is connected to the input of another flip-flop. With ring counters, the output of the last flip-flop is fed to the input of the first flip-flop. Ring counters do not count using normal binary code, but their internal state can be used to decode to any output sequence wanted. There are two types of ring counters:
**a) Straight ring counter**
**b) Johnson counter**
There is a third counter type using a shift register, the Linear Feedback Shift Register (LFSR). It has a more elaborate feedback circuit than the other two ring counters.

**a. Straight Ring Counter**
A straight ring counter or Overbeck counter connects the output of the last flip-flop to the first flip-flop input and circulates a single one bit around the ring. It provides a one-hot counting sequence. For example, in a 4-register ring counter, with initial register values of 1000, the repeating sequence is 1000, 0100, 0010, 0001. Note that one of the flip-flops must be pre-loaded with a logic 1 in order for it to operate properly. Also note that an n-bit ring counter cycles through exactly n states. A schematic of a 4-bit straight ring counter is given below. The asynchronous reset will set the initial contents of the counter to 1000 (note that the least significant flip-flop is on the left side).




**b. Johnson Counter**
Another form of ring counter is created by feeding back the complement of the contents of the last flip-flop to the input of the first flip-flop. This is called a twisted ring counter, but is better known as the Johnson counter. The alternative term Möbius counter is found in many books and articles because the Johnson counter resembles the famous Möbius strip. For example, in a 5-flip-flop Johnson counter with an initial register contents (or state) of 00000, the repeating sequence is 00000, 10000, 11000, 11100, 11110, 11111, 01111, 00111, 00011, 00001. When observing the pattern, it can be seen that any changes between succeeding states, only one flip-flop changes state. As a result, any of these states is directly, spike-free decodable with only a two-input gate [5, 6].



## PRACTICE SET

1. In computers, subtraction is generally carried out by?
(a) 9's complement (b) 10's complement
(c) 1's complement (d) 2's complement
(e) None of these
2. The number of bus controllers that are used for interfacing of memory and I/O devices is?
(a) 1 (b) 2
(c) 3 (d) None of the mentioned
(e) 5
3. If MBYTES input is high, then the pin serves as?
(a) AEN (b) CEN
(c) AEN and CEN (d) None of the mentioned
(e) GEN
4. What characteristic of RAM memory makes it not suitable for permanent storage?
(a) too slow (b) unreliable
(c) it is volatile (d) too bulky
(e) None of these
5. The average time required to reach a storage location in memory and obtain its contents is called the?
(a) seek time (b) turnaround time
(c) access time (d) transfer time
(e) None of these
6. Which of the following is not a weighted code?
(a) Decimal Number system (b) Excess 3-cod
(c) Binary number System
(d) Machine number system (e) None of these
7. In the application where all the interrupting devices are of equal priority, the mode used is?
(a) automatic rotation (b) automatic EOI mode
(c) specific rotation (d) EOI
(e) None of these
8. In cascaded mode, the number of vectored interrupts provided by 8259A is?
(a) 4 (b) 8
(c) 16 (d) 64
(e) None of these
9. The circuit used to store one bit of data is known as?
(a) Register (b) Encoder
(c) Decoder (d) Flip Flop
(e) None of these
10. The pin that requests the access of the system bus is?
(a) HLDA (b) HRQ
(c) ADSTB (d) none of the mentioned
(e) CEN
11. In a memory-mapped I/O system, which of the following will not be there?
(a) LDA (b) IN
(c) ADD (d) OUT
(e) None of these
12. The number of hardware interrupts that the processor 8085 consists of is?
(a) 1 (b) 3
(c) 5 (d) 7
(e) None of these
13. Write Through technique is used in which memory for updating the data?
(a) Virtual memory (b) Main memory
(c) Auxiliary memory (d) Cache memory
(e) None of these
14. Generally Dynamic RAM is used as main memory in a computer system as it?
(a) Consumes less power (b) has higher speed
(c) has lower cell density (d) needs refreshing circuitry
(e) None of these
15. The 8257 is able to accomplish the operation of?
(a) verifying DMA operation
(b) write operation (c) read operation
(d) all of the mentioned (e) Only (b) and (c)
16. Virtual memory consists of?
(a) Static RAM (b) Dynamic RAM
(c) Magnetic memory (d) Cache memory
(e) None of these
17. A Stack-organised Computer uses instruction of?
(a) Indirect addressing (b) Two-addressing
(c) Zero addressing (d) Index addressing
(e) None of these
18. The bus is available when the DMA controller receives the signal?
(a) HRQ (b) HLDA
(c) DACK (d) all of the mentioned
(e) None of these
19. The register of 8257 that can only be written in is?
(a) DMA address register
(b) terminal count register (c) mode set register
(d) status register (e) None of these
20. An n-bit microprocessor has?
(a) n-bit program counter
(b) n-bit address register
(c) n-bit ALU (d) n-bit instruction register
(e) None of these
21. In 8257 register format, the selected channel is disabled after the terminal count condition is reached when
(a) auto load is set (b) auto load is reset
(c) TC STOP bit is reset (d) TC STOP bit is set
(e) None of these
22. The multiplicand register & multiplier register of a hardware circuit implementing booth's algorithm have (11101) & (1100). The result shall be
(a) (812) 10 (b) (-12) 10
(c) (12) 10 (d) (-812) 10

(e) None of these

23. PSW is saved in stack when there is a
(a) interrupt recognised
(b) execution of RST instruction
(c) Execution of CALL instruction
(d) All of these (e) None of these

24. The IOW (active low) in its slave mode loads the contents of data bus to
(a) 8-bit mode register
(b) upper/lower byte of 16-bit DMA address register
(c) terminal count register
(d) all of the mentioned
(e) None of these

25. A k-bit field can specify any one of
(a) 3k registers (b) 2k registers
(c) K2 registers (d) K3 registers
(e) None of these

26. In reading the columns of a keyboard matrix, when no key is pressed then all the pins show
(a) 0 (b) 1
(c) F (d) 7
(e) None of these

27. The instructions which copy information from one location to another either in the processor's internal register set or in the external main memory are called
(a) Data transfer instructions.
(b) Program control instructions.
(c) Input-output instructions.
(d) Logical instructions.
(e) None of these

28. What are the actual steps that are followed in identifying any key that is being pressed?
(a) wait for the debounce time
(b) identify the key that is pressed
(c) initially no key should be pressed
(d) all of the mentioned
(e) None of these

29. Memory access in RISC architecture is limited to instructions
(a) CALL and RET (b) PUSH and POP
(c) STA and LDA (d) MOV and JMP
(e) None of these

30. To identify that the key is present in which row and the column
(a) we ground the bits of the row one by one
(b) we ground the bits of the column one by one
(c) we connect the bits of the row to the logic level 1 one by one
(d) we can connect the columns to the logic level 1 one by one
(e) None of these

31. PC Program Counter is also called ………………
(a) instruction pointer (b) memory pointer
(c) data counter (d) file pointer
(e) None of these

32. The registers that store the keyboard and display modes and operations programmed by CPU are
(a) I/O control and data buffers
(b) control and timing registers
(c) return buffers
(d) display address registers
(e) None of these

33. CPU does not perform the operation ……………..
(a) data transfer (b) logic operation
(c) arithmetic operation (d) all of the above
(e) None of these

34. The access time of memory is ………….. the time required for performing any single CPU operation.
(a) Longer than (b) Shorter than
(c) Negligible than (d) Same as
(e) None of these

35. The sensor RAM acts as 8-byte first-in-first-out RAM in
(a) keyboard mode (b) strobed input mode
(c) keyboard and strobed input mode
(d) scanned sensor matrix mode
(e) None of these

36. The data that is entered from the left side of the display unit is of
(a) left entry mode (b) right entry mode
(c) left and right entry modes
(d) Upper entry mode (e) None of these

37. Data hazards occur when ……
(a) Greater performance loss
(b) Pipeline changes the order of read/write access to operands
(c) Some functional unit is not fully pipelined
(d) Machine size is limited
(e) None of these

38. Which of the following is not a mode of data transmission?
(a) simplex (b) duplex
(c) semi duplex (d) half duplex
(e) None of these

39. Interrupts which are initiated by an instruction are
(a) internal (b) external
(c) hardware (d) software
(e) None of these

40. TXD(Transmitted Data Output) pin carries serial stream of the transmitted data bits along with
(a) start bit (b) stop bit
(c) parity bit (d) all of the mentioned
(e) None of these

41. In 8257 (DMA), each of the four channels has
(a) a pair of two 8-bit registers
(b) a pair of two 16-bit registers
(c) one 16-bit register
(d) one 8-bit register
(e) None of these

42. Logic gates with a set of input and outputs is arrangement of
(a) Computational circuit
(b) Logic circuit (c) Design circuits
(d) Register (e) None of these

43. A micro program sequencer
(a) generates the address of next micro instruction to be executed.
(b) generates the control signals to execute a microinstruction.
(c) sequentially averages all microinstructions in the control memory.
(d) enables the efficient handling of a micro program subroutine.
(e) None of these

44. The common register(s) for all the four channels of 8257 are
(a) DMA address register (b) terminal count register
(c) mode set register and status register
(d) none of the mentioned (e)Both (a) and (b)

45. In Reverse Polish notation, expression A*B+C*D is written as
(a) AB*CD*+ (b) A*BCD*+
(c) AB*CD+* (d) A*B*CD+
(e) None of these

46. Suppose that a bus has 16 data lines and requires 4 cycles of 250 nsecs each to transfer data. The bandwidth of this bus would be 2 Megabytes/sec. If the cycle time of the bus was reduced to 125 nsecs and the number of cycles required for transfer stayed the same what would the bandwidth of the bus?
(a) 1 Megabyte/sec (b) 4 Megabytes/sec
(c) 8 Megabytes/sec (d) 2 Megabytes/sec
(e) None of these

47. The amount of time required to read a block of data from a disk into memory is composed of seek time, rotational latency, and transfer time. Rotational latency refers to
(a) the time its takes for the platter to make a full rotation
(b) the time it takes for the read-write head to move into position over the appropriate track
(c) the time it takes for the platter to rotate the correct sector under the head
(d) none of the above
(e) All of the above

48. The IOR (active low) input line acts as output in
(a) slave mode (b) master mode
(c) master and slave mode (d) none of the mentioned
(e) All of the mentioned

49. Computers use addressing mode techniques for ___.
(a) giving programming versatility to the user by providing facilities as pointers to memory counters for loop control
(b) to reduce no. of bits in the field of instruction
(c) specifying rules for modifying or interpreting address field of the instruction
(d) All the above
(e) None of these

50. The pin that disables all the DMA channels by clearing the mode registers is
(a) MARK (b) CLEAR
(c) RESET (d) READY
(e) None of these

51. (2FAOC) 16 is equivalent to
(a) (195 084) 10 (b) (001011111010 0000 1100) 2
(c) Both (a) and (b)
(d) (194 085)10 (e) None of these

52. The pin that is used to write data to the addressed memory location, during DMA write operation is
(a) MEMR (active low) (b) AEN
(c) MEMW (active low) (d) IOW (active low)
(e) None of these

53. The register that stores all the interrupt requests in it in order to serve them one by one on priority basis is
(a) Interrupt Request Register
(b) In-Service Register (c) Priority resolver
(d) Interrupt Mask Register (e) None of these

54. If memory access takes 20 ns with cache and 110 ns without it, then the ratio (cache uses a 10 ns memory) is
(a) 93% (b) 90%
(c) 88% (d) 87%
(e) None of these

55. If the main memory is of 8K bytes and the cache memory is of 2K words. It uses associative mapping. Then each word of cache memory shall be
(a) 11 bits (b) 21 bits
(c) 16 bits (d) 20 bits
(e) None of these

56. A-Flip Flop can be converted into T-Flip Flop by using additional logic circuit
(a) n TQD =• (b) T D =
(c) D = T . Q n (d) n TQD =?
(e) None of these

57. Once the ICW1 is loaded, then the initialization procedure involves
(a) edge sense circuit is reset (b) IMR is cleared
(c) slave mode address is set to 7
(d) all of the mentioned (e) None of these

58. The signal that is applied to the decoding logic, to differentiate between interrupt, code fetch and data bus cycles is
(a) COD (b) INTA (active low)
(c) M/IO (active low) (d) all of the mentioned
(e) None of these

59. 'Aging registers' are
(a) Counters which indicate how long ago their associated pages have been referenced.

(b) Registers which keep track of when the program was last accessed.

(c) Counters to keep track of last accessed instruction.

(d) Counters to keep track of the latest data structures referred.

(e) None of these

60. SIMD represents an organization that __________.

(a) refers to a computer system capable of processing several programs at the same time.

(b) represents organization of single computer containing a control unit, processor unit and a memory unit.

(c) includes many processing units under the supervision of a common control unit

(d) none of the above.

(e) All (a), (b) and (c)

## SOLUTIONS

| | | | | | | |
|---|---|---|---|---|---|---|
| 1. (d) | 2. (b) | 3. (a) | 4. (c) | 5. (c) | 6. (b) | 7. (a) |
| 8. (d) | 9. (d) | 10. (b) | 11. (a) | 12. (c) | 13. (d) | 14. (b) |
| 15. (d) | 16. (a) | 17. (c) | 18. (b) | 19. (c) | 20. (d) | 21. (d) |
| 22. (a) | 23. (a) | 24. (d) | 25. (b) | 26. (b) | 27. (a) | 28. (d) |
| 29. (c) | 30. (a) | 31. (a) | 32. (b) | 33. (a) | 34. (a) | 35. (c) |
| 36. (a) | 37. (b) | 38. (c) | 39. (d) | 40. (d) | 41. (b) | 42. (a) |
| 43. (a) | 44. (c) | 45. (a) | 46. (d) | 47. (a) | 48. (b) | 49. (d) |
| 50. (c) | 51. (b) | 52. (c) | 53. (a) | 54. (b) | 55. (c) | 56. (d) |
| 57. (d) | 58. (d) | 59. (a) | 60. (c) | | | |

# 9 Data Structure

## 1. Introduction

Data Structure is a systematic way to organize data in order to use it efficiently. Following terms are the foundation terms of a data structure.

- Interface − Each data structure has an interface. Interface represents the set of operations that a data structure supports. An interface only provides the list of supported operations, type of parameters they can accept and return type of these operations.
- Implementation − Implementation provides the internal representation of a data structure. Implementation also provides the definition of the algorithms used in the operations of the data structure.

## 2. Asymptotic Notation

Asymptotic analysis of an algorithm, refers to defining the mathematical boundation/framing of its
run-time performance. Using asymptotic analysis, we can very well conclude the best case,
average case and worst-case scenario of an algorithm.

When we study algorithms, we are interested in characterizing them according to their efficiency. We are usually interested in the order of growth of the running time of an algorithm, not in the
exact running time. This is also referred to as the asymptotic running time. We need to develop a way to talk about rate of growth of functions so that we can compare
algorithms. Asymptotic notation gives us a method for classifying functions according to their rate of
growth.
Usually, time required by an algorithm falls under three types:

1. **Best Case** − Minimum time required for program execution.
2. **Average Case −** Average time required for program execution.
3. **Worst Case −** Maximum time required for program execution.

Following are commonly used asymptotic notations used in calculating running time complexity of an algorithm.

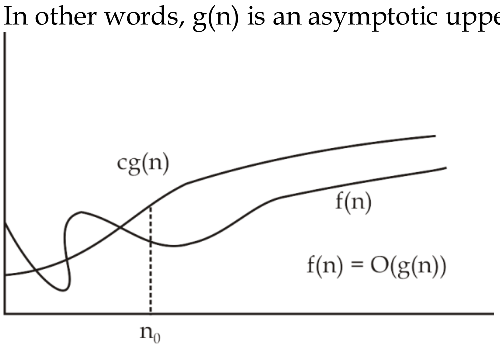
- O Notation
- Ω Notation
- θ Notation

### 2.1 Big Oh Notation, O

Definition: Given a function g(n), we denote (g(n)) to be the set of functions
{ f(n) | there exists positive constants c and $n_0$ such that 0 · f(n) · c g(n) for all n ¸ $n_0$ }

i.e. (g(n)) includes all functions that are upper bounded by g(n)

We say that "f(n) is big-O of g(n)."
As n increases, f(n) grows no faster than g(n).
In other words, g(n) is an asymptotic upper bound on f(n).



Example: $n^2 + n = O(n^3)$

**Proof:**

- Here, we have $f(n) = n^2 + n$, and $g(n) = n^3$
- Notice that if $n \geq 1$, $n \leq n^3$ is clear.
- Also, notice that if $n \geq 1$, $n^2 \leq n^3$ is clear.
- In general, if $a \leq b$, then $n^a \leq n^b$ whenever $n \geq 1$. This fact is used often in these types of proofs.
- Therefore, $n^2 + n \leq n^3 + n^3 = 2n^3$
- We have just shown that $n^2 + n \leq 2n^3$ for all $n \geq 1$
- Thus, we have shown that $n^2 + n = O(n^3)$
  (by definition of Big-O, with $n_0 = 1$, and $c = 2$.)

The constant multiplier c is what allows functions that differ only in their largest coefficient to have the same asymptotic complexity

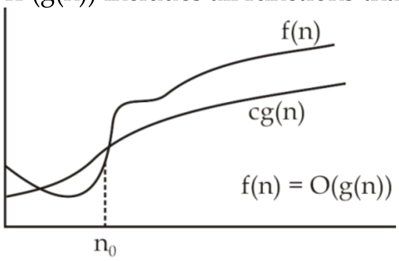
Example: $g(n) = 7n+5$ and $f(n) = n$
− For any choice of $n_0$, need a $c > 7$ (or more) to show $g(n)$ is in $O( f(n) )$

## 2.2 Big-Ω notation

The Ωn is the formal way to express the lower bound of an algorithm's running time. It measures the best case time complexity or best amount of time an algorithm can possibly take to complete.

Definition: $f(n) = \Omega(g(n))$ iff there are two positive constants c and n0 such that
$|f(n)| \geq c\ |g(n)|$ for all $n \geq n_0$

• If f(n) is nonnegative, we can simplify the last condition to
$0 \leq c\ g(n) \leq f(n)$ for all $n \geq n_0$
• We say that "f(n) is omega of g(n)."
• As n increases, f(n) grows no slower than g(n). In other words, g(n) is an asymptotic lower bound on f(n).
Ω (g(n)) includes all functions that are lower bounded by g(n)




Similar to Big-O, we will slightly change the notation, and write $f(n) = \Omega(g(n))$ to mean
$f(n)\ 2 \in \Omega(g(n))$
Relationship between Big-O and Big- Ω: $f(n) = \Omega(g(n)) \Leftrightarrow g(n) = O(f(n))$

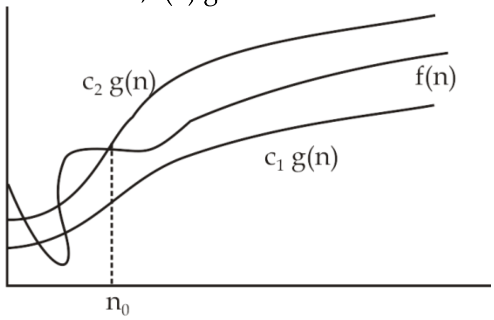
*Example:* $n^3 + 4n^2 = \Omega(n^2)$
Proof:
• Here, we have $f(n) = n^3 + 4n^2$, and $g(n) = n^2$
• It is not too hard to see that if $n \geq 0$, $n^3 \leq n^3 + 4n^2$
• We have already seen that if $n \geq 1$, $n^2 \leq n^3$
• Thus, when $n \geq 1$, $n^2 \leq n^3 \leq n^3 + 4n^2$
• Therefore,$1n^2 \leq n^3 + 4n^2$ for all $n \geq 1$
• Thus, we have shown that $n^3 + 4n^2 = \Omega(n^2)$
(by definition of Big-Ω, with $n_0 = 1$, and $c = 1$.)

## 2.3 Big Theta Notation -Θ

Definition (Big–Theta, Θ()): Let f(n) and g(n) be functions that map positive integers to positive real numbers. We say that f(n) is $\Theta(g(n))$ (or $f(n) \in \Theta(g(n))$) if and only if $f(n) \in O(g(n))$ and $f(n) \in \Omega(g(n))$.

$f(n) = \Theta(g(n))$ iff there are three positive constants $c_1$, $c_2$ and $n_0$ such that
$c_1|g(n)| \leq |f(n)| \leq c_2|g(n)|$ for all $n \geq n_0$

• If f(n) is nonnegative, we can simplify the last condition to
$0 \leq c_1 g(n) \leq f(n) \leq c_2 g(n)$ for all $n \geq n_0$
• We say that "f(n) is theta of g(n)."
• As n increases, f(n) grows at the same rate as g(n). In other words, g(n) is an asymptotically tight bound on f(n).




$f(n) = O(g(n)) \Rightarrow f \leq g$
$f(n) = \Omega(g(n)) \Rightarrow f \geq g$
$f(n) = \Theta(g(n)) \Rightarrow f \approx g$

- It is important to remember that a Big-O bound is only an upper bound. So, an algorithm that is $O(n^2)$ might not ever take that much time. It may actually run in O(n) time.
- Conversely, an Ω bound is only a lower bound. So an algorithm that is $\Omega(n \log n)$ might actually be $\Theta(2^n)$.
- Unlike the other bounds, a Θ-bound is precise. So, if an algorithm is $\Theta(n^2)$, it runs in quadratic time.

## 3. Arrays & Stack

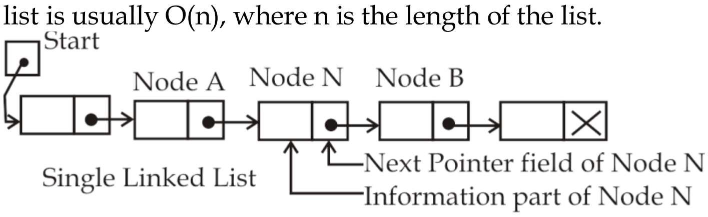
Array is a container which can hold a fix number of items and these items should be of the same type. Most of the data structures make use of arrays to implement their algorithms. Following are the important terms to understand the concept of Array.

- **Element −** Each item stored in an array is called an element.
- **Index −** Each location of an element in an array has a numerical index, which is used to identify the element.

Following are the basic operations supported by an array.

- Traverse − print all the array elements one by one.
- Insertion − Adds an element at the given index.
- Deletion − Deletes an element at the given index.
- Search − Searches an element using the given index or by the value.
- Update − Updates an element at the given index.

**Linked Lists**
Linked lists are a common alternative to arrays in the implementation of data structures. Each item in a linked list contains a data element of some type and a pointer to the next item in the list. It is easy to insert and delete elements in a linked list, which is not a natural operation on arrays. On the other hand access to an element in the middle of the list is usually O(n), where n is the length of the list.

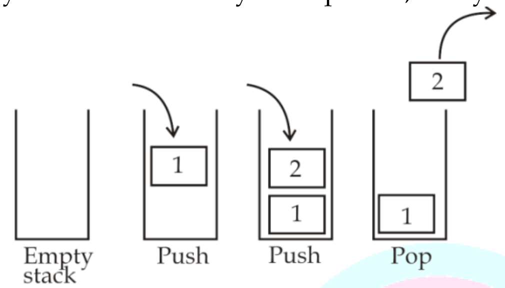



An item in a linked list consists of a struct containing the data element and a pointer to another linked list. This gives rise to the following definition:
struct list {
string data;

struct list* next;
};
typedef struct list* list;

This definition is an example of a recursive type. A struct of this type contains a pointer to another struct of the same type, and so on. We usually use the special element of type t*, namely NULL, to indicate that we have reached the end of the list. Sometimes (as will be the case for queues introduced next), we can avoid the explicit use of NULL and obtain more elegant code. The type definition is there to create the type name list, which stands for a pointer to a struct list.

A **stack** is a container of objects that are inserted and removed according to **the last-in first-out (LIFO)** principle. In the pushdown stacks only two operations are allowed: push the item into the stack, and pop the item out of the stack. A stack is a limited access data structure - elements can be added and removed from the stack only at the top. push adds an item to the top of the stack, pop removes the item from the top. A helpful analogy is to think of a stack of books; you can remove only the top book, also you can add a new book on the top.




**Basic Operations**

- push() − Pushing (storing) an element on the stack.
- pop() − Removing (accessing) an element from the stack.

To use a stack efficiently, we need to check the status of stack as well. For the same purpose, the following functionality is added to stacks −

- peek() − get the top data element of the stack, without removing it.
- isFull() − check if stack is full.
- isEmpty() − check if stack is empty.

**Algorithm for PUSH Operation**

```
begin procedure push: stack, data

      if stack is full
         return null
      endif

      top ← top + 1

      stack[top] ← data

   end procedure
```

**Algorithm for Pop Operation**

```
   begin procedure pop: stack

      if stack is empty
         return null
      endif

      data ← stack[top]

      top ← top - 1
```

```
    return data

end procedure
```

**Stack Program using Linked List (in C)**

**Linked stack source code file (stack.c)**

```c
#include <stdio.h>

struct node
{
   int data;
   struct node* next;
};

/*
   init the stack
*/
void init(struct node* head)
{
   head = NULL;
}

/*
   push an element into stack
*/
struct node* push(struct node* head,int data)
{
   struct node* tmp = (struct node*)malloc(sizeof(struct node));
   if(tmp == NULL)
   {
      exit(0);
   }
   tmp->data = data;
   tmp->next = head;
   head = tmp;
   return head;
}
/*
   pop an element from the stack
*/
struct node* pop(struct node *head,int *element)
{
   struct node* tmp = head;
   *element = head->data;
   head = head->next;
   free(tmp);
   return head;
}
/*
   returns 1 if the stack is empty, otherwise returns 0
*/
int empty(struct node* head)
{
   return head == NULL ? 1 : 0;
}
```

```
/*
   display the stack content
*/
void display(struct node* head)
{
   struct node *current;
   current = head;
   if(current!= NULL)
   {
      printf("Stack: ");
      do
      {
         printf("%d ",current->data);
         current = current->next;
      }
      while (current!= NULL);
      printf("\n");
   }
   else
   {
      printf("The Stack is empty\n");
   }

}
```

**Linked stack test program:**

```
#include <stdio.h>

#include "linkedstack.h"

int main()
{
   struct node* head = NULL;
   int size, element;
   int counter = 0;

   printf("Enter the number of stack elements:");
   scanf("%d",&size);

   printf("--- Push elements into the linked stack ---\n");

   init(head);

   while(counter < size)
   {

      printf("Enter a number to push into the stack:");
      scanf("%d",&element);
      head = push(head,element);
      display(head);
      counter++;
   }

   printf("--- Pop elements from the linked stack --- \n");
   while(empty(head) == 0)
   {
      head = pop(head,&element);
      printf("Pop %d from stack\n",element);
      display(head);
```

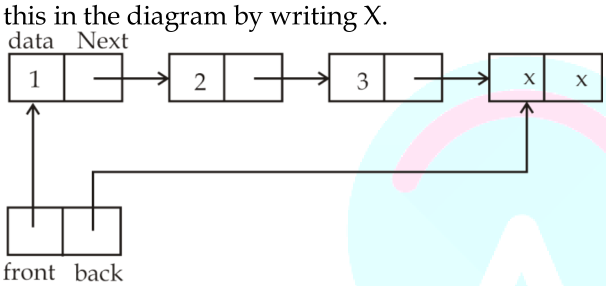
```
            }

            return 0;
        }
```

## 4. Queue

A queue is a data structure where we add elements at the back and remove elements from the front. In that way a queue is like "waiting in line": the first one to be added to the queue will be the first one to be removed from the queue. This is also called a FIFO (First In First Out) data structure.

Queues are common in many applications. For example, when we read a book from a file as in Assignment 2, it would be natural to store the the words in a queue so that when we are finished reading the file the words are in the order they appear in the book.

A queue is implemented as a struct with a front and back field. The front field points to the front of the queue, the back field points to the back of the queue. In arrays, we often work with the length which is one greater than the index of the last element in the array. In queues, we use a similar
strategy, making sure the back pointer points to one element past the end of the queue. Unlike arrays, there must be something in memory for the pointer to refer to, so there is always one extra element at the end of the queue which does not have valid data or next pointer. We have indicated
this in the diagram by writing X.




The above picture yields the following definition.
struct queue {
list front;
list back;
};

- enqueue() − add (store) an item to the queue.
- dequeue() − remove (access) an item from the queue.
- peek() − Gets the element at the front of the queue without removing it.
- isfull() − Checks if the queue is full.
- isempty() − Checks if the queue is empty.

In queue, we always dequeue (or access) data, pointed by front pointer and while enqueing (or storing) data in the queue we take help of rear pointer.

**Enqueue Operation**
The following steps should be taken to enqueue (insert) data into a queue −
Step 1 − Check if the queue is full.
Step 2 − If the queue is full, produce overflow error and exit.
Step 3 − If the queue is not full, increment rear pointer to point the next empty space.
Step 4 − Add data element to the queue location, where the rear is pointing.
Step 5 − return success.

**Algorithm for enqueue operation**
**1.** procedure enqueue(data)
**2.** if queue is full
**3.** return overflow
**4.** endif
**5.** rear ← rear + 1

6. queue[rear] ← data
7. return true
8. end procedure

**Dequeue Operation**

Step 1 − Check if the queue is empty.
Step 2 − If the queue is empty, produce underflow error and exit.
Step 3 − If the queue is not empty, access the data where front is pointing.
Step 4 − Increment front pointer to point to the next available data element.
Step 5 − Return success.

**Algorithm for dequeue operation**

1. procedure dequeue
2. if queue is empty
3. return underflow
4. end if
5. data = queue[front]
6. front ← front + 1
7. return true
8. end procedure

Queues can be implemented as linked lists. Linked list implementations of queues often require two pointers or references to links at the beginning and end of the list. Using a pair of pointers or references opens the code up to a variety of bugs especially when the last item on the queue is dequeued or when the first item is enqueued.

Under the list representation, a queue q consists of a list and two pointers, q.front and q.rear. The operations are insertion and deletion. Special attention is required when the last element is removed from a queue. In that case , q.rear must also be set to null, Since in an empty queue both r.front and q.rear must be null.

The pseudo code for deletion is below:

```
 1.          if (empty(q))
    {
printf("Queue is Underflow");
exit(1);
    }
   f = q.front;
   t = info(f);
   q.front = next(f);
      if (q.front == null)
   q.rear = null;
   freenode(f);
return(t);
```

The operation insert algorithm is implemented

```
  f = getnode();
   info(f) = x;
   next(f) = null;
        if (q.rear == null)
          q.front =f;
        else
      next(q.rear) = f;
      q.rear = f;
```

We can use a list to represent a priority queue in ordered list or unordered list. For an ascending Priority queue, insertion is implemented by the place operation, which keeps the list ordered, and deletion of the minimum element is implemented by the delete operation, which removes the first element from the list. A Descending priority queue can be implemented by keeping the list in decending order rather than ascending, or by using remove to delete the minimum element. In an ordered list, if you want to insert an element to the priority queue, it will require examining an average of approximately n/2 nodes but only one search for deletion.

An unordered list may also be used as a priority queue. If you want to insert an element to the list always requires examining only one node but always requires examining n elements for removal of an element.

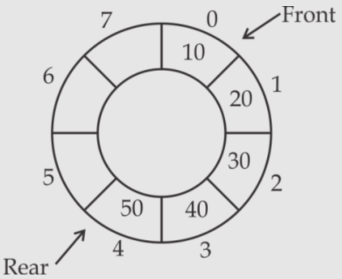
The advantage of a list over an array for implementing a priority queue is that an element can be inserted into a list without moving any other elements, where as this is impossible for an array unless extra space is left empty.

**Priority Queue** is more specialized data structure than Queue. Like ordinary queue, priority queue has same method but with a major difference. In Priority queue items are ordered by key value so that item with the lowest value of key is at front and item with the highest value of key is at rear or vice versa. So we're assigned priority to item based on its key value. Lower the value, higher the priority.

A typical priority queue supports following operations.
- insert(item, priority): Inserts an item with given priority.
- getHighestPriority(): Returns the highest priority item.
- deleteHighestPriority(): Removes the highest priority item.

A **circular queue** is an abstract data type that contains a collection of data which allows addition of data at the end of the queue and removal of data at the beginning of the queue. Circular queues have a fixed size. Circular queue follows FIFO principle. Queue items are added at the rear end and the items are deleted at front end of the circular queue.

**Circular Queue operations:**




- Two pointers called FRONT and REAR are used to keep track of the first and last elements in the queue.
- When initializing the queue, we set the value of FRONT and REAR to -1.
- On enqueing an element, we circularly increase the value of REAR index and place the new element in the position pointed to by REAR.
- On dequeueing an element, we return the value pointed to by FRONT and circularly increase the FRONT index.
- Before enqueing, we check if queue is already full.
- Before dequeuing, we check if queue is already empty.
- When enqueing the first element, we set the value of FRONT to 0.
- When dequeing the last element, we reset the values of FRONT and REAR to -1.

However, the check for full queue has a new additional case:

Case 1: FRONT = 0 && REAR == SIZE - 1
Case 2: FRONT = REAR + 1

The second case happens when REAR starts from 0 due to circular increment and when its value is just 1 less than FRONT, the queue is full.

**Algorithm for Insertion in a circular queue**
Insert CircularQueue ( )

1. If (FRONT == 1 and REAR == N) or (FRONT == REAR + 1) Then
2. Print: Overflow
3. Else

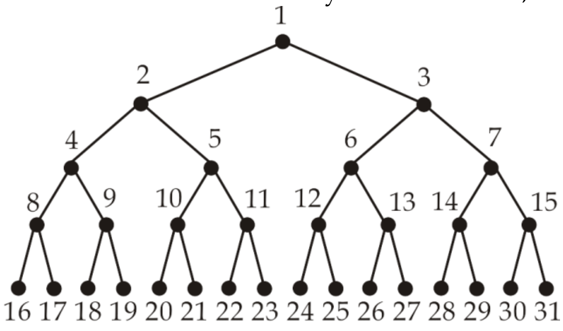
4. If (REAR == 0) Then [Check if QUEUE is empty]
5. (a) Set FRONT = 1
6. (b) Set REAR = 1
7. Else If (REAR == N) Then [If REAR reaches end if QUEUE]
8. Set REAR = 1
9. Else
10. Set REAR = REAR + 1 [Increment REAR by 1]
11. [End of Step 4 If]
12. Set QUEUE[REAR] = ITEM
13. Print: ITEM inserted
14. [End of Step 1 If]
15. Exit

**Algorithm for Deletion in a circular queue**
Delete CircularQueue ( )

1. If (FRONT == 0) Then [Check for Underflow]
2. Print: Underflow
3. Else
4. ITEM = QUEUE[FRONT]
5. If (FRONT == REAR) Then [If only element is left]
6. (a) Set FRONT = 0
7. (b) Set REAR = 0
8. Else If (FRONT == N) Then [If FRONT reaches end if QUEUE]
9. Set FRONT = 1
10. Else
11. Set FRONT = FRONT + 1 [Increment FRONT by 1]
12. [End of Step 5 If]
13. Print: ITEM deleted
14. [End of Step 1 If]
15. Exit

## 5. Binary Trees

A tree is a finite set of nodes having a distinct node called root.
Binary Tree is a tree which is either empty or has at most two subtrees, each of the subtrees also being a binary tree. It means each node in a binary tree can have 0, 1 or 2 subtrees. A left or right subtree can be empty.




A binary tree is made of nodes, where each node contains a "left" pointer, a "right" pointer, and a data element. The "root" pointer points to the topmost node in the tree. The left and right pointers point to smaller "subtrees" on either side. A null pointer represents a binary tree with no elements -- the empty tree. The formal recursive definition is: a binary tree is either empty (represented by a null pointer), or is made of a single node, where the left and right pointers (recursive definition ahead) each point to a binary tree.

It has a distinct node called root. And every node has either 0,1 or 2 children. So in a binary tree every node has a maximum of 2 children.

If A is the root of a binary tree & B the root of its left or right subtree, then A is the parent or father of B and B is the left or right child of A. Those nodes having no children are leaf nodes. Any node say A is the ancestor of node B and B is the descendant of A if A is either the father of B or the father of some ancestor of B. Two nodes having same father are called brothers or siblings.

**Going from leaves to root is called climbing the tree & going from root to leaves is called descending the tree.**

A binary tree in which every non leaf node has non empty left & right subtrees is called a strictly binary tree.

- The number of subtrees of a node is called the degree of the node. In a binary tree, all nodes have degree 0, 1, or 2.
- A node of degree zero is called a terminal node or leaf node.
- A non-leaf node is often called a branch node.
- The degree of a tree is the maximum degree of a node in the tree. A binary tree is degree 2.
- A directed path from node $n_1$ to $n_k$ is defined as a sequence of nodes $n_1$, $n_2$, ..., $n_k$ such that $n_i$ is the parent of $n_i$+1 for 1 <= i < k. An undirected path is a similar sequence of undirected edges. The length of this path is the number of edges on the path, namely k – 1 (i.e., the number of nodes – 1). There is a path of length zero from every node to itself. Notice that in a binary tree there is exactly one path from the root to each node.
- The level or depth of a node with respect to a tree is defined recursively: the level of the root is zero; and the level of any other node is one higher than that of its parent. Or to put it another way, the level or depth of a node $n_i$ is the length of the unique path from the root to $n_i$.
- The height of $n_i$ is the length of the longest path from $n_i$ to a leaf. Thus, all leaves in the tree are at height 0.
- The height of a tree is equal to the height of the root. The depth of a tree is equal to the level or depth of the deepest leaf; this is always equal to the height of the tree.
- If there is a directed path from n1 to n2, then n1 is an ancestor of n2 and n2 is a descendant of n1.

**The structure defining a node of binary tree in C is as follows.**

```
Struct node
{
struct node *lc ; /* points to the left child */
int data; /* data field */
struct node *rc; /* points to the right child */
}
```

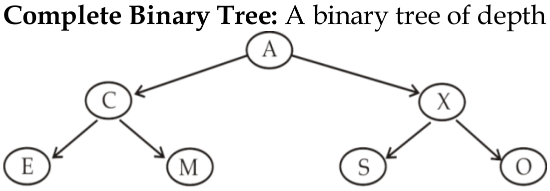
## 5.1 Linked Representation of Binary Tree:

Binary trees can be represented by links where each node contains the address of the left child and the right child. If any node has its left or right child empty then it will have in its respective link field, a null value. A leaf node has null value in both of its links.
The structure defining a node of binary tree in C is as follows.

```
Struct node
{
struct node *lc ; /* points to the left child */
int data; /* data field */
struct node *rc; /* points to the right child */
}
```

## 5.2 Types of Binary Tree

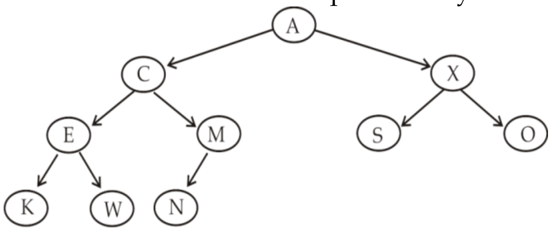
**Complete Binary Tree:** A binary tree of depth d is an almost complete binary tree if:




- Each leaf in the tree is either at level d or at level d – 1.

- For any node $n_d$ in the tree with a right descendant at level d, all the left descendants of $n_d$ that are leaves are also at level d.

**Almost complete binary tree:** An almost complete strictly binary tree with N leaves has 2N – 1 nodes (as does any other strictly binary tree). An almost complete binary tree with N leaves that is not strictly binary has 2N nodes. There are two distinct almost complete binary trees with N leaves, one of which is strictly binary and one of which is not.




There is only a single almost complete binary tree with N nodes. This tree is strictly binary if and only if N is odd.

**Degenerate (or pathological) tree:** A Tree where every internal node has one child. Such trees are performance-wise same as linked list.

## 5.3 Binary Search Tree

A binary search tree is a data structure that allows for key lookup, insertion, and deletion. It is a binary tree, meaning every node of the tree has at most two child nodes, a left child and a right child. Each node of the tree holds the following information:

- x.key - Value stored in node x
- x.left- Pointer to the left child of node x. NIL if x has no left child
- x.right - Pointer to the right child of node x. NIL if x has no right child
- x.parent - Pointer to the parent node of node x. NIL if x has no parent, i.e. x is the root of the tree

Each node has the following attributes:

- p, left, and right, which are pointers to the parent, the left child, and the right child, respectively, and key, which is key stored at the node.
- Here k is the key that is searched for and x is the start node.

BST-Search(x, k)
1. y ← x
2. while y ≠ nil do
3. if key[y] = k then return y
4. else if key[y] < k then y ← right[y]
5. else y ← left[y]
6. return ("NOT FOUND")

**The Maximum and the Minimum**
To find the minimum identify the leftmost node, i.e. the farthest node you can reach by following only left branches. To find the maximum identify the rightmost node, i.e. the farthest node you can reach by following only right branches.

**BST-Minimum(x)**
1. if x = nil then return ("Empty Tree")
2. y ← x
3. while left[y] ≠ nil do y ← left[y]
4. return (key[y])

**BST-Maximum(x)**
1. if x = nil then return ("Empty Tree")
2. y ← x
3. while right[y] ≠ nil do y ← right[y]
4. return (key[y])

**Insertion in BST**
Suppose that we need to insert a node z such that k = key[z]. Using binary search we find a nil such that replacing it by z does not break the BST-property.

BST-Insert(x, z, k)
1. if x = nil then return "Error"
2. y ← x
3. while true do {
4. if key[y] < k
5. then z ← left[y]
6. else z ← right[y]
7. if z = nil break
8. }
9. if key[y] > k then left[y] ← z
10. else right[p[y]] ← z

**Predecessor and Successor in BST**
The idea for finding the successor of a given node x.

- If x has the right child, then the successor is the minimum in the right subtree of x.
- Otherwise, the successor is the parent of the farthest node that can be reached from x by following only right branches backward.

**BST-Successor(x)**
1. if right[x] ≠ nil then
2. { y ← right[x]
3. while left[y] ≠ nil do y ← left[y]
4. return (y) }
5. else
6. { y ← x
7. while right[p[x]] = x do y ← p[x]
8. if p[x] ≠ nil then return (p[x])
9. else return ("No Successor.") }

The predecessor can be found similarly with the roles of left and right exchanged and with the roles of maximum and minimum exchanged.

**Deletion in BST**
The following algorithm deletes z from BST T.

BST-Delete(T, z)
1. if left[z] = nil or right[z] = nil
2. then y ← z
3. else y ← BST-Successor(z)
4. y is the node that's actually removed.
5. Here y does not have two children.
6. if left[y] ≠ nil
7. then x ← left[y]
8. else x ← right[y]
9. x is the node that's moving to y's position.
10. if x ≠ nil then p[x] ← p[y]
11. p[x] is reset If x isn't NIL.
12. Resetting is unnecessary if x is NIL.

## 5.4 Traversal

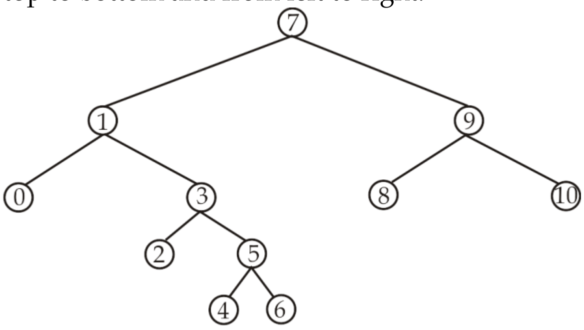
A traversal is a process that visits all the nodes in the tree. Since a tree is a nonlinear data structure, there is no unique traversal. Traversing a tree means visiting each node in a specified order. This process is not as commonly used as finding, inserting, and deleting nodes. One reason for this is that traversal is not particularly fast. Following are two kinds of traversal:

- **Depth-first traversal**
- **Breadth-first traversal**

There are three different types of depth-first traversals:

- **PreOrder traversal -** visit the parent first and then left and right children;
- **InOrder traversal -** visit the left child, then the parent and the right child;
- **PostOrder traversal** - visit left child, then the right child and then the parent;

There is only one kind of breadth-first traversal--the level order traversal. This traversal visits nodes by levels from top to bottom and from left to right.



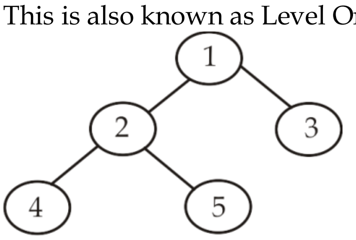

**Pre-order traversal:** To traverse a binary tree in Pre-order, following operations are carried-out
(i) Visit the root,
(ii) Traverse the left subtree, and
(iii) Traverse the right subtree.
Therefore, the Preorder traversal of the above tree will output:
7, 1, 0, 3, 2, 5, 4, 6, 9, 8, 10

**In-order traversal:** To traverse a binary tree in In-order, following operations are carried-out
(i) Traverse the left most subtree starting at the left external node,
(ii) Visit the root, and
(iii) Traverse the right subtree starting at the left external node.
Therefore, the In-order traversal of the above tree will output:
0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10

**Post-order traversal:** To traverse a binary tree in Post-order, following operations are carried-out
(i) Traverse all the left external nodes starting with the left most subtree which is then followed by bubble-up all the internal nodes,
(ii) Traverse the right subtree starting at the left external node which is then followed by bubble-up all the internal nodes, and
(iii) Visit the root.
Therefore, the Post-order traversal of the above tree will output:
0, 2, 4, 6, 5, 3, 1, 8, 10, 9, 7

**<u>Breadth-First Traversal</u>**

This is also known as Level Order Tree Traversal.




Level order traversal of the above tree is 1 2 3 4 5

**Algorithm:**
printLevelorder(tree)

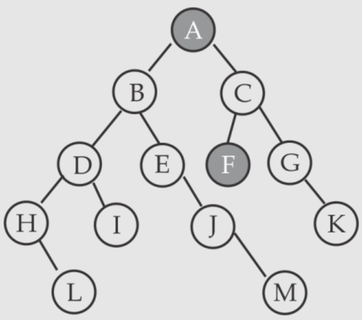
1) Create an empty queue q
2) temp_node = root /*start from root*/
3) Loop while temp_node is not NULL
   a) print temp_node->data.
   b) Enqueue temp_node's children (first left then right children) to q
   c) Dequeue a node from q and assign it's value to temp_node

For BFS algorithm, visit a node's siblings before its children, while in DFS algorithm, visit a node's children before its siblings

BFS is useful in finding shortest path. BFS can be used to find the shortest distance between some starting node and the remaining nodes of the graph. DFS in not so useful in finding shortest path. It is used to perform a traversal of a general graph and the idea of DFS is to make a path as long as possible, and then go back (backtrack) to add branches also as long as possible.

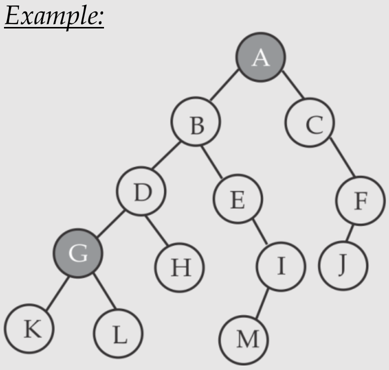
BFS starts traversal from the root node and then explore the search in the level by level manner i.e. as close as possible from the root node. DFS starts the traversal from the root node and explore the search as far as possible from the root node i.e. depth wise.

*Example:*




**Before countering goal node F:**
BFS algorithm encounters nodes: ABCDE
DFS algorithm encounters nodes: ABDHLIEJMC

*Example:*




**Before countering goal node G:**
BFS algorithm encounters nodes: ABCDEF
DFS algorithm encounters nodes: ABD

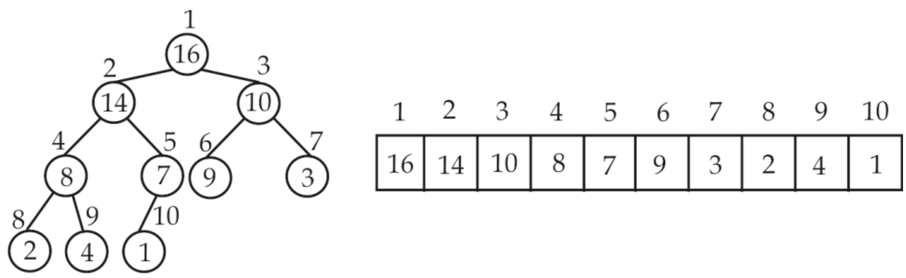
## 5.5 Heap and Heap Sort

The (Binary) heap data structure is an array object that can be viewed as a nearly complete binary tree. A binary tree with n nodes and depth k is complete iff its nodes correspond to the nodes numbered from 1 to n in the full binary tree of depth k

## Attributes of a Heap

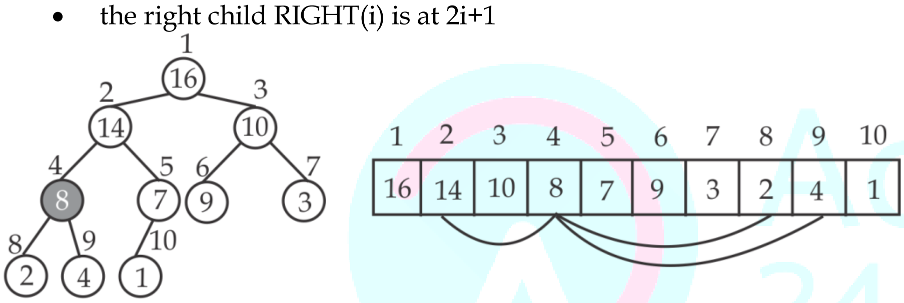
An array A that presents a heap with two attributes:

- length[A]: the number of elements in the array.
- heap-size[A]: the number of elements in the heap stored with array A.
- length[A]≥ heap-size[A]




If a complete binary tree with n nodes is represented sequentially, then for any node with index i, $1 \le i \le n$, we have

- A[1] is the root of the tree
- the parent PARENT(i) is at $\lfloor i/2 \rfloor$ if $i \neq 1$
- the left child LEFT(i) is at 2i
- the right child RIGHT(i) is at 2i+1




There are two kinds of binary heaps: max-heaps and min-heaps.

- In a max-heap, the max-heap property is that for every node i other than the root, the largest element in a max-heap is stored at the root and the subtree rooted at a node contains values no larger than that contained at the node itself
  A[Parent(i)] ≥ A[i]
- In a min-heap, the min-heap property is that for every node i other than the root, the smallest element in a min-heap is at the root and the subtree rooted at a node contains values no smaller than that contained at the node itself.
  A[Parent(i)] ≤ A[i]

**Algorithm**

MAX-HEAPIFY(A, i)
1. l ← LEFT(i)
2. r ← RIGHT(i)
3. if l ≤ heap-size[A] and A[l] > A[i]
4. then largest ← l
5. else largest ← i
6. if r ≤ heap-size[A] and a[r] > A[largest]
7. then largest ← r
8. if largest ≠ i
9. then exchange A[i] A[largest]
10. MAX-HEAPIFY (A, largest)

We can use the MAX-HEAPIFY procedure to convert an array A=[1..n] into a max-heap in a bottom-up manner. The elements in the subarray A[(⌊n/2⌋+1)…n ] are all leaves of the tree, and so each is a 1-element heap. The procedure BUILD-MAX-HEAP goes through the remaining nodes of the tree and runs MAX-HEAPIFY on each one.

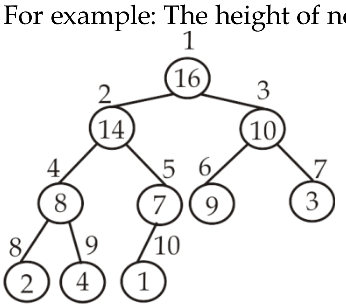
BUILD-MAX-HEAP(A)
1. heap-size[A] ← length[A]
2. for i ← ⌊length[A]/2⌋ downto 1
3. do MAX-HEAPIFY(A,i)

Since the maximum element of the array is stored at the root, A[1] we can exchange it with A[n].
If we now "discard" A[n], we observe that A[1...(n−1)] can easily be made into a max-heap.
The children of the root A[1] remain max-heaps, but the new root A[1] element may violate the max-heap property, so we need to readjust the max-heap. That is to call MAX-HEAPIFY(A, 1).

HEAPSORT(A)
1. BUILD-MAX-HEAP(A)
2. for i ← length[A] downto 2
3. do exchange A[1] A[i]
4. heap-size[A]← heap-size[A]−1
5. MAX-HEAPIFY(A, 1)

**The height of a heap**

The height of a node in a heap is the number of edges on the longest simple downward path from the node to a leaf, and the height of the heap to be the height of the root.
For example: The height of node 2 is 2 and the height of the heap is 3




**Time complexity**

- The HEAPSORT procedure takes O(nlgn) time
- The call to BUILD-MAX-HEAP takes O(n) time
- Each of the n−1 calls to MAX-HEAPIFY takes O(lgn) time

## 6. Graph

Graphs are one of the most interesting data structures in computer science. Graphs and the trees are somewhat similar by their structure. In fact, tree is derived from the graph data structure. However there are two important differences between trees and graphs.

- Unlike trees, in graphs, a node can have many parents.
- The link between the nodes may have values or weights.

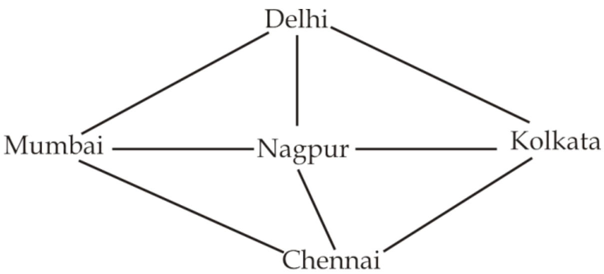
Graph consists of a non-empty set of points called vertices and a set of edges that link vertices.
**Formal Definition :** A graph G= (V, E) consists of
- a set V={v1 , v2 .....,vn } of n >1 vertices and
- a set of E={e1 ,e2 ,.....,em } of m>0 edges

such that each edge ek is corresponds to an un ordered pair of vertices (vi ,vj) where 0 < i,j <= n and 0 < k <= m.

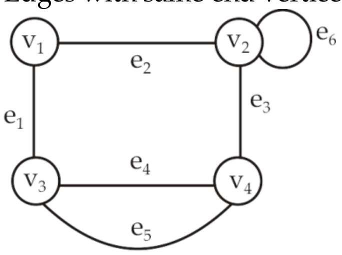
A road network is a simple example of a graph, in which vertices represents cities and road connecting them are correspond to edges.
V= { Delhi, Chennai, Kolkata, Mumbai, Nagpur }

E= { (Delhi, Kolkata), (Delhi, Mumbai}, (Delhi, Nagpur), (Chennai, Kolkata), (Chennai, Mumbai), (Chennai, Nagpur), (Kolkata, Nagpur), (Mumbai, Nagpur) }



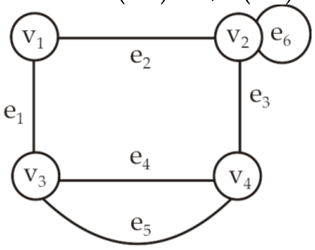

Loop is an edge that connects a vertex to itself. Edge e6 in the figure below is a loop.
Edges with same end vertices are called parallel edges. Edges e4 and e5 are parallel edges in the below figure.




- A Graph without loops and parallel edges is called a simple graph.
- A graph with isolated vertices (no edges) is called null graph.
- Set of edges E can be empty for a graph but not set of vertices V.

Incidence: if a vertex vi is an end vertex of an edge ek, we say vertex vi is incident on ek and ek is incident on vi.
- e1 is incident on v1 and v3 in the below figure.
- v4 is incident on e3, e4, and e5 in the figure below.

Degree: Degree of a vertex is number of edges incident on it, with loops counted twice.
- d(v1 ) = 2, d(v2 ) = 4, d(v3 ) = 3, and d(v4 ) = 3 in the figure below.




**Adjacent Edges:** Two non-parallel edges are adjacent if they have a vertex in common. e1 and e2, e2 and e6, e2 and e3, e1 and e4 are adjacent edges.

**Adjacent vertices:** Two vertices are adjacent if they are connected by an edge. v1 and v3, v1 and v2, v2 and v4 are adjacent vertices.
Any graph problems can be solved by visiting each and every vertex in a systematic manner. There are two main method to traversal graphs, Depth First Search (DFS) and Breadth First Search (BFS).

**Depth First Search (DFS):** Initially all vertices of the graph are unvisited. Start visiting the graph from any vertex, say v. For each unvisited adjacent vertex of v, search recursively. This process stops when all vertices reachable from v are visited. If there are any more unvisited vertices are present, select any unvisited vertex and repeat the same search process. This method is called depth first search since searching is done forward (deeper) from current node. The distance from start vertex is called depth.

**Algorithm DFS ( G )**
1. for i = 1 to n do // Initialize all vertices are unvisited
2. status[i] = unvisited
3. parent[i] = NULL

4. for i = 1 to n do
5. if (status[i] == unvisited) // If there exits an unvisited vertex, start traversal
6. DF-Travel(i)

**Algorithm DF-Travel ( v )**

1. status[v] = visited
2. for each vertex u adjacent to v do
3. if status[u] == unvisited then
4. parent[u] = v
5. DF-Travel ( u )

**Breadth First Search (BSF):** Initially all vertices of the graph are unvisited. Start visiting the graph from any vertex, say v. Visit each unvisited adjacent vertex of v. Repeat the process for each vertex visited. This process stops when all vertices reachable from v are visited. If there is any more unvisited vertices are present select any unvisited vertex and repeat the same search process. This method is called breadth first search, since it works outward from a center point, much like the ripples created when throwing a stone into a pond. It moves outward in all directions, one level at a time.

Step 1: Push the root node in the Queue.
Step 2: Loop until the queue is empty.
Step 3: Remove the node from the Queue.
Step 4: If the removed node has unvisited child nodes, mark them as visited and insert the unvisited children in the queue.

## 7. Sorting

Sorting refers to the operation or technique of arranging and rearranging sets of data in some specific order. A collection of records called a list where every record is having one or more fields. The fields which contains unique value for each record, is termed as the key field. For example, a phone number directory can be thought of as a list where each record has three fields – 'name' of the person, 'address' of those person, and their 'phone numbers'. Being unique phone number can work as a key to locate any record in the list.

Sorting is the operation performed to arrange the records of a table or list in some order according to some specific ordering criterion. Sorting is performed according to some key value of each record.

The records are either sorted either numerically or alphanumerically. The records are then arranged in ascending or descending order depending on the numerical value of the key.

**Complexity of Sorting Algorithms**

The complexity of sorting algorithm calculates the running time of a function in which 'n' number of items are to be sorted. The choice for which sorting method is suitable for a problem depends on several dependency configurations for different problems. The most noteworthy of these considerations are:

The length of time spent by the programmer in programming a specific sorting program

- Amount of machine time necessary for running the program
- The amount of memory necessary for running the program

**Efficiency of Sorting Techniques**

To get the amount of time required to sort an array of 'n' elements by a particular method, the normal approach is to analyze the method to find the number of comparisons (or exchanges) required by it. Most of the sorting techniques are data sensitive and so the metrics for them depends on the order in which they appear in an input array.

Various sorting techniques are analyzed in various cases and named these cases as follows:

- Best case
- Worst case

- Average case

Hence, the result of these cases is often a formula giving the average time required for a particular sort of size 'n'. Most of the sort methods have time requirements that range from O(nlog n) to O($n^2$).

**Types of Sorting**

**Internal Sorts:-** This method uses only the primary memory during sorting process. All data items are held in main memory and no secondary memory is required this sorting process. If all the data that is to be sorted can be accommodated at a time in memory is called internal sorting. There is a limitation for internal sorts; they can only process relatively small lists due to memory constraints. There are 3 types of internal sorts.

- Selection Sort: - Ex: - Selection sort algorithm, Heap Sort algorithm
- Insertion Sort: - Ex: - Insertion sort algorithm, Shell Sort algorithm
- Exchange Sort: - Ex: - Bubble Sort Algorithm, Quick sort algorithm

**External Sorts:-** Sorting large amount of data requires external or secondary memory. This process uses external memory such as HDD, to store the data which is not fit into the main memory. So, primary memory holds the currently being sorted data only. All external sorts are based on process of merging. Different parts of data are sorted separately and merged together.

- Ex: - Merge Sort

## 7.1 Insertion Sort

One of the simplest methods to sort an array is an insertion sort. An example of an insertion sort occurs in everyday life while playing cards. To sort the cards in your hand you extract a card, shift the remaining cards, and then insert the extracted card in the correct place. This process is repeated until all the cards are in the correct sequence. Both average and worst-case time is O($n^2$).

Assuming there are n elements in the array, we must index through n – 1 entries. For each entry, we may need to examine and shift up to n – 1 other entries, resulting in a O($n^2$) algorithm.
The insertion sort is an in-place sort. That is, we sort the array in-place. No extra memory is required. The insertion sort is also a stable sort. Stable sorts retain the original ordering of keys when identical keys are present in the input data.

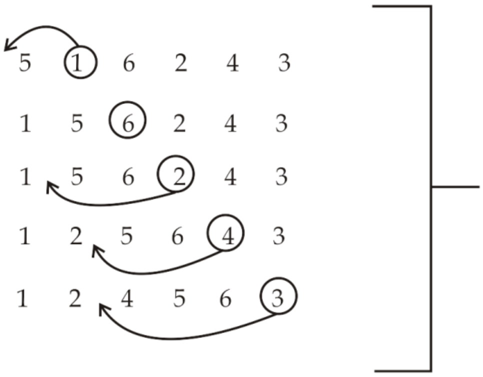
**Algorithm**
Step 1 − If it is the first element, it is already sorted. return 1;
Step 2 − Pick next element
Step 3 − Compare with all elements in the sorted sub-list
Step 4 − Shift all the elements in the sorted sub-list that is greater than the value to be sorted
Step 5 − Insert the value
Step 6 − Repeat until list is sorted

**INSERTION_SORT (A)**

```
1. FOR j ← 2 TO length[A]
2. DO key ← A[j]
3. {Put A[j] into the sorted sequence A[1 . . j − 1]}
4. i ← j − 1
5. WHILE i > 0 and A[i] > key
6. DO A[i +1] ← A[i]
7. i ← i − 1
8. A[i + 1] ← key
```

| | | | | | |
|---|---|---|---|---|---|
| 5 | 1 | 6 | 2 | 4 | 3 |

Lets take this Array




As we can see here, in insertion sort, we pick up a insertion sort, we pick up a key, and compares it with elemnts ahead of it, and puts the key in the right place

5 has nothing before it

1 is compared to 5 and is inserted before 5.

6 is greater than 5 and 1.

2 is smaller than 6 and 5, but greater than 1, so its is inserted after 1

Ant this goes on

**Time Complexity:** O(n*n)
**Auxiliary Space:** O(1)
**Boundary Cases:** Insertion sort takes maximum time to sort if elements are sorted in reverse order. And it takes minimum time (Order of n) when elements are already sorted.
**Algorithmic Paradigm:** Incremental Approach

## 7.2 Bubble Sort

Bubble Sort is the simplest sorting algorithm that works by repeatedly swapping the adjacent elements if they are in wrong order.

Bubble sort algorithms cycle through a list, analyzing pairs of elements from left to right, or beginning to end. If the leftmost element in the pair is less than the rightmost element, the pair will remain in that order. If the rightmost element is less than the leftmost element, then the two elements will be switched. This cycle repeats from beginning to end until a pass in which no switch occurs.

*Example:*
First Pass:
( 5 1 4 2 8 ) –> ( 1 5 4 2 8 ), Here, algorithm compares the first two elements, and swaps since 5 > 1.
( 1 5 4 2 8 ) –> ( 1 4 5 2 8 ), Swap since 5 > 4
( 1 4 5 2 8 ) –> ( 1 4 2 5 8 ), Swap since 5 > 2
( 1 4 2 5 8 ) –> ( 1 4 2 5 8 ), Now, since these elements are already in order (8 > 5), algorithm does not swap them.

Second Pass:
( 1 4 2 5 8 ) –> ( 1 4 2 5 8 )
( 1 4 2 5 8 ) –> ( 1 2 4 5 8 ), Swap since 4 > 2
( 1 2 4 5 8 ) –> ( 1 2 4 5 8 )
( 1 2 4 5 8 ) –> ( 1 2 4 5 8 )
Now, the array is already sorted, but our algorithm does not know if it is completed. The algorithm needs one whole pass without any swap to know it is sorted.

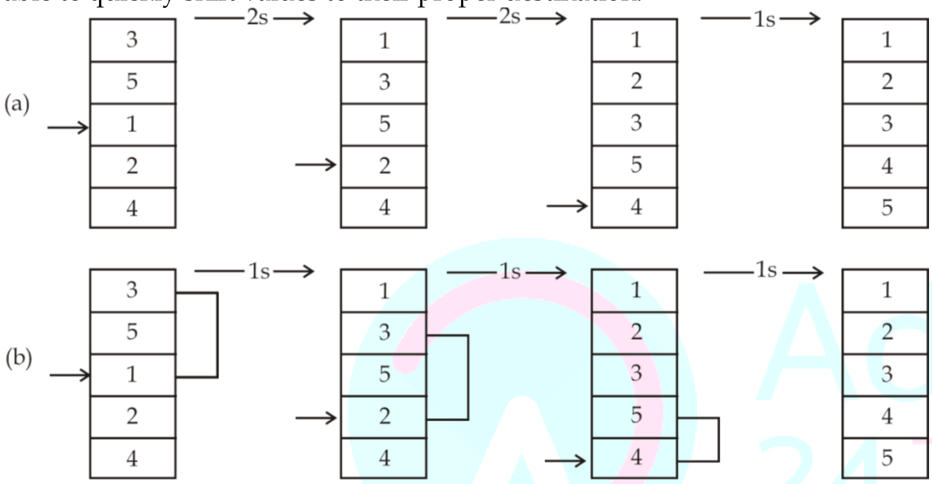
Third Pass:
( 1 2 4 5 8 ) –> ( 1 2 4 5 8 )
( 1 2 4 5 8 ) –> ( 1 2 4 5 8 )
( 1 2 4 5 8 ) –> ( 1 2 4 5 8 )
( 1 2 4 5 8 ) –> ( 1 2 4 5 8 )

**Worst and Average Case Time Complexity:** O(n*n). Worst case occurs when array is reverse sorted.
**Best Case Time Complexity:** O(n). Best case occurs when array is already sorted.
**Auxiliary Space:** O(1)
**Boundary Cases:** Bubble sort takes minimum time (Order of n) when elements are already sorted.

## 7.3 Shell Sort

Shell sort, developed by Donald L. Shell, is a non-stable in-place sort. Shell sort improves on the efficiency of insertion sort by quickly shifting values to their destination. Average sort time is O(n1.25), while worst-case time is O(n1.5).

We begin by doing an insertion sort using a spacing of two. In the first frame we examine numbers 3-1. Extracting 1, we shift 3 down one slot for a shift count of 1. Next, we examine numbers 5-2. We extract 2, shift 5 down, and then insert 2. After sorting with a spacing of two, a final pass is made with a spacing of one. This is simply the traditional insertion sort. The total shift count using shell sort is 1+1+1 = 3. By using an initial spacing larger than one, we were able to quickly shift values to their proper destination.




## 7.4 Merge Sort

Merge sort is a divide-and-conquer algorithm based on the idea of breaking down a list into several sub-lists until each sub-list consists of a single element and merging those sub-lists in a manner that results into a sorted list.

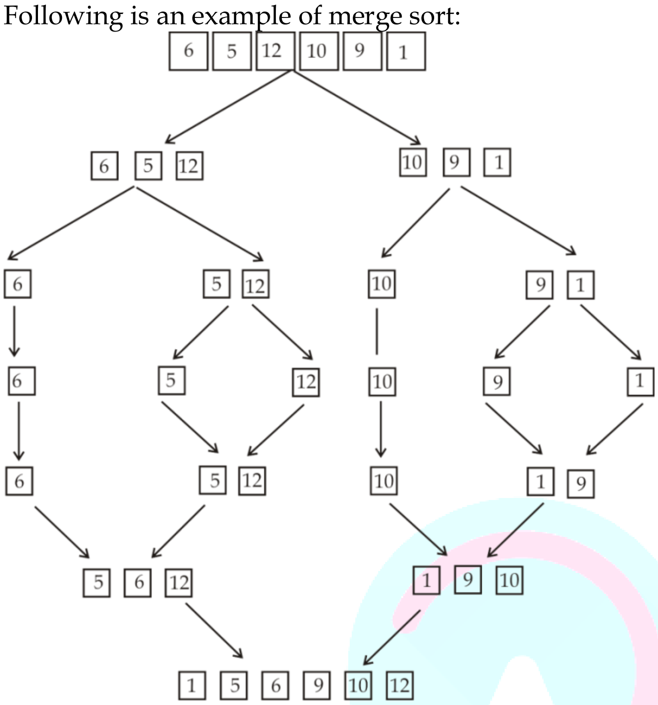
- Divide the unsorted list into N sub-lists, each containing 1 element.
- Take adjacent pairs of two singleton lists and merge them to form a list of 2 elements. N will now convert into N/2
- N/2 lists of size 2.
- Repeat the process till a single sorted list of obtained.

**MERGE-SORT(A,p,q,r)**

```
if p <r
then q ← (r + p)/2
MERGE-SORT(A, p, q )
MERGE-SORT(A,q+1,r)
MERGE(A, p, q, r)
n1←q-p+1
n2←r-q
create arrays L[1...N1+1] and R[1...N2+1]
for i←1 to N1
do L[i] ← A[p+i-1]
for j ← 1 to n2
do R[j] ← A[q+j]
L[N1+1] ← ∞
R[N2+1] ← ∞
i ← 1
j ← 1
```

```
for k ← p to r
do if L[i] ≤ R[j]
then A[k] ← L[i]
i ← i+1
else A[k] ← R[j]
j ← j+1
```

Following is an example of merge sort:




**Analysis of Merge sort**

Analyze the asymptotic complexity of merge sort:

```
void mergesort (int[] A, int lower, int upper)
//@requires 0 <= lower && lower <= upper && upper <= \length(A);
// modifies A;
//@ensures is_sorted(A, lower, upper);
{
if (upper-lower <= 1) return;
else {
int mid = lower + (upper-lower)/2;
mergesort(A, lower, mid); //@assert is_sorted(A, lower, mid);
mergesort(A, mid, upper); //@assert is_sorted(A, mid, upper);
merge(A, lower, mid, upper);
}
}
```

If we call mergesort with an range (upper − lower ) of size n, it makes recursive calls of size n/2, possibly rounding one or the other down. To avoid special cases and rounding, we just assume that the original range n is a power of 2.

| | Total work |
|---|---|
| *0(n) | 0(n) |
| *0(n/2) *0(n/2) | *0(n) |
| *0(n/4) *0(n/4) *0(n/4) *0(n/4) | *0(n) |
| * * * * * * * * | 0(n) |

When each pair of recursive calls returns, we have to perform a merge operation which takes O(m), where m is the size of resulting sorted range. For example, at the root node we need O(n) operations to perform the merge. At the next level of recursion, both left and right subrange are of size n/2, each requiring O(n/2) operations to merge, which means that the second level in the tree is also O(n). At the next level subranges have size
O(n/4), but there are 4 of them, so again processing this layer requires O(n) operations. Overall, there will be log(n) layers, each layer requiring a total amount of work of O(n), leading to an overall complexity of O(n ∗ log(n)).

**Time Complexity:** Sorting arrays on different machines. Merge Sort is a recursive algorithm and time complexity can be expressed as following recurrence relation.
$T(n) = 2T(n/2) + \Theta(n)$

Time complexity of Merge Sort is $\Theta$ (nLogn) in all 3 cases (worst, average and best) as merge sort always divides the array in two halves and take linear time to merge two halves.
**Auxiliary Space:** O(n)
**Algorithmic Paradigm:** Divide and Conquer

### 7.5 Quick Sort

Quick sort is based on the divide-and-conquer approach based on the idea of choosing one element as a pivot element and partitioning the array around it such that: Left side of pivot contains all the elements that are less than the pivot element Right side contains all elements greater than the pivot.

It reduces the space complexity and removes the use of the auxiliary array that is used in merge sort. Selecting a random pivot in an array results in an improved time complexity in most of the cases.

1. Pick an arbitrary element of the array (the pivot).
2. Divide the array into two subarrays, those that are smaller and those that are greater (the partition phase).
3. Recursively sort the subarrays.
4. Put the pivot in the middle, between the two sorted subarrays to obtain the final sorted array.

The pivot value divides the list into two parts. And recursively, we find the pivot for each sub-lists until all lists contains only one element.

**Quick Sort Pivot Algorithm**
Step 1 − Choose the highest index value has pivot
Step 2 − Take two variables to point left and right of the list excluding pivot
Step 3 − left points to the low index
Step 4 − right points to the high
Step 5 − while value at left is less than pivot move right
Step 6 − while value at right is greater than pivot move left
Step 7 − if both step 5 and step 6 does not match swap left and right
Step 8 − if left ≥ right, the point where they met is new pivot

```
function partitionFunc(left, right, pivot)
   leftPointer = left
   rightPointer = right - 1

   while True do
      while A[++leftPointer] < pivot do
         //do-nothing
      end while

      while rightPointer > 0 && A[--rightPointer] > pivot do
         //do-nothing
      end while

      if leftPointer >= rightPointer
```

```
            break
        else
            swap leftPointer,rightPointer
        end if

    end while

    swap leftPointer,right
    return leftPointer

end function
```

**Quick Sort Algorithm**
Step 1 − Make the right-most index value pivot
Step 2 − partition the array using pivot value
Step 3 − quicksort left partition recursively
Step 4 − quicksort right partition recursively

```
procedure quickSort(left, right)

    if right-left <= 0
        return
    else
        pivot = A[right]
        partition = partitionFunc(left, right, pivot)
        quickSort(left,partition-1)
        quickSort(partition+1,right)
    end if

end procedure
```

Complexity The worst case time complexity of this algorithm is $O(N^2)$ , but as this is randomized algorithm, its time complexity fluctuates between $O(N^2)$ and O(NlogN) and mostly it comes out to be O(NlogN).

## 7.6 Selection Sort

The selection sort algorithm sorts a list of values by successively putting values in their final, sorted positions.

Consider a list of numeric values that should be in ascending (increasing) order:
1. Scan the entire list and find the smallest value.
2. Exchange that value with the value in the first position of the list.
3. Scan the list, omitting the first value, and find the smallest value.
4. Exchange that value with the value in the second position of the list.
5. Scan the list, omitting the first two values, and find the smallest value.
6. Exchange that value with the value in the third position of the list.
7. Continue this process for all but the last position in the list, which will contain the largest value.

```
Procedure selection sort
    list : array of items;      n : size of list

    for i = 1 to n - 1
    /* set current element as minimum*/
        min = i

        /* check the element to be minimum */

        for j = i+1 to n
```

```
            if list[j] < list[min] then
                min = j;
            end if
        end for

        /* swap the minimum element with the current element*/
        if indexMin != i    then
            swap list[min] and list[i]
        end if

    end for

end procedure
```

**Time Complexity:** $O(n^2)$ as there are two nested loops.

## 7.7 Radix Sort

Radix sort is an integer sorting algorithm that sorts data with integer keys by grouping the keys by individual digits that share the same significant position and value (place value). Radix sort uses counting sort as a subroutine to sort an array of numbers. Because integers can be used to represent strings (by hashing the strings to integers), radix sort works on data types other than just integers.

**Example:**

| | | | |
|---|---|---|---|
| **327** | **470** | **418** | **146** |
| **476** | **382** | **327** | **173** |
| **285** | **173** | **146** | **259** |
| **418** | **285** | **259** | **285** |
| **568** | **476** | **568** | **327** |
| **382** | **146** | **470** | **382** |
| **146** | **327** | **173** | **418** |
| **259** | **418** | **476** | **470** |
| **173** | **568** | **382** | **476** |
| **470** | **259** | **285** | **568** |
| **Input** | **Sorted by LS digit** | **Sorted by Middle digit** | **Sorted by Most sig digit** |



**Algorithm radix-sort (a)**

```
Step 1:    for i = 0 to 9 do
Step 2:      empty-bin(i);
Step 3:      for position = least significant digit to most significant digit
Step 4:        for i = 1 to n do
Step 5:          x = digit in the position of a[i];
Step 6:            put a[i] in BIN x;
Step 7:              i =1;
Step 8:             for j = 0 to 9 do
Step 9:            while (BIN(j) <> empty) do
Step 10:     a[i] = get-element(BIN(j));
Step 11:         i = i + 1;
```

**Time Complexity comparison of Sorting Algorithms**

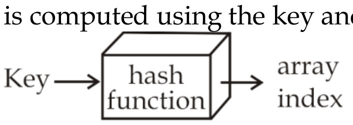
| **Algorithm** | **Time Complexity** | | |
|---|---|---|---|
| | **Best** | **Average** | **Worst** |
| Quick Sort | O(n log(n)) | O(n log(n)) | $O(n^2)$ |
| Merge Sort | O(n log(n)) | O(n log(n)) | O(n log(n)) |
| Heap Sort | O(n log(n)) | O(n log(n)) | O(n log(n)) |

| | | | |
|---|---|---|---|
| Bubble Sort | O(n) | $O(n^2)$ | $O(n^2)$ |
| Insertion Sort | O(n) | $O(n^2)$ | $O(n^2)$ |
| Select Sort | $O(n^2)$ | $O(n^2)$ | $O(n^2)$ |
| Bucket Sort | O(n+k) | O(n+k) | $O(n^2)$ |
| Radix Sort | O(nk) | O(nk) | O(nk) |

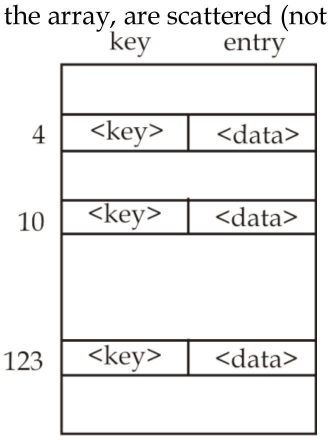
## 8. Hashing

Now-a-days it is impossible to find anything in the internet, unless we develop new data structures and algorithms for storing and accessing data. So, what is wrong with traditional data structures like Arrays and Linked Lists? Suppose we have a very large data set stored in an array. The amount of time required to look up an element in the array is either O(log n) or O( n) based on whether the array is sorted or not. If the array is sorted then a technique such as binary search can be used to search the array. Otherwise, the array must be searched linearly. Either case may not be desirable if we need to process a very large data set. Therefore we have a new technique called hashing that allows us to update and retrieve any entry in constant time O(1). The constant time or O(1) performance means, the amount of time to perform the operation does not depend on data size n.

**Hashing using Arrays**
When implementing a hash table using arrays, the nodes are not stored consecutively, instead the location of storage is computed using the key and a hash function. The computation of the array index can be visualized as shown below:




The value computed by applying the hash function to the key is often referred to as the hashed key. The entries into the array, are scattered (not necessarily sequential) as can be seen in figure below.




**Hashing is implemented in two steps:**

- An element is converted into an integer by using a hash function. This element can be used as an index to store the original element, which falls into the hash table.
- The element is stored in the hash table where it can be quickly retrieved using hashed key.
  hash = hashfunc(key)
  index = hash % array_size

In this method, the hash is independent of the array size and it is then reduced to an index (a number between 0 and array_size − 1) by using the modulo operator (%).

A hash function is any function that can be used to map a data set of an arbitrary size to a data set of a fixed size, which falls into the hash table. The values returned by a hash function are called hash values, hash codes, hash sums, or simply hashes. To achieve a good hashing mechanism, It is important to have a good hash function with the following basic requirements:

- Easy to compute: It should be easy to compute and must not become an algorithm in itself.
- Uniform distribution: It should provide a uniform distribution across the hash table and should not result in clustering.
- Less collisions: Collisions occur when pairs of elements are mapped to the same hash value. These should be avoided.

Note: Irrespective of how good a hash function is, collisions are bound to occur. Therefore, to maintain the performance of a hash table, it is important to manage collisions through various collision resolution techniques.

Our goal is to create a hash function that minimizes the number of collisions, is easy to compute, and evenly distributes the items in the hash table. There are a number of common ways to extend the simple remainder method. We will consider a few of them here.

**Folding method:**
The folding method for constructing hash functions begins by dividing the item into equal-size pieces (the last piece may not be of equal size). These pieces are then added together to give the resulting hash value. For example, if our item was a sequence of numbers 436-555-4601, we would take the digits and divide them into groups of 2 (43,65,55,46,01).

After the addition, 43+65+55+46+0143+65+55+46+01, we get 210.
If we assume our hash table has 11 slots, then we need to perform the extra step of dividing by 11 and keeping the remainder. In this case 210 % 11210 % 11 is 1, so the number sequence 436-555-4601 hashes to slot 1. Some folding methods go one step further and reverse every other piece before the addition. For the above example, we get 43+56+55+64+01=21943+56+55+64+01=219 which gives 219 % 11=10219 % 11=10.

**Mid-Square method:**
Another numerical technique for constructing a hash function is called the **mid-square method**. We first square the item, and then extract some portion of the resulting digits. For example, if the item were 44, we would first compute 442=1,936442=1,936. By extracting the middle two digits, 93, and performing the remainder step, we get 5 (93 % 1193 % 11).

**Collision Removal**
When two items hash to the same slot, we must have a systematic method for placing the second item in the hash table. This process is called collision resolution. One method for resolving collisions looks into the hash table and tries to find another open slot to hold the item that caused the collision. A simple way to do this is to start at the original hash value position and then move in a sequential manner through the slots until we encounter the first slot that is empty. We may need to go back to the first slot (circularly) to cover the entire hash table. This collision resolution process is referred to as **open addressing** in that it tries to find the next open slot or address in the hash table. By systematically visiting each slot one at a time, we are performing an open addressing technique called **linear probing**.

A disadvantage to linear probing is the tendency for clustering; items become clustered in the table. This means that if many collisions occur at the same hash value, a number of surrounding slots will be filled by the linear probing resolution. This will have an impact on other items that are being inserted. One way to deal with clustering is to extend the linear probing technique so that instead of looking sequentially for the next open slot, we skip slots, thereby more evenly distributing the items that have caused collisions. This will potentially reduce the clustering that occurs.

## PRACTICE SET

1. A binary search tree whose left sub tree and right sub tree differ in height by at most 1 unit is called ……
(a) AVL tree (b) Red-black tree
(c) Lemma tree (d) None of the above
(e) Both (a) and (b)
2. ……………….. level is where the model becomes compatible executable code
(a) Abstract level (b) Application level
(c) Implementation level
(d) All of the above (e) None of these
3. Stack is also called as
(a) Last in first out (b) First in last out
(c) Last in last out (d) First in first out
(e) None of these
4. Which of the following is true about the characteristics of abstract data types?
(i) It exports a type.
(ii) It exports a set of operations
(a) True, False (b) False, True
(c) True, True (d) False, False
(e) None of these
5. …………… is not the component of data structure.
(a) Operations (b) Storage Structures
(c) Algorithms (d) None of above
(e) Programming
6. Inserting an item into the stack when stack is not full is called …………. Operation and deletion of item form the stack, when stack is not empty is called ………..operation.
(a) Push, pop (b) Pop, push
(c) Insert, delete (d) Delete, insert
(e) None of these
7. ……………. Is a pile in which items are added at one end and removed from the other.
(a) Stack (b) Queue
(c) List (d) None of the above
(e) Zoom
8. ………… is very useful in situation when data have to stored and then retrieved in reverse order.
(a) Stack (b) Queue
(c) List (d) Link list
(e) None of these
9. Which data structure allows deleting data elements from and inserting at rear?
(a) Stacks (b) Queues
(c) Dequeues (d) Binary search tree
(e) None of these
10. Which of the following data structure can't store the non-homogeneous data elements?
(a) Arrays (b) Records
(c) Pointers (d) Stacks
(e) None of these
11. Which of the following is non-liner data structure?
(a) Stacks (b) List
(c) Strings (d) Trees
(e) None of these
12. To represent hierarchical relationship between elements, Which data structure is suitable?
(a) Dequeue (b) Priority
(c) Tree (d) Graph
(e) None of these
13. Which data structure is used in breadth first search of a graph to hold nodes?
(a) Stack (b) Queue
(c) Tree (d) Array
(e) None of these
14. In a stack, if a user tries to remove an element from empty stack it is called _________
(a) Underflow (b) Empty collection
(c) Overflow (d) Garbage Collection
(e) None of these
15. The disadvantage in using a circular linked list is ……………………
(a) It is possible to get into infinite loop.
(b) Last node points to first node.
(c) Time consuming
(d) Requires more memory space
(e) None of these
16. Deletion operation is done by using ……… in a queue.
(a) Front (b) Rear
(c) Top (d) List
(e) None of these
17. Heap can be used as…
(a) Priority queue (b) Stack
(c) A decreasing order array
(d) None of these (e) A increasing order array
18. Which of the following is true?
(a) A graph may contain no edges and many vertices
(b) A graph may contain many edges and no vertices
(c) A graph may contain no edges and no vertices
(d) None of these
(e) All of the above
19. Quick Sort can be categorized into which of the following?
(a) Brute Force technique (b) Divide and conquer
(c) Greedy algorithm
(d) Dynamic programming (e) None of these
20. What is the worst case complexity of Quick Sort?
(a) O(nlogn) (b) O(logn)
(c) O(n) (d) O(n2)
(e) None of these
21. The number of comparisons done by sequential search is ………………
(a) (N/2)+1 (b) (N+1)/2
(c) (N-1)/2 (d) (N+2)/2
(e) None of these
22. In ……………, search start at the beginning of the list and check every element in the list.
(a) Linear search (b) Binary search
(c) Hash Search (d) Binary Tree search

(e) None of these

23. State True or False.
(i) Binary search is used for searching in a sorted array.
(ii) The time complexity of binary search is O(log n).
(a) True, False (b) False, True
(c) False, False (d) True, True
(e) None of these

24. Which of the following is not the internal sort?
(a) Insertion Sort (b) Bubble Sort
(c) Merge Sort (d) Heap Sort
(e) None of these

25. State True or False.
(i) An undirected graph which contains no cycles is called forest.
(ii) A graph is said to be complete if there is an edge between every pair of vertices.
(a) True, True (b) False, True
(c) False, False (d) True, False
(e) None of these

26. What is the time complexity of inserting at the end in dynamic arrays?
(a) O(1) (b) O(n)
(c) O(log n) (d) Either O(1) or O(n)
(e) None of these

27. Linked lists are not suitable to for the implementation of?
(a) Insertion sort (b) Radix sort
(c) Polynomial manipulation
(d) Binary search (e) None of these

28. Linked list data structure offers considerable saving in
(a) Computational Time (b) Space Utilization
(c) Space Utilization and Computational Time
(d) None of these (e) Both (a) and (c)

29. Which of the following sorting algorithms can be used to sort a random linked list with minimum time complexity?
(a) Insertion Sort (b) Quick Sort
(c) Heap Sort (d) Merge Sort
(e) None of these

30. The time complexity of quick sort is …………..
(a) O(n) (b) O(n2)
(c) O(n log n) (d) O(log n)
(e) None of these

31. In a priority queue, insertion and deletion takes place at ………………
(a) Front, rear end (b) Only at rear end
(c) Only at front end (d) Any position
(e) None of these

32. While deleting nodes from a binary heap, ………. node is replaced by the last leaf in the tree.
(a) Left leaf (b) Right leaf
(c) Root (d) Cycle
(e) None of these

33. The worst case complexity of deleting any arbitrary node value element from heap is
(a) O(log n) (b) O(n)
(c) O(n log n) (d) O(n2)
(e) None of these

34. Which algorithmic technique does Fibonacci search use?
(a) Brute force (b) Divide and Conquer
(c) Greedy Technique (d) Backtracking
(e) None of these

35. Depth First Search is equivalent to which of the traversal in the Binary Trees?
(a) Pre-order Traversal (b) Post-order Traversal
(c) Level-order Traversal
(d) In-order Traversal (e) None of these

36. Time Complexity of DFS is? (V – number of vertices, E – number of edges)
(a) O(V + E) (b) O(V)
(c) O(E) (d) None of these
(e) Both (b) and (c)

37. The Data structure used in standard implementation of depth first search is?
(a) Stack (b) Queue
(c) Linked List (d) None of these
(e) Both (a) and (b)

38. The complexity of sorting algorithm measures the ..…… as a function of the number n of items to be sorter.
(a) Average time (b) Running time
(c) Average-case complexity
(d) Case-complexity (e) None of these

39. If the number of records to be sorted is small, then …… sorting can be efficient.
(a) Merge (b) Heap
(c) Selection (d) Bubble
(e) None of these

40. Partition and exchange sort is ……..
(a) Quick sort (b) Tree sort
(c) Heap sort (d)Bubble sort
(e) None of these

41. In ……….. ; for any node n, every descendant node's value in the left sub tree of n is less than the value of n and every descendant node's value in the right sub tree is greater than the value n.
(a) Binary tree (b) Binary search tree
(c) AVL tree (d) Binary heap tree
(e) None of these

42. In the best case of BST, the time is on the order of ………., but in the worst case it requires linear time.
(a) $\log_2 n$ (b) n
(c) $\log_2(n+1)$ (d) n+1
(e) None of these

43. ………………….. of binary search tree starts by visiting the current node, then its left child and then its right child.
(a) Preorder traversal (b) In-order traversal
(c) Linear traversal (d) Post-order traversal
(e) None of these

44. State True or False for internal sorting algorithms.

(i) Internal sorting are applied when the entire collection if data to be sorted is small enough that the sorting can take place within main memory.
(ii) The time required to read or write is considered to be significant in evaluating the performance of internal sorting.
(a) i-True, ii-True (b) i-True, ii-False
(c) i-False, ii-True (d) i-False, ii-False
(e) None of these

45. The height of a tree is the length of the longest root-to-leaf path in it. The maximum and minimum number of nodes in a binary tree of height 5 is.
(a) 63 and 6, respectively
(b) 64 and 5, respectively
(c) 32 and 6, respectively
(d) 31 and 5, respectively
(e) None of these

46. In a binary tree, the number of internal nodes of degree 1 is 5, and the number of internal nodes of degree 2 is 10. The number of leaf nodes in the binary tree is
(a) 10 (b) 11
(c) 12 (d) 15
(e) None of these

47. Breadth First Search (BFS) is started on a binary tree beginning from the root vertex. There is a vertex t at a distance four from the root. If t is the n-th vertex in this BFS traversal, then the maximum possible value of n is ________
(a) 15 (b) 16
(c) 31 (d) 32
(e) None of these

48. The maximum number of binary trees that can be formed with three unlabeled nodes is:
(a) 1 (b) 5
(c) 4 (d) 3
(e) None of these

49. When does top value of the stack changes?
(a) Before deletion
(b) While checking underflow
(c) At the time of deletion
(d) After deletion (e) None of these

50. ……………… is known as a greedy algorithm, because it chooses at each step the cheapest edge to add to sub graph S.
(a) Kruskal's algorithm (b) Prim's algorithm
(c) Dijkstra algorithm (d) Bellman ford algorithm
(e) None of these

51. The result of prim's algorithm is a total time bound of ………………
(a) O(log n) (b) O(m + n log n)
(c) O(mn) (d) O(m log n)
(e) None of these

52. Which function plays an important role in returning the address of memory block allocated to locate / store a node especially while declaring ' top ' in the linked representation of the Stack?
- Published on 28 Aug 15
(a) galloc () (b) falloc ()
(c) malloc () (d) calloc ()
(e) None of these

53. Which of the following statements is/are TRUE for an undirected graph?
(i) Number of odd degree vertices is even
(ii) Sum of degrees of all vertices is even
(a) Only I (b) Only ii
(c) Both i and ii (d) Neither i nor ii
(e)None of these

54. The Average case occurs in linear search algorithm ……….
(a) When item is somewhere in the middle of the array
(b) When item is not the array at all
(c) When item is the last element in the array
(d) Item is the last element in the array or item is not there at all
(e) None of these

55. Which of the following is not the required condition for binary search algorithm?
(a) The list must be sorted
(b) There should be the direct access to the middle element in any sub list
(c) There must be mechanism to delete and/or insert elements in list.
(d) Both (b) and (a)
(e) None of these

56. For a linear search in an array of n elements the time complexity for best, worst and average case are ……. , …….. and …….. respectively.
(a) O(n), O(1), and O(n/2)
(b) O(1), O(n) and O(n/2) (c) O(1),O(n) and O(n)
(d) O(1), O(n) and (n-1/2) (e) None of these

57. Which of the following sorting methods will be the best if number of swapping done, is the only measure of efficiently?
(a) Bubble sort (b) Selection sort
(c) Insertion sort (d) Quick sort
(e) None of these

58. The maximum number of comparisons needed to sort 7 items using radix sort is (assume each item is 4 digit decimal number)
(a) 280 (b) 40
(c) 47 (d) 38
(e) None of these

59. Which of the following statements are correct?
I. If each tree node contains a father field, then it's not necessary to use either stack or threads
II. Traversal using father pointers is more time efficient than traversal of a threaded tree
III. A in-threaded binary tree is defined as binary tree that is both left-in threaded and right-in threaded.
(a) II and III (b) I and III
(c) I and II (d) None of these
(e) Both (a) and (b)

60. If each node in a tree has value greater the every value in its left sub tree and has value less than every value in its right sub tree, the tree is called
(a) Complete tree (b) Full binary tree
(c) Binary search tree (d) AVL tree
(e) None of these

## SOLUTIONS

1. (a)
2. (c)
3. (a)
4. (c)
5. (d)
6. (a) push, pop
7. (b) Queue
8. (a)
9. (b) Queues
10. (a)
11. (d)
12. (c)
13. (b)
14. (a)
15. (a) It is possible to get into infinite loop.
16. (a)
17. (a); The property of heap that the value of root must be either greater or less than both of its children makes it work like a priority queue.
18. (b)
19. (b)
20. (d)
21. (b)
22. (a)
23. (d)
24. (c)
25. (a)
26. (d)
27. (d)
28. (c)
29. (d)
30. (c)
31. (d)
32. (c)
33. (a)
34. (b)
35. (a); In Depth First Search, we explore all the nodes aggressively to one path and then backtrack to the node. Hence, it is equivalent to the pre-order traversal of a Binary Tree.
36. (a); The Depth First Search explores every node once and every edge once (in worst case), so it's time complexity is O(V + E).
37. (a); The Depth First Search is implemented using recursion. So, stack can be used as data structure to implement depth first search.
38. (b)
39. (c)
40. (a)
41. (b)
42. (a)
43. (a); Preorder traversal
44. (b)
45. (a); Number of nodes is maximum for a perfect binary tree.
A perfect binary tree of height h has $2^{h+1}$ - 1 nodes
Number of nodes is minimum for a skewed binary tree.
A perfect binary tree of height h has h+ 1 node.
46. (b); In a binary tree, the number of leaf nodes is always 1 more than number of internal nodes with 2 children
So,
Number of Leaf Nodes = Number of Internal nodes with 2 children + 1
Number of Leaf Nodes = 10 + 1
Number of Leaf Nodes = 11
47. (c); t is the nth vertex in this BFS traversal at distance four from the root. So height of tree is 4.
Max number of nodes = $2^{h+1} - 1 = 31$
At distance four, last node is 31.
48. (b)
49. (d)
50. (b)
51. (a)
52. (c)
53. (c)
54. (a)
55. (c)
56. (c)
57. (b); Because in selection sort algorithm we randomly access data rather than a list in which we can easily swap data which we want to swap and it takes less time. So, answer is option 'b'
58. (a); The maximum number of comparison is number of items ´ radix ´ number of digits i.e., 7 x 10 x 4 = 280.
59. (b)
60. (c)

# 10 Software Engineering

## 1. Introduction

Software engineering is a systematic and disciplined approach to developing software. It applies both computer science and engineering principles and practices to the creation, operation, and maintenance of software systems.

**Significance of Software Engineering**

- software systems are the most complex artefacts ever constructed by humans
- software is developed for clients who do not understand how software works
- software developers often do not understand the application domain of the software
- many software projects are never completed, and many software systems are never used after being delivered
- software increasingly pervades our daily lives
- bad software endangers lives and property

**What is Software Development Life Cycle?**

Software development life cycle is a step-by-step process involved in the development of a software product. It is also denoted as software development process in certain parts of the world. The whole process is generally classified into a set of steps and a specific operation will be carried out in each of the steps.

**Classification**

The basic classification of the whole process is as follows

- Planning
- Analysis
- Design
- Development
- Implementation
- Testing
- Deployment
- Maintenance

Each of the steps of the process has its own importance and plays a significant part in the product development. The description of each of the steps can give a better understanding.

**Planning**

This is the first and foremost stage in the development and one of the most important stages. The basic motive is to plan the total project and to estimate the merits and demerits of the project. The planning phase includes the definition of the intended system, development of the project plan, and parallel management of the plan throughout the proceedings. A good and matured plan can create a very good initiative and can positively affect the complete project.

The main aim of the analysis phase is to perform statistics and requirements gathering. Based on the analysis of the project and due to the influence of the results of the planning phase, the requirements for the project are decided and gathered. Once the requirements for the project are gathered, they are prioritized and made ready for further use. The decisions taken in the analysis phase are out-and-out due to the requirements analysis. Proceedings after the current phase are defined.

**Design**

Once the analysis is over, the design phase begins. The aim is to create the architecture of the total system. This is one of the important stages of the process and serves to be a benchmark stage since the errors performed until this stage and during this stage can be cleared here. Most of the developers have the habit of developing a prototype of the entire software and represent the software as a miniature model. The flaws, both technical and design, can be found and removed and the entire process can be redesigned.

**Development and Implementation**
The development and implementation phase is the most important phase since it is the phase where the main part of the project is done. The basic works include the design of the basic technical architecture and the maintenance of the database records and programs related to the development process.
One of the main scenarios is the implementation of the prototype model into a full-fledged working environment, which is the final product or software.

**Testing**
The testing phase is one of the final stages of the development process and this is the phase where the final adjustments are made before presenting the completely developed software to the end-user.
In general, the testers encounter the problem of removing the logical errors and bugs. The test conditions which are decided in the analysis phase are applied to the system and if the output obtained is equal to the intended output, it means that the software is ready to be provided to the user.

**Maintenance**
The toughest job is encountered in the maintenance phase which normally accounts for the highest amount of money. The maintenance team is decided such that they monitor on the change in organization of the software and report to the developers, in case a need arises.

## 2. Software Process Models

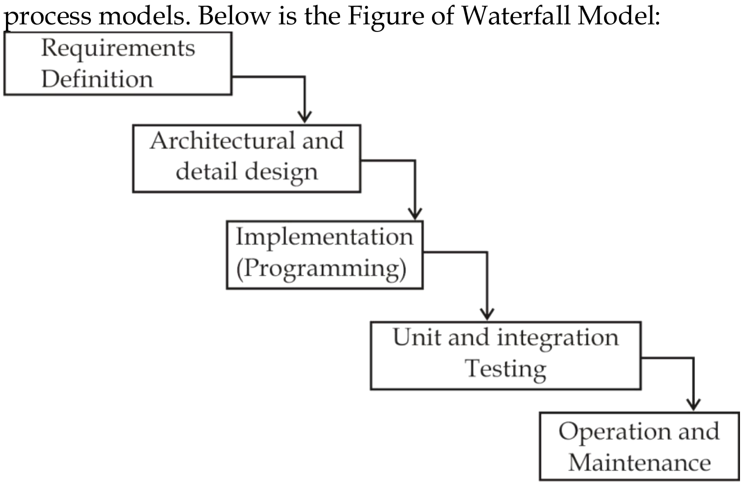
A software life cycle model defines entry and exit criteria for every phase. A phase can start only if its phase-entry criteria have been satisfied. So, without software life cycle model the entry and exit criteria for a phase cannot be recognized.

**Different software life cycle models**
Many life cycle models have been proposed so far. Each of them has some advantages as well as some disadvantages. A few important and commonly used life cycle models are as follows:

- Waterfall Model
- Prototyping Model
- Evolutionary Model
- Spiral Model

### 2.1 The Waterfall Model

As the original software process model, it can be viewed as a "first approximation" of the activities needed in the software development process. Some versions provide feedback loops from each stage to the previous ones, but it is most used as a simple linear model as shown in Figure. In reality, it is only suitable for projects in which all the customer requirements are known at the outset of the project, a rare condition, even for small scale projects. Also note that a working version of the system is only available late in the project, a problem addressed by the incremental process models. Below is the Figure of Waterfall Model:



## 2.2 Evolutionary Process Models

These models are designed to grow the final software system by iterative cumulative development. For example, **Rapid Prototyping** is sometimes used at the beginning of projects to obtain improved understanding of the customer requirements. The prototype can either be a throw-away or can be extended to the development of the entire system.

**Prototype** is a working model of software with some limited functionality. The prototype does not always hold the exact logic used in the actual software application and is an extra effort to be considered under effort estimation.

Prototyping is used to allow the users evaluate developer proposals and try them out before implementation. It also helps understand the requirements which are user specific and may not have been considered by the developer during product design.

**Advantages of prototyping models**

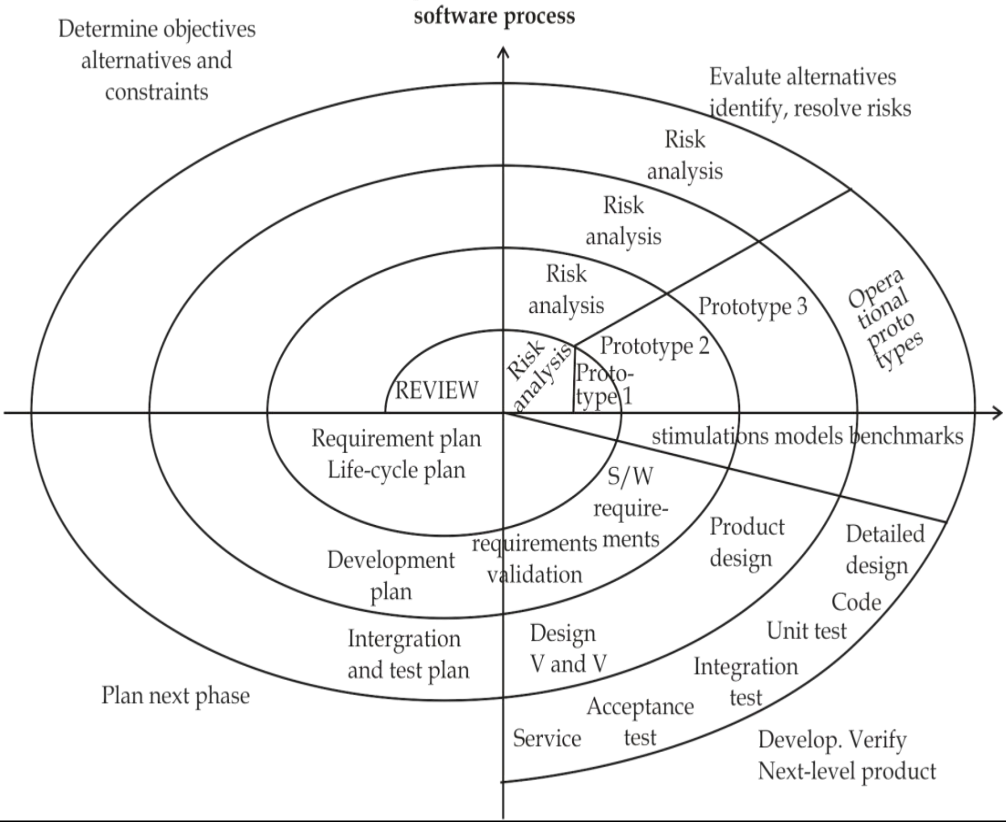
- Important functionalities can be considered separately.
- Requirements can be "discovered" at the prompting of
- an iteration.
- Customers & managers feel a sense of progress.
- Prototypes are often used to win customer contracts (especially in bidding).

**Challenges of Prototyping**

- Identifying the subset of requirements for the next iteration can be difficult.
- Establishing consistency is a repetitive work, particularly when a new subset of requirements bears
- no relationships with the existing ones
- Moreover, it is a tedious and time-consuming work
- The project deadline is difficult to predict.
- No visible end to the set of iterations
- Poses "maintenance" problems

**The Spiral Model** combines elements of the waterfall model and rapid prototyping to implement evolutionary development as shown in Figure 2 below. Each traversal around the spiral, beginning with Objectives, represents a new more complete version of the system with a risk assessment each time around. Each version can be viewed as a system prototype during any phase of the evolutionary development. For example, one of the spiral traversals might represent the system design and another might focus on integration testing.

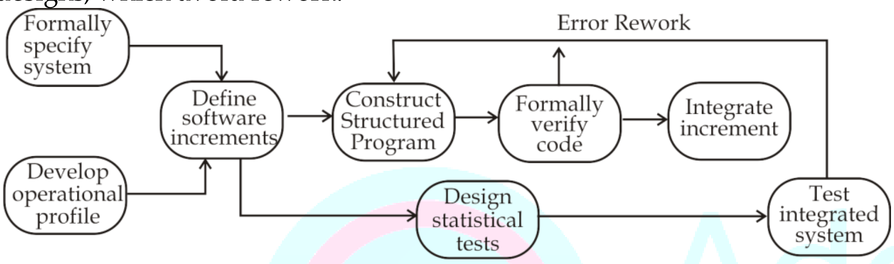
**Spiral Model of the software process**



In the **Concurrent Development Model,** all activities in the process model exist concurrently in various states such as "awaiting changes", "none" and "under development". This model describes the software process to be implemented as a network of activities rather than a simple linear process. **Component-Based Development** uses off-the-shelf software packages as the defining characteristic of the process. This is a software reuse approach and often leads to large reductions in cost and development time but depends on the nature and quality of the components library available.

## 3. Clean Room Software Engineering

Clean room software engineering (CSE) is a process model that removes defects before they can precipitate serious hazards. It is a team-oriented, theory based software, which is developed using the formal methods, correctness verification and Statistical Quality Assurance (SQA). Clean room management is based on the incremental model of software development, which accumulates into the final product. The approach combines mathematical-based methods of software specification, design and correctness verification with statistical, usage-based testing to certify software fitness for use. The main goal of clean room engineering is to produce zero error-based software by allowing correct designs, which avoid rework.




**Clean Room Management process:**

1. Project Planning: The main objective of this process is to ensure the software engineering processes are tailored for the project and they are well documented. This also ensures that project plan is well documented and defined. Another important objective is to ensure that the customer, the project team and the peer groups agree to the review.

2. Project Management Process: The main objective of the project management process is to manage the project which is implemented using cleanroom to deliver the software on schedule and within the budget. To establish and train Cleanroom teams, initiate tracking, control planned Cleanroom processes, eliminate or reduce risks, revise plans as necessary to accommodate changes and actual results, and continually improve Cleanroom team performance.

3. Performance Improvement Process: The main objectives of the performance improvement process is to continually evaluate and improve team performance by causal analysis of deviations from plans and faults found through the Correctness Verification and the Statistical Testing and Certification processes and evaluation and introduction of appropriate new technologies and processes.

**Aspect-Oriented Software Development (AOSD)**

AOSD is a new concept intended to model localized features, functions and information content that have impact across the software system. It is sometimes called crosscutting concerns that go beyond mechanisms such as subroutines.

Aspect-Oriented Programming was built as a response to limitations of Object-Oriented Programming (OOP). Aspect-Oriented Programming (AOP) helps to provide a solution to the previous problem. Instead of objects, AOP deals with aspects. An aspect is a behavior that cuts through multiple objects.

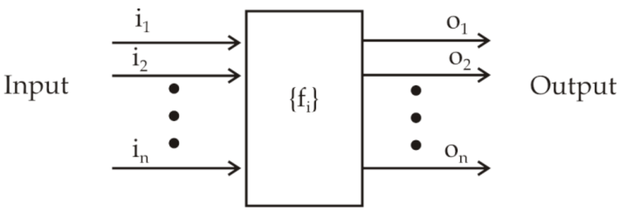
One of the drawbacks to OOP is that certain things will crosscut objects and methods. That is, a single behavior will apply to multiple classes. An example is logging or monitoring; you would want to monitor/log updates to several classes in a financial program. In OOP, you have to add the logging to each class or method. AOP avoids this by treating this crosscutting behavior as an aspect, as opposed to an object.

**Agile Development**
A somewhat controversial approach to the software process is to compress and overlap the traditional life cycle phases as much as possible with close customer partnership. The overriding objective is rapid time to delivery. "Extreme Programming" is also a term that is associated with this strategy. Some have said that agile development models are appropriate for web applications developers, a group that has tended to resist the discipline of software processes.

## 4. Requirement Analysis and Modelling

**Functional requirements:** The functional requirements part discusses the functionalities required from the system. The system is considered to perform a set of high level functions {fi}.
Each function fi of the system can be considered as a transformation of a set of input data (ii) to the corresponding set of output data (oi). The user can get some meaningful piece of work done
using a high-level function.




**Non-functional requirements:** Non-functional requirements deal with the characteristics of the system which cannot be expressed as functions - such as the maintainability of the system, portability of the system, usability of the system, etc.

Non-functional requirements may include:
- reliability issues,
- accuracy of results,
- human - computer interface issues,
- constraints on the system implementation, etc.

**Software Process Activities**

**Requirements Engineering**
Requirements engineering is about communicating with the customer. The objective is to arrive at a written agreement describing the functionality of the software to be developed. The final product of this activity is usually called a specification and forms the basis for all the development activities that follow. It turns out that the activities of this stage of the development life cycle are among the most difficult, but perhaps the most important.
The problem of developing a clear understanding of what the customer needs has become a classic challenge. Newcomers to the field are sometimes surprised to learn that the customer representatives themselves find it difficult to communicate their needs clearly. Given this state of affairs, projects frequently find themselves changing requirements throughout the development period. The later in the development process that a requirement is changed, the more difficult and more expensive that change becomes. Therefore, what may seem to be excessive time spent on this first phase is in fact an excellent investment in risk reduction for the entire project. Many tools are used at this stage to increase confidence in understanding what the system should be like, including rapid prototyping, user scenarios and functions/feature lists. Bringing modelling and design tools into this phase is not unusual. Not surprisingly, this phase is characterized by lengthy meetings between the developer and customer.
The focus of this activity is on defining the operational characteristics of the software and has three primary goals of providing the following:
- behavioral description of customer requirements with the emphasis on the *what* rather than the *how*
- foundation for the software design
- operational system definition that can be used for system validation after the software development is complete.

The following sub activities are identified with requirements analysis and modelling:
**Domain analysis**
Reusability is an important goal in software development since it reduces development costs, increases reliability and reduces development time. Domain analysis is the process of identifying patterns that can be reused. These patterns can be any common functions or features that have the potential for broad use across an application domain.

An application domain is typically a class of problem such as financial, medical or aerospace; however, the broader the reuse the better.

**Data modelling**
Analysis modelling sometimes begins with the identification of all data objects that are to be processed in the system and the relationships between these objects. Data modelling is used for large database and information systems applications.

**Object oriented analysis**
The object-oriented (OO) approach to analysis represents the latest "paradigm shift" in analysis methodology and is epitomized by the Java language at the implementation stage. Some of the claims which have made this approach popular are as follows:

- Customers can understand OO models with no programming knowledge thus facilitating the all-important early phases of communication.
- OO languages promote code reuse and thus programmer productivity
- The OO design and analysis methods are accommodating to change

The OO approach is based on modelling of the problem domain using classes and objects.
**Class:** defines the data and procedural abstractions for the information content and behaviour of some system entity.
**Method:** representation of one of the behaviours of a class.
**Object:** instance of a specific class. Objects can inherit the attributes and operations defined for a class. Classes are sometimes illustrated as "cookie cutters" and the associated objects as "cookies".

The goal of object-oriented analysis is the design of all classes and associated methods that are appropriate for the system being developed. The unified modelling language (UML) has been developed for the modelling and development of object-oriented (OO) systems. UML has become an industry standard for OO development.

**Scenario-based modelling**
End-user involvement in a software project is critical to its success. Scenario-based modelling provides mechanisms for capturing information on how end-users desire to interact with the system. UML provides support for the development of interaction scenarios that begin with the writing of use-cases that describe a use of the system by a specific end-user. The dynamics of these use-cases can be represented in UML activity diagrams similar to flow charts. More complex interactions can be captured in UML swimlane diagrams that can model concurrent activities.

**Flow oriented modelling**
Although not part of UML, the input-process-output data flow diagrams (DFD) continue to be a very popular analysis modelling tool and can be used to augment UML diagrams. Data flow in the system can be modelled in a hierarchical fashion with DFDs with higher level context diagrams being refined with greater detailed DFDs at lower levels.

**Dynamic modelling**
After static data and attribute relationships have been established, it is useful to create behavioural models to represent the systems response to external events. Use-cases can be used to identify events and UML sequence diagrams can be used to model how events trigger transitions from one object to another.

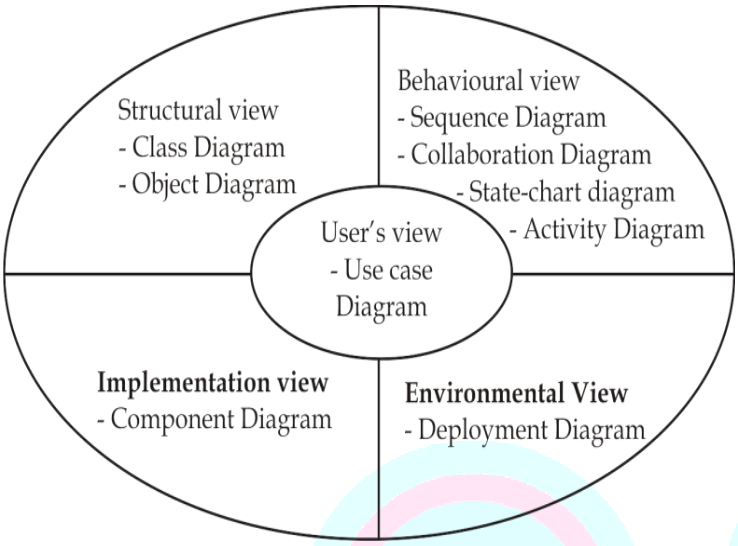
## 5. Unified Modelling Language (UML)

UML, as the name implies, is a modelling language. It may be used to visualize, specify, construct, and document the artefacts of a software system. It provides a set of notations (e.g. rectangles, lines, ellipses, etc.) to create a visual model of the system. Like any other language, UML has its own syntax (symbols and sentence formation rules) and semantics (meanings of symbols and sentences).
Also, we should clearly understand that UML is not a system design or development methodology, but can be used to document object-oriented and analysis results obtained using some methodology.

**UML diagrams**
UML can be used to construct nine different types of diagrams to capture five different views of a system. Just as a building can be modelled from several views (or perspectives) such as ventilation perspective, electrical perspective,

lighting perspective, heating perspective, etc.; the different UML diagrams provide different perspectives of the software system to be developed and facilitate a comprehensive understanding of the system. Such models can be refined to get the actual implementation of the system.
The UML diagrams can capture the following five views of a system:

- User's view
- Structural view
- Behavioural view
- Implementation view
- Environmental view




Diagrams and views in UML

**User's view:** This view defines the functionalities (facilities) made available by the system to its users. The users' view captures the external users' view of the system in terms of the functionalities offered by the system. The users' view is a black-box view of the system where the internal structure, the dynamic behaviour of different system components, the implementation etc. are not visible.

**Structural view:** The structural view defines the kinds of objects (classes) important to the understanding of the working of a system and to its implementation. It also captures the relationships among the classes (objects). The structural model is also called the static model, since the structure of a system does not change with time.

**Behavioural view:** The behavioural view captures how objects interact with each other to realize the system behaviour. The system behaviour captures the time-dependent (dynamic) behaviour of the system.

**Implementation view:** This view captures the important components of the system and their dependencies.

**Environmental view:** This view models how the different components are implemented on different pieces of hardware.

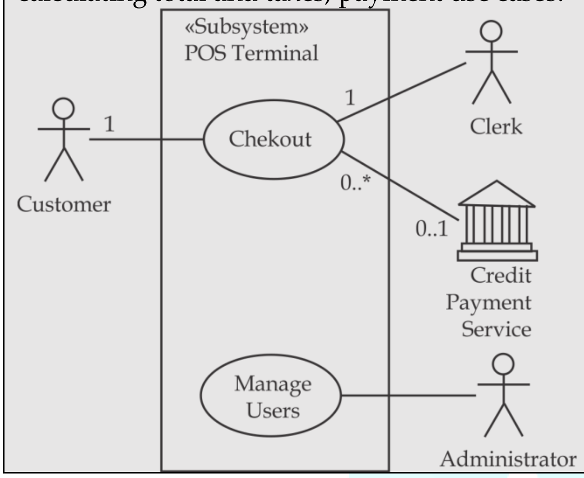
**Use Case Model**
The purpose of a use case is to define a piece of coherent behaviour without revealing the internal structure of the system. The use cases do not mention any specific algorithm to be used or the internal data representation, internal structure of the software, etc. A use case typically represents a sequence of interactions between the user and the system.

**Representation**

Use cases can be represented by drawing a use case diagram and writing an accompanying text elaborating the drawing. In the use case diagram, each use case is represented by an ellipse with the name of the use case written inside the ellipse. All the ellipses (i.e. use cases) of a system are enclosed within a rectangle which represents the system boundary. The name of the system being modelled (such as Library Information System) appears inside the rectangle. The different users of the system are represented by using the stick person icon. Each stick person icon is normally referred to as an actor. An actor is a role played by a user with respect to the system use. It is possible that the same user may play the role of multiple actors. Each actor can participate in one or more use cases. The line connecting the actor and the use case is called the communication relationship. It indicates that the actor makes use of the functionality provided by the use case. Both the human users and the external systems can be represented by stick person icons.

**An example of UML use case diagram for Point of Sale (POS) Terminal or Checkout.**
A retail POS system typically includes a computer, monitor, keyboard, barcode scanners, weight scale, receipt printer, credit card processing system, etc. and POS terminal software.

Checkout use case involves Customer, Clerk and Credit Payment Service actors and includes scanning items, calculating total and taxes, payment use cases.




## 6. User Interface Design

A common failure of software projects is to spend too little time communicating with the user. It is easy for software experts to fall into the subconscious trap of "knowing what is good for the user". What may seem to be "clearly good for the user" is all too frequently not the case from the perspective of the user herself. The use of user scenarios and very early and iterative prototype screen designs can help to assure that the user is being understood. It has been said that you should plan on building one to throw away. Three good guidelines are the following:

- Put the user in control
- Reduce the user's memory load
- Make the interface consistent

**Types of user interfaces**
User interfaces can be classified into the following three categories:
- Command language based interfaces
- Menu-based interfaces
- Direct manipulation interfaces

**Command Language-based Interface:** A command language-based interface – as the name itself suggests, is based on designing a command language which the user can use to issue the commands. The user is expected to frame the appropriate commands in the language and type them in appropriately whenever required. A command language-based interface can be made concise requiring minimal typing by the user. Command language-based interfaces allow fast interaction with the computer and simplify the input of complex commands.

**Menu-based Interface:** An important advantage of a menu-based interface over a command language-based interface is that a menu-based interface does not require the users to remember the exact syntax of the commands. A menu-based interface is based on recognition of the command names, rather than recollection. Further, in a menu-based interface the typing effort is minimal as most interactions are carried out through menu selections using a pointing device.

**Direct Manipulation Interfaces:** Direct manipulation interfaces present the interface to the user in the form of visual models (i.e. icons or objects). For this reason, direct manipulation interfaces are sometimes called as iconic interface. In this type of interface, the user issues commands by performing actions on the visual representations of the objects, e.g. pull an icon representing a file into an icon representing a trash box, for deleting the file. However, direct manipulation interfaces can be considered slow for experienced users. Also, it is difficult to give complex commands using a direct manipulation interface.

## 7. Software Testing

After the software system is coded into a deliverable product, testing strategies are used to validate system requirements. Testing strategies are designed to detect errors in the system.

There are many types of testing like:

- Unit Testing
- Integration Testing
- Functional Testing
- System Testing
- Stress Testing
- Performance Testing
- Usability Testing
- Acceptance Testing
- Regression Testing
- Beta Testing

Debugging is the process of finding the source of the errors for correction. Exhaustive testing is impractical. Therefore, no matter how much testing is done, it is never known with certainty if all bugs have been detected. Since testing is a process of detecting the presence of errors, the absence of all errors cannot be guaranteed by the testing process. A high percentage of project resources are expended on the testing phase. Testing usually proceeds in two phases, first at the component level sometimes called unit testing. Unit testing is followed by integration testing in which increasingly larger groups of components are tested culminating in the total system. Unit testing is usually done by the developer and integration testing by an independent test group. Testing strategies for conventionally designed software differ somewhat from those for object-oriented systems.

**Conventional software**: unit testing focuses on execution paths through component program logic with the goal of maximizing error detection by path coverage; whereas integration testing usually involves input and output values.

**Object-oriented software**: unit testing is done with classes, whose definition involves not only internal program logic but also attributes and operations as well as communication and collaboration. Operations must be tested in the context of a class. Two approaches to integration testing of object-oriented systems are common, thread-based and use-based testing. The thread-based approach tests the set of classes that respond to a given system input or event. Use-based testing begins with by testing classes that are relatively independent of all others and continues in stages with each stage defined by the addition of a layer of dependent classes until the entire system is encompassed.
After unit and integration testing, the entire system is tested in accordance with customer requirements. This final testing phase is usually called validation testing and includes alpha and beta tests. Alpha tests are performed at the developer site and beta tests occur later at end user sites. Final release of the software is scheduled after the beta tests are complete.

### 7.1 Clean room testing

Clean room testing was pioneered by IBM. This type of testing relies heavily on walk throughs, inspection, and formal verification. The programmers are not allowed to test any of their code by executing the code other than doing some syntax testing using a compiler. The software development philosophy is based on avoiding software defects by using a rigorous inspection process. The objective of this software is zero-defect software. The name 'clean room' was derived from the analogy with semi-conductor fabrication units. In these units (clean rooms), defects are avoided by manufacturing in ultra-clean atmosphere. In this kind of development, inspections to check the consistency of the components with their specifications has replaced unit-testing.
This technique reportedly produces documentation and code that is more reliable and maintainable than other development methods relying heavily on code execution-based testing.

The clean room approach to software development is based on five characteristics:

- **Formal specification:** The software to be developed is formally specified. A state-transition model which shows system responses to stimuli is used to express the specification.
- **Incremental development:** The software is partitioned into increments which are developed and validated separately using the clean room process. These increments are specified, with customer input, at an early stage in the process.
- **Structured programming:** Only a limited number of control and data abstraction constructs are used. The program development process is process of stepwise refinement of the specification.
- Static verification: The developed software is statically verified using rigorous software inspections. There is no unit or module testing process for code components.
- **Statistical testing of the system:** The integrated software increment is tested statistically to determine its reliability. These statistical tests are based on the operational profile which is developed in parallel with the system specification.

The main problem with this approach is that testing effort is increased as walk throughs, inspection, and verification are time consuming.

## 7.2 Unit Testing

Unit testing is undertaken after a module has been coded and successfully reviewed. Unit testing (or module testing) is the testing of different units (or modules) of a system in isolation. In order to test a single module, a complete environment is needed to provide all that is necessary for execution of the module. That is, besides the module under test itself, the following steps are needed in order to be able to test the module:

- The procedures belonging to other modules that the module under test calls.
- Nonlocal data structures that the module accesses.
- A procedure to call the functions of the module under test with appropriate parameters.

Modules required to provide the necessary environment (which either call or are called by the module under test) is usually not available until they too have been unit tested, stubs and drivers are designed to provide the complete environment for a module.
A **stub** procedure is a dummy procedure that has the same I/O parameters as the given procedure but has a highly simplified behaviour. A driver module contains the nonlocal data structures accessed by the module under test, and would also have the code to call the different functions of the module with appropriate parameter values.

## 7.3 Black-Box Testing

This testing methodology looks at what are the available inputs for an application and what the expected outputs are that should result from each input. It is not concerned with the inner workings of the application, the process that the application undertakes to achieve a particular output or any other internal aspect of the application that may be involved in the transformation of an input into an output. Most black-box testing tools employ either coordinate based interaction with the applications graphical user interface (GUI) or image recognition. Black box testing tends to find different kinds of errors than white box testing- Missing functions, Usability problems, Performance problems, Concurrency and timing errors, Initialization and termination errors etc. Unlike white box testing, black box testing tends to be applied later in the development process.

## 7.4 White-box Testing

This testing methodology looks under the covers and into the subsystem of an application. Whereas black-box testing concerns itself exclusively with the inputs and outputs of an application, white-box testing enables you to see what is happening inside the application. White-box testing provides a degree of sophistication that is not available with black-box testing as the tester is able to refer to and interact with the objects that comprise an application rather than only having access to the user interface. An example of a white-box system would be in-circuit testing where someone is looking at the interconnections between each component and verifying that each internal connection is working properly. Another example from a different field might be an auto-mechanic who looks at the inner-workings of a car to ensure that all of the individual parts are working correctly to ensure the car drives properly.

The main difference between black-box and white-box testing is the areas on which they choose to focus. In simplest terms, black-box testing is focused on results. If an action is taken and it produces the desired result then the process that was actually used to achieve that outcome is irrelevant. White-box testing, on the other hand, is concerned with the details. It focuses on the internal workings of a system and only when all avenues have been tested and the sum of an application's parts can be shown to be contributing to the whole is testing complete.

### 7.5 Integration testing

The primary objective of integration testing is to test the module interfaces, i.e. there are no errors in the parameter passing, when one module invokes another module. During integration testing, different modules of a system are integrated in a planned manner using an integration plan. The integration plan specifies the steps and the order in which modules are combined to realize the full system.

After each integration step, the partially integrated system is tested. An important factor that guides the integration plan is the module dependency graph. The structure chart (or module dependency graph) denotes the order in which different modules call each other. By examining the structure chart the integration plan can be developed.

Integration test approaches There are four types of integration testing approaches. Any one (or a mixture) of the following approaches can be used to develop the integration test plan. Those approaches are the following:
- Big bang approach
- Top-down approach
- Bottom-up approach
- Mixed-approach

### 7.6 System testing

System tests are designed to validate a fully developed system to assure that it meets its requirements. There are essentially three main kinds of system testing:
- **Alpha Testing:** refers to the system testing carried out by the test team within the developing organization.
- **Beta testing:** is the system testing performed by a select group of friendly customers.
- **Acceptance Testing:** is the system testing performed by the customer to determine whether he should accept the delivery of the system.

## 8. Debugging

Once errors are identified in a program code, it is necessary to first identify the precise program statements responsible for the errors and then to fix them. Identifying errors in a program code and then fix them up are known as debugging.

1. **Brute Force Method:** This is the most common method of debugging but is the least efficient method. In this approach, the program is loaded with print statements to print the intermediate values with the hope that some of the printed values will help to identify the statement in error. This approach becomes more systematic with the use of a symbolic debugger (also called a source code debugger), because values of different variables can be easily checked and break points and watch points can be easily set to test the values of variables effortlessly.

2. **Backtracking:** This is also a fairly common approach. In this approach, beginning from the statement at which an error symptom has been observed, the source code is traced backwards until the error is discovered. Unfortunately, as the number of source lines to be traced back increases, the number of potential backward paths increases and may become unmanageably large thus limiting the use of this approach.

3. **Cause Elimination Method:** In this approach, a list of causes which could possibly have contributed to the error symptom is developed and tests are conducted to eliminate each. A related technique of identification of the error from the error symptom is the software fault tree analysis.

4. **Program Slicing:** This technique is similar to back tracking. Here the search space is reduced by defining slices. A slice of a program for a particular variable at a particular statement is the set of source lines preceding this statement that can influence the value of that variable

## 9. Error Seeding

In error-seeding technique, a predefined number of artificially generated errors is "sown" in the program code. After that, test runs are used to detect errors and to examine the ratio between actual and artificial errors based on the total number of detected errors. The testers do not know the artificially generated errors.

Error seeding, as the name implies, seeds the code with some known errors. In other words, some artificial errors are introduced into the program artificially. The number of these seeded errors detected in the course of the standard testing procedure is determined.

These values in conjunction with the number of unseeded errors detected can be used to predict:
• The number of errors remaining in the product.
• The effectiveness of the testing strategy.

> Let N be the total number of defects in the system and let n of these defects be found by testing.
> Let S be the total number of seeded defects, and let s of these defects be found during testing.
> $n/N = s/S$
> or
> $N = S \times n/s$
> Defects still remaining after testing = $N–n = n\times(S – s)/s$

Error seeding works satisfactorily only if the kind of seeded errors matches closely with the kind of defects that actually exist. However, it is difficult to predict the types of errors that exist in a software. To some extent, the different categories of errors that remain can be estimated to a first approximation by analysing historical data of similar projects. Due to the shortcoming that the types of seeded errors should match closely with the types of errors actually existing in the code, error seeding is useful only to a moderate extent.

**Mutation Testing**
A mutation is a small change in a program. Such small changes are intended to model low level defects that arise in the process of coding systems. Ideally mutations should model low-level defect creation.

Mutation testing is a structural testing method aimed at assessing/improving the adequacy of test suites, and estimating the number of faults present in systems under test.

The process, given program P and test suite T, is as follows:
- We systematically apply mutations to the program P to obtain a sequence P1, P2,... Pn of mutants of P. Each mutant is derived by applying a single mutation operation to P.
- We run the test suite T on each of the mutants, T is said to kill mutant Pj if it detects an error.
- If we kill k out of n mutants the adequacy of T is measured by the quotient k/n. T is mutation adequate if k = n.

One of the benefits of the approach is that it can be almost completely automated.

## 10. Software Project Management

In many ways managing a software project is like managing any other engineering project. However, it is also true that software project management is more difficult. Perhaps the most important reason is the product is intangible. Monitoring the completion status of an entity that one cannot see or feel is a formidable challenge. Also standard processes and designs do not exist in the software field in the same way that they are found in handbooks of other engineering disciplines. People are the crucial element in project management and can be organized in teams that vary in their level of autonomy from traditional hierarchical structures to the “self-organizing” teams of the new agile paradigm.

**People Capability Maturity Model (P-CMM)**
P-CMM is a framework for continuously improving the management and development of the human assets of an organisation. It provides guidance in the following:
- Characterizing the maturity of workforce practices
- Setting priorities for immediate action
- Integrating workforce development with process improvement
- Becoming an employer of choice

P-CMM can be used alone or in conjunction with the SEI software process improvement programs.

**Project Estimation**

Early in a project the software development group and management must establish estimates for resources required, work to be done and time to delivery. Project planning is crucial to success. Technical people frequently do not take planning activities as seriously as they should and project cost; quality and time to delivery are often affected as a result. Cost estimating techniques are available based on metrics accumulated from past similar project experiences. The usual approach is to use several methods and compare values. If these values vary widely, then this variance is taken as an indication of the need for more information. **CO**nstructive **CO**st **MO**del (COCOMO) II is a popular estimation model for conventional software. The parameters in this model have been derived from data from over 4,000 software projects. Estimation methods have also been developed for object-oriented and agile development.
An important management observation is that software development time is not solely a function of the number of people on the project. One should not succumb quickly to the temptation to add more people to a late project. Adding more people could actually make it later.

## 11. COCOMO

Constructive Cost Estimation Model was proposed by Boehm [1981]. According to Boehm, software cost estimation should be done through three stages: Basic COCOMO, Intermediate COCOMO, and Complete COCOMO.

The basic COCOMO model gives an approximate estimate of the project parameters. The basic COCOMO estimation model is given by the following expressions:

$$\textbf{Effort} = a_1 \times (\text{KLOC})^{a_2} \text{ PM}$$

$$\textbf{Tdev} = b_1 \times (\text{Effort})^{b_2} \text{ Months}$$

- KLOC is the estimated size of the software product expressed in Kilo Lines of Code,
- a1, a2, b1, b2 are constants for each category of software products,
- Tdev is the estimated time to develop the software, expressed in months,
- Effort is the total effort required to develop the software product, expressed in person months (PMs).

The effort estimation is expressed in units of person-months (PM). It is the area under the person-month plot. It should be carefully noted that an effort of 100 PM does not imply that 100 persons should work for 1 month nor does it imply that 1 person should be employed for 100 months, but it denotes the area under the person-month curve.

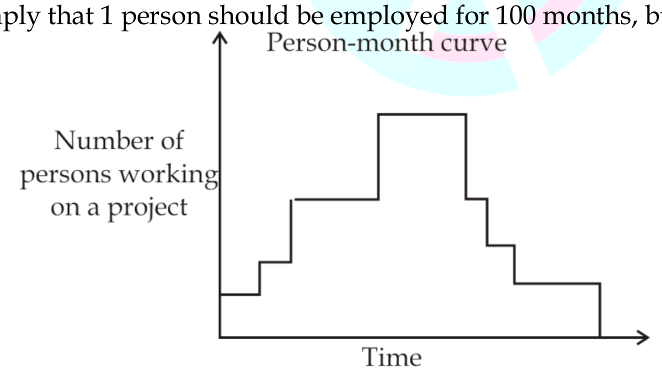



Fig. Person Month Curve

- Person-months (PM) denoted by the area under the person-month curve

## 12. Risk Management

Anticipating and having a plan for potential project problems will help avoid "crisis management" when problems do occur. The Spiral Process Model discussed earlier provides a framework for dealing with this issue. An investment of project management time in risk identification and monitoring can help keep potential problems to a minimum. Some examples of project risks are: changing requirements, low estimates of components reuse, high technical staff turnover and change in delivery deadline.

**Risk assessment**

The objective of risk assessment is to rank the risks in terms of their damage causing potential. For risk assessment, first each risk should be rated in two ways:

- The likelihood of a risk coming true (denoted as r).

- The consequence of the problems associated with that risk (denoted as s).

Based on these two factors, the priority of each risk can be computed:

**p = r * s**
Where, p is the priority with which the risk must be handled, r is the probability of the risk becoming true, and s is the severity of damage caused due to the risk becoming true. If all identified risks are prioritized, then the most likely and damaging risks can be handled first and more comprehensive risk abatement procedures can be designed for these risks.

**Risk containment**
After all the identified risks of a project are assessed, plans must be made to contain the most damaging and the most likely risks. Different risks require different containment procedures. In fact, most risks require ingenuity on the part of the project manager in tackling the risk.

Risk containment Strategies

1. **Avoid the risk:** This may take several forms such as discussing with the customer to change the requirements to reduce the scope of the work, giving incentives to the engineers to avoid the risk of manpower turnover, etc.
2. **Transfer the risk**: This strategy involves getting the risky component developed by a third party, buying insurance cover, etc.
3. **Risk reduction:** This involves planning ways to contain the damage due to a risk. For example, if there is risk that some key personnel might leave, new recruitment may be planned.

**Risk leverage**
To choose between the different strategies of handling a risk, the project manager must consider the cost of handling the risk and the corresponding reduction of risk. For this the risk leverage of the different risks can be computed.

Risk leverage is the difference in risk exposure divided by the cost of reducing the risk. More formally,
**risk leverage = (risk exposure before reduction – risk exposure after reduction) / (cost of reduction**)

## 13. ISO 9000 Certification vs. SEI/CMM

ISO 9000 is a set of standards for quality assurance systems. The standards were developed by the International Organization for Standardization (ISO). First published in 1987, the standards were revised in 1994. They provide a foundation for organizations to develop or improve their quality assurance systems. ISO 9000 describes the elements of a quality assurance system in general terms. These elements include the organizational structure, procedures, processes, and resources needed to implement quality planning, quality control, quality assurance, and quality improvement. However, ISO 9000 does not describe how an organization should implement these quality system elements. Consequently, the challenge lies in designing and implementing a quality assurance system that meets the standard and fits the company's products, services, and culture.

**SEI CMM: Key process areas (KPA) of a software organization**
Except for SEI CMM level 1, each maturity level is characterized by several Key Process Areas (KPAs) that includes the areas an organization should focus to improve its software process to the next level. The focus of each level and the corresponding key process areas

| CMM Level | Focus | Key Process Ares |
|---|---|---|
| 1. Initial | Competent people | |
| 2. Repeatable | Project management | Software project planning Software configuration management |
| 3. Defined | Definition of processes | Process definition training program peer reviews |
| 4. Managed | Product and process quality | Quantitative process metrics Software quality management |
| 5. Optimizing | Continuous process improvement | Defect prevention process change management technology change management |

**Difference between of ISO 9000 certification and the SEI CMM**

For quality appraisal of a software development organization, the characteristics of ISO 9000 certification and the SEI CMM differ in some respects. The differences are as follows:

- ISO 9000 is awarded by an international standards body. Therefore, ISO 9000 certification can be quoted by an organization in official documents, communication with external parties, and the tender quotations. However, SEI CMM assessment is purely for internal use.
- SEI CMM was developed specifically for software industry and therefore addresses many issues which are specific to software industry alone.
- SEI CMM goes beyond quality assurance and prepares an organization to ultimately achieve Total Quality Management (TQM). In fact, ISO 9001 aims at level 3 of SEI CMM model.
- SEI CMM model provides a list of key process areas (KPAs) on which an organization at any maturity level needs to concentrate to take it from one maturity level to the next. Thus, it provides a way for achieving gradual quality improvement.

## 14. Software Security

Everyone is aware that security is one of the most important issues in the computer field today. What is not apparent to everyone is that the security challenges today are frequently software problems. The weak points are the applications at the ends of the communications link and therefore represent the points of greatest vulnerability to attack.

Three trends are often cited as introducing security risks into systems and contributing to the magnitude of the security problem today.

1. The increasing complexity of systems make them more difficult to understand and hence more difficult to secure.
2. Increasing access to applications through various computer network technologies adds considerably to the security risks.
3. Software is being increasingly designed to be extensible with the incremental addition of functionality making it impossible to anticipate the kind of mobile code (updates) that may be downloaded.

**Can Security be Defined?**

A good question is "Can we ever declare a software application secure?". Unfortunately security, like many other engineering goals, is a relative quantity and 100 percent security is unachievable. A better question is to be more specific and ask, "Secure against what and from whom?". Some consider security to be a subset of reliability.

### 14.1 Approaches to the Security Problem

**Penetrate and Patch:** Often software is developed in an "Internet time" highly compressed schedule in order to be first to market. This approach considers security as an add-on feature after delivery. When vulnerabilities are found, frequently as a result of an attack, patches are developed and issued to the user community. There are many problems with this "penetrate-and-patch" approach to security. Here are a few:

- "Developers can only patch problems which they know about. Attackers may find problems that they never report to developers.
- Patches are rushed out as a result of market pressures on vendors, and often introduce new problems of their own to a system.
- Patches often only fix the symptom of a problem, and do nothing to address the underlying cause.
- Patches often go unapplied, because system administrators tend to be overworked and often do not wish to make changes to a system that "works [7]". It should also be noted that system administrators are often not security experts.

### 14.2 Build Security into the Software Development Life Cycle

The recommended approach is to incorporate software security as an engineering goal throughout the software engineering life cycle. Since many of the issues of software security are issues of risk management, the spiral model of software development is often mentioned as appropriate, with the repetitive spiral refining and converging security considerations toward the final goal. Some activities that should be added to each life cycle stage are listed below.

- **Requirements:** Add security specifications.
- **Design:** Develop threat models by viewing the system form an adversary's perspective and apply security design principles, e.g. "Design with the Enemy in Mind"

- **Implementation:** Add secure coding standards and language subsets
- **Testing:** Add Security test plans and use random input testing (e.g. Fuzz Testing) or vulnerability analysis using penetration testing.

**Principles for Software Security**

It has been said that 90% of security problems can be avoided if the following principles are followed:

1. Secure the weakest link: security is a chain
2. Practice defence in depth: manage risk with diverse defensive strategies
3. Fail securely: Failures are unavoidable and should be planned for
4. Follow the principle of least privilege: minimum access required to perform an operation and only for the minimum time necessary
5. Compartmentalize: minimize the amount of potential damage by organizing the system into the smallest number of units as possible.
6. Keep it simple
7. Promote privacy
8. Remember that hiding secrets is inherently difficult
9. Be reluctant to trust: Servers should be designed to distrust clients and conversely.
10. Use your community resources: Use security libraries and cryptographic algorithms that have been widely used and evaluated

### 14.3 Some Important Specific Software Security Issues

**Language Selection:** Many factors influence the choice of a programming language to use for implementation. It is common for efficiency considerations to dominate the language selection process. One of the factors should be security considerations. For example, choosing the C programming language for efficiency should take into account the inherent security risks associated with a language that has no bounds checks on array and pointer references. The programmer must build these checks into the program code. C program efficiencies and low-level data manipulation capabilities come at the high risk of security vulnerabilities and very special diligence is required. Using a language like Java can greatly reduce these risks, since it performs bounds checking. However, the system requirements must tolerate a lower level of run-time performance for this to be a viable option.

**Buffer Overflows:** Buffer overflows as a security vulnerability have been discussed for 40 years and yet this type of software problem continues to be one of the most frequently reported instances of system attacks. A buffer overflow is a condition caused by a write operation into a fixed-sized buffer in which the size of the data is larger than the size of the buffer. Most buffer overflows are the result of the properties of the C language mentioned in the last section above. This is the case with C++ as well.

## PRACTICE SET

1. Which of the following term describes testing?
(a) Finding broken code
(b) Evaluating deliverable to find errors
(c) A stage of all projects
(d) None of the mentioned
(e) Both (a) and (c)
2. Lower and upper limits are present in which chart?
(a) Run chart (b) Bar chart
(c) Control chart (d) None of the mentioned
(e) Both (b) and (c)
3. What are the various Testing Levels?
(a) Unit Testing (b) System Testing
(c) Integration Testing (d) All of the mentioned
(e) None of these
4. The testing in which code is checked
(a) Black box testing (b) White box testing
(c) Red box testing (d) Green box testing
(e) None of these
5. Acceptance testing is also known as
(a) Grey box testing (b) White box testing
(c) Alpha Testing (d) Beta testing
(e) None of these
6. Beta testing is done at
(a) User's end (b) Developer's end
(c) User's & Developer's end
(d) None of the mentioned
(e) Tester's end
7. Which of the following testing types is not a part of system testing?
(a) Recovery testing (b) Stress testing
(c) System testing (d) Random testing

(e) None of these

8. What is testing process' first goal?
(a) Bug prevention (b) Testing
(c) Execution (d) Analyses
(e) None of these

9. Software mistakes during coding are known as
(a) errors (b) failures
(c) bugs (d) defects
(e) None of these

10. Effective testing will reduce _______ cost.
(a) maintenance (b) design
(c) coding (d) documentation
(e) None of these

11. Which model in system modelling depicts the static nature of the system ?
(a) Behavioral Model (b) Context Model
(c) Data Model (d) Structural Model
(e) None of these

12. The UML supports event-based modeling using ______ diagrams.
(a) Deployment (b) Collaboration
(c) State chart (d) All of the mentioned
(e) None of these

13. Selection of a model is based on
(a) Requirements
(b) Development team & Users
(c) Project type and associated risk
(d) All of the mentioned
(e) None of these

14. The spiral model was originally proposed by
(a) IBM (b) Barry Boehm
(c) Pressman (d) Royce
(e) None of these

15. ________ allows us to infer that different members of classes have some common characteristics.
(a) Realization (b) Aggregation
(c) Generalization (d) dependency
(e) None of these

16. Which of the following diagram is not supported by UML considering Data-driven modeling?
(a) Activity (b) Data Flow Diagram (DFD)
(c) State Chart (d) Component
(e) None of these

17. Which model in system modelling depicts the dynamic behaviour of the system?
(a) Context Model (b) Behavioral Model
(c) Data Model (d) Object Model
(e) None of these

18. Unit testing is done by..
(a) Users (b) Developers
(c) Customers (d) None of the mentioned
(e) Tester

19. Efficiency in a software product does not include ___
(a) responsiveness (b) licensing
(c) memory utilization (d) processing time
(e) None of these

20. Which of these does not account for software failure?
(a) Increasing Demand (b) Low expectation
(c) Increasing Supply
(d) Less reliable and expensive
(e) None of these

21. What is Cyclomatic complexity?
(a) Black box testing (b) White box testing
(c) Yellow box testing (d) Green box testing
(e) None of these

22. Maintenance testing is performed using which methodology?
(a) Retesting (b) Sanity testing
(c) Breadth test and depth test
(d) Confirmation testing(e) None of these

23. Which of the following is non-functional testing?
(a) Black box testing (b) Performance testing
(c) Unit testing (d) None of the mentioned
(e) White box testing

24. SPICE stands for
(a) Software Process Improvement and Compatibility Determination
(b) Software Process Improvement and Control Determination
(c) Software Process Improvement and Capability dEtermination
(d) None of the mentioned
(e) Simulation Process Impliment and Capability Determination

25. Which of the following is not used in measuring the size of the software?
(a) KLOC (b) Function Points
(c) Size of module (d) None of the mentioned
(e) Both (a) and (b)

26. Which testing integrates the set of classes required to respond to one input or event for the system?
(a) cluster testing (b) thread-based testing
(c) use-based testing (d) Both (c) and (a)
(e) None of these

27. Which of the following is a part of testing OO code?
(a) Validation tests (b) Integration tests
(c) Class tests (d) System tests
(e) All of the mentioned

28. In which of the following testing strategies, a smallest testable unit is the encapsulated class or object?
(a) Unit testing (b) Integration testing
(c) System testing (d) Machine testing
(e) None of these

29. Test should be conducted for every possible
(a) data (b) case
(c) variable (d) All of the mentioned
(e) None of these

30. Which of the following is not a part of bug report?
(a) Test case (b) Output

(c) Software Version (d) LOC
(e) None of these

31. Boundary value analysis belongs to?
(a) White Box Testing (b) Black Box Testing
(c) White Box & Black Box Testing
(d) Silver box testing (e) None of these

32. Which of the following life cycle model can be chosen if the development team has less experience on similar projects?
(a) Spiral (b) Waterfall
(c) RAD
(d) Iterative Enhancement Model
(e) None of these

33. The spiral model has two dimensions namely _______ and ____________
(a) diagonal, angular (b) radial, perpendicular
(c) radial, angular (d) diagonal, perpendicular
(e) None of these

34. Which of the following is/are White box technique?
(a) Statement Testing (b) Decision Testing
(c) Condition Coverage (d) All of the mentioned
(e) None of these

35. Testing done without planning and Documentation is called
(a) Unit testing (b) Regression testing
(c) Adhoc testing (d) All of the mentioned
(e) None of these

36. Behavioral testing is
(a) White box testing (b) Black box testing
(c) Grey box testing (d) All of the mentioned
(e) None of these

37. The construction of object-oriented software begins with the creation of
(a) design model (b) analysis model
(c) code levels
(d) both design and analysis model
(e) None of these

38. ____ methods can be used to drive validations tests
(a) Yellow-box testing (b) Black-box testing
(c) White-box testing (d) All of the mentioned
(e) None of these

39. Name an evaluation technique to assess the quality of test cases.
(a) Mutation analysis (b) Validation
(c) Verification (d) Performance analysis
(e) None of these

40. Which of the following is a common pointwer problem?
(a) Data sharing errors
(b) Accessing data elements of the wrong type
(c) Attempting to use memory areas after freeing them
(d) All of the mentioned
(e) None of these

41. Which perspective in system modelling shows the system or data architecture?
(a) Structural perspective(b) Behavioral perspective
(c) External perspective (d) All of the mentioned
(e) None of these

42. Which level of Entity Relationship Diagram (ERD) models all entities and relationships?
(a) Level 1 (b) Level 2
(c) Level 3 (d) Level 4
(e) None of these

43. Which of the following statement is incorrect regarding the Class-responsibility-collaborator (CRC) modeling?
(a) All use-case scenarios (and corresponding use-case diagrams) are organized into categories in CRC modelling
(b) The review leader reads the use-case deliberately
(c) Only developers in the review (of the CRC model) are given a subset of the CRC model index cards
(d) All of the mentioned
(e) None of these

44. Which two models doesn't allow defining requirements early in the cycle?
(a) Waterfall & RAD (b) Prototyping & Spiral
(c) Prototyping & RAD (d) Waterfall & Spiral
(e) None of these

45. Which two of the following models will not be able to give the desired outcome if user's participation is not involved?
(a) Waterfall & Spiral (b) RAD & Spiral
(c) RAD & Waterfall (d) RAD & Prototyping
(e) None of these

46. Choose the correct option from given below:
(a) Prototyping Model facilitates re-usability of components
(b) RAD Model Model facilitates re-usability of components
(c) Both RAD & Prototyping Model facilitates re-usability of components
(d) None of these
(e) All (a), (b) and (c)

47. What is the major advantage of using Incremental Model?
(a) Customer can respond to each increment
(b) Easier to test and debug
(c) It is used when there is a need to get a product to the market early
(d) Easier to test and debug & It is used when there is a need to get a product to the market early
(e) None of these

48. Identify the disadvantage of Spiral Model.
(a) Doesn't work well for smaller projects
(b) High amount of risk analysis
(c) Strong approval and documentation control

(d) Additional Functionality can be added at a later date
(e) None of these

49. White Box techniques are also classified as
(a) Design based testing (b) Structural testing
(c) Error guessing technique
(d) None of the mentioned (e) Solution testing

50. The object of ___________within an OO system is to design tests that have a high likelihood of uncovering plausible bugs.
(a) Fault-based testing (b) Integration testing
(c) Use-based testing (d) Scenario-based testing
(e) None of these

51. ____________ categorizes class operations based on the generic function that each performs
(a) Category-based partitioning
(b) Attribute-based partitioning
(c) State-based partitioning
(d) None of the mentioned
(e) Both (b) and (c)

52. What refers to the externally observable structure of an OO program?
(a) Deep structure (b) Surface structure
(c) Core structure (d) All of the above
(e) None of these

53. Cyclomatic Complexity method comes under which testing method.
(a) Yellow box (b) White box
(c) Gray box (d) Black box
(e) None of these

54. Which of the following is the way of ensuring that the tests are actually testing code?
(a) Control structure testing
(b) Complex path testing
(c) Code coverage
(d) Quality assurance of software
(e) None of these

55. The Unified Modeling Language (UML) has become an effective standard for software modelling. How many different notations does it have?
(a) Three (b) Four (c) Six
(d) Nine (e) None of these

56. ______ & ________ diagrams of UML represent Interaction modeling.
(a) Use Case, Sequence (b) Class, Object
(c) Activity, State Chart (d) All of the mentioned
(e) None of these

57. If you were a lead developer of a software company and you are asked to submit a project/product within a stipulated time-frame with no cost barriers, which model would you select?
(a) Waterfall (b) Spiral (c) RAD
(d) Incremental (e) None of these

58. A company is developing an advance version of their current software available in the market, what model approach would they prefer ?
(a) RAD (b) Iterative Enhancement
(c) Both RAD & Iterative Enhancement
(d) Spiral (e) None of these

59. The Incremental Model is a result of combination of elements of which two models?
(a) Build & FIX Model & Waterfall Model
(b) Linear Model & RAD Model
(c) Linear Model & Prototyping Model
(d) Waterfall Model & RAD Model
(e) None of these

60. If you were to create client/server applications, which model would you go for?
(a) WINWIN Spiral Model (b) Spiral Model
(c) Concurrent Model (d) Incremental Model
(e) None of these

## SOLUTIONS

| | | | | | | |
|---|---|---|---|---|---|---|
| 1. (b) | 2. (a) | 3. (d) | 4. (b) | 5. (d) | 6. (a) | 7. (d) |
| 8. (a) | 9. (c) | 10. (a) | 11. (d) | 12. (c) | 13. (d) | 14. (b) |
| 15. (c) | 16. (b) | 17. (b) | 18. (b) | 19. (b) | 20. (c) | 21. (b) |
| 22. (c) | 23. (b) | 24. (c) | 25. (c) | 26. (b) | 27. (c) | 28. (a) |
| 29. (d) | 30. (d) | 31. (b) | 32. (a) | 33. (c) | 34. (d) | 35. (c) |
| 36. (b) | 37. (d) | 38. (b) | 39. (a) | 40. (d) | 41. (a) | 42. (b) |
| 43. (c) | 44. (b) | 45. (d) | 46. (c) | 47. (d) | 48. (a) | 49. (b) |
| 50. (a) | 51. (a) | 52. (b) | 53. (b) | 54. (c) | 55. (d) | 56. (a) |
| 57. (c) | 58. (c) | 59. (c) | 60. (c) | | | |

# 11 Programming Languages

## 1. Introduction

A language is the main medium of communicating between the Computer systems and the most common are the programming languages. As we know a Computer only understands binary numbers that is 0 and 1 to perform various operations but the languages are developed for different types of work on a Computer. They are based on certain syntactic and semantic rules, which define the meaning of each of the programming language constructs.

### 1.1 Some Terms Related to Programming Language

**Algorithm:** An algorithm is a well-defined procedure that allows a computer to solve a problem. Another way to describe an algorithm is a sequence of unambiguous instructions. The use of the term 'unambiguous' indicates that there is no room for subjective interpretation. Every time you ask your computer to carry out the same algorithm, it will do it in exactly the same manner with the exact same result.

**Syntax:** the syntax of a computer language is the set of rules that defines the combinations of symbols that are considered to be a correctly structured document or fragment in that language. Text-based computer languages are based on sequences of characters, while visual programming languages are based on the spatial layout and connections between symbols (which may be textual or graphical). Documents that are syntactically invalid are said to have a syntax error.

**Semantics:** semantics is the field concerned with the rigorous mathematical study of the meaning of programming languages. It does so by evaluating the meaning of syntactically legal strings defined by a specific programming language, showing the computation involved.
Syntax is about the structure or the grammar of the language. It answers the question: how do I construct a valid sentence? All languages, even English and other human languages have grammars, that is, rules that define whether or not the sentence is properly constructed. Semantics is about the meaning of the sentence. It answers the questions: is this sentence valid? If so, what does the sentence mean?
In summary, syntax is the concept that concerns itself only whether or not the sentence is valid for the grammar of the language. Semantics is about whether or not the sentence has a valid meaning.

**Source Code and Object Code:**
Source code is the fundamental component of a computer program that is created by a programmer. It can be read and easily understood by a human being. When a programmer types a sequence of C language statements into Windows Notepad, for example, and saves the sequence as a text file, the text file is said to contain the source code. Source code and object code refer to the "before" and "after" versions of a computer program that is compiled (see compiler) before it is ready to run in a computer. Object code, or sometimes an object module, is what a computer compiler produces. In a general sense object code is a sequence of statements or instructions in a computer language.

**Debugging:** Debugging is the art of diagnosing errors in programs and determining how to correct them. "Bugs" come in a variety of forms, including: coding errors, design errors, complex interactions, poor user interface designs, and system failures. Debugging is the routine process of locating and removing computer program bugs, errors or abnormalities, which is methodically handled by software programmers via debugging tools. Debugging checks, detects and corrects errors or bugs to allow proper program operation according to set specifications.

**Patch:** A patch is a piece of software designed to update a computer program or its supporting data, to fix or improve it. This includes fixing security vulnerabilities and other bugs, with such patches usually called bugfixes or bug fixes, and improving the usability or performance.

Patch (sometimes called a "fix") is a quick-repair job for a piece of programming. During a software product's beta test distribution or try-out period and later after the product is formally released, problems (called bug) will almost invariably be found. A patch is the immediate solution that is provided to users.

## 1.2 Types of Languages

Broadly there are two types that is Low Level Language and High-Level Language:
**Low Level Language:** Low-level languages those languages which are extremely close to machine language. A low-level programming language is a programming language that provides little or no abstraction from a computer's instruction set architecture—commands or functions in the language map closely to processor instructions. Generally, this refers to either machine code or assembly language. Low-level languages are more appropriate for developing new operating systems or writing firmware codes for micro-controllers.

**Assembly Language:** An assembly language is a group of languages that implements a symbolic representation of the machine code required to program certain CPU architecture. It is a programming language for microprocessors and other programming devices, and it is the most basic programming language available for any processor. Generally, assemblers produce object files, and most provide macros. Unlike high-level languages, assembly languages lack variables and functions, but they have the same structure and set of commands, much like machine languages. This programming language is helpful to programmers when speed is required and when they need to perform an operation that cannot be done in high-level languages.
Machine code is the only language a computer can process directly without a previous transformation.

**High Level Language:** A high-level language is a programming language such as C, FORTRAN, or Pascal that enables a programmer to write programs that are more or less independent of a particular type of computer. Such languages are considered high-level because they are closer to human languages and further from machine languages. In contrast, assembly languages are considered low-level because they are very close to machine languages. The main advantage of high-level languages over low-level languages is that they are easier to read, write, and maintain. Ultimately, programs written in a high-level language must be translated into machine language by a compiler or interpreter.
It is important to know that a "computer language" and a "programming language" are not quite the same. A computer language is code that can be read by a computer. A programming language (also called software languages) is used to make a program. For instance, HTML, CSS, XML, SQL, and Latex are examples of computer languages that are not programming languages. C/C++, Python, Lua, Scala, and Java are examples of computer languages that are programming languages.
Some programming languages are **scripting languages**. Scripting languages are computer languages that are not compiled (converted to binary/machine-code). In other words, the computer reads the plain text, usually by using a program called an interpreter. Examples of scripting languages include Python, Lua, Perl, Ruby, and JavaScript. **Markup languages** are computer languages that are not compiled (like scripting languages). However, markup languages are not used to make programs. Rather, they can be used to make web-pages (like HTML) or databases (like XML). Markup languages use "tags" instead of commands (such as "<br />").

**Procedural Language:** Procedural language is a type of computer programming language that specifies a series of well-structured steps and procedures within its programming context to compose a program. It contains a systematic order of statements, functions and commands to complete a computational task or program.
Procedural language is also known as imperative language. Procedural programming languages include C, Go, Fortran, Pascal, Ada, and BASIC.

**Object-Oriented Programming Language:** Object-oriented programming (OOP) is a programming paradigm based on the concept of "objects", which may contain data, in the form of fields, often known as attributes; and code, in the form of procedures, often known as methods. A feature of objects is that an object's procedures can access and often modify the data fields of the object with which they are associated. Significant object-oriented languages include Java, C++, C#, Python, PHP, Ruby, Perl, Delphi, Objective-C, Swift, Common Lisp, and Smalltalk. In your examination they can ask a question that which among the following is a type of object oriented programming language. So it's beneficial for you to know that which follow OOPS concept and which do not does so.

## 1.3. Compiler Interpreter and Assembler

Both compiler and interpreter convert human readable high level language like Java, C++ etc into machine language but there is difference in the way both function.
Compiler scans the entire program once and then converts it into machine language which can then be executed by computer's processor. In short compiler translates the entire program in one go and then executes it. Interpreter on the other hand first converts high level language into an intermediate code and then executes it line by line. This intermediate code is executed by another program. The execution of program is faster in compiler than interpreter as

in interpreter code is executed line by line. Compiler generates error report after translation of entire code whereas in case of interpreter once an error is encountered it is notified and no further code is scanned.
Whereas Assembler is used for converting the code of low level language (assembly language) into machine level language.

## 2. C Language

### 2.1. Introduction:

C is a programming language developed at AT & T's Bell Laboratories of USA in 1972. It was designed and written by Dennis Ritchie. It is one of the most popular computer languages today because of its structure, high-level abstraction, machine independent feature. C language was developed with UNIX operating system, so it is strongly associated with UNIX, which is one of the most popular network operating system.

### 2.2. C Keyword:

Keywords are the words whose meaning has already been explained to the C compiler. The keywords cannot be used as variable names because if we do so we are trying to assign a new meaning to the keyword, which is not allowed by the computer. Some C compilers allow you to construct variable names that exactly resemble the keywords. However, it would be safer not to mix up the variable names and the keywords. **The keywords are also called 'Reserved words'.**

**There are only 32 keywords available in C.**

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| auto | double | int | struct | break | else | long | switch | Case |
| enum | register | typedef | char | extern | return | union | const | Float |
| short | unsigned | continue | for | signed | void | default | goto | sizeof |
| volatile | do | if | static | while | | | | |

### 2.3. Data Types in C:

Data types simply refer to the type and size of data associated with variables and functions.
There are five basic data type use in C:

- int - integer: a whole number.
- float - floating point value: ie a number with a fractional part.
- double - a double-precision floating point value.
- char - a single character.
- void - valueless special purpose type which we will examine closely in later sections.

### 2.4. Header files in C:

Header files contain the set of predefined standard library functions that we can include in our c programs. But, to use these various library functions, we have to include the appropriate header files.

| Name | Description |
|---|---|
| stdio.h | Input/ Output Functions |
| conio.h | console input/output |
| assert.h | Diagnostics Functions |
| ctype.h | Character Handling Functions |
| cocale.h | Localization Functions |
| math.h | Mathematics Functions |
| setjmp.h | Nonlocal Jump Functions |
| signal.h | Signal Handling Functions |
| stdarg.h | Variable Argument List Functions |
| stdlib.h | General Utility Functions |
| string.h | String Functions |
| time.h | Date and Time Functions |

### 2.5. Decision making in C:

Decision making is about deciding the order of execution of statements based on certain conditions or repeat a group of statements until certain specified conditions are met. C language handles decision-making by supporting the following statements,

**1-if statement**: If statements in C is used to control the program flow based on some condition, it's used to execute some statement code block if expression is evaluated to true, otherwise it will get skipped. This is an simplest way to modify the control flow of the program. There are four different types of if statement in C. These are:

- Simple if Statement
- if-else Statement
- Nested if-else Statement
- else-if Ladder

**2-Switch statement:** switch statement is used when you have multiple possibilities for the if statement.
The basic format of switch statement is:

```
Syntax:
switch(variable)
{
case 1:
    //execute your code
break;
case n:
    //execute your code
break;
default:
    //execute your code
break;
}
```

After the end of each block it is necessary to insert a break statement because if the programmers do not use the break statement, all consecutive blocks of codes will get executed from each and every case onwards after matching the case block.

**3-goto statement:** C supports a special form of statement that is the goto Statement which is used to branch unconditionally within a program from one point to another. goto statement is used by programmers to change the sequence of execution of a C program by shifting the control to a different part of the same program.

**4-Conditional operator statement:** Conditional Operator returns the statement depends upon the given expression result.

## 2.6. C Loops:

Loops are used to execute a set of statements repeatedly until a particular condition is satisfied.
A loop statement allows us to execute a statement or group of statements multiple times.
There are 3 type of Loops in C language:

**1-while loop:** The most basic loop in C is the while loop and it is used is to repeat a block of code. while loop has one control expression (a specific condition) and executes as long as the given expression is true. Here is the syntax :

```
Syntax :
variable initialization ;
while (condition)
{
  statements ;
  variable increment or decrement ;
}
```

**2-for loop:** for loop is used to execute a set of statements repeatedly until a particular condition is satisfied. we can say it an open ended loop.

```
Syntax:
for(initialization; condition ; increment/decrement)
{
   statement-block;
}
```

In for loop we have exactly two semicolons, one after initialization and second after condition. In this loop we can have more than one initialization or increment/decrement, separated using comma operator.

**3-do-while loop:** In some situations it is necessary to execute body of the loop before testing the condition. Such situations can be handled with the help of do-while loop. do statement evaluates the body of the loop first and at the end, the condition is checked using while statement. It means that for at least one time, the body of the loop will be executed, even though the starting condition inside while is initialized to false.

Syntax:

```
do
{
  ....
  .....
}
while(condition)
```

### 2.6.1. C Loop Control Statements:

Loop control statements are used to change normal sequence of execution of loop.

| Statement | Syntax | Description |
|---|---|---|
| break statement | break; | Is used to terminate loop or switch statements. |
| continue statement | continue; | Is used to suspend the execution of current loop iteration and transfer control to the loop for the next iteration. |
| goto statement | goto labelName;<br>labelName: statement; | It's transfer current program execution sequence to some other part of the program. |

### 2.7. Functions in C Programming:

*Functions* are used to provide modularity to the software. By using functions, you can divide complex tasks into small manageable tasks. The use of functions can also help avoid duplication of work. For example, if you have written the function for calculating the square root, you can use that function in multiple programs.

**Syntax of function**

```
Arguments list
Function header
Function (argl, arg2, arg3)
type argl, arg2, arg3 Arguments declarations
{
statement 1;
statement2;
statement3; Function body
statement4;
}
```

**The general format of a function is**

```
<Return type> <Function name> <Parameter list>
{
<local definitions>
executable statements;
Return (expression);
}
```

**Example:**

```
int f1 (int j, float f)
{
int k;
k = l;
return (k);
}
```

**Formal Arguments:**

The arguments specify in function header is called as formal arguments because they represent the names of data items that *are* transferred in to the function from the calling portion of the program. They *are* also called as formal parameters.

**Actual Arguments:**

The arguments specify in function call is called as actual arguments. Since they define the data items that are actually transferred. They are also called as actual parameters or simply arguments.

**Example:**

```
main ()
{
int a, b, c;
printf ("\n Enter two number");
scanf ("\%d %d ", &a ,&b);
Actual arguments
c = add (a, b); Function Call
printf("The addition of two no is %d ",c)
Formal arguments
Add (int x, inty)
{
int z;
z = x +y;
return (z);
}
```

**Program**

```
#include <stdio.h>
int add (int x, int y) //A
{
int z; //B
z = x + y;
return (z); //C
}
main ()
{
int i, j, k;

i = 10;
j = 20;
k = add(i, j); //D
printf ("The value of k is%d\n", k); //E
}
```

**Explanation**

- This function adds two integers and returns their sum.
- When defining the name of the function, its return data type and parameters must also be defined. For example, when you write
- int add (int x, int y)
- int is the type of data to be returned, add is the name of the function, and x and y are the parameters of the type int. These are called *formal parameters.*
- The body of a function is just like the body of main. That means you can have variable declarations and executable statements.
- A function should contain statements that return values compatible with the function's return type.
- The variables within the function are called *local variables*.
- After executing the return statement, no further statements in the function body are executed.
- The name of the function can come from the arguments that are compatible with the formal parameters, as indicated in statement D.
- The arguments that are used in the call of the function are called *actual parameters.*
- During the call, the value of the actual parameter is copied into the formal parameter and the function body is executed.
- After the return statement, control returns to the next statement which is after the call of the function.

### 2.7.1 Function call

Functions cannot execute themselves with exception to main (). In a C program only main is executable. All other functions are not. If this is true, then there must be some way by which the functions executes. This is known as function call. A function can be called from another function with exception to main() function. This means main() functions can call other functions but other functions cannot call main(). If by curiosity you happen to call main() function and it compiled well then you may end up in infinite loop.

A function is called from within another function by just putting its name inside that function with proper parameters and if it returns anything then you must supply a proper variable to catch the value. Before you can call the function, it should be defined or at least prototype.

## 2.8. Parameter Passing

Information can be passed from one function to another using parameters.

**Program**

```
Main ( )
{
int i;
i = 0;
printf (" The value of i before call %d \n", i);
f1 (i);
printf (" The value of i after call %d \n", i);
}
void f1 (int k)
{
k = k + 10;
}
```

**Explanation**

- The parameter used for writing the function is called the formal parameter, k in this case.
- The argument used for calling the function is called the actual parameter.
- The actual and formal parameters may have the same name.
- When the function is called, the value of the actual parameter is copied into the formal parameter. Thus k gets the value 0. This method is called *parameter passing by value*.
- Since only the value of i is passed to the formal parameter k, and k is changed within the function, the changes are done in k and the value of i remains unaffected.
- Thus i will equal 0 after the call; the value of i before and after the function call remains the same.

## 2.9. Categories of function

A Function in C belongs to one of the following categories:

- Function with no arguments and no return values
- Function with arguments and no return values
- Function with no arguments and return values
- Function with arguments and return values

## 2.10. Scope rules:

**Local Variables:**

**A local variable** is a variable that is given *local scope.* Such variables are accessible only from the function or block in which it is declared. Local variables are contrasted with global variables. Local variables *are* special because in most languages they are automatic variables stored on the call stack directly. This means that when a recursive function calls itself, local variables in each instance of the function are given separate memory address space. Hence variables of this scope can be declared, written to, and read, without any risk of side-effects to processes outside of the block in which they are declared.

**Global variables:**

These variables can be accessed (i.e. known) by any function comprising the program. They are implemented by associating memory locations with variable names. They do not get recreated if the function is recalled. Global variables are used extensively to pass information between sections of code that don't share a caller/callee relation like concurrent threads and signal handlers. In Other word Global variables are those, which are available throughout the program's execution. They are not defined inside any block. They exist in memory until the program is in memory.

## 2.11. Parameter passing technique

Parameters are the values upon which we want the function should operate These are passed to functions within parenthesis. The mechanism by which parameters are passed to function is known as parameter passing. There are basically 2-types of techniques

- **Call by value**
- **Call by reference**

**Call by value**

In call by value method if the formal parameters are changed then the values in calling function will not change. This is because formal parameters are local to the function, and actual parameters are local to the calling function. But there are certain cases we need to change the value of the variable from inside a function. This is usually done in *pass by reference* method or else it can be done only in case of global variable. Only global variables can be changed from within function. Here an interesting thing comes to mind. If we have defined a global variable and a local variable with same name then if we change the value then whose value will be changed, let us see it. Consider the following examples.

**Example:**

```
main ()
{
int a = 10, b=20;
swapy (a,b);
printf ("\na = % d b = % d", a,b);
}
swapy (int x, int y)
{
int t;
t = x;
x = y;
y = t;
printf ("\n x = % d y = % d", x, y);
}
*-------------------Output----------------------*
x = 20 y = 10
a =10 b=20
```

**Call by Reference**

Call by reference is implemented indirectly by passing the address of the variable. implement the function indirectly. This is done by passing the address of the variable and changing the value of the variable through its address.

**Example:**

```
main ()
{
int a = 10, b =20,
swapr (&a, &b);
printf ("\n a = %d b= %d", a, b);
}
swapr (int *x, int * y)
{
int t;
t= *x
*x = *y;
*y = t;
printf ("\n x = % d y = % d", *x, *y);
}
*------------------Output-----------------------*
x =20 y =10
a = 20 b=10
```

### 2.12. Recursion

Recursion is the phenomenon of function calling itself repeatedly. Consider a function calls itself from within itself. Recursion is useful in the situation, where some series of similar and dependent calculations are needed. The best example is
factorial. *Factorial of n(n!)* can be written as *n*(n-l)*(n-2)*(n-3)*......*2*1*
So, it is a continued product from n to 1. Hence, we can rewrite it as *n*factorial of(n-l).* From this we can write a function to calculate the factorial of a number

```
int factorial(int n)
{
if(n<=l)
return(l);
else
return(n*factorial(n-l));
}
```

**Example**.

```
/* to calculate the factorial through recursion */
#include<stdio.h>
#include<conio.h>
int fact(int);
void main()
{
int n,result;
clrscr();
printf("Enter the number");
scanf("%d",&n);
result=fact(n);
printf("\n The factorial of %d is %d",n,result);
getch();
}
int fact(int x)
{
int f;
if(x==1)
{
return(x);
}
else
{
return(x*fact(x-1));
}
}
```

### 2.13. Pointer

A pointer is a variable that holds a memory address, usually the location of another variable in memory. Pointer is a variable, which contains memory address. This address is the location of another variable in memory. Consider 2 variables. If the first variable contains the address of second variable, then it is said the first variable points to second variable.
The declaration of a variable consists of type (any valid type), '*' and variable name. The syntax for this purpose is as follows
**<type> * <variable name>;**
The type defines here which type of data the pointer is going to point. This means which type of variables address this pointer will point to. The "&" operator gets the address of any variable and the '*' operator counters the '&' operator. That is gets the value at that address taking into consideration the base type. Let us see an example

```
int a, b ;
int *ptr1, *ptr2;
/*Here ptr1 and ptr2 are two pointer type variables, which can contain the address of 2 integer variables*/
ptr1=&a;
ptr2=&b;
```

In the above example **ptr1** contains the **address of a** and **ptr2** contains the **address of b**. By address it is meant the memory location of the variable at the run time. Suppose at run time **a** is at **location 1004** and it has a value of say 50 then **ptr1** will contain **1004 not 50.** Let us now check the * operator. This operator gets the value. So, to get the value stored at any location we can write

*m=*ptr1;*

Here type of m and base type of ptr1 must be same. Now contains the **value** stored **at address** pointed by **ptr1,** that is value of **a.**

A pointer can be assigned to another pointer type variable of same base type Consider the following code.

```
int a
int *ptr1, *ptr2;
ptrl=&a; /* ptr1 is assigned the address of a */
ptr2=ptr1; /* ptr2 also stores the address of a */
printf("%p%p", ptr1, ptr2);
/* will print address of 2 times */
```

The above code shows assigning of the address of a variable in to a pointer type variable. Then we store the address stored in one pointer type variable in another pointer type variable. Then in last statement we are printing the address stored in those pointers. You can notice we are using a %p a here to print the address.

**Array of Pointer**

If array of any data type is possible then array of pointers is also possible. In this case each element of the array is a pointer to any data type. Suppose we want to create an array of integer pointer then we should write in following manner.

```
int k,l,m,n;
int *a[4]; /* here a is an array of pointe to integer */
/*hence we can write*/
a[0]=&k;
a[l]=&l
a[2]=&m
a[3]=&n;
```

**Multidimensional Array**

We read earlier how to represent multi-dimensional arrays using pointers. Now we will look into how to access the elements of multi-dimensional array. Follow the example bellow.

```
Int a[3][3][3]; /* is a multi-dimensional array*/
Int p,x,y,z,;
Int ***k; /* k is a pointer to pointe to pointer to integer */
/* hence k can hold the address of 3-diensional array */
k=a;
/*now we will store values into the array*/
for(x=0;x<3;x+ +)
{
for(y=0;y<3;y++)
{
for(z=0;z<3;z++)
{
printf'("Enter element %d %d %d\n", x, y, z );
scanf("%d",&p); /* read the element */
*(*(*(k+x)+y)+z)=p; /* store the element in array */
}
}
}
/*now to print the values */
for(x=0;x<3;x+ +)
{
for(y=0;y<3;y++)
{
for(z=0; z<3;z++)
{
p=*(*(*(k+x)+y)+z);
```

```
printf ("The Element %d %d %d= %d\n", x, y,z , p);

}}
}
```

**Function and pointer**

We have discussed pass by value method in functions. In that case the value is passed to the function. But in case of pass by reference we pass the address of variables to the called function. As the address of variable is passed, any modification of value at that address will be reflected in the calling function. Hence this should be considered seriously.

The pass by reference is advantageous when we pass array, or we need the variables value should be changed by the function. Consider the case of a swapping function we need to alter the values and that should be available in calling function. We know a function can return only one value hence pass by value is not possible. So, it is done using pass by reference. Consider the following swapping example

```
#include<stdio.h>
void swap (int *p, int *q) /* we take two pointers */
int * temp /* temporary pointer variable */
*temp= * /*'store the value stored in p*/
*p= *q; /* store value stored in p into q */
*q= *temp;
}
main()
{
int a=10 b=30;
printf("value of a =%d and of b =%d before swap( ) \ n " , a , b );
swap (&a, &b); /* passing addresses*/
printf("value of a =%d and of b =%d before swap( ) \ n " , a , b );
}
```

## 2.14. Arrays

Array means a group of similar type of things. In C programming it means a collection of similar type of data in consecutive memory locations, as an example array of integers array of floats etc.

**Types of Arrays**

- Single-dimensional array
- Multi-dimensional array

**Single-dimensional Array:** The simplest form of an array is a single dimensional array. The array is given a name and is elements are referred to by their subscripts or indices.

**Introduction to Array Variables**

- Arrays are a data structure which **hold multiple values of the same data type**. Arrays are an example of a **structured variable** in which 1) there are a number of pieces of data contained in the variable name, and 2) there is an ordered method for extracting individual data items from the whole.
- Consider the case where a programmer needs to keep track of the ID numbers of people within an organization. Her first approach might be to create a specific variable for each user. This might look like
  **int id1 = 101; int id2 = 232; int id3 = 231;**
- It becomes increasingly more difficult to keep track of the IDs as the number of variables increase. Arrays offer a solution to this problem.

**Array Variables Example**

- An array is a multi-element box, a bit like a filing cabinet, and uses an indexing system to find each variable stored within it. **In C, indexing starts at zero**. Arrays, like other variables in C, must be declared before they can be used.
- The replacement of the previous example using an array looks like this:
  **int id[3];** /* declaration of array id */
  **id[0] = 101;**
  **id[1] = 232;**
  **id[2] = 231;**

- In the first line, we declared an array called id, which has space for three integer variables. Each piece of data in an array is called an element. Thus, array id has three elements. After the first line, each element of id is initialized with an ID number.

**Memory Representation of Single Dimension Arrays**
- Single –dimension arrays are essentially lists of information of the same type and their elements are stored in contiguous memory location in their index order. For instance, an array grade of type char with 8 elements declared as char grade[8]; will have the element grade[0] at the first allocated memory location, grade[1] at the next contiguous memory location, grade[2] at the next, and so forth. Since grade is a char type array, each element size is 1 byte(A character size is 1 byte) and it will be represented in memory as shown below

grade[0] grade[1]…………………………………………………..grade[7]
Address 101 102 103 104 105 106 107 108
If you have an int array age with 5 elements declared as int age[5]; Its each element will have size of(int)1 bytes for its storage. On a system with 2 bytes integer size, the array age's each element (being an integer ) will be having 2 bytes for its storage. If the starting memory location of array age is 100, then it will be represented in the memory as shown below
age[0] age[1] age[2] age[3] age[4]
Address 100 102 104 106 108

**Declaring Arrays**
- Arrays may consist of any of the valid data types. Arrays are declared along with all other variables in the declaration section of the program and the following syntax is used
  ***type array_name*[*n*];**
  grade
  age
- where ***n*** is the **number of elements** in the array. Some examples are
  int final[160];
  float distance[66];
- During declaration **consecutive memory locations** are reserved for the array and all its elements. After the declaration, you cannot assume that the elements have been initialized to zero. Random junk is at each element's memory location.

**Initializing Arrays during Declaration**
If the declaration of an array is preceded by the word static, then the array can be initialized at declaration. The initial values are enclosed in braces. e.g., static int value[9] = {1,2,3,4,5,6,7,8,9}; static float height[5]={6.0,7.3,2.2,3.6,19.8};
- Some rules to remember when initializing during declaration
- If the list of initial elements is shorter than the number of array elements, the remaining elements are initialized to zero.
- If a static array is not initialized at declaration manually, its elements are automatically initialized to zero.
- If a static array is declared without a size specification, its size equals the length of the initialization list. In the following declaration, a has size 5.
  static int a[]={-6,12,18,2,323};

**Using Arrays**
- Recall that indexing is the method of accessing individual array elements. Thus **grade[89]** refers to the **90**th element of the **grade**s array. A common programming error is **out-of-bounds array indexing**. Consider the following code:
  **int grade[3];**

**grade[5] = 78;**
- The result of this mistake is unpredictable and machine and compiler dependent. You could write over important memory locations, for example. Often run-time errors result.
- **Array variables and for loops often work hand-in-hand** since the for loop offers a convenient way to successively access array elements and perform some operation with them. Basically, the for loop counter can do double duty and act as an index for the array, as in the following summation example:
  **int total=0,i;**
  **int grade[4]={93,94,67,78};**
  **for (i=0; i<4; ++i)**
  **total += grade[i];**

*int num[3] = {7, 51, -22};*
*float values[5] = {2.1, 6.7};*
*char name[] = {'F', 'r', 'e', 'd'};*
- The first element of the array is index 0. If there are N elements, the last on is index (N-1).

**Example :**

```
#include <stdio.h>
main()
{
int a[5]; \\A
for(int i = 0;i<5;i++)
{
a[i]=i; \\B
}
print_elements(a);
}
void print_elements(int a[])
{
Remaining 3 values get initialized to zeros The compiler fills in the 4 after counting the number of initializers.
Beware: not a null-terminated string!
for(int i = 0;i<5;i++)
{
printf("value in array %d\n",a[i]);
}
}
```

**Explanation**

- Statement A defines an array of integers. The array is of the size 5—that means you can store 5 integers.
- Array elements are referred to using subscript; the lowest subscript is always 0 and the highest subscript is (size – 1). If you refer to an array element by using an out-of-range subscript, you will get an error. You can refer to any element as a[0], a[1], a[2], etc.
- Generally, you can use a for loop for processing an array. For the array, consecutive memory locations are allocated, and the size of each element is same.
- The array name, for example, a, is a pointer constant, and you can pass the array name to the function and manipulate array elements in the function. An array is always processed element by element.
- When defining the array, the size should be known.

**Two-dimensional (2D) Array**

A two-dimensional array is an array in which each element is itself an array. For instance, an array A[m][n] is an m by n table with m rows and n columns containing m x n elements.
The number of elements in a 2-D array can be determined by multiplying number of rows with number of columns. For example, the number of elements in an array A[7][9] is calculated as 7x 9 =63
The simplest form of a multidimensional array, the two-dimensional array, is an array having single-dimensional arrays as its elements.

**Multi-dimensional arrays**

Up to this point what we have discussed is single dimensional arrays. But arrays can have more than one dimension. Consider a matrix or a determinant They are 2-dimensional arrays. Consider a vector in 3-D plane. It is an array of 3-dimensions. Think of Einstein he took another dimension time. So to represent vector in his calculations a 4-D array is needed. More than 4-D arrays are rarely used.
As an example *int abc[3][4][2]* defines *abc* to be a *3-dimensional array* of *integer.* To access cells of this array we need 3 loops one for each dimension. Let us see how we can access the cells

```
int i j, k, temp;
int pqr[5][6][9];
/* input data into array…………………………...*/
for(i=0;i<5;i++)
{
for(j=0;j<6;j++)
{
```

```
for(k=0;k<9;k++)
{
printf("Enter a number ");
scanf("%d", &temp);
pqr[i][j][k]=temp;
}
}
}
/* output(display) elements of array ....................*/
for(i=0;i<5;i++)
{
for(j=0;j<6;j++)
{
for(k=0;k<9;k++)
{
printf("value at cell %d %d %d = %d\n",pqr[i][j][k]);
}
}
}
```

above code fragment is capable of reading a number from user store it in a multidimensional array and display it Here is a point to note. Watch the above code carefully, you will see pqr is an array of array of arrays. Just watch array is repeated 3-times. Similarly, a 2-D array is a single dimensional array of single dimensional arrays.

## 3. Object Oriented Programming Concepts

Object-Oriented Programming (OOP) uses "objects" to model real-world objects. Object-Oriented Programming (OOP) consist of some important concepts namely Encapsulation, Polymorphism, Inheritance and Abstraction. These features are generally referred to as the OOPS concepts.

An object in OOP has some state and behavior. In Java, the state is the set of values of an object's variables at any particular time and the behaviour of an object is implemented as methods. Class can be considered as the blueprint or a template for an object and describes the properties and behavior of that object, but without any actual existence. An object is a particular instance of a class which has actual existence and there can be many objects (or instances) for a class. Static variables and methods are not purely object oriented because they are not specific to instances (objects) but common to all instances.

**Objects:** It can represent a person, a bank account or any item that a program can handle. When a program is executed, the objects interact by sending messages to one another. For example, if 'customer' and 'account' are two objects in a program, then the customer object may send message to account object requesting for a bank balance. Each object contains data and code to manipulate data. Objects can interact without having to know details of each other's data or code. It is sufficient to know the type of massage accepted and the type of response returned by the objects

**Classes:** The entire set of data and code of an object can be made a user-defined data type with the help of a class. In fact objects are variables of type class. Once a class has been defined, we can create any number of objects associated with that class. For example, mango, apple and orange are members of class fruit. If fruit has been defined as a class, then the statement fruit mango, will create an object mango belonging to the class fruit.

A class can contain any of the following variable types.

- Local variables: Variables defined inside methods, constructors or blocks are called local variables. The variable will be declared and initialized within the method and the variable will be destroyed when the method has completed.
- Instance variables: Instance variables are variables within a class but outside any method. These variables are initialized when the class is instantiated. Instance variables can be accessed from inside any method, constructor or blocks of that particular class.
- Class variables: Class variables are variables declared within a class, outside any method, with the static keyword.

A class can have any number of methods to access the value of various kinds of methods. In the above example, barking(), hungry() and sleeping() are methods.

**Data Types**

Variables are nothing but reserved memory locations to store values. This means that when you create a variable you reserve some space in the memory.
Based on the data type of a variable, the operating system allocates memory and decides what can be stored in the reserved memory. Therefore, by assigning different datatypes to variables, you can store integers, decimals, or characters in these variables.
There are two data types available in Java:

- Primitive Datatypes
- Reference/Object Datatypes

There are 8 primitive data types: byte, short, int, long, float, double, char, Boolean
Primitive data are only single values; they have no special capabilities.

| **Common Primitive Types** | | |
|---|---|---|
| **Type** | **Description** | **Example of Literals** |
| Int | Integers (whole numbers) | 42, 60634, –8, 0 |
| Double | Real numbers | 0.039, –10.2, 4.2E+72 |
| Char | Single characters | ‘a’, ‘B’, ‘&’, ‘6’ |
| Boolean | Logical values | True, false |

| **Number** | | |
|---|---|---|
| **Type** | **Storage** | **Range of Values** |
| Byte | 8 bits | – 128 to 127 |
| Short | 16 bits | –32,768 to 32.727 |
| Int | 32 bits | –2,147,483,648 to 2,147,483,647 |
| Long | 64 bits | $-9\times10^{18}$ to $9 \times 10^{18}$ |
| Float | 32 bits | $\pm10^{-45}$ to $\pm10^{38}$,7 significant digits |
| Double | 64 bits | $\pm10^{-324}$ to ± 10308, 15 significant digits |

There are 8 primitive data types: byte, short, int, long, float, double, char, Boolean Primitive data are only single values; they have no special capabilities.

**Variables**

- A variable is a name for a location in memory used to store a data value.
- We use variables to save and restore values or the results of calculations.
- The programmer has to tell Java what type of data will be store in the variable’s memory location. Its type cannot change.
- During the program execution the data saved in the memory location can change; hence the term “variable”.

**Variable Declaration**

- Before you can use a variable, you must declare its type and name.
- You can declare a variable only once in a method.

**For Example:**

```
int numDimes;
double length;
char courseSection;
boolean done;
String lastName;
```

## 3.1 Encapsulation

Encapsulation is the process of wrapping up of data (properties) and behavior (methods) of an object into a single unit; and the unit here is a Class (or interface). Encapsulate in plain English means to enclose or be enclosed in or as if in a capsule. In Java, everything is enclosed within a class or interface, unlike languages such as C and C++ where we can have global variables outside classes. Encapsulation enables data hiding, hiding irrelevant information from the users of a class and exposing only the relevant details required by the user.

### 3.2 Setters & Getters

A setter is a method used to change the value of an attribute and a getter is a method used to get the value of an attribute. There is a standard naming convention for getters and setters, but Java compiler won't complain even otherwise.

```
private String name;
public String getName() {
return name;
}
public void setName(String name) {
this.name=name;
}
```

Here getName and setName are the getter and setter for the variable 'name' respectively. Since this is a standard naming convention, many IDEs like eclipse will generate it for you in this form.

### 3.3 Inheritance

Inheritance describes the parent child relationship between two classes. A class can get some of its characteristics from a parent class and then add more unique features of its own. For example, consider a Vehicle parent class and a child class Car. Vehicle class will have properties and functionalities common for all vehicles. Car will inherit those common properties from the Vehicle class and then add properties which are specific to a car. Vehicle parent class is known as base class or superclass. Car is known as derived class, Child class or subclass.

Inherited fields can be accessed just like other normal fields (even using "this" keyword). Java specification says that "A subclass does not inherit the private members of its parent class.". This means that subclass cannot access private member (using "this" keyword) from a subclass like other members. When you create an object, it will call its super constructor and the super class object is created with all fields including private; however, only inherited fields can be accessed. We can still access the private
variables of the parent class using an accessible parent method like a setter or getter.

```
public class Parent {
private int i;
public int j;
Parent(int i, int j) {
this.i = i;
this.j = j;
}
public static void main(String[] args) {
Child c = new Child(5, 10);
c.printIJ();
}
void printIJ() {
System.out.println(i + " " + j);
}
}
class Child extends Parent {
Child(int i, int j) {
super(i, j);
// this.i=25;
this.j = 15;
}
}
```

In general, Java supports single-parent, multiple-children inheritance and multilevel inheritence (Grandparent-> Parent -> Child) for classes and interfces. Java supports multiple inheritance (multiple parents, single child) only through interfaces. This is done to avoid some confusions and errors.

### 3.4 Abstraction

Abstraction in Java or Object-oriented programming is a way to segregate implementation from interface and one of the five fundamentals along with Encapsulation, Inheritance, Polymorphism, Class and Object.

- An essential component of object oriented programming is Abstraction
- Humans manage complexity through abstraction.
- For example, people do not think a car as a set of tens and thousands of individual parts. They think of it as a well-defined object with its own unique behavior.
- This abstraction allows people to use a car ignoring all details of how the engine, transmission and braking systems work.
- In computer programs the data from a traditional process oriented program can be transformed by abstraction into its component objects.
- A sequence of process steps can become a collection of messages between these objects. Thus, each object describes its own behavior.

### 3.5 Overriding

- In a class hierarchy when a sub class has the same name and type signature as a method in the super class, then the method in the subclass is said to override the method in the super class.
- When an overridden method is called from within a sub class, it will always refer to the version of that method defined by the sub class.
- The version of the method defined by the super class will be hidden.

### 3.6 Exceptions

- An exception is an abnormal condition that arises in a code sequence at run time.
- In other words, an exception is a run time error.
- A java exception is an object that describes an exceptional condition that has occurred in a piece of code.
- When an exceptional condition arises, an object representing that exception is created and thrown in the method that caused the error.
- Now the exception is caught and processed.

### 3.7 Polymorphism

The steering wheel (i.e., the interface) is the same no matter what type of actual steering mechanism is used. That is, the steering wheel works the same whether your car has manual steering, power steering, or rack-and-pinion steering. Therefore, once you know how to operate the steering wheel, you can drive any type of car. The same principle can also apply to programming. For example, consider a stack (which is a first-in, last-out list). You might have a program that requires three different types of stacks. One stack is used for integer values, one for floating-point values, and one for characters. In this case, the algorithm that implements each stack is the same, even though the data being stored differs. In a non-object-oriented language, you would be required to create three different sets of stack routines, with each set using different names. However, because of polymorphism, in Java you can create one general set of stack routines that works for all three specific situations. This way, once you know how to use one stack, you can use them all.

More generally, the concept of polymorphism is often expressed by the phrase "one interface, multiple methods." This means that it is possible to design a generic interface to a group of related activities. Polymorphism helps reduce complexity by allowing the same interface to be used to specify a general class of action. It is the compiler's job to select the specific action (i.e., method) as it applies to each situation. You, the programmer, don't need to do this selection manually. You need only remember and utilize the general interface.

When properly applied, polymorphism, encapsulation, and inheritance combine to produce a programming environment that supports the development of far more robust and scalable programs than does the process-oriented model. A well-designed hierarchy of classes is the basis for reusing the code in which you have invested time and effort developing and testing. Encapsulation allows you to migrate your implementations over time without breaking the code that depends on the public interface of your classes. Polymorphism allows you to create clean, sensible, readable, and resilient code.

# 4. Java Basics

Java was conceived by James gosling, Patrick Naughton, Chris Warth, Ed frank and Mike Sheridan at sun Microsystems.

### 4.1 Java's Byte code

The key that allows java to solve the both security and portability problems is that the output of a java compiler is not executable code rather it is byte code. Byte code is highly optimized set of instructions designed to be executed by java runtime systems, which is called Java Virtual Machine (JVM). JVM is interpreter for byte code. Translating a java program into byte code helps makes it much easier to run a Program in a wide variety of environments. The reason is straightforward: only the JVM needs to be implemented for each platform. Once the run-time package exists for a given system, any Java program can run on it.

### 4.2 Data Types

Java defines eight simple types of data: byte, short, int, long, char, float, double, and Boolean. These can be put in four groups:

- Integers this group includes byte, short, int, and long, which are for whole valued signed numbers.
- Floating-point numbers this group includes float and double, which represent numbers with fractional precision.
- Characters this group includes char, which represents symbols in a character set, like letters and numbers.
- Boolean this group includes Boolean, which is a special type for representing true/false values.

### 4.3 The Scope and Lifetime of Variables

All of the variables used till now have been declared at the start of the main( ) method. However, Java allows variables to be declared within any block. A block is begun with an opening curly brace and ended by a closing curly brace. A block defines a scope.

Thus, each time you start a new block, you are creating Scope determines what objects are visible to other parts of your program. It also determines the lifetime of those objects. Most other computer languages define two general categories of scopes: **global and local**. The scope defined by a method begins with its opening curly brace. However, if that method has parameters, they too are included within the method's scope. variables declared inside a scope are not visible (that is, accessible) to code that is defined outside that scope. Thus, when you declare a variable within a scope, you are localizing that variable and protecting it from unauthorized access and/or modification. Scopes can be nested. For example, each time you create a block of code, you are creating a new, nested scope. When this occurs, the outer scope encloses the inner scope. This means that objects declared in the outer scope will be visible to code within the inner scope. However, the reverse is not true. Objects declared within the inner scope will not be visible outside it.

### 4.4 Operators

Operators are special symbols that perform specific operations on one, two, or three operands, and then return a result. The operators in the following table are listed according to precedence order. The closer to the top of the table an operator appears, the higher its precedence. Operators with higher precedence are evaluated before operators with relatively lower precedence. Operators on the same line have equal precedence. When operators of equal precedence appear in the same expression, a rule must govern which is evaluated first. All binary operators except for the assignment operators are evaluated from left to right; assignment operators are evaluated right to left.

| **Operator** ***Precedence*** | |
|---|---|
| **Operators** | ***Precedence*** |
| Postfix | expr$^{++}$, *expr* - - |
| Unary | ++*expr* --*expr* +*expr* –*expr* ~ *!- !* |
| multiplicative | * / % |
| additive | + - |
| shift | < < > >  > > > |
| relational | < > <= >= instance of |
| equality | = = ! = |
| bitwise AND | & |
| bitwise exclusive OR | **^** |
| Bitwise inclusive OR | I |
| logical AND | && |

<table>
<tr><td>logical OR</td><td>Ⅱ</td></tr>
<tr><td>ternary</td><td>? :</td></tr>
<tr><td>assignment</td><td>= + =–=*=/=%=&=^= |=<⇔>=<br>>>>=</td></tr>
</table>

## 4.5 Classes and Objects

In the real world, you'll often find many individual objects all of the same kind. There may be thousands of other bicycles in existence, all of the same make and model. Each bicycle was built from the same set of blueprints and therefore contains the same components. In object-oriented terms, we say that your bicycle is an instance of the class of objects known as bicycles. A class is the blueprint from which individual objects are created.

**Constructors:** A class contains constructors that are invoked to create objects from the class blueprint. Constructor declarations look like method declarations—except that they use the name of the class and have no return type. A constructor initializes an object immediately upon creation. Constructors can be default or parameterized constructors. A default constructor is called when an instance is created for a class.

**For Example:**

```
Class demo
{
int x;
int y;
float z;
demo()
{
  X=1;
Y=2;
Z=3;
}
void display()
{
System.out.println(—values of x, y and z are:||+x+|| —+y||+|| —+z);
}
}
Class demomain
{
Public static void main(String args[])
{
demo d1=new demo(); // this is a call for the above default constructor
d1.display();
}
}
Parameterized constructor:
Class demo
{
int x;
int y;
float z;
demo(int x1,int y1,int z1)
{
  x=x1;
  y=y1;
  z=z1;
}
void display()
{
System.out.println(—values of x, y and z are:||+x+|| —+y||+|| —+z);
```

```
}
}
Class demomain
{
Public static void main(String args[])
{
demo d1=new demo(1,2,3); // this is a call for the above parameterized constructor
d1.display();
}
}
```

**This Keyword:** Sometimes a method will need to refer to the object that invoked it. To allow this, Java defines the this keyword. this can be used inside any method to refer to the current object. That is, this is always a reference to the object on which the method was invoked. You can use this anywhere a reference to an object of the current class' type is permitted.

***For Example:***

```
Class demo
{
int x;
int y;
float z;
demo(int x,int y,int z)
{
 this.x=x;
 this.y=y;
 this.z=z;
}
void display()
{
System.out.println(—values of x, y and z are:||+x+|| —+y||+|| —+z);
}
}
Class demomain
{
Public static void main(String args[])
{
demo d1=new demo(1,2,3); // this is a call for the above parameterized constructor
```

**Whitespace:** Java is a free-form language. This means that you do not need to follow any special indentation rules. For instance, the Example program could have been written all on one line or in any other strange way you felt like typing it, as long as there was at least one whitespace character between each token that was not already delineated by an operator or separator. In Java, whitespace is a space, tab, or newline.

**Identifiers:** Identifiers are used for class names, method names, and variable names. An identifier may be any descriptive sequence of uppercase and lowercase letters, numbers, or the underscore and dollar-sign characters. They must not begin with a number, lest they be confused with a numeric literal. Again, Java is case-sensitive, so VALUE is a different identifier than Value.

**Literals:** A constant value in Java is created by using a literal representation of it.

**Comments:** there are three types of comments defined by Java- single-line, multiline and documentation comment. The documentation comment begins with a /** and ends with a */

**Separators:** In Java, there are a few characters that are used as separators. The most commonly used separator in Java is the semicolon.

| Symbol | Name | Purpose |
|---|---|---|

| | | |
|---|---|---|
| ( ) | Parentheses | Used to contain lists of parameters in method definition and invocation. Also used for defining precedence in expressions, containing expressions in control statements, and surrounding cast types. |
| { } | Braces | Used to contain the values of automatically initialized arrays. Also used to define a block of code, for classes, methods, and local scopes. |
| [] | Brackets | Used to declare array types. Also used when dereferencing array values. |
| ; | Semicolon | Terminates statements. |
| . | Comma | Separates consecutive identifiers in a variable declaration. Also used to chain statements together inside a for statement. |
| . | Period | Used to separate package names from subpackages and classes. Also used to separate a variable or method from a reference variable. |

**The Java Keywords:** There are 50 keywords currently defined in the Java language. These keywords, combined with the syntax of the operators and separators, form the foundation of the Java language. These keywords cannot be used as names for a variable, class, or method.

| | | | | |
|---|---|---|---|---|
| abstract | continue | for | new | switch |
| assert | default | goto | package | synchronized |
| boolean | do | if | private | this |
| break | double | implements | protected | throw |
| byte | else | import | public | throws |
| case | enum | instanceof | return | transient |
| catch | extends | int | short | try |
| char | final | interface | static | void |
| class | finally | long | strictfp | volatile |
| const | float | native | super | while |

## 4.6 Control Statements

Java supports two selection statements: if and switch. These statements allow you to control the flow of your program's execution based upon conditions known only during run time.

The **if statement** is Java's conditional branch statement. It can be used to route program execution through two different paths. Here is the general form of the if statement:

```
if (condition) statement1;
else statement2;
```

Here, each statement may be a single statement or a compound statement enclosed in curly braces (that is, a block). The condition is any expression that returns a boolean value. The else clause is optional. The if works like this: If the condition is true, then statement1 is executed. Otherwise, statement2 (if it exists) is executed. In no case will both statements be executed. For example, consider the following:

```
int a, b;
// ...
if(a < b) a = 0;
else b = 0;
```

Here, if a is less than b, then a is set to zero. Otherwise, b is set to zero. In no case are they both set to zero.

**Nested ifs**

A nested if is an if statement that is the target of another if or else. Nested ifs are very common in programming. When you nest ifs, the main thing to remember is that an else statement always refers to the nearest if statement that is within the same block as the else and that is not already associated with an else. Here is an example:

```
if(i == 10) {
if(j < 20) a = b;
if(k > 100) c = d; // this if is
else a = c; // associated with this else
}
else a = d; // this else refers to if(i == 10)
```

**Switch**
The switch statement is Java's multiway branch statement. It provides an easy way to dispatch execution to different parts of your code based on the value of an expression. As such, it often provides a better alternative than a large series of if-else-if statements. Here is the general form of a switch statement:

```
switch (expression) {
case value1:
// statement sequence
break;
case value2:
// statement sequence
break;
.
.
.
case valueN:
// statement sequence
break;
default:
// default statement sequence
}
```

The switch statement works like this: The value of the expression is compared with each of the literal values in the case statements. If a match is found, the code sequence following that case statement is executed. If none of the constants matches the value of the expression, then the default statement is executed. However, the default statement is optional. If no case matches and no default is present, then no further action is taken. The break statement is used inside the switch to terminate a statement sequence. When a break statement is encountered, execution branches to the first line of code that follows the entire switch statement. This has the effect of "jumping out" of the switch.

## 4.7 Iteration Statements

Java's iteration statements are for, while, and do-while. These statements create what we commonly call loops. As you probably know, a loop repeatedly executes the same set of instructions until a termination condition is met.

**While**
The while loop is Java's most fundamental loop statement. It repeats a statement or block while its controlling expression is true. Here is its general form:

```
while(condition) {
// body of loop
}
```

The condition can be any Boolean expression. The body of the loop will be executed as long as the conditional expression is true. When condition becomes false, control passes to the next line of code immediately following the loop. The curly braces are unnecessary if only a single statement is being repeated.

```
Example:// Demonstrate the while loop.
class While {
public static void main(String args[]) {
int n = 10;
while(n > 0) {
System.out.println("tick " + n);
n--;
}
}
}
```

Since the while loop evaluates its conditional expression at the top of the loop, the body of the loop will not execute even once if the condition is false to begin with. For example, in
the following fragment, the call to println( ) is never executed:

```
int a = 10, b = 20;
while(a > b)
System.out.println("This will not be displayed");
```

**do-while**

There are times when you would like to test the termination expression at the end of the loop rather than at the beginning. Fortunately, Java supplies a loop that does just that: the do-while. The do-while loop always executes its body at least once, because its conditional expression is at the bottom of the loop. Its general form is

```
do {
// body of loop
} while (condition);
```

Each iteration of the do-while loop first executes the body of the loop and then evaluates the conditional expression. If this expression is true, the loop will repeat. Otherwise, the loop terminates. As with all of Java's loops, condition must be a Boolean expression. The do-while loop is especially useful when you process a menu selection, because you will usually want the body of a menu loop to execute at least once. In a program, the do-while loop is used to verify that the user has entered a valid choice.

**for**

Here is the general form of the traditional for statement:

```
for(initialization; condition; iteration) {
// body
}
```

If only one statement is being repeated, there is no need for the curly braces. The for loop operates as follows. When the loop first starts, the initialization portion of the loop is executed. Generally, this is an expression that sets the value of the loop control variable, which acts as a counter that controls the loop. It is important to understand that the initialization expression is only executed once. Next, condition is evaluated. This must be a Boolean expression. It usually tests the loop control variable against a target value. If this expression is true, then the body of the loop is executed. If it is false, the loop terminates.

Next, the iteration portion of the loop is executed. This is usually an expression that increments or decrements the loop control variable. The loop then iterates, first evaluating the conditional expression, then executing the body of the loop, and then executing the iteration expression with each pass. This process repeats until the controlling expression is false.

## 4.8 Recursion

Recursion Java supports recursion. Recursion is the process of defining something in terms of itself. As it relates to Java programming, recursion is the attribute that allows a method to call itself. A method that calls itself is said to be recursive. The classic example of recursion is the computation of the factorial of a number. The factorial of a number N is the product of all the whole numbers between 1 and N. For example, 3 factorial is 1 × 2 × 3, or 6. Here is how a factorial can be computed by use of a recursive method.

```
// A simple example of recursion.
class Factorial {
// this is a recursive method
int fact(int n) {
int result;
if(n==1) return 1;
result = fact(n-1) * n;
return result;
}
}
class Recursion {
public static void main(String args[]) {
Factorial f = new Factorial();
System.out.println("Factorial of 3 is " + f.fact(3));
System.out.println("Factorial of 4 is " + f.fact(4));
System.out.println("Factorial of 5 is " + f.fact(5));
}
```

```
}
```

**The output from this program is shown here:**
Factorial of 3 is 6
Factorial of 4 is 24
Factorial of 5 is 120

## 4.9 Access Control

Encapsulation links data with the code that manipulates it. However, encapsulation provides another important attribute: access control. Through encapsulation, you can control what parts of a program can access the members of a class. By controlling access, you can prevent misuse. For example, allowing access to data only through a well defined set of methods, you can prevent the misuse of that data. Thus, when correctly implemented, a class creates a "black box" which may be used, but the inner workings of which are not open to tampering. Java's access specifiers are public, private, and protected. Java also defines a default access level. protected applies only when inheritance is involved.

When a member of a class is modified by the **public** specifier, then that member can be accessed by any other code. When a member of a class is specified as **private**, then that member can only be accessed by other members of its class. Now you can understand why main( ) has always been preceded by the public specifier. It is called by code that is outside the program—that is, by the Java run-time system. When no access specifier is used, then by default the member of a class is public within its own package, but cannot be accessed outside of its package.

**Static:** There will be times when you will want to define a class member that will be used independently of any object of that class. Normally, a class member must be accessed only in conjunction with an object of its class. However, it is possible to create a member that can be used by itself, without reference to a specific instance. To create such a member, precede its declaration with the keyword static. When a member is declared static, it can be accessed before any objects of its class are created, and without reference to any object. Instance variables declared as static are, essentially, global variables. When objects of its class are declared, no copy of a static variable is made. Instead, all instances of the class share the same static variable.

**Final:**
A variable can be declared as final. Doing so prevents its contents from being modified. This means that you must initialize a final variable when it is declared. For example:
final int FILE_NEW = 1;
final int FILE_OPEN = 2;
final int FILE_SAVE = 3;
final int FILE_SAVEAS = 4;
final int FILE_QUIT = 5;

Subsequent parts of your program can now use FILE_OPEN, etc., as if they were constants, without fear that a value has been changed. It is a common coding convention to choose all uppercase identifiers for final variables. Variables declared as final do not occupy memory on a per-instance basis. Thus, a final variable is essentially a constant

## 4.10 Exception Handling

All exception types are subclasses of the built-in class Throwable. Thus, Throwable is at the top of the exception class hierarchy. Immediately below Throwable are two subclasses that partition exceptions into two distinct branches. One branch is headed by Exception. This class is used for exceptional conditions that user programs should catch. This is also the class that you will subclass to create your own custom exception types. There is an important subclass of Exception, called Runtime Exception. Exceptions of this type are automatically defined for the programs that you write and include things such as division by zero and invalid array indexing.

The other branch is topped by Error, which defines exceptions that are not expected to be caught under normal circumstances by your program. Exceptions of type Error are used by the Java run-time system to indicate errors having to do with the run-time environment, itself. Stack overflow is an example of such an error. This chapter will not be dealing with exceptions of type Error, because these are typically created in response to catastrophic failures that cannot usually be handled by your program.

## 4.11 Java Applets

An applet is a special kind of Java program that is designed to be transmitted over the Internet and automatically executed by a Java-compatible web browser. Furthermore, an applet is downloaded on demand, without further interaction with the user. If the user clicks a link that contains an applet, the applet will be automatically downloaded and run in the browser. Applets are intended to be small programs. They are typically used to display data provided by the server, handle user input, or provide simple functions, such as a loan calculator, that execute locally, rather than on the server. In essence, the applet allows some functionality to be moved from the server to the client.

- An applet program is a written as an inheritance of the java.Applet class
- There is no main() method in an Applet.
- An applet uses AWT for graphics.

**Life Cycle of an Applet:**

- init: This method is intended for whatever initialization is needed for an applet.
- start: This method is automatically called after init method. It is also called whenever user returns to the page containing the applet after visiting other pages.
- stop: This method is automatically called whenever the user moves away from the page containing applets. This method can be used to stop an animation.
- destroy: This method is only called when the browser shuts down normally.

## PRACTICE SET

1. Which of the following is true about NULL pointer?
(a) A pointer which points nothing.
(b) A pointer which points a single value
(c) Both (a) and (b)
(d) A pointer which points a double value
(e) None of these

2. Which of the following is not a valid variable name declaration?
(a) int __a3; (b) int __3a;
(c) int __A3; (d) None of these
(e) Both (a) and (c)

3. All keywords in C are in
(a) Lower Case letters (b) Upper Case letters
(c) Camel Case letters (d) None of these
(e) Both (a) and (b)

4. Which of the following is not a valid C variable name?
(a) int number; (b) float rate;
(c) int variable_count; (d) int $main;
(e) None of these

5. The format identifier '%i' is also used for _____ data type?
(a) char (b) int
(c) float (d) double
(e) None of these

6. What is the size of an int data type?
(a) 4 Bytes (b) 8 Bytes
(c) Depends on the system/compiler
(d) Cannot be determined (e) None of these

7. Which of the following will not return a value?
(a) null (b) void
(c) empty (d) free
(e) None of these

8. What does the following statement mean?
void a;
(a) variable a is of type void
(b) a is an object of type void
(c) declares a variable with value a
(d) flags an error
(e) None of these

9. The name of the variable used in one function cannot be used in another function
(a) True (b) False
(c) Either (a) or (b) (d) None of these
(e) It will show compile time error

10. Which is correct with respect to size of the data types?
(a) char > int > float (b) int > char > float
(c) char < int < double (d) double > char > int
(e) None of these

11. Which of these values can a Boolean variable contain?
(a) True & False (b) 0 & 1
(c) Any integer value (d) Only True
(e) None of these

12. What is the process of defining a method in terms of itself, that is a method that calls itself?
(a) Polymorphism (b) Abstraction
(c) Encapsulation (d) Recursion
(e) None of these

13. Which of these keywords is used to make a class?
(a) class (b) struct
(c) int (d) constructor
(e) abstract

14. Which of these operators is used to allocate memory for an object?
(a) malloc (b) alloc
(c) new (d) give

(e) None of these

15. Which keyword can be used for coming out of recursion?
(a) break (b) return
(c) exit (d) Both break and return
(e) None of these
16. int[ ] ={5,6,7,8,9} What is the value of a[3]?
(a) 9 (b) 8
(c) 7 (d) 6
(e) None of these
17. Array is a..
(a) Both (b) and (c)
(b) Pointer data type
(c) Heterogeneous data type
(d) Homogeneous data type
(e) None of these
18. Which symbol is used as a statement terminator in C?
(a) ! (b) ~
(c) # (d) ;
(e) None of these
19. If the size of the array is less than the number of initializes then, ……….
(a) extra values are being ignored
(b) generates an error message
(c) size of array is increased
(d) size is neglected when values are given
(e) None of these
20. A compiler …………….
(a) Is a single computer program
(b) Translates a high level language into machine language
(c) Is a part of software and hardware
(d) Editor
(e) None of these
21. Which of the following is not a type of constructor?
(a) Copy constructor (b) Friend constructor
(c) Default constructor
(d) Parameterized constructor
(e) None of these
22. Which one is a valid declaration of a boolean?
(a) boolean b1 = 1; (b) boolean b2 = 'false';
(c) boolean b3 = 0; (d) boolean b4 = 'true'
(e) None of these
23. Input/output function prototypes and macros are defined in which header file?
(a) conio.h (b) stdlib.h
(c) stdio.h (d) dos.h
(e) None of these
24. What does the following declaration mean?
int(*ptr)[10].
(a) ptr is array of pointers to 10 integers.
(b) ptr is a pointer to an array of 10 integers
(c) ptr is an array of 10 integers
(d) ptr is an pointer to array
(e) None of these
25. The constructor and the destructor of a class are automatically invoked when memory is allocated and deallocated to an ….
(a) Data type (b) Copy constructor
(c) Object (d) None of the above
(e) Both (a) and (b)
26. A destructor is used to destroy the objects that have been created by a ………………..
(a) object (b) class
(c) function (d) constructor
(e) None of these
27. When the base class is publicly inherited, public members of the base class become …………. of the derived class.
(a) private members (b) protected members
(c) Public members (d) Not inherited
(e) None of these
28. Which of these keyword must be used to inherit a class?
(a) super (b) this
(c) extent (d) extends
(e) None of these
29. Which of these keywords is used to refer to member of base class from a sub class?
(a) upper (b) super
(c) this (d) None of these
(e) Both (a) and (b)
30. Which of these is supported by method overriding in Java?
(a) Abstraction (b) Encapsulation
(c) Polymorphism (d) None of these
(e) Both (a) and (b)
31. Which of these keywords can be used to prevent Method overriding?
(a) static (b) constant
(c) protected (d) final
(e) None of these
32. What is cfront in C++ ?
(a) is the front end of a C compiler
(b) is the pre-processor of a C compiler
(c) is a tool that translates a C++ code to its equivalent C code
(d) none of the above
(e) Both (a) and (c)
33. Which of the following are procedural languages?
(a) Pascal (b) Smalltalk
(c) C (d) Both (a) and (c)
(e) None of these
34. Which of the following concepts means wrapping up of data and functions together?
(a) Abstraction (b) Encapsulation
(c) Inheritance (d) Polymorphism
(e) None of these
35. Which of the following mechanisms is/are provided by Object Oriented Language to implement Object Oriented Model?
(a) Encapsulation (b) Inheritance
(c) Polymorphism (d) All of the above

(e) None of these

36. We can convert numeric string to primitive number using ……………………..
(a) parsing methods (b) wrapper classes
(c) constructor methods (d) abstract classes
(e) None of these

37. A package is a collection of
(a) Classes (b) interfaces
(c) editing tools (d) classes and interfaces
(e) None of these

38. Which of the following is true about multithreading in Java?
(a) Multithreading in java is a process of executing multiple threads simultaneously.
(b) Thread is basically a lightweight sub-process, a smallest unit of processing.
(c) Java Multithreading is mostly used in games, animation etc.
(d) None of these
(e) All (a), (b), and (c)

39. Which of the following is not an Advantage of Garbage Collection in java?
(a) It makes java memory efficient because garbage collector removes the unreferenced objects from heap memory.
(b) It is automatically done by the garbage collector (a part of JVM) so we don't need to make extra efforts.
(c) Both (a) and (b) (d) None of these
(e) Either (a) or (b)

40. Which of the following keywords are used to control access to a class member?
(a) default (b) break
(c) protected (d) goto
(e) None of these

41. What will be output of the following C program?
```
#include int xyz=10;
int main()
{
int xyz=20;
printf (“%d”, xyz);
return 0;
}
```
(a) 10 (b) 20
(c) 30 (d) compilation error
(e) None of these

42. If a=8 and b=15 then the statement x=(a>b) ? a:b;
(a) assigns a value 8 to x
(b) gives an error message
(c) assigns a value 15 to x
(d) assigns a value 7 to x
(e) None of these

43. What is the output of this program?
```
final class A {
        int i;
 }
 class B extends A {
      int j;
      System.out.println (j + " " + i);
 }
 class inheritance {
      public static void main(String args[])
      {
            B obj = new B();
            obj.display();
      }
}
```
(a) 2 2 (b) 3 3
(c) Runtime Error (d) Compilation Error
(e) None of these

44. What is the output of this program?
```
class Alligator
 {
  public static void main(String[] args)
   {
   int []x[] = {{1,2}, {3,4,5}, {6,7,8,9}};
   int [][]y = x;
   System.out.println (y[2][1]);
   }
 }
```
(a) 2 (b) 3
(c) 7 (d) Compilation Error
(e) None of these

45. The output of the code below is
```
#include <stdio.h>
void main()
{
      int x = 0;
      if (x == 0)
            printf("hi");
      else
            printf("how are u");
            printf("hello");
}
```
(a) hi (b) how are you
(c) hello (d) hihello
(e) None of these

46. Which of the following is FALSE about arrays on Java?
(a) A java array is always an object
(b) Length of array can be changed after creation of array
(c) Arrays in Java are always allocated on heap
(d) None of these
(e) Both (a) and (c)

47. What is the use of final keyword in Java?
(a) When a class is made final, a subclass of it cannot be created.
(b) When a method is final, it cannot be overridden.
(c) When a variable is final, it can be assigned value only once.
(d) None of these
(e) All (a), (b), and (c)

48. What will be the output of following Java program?
```
class Main {
  public static void main(String args[]){
```

```
    final int i;
    i = 20;
    System.out.println(i);
  }
}
```

(a) 20 (b) Compiler time error
(c) 0 (d) None of these
(e) Run time error

49. In C, parameters are always
(a) Passed by value
(b) Passed by reference
(c) Non-pointer variables are passed by value and pointers are passed by reference
(d) Passed by value result
(e) None of these

50. Which of the following is true about FILE *fp ?
(a) FILE is a keyword in C for representing files and fp is a variable of FILE type.
(b) FILE is a structure and fp is a pointer to the structure of FILE type
(c) FILE is a stream
(d) FILE is a buffered stream
(e) File Handling

51. Which of the data types have size that is variable?
(a) int (b) struct
(c) float (d) double
(e) None of these

52. What will happen when the structure is declared?
(a) it will not allocate any memory
(b) it will allocate the memory
(c) it will be declared and initialized
(d) none of these
(e) Both (a) and (b)

53. The data elements in structure are also known as what?
(a) objects (b) members
(c) data (d) None of these
(e) Both (a) and (c)

54. Which of these keywords is used to define packages in Java?
(a) pkg (b) Pkg
(c) package (d) Package
(e) None of these

55. Which of the following is correct way of importing an entire package 'pkg'?
(a) import pkg. (b) Import pkg.
(c) import pkg.* (d) Import pkg.*
(e) None of these

56. Which of the following is correct way of implementing an interface salary by class manager?
(a) class manager extends salary {}
(b) class manager implements salary {}
(c) class manager imports salary {}
(d) None of these
(e) Both (a) and (b)

57. In which stage the following code #include<stdio.h> gets replaced by the contents of the file stdio.h
(a)During Preprocessing (b) During Execution
(c) During linking (d) During Editing
(e) None of these

58. An array elements are always stored in ________ memory locations.
(a) Sequential (b) Random
(c) Sequential and Random (d) None of the above
(e) Both (a) and (b)

59. What is function?
(a) Function is a block of statements that perform some specific task.
(b) Function is the fundamental modular unit. A function is usually designed to perform a specific task.
(c) Function is a block of code that performs a specific task. It has a name and it is reusable.
(d) All (a), (b), and (c)
(e) None of the above

60. Which of the following correctly accesses the seventh element stored in arr, an array with 100 elements?
(a) arr[6] (b) arr[7]
(c) arr [5] (d) arr [8]
(e) None of these

## SOLUTIONS

1. (a)
2. (d)
3. (a)
4. (d)
5. (b); Both %d and %i can be used as a format identifier for int data type.
6. (c); The size of the data types depend on the system.
7. (b)
8. (d); There are no void objects.
9. (b)
10. (c)
11. (a)
12. (d)
13. (a)
14. (c)
15. (b)
16. (b)
17. (d)
18. (d)
19. (b)
20. (b)
21. (b)
22. (b)
23. (c)
24. (b)
25. (c)
26. (d)
27. (c)

28. (d)
29. (b)
30. (c)
31. (d)
32. (c)
33. (d); Procedural languages sequentially execute a set of imperative statements to achieve the desired effect.
34. (b)
35. (d)
36. (a)
37. (d)
38. (e)
39. (c)
40. (c)
41. (b)
42. (c)
43. (d); class A has been declared final hence it cannot be inherited by any other class. Hence class B does not have member i, giving compilation error.
44. (c)
45. (d)
46. (b); In Java, arrays are objects, they have members like length. The length member is final and cannot be changed. All objects are allocated on heap in Java, so arrays are also allocated on heap.
47. (e)
48. (a)
49. (a); In C, function parameters are always passed by value. Pass-by-reference is simulated in C by explicitly passing pointer values.
50. (b)
51. (b); Since the size of the structure depends on its fields, it has a variable size.
52. (a); While the structure is declared, it will not be initialized, So it will not allocate any memory.
53. (b);
54. (c)
55. (c)
56. (b)
57. (a)
58. (a)
59. (d)
60. (a)

# Professional Knowledge Practice Set: 01

1. Which SQL command is used to delete complete table from the database?
(a) Delete (b) Truncate (c) Drop
(d) Remove (e) Select
2. Bit stuffing technique used in which method?
(a) Checking error (b) flow control (c) framing
(d) Route finding (e) Update Anomaly
3. The phase of the SDLC in which an information system is systematically repaired and improved is referred to as________
(a) Analysis (b) Implementation
(c) Maintenance (d) Testing (e) Coding
4. What is loopback address?
(a) 127.0.0.1 (b) 255.0.0.0 (c) 255.255.0.0
(d) 127.0.0.0 (e) 111.0.0.128
5. Congestion Control is done by which layer?
(a) Data link Layer (b) Network Layer
(c) Transport Layer (d) Application Layer
(e) Presentation Layer
6. Which of the following is a form of virus explicitly designed to hide itself from detection by antivirus software.
(a)Stealth Virus (b)Macro Virus
(c)Polymorphic Virus
(d) Parasitic Virus (e)Both (c) and (d)
7. Which of the following variable is used to hold the address of another object?
(a) Integer (b) Pointer (c) Constant
(d) Memory Variable (e) Interrupt
8. Alpha testing is done on which side of software development?
(a) Customer side (b) Developer side
(c) Both Customer side and developer side
(d) Server side (e) None of these
9. What is Dirty Bit in operating system?
(a) Page with corrupted data
(b) Wrong page in the memory
(c) Page that is modified after being loaded into cache memory
(d) Page that is less frequently accessed
(e) Page contain high memory location
10. Which cloud computing service is provided by Microsoft?
(a)Azure (b)Elastic (c)Smart
(d)Grid (e)Coral
11. Shell is the exclusive feature of:
(a) UNIX (b) DOS (c) VMWare
(d) Application software (e) Utility Software
12. Which of the following is a keyword used for a storage class in C programming?
(a) printf (b) external (c) auto
(d) scanf (e) readf
13. Ethernet uses a ____________ physical address that is imprinted on the network interface card.
(a) 64 bit (b) 68 byte (c) 6 bit
(d) 6 byte (e) 32 bit
14. Which of the following is the time interval between the submission and completion of job?
(a) Waiting time (b) Turnaround time
(c) Throughput (d) Response time
(e) Output time
15. Which of the following is used to convert infix notation to postfix notation in data structure?
(a) Branch (b) Queue (c) Tree
(d) Stack (e) List
16. Which of the memories has the shortest access time?
(a) Cache memory (b) Magnetic bubble memory
(c) Magnetic core memory
(d) RAM (e) Optical
17. Foreign key can take which type of value?
(a) Same value as the primary key it refers
(b) Any New value
(c) NULL value
(d) Value which is different from primary key
(e) All the values of a relation
18. What is the process of defining a method in terms of itself, that is a method that calls itself?
(a) Polymorphism (b) Abstraction
(c) Encapsulation (d) Recursion (e) Inheritance
19. Which of the following testing method is used to test the software without knowing the internal structure of code or program?
(a)White box Testing (b)Alpha testing
(c)Black box Testing (d)Grey box testing
(e)Beta testing
20. What is the file extension of compiled java program?
(a) (.class) (b) (.java) (c) (.css)
(d) (.html) (e) (.xml)
21. Which of the following is an authorization command in SQL?
(a) Access (b) Allow (c) Grant
(d)Revoke (e) Permission
22. What is the use of HTTP protocol?
(a) Information access
(b) Telnet (c) E- mail
(d) FTP (e) Routing
23. Which tool is use for structured designing?
(a) Program flowchart (b) Structure chart
(c) Data-flow diagram (d) Module
(e) Design Manual
24. Which of the following is true for CIDR (Classless Inter Domain Routing) ?

(a) It is used in class C Networks
(b) It is used in class B Networks
(c) It is used in class A Networks
(d) All of the above (e) None of these

25. What is the use of "AS" Clause in SQL?
(a) Selection operation (b) Rename operation
(c) Join operation (d) Projection operation
(e) Union operation

26. Which of the following is also known as PC Program Counter?
(a)Instruction pointer (b)Memory pointer
(c)Data counter (d)File pointer
(e) Comparison counter

27. Which C++ which among the following OOP feature is related to re-usability?
(a) Encapsulation (b) Inheritance
(c) Abstraction (d) Polymorphism
(e) Overloading

28. Dequeue also called which of the following terminology?
(a) Double ended queue (b) Single ended queue
(c) A Stack (d) Push
(e) Pop

29. Which of the following is use of COCOMO Model in Software Engineering?
(a) Finding cost of the project
(b) Design software (c) Scheduling software
(d) Time complexity of the project
(e) Maintenance of the project

30. Which of the following is true in any C program?
(a) must contain at least one function
(b) need not contain any function
(c) needs input data
(d) both (A) and (B) (e) None of these

31. Pretty good privacy (PGP) is used in
(a) browser security (b) email security
(c) FTP security (d) IMAP
(e) none of the mentioned

32. The schema at the intermediate level according to the level of abstraction in DBMS called____.
(a)Conceptual Schema
(b)Logical Schema (c)Physical Schema
(d)Sub schema (e)Super Schema

33. OLAP applications are widely used by________
(a)RDBMS (b)JAVA
(c)Data Mining Technique
(d)Operating System (e) Software designing

34. What is an applet?
(a) An applet is a Java program that runs in a Web browser.
(b) Applet is a standalone java program.
(c) Applet is a tool.
(d) Applet is a run time environment.
(e) None of these

35. Which of the following company developed core banking solution software "**BaNCS**"?
(a) TCS (b) Infosys (c) IBM
(d) Oracle (e) Microsoft

36. Which of the following is not proper state of transaction?
(a)Partially aborted (b)Partially committed
(c)Aborted (d)Committed
(e)Active

37. 803.11 is use in which technology?
(a) Wireless LAN (b) ARPANET
(c) ETHERNET (d) BITNET (e) Token Ring

38. Which of the following algorithms solves the all pair shortest path problem?
(a) Dijkstra's algorithm
(b) Floyd's algorithm (c) Prims algorithm
(d) Wound Die and Wound Wait
(e) Topological Graph

39. A switch uses _______ address of machine to transfer data.
(a) MAC (b) IP (c) Header
(d) Both (a) & (b) (e) HTTP

40. Which of the following protocols is used to monitor network devices such as hubs, switches, and routes?
(a) SMTP (b) RIP (c) SNMP
(d) OSPF (e) IMAP

41. SRS in software engineering stands for________
(a) System recourses Specification
(b) System Recursive specification
(c) System requirement specification
(d) Software requirement specification
(e) Strategic Resource System

42. The number of attributes in a relation is called :
(a) Table (b) Domain (c) Degree
(d) Column (e) Tuple

43. Which of the following looks like a real program but it is actually malicious and designed to cause harm your system?
(a) Worm (b) Stealth virus (c) Trojan
(d) Virus (e) All are these

44. Which of the following is the average and worst case complexity of bubble sort?
(a) $O(n^2)$ (b) O(n) (c) O(nlogn)
(d) O(1) (e) O(n-1)

45. What is KPA in cryptanalysis?
(a) Known Phase Adjective
(b) Known Plaintext Attack
(c) Known Popular Analysis
(d) Key Phase Adjective (e) Key Popular Analysis

46. IPsec uses standards suite protocol that provides integrity, authentication and security. It is standardised by which organisation?
(a) IETE (b) EMFT (c) IKE
(d) ISA (e) ISE

47. Kerberos uses which cryptographic technique for authentication?
(a) Asymmetric (b) Symmetric (c) Dissymmetric
(d) Plain (e) System

48. Which of the following is not a class A host address?
(a) 128.4.5.6 (b) 117.4.5.1 (c) 117.0.0.0
(d) 117.8.0.0 (e) 111.4.3.1

49. Queue is which type of data structure?

(a) Linear data structure
(b) Nonlinear data structure
(c) Random access data structure
(d) Self-sorted data structure
(e) Abstract data structure

50. In which format records are arranged in Hierarchical Database Model?
(a)Array (b)List (c)Links
(d)Tree (e)Queue

51. Databases that support OLTP are most often referred to as _____.
(a) operational databases
(b) relational databases (c) data warehouses
(d) data dictionaries (e) network databases

52. Which of these operators is used to allocate memory for an object?
(a) malloc (b) alloc (c) new
(d) give (e) start

53. Which protocol is used to eliminate loops?
(a) TCP/IP (b) Spanning Tree Protocol
(c) SMTP (d) UDP (e) FTP

54. Which of the following sorting methods will be best if number of swapping done, is the only measure of efficiency?
(a) Bubble sort (b) Quick sort (c) Merge sort
(d) Selection sort (e) Radix Sort

55. WPA is a type of security used for what type of connection?
(a) Ethernet (b) Wi-Fi (c) Bluetooth
(d) Topology (e) Remote

56. Object DBMS add database functionality to_______ programming language.
(a) Object (b) Row (c) Data
(d) Field (e) Relational

57. UDP Protocol is defined at which layer?
(a) Transport layer of OSI Model
(b) Network layer of OSI Model
(c) Data link layer of OSI Model
(d) Application layer OSI Model
(e) Presentation Layer

58. A circular list can be used to represent:
(a) A queue (b) B-tree (c) B++ tree
(d) Stack (e) Recursion

59. Which of the following are the two mode of IP security?
(a) Certificate and tunnel
(b) Certificate and transport(c) Transport and tunnel
(d) Security and phases (e) Security and Tunnel

60. ______________ allows to create complex entities from existing entities and relationships.
(a) Specialization (b) Generalization
(c) Aggregation (d) Inheritance
(e) None of these

## SOLUTIONS

1. (c); The SQL DROP command is used to remove an object from the database. If you drop a table, all the rows in the table is deleted and the table structure is removed from the database.
2. (c); Bit stuffing is the insertion of one or more bits into a transmission unit as a way to provide signaling information to a receiver. The receiver knows how to detect and remove or disregard the stuffed bits. This technique is used in framing.
3. (c); The Maintenance Phase occurs once the system is operational. It includes implementation of changes that software might undergo over a period of time, or implementation of new requirements after the software is deployed at the customer location.
4. (a); 127.0.0.1 is the loopback address in IP. Message sent to 127.0.0.1 do not appear in the network . The network adapter loopback (send back) all the messages to the sending application. Loopback address is used to test the self connectivity.
5. (b); **Congestion** is a state occurring in network layer when the message traffic is so heavy that it slows down network response time. Congestion control is done by network layer by **Leaky Bucket Algorithm** and **Token Bucket Algorithm**.
6. (a); A **stealth virus** is a hidden computer **virus** that attacks operating system processes and averts typical anti-**virus** or anti-malware scans. **Stealth viruses** hide in files, partitions and boot sectors and are adept at deliberately avoiding detection.
7. (b); A pointer is a variable which contains the address in memory of another variable.
8. (b); Alpha testing takes place at the developer's site by the internal teams, before release to external customers. Alpha testing performed to identify all possible ***issues/bugs*** before releasing the product to everyday users or public.
9. (c); A dirty bit is a flag that indicates whether an attribute needs to be updated. Such situations usually occur when a bit in a memory cache or virtual memory page that has been changed by a processor but has not been updated in the storage.
10. (a); **Microsoft Azure** is a cloud computing service created by Microsoft for building, testing, deploying, and managing applications and services through a global network of Microsoft-managed data centers.
11. (a); Shell is a UNIX term for the interactive user interface with an operating system. The shell is the layer of programming that understands and executes the commands a user enters.
12. (c); Auto is the default storage class for all the variables declared inside a function or a block.
13. (d); Ethernet uses a 6 -byte (48-bit) physical address that is imprinted on the network interface card (NIC). The physical address also known as the link address is the address of a node physical address, also known as the link address, is the address of a node as defined by its LAN or WAN. It is included in the frame used by the data link layer.

14. (b); Turnaround time is the time interval from the time of submission of a process to the time of the completion of the process.
15. (d); Stack is used to expression conversion. There are three type of expression- infix, postfix, and prefix. Stack converts one type of expression to another like infix to postfix, infix to prefix, postfix to prefix and vice versa.
16. (a); Cache memory is a small-sized type of volatile computer memory that provides high-speed data access to a processor and stores frequently used computer programs, applications and data. Cache is the fastest memory in a computer.
17. (a); A **foreign key** is a column or group of columns in a relational database table that provides a link between data in two tables. It acts as a cross-reference between tables because it references the primary **key** of another table, thereby establishing a link between them.
18. (d); Recursion is a process in which a function calls itself as a subroutine. This allows the function to be repeated several times, since it calls itself during its execution.
19. (c); **Black-box testing** is a method of software testing that examines the functionality of an application without peering into its internal structures or workings.
20. (a); (.class) is the Compiled java source code file, and (.java) is the extension of source code of a java program.
21. (c); SQL GRANT is a command used to provide access or privileges on the database objects to the users.
22. (a); HTTP means Hyper Text Transfer Protocol. HTTP is the underlying protocol used by the World Wide Web and this protocol defines how messages are formatted and transmitted, and what actions Web servers and browsers should take in response to various commands.
23. (b); A Structure Chart (SC) in software engineering and organizational theory, is a chart which shows the breakdown of a system to its lowest manageable levels.
24. (d); Short for **C***lassless* **I***nter*-**D***omain* **R***outing,* an IP addressing scheme that replaces the older system based on classes A, B, and C. With CIDR, a single IP address can be used to designate many unique IP addresses. A CIDR IP address looks like a normal IP address except that it ends with a slash followed by a number, called the *IP network prefix.* For example: 172.200.0.0/16
25. (b); "AS" clause is used for rename operation in SQL.
26. (a); A **program counter** is a register in a computer processor that contains the address (location) of the instruction being executed at the current time. It is also known Instruction Pointer.
27. (b); Inheritance feature is used for concept of code reusability as, in inheritance a class can inherit properties and function of existing well written class.
28. (a); A **deque**, also known as a double-ended queue, is an ordered collection of items similar to the queue. It has two ends, a front and a rear, and the items remain positioned in the collection. In deque, new items can be added at either the front or the rear. Likewise, existing items can be removed from either end.
29. (a); The Constructive Cost Model (COCOMO) is the most common and widely **used** cost estimation **models** for most **software** projects.
30. (a); Any C program contains at least one function. If a program contains only one function, it must be main(). If a C program contains more than one function, then one (and only one) of these functions must be main(), because program execution always begins with main().
31. (b); Pretty Good Privacy (PGP) is an encryption program that provides cryptographic privacy and authentication for data communication. PGP is used for signing, encrypting, and decrypting texts, e-mails, files, directories, and whole disk partitions and to increase the security of e-mail communications.
32. (a); Schema at the intermediate level is called Conceptual Schema
33. (c); OLAP (***O****nline* ***A****nalytical* ***P****rocessing) is* a category of software tools that provides analysis of data stored in a database. OLAP tools enable users to analyze different dimensions of multidimensional data. For example, it provides time series and trend analysis views. OLAP often is used in data mining.
34. (a); An **applet** is a Java program that runs in a Web browser. Applets are small Java applications that can be accessed on an Internet server, transported over Internet, and can be automatically installed and run as apart of a web document.
35. (a); **BaNCS** core banking solution is designed by Tata Consultancy Services. This is a very popular core banking platform across the world. BaNCS is used by over 280 institutions in more than 80 countries. Here is a list of banks that use BaNCS Software Packages- State Bank of India, Allahabad Bank, Bank of Maharashtra, Central Bank of India, Indian Bank.
36. (a); **Transaction** is a logical unit of work that represents real-world events of any organization or an enterprise whereas concurrency control is the management of concurrent transaction execution. States of transaction are- Active, partially committed, failed, aborted, committed.
37. (a); 802.11 and *802.11x* refers to a family of specifications developed by the IEEE for *wireless* ***LAN*** (WLAN) technology.

802.11 specifies an over-the-air interface between a wireless client and a base station or between two wireless clients.

38. (b); Floyd-Warshall algorithm is a procedure, which is used to find the shortest path among all pairs of nodes in a graph, which does not contain any cycles of negative length. The main advantage of Floyd Warshall algorithm is its simplicity.
39. (a); Switch work on data link layer and it use MAC Address of machine to transfer data. A media access control **address** (**MAC address**) of a computer is a unique identifier assigned to network interfaces for communications at the data link layer of a network segment.
40. (c); Simple Network Management Protocol (SNMP) is a popular protocol for network management. It is used for collecting information from, and configuring, network devices, such as servers, printers, hubs, switches, and routers on an Internet Protocol (IP) network.
41. (d); A software requirements specification (SRS) is a description of a software system to be developed. It lays out functional and non-functional requirements, and may include a set of use cases that describe user interactions that the software must provide.
42. (c); Degree of relationship refers to the number of participating entities in a relationship.
43. (c); Trojan disguises itself as a useful program. But can comprise computer security & cause much damage. e.g. It appears as a perfectly legitimate e-mail or computer update.
44. (a); Complexity of bubble sort is $O(n^2)$
45. (b); A known plaintext attack is that if you know any of the plaintext that has been encrypted and have the resulting encrypted file, with a flawed encryption algorithm you can use that to break the rest of the encryption.
46. (a); IPSec (IP Security) is a suite of protocols which was designed by Internet Engineering Task Force (IETF) to protect data by signing and encrypting data before it is transmitted over public networks.
47. (b); Symmetric-key algorithms are algorithms for cryptography that use the same cryptographic keys for both encryption of plaintext and decryption of ciphertext. Kerberos encryption methods are based on keys that can be created only by the KDC and the client, or by the KDC and a network service, the Kerberos V5 protocol is said to use symmetric encryption. That is, the same key is used to encrypt and decrypt messages.
48. (a); Range of Class A address is 1.0.0.0 – 127.0.0.0
49. (a); Queue is a linear data structure used to represent a linear list. It allows insertion of an element to be done at one end and deletion of an element to be performed at the other end.
50. (d); A hierarchical database model is a data model in which the data is organized into a tree-like structure. The data is stored as records which are connected to one another through links. A record is a collection of fields, with each field containing only one value.
51. (a); Databases that support OLTP are most often referred to as operational databases.
52. (c); Operator new dynamically allocates memory for an object and returns a reference to it. This reference is address in memory of the object allocated by new.
53. (b); Bridges use Spanning Tree Protocol (STP) to eliminate loops in an extended network. If you choose to enable this feature, bridges exchange bridge protocol data unit (BPDU) messages with other bridges to detect loops, and the bridges shut down selected port to stop the loops.
54. (d); Selection sort is a simple sorting algorithm. This sorting algorithm is an in-place comparison-based algorithm in which the list is divided into two parts, the sorted part at the left end and the unsorted part at the right end. The smallest element is selected from the unsorted array and swapped with the leftmost element, and that element becomes a part of the sorted array. This process continues moving unsorted array boundary by one element to the right.
55. (b); Wi-Fi Protected Access (WPA) is a security standard for users of computing devices equipped with wireless internet connections, or Wi-Fi.
56. (a); An object database (also object-oriented database management system, OODBMS) is a database management system in which information is represented in the form of objects as used in object-oriented programming. Object databases are different from relational databases which are table-oriented.
57. (a); The User Datagram Protocol (UDP) is simplest Transport Layer communication protocol available of the TCP/IP protocol suite. It involves minimum amount of communication mechanism. UDP is said to be an unreliable transport protocol but it uses IP services which provides best effort delivery mechanism.
58. (a); A circular list can be used to represent a queue. Nodes can be accessed easily and deletion of nodes is easier.
59. (c); IPSec (IP Security) supports transport and tunnel mode, both of which can use either ESP or AH packets. Transport mode secures packets between two endpoints, typically in a client-to-gateway scenario, and leaves the original IP header unchanged. Tunnel mode encapsulates the IP header and payload into a new IPSec packet for transfer between two endpoints, typically two IPSec gateway devices.
60. (c); Aggregation allows creating complex entities from existing entities and relationship. Aggregation is the way used to abstracting a larger amount object.

# Professional Knowledge Practice Set: 02

1. Which of the following HTML tag is used to define a paragraph?
(a) <p> (b) <pr> (c) <c>
(d) <pl> (e) <style>
2. An Abstract class is:
(a) A class that must contain all pure virtual functions
(b) A class that must contain at least one pure virtual function
(c) A class that may not contain pure virtual function.
(d) A class that must contain pure virtual function defined outside the class.
(e) A class that must contain at least three pure virtual function
3. What is the correct HTML for adding a background color?
(a) <body bg="yellow">
(b) <background>yellow</background>
(c) <body style="background-color:yellow;">
(d) <body color: "background:yellow>
(e) <body background="yellow">
4. An IP address is a _____ bit number.
(a) 32 (b) 16 (c) 8
(d) 4 (e) 34
5. Which of the following is an open source server site development environment to develop dynamic web pages?
(a) PHP (b) XML (c) VB Script
(d) Only (b) and (c) (e) All of these
6. Swap space reside in_________.
(a) Disk (b) RAM (c) ROM
(d) Chip (e) Peripheral
7. Which instructions should be allowed only in kernel mode?
(a) Disable all interrupts
(b) Switch user mode to Kernel mode
(c) Read the time
(d) All of the above
(e) None of these
8. Round Robin scheduling algorithm falls under which of the following category?
(a) Preemptive scheduling
(b) Non-preemptive scheduling
(c) Sometime preemptive sometimes non-preemptive
(d) Deadlock Detection
(e) None of these
9. If a relation schema is in BCNF, then it is also in which normal form?
(a) 1st normal form (b) Second normal form
(c) 3NF (d) All of the above
(e) Both (a) and (b)
10. In fiber optics, the attenuation of light through glass depends on which among the following factors?
(a) Wave length of light (b) Frequency of light
(c) Speed of light (d) All of the above
(e) None of these
11. Network layer is responsible for ______ delivery of a packet.
(a) source-to-destination (b) end-to-end
(c) point-to-point (d) error free
(e) peer-to peer
12. When a packet with the code is transmitted; it is received and processed by every machine on the network. What is this mode of operation called?
(a) Broadcasting (b) Multi-casting (c) Unicasting
(d) Point-to-point (e) Both (b) and (d)
13. ADSL is the abbreviation for _______ .
(a) Asymmetric Dual Subscriber Line
(b) Asynchronous Digital System Line
(c) Asymmetric Digital Subscriber Line
(d) Asynchronous Dual System Line
(e) Asynchronous Dual Subscriber Line
14. What is the correct HTML for inserting an image?
(a) <image src="image.gif" alt="MyImage">
(b) <img href="image.gif" alt="MyImage">
(c) <img src="image.gif" alt="MyImage">
(d) <img alt="MyImage">image.gif</img>
(e) <ims herf="image">alt<MyImage>
15. Protocols in which the sender sends one frame and then waits for an acknowledgement before proceeding falls under which category?
(a) Sliding window (b) Stop and wait
(c) Frame buffer (d) Analog (e) Discrete
16. If a computer on the network provides shared resources for others to use, it is called ______ .
(a) Server (b) Client (c) Mainframe
(d) Microcomputer (e) User
17. The division of a message into segments, is a function of which of the following layer?
(a) Physical (b) Data link (c) Transport
(d) Network (e) Application
18. Which of the following is used to interconnect two IP classes, class A and C networks?
(a) A bridge (b) A class B network
(c) A router (d) A cable (e) A Port
19. Which among the following page replacement policy sometimes leads to more page faults when size of the memory is increased?
(a) FIFO (b) LRU (c) Optimal page
(d) No such policy exists (e) Both (a) and (b)
20. Which of the following is a service not supported by the operating system?
(a) Protection (b) Accounting
(c) Compilation (d) I/O operation
(e) Resource Allocation
21. What is thrashing?
(a) Reduces page I/O
(b) Decreases the degree of multi programming
(c) Implies excessive page I/O
(d) Improves the system performance
(e) Process takes less time

22. Which topology requires a multipoint connection in networking?
(a) Ring topology (b) Bus topology
(c) Mesh topology (d) Star topology
(e) Both (c) and (d)
23. Which of the following indicated by double line in ER diagram?
(a)Total participation
(b) Multiple participation
(c) Cardinality
(d) Entity
(e) Weak Entity
24. Which clause is used to specify condition while retrieving data from table in SQL?
(a) WHERE (b) LIKE (c) HAVING
(d) Group By (e) SELECT
25. In airline reservation system, the entities are date, Flight number, place of departure, destination, types of plane and seat available. The primary key is
(a) Flight number
(b) Flight number + place of departure
(c) Flight number + date
(d) Flight number+ destination
(e) Types of Plane + Flight number
26. What is the use of DML in database?
(a) Description of logical structure of database
(b) Addition of new structure in the database system
(c) Manipulation & processing of database
(d) Definition of physical structure of database system
(e) Definition Meta Data
27. What is ODBC?
(a) Object Database Connectivity
(b) Oral Database Connectivity
(c) Oracle Database Connectivity
(d) Open database Connectivity
(e) None of these
28. Which key represents relationship between tables?
(a) Primary Key (b) Secondary Key
(c) Foreign Key (d) Super key
(e) Candidate Key
29. In RDBMS terminology rows is synonyms of:
(a) Attribute (b) Relation (c) Matrix
(d) Tuple (e) Cardinality
30. Which network topology supports multiple paths from source to destinations between each possible node?
(a) Ring topology (b) Bus topology
(c) Mesh topology (d) Star topology
(e) Both (b) and (d)
31. Which protocol locates the hardware address of a local device?
(a) ARP (b) RARP (c) PING
(d) ICMP (e) POP
32. IP address 253.45.67.9 belongs to which class?
(a) A class (b) B class (c) D class
(d) E class (e) C Class
33. Logical link Control (LLC) is a Sub-layer of which layer in the OSI model?
(a) Network layer (b) Data link (c) Transport
(d) Presentation (e) Physical
34. Which one is the least expensive that can support 100 Mbps?
(a) Coaxial (b) UTP (c) Fiber-optic
(d) STP (e) Both (a) and (c)
35. Which IEEE standard is used in Wireless LAN?
(a) 802.3 (b) 802.4 (c) 802.5
(d) 802.11 (e) 802.12
36. By default, members of a C++ Class are:
(a)Private (b)Public (c)Protected
(d)Inverted (e)Inherited
37. What is data integrity?
(a) It is the data contained in the database that is non-redundant.
(b) It is the data contained in the database that is accurate and consistent.
(c) It is the data contained in the database that is secured
(d) It is the data contained in the database that is shared
(e) It is the update anomalies in database
38. With SQL, how do you select all the records from a table named "Persons" where the value of the column "FirstName" is "Navika"?
(a) SELECT * FROM Persons WHERE FirstName='Navika'
(b) SELECT [all] FROM Persons WHERE FirstName='Navika'
(c) SELECT [all] FROM Persons WHERE FirstName LIKE 'Navika'
(d) SELECT * FROM Persons WHERE FirstName<>'Navika'
(e) SELECT FROM Person WHERE FirstName LIKE 'Navika'
39. Which of the following parameter passing mechanism(s) is/are supported by C++, but not by C?
(a) Pass by value (b) Pass by reference
(c) Pass by value-result
(d) Both (a) and (b) (e) All of the above
40. Which of the following language is not used in Artificial Intelligence?
(a) LISP (b) PYTHON (c) PROLOG
(d) FORTRAN (e) C++
41. Which of the following process is is the process of discovering the principles of functioning of a software system through analysis of its operation?
(a) Software engineering
(b) Software re-engineering
(c) Reverse engineering
(d) Re-engineering
(e) Software Development Life Cycle
42. Which of these is NOT a valid keyword or reserved word in Java ?
(a) default (b) null (c) String
(d) volatile (e) Try
43. Which of these languages uses both interpreter and compiler?
(a) COBOL (b) BASIC (c) JAVA
(d) C++ (e) None of these

44. What is the name of new built in browser includes in windows 10?
(a) Opera (b) Cortana (c) Edge
(d) Super internet explorer pro (e) Chrome
45. Which of the following is related to 4th Normal Form?
(a)Single value dependency
(b) Multi valued dependency
(c) Join dependency
(d) Tri valued dependency
(e)None of these
46. Which of the following testing is done without planning and documentation?
(a) Unit Testing (b) Regression Testing
(c) Alpha Testing (d) Adhoc Testing
(e) Both (a) and (b)
47. After compilation of _____ code, _____ code is obtained to execute the any C program.
(a) Source, Object (b) Main, Object
(c) Object, Main (d) Object, Source
(e) Encrypt, Object
48. Ethernet use which Topology?
(a)Bus (b)Star (c)Ring
(d)Both (a) and (b) (e) Mesh
49. 'Like' clause is used to find which of the following in SQL?
(a) Distinct Record (b) Specified Pattern
(c) Meta Data (d) Any type of record
(e) All records from a table
50. Pharming, phishing and other malicious acts are known as_______
(a) Social Trimming (b) Social networking
(c) Social Engineering (d) Social phishing
(e) None of the Above
51. The term 'page traffic' describes:
(a) number of pages in memory at a given instant.
(b) number of papers required to be brought in at a given page request.
(c) the movement of pages in and out of memory.
(d) number of pages of executing programs loaded in memory.
(e) None of these
52. Which among the following can be used to make a copy of one class object from another class object of the same class type?
(a) constructor (b) copy constructor
(c) destructor (d) default constructor
(e) copy destructor
53. A special method that is used to initialize a class object?
(a) abstract method (b) static method
(c) Constructor (d) overloaded method
(e) override method
54. What is Unix?
(a) Single User, Single tasking OS
(b) Single User, Multi-tasking OS
(c) Multi-User, Multi- tasking OS
(d) Multi-User, Single tasking OS
(e) Both (a) and (b)
55. The process of defining two or more methods within the same class that have same name, but different parameters list is:
(a) method overloading (b) method overriding
(c) Encapsulation (d) Inheritance
(e) None of these
56. Process of removing an element from stack is called __________.
(a) Create (b) Push (c) Evaluation
(d) Pop (e) Eliminate
57. The data to be encrypted at the sender site is called _____ and the data to be decrypted at the receiver end is called _____.
(a) Secret key, public key (b) Public key, secret key
(c) Plain text, cipher text (d) Cipher text, plain text
(e) None of these
58. Data block of a very large file in the Unix file system are allocated using which type of allocation method?
(a) contiguous allocation
(b) linked allocation
(c) indexed allocation
(d) an extension of indexed allocation
(e) Filter allocation
59. In OSI network architecture, the dialog control and token management are responsibility of ________
(a) Session Layer (b) Network Layer
(c) Transport Layer (d) Data Link Layer
(e) Physical layer
60. Which of the following is not a Data Definition Language Statement?
(a) Delete (b) Create (c) Drop
(d) Alter (e) None

## SOLUTIONS

1. (a);
2. (b); It is sufficient to have one pure virtual function in the class to make it as an abstract class.
3. (c);
4. (a); IP address is a numerical label assigned to each device participating in a computer network that uses the Internet Protocol for communication, IP address is defined as a 32 bit number and this system is known as IPv4.
5. (a); PHP is a server scripting language, and a powerful tool for making dynamic and interactive Web pages.
6. (a); Swap space is an area on disk that temporarily holds a process memory image. When physical memory demand is sufficiently low, process memory images are brought back into physical memory from the swap area. Having sufficient swap space enables the system to keep some physical memory free at all times.
7. (a);
8. (a); In Non-preemptive scheduling, the CPU is allocated to the process till it terminates or switches to waiting state. While in preemptive scheduling the CPU is allocated to the processes for the limited time. Round Robin scheduling is preemptive scheduling.

Round Robin scheduling algorithm is designed especially for time sharing system.

9. (d); Boyce and Codd Normal Form is a higher version of the Third Normal form. This form deals with certain type of anomaly that is not handled by 3NF. A 3NF table which does not have multiple overlapping candidate keys is said to be in BCNF.
10. (a); Light travelling in an optical fiber loses power over distance. The loss of power depends on the wavelength of the light and on the propagating material. For silica glass, the shorter wavelengths are attenuated the most. The lowest loss occurs at the 1550-nm Wavelength, which is commonly used for long transmissions. So, the amount of attenuation for a given cable is wavelength dependent.
11. (a); Network layer helps in routing of frames (packets) through the network by defining the most optimum path the packet should take from the source to the destination.
12. (a); Broadcasting is the simultaneous transmission of the same message to multiple recipients. In networking, broadcasting occurs when a transmitted data packet is received by all network devices.
13. (c); Asymmetric digital subscriber line (ADSL) is a type of DSL technology that provides greater bandwidth and provides higher-speed transmission over traditional copper telephone wires than conventional voice band modem. ADSL is characterized by "high speeds" and "always on" connectivity.
14. (c); <img src="image.gif" alt="MyImage"> is used to inserting image.
15. (b); In this method of flow control, the sender sends a single frame to receiver & waits for an acknowledgment. The next frame is sent by sender only when acknowledgment of previous frame is received. This process of sending a frame & waiting for an acknowledgment continues as long as the sender has data to send. The main advantage of stop & wait protocols is its accuracy. Next frame is transmitted only when the first frame is acknowledged. So there is no chance of frame being lost.
16. (a); A server is a computer designed to process requests and deliver data to other (client) computers over a local network or the internet.
17. (c);
18. (c); Router is used to connect two different classes of networks.
19. (a); It is due to Belady's Anomaly in which
    More frames = more page fault
20. (c)
21. (c)
22. (b); Multipoint connection is used in BUS Topology. All the devices are connected to a single transmission medium, which acts as the Backbone of the connection. This links all the devices in the network. Here each node has its unique address.
23. (a); In total participation, each entity is involved in the relationship. Total participation is represented by double lines.
24. (a); The WHERE clause is used to extract only those records that fulfill a specified condition.
25. (c); A primary key is a special relational database table column (or combination of columns) designated to uniquely identify all table records.
26. (c); Data Manipulation Language (DML) statements are used for managing data in database. DML commands are not auto-committed. It means changes made by DML command are not permanent to database, it can be rolled back.
27. (d); Open Database Connectivity (ODBC) is an open standard application programming interface (API) for accessing a database.
28. (c); A foreign key is a column or group of columns in a relational database table that provides a link between data in two tables. It acts as a cross-reference between tables because it references the primary key of another table, thereby establishing a link between them.
29. (d);
30. (c); Mesh topology supports bidirectional links between each possible nodes in a network.
31. (a); Address Resolution Protocol (ARP) is a protocol for mapping the internet protocol address to a physical machine address that is recognized in local network.
32. (d); 253.45.67.9 is a class E address. Class E address range is reserved for future or experimental purpose. This start address is 240.0.0.0 and end address is 255.255.255.255 .
33. (b); The Logical Link Control (LLC) layer is one of two sub layers that make up the Data Link Layer of the OSI model. The Logical Link Control layer controls frame synchronization, flow control and error checking.
34. (b); Unshielded twisted pair (UTP) a popular type of cable that consists of two unshielded wires twisted around each other. Due to its low cost, UTP cabling is used extensively for local-area networks (LANs) and telephone connections.
35. (d); IEEE 802.11 is a set of media access control (MAC) and physical layer (PHY) specifications for implementing wireless local area network (WLAN).
36. (a); C++ is an object-oriented programming language. There are three access modifiers in C++ Public, protected and private. **Default** access to **members** of a class is **private.**
37. (b); Data integrity is the overall completeness, accuracy and consistency of data. This can be indicated by the absence of alteration between two instances or between two updates of a data record, meaning data is intact and unchanged.
38. (a); Correct format is SELECT * FROM Persons WHERE FirstName='Navika'
39. (b); As such C does not support pass by reference. But it can be simulated by using pointers.

40. (d); FORTRAN is not an Artificial Intelligence language. It is used for scientific numerical computation.
41. (c); Reverse engineering is the process followed in order to find difficult, unknown and hidden information about a software system.
42. (c); String is not a keyword or a reserved word in java.
43. (c); Java uses a two-step compilation process. Java uses a compiler to convert source code into byte code. This byte code is not a machine code. Later for execution JVA platform uses an interpreter to convert these byte codes into machine language.
44. (c); Microsoft Edge is the faster, safer browser designed for Windows 10.
45. (b); Fourth normal form (4NF) is a level of database normalization where there are no non-trivial multivalued dependencies other than a candidate key.
46. (d); Adhoc testing is used term for software testing performed without planning and documentation. The tests are intended to be run only once, unless a defect is discovered.
47. (a); **Source code** is the version of a computer program as it is originally written (i.e., typed into a computer) by a human in a programming language. **Object code** is the output of a compiler after it processes **source code**. Object code is a machine understandable code.
48. (d); Ethernet use two type of topology- bus and star. Use a bus topology for a large network with many users and longer segments. With repeaters or media converters, you can easily interconnect to other networks with different topologies. Use a star topology when you want to use twisted-pair cabling (10BASE-T Ethernet) for a multiple-building campus setup (you might already have twisted pair-telephone wire-installed on your premises). Use a star topology for your fiber optic links.
49. (b); The 'LIKE' operator is used in a WHERE clause to search for a specified pattern in a column.
50. (c); Social engineering is the art of manipulating people so they give up confidential information. The types of information these criminals are seeking can vary, but when individuals are targeted the criminals are usually trying to trick you into giving them your passwords or bank information, or access your computer to secretly install malicious software–that will give them access to your passwords and bank information as well as giving them control over your computer.
51. (c); Page Traffic in operating system describes the movement of pages in and out of memory.
52. (b); Copy constructor used to make a copy of one class object from another class object of the same class type.
53. (c); Constructor is a special method that is used to initialize an object. Every class has a constructor, if we don't explicitly declare a constructor for any Java class the compiler builds a default constructor for that class.
54. (c); UNIX is a multi-user, multi-tasking operating system. Multiple users may have multiple tasks running simultaneously.
55. (a); Two or more methods can have same name as long as their parameters declaration and definitions are different, the methods are said to be overloaded and the process is called method overloading. Method overloading is used when methods are required to perform similar tasks using different input parameters.
56. (d); Stack is an ordered list of similar data type. Stack is a LIFO structure. (Last in First out). push() function is used to insert new elements into the Stack and pop() function is used to delete an element from the stack. Both insertion and deletion are allowed at only one end of Stack called Top.
57. (c); Cipher text is encrypted text. Plaintext is what you have before encryption, and cipher text is the encrypted result.
58. (d); The Unix file system uses an extension of indexed allocation. It uses direct blocks, single indirect blocks, double indirect blocks and triple indirect blocks.
59. (a); The session layer in a network model is the network dialog controller. It establishes, maintains, and synchronizes the interaction among communicating systems. Session layer also responsible for token management.
60. (a); DDL is short name of Data Definition Language, which deals with database schemas and descriptions, of how the data should reside in the database. DDL commands are-CREATE, ALTER, DROP, TRUNCATE, RENAME.

# Professional Knowledge Practice Set:03

1. Which among the following is the full form of SOA? It a method to improve availability in ITL.
(a) Service offer Adjustment
(b) Service outage Analysis
(c) Service of Applications
(d) System Optimization Approach
(e) Systematic Operational Adjustment
2. Identify that, which among the following is included in SQL security issues?
(a) The complexity of key structures
(b) The number of records
(c) The size of databases
(d) The number of candidate key elements
(e) granularity of authorisation
3. What will be the output of the program?
```
#include<stdio.h>
int main()
{
 int m=1;
 printf("%d == 1 is" "%s\n", m, m==1?"TRUE":"FALSE");
 return 0;
}
```
(a) m == 1 is TRUE (b) 1 == 1 is TRUE
(c) 1 == 1 is FALSE (d) m == 1 is FALSE
(e) M== 0
4. Which of the following relates with a CASE SQL statement?
(a) It is a way to establish a data definition in SQL
(b) It is a way to establish an IF-THEN-ELSE in SQL.
(c) It establishes a loop in SQL or establish an IF-THEN-ELSE in SQL.
(d) It is a way to establish a look in SQL
(e) It is a way to establish a data definition in SQL or establish a loop in SQL.
5. Identify a dynamic analysis tool to test software modules?
(a) Desk Checking (b) Design and code
(c) Structured (d) Green box testing
(e) Black box test
6. Which one of these lists contains only Java programming language keywords?
(a) class, if, void, long, Int, continue
(b) strictfp, constant, super, implements, do
(c) try, virtual, throw, final, volatile, transient
(d) goto, instanceof, native, finally, default, throws
(e) byte, break, assert, switch, include
7. Data mining is used to aid in:
(a) analyzing current decisions made by managers.
(b) operational management
(c) retrieving archival data
(d) analyzing past decisions made by managers
(e) detecting patterns in operational data
8. Which of the following are SQL wildcard?
(a) asterisk(*) ; percentage sign(%)
(b)percentage sign (%); underscore(_)
(c)underscore(_); question mark(?)
(d)question mark(?); asterisk(*)
(e)percentage sign(%); question mark(?)
9. Which of the following system use more than one CPU to execute any program?
(a) time sharing system (b) desktop system
(c) client-server system (d) parallel system
(e) Embedded System
10. A server farm is an example of which among the following?
(a) Idle servers (b) Serve tolerance
(c) Stand-alone server (d) Server clustering
(e) Redundant servers
11. Which of the following is a cloud concept that is related to posting and sharing of resources?
(a) Visualization (b) Abstraction
(c) Virtualization (d) Polymorphism
(e) Encapsulation
12. Which is a connection-oriented protocol?
(a) ARP (b) ICMP (c) UDP
(d) UDP and ARP (e) TCP
13. What value must a C++ destructor return?
(a) Destructors do not return a value
(b) An object of the class
(c) A status code determining whether the class was destructed correctly
(d) A pointer or status code
(e) A pointer to the class
14. Identify which is used to write small programs to control Unix functionalities?
(a) Filters (b) C language (c) Method
(d) Special script (e) Shell
15. Which among the following closely relates to use of PPP?
(a) It is used to provide communication between two computers.
(b) It is used to assign IP address.
(c) It is used to send packets of data from one computer systems to the other.
(d) It is used to provide communication between two computers over a serial cable.
(e) It is used to join two networks over a serial port.
16. In Java, when overriding a predefined method, both the original method and the new method must have the same:
(a) Method signature (b) Method body
(c) Method inheritance
(d) Method signature and inheritance
(e) Parameter names
17. A computer host or small network inserted as a 'neutral zone' between a company's private network and the outside public network is called__________.
(a) sheep dip computer

(b) VPN (c) Firewall
(d) Man in the middle (e) DMZ

18. Data dictionary is a tool used to centrally manage parts of a database by controlling which among the following?
(a) Portion of data (b) Programming
(c) Keys (d) Metadata (e) Figures

19. Which feature provides a cloud storage area where you can store files in a private and secure online that you can access from any computer?
(a) OneDrive (b) Wireless Drop
(c) Network (d) Over Drive
(e) Cloud Connect

20. Software Development Life Cycle is a process that produces software with the highest quality and lowest cost in the shortest time. Identify the phase in SDLC, that involves acquiring any new hardware that might be required?
(a) systems design (b) systems development
(c) systems implementation
(d) Systems analysis (e) Systems maintenance

21. A software program is commonly defined as a set of instructions, or a set of modules or procedures, that allow for a certain type of computer operation. Which among the following is a Software programs that close potential security breaches in an operating system?
(a) fresh patches (b) security breach fixes
(c) security patches (d) security repairs
(e) refresh patches

22. Identify the valid declaration of a Boolean?
(a) boolean b1 = false; (b) boolean b1 = 'false';
(c) boolean b1 = 0;
(d) boolean b1 = Boolean.false();
(e) boolean b1 = no;

23. Which among the following are Common UDP Applications?
1. DHCP 2. SNMP
3. TFTP 4. HTTP
(a) 1, 3, 4 (b) 1, 2, 3 (c) 2, 3, 4
(d) Only 2 and 4 (e) Only 4

24. Which of the following technique is an alternative to log based recovery?
(a) Check points (b) Shadow paging
(c) Locks (d) Page based
(e) Recovery with current transactions

25. A coding scheme recognized by system software for representing organizational data best defines a ________.
(a) Data type (b) Data size (c) Tuple
(d) Hyperlink (e) Hashing algorithms

26. Which of the following CASE tools generate reports that help you identify possible inconsistencies, redundancies and omissions in diagrams, forms and reports?
(a) Form generators (b) Report generators
(c) Diagramming tools (d) Analysis tools
(e) Documentation generators

27. An attribute whose value is unique across all occurrences:
(a) primary key (b) Recursive key
(c) Join attribute (d) Data marker
(e) Single-valued key

28. What is the job of Medium term scheduler?
(a) Control degree of multiprogramming
(b) Re-introduce the process
(c) Perform swapping of processes
(d) Use to deletion of processes
(e) All of the above

29. Collecting personal information and effectively posing as another individual is a crime known as________.
(a) Spoofing (b) Identity theft
(c) Hacking (d) Personality theft
(e) Spooling

30. Which command is used to remove a directory in unix?
(a) rd (b) rmdir (c) dldir
(d) rdir (e) deldir

31. Which connectionless protocol is used for its low overhead and speed?
(a) ARP (b) TCP and ARP
(c) ICMP (d) UDP (e) TCP

32. Which among the following Linux system call is used to bias existing property of process?
(a) Exit (b) Fork (c) Wait
(d) Nice (e) Exec

33. Business intelligence leverages software and services to transform data into actionable intelligence that informs an organization's strategic and tactical business decisions. Business intelligence application gather data from:
(a) Data warehouse (b) Dashboards
(c) OLAP (d) Data analytics
(e) Decision support system

34. Which among the following is a functionality in HTML that facilitates connection of one or more pages in web pages?
(a) Links (b) Anchors (c) Navigation
(d) Browsing (e) tracking

35. Which among the following is an empty element?
(a) <img> (b) <body> (c) <center>
(d) <marquee> (e) <br>

36. For a weak entity set to be meaningful, it must be associated with another entity set. Which among the following is that entity set?
(a) Any Existing set (b) Owner set
(c) Neighbouring set (d) Strong entity set
(e) Relationship set

37. Cryptanalysis is used to:
(a) Breach cryptographic security systems and gain access to the contents of encrypted messages.
(b) Get information like passwords, networks, devices, firewall information and much more By impersonating someone with the authority to access documents or configurations.

(c) Capture the amount data being transmitted on a network.
(d) Leverage man-in-the-middle attacks against hosts on a computer network.
(e) Provide secrecy and integrity to the data, and both authentication and anonymity to the communication

38. A VPN, or Virtual Private Network, is a service that allows you to connect to the internet via a server run by a VPN provider. Another name for a VPN is a:
(a) Tunnel (b) One-time Password
(c) Pipeline (d) bypass (e) Switch

39. The RSA algorithm is an example of what type of cryptography?
(a) Asymmetric key (b) Symmetric key
(c) Secret key (d) Private key
(e) Logic key

40. If a table has been normalized so that all determinants are candidate keys, then that table is in:
(a) 1NF (b) BCNF (c) 2 NF
(d) PJNF (e) Both (a) and (c)

41. Disk scanners are?
(a) Anti-virus protection systems
(b) Aftereffects of virus (c) 8086's supplementary IC
(d) All of these (e) None of these

42. Array is a set of _____ objects.
(a) different type of (b) same type of
(c) Unique (d) Unwanted
(e) None of these

43. Top to down traversing of tree is called________.
(a) Preorder (b) Postorder (c) Inorder
(d) Searching (e) None of these

44. When a language has the capability to produce new data types, it is said to be_______.
(a) extensible (b) overloaded
(c) reprehensible (d) encapsulated
(e) None of these

45. What is the software called which when get downloaded on computer scans your hard drive for personal information and your internet browsing habits?
(a)Backdoors (b) Key-logger (c) Malware
(d) hackers (e) Spyware

46. ________message denotes the request has been fulfilled and a new source is created.
(a) 202 Accepted (b) 201 Created
(c) 403 Forbidden (d) 404 Not Found
(e) None of these

47. Which of the following is a technique of improving the priority of process waiting in Queue for CPU allocation?
(a)Aging (b)Starvation (c)Revocation
(d)Relocation (e) None of these

48. Which is not a software life cycle model?
(a) Waterfall model (b) Spiral model
(c) Prototyping model
(d) Capability Maturity Model
(e) None of the above

49. Which key is represents a relationship between tables.
(a) Primary Key (b) Secondary Key
(c) Foreign Key (d) None of these
(e) All of these

50. How many bits are in an IPv6 address?
(a) 128 bit (b) 32 bit (c) 64 bit
(d) 16 bit (e) None of these

51. Which command is used to execute compiled java program?
(a) Java (b) Javac (c) Run
(d) Javaw (e) Jcomp

52. Which of the following are layers in the TCP/IP model?
A. Application B. Session C. Transport
D. Internet E. Data Link F. Physical
(a) A and B (b) A, C and D
(c) B, C and E (d) C, D and E (e) D, E and F

53. Which of the following is the phase of SDLC in which the identified requirements are incorporated into the new system's design?
(a) Maintenance (b) Physical layout
(c) Requirement design (d) Systems design
(e) Systems blueprint

54. This is one of the older and slower methods where an intruder might use to scan a network to determine which of a range of IP addresses map to live hosts:
(a) ping sweep (b) phreaking (c) smurfing
(d) war driving (e) brute force cracking

55. "CREATE TABLE …" command is used to create which type of table in Mysql?
(a) Permanent Tables (b) Virtual tables
(c) Temporary tables
(d) Both (b) and (c)
(e) All of the mentioned

56. When you drop a ________ corresponding indexes are also dropped.
(a) Column (b) Sequence (c) View
(d) Table (e) Tuple

57. Which of the following can be determined only from the meaning of the attributes in the entity type?
(a) Domain (b) Cardinality (c) Key
(d) Relation (e) Constant

58. Student and Courses are example of _________.
(a) One to One (b) Many to Many
(c) One to Many (d) Many to One
(e) None of these

59. Multiple inheritance is not supported in Java because?
(a) To remove ambiguity and provide more maintainable and clear design.
(b) Java is an Object oriented language.
(c) Multiple inheritance is not an important feature.
(d) Both(b)and(c)
(e) All of the above

60. Cross-compiler is a compiler:

(a) Which is written in a language that is different from the source language.
(b) That generates object code for its host machine
(c) Which is written in a language that is same as the source language.
(d) That runs on one machine but produces object code for another machine.
(e) Both (a) and (b)

## SOLUTIONS

1. (b); The IT infrastructure Library (ITIL) refers to service or systems outage analysis (SOA) as a method to improve availability. Presented as an availability management process tool or technique, SOA is a powerful management tool to improve quality.
2. (e); SQL security issues include the granularity of authorization and the number of different ways you can execute the same query.
3. (b); int m=1; The variable m is declared as an integer type and initialized to '1'.
Step 2: printf("%d == 1 is" "%s\n", m, m==1?"TRUE":"FALSE"); becomes
=> m==1?"TRUE":"FALSE"
=> 1==1?"TRUE":"FALSE"
=> "TRUE"
Therefore the output of the program is 1 == 1 is TRUE
4. (b); The CASE statement is SQL's way of handling if /then else logic. The CASE statement is followed by at least one pair of WHEN and THEN statements-SQL's equivalent of IF/THEN. It must end with the END statement. The ELSE statement is optional, and provides a way to capture values not specified in the WHEN/THEN statements.
5. (e); A blackbox test is a dynamic analysis tool for testing software modules. During the testing of software modules a blackbox test works first in a cohesive manner as one single unit/entity, consisting of numerous modules and second, with the user data that flows across software modules. In some cases, this even drives the software behavior. In choices B, C and D, the software (design or code) remains static and somebody simply closely examines it by applying his/her mind, without actually activating the software. Hence, these cannot be referred to as dynamic analysis tools.
6. (d); All the words in option d are among the 49 Java keywords.
int starts with a lowercase i.
"virtual" is a keyword in C++, but not Java.
"constant" is not a keyword. Constants in Java are marked static and final.
"include" is a keyword in C, but not in Java.
7. (e); Data mining software is one the analytical tool for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases.
8. (b); The SQL WILDCARD Operators are used in conjunction with LIKE operator to enhance the search in a table. The two main Wild Card operator used in SQL are:
The percent sign(%) represents zero, one, or multiple characters and The underscore(_) represents a single character.
9. (d); Parallel system simultaneous use more than one CPU to execute a program. Ideally, parallel processing makes a program run faster because there are more engines (CPUs) running it.
10. (d); A server farm is a group of computers acting as servers and housed together in a single location. A server farm is sometimes called a server cluster. A Web server farm can be either (a) a Web site that has more than one server, or (b) an Internet service provider (ISP) that provides Web hosting services using multiple servers.
11. (c); Virtualization is a foundational element of cloud computing and helps deliver on the value of cloud computing," Adams said. "Cloud computing is the delivery of shared computing resources, software or data – as a service and on demand through the internet.
12. (e); TCP is a connection-oriented protocol, which means a connection is established and maintained until the application programs at each end have finished exchanging messages. It determines how to break application data into packets that networks can deliver, sends packets to and accepts packets from the network layer, manages flow control, and–because it is meant to provide error-free data transmission –handles retransmission of dropped or garbled packets as well as acknowledgement of all packets that arrive.
13. (a); Constructors and destructors are special member functions of classes that are used to construct and destroy class objects. Construction may involve memory allocation and initialization for objects. Destruction may involve cleanup and deallocation of memory for objects. The following restrictions apply to constructors and destructors:
    - Constructors and destructors do not have return types nor can they return values.
    - References and pointers cannot be used on constructors and destructors because their addresses cannot be taken.
    - Constructors cannot be declared with the keyword virtual.
    - Constructors and destructors cannot be declared static, const, or volatile.
    - Unions cannot contain class objects that have constructors or destructors.

Constructors and destructors obey the same access rules as member functions. For example, if you declare a constructor with protected access, only derived classes and friends can use it to create class objects.

14. (e); The shell provides you with an interface to the UNIX system. It gathers input from you executes programs based on that input. When a program finishes executing. it displays that program's output. A shell is an environment in which we can run our commands, programs. and shell scripts. The basic concept of a shell script is a list of commands, which are listed in the order of execution. A good shell script will have comments, preceded by a pound sign, #, describing the steps.
15. (a); PPP (Point-to-Point Protocol) is a protocol for communication between two computers using a serial interface, 24. typically a personal computer connected by phone line to a server.
16. (a); Method overloading and method overriding are two important concepts in Java which allows Java programmer to declare method with same name but different behaviour. **For overriding both name and signature of method must remain same**, but in for overloading method signature must be different. In case of method overloading, method with same name co-exists in same class but they must have different method signature, while in case of method overriding, method with same name is declared in derived class or sub class.
17. (e); A firewall is a set of related programs, located at a network gateway server that protects the resources of a private network from users from other networks. The term also implies the security policy that is used with the programs. Demilitarized zone (DMZ) is a computer host or small network inserted as a "neutral zone" between a company's private network and the outside public network to prevent outside users from gaining direct access to a server that stores company data.
18. (d); Data dictionary is a tool used to centrally manage parts of a database by controlling the metadata within. We can define a data dictionary as a DBMS component that stores the definition of data characteristics and relationships. You may recall that such "data about data" were labeled metadata. The metadata stored in the data dictionary is often the bases for monitoring the database use and assignment of access rights to the database users.
19. (a); OneDrive provides a cloud storage area where you can store files in a private and secure online location that you can access from any computer.
20. (a); The third phase of the SDLC is called systems design. The output of the analysis phase is a description of the alternative solution recommended by the analysis team. Once the recommendation is accepted by the organization, you can make plans to acquire any hardware and system software necessary to build or operate the system as proposed. System design refers to the technical specifications that will be applied in implementing the candidate system. It also includes the construction of programmers and program testing.

It has the following stages: Acquiring hardware and software, if necessary; Database design; Developing system processes; Coding and testing each module.

21. (c); A patch is a small piece of software that is issued whenever a security flaw is uncovered.
22. (a); A Boolean can only be assigned the literal true or false.
23. (b); The DHCP employs a connectionless service model, using the User Datagram Protocol (UDP). TFTP is designed for the quick and easy transfer of small files and is a great example of a protocol that was specifically designed for UDP. SNMP also uses UDP as it is an administrative protocol that uses relatively short messages.
24. (b); Shadow paging is an alternative to transaction log based recovery technique. This is the method where all the transactions are executed in the primary memory or the shadow copy of database. Once all the transactions completely executed, it will be updated to the database.
25. (a); Data type is a detailed coding scheme recognized by system software, such as a DBMS. for representing organizational data. In SQL Server, each column, local variable, expression, and parameter has a related data type. A data type is an attribute that specifies the type of data that the object can hold: integer data, character data, monetary data, date and time data, binary strings, and so on.
26. (d); The general types of CASE tools are listed below:
    Diagramming tools: enable system process, data and control structures to be represented graphically.
    - Computer display and report generators: help prototype how systems look and feel. It makes it easier for the systems analyst to identify data requirements and relationship.
    - Analysis tools: automatically check for importance, inconsistent, or incorrect specifications in diagrams, forms, and reports.
    - Central repository: enables the integrated storage of specifications, diagrams, reports and project management information.
    - Documentation Generators: produce technical and user documentation in standard formats.
    - Code generators: enable the automatic generation of program and data base definition code directly from the design documents, diagrams, forms, and reports.
27. (a); The primary key attribute — that attribute whose value is unique across all occurrences of the relation - is indicated by an underline and an

attribute of a relation that is the primary key of another relation is indicated by a dashed underline.

28. (e); Medium-term scheduling is a part of swapping. It removes the processes from the memory. It reduces the degree of multiprogramming. The medium-term scheduler is in-charge of handling the swapped out-processes. It can re-introduce the process into memory and execution can be continued.
29. (b); Identity theft is the deliberate use of someone else's identity, usually as a method to gain a financial advantage or obtain credit and other benefits in the other person's name, and perhaps to the other person's disadvantage or loss.
30. (b); rmdir is used to remove a directory in unix.
31. (d); User Data Protocol (UDP) is a connectionless protocol. it means that a datagram can be sent at any moment without prior advertising, negotiation or preparation.
32. (d); System calls are used for process management in Linux Nice():- to bias the existing property to process. Fork () is used to create a new process; Exec() Execute a new program; Wait () wait until the process finishes execution; Exit() Exit from the process.
33. (a); Business intelligence is the ways in which can be stored and used business information. It encompasses the technologies, application, and means for collecting, integrating, analyzing, and presenting business data. Using data that has been stored in a data warehouse, software applications are able to use this data to report past business information as well as predict future business information, including trends, threats, opportunities and patterns.
34. (a); A link is a connection from on Web resource to another. A link has two ends – called anchors – and a direction.
35. (e); HTML elements with no content are called empty elements. <br> is an empty element without a closing tag (the <br> tag defines a line break).
36. (d); The existence of a weak entity set depends on the existence of a strong entity set; it must relate to the strong set via a one-to-many relationship set. The discriminator (or partial key) of a weak entity set is the set of attributes that distinguishes among all the entities of a weak entity set.
37. (a); Cryptanalysis is the study of analyzing information systems in order to study the hidden aspects of the systems. Cryptanalysis is used to breach cryptographic security systems and gain access to the contents of encrypted message, even if the cryptographic key is unknown.
38. (a); A VPN is designed to provides a secure, encrypted tunnel in which to transmit the data between the remote user and the company network.
39. (a); The most common asymmetric cipher currently in use is RSA, which is fully supported by the. NET Security Framework. Asymmetric means that there are two different keys. This is also called public key cryptography, because one of them can be given to everyone. The other key must be kept private.
40. (b); BCNF is based on the concept of a determinant. A determinant is any attribute (simple or composite) on which some other attribute is fully functionally dependent. A relation is in BCNF is, and only if, every determinant is a candidate key.
41. (a); Disk scanners eg. Memory resident, inoculation, start-up, Behavior-based, detection disk scanners are the utility that comes with the antivirus software. They all constitute protection systems against malwares.
42. (b); An array is a collection of data items of the same type.
43. (a); Pre-order traversing involves visit the root node first (Top of the tree), traverse to left subtree & then traversing to right subtree. Therefore, it is top to down traversing.
44. (a); An extensible system is one that can be upgraded to fully handle the new data in the new input format.
45. (e); Spyware is a type of malware that is installed on a computer without the knowledge of the owner in order to collect the owner's private information. Spyware is often hidden from the user in order to gather information about internet interaction, keystrokes (also known as key logging), passwords, and other valuable data.
46. (b); 201 created shows when the request has been fulfilled, resulting in the creation of a new resource.
47. (a); Ageing is a scheduling technique used to avoid starvation. Fixed priority scheduling is a scheduling discipline, in which tasks queued for utilizing a system resource are assigned a priority each.
48. (d); Capability Maturity Model because CMM is not a software life cycle model. Instead, it is a strategy for improving the software process.
49. (c); A foreign key is a column or group of columns in a relational database table that provides a link between data in two tables. It acts as a cross-reference between tables because it references the primary key of another table, thereby establishing a link between them.
50. (a); Internet Protocol version 6 (IPv6) is the most recent version of the Internet Protocol (IP), the communications protocol that provides an identification and location system for computers on networks and routes traffic across the Internet. IPv6 uses a 128-bit address.
51. (b); Javac command is used to compile a java file. You can directly run javac in command prompt if environment variable is set correctly.
52. (b); The Session layer is not in the TCP/IP model; neither are the Data Link and Physical layers. Transport layer (Host-to-Host in the DoD model), Internet layer (Network layer in the OSI), and

Application layer (Application/Process in the DoD) and Network Interface Layer are in TCP/IP Model.

53. (d); The design phase takes as its initial input- the requirements identified in SRS and includes translation of the requirements specified in the SRS into a logical structure that can be implemented in a programming language. The output of a design phase is a design document that acts as an input for all the subsequent SDLC phases.
54. (a); Ping sweep (also known as an ICMP sweep) is a basic network scanning technique used to determine which range of IP addresses map to live end hosts (computers). You can use a single ping to find out whether a specific end hosts exists on the network. A Ping Sweep consists of ICMP (Internet Control message Protocol) ECHO requests sent to multiple hosts. If a given address is live, it will return an ICMP ECHO reply. Ping sweeps are among the older and slower methods used to scan a network.
55. (a); CREATE TABLE command is used to create permanent tables. To create a virtual table CREATE VIEW command is used
56. (b); Sequence
57. (c); A key is an attribute or a collection of attributes whose value(s) uniquely identify an entity in the entity set. Keys can be determined only from the meaning of the attributes in the entity type.
58. (b); The following relationship between student and Course entity is a many-to-many relationship According to the relationship a student can enrol in any number of courses and the course can have any number of students.
59. (a); Multiple Inheritance is the process whereby a sub-class can be derived from more than one super class.
    Multiple inheritance can lead to a lot of confusion when two base classes implement a method with the same name. That's because multiple inheritance not support java.
60. (d); A cross compiler that runs on one computer but produces object code for a different type of computer. Cross compilers are used to generate software that can runs on computers with a new architecture or on special-purpose devices that cannot host their own compilers.

# Professional Knowledge Practice Set:04

1. The Key size in DES (Data Encryption standard) is:
   (a) 24 bit (b) 32 bit (c) 56 bit
   (d) 64 bit (e) 8 bit
2. Which of the following is the latest version of Android?
   (a)Nougat (b)Marshmallow (c)Oreo
   (d)Jelly Bean (e) Lollipop
3. Which of the following characteristics include in the spiral model of software development?
   (a) Ends with the delivery of the software product
   (b) Is more chaotic than the incremental model
   (c) Includes project risks evaluation during each iteration
   (d) work well for smaller project
   (e) All of the above
4. What is ICMP Protocol?
   (a) A protocol used to dynamically bind a high level IP Address to a low-level physical hardware address
   (b) A protocol for transferring files from one machine to another
   (c) A protocol used to monitor computers
   (d) A protocol that handles error and control messages
   (e) A protocol that used for internet access
5. Which one of the following is a cryptographic protocol used to secure HTTP connection?
   (a) stream control transmission protocol (SCTP)
   (b) transport layer security (TSL)
   (c) explicit congestion notification (ECN)
   (d) resource reservation protocol
   (e) Routing Protocol.
6. What is the output of this C code?

```
#include <stdio.h>
 int main()
 {
      j = 10;
      printf("%d\n", j++);
      return 0;
 }
```

   (a) 10 (b) 11 (c) Compile time error
   (d) 0 (e) None of these
7. What are the three types of addresses in IPV6?
   (a) Class A, Class B, Class C
   (b) Unicast, anycast, multicast
   (c) Unicast, dualcast, multicast
   (d) Unicast ,dualcast, anycast
   (e) None of these
8. "PaaS" stand for in cloud computing______
   (a) Protocol as a service
   (b) Platform as a service
   (c) Platform as a system
   (d)Performance as a service
   (e) None of these
9. Which type of malicious code is embedded with some legitimate program that is set to "explode" when certain conditions met?
   (a) Trap doors (b) Trojan horse
   (c) Logic Bomb (d) Virus
   (e) None of these
10. With SQL, how can you insert "saxena" as the "LastName" in the "Persons" table?
   (a) INSERT INTO Persons (LastName) VALUES ('saxena')
   (b) INSERT INTO Persons ('saxena') INTO LastName
   (c) INSERT ('saxena') INTO Persons (LastName)
   (d) INSERT (saxena) TO Person(LastName)
   (e) INSERT TO Person(saxena)INTO(LastName)
11. IP Address 168.2.7.6 belong to which Class?
   (a) Class A (b) Class B (c) Class C
   (d) Class D (e) Class E
12. Which one of the following is a visual (mathematical) way to determine the deadlock occurrence?
   (a) Resource allocation graph
   (b) Starvation graph (c) Inversion graph
   (d) Recursion Graph (e) None of these
13. When a program tries to access a page that is mapped in address space but not loaded in physical memory, then
   (a) Segmentation fault occurs
   (b) Fatal error occurs (c) Page fault occurs
   (d) No error occurs (e) Systematic error occur
14. Which of the following create metadata?
   (a)DML Compiler (b) DML pre-processor
   (c)DDL interpreter (d) Query interpreter
   (e) DCL compiler
15. Unix is originally written in which language?
   (a) C (b) Machine language
   (c) Assembly language
   (d) Perl (e) JAVA
16. Which of the following is a keyword used for a storage class in C programming?
   (a) Printf (b) external (c) auto
   (d) scanf (e) None of these
17. Which of the following is the latest version of Oracle?
   (a)Oracle 11g (b)Oracle 12c (c)Oracle 10g
   (d)Oracle 9i (e)Oracle 8i
18. What is the output of this program?

```
#include <iostream>
using namespace std;
int main()
{
      int a = 10;
      if (a < 10) {
            for (i = 0; i < 10; i++)
                  cout << i;
      }
```

```
    else {
        cout << i;
    }
    return 0;
}
```

(a) 0123456789 (b) 123456789 (c) 0
(d) error (e) 1234098765

19. Which of the following is a mechanism by which object acquires the properties of another object?
a) Encapsulation b) Abstraction
c) Inheritance d) Polymorphism
(e) None of these

20. Which data structure represents hierarchical relationship between various elements?
(a) Linked List (b) Tree (c) Array
(d) Stack (e) Queue

21. In the Ethernet network, what method is used to access the media?
(a) Demand priority (b) CSMA/CD
(c) Polling (d) CSMA/CA
(e) None of these

22. When an operating system or computer hardware is fails, then what type of failure occur in oracle database.
(a)Application Failure (b)Instance failure
(c) Media Failure (d) Rollback Failure
(e)None of these

23. Which of the following is the biometric login featured of Windows 10.
(a)Windows Instant (b)Windows Bio
(c)Windows Hello (d)Windows Padlock
(e)None of these

24. Doubles ellipses in E-R diagram represent________
(a) Entity set (b) multi valued
(c) derived attributes (d) relationship
(e) Weak Entity

25. What is trap door?
(a) IT is trap door in War Games
(b) It is a hole in software left by designer.
(c) It is a Trojan horse
(d) It is a virus which traps and locks user terminal
(e)None of these

26. Which of the following is a subject- oriented, integrated, time variant, non-volatile collection of data in support of management decisions.
(a)Data Mining (b) Data Warehousing
(c)Web Mining (d) Text Mining
(e)None of these

27. Data stored in an electronic memory cell can be accessed at random and on demand using:
(a) Memory addressing (b) Direct addressing
(c) Indirect addressing (d) Control Unit
(e) None of these

28. Which of the following is not one of the 3 general components of the Unix operating system?
(a) the kernel (b) the shell (c) the GUI
(d) the file system (e) None of these

29. Which algorithm is that part of the network layer software responsible for deciding which output line an incoming packet should be transmitted on
(a) Routing (b) Session (c) Adaptive
(d) Non-adaptive (e) None of these

30. What is private cloud?
(a)A standard cloud service offered via the Internet
(b)A cloud architecture maintained within an enterprise data center
(c)A cloud service inaccessible to anyone but the cultural elite
(d) All of these
(e) None of these

31. What are the qualities of good software?
(a)Reliability (b)Correctness (c)Efficiency
(d)Portability (e) All of the above

32. JavaScript was developed by __________ to run inside of web pages.
(a) Microsoft (b) Sun (c) Oracle
(d) Netscape (e) None of these

33. Which vendor recently launched a cloud based test and development service for enterprises?
(a) TCS (b) Cisco (c) IBM
(d) Oracle (e) None of these

34. What is VSAT in networking?
(a) Very small Aperture terminal
(b) Virtually Small Aperture Terminal
(c) Value small Aperture terminal
(d) Very small amplified terminal
(e) Virtual small amplified terminal

35. If K is a foreign Key in relation R1 then
(a) Every tuple of R1 has a distinct value for K
(b) K cannot have a null value for tuples in R1
(c) K is a key for some other relation.
(d) K is a primary key for R1.
(e) None of these

36. What is fragmentation in operating system?
(a) The condition of a disk in which files are divided into pieces scattered around the disk.
(b) Removing viruses from the computer
(c) Cleaning up of computers memory
(d) Increasing the size of primary memory
(e) All of the above

37. Hierarchical Database Model does not support.
(a) Many to Many relation
(b) Tree like structure
(c) Relation between nodes
(d) Both (a) and (b)
(e) None of these

38. Select * from employee where salary>10000 and dept_id=101;
Which of the following fields are displayed as output?
(a) Salary, dept_id from employee relation
(b) Employee from employee relation
(c) Salary from employee relation
(d) All the field of employee relation that meets the criteria
(e) It will not display any relation

39. What is RJ45?
(a) Connector (b) wire (c) Interface device
(d) Converter (e) None of the above
40. In Round Robin CPU scheduling, as the time quantum is increased, the average turnaround time:
(a) increases (b) decreases
(c) remains constant (d) varies irregularly
(e) None of these
41. In a 'for loop' if the condition is missing, then infinite looping can be avoided by a________
(a) Continue statement (b) goto statement
(c) return statement (d) break statement
(e) only (b), (c) and (d)
42. Which traversal method can be done by starting from the root and performing?
(a)Preorder traversal (b)In order traversal
(c) DFS (d) BFS (e) None
43. Which is the example of embedded Operating system?
(a) Window 8 (b) Linux (c) Android
(d) Dos (e) None of these
44. Which protocol is used to convert IP addresses to MAC Address?
(a) IP (b) RARP (c) In ARP
(d) ARP (e) None of these
45. In which database model the rule of Normalization can apply.
(a) Network Model (b) Hierarchical Model
(c) Relational Model (d) Both c and b
(e) None of these
46. Suppose we have a system in which processes is in hold and wait condition then which of the following approach prevent the deadlock.
(a) Request all resources initially
(b) Spool everything
(c) Take resources away
(d) Order resources numerically
(e) None of these
47. Which of the following mode is used whenever either end of a security association is gateway.
(a) Tunnel (b) Encapsulating
(c) Transport (d) Gateway
(e) None of these
48. All exceptions in Java are subclasses of built in class called_________
(a) Exception (b) Error (c) Throwable
(d) Raise (e) none of these
49. What is the amount of time takes for the desired sector of a disk to rotate under the read-write heads of the disk drive?
(a) Positioning time (b) Random access time
(c) Seek time (d) Rotational latency
(e) None of these
50. Which of the following are shows by using ER diagram?
(a) Attributes of the Entities
(b) Entities of the database
(c) Relationship between the entities
(d) all of the above
(e) None of these
51. Which of them take care of email at client side?
(a) POP (b) MAP (c) IMAP
(d) Both (a) and (c) (e) None of the mentioned
52. RAD Software process model stands for:
(a) Rapid Application Development.
(b) Relative Application Development.
(c) Rapid Application Design.
(d) Recent Application Development.
(e) None of these
53. What is the other name for a LAN Card?
(a) NIC (b) Network Connector
(c) MODEM (d) Internet card
(e) None of these
54. The length of the one-byte instruction is________.
(a) 2 bytes (b) 1 byte (c) 3 bytes
(d) 4 bytes (e) None of these
55. Operation code field is present in which type of instruction?
(a) Programming language instruction
(b) Assembly language instruction
(c) Machine language instruction
(d) Logical instruction
(e) None of the mentioned
56. To promote encapsulation, Java provides which type of access modifier?
(a) Public (b) Private (c) Protected
(d) All of the above (e) None of the above
57. A distributed network configuration in which all data information pass through a central computer is ________.
(a) Bus Network (b) Ring Network
(c) Star Network (d) Font to Point Network
(e) None of the above
58. **Flexcube** , core banking solution software developed by which company?
(a)TCS (b)Infosys (c)IBM
(d)Oracle (e)None of these
59. Which of these coding types is used for data type characters in Java?
(a) ASCII (b) ISO-LATIN-1 (c) UNICODE
(d) Both (a) and (c) (e) None of these
60. What is the complexity of Merge Sort?
(a) O (1) (b) O ($n^2$) (c) O (log n)
(d) O (n log n) (e) None of these

## SOLUTIONS

1. (c); The data encryption standard (DES) is a common standard for data encryption and a form of secret key cryptography (SKC), which uses only one key for encryption and decryption. Public key cryptography (PKC) uses two keys, i.e., one for encryption and one for decryption. Key size in DES is -56 bit and Block size 64 bit.

2. (c); Android Oreo is the official name of the next version of Google's mega-popular mobile operating system that's available now for select devices. Oreo focuses on speed and efficiency.3.(c); The main objective of the spiral model is to emphasize management to evaluate and resolve risks in the software project.
4. (d); Internet Control Message Protocol (ICMP) is a TCP/IP network layer protocol that provides troubleshooting, control and error message services. It is a supporting protocol and used by networks devices like routers for sending the error messages and operations information.
5. (b); Transport layer security (TLS) is a protocol that provides communication security between client/server applications that communicate with each other over the Internet. It enables privacy, integrity and protection for the data that's transmitted between different nodes on the Internet. TLS is a successor to the secure socket layer (SSL) protocol.
6. (c); Variable j is not defined in the given program.
7. (b); Internet Protocol version 6 (IPv6) is the most recent version of the Internet Protocol (IP), the communications protocol that provides an identification and location system for computers on networks and routes traffic across the Internet. IPv6 are divided into three addresses- Unicast, anycast, multicast.
8. (b); Platform as a Service, often simply referred to as PaaS, is a category of cloud computing that provides a platform and environment to allow developers to build applications and services over the internet. PaaS services are hosted in the cloud and accessed by users simply via their web browser.
9. (c); The logic bomb is code embedded in some legitimate program that is set to "explode" when certain condition are met.
10. (a); When we want to insert any value in the table, the correct format of query is-
INSERT INTO table_name(column1, column2, column3,….) VALUES(value1, value2, value3,….)
11. (b); Range of Class B IP address 128.0.0.0 to 191.255.255.255
12. (a); Resource allocation graph is a mathematical way to determine the deadlock occurrence. A resource allocation graph tracks which resource is held by which process and which process is waiting for a resource of a particular type. It is very powerful and simple tool to illustrate how interacting processes can deadlock.
13. (c); **page fault** error occurs when the **operating system** cannot find the data in virtual memory.
14. (c); DDL interprets translate the DDL(data definition language) statement and record them in tables containing metadata.
15. (c); UNIX is originally written in assembly language, later rewritten to C.
16. (c); Auto is the default storage class for all the variables declared inside a function or a block.
17. (b); Oracle 12c is a version of the Oracle Database. The "c" stands for "cloud" to indicate that 12c is "cloud enabled". It features a new multi-tenant option that will help companies to consolidate databases into private or public clouds.
18. (d); We will get compilation error because 'i' is an undeclared identifier.
19. (c); 'Inheritance' is the mechanism provided by Object Oriented Language, which helps an object to acquire the properties of another object usually child object from parent object.
20. (b); Tree is a hierarchical data structure, with a root value and sub trees of children with a parent node, represented as a set of linked nodes.
21. (b); All versions of Ethernet use a Media Access Control (MAC) protocol called CSMA/CD. CSMA/ CD (Carrier Sense Multiple Access /Collision Detection) a set of rules determining how network devices respond when two devices attempt to use a data channel simultaneously.
22. (b); A database instance is a set of memory structures that manage database files. Causes of instance Failure- a power outage, a server hardware failure, failure of an oracle background process, emergency shutdown procedures.
23. (c); Windows Hello is the biometric authentication feature that helps strengthen authentication and helps to guard against potential spoofing through fingerprint matching and facial recognition.
24. (b); A multi valued attribute can have more than one value at a time for an attribute. For ex., skills of a surgeon are a multi valued attribute since a surgeon can have more than one skill. It represents double ellipses in E-R diagram.
25. (b); **Trap door** is the new jargon for Backdoor Programs or Backdoor **virus** in Software field. A **trap door** is a secret entry point into a program that allows someone that is aware of the **trap door** to gain access without going through the usual security access procedures.
26. (b); Data warehousing is the process of constructing and using a data warehouse. A data warehouse is constructed by integrating data from multiple heterogeneous sources that support analytical reporting, structured and/or ad hoc queries, and decision making. Data warehousing involves data cleaning, data integration, and data consolidations.
27. (b); Direct addressing eliminates the need to process a large stream of irrelevant data in order to the desired data word.
28. (c); the **UNIX operating system** is made up of **three** parts; the kernel, the shell, and the file system(program).
29. (a); The **network layer** is responsible for **routing** packets from the source to destination.

The **routing algorithm** is the piece of software that decides where a packet goes next (e.g., which output line, or which node on a broadcast channel).

30. (b);Private clouds are data center architectures owned by a single company that provides flexibility, scalability, provisioning, automation and monitoring. The goal of a private cloud is not sell "as-a-service" offerings to external customers but instead to gain the benefits of cloud architecture without giving up the control of maintaining your own data center.
31. (e); Quality of software is not just concern with what software does, it also include the software's behavior while it is executing and the structure and organization of the system programs and associated documentation. Quality of good software is- reliability, correctness, efficiency, portability, flexibility, maintainability etc.
32. (d); Java script was developed by Netscape. JavaScript is a powerful and popular language for programming on the web.
33. (c); IBM has launched a cloud-based commercial software development and test service for partners and customers that the company said can dramatically reduce IT labor and infrastructure costs.
34. (a); A very small aperture terminal (VSAT) is a small telecommunication earth station that receives and transmits real-time data via satellite. A VSAT transmits narrow and broadband signals to orbital satellites. The data from the satellites is then transmitted to different hubs in other locations around the globe.
35. (c); A **foreign key** is a column or group of columns in a relational database table that provides a link between data in two tables.
36. (a); Fragmentation is the condition of a disk in which files are divided into pieces scattered around the disk. Fragmentation occurs naturally when you use a disk frequently, creating, deleting, and modifying files.
37. (a); A **hierarchical database model** is a data model in which the data is organized into a tree-like structure. The data is stored as **records** which are connected to one another through **links**.
38. (d); Here *is used to select all the fields of the relation that will meet the criteria.
39. (a); **Registered Jack 45 (RJ45)** is a standard type of physical connector for network cables. RJ45 connectors are most commonly seen with Ethernet cables and networks.
40. (d); In Round Robin algorithm, the value of time quantum or the time slice, plays a crucial role in deciding how effective the algorithm is. If the time quantum is too small, there could be lot of context switching happening which could slow down the performance. If the time quantum is too high, then RR behaves like FCFS. If the time quantum is increased, the average response time varies irregularly.
41. (e); An **infinite loop** (sometimes called an endless **loop** ) is a piece of coding that lacks a functional exit so that it repeats indefinitely. In computer programming, a **loop** is a sequence of instruction s that is continually repeated until a certain condition is reached. In a for loop if condition is missing , then infinite looping can be avoided by goto, return, and break statement.
42. (d); **Breadth-first search** (**BFS**) is an algorithm for traversing or **searching** tree or graph data structures. It starts at the tree root (or some arbitrary node of a graph, sometimes referred to as a '**search** key') and explores the neighbor nodes **first**, before moving to the next level neighbors.
43. (c); An embedded operating system is a type of operating system that is embedded and specifically configured for a certain hardware configuration. Embedded operating systems are design for the many specialized devices and do specific task with efficiency. Embedded operating systems are also known as real-time operating systems (RTOS). Example of Embedded Operating Systems are- Car, Security Camera, Smartphone, Home appliances like fridges, microwaves, and washing machines etc.
44. (d); The **Address Resolution Protocol** (**ARP**) feature performs a required **function** in IP routing. **ARP** finds the hardware address, also known as Media Access Control (MAC) address, of a host from its known IP address. **ARP** maintains a cache (table) in which MAC addresses are mapped to IP addresses.
45. (c); **Normalization** is the process of organizing the columns (attributes) and tables (relations) of a **relational database** to reduce data redundancy and improve data integrity.
46. (a); There are two approach to eliminate hold and wait condition first is allocate all required resources to the process before start of its execution, this way hold and wait condition is eliminated but it will lead to low device utilization and second is require processes to request resources only when it has none. Thus, first they must release all their currently held resources before requesting all the resources they will need from scratch.
47. (a); The Tunnel mode is used whenever either end of a security association is a gateway. Therefore, between two firewalls the tunnel mode is always used.
48. (c); All **Exception** types are **subclasses** of the **built-in class** Throwable. **Exception** is used for exceptional conditions that user program should catch. Catch is not **called**, so control never returns to the try block. You can describe the **Exception** by passing it as an argument to println() statement.

49. (d); A **rotational delay or rotational latency** is the amount of time between information requests and how long it takes the hard drive to move to the sector where the requested data is located.
50. (d); The **entity-relationship model** (or **ER model**) is a way of graphically representing the logical relationships of entities (or objects) in order to create a **database**. ER diagram contain entity, attributes and relationship between the entities.
51. (d); **POP** is short for **Post Office Protocol**, a protocol used to retrieve e-**mail** from a **mail** server. Most e-**mail** applications (sometimes called an e-**mail** client) use the **POP** protocol, although some can use the newer IMAP (Internet Message Access Protocol).
    "IMAP" stands for "**Internet Message Access Protocol**." It is better than POP for accessing mail from multiple locations because it leaves the mail messages on the mail server. This lets you read the same messages regardless of what computer you use to read your e-mail.
52. (a); RAD model is Rapid Application Development model. It is a type of **incremental model**. In RAD model the components or functions are developed in parallel as if they were mini projects. The developments are time boxed, delivered and then assembled into a working prototype. This can quickly give the customer something to see and use and to provide feedback regarding the delivery and their requirement.
53. (a); A network interface controller (NIC, also known as a network interface card, network adapter, LAN adapter or physical network interface, and by similar terms) is a computer hardware component that connects a computer to a computer network.
54. (b); This format is only one byte long.
55. (c); Machine language instruction format has one or more fields. The first one is the operation code field.
56. (b); Encapsulate provides contract to other object to what to hide and what to expose which can be accessed other objects. In java, we use private access modifier to hide method and variable to restrict from outer world.
57. (c); Star networks are one of the most common computer network topologies. In its simplest form, a star network consists of one central hub which acts as a conduit to transmit messages.
58. (d); **Flexcube** is designed by Oracle Corporation. This software package is used by some of the reputed Banks of the world like City Bank. Here is the list of Indian banks that use Flexcube Software Packages- Canara Bank, Karur Vysya Bank, Lakshmi Vilash Bank, Syndicate Bank, Yes Bank.
59. (c); Unicode defines fully international character set that can represent all the characters found in all human languages. Its range is from 0 to 65536.
60. (d); **Merge Sort** is an efficient, general-purpose, comparison-based sorting algorithm. Complexity of merge sort is O (n log n).

1. In which topology, if there are n devices in a network, each device has n-1 ports for cable.
(a) Mesh (b) Star (c) Bus
(d) Ring (e) Hybrid
2. Error detection and control is the function of which layer of the OSI model
(a) Physical layer (b) Data link layer
(c) Application layer (d) Network layer
(e) Transport layer
3. PJNF Stand for
(a) Practically – join Normal form
(b) Project – join normal form
(c) Pages – Join normal form
(d) Pages – join normal form
(e) None of these
4. Which of the following is suitable for 'thread'?
(a) Heavyweight process (b) Process
(c) Lightweight process (d) A task
(e) None of these
5. In an E-R diagram ellipses represent:
(a) Entity sets
(b) Relationship among entity sets
(c) Attributes
(d) Link between attributes and entity sets
(e) None of the above
6. Bridge works at which layer of OSI Model.
(a) Network layer (b) Data link layer
(c) Physical layer (d) Application layer
(e) Presentation layer
7. What is the full form of ACID
(a) Atomicity, consistency, Isolation and durability
(b) Atomicity, content, Isolation and duration
(c) Atom, content, isolation, duration
(d) Application, consistencies, isolation and duration
(e) None of the above
8. Which one of the following deletes all the entries but keeps the structure of the relation?
(a) Delete from relations where P;
(b) Delete from dept where dept name= 'Finance';
(c) Delete from emp where salary between 13000 and 15000;
(d) Delete from emp;
(e) Drop relations;
9. Which of the following is the ability to modify the internal schema without causing any change to the external schema in database?
(a) Physical data independence
(b) Logical data independence
(c) External data independence
(d) Both Physical and External data independence
(e) None of these
10. What is Adwords?
(a) Advertising Service by Microsoft
(b) Advanced Search Engine
(c) Advertising Service by Google
(d) Automatic words Search Engine by Yahoo
(e) Advanced search engine by IBM
11. Which is the malicious action performs collecting personal information and effectively posing as another individual.
(a) Scareware (b) Pharming (c) Hacking
(d) Identity theft (e) Cracking
12. Find the decimal value of $(111001)_2$?
(a) $(56)_{10}$ (b) $(57)_{10}$ (c) $(58)_{10}$
(d) $(55)_{10}$ (e) $(60)_{10}$
13. If Δ represents '1' and ○ represents '0'. What will be the one's complement of - ○ΔΔ○○Δ?
(a) 011001 (b) 100110 (c) 101010
(d) 000000 (e) 111111
14. Which of the following memory chip is faster?
(a) DRAM (b) RAM (c) SRAM
(d) PRAM (e) None of these
15. Which of the following is a set of Flip flops integrated together?
(a) Adder (b) Memory (c) Counter
(d) Register (e) Clock
16. Which of the following key combination is used to restart your computer?
(a) Shift+Ctrl+Del (b) Ctrl+Alt+Del
(c) Alt+Ctrl+Shift (d) Ctrl+Del (e) Alt+Ctrl
17. Which of the following is used to specify the layout of web pages?
(a) HTML (b) CSS (c) Java Script
(d) XML (e) None of these
18. Which of the following is not a transmission medium?
(a) telephone lines (b) coaxial cables
(c) modem (d) microwave systems
(e) None of these
19. The number of processes completed per unit time is known as __________
(a) Output (b) Throughput (c) Efficiency
(d) Capacity (e) Accuracy
20. Which of the following language defines commands GRANT and REVOKE.
(a) DDL (b) DML (c) DCL
(d) DQL (e) None of these
21. Which of the following is correct acronym of VGA?
(a) Video Graphics Adapter
(b) Visual Graphics Array
(c) Volatile Graphics Array
(d) Video Graphics Array
(e) Virtual Graphics Array
22. Which of the following addressing mode, the operand is available in the instruction itself?
(a) Direct addressing (b) Immediate addressing
(c) Implied addressing (d) Indirect addressing
(e) None of these

23. What is linker in an operating system?
(a) is the same a loader
(b) is required to create a load module
(c) user source code as input
(d) is always used before programs are executed
(e) None of these

24. Which of the following is related to simplex communication?
(a) Single wire and two-sided communication
(b) Television
(c) Two-wire communication
(d) Walkie-talkie
(e) Mobile phone

25. Which of the following computers exhibits the features of both analog and digital computers?
(a) Hybrid computer (b) Mini computer
(c) Laptop (d) Mainframe computer
(e) None of these

26. Which of the following allows making copies of the database periodically to help in the cases of crashes and disaster?
(a) Recovery Utility (b) Monitoring Utility
(c) Backup Utility (d) Data loading Utility
(e) None of these

27. Which of the following type of file is part of the Oracle database?
(a)Control File (b)Password File
(c)Parameter File (d)Archived log File
(e)None of these

28. Which of the following allows you to upload web pages?
(a) HTTP (b) TCP (c) HTML
(d) UDP (e) None of these

29. 802.5 topology is called______.
(a) Ethernet (b) Token bus (c) Star
(d) Token ring (e) None of these

30. Which one of the following has to be added to the blank to select the dept_name which has Engineering as its ending string?
SELECT teacher_name
FROM department
WHERE dept_name LIKE ' _____ Engineering';
(a) % (b) _ (c) #
(d) $ (e) *

31. When the maximum clock rate is quoted for a logic family, it applies to a_______.
(a) Shift register (b) Flip-flop
(c) Counter (d) Single logic gate
(e) None of these

32. How many types of subprogram in PL/SQL?
(a)Two (b)Three (c)Four
(d)Six (e)One

33. In OSI terminal emulation is done in -
(a) Sessions layer (b) Application layer
(c) Presentation layer (d) Transport layer
(e) None of these

34. An important application of cryptography, used in computerized commercial and financial transaction-
(a) Data mining (b) Data warehousing
(c) Digital signature (d) Media convergence
(e) None of these

35. In Which type of network service, user first establishes and uses a connection, and then terminates the connection?
(a) connection-oriented (b) connection-less
(c) service-oriented (d) service-less
(e) Message oriented

36. Sockets are used in which layer of OSI Model?
(a) Physical layer (b) Datalink layer
(c) Network layer (d) Transport layer
(e) NOne of these

37. 253.45.67.9 is a class __________ address
(a) D (b) A (c) C
(d) E (e) B

38. Which is the fastest replicating virus?
(a) Stealth virus (b) Trojan (c) Worm
(d) All of these (e) None of these

39. The languages that come under Third Generation are ____.
(a) Java (b) C (c) VB
(d) All of these (e) None of these

40. A webpage cannot be created in :
(a) Java Server Pages (b) Active Server Pages
(c) Active USP Pages (d) All of these
(e) None of these

41. In Java, communication between object is done by:
(a) Methods (b) Procedures (c) Instances
(d) Contractors (e) Packages

42. Which among the following is a type of password recovery is considered more difficult and for this process one must work through all possible combinations of numbers and characters?
(a) Passive (b) Active (c) Dictionary
(d) Brute force (e) Hybrid

43. What is the most common security stance employed on firewalls?
(a) Allowing by default
(b) Custom configuring of access based on user account
(c) Caching internet content
(d) Denying by default, allowing by exception
(e) Using best available path

44. What type of join includes only those tuples with matching attributes and discards the rest?
(a) Equi-join (b) Natural join
(c) Outer join (d) Both (a) and (b)
(e) All of the above join operations can perform this function

45. Which of the following locks the item from change but not from read?
(a) Implicit lock (b) Explicit lock
(c) Exclusive lock (d) Shared lock
(e) Mutual lock
46. The storing of identical data in more than one place is called:
(a) Data inflexibility
(b) Program-data dependence
(c) Data inconsistency
(d) Data redundancy.
(e) Data encapsulation
47. A data structure where elements can be added or removed at either ends but not in the middle is called:
(a) Requeue (b) Queues (c) Dequeue
(d) inked lists (e) stacks
48. Frame relay:
(a) does not use error correction routines.
(b) is used by facsimile machines.
(c) uses error correction routines
(d) does not work well on the internet
(e) is an outdated technology.
49. Which of the following are five built-in functions provided by SQL?
(a) SUM, AVG, MIN, MAX, NAME
(b) MIN, MAX, MULT, DIV, NAME
(c) SUM, AVG, MULT, DIV, MIN
(d) SUM, AVG, MIN, MAX, MULT
(e) COUNT, SUM, AVG, MAX, MIN
50. Common Gateway Interface (CGI) is which of the following?
(a) An interface that accepts and returns data that may be written in SQL.
(b) A small program that executes within other application and is stared on the server.
(c) A small program that executes within other application and is stored on the client
(d) A small program that executes within other application and is stored the client or the server.
(e) An interface that accepts and returns data that may be written in any language that produces executable file.
51. AJAX stands for______
(a) Asynchronous JavaScript and xml
(b) Advanced JSP and xml
(c) Asynchronous JSP and xml
(d) Advanced JavaScript and xml
(e) Associative Java and XML
52. Where are cookies stored?
(a) On the server (b) In web.xml
(c) On the client (d) In HTML
(e)None of these
53. What happens when you push a new node onto a stack?
(a) The new node is placed at the front of the linked list
(b)The new node is placed at the back of the linked list
(c) The new node is placed at the middle of the linked list
(d)No Changes happens
(e)None of these
54. Parsing is also known as:
(a) Lexical Analysis (b) Syntax Analysis
(c) Semantic Analysis (d) Code Generation
(e) Interpreting
55. Which of the following relationship represent weak entity set
(a) one to one relationship
(b) one to many relationship
(c) many to many relationship
(d) All of the above
(e) None of the above
56. Which oops concept exposing only necessary information to user or clients?
(a) Abstraction (b) Encapsulation
(c) Data hiding (d) Hiding complexity
(e) None of these
57. Context-free language can be recognized by which type of automata
(a) Finite state automaton
(b) Linear bounded automata
(c) Push down automata
(d) Both (b) and (c) above
(e) None of these
58. Port number is consisting of how many bits?
(a) 32 bits (b) 16 bits (c) 8 bits
(d) 64 bits (e) None of these
59. Identify among the following which type conversion is NOT accepted?
(a) From char to int
(b) From float to char pointer
(c) From negative int to char
(d) From double to char
(e) All these conversions are accepted
60. Which SQL command used to avoid the selection of duplicate rows?
(a) UNIQUE (b) DISTINCT
(c) EXCLUSIVE (d) DISCRETE
(e) None of these

## SOLUTIONS

1. (a); Mesh Topology is a point-to-point connection to other nodes or devices. All the network nodes are connected to each other. Mesh has n (n-1)/2 physical channels to link n devices.
2. (b); the bit stream transmitted by the physical layer is not guaranteed to be error free. The data link layer is responsible for error detection and correction. The most common error control method is checksum.
3. (b); 5NF is also knows as Project-Join Normal Form (PJ/NF). It is designed for reducing the redundancy in relational databases. In 5NF, if an attribute is

multivalued attribute, then it must be taken out as a separate entity.

4. (c); Thread is a lightweight process because it works on the common address space. As a result, it needs less system resources to execute the threads. So, thread is a lightweight process.
5. (c); Attributes are represented by means of ellipses. Every ellipse represents one attribute and is directly connected to its entity (rectangle).
6. (b); Bridges operate at the Data-Link Layer of the OSI Model. They can distinguish between local and remote data, so data traveling from one workstation to another in the same segment doesn't have to cross the bridge. Bridges operate on MAC-Layer addresses.
7. (a); A transaction is a very small unit of a program and it may contain several low level tasks. A transaction in a database system must maintain Atomicity, Consistency, Isolation, and Durability − commonly known as ACID properties − in order to ensure accuracy, completeness, and data integrity.
8. (d); The absence of condition deletes all rows. Delete statement performs the conditional based deletion, whereas Drop command deletes entire records in the table. Delete statement removes only the rows in the table and it preserves the table structure as same, and Drop command removes all the data in the table and the table structure.
9. (a); Physical Data Independence the ability to modify the way database stored (data structures, file organization, etc. - internal view of a database) without affecting the next higher level conceptual schema is the physical data independence.
10. (c); Google AdWords is an online advertising service developed by Google, where advertisers pay to display brief advertising copy, product listings, and video content within the Google ad network to web users.
11. (d); Identity theft, also known as identity fraud, is a crime in which an imposter obtains key pieces of personally identifiable information, such as Social Security or driver's license numbers, in order to impersonate someone else.
12. (b);

| Binary Number | 1 | 1 | 1 | 0 | 0 | 1 |
|---|---|---|---|---|---|---|
| Power of 2 | $2^5$ | $2^4$ | $2^3$ | $2^2$ | $2^1$ | $2^0$ |

$(111001)_2 = 1\times2^5+1\times2^4+1\times2^3+0\times2^2+0\times2^1+1\times2^0= (57)_{10}$

13. (b); ○ΔΔ○○Δ represents 011001. For finding one's complement we change each 1 to 0 and each 0 to 1. Hence One's complement of 011001 will be 100110.
14. (c); SRAM a type of memory chip which is faster and requires less power than dynamic memory.
15. (d); The flip-flop is the basic unit of digital memory. A flip-flop can remember one bit of data. Sets of flip-flops are called registers, and can hold bytes of data. Sets of registers are called memories, and can hold many thousands of bits, or more.
16. (b); Ctrl+Alt+Del is used to restart your computer.
17. (b); CSS describes how HTML elements are to be displayed on screen, paper, or in other media.
18. (c); Modem is not a transmission medium , it is a device or program that enables a computer to transmit data over, for example, telephone or cable lines.
19. (b);
20. (c); DCL (Data Control Language) is used to control the user access to the database, tables, views, procedures, functions and packages. Using DCL command, it allows or restricts the user from accessing data in database schema. DCL use two commands to control access- GRANT and REVOKE.
21. (d); Video Graphics Array (VGA) refers specifically to the display hardware first introduced with the IBM PS/2 line of computers, but through its widespread adoption has also come to mean an analog computer display standard, the 15-pin D-subminiature VGA connector or the 640x480 resolution itself.
22. (b); The addressing mode in which the data operand is a part of the instruction itself is known as immediate addressing mode.
23. (b); Linker is a program that takes one or more objects generated by a compiler and combines them into a single executable program.
24. (b); Television is an example of simplex communication.
25. (a); The computer that exhibits the features of both digital and analog computers, is known is a hybrid computer
26. (c); Backup Utility software is an application used to create a duplicate copy of data to safeguard it and enable recovery in the event it is lost, corrupted or infected by malware.
27. (a); oracle database is a set of three type of data files- Data files and temp file, control file and online redo log files.
28. (a); Web pages may be retrieved from a local computer or from a remote web server. web pages are requested and served from web server using Hypertext Transfer Protocol(HTTP).

29. (d); In IEEE 802.5, the token passing scheme is used in place of CSMA/CD on a ring topology LAN.
30. (a); The % character matches any substring.
31. (a); Shift registers are type of sequential logic circuit, mainly for storage of digital data. They are a group of flip-flop connected in a chain so that the output from one F/F becomes the input of the next F/F. All F/F is driven by a common clock, and all are set or reset simultaneously.
32. (a); PL/SQL (Procedural Language/Structured Query Language) is Oracle Corporation's procedural extension for SQL and the Oracle relational database. PL/SQL has two types of subprograms called procedures and functions. PL/SQL use a procedure to perform an action and a function to compute a value.
33. (b); One of the well-known Application layer protocol is telnet. Telnet is a terminal emulation protocol which is used for logging on remotely to network hosts.
34. (c); Digital signature is an electronic signature equivalent of the hand written signature, with data in electronic from being attached to other, electronic subject data (invoice, payment slip, contract, etc) as a mean of authentication.
35. (a); In connection oriented service we have to establish a connection before starting the communication. When connection is established we send the message or the information and then we release the connection.
36. (d); A network socket is an endpoint of an inter process communication flow across a communication flow across a computer network. The socket is primarily a concept used in Transport layer of the Internet model. A socket is a interface between application layer and transport layer within a host. Transport layer and transport layer within a host. Transport layer within a host. Transport layer of the receiving host delivers data to the socket.
37. (d); 253.45.67.9 is a class E address. Class E address range is reserved for future or experimental purposes. It this start address is 240.0.0.0 and end address is 255.255.255.255. but 253.45.67.9 fall between the two.
38. (c); A worm replicates itself and spread itself automatically to other computers.
virus replicates or spreads it-self by infecting other programs on the same computer. A stealth virus is a resident virus that attempts to evade detection by concealing its presence in infected files. Trojan is a malicious program embedded inside a normal, safe-looking program. Trojan is embedded whereas virus is attached to the file or executable. Trojan can't replicate.
39. (d); VB is a third generation event driven programming language.
C and Java are 3rd generation high level languages.
40. (c); Both ASP and JSP can be used to create dynamically generated pages ASP – Microsoft; JSP – sun Microsystems.
41. (a); In java, methods operate on the internal state of an object and the object-to-object communication is done via methods.
42. (d); Brute Force Attack is the most widely known password cracking method. This attack simply tries to use every possible character combination as a password.
43. (d); A default/ deny stance assumes that all traffic is potentially malicious or at least unwanted or unauthorized. Thus, everything is prohibited by default. Then, as benign, desired, and authorized traffic is identified, an exception rule grants it access. This is known as deny by default/allow by exception. These two rules define the foundation for rules governing traffic crossing the firewall.
44. (d); Join is a combination of a Cartesian product followed by a selection process. A Join operation pairs two tuples from different relations, if and only if a given join condition is satisfied. Theta Join, Equijoin, and Natural Join are called inner joins. An inner join includes only those tuples with matching attributes and the rest are discarded in the resulting relation. Outer joins include all the tuples from the participating relations in the resulting relation. There are three kinds of outer joins − left outer join, right outer join, and full outer join.
45. (d); When a process places a shared lock on a data item, the process declares that it is just viewing the data and it allows others to look, but no other processes can change the data item. Shared locks are used by processes that are reading data that should not be changed while the process is reading it.
46. (d); Data redundancy is a condition created within a database or data storage technology in which the same piece of data is held in two separate places. This can mean two different fields within a single database, or two different spots in multiple software environments or platforms. Whenever data is repeated, this basically constitutes data redundancy. This can occur by accident, but is also done deliberately for backup and recovery purposes.

47. (c); A deque is a linear list in which elements can be added or removed at either end but not in the middle. The term deque refers to the name double ended refers to the name double ended queue. An input restricted deque is one which allows insertions at only one end but does not allow of the list. An output restricted deque is one which allows deletions at only one end of the list but allow insertions at both the ends of the list.
48. (a); Frame relay is shared network technology that packages data into bundles for transmission but does not use error correction routines. When a frame fails the error checking routine, it is dropped. The frame relay service has a bit command/ response flag in the header.
49. (e); SQL has many built-in functions for performing processing on string or numeric data. Following is the list of all us useful SQL built-in functions: COUNT, MAX, MIN, AVG, SUM, SQRT, RAND, CONCAT, numeric and string functions.
50. (e); CGI is a standard script for running programs on a server from a web page, CGI programs can be run independently and were designed to be external, so they can by run under various servers interchangeably. Gateways conforming to this specification can be written in any language that produces an executable file. Some of the language that use CGI are C, C++, Perl, Python.
51. (a); AJAX stands for Asynchronous JavaScript and XML. AJAX is a new technique for creating better, faster, and more interactive web applications with the help of XML, HTML, CSS, and Java Script.
52. (c); A cookie is information stored on your computer by a website you visit. Cookies are stores in client side.
53. (a); Stack is Last In First Out data structure, so when new node is add in linked list it placed in front.
54. (b); Parsing or syntactic analysis is the process of analysing a string of symbols and conforming to the rules of grammar.
55. (B); one to many relationship is the example of weak entity set
56. (a); Abstraction in Object Oriented Programming helps to hide the irrelevant details of an object.
57. (d); Context free language can be recognized by both Linear Bounded automata and Push down automata.
58. (b); A port number is the logical address of each application or process that uses a network or the Internet to communicate. A port number uniquely identifies a network-based application on a computer. Each application/program is allocated a 16-bit integer port number. This number is assigned automatically by the OS, manually by the user or is set as a default for some popular applications.
59. (b); Conversion of a float to pointer type is not allowed.
60. (b); The SQL DISTINCT keyword is used in conjunction with the SELECT statement to eliminate all the duplicate records and fetching only unique record.

# Professional Knowledge Practice Set:06

1. Which of the following is a collection of similar Entities?
(a) Attribute set (b) Relation set
(c) Entities set (d) Relationship set
(e) Tuple set
2. Which of these keywords can be used to prevent Method overriding?
(a) Static (b) Constant (c) Protected
(d) Final (e) None of these
3. Which of the following occurs due to the software mistakes during coding?
(a) Bugs (b) Failure (c) Defects
(d) Error (e) Crash
4. White Box techniques are also classified as which of the following methods?
(a) Error guessing technique
(b) Design based testing
(c) Structural testing
(d) Domain analysis
(e) Decision table testing
5. In which type of testing technique individual units are combined and tested as a group?
(a) Unit testing (b) Integrated testing
(c) System testing (d) Acceptance testing
(e) None of these
6. What is SIM in microprocessor?
(a) Signal Interrupt Mask
(b) Sorting Interrupt Mask
(c) Select Interrupt Mask
(d)Set Interrupt Mask
(e) Semi Interrupt Mask
7. Proxy firewall filters at which layer?
(a) Network layer (b) Transport layer
(c) Application layer (d) Datalink layer
(e) Presentation layer
8. Which of the following cloud deployment are owned and operated by a third-party cloud service provider, which deliver their computing resources like servers and storage over the Internet?
(a) Public cloud (b) Private cloud
(c) Hybrid cloud (d) Both (a) and (b)
(e) PaaS
9. In the asymmetric-key method of cryptography, which key is publicly known?
(a) Encryption key only (b) Decryption key only
(c) Both keys (d) Cipher
(e) None of these
10. Which of the following is used to defines a hyperlink in HTML?
(a) <a> (b) <br> (c) <h1>
(d) <b> (e) <f>
11. Which of the following loop structure repeats a statement or group of statements while a given condition is true?
(a) for loop (b) while loop (c) nested loop
(d) both (a) and (b) (e) do loop
12. Which of the following malicious software that can change its code and signature patterns with each iteration?
(a) Keylogger (b) Mutation Engine
(c) Morris worm (d) Metamorphic Virus
(e) Stealth Virus
13. In which type of SQL join returns all the rows from left table combine with the matching rows of the right table
(a) Inner join (b) Left outer join
(c) Right outer join (d) Full join
(e) Self join
14. Which of the following sorting algorithm is best if a list is already sorted?
(a) Insertion sort (b) Quick Sort (c) Merge Sort
(d) Heap sort (e) Bubble Sort
15. If person A wants to send sensitive data to person B, and wants to be sure that only person B may be able to read it, he will encrypt the data with person B's Public Key. Only Person B has access to his corresponding Private key and as a result Person B is the only one, with the capability of decrypt the encrypted data back to its original form. What is the name of this cryptographic process?
(a) Secret key Cryptography
(b) Symmetric key Cryptography
(c) Public key Cryptography
(d) Asymmetric key Cryptography
(e) Both (c) and (d)
16. Which of the following can lead to network failure?
(a) Configuration BPDUs
(b) Slow LACP negotiation
(c) Broadcast storm
(d) Failure to establish an EtherChannel
(e) None of these
17. The number of usable host addresses on a subnet with mask 255.255.255.0 is:
(a) 32 (b) 254 (c) 65534
(d) 8 (e) 14
18. Which OSI Model layer is responsible for character code translation?
(a) Layer 1 (b) Layer 5 (c) Layer 2
(d) Layer 6 (e) Layer 4
19. Which of the following property refers- when data should be sent and how fast it can be sent?
(a) Semantics (b) Timing (c) Syntax
(d) Topology (e) Channel
20. Which of the following provides authentication, integrity and confidentially of network packets data/payload in IPv4 and IPv6 networks?
(a) AH (b) ESP (c) PGP
(d) SSL (e) ICMP
21. Which of the following HTTP error message shows when server is currently unavailable
(a) 503 (b) 404 (c) 406
(d) 400 (e) 202
22. Which HTTP request method are used to submits data to be processed to a specified resource?
(a) GET (b) POST (c) HEAD
(d) PUT (e) OPTIONS

23. Which of the following phase produce intermediate representations of the source program in compiler design?
(a) Lexical Analyzer
(b) Syntax Analyzer
(c) Semantic Analyzer
(d) Intermediate code generator
(e) Machine Independent code optimiser

25. C++ encourages structure of software as a collection of components that are ____
(a) Highly cohesive and loosely coupled
(b) Not highly cohesive but loosely coupled
(c) Highly cohesive and tightly coupled
(d) Not highly cohesive but tightly coupled
(e) None of these

26. In which type of communication, a message is sent out on the network and is destined for a group of systems?
(a) Unicast (b) Multicast (c) Broadcast
(d) Both (b) and (c) (e) Semicast

27. Which of the following is not a transaction state in DBMS?
(a) Active (b) Aborted (c) Committed
(d) Failed (e) Terminated

28. Someone who breaks into the telephone network illegally, typically to make free long-distance phone calls or to tap phone lines:
(a) Phreaker (b) Hacker (c) Cracker
(d) Spoofing (e) Hacktivist

29. Which of the network security attacks are a series of minor data security attacks that together result in a larger attack?
(a) Data diddling attacks
(b) Trust relationship attacks
(c) Salami attacks
(d) Man-in-the-middle attacks
(e) Session hijacking attacks

30. Which operator is used to combine the result-set of two or more SELECT statements?
(a) Join (b) Having (c) Union
(d) Combine (e) Unique

31. What is XaaS in cloud computing?
(a) Anything as a Service
(b) Example as a Service
(c) Anything as a System
(d) Explore as a Service
(e) Explore as a System

32. What is the output of the given program:
```
#include <iostream>
using namespace std;
int main()
{// local variable declaration:
int x = 1;
switch(x)
{case 1 :
cout << "Hi!" << endl;
break;
default :
cout << "Hello!" << endl;}
return 0;}
```
(a) Hi (b) Hello (c) HelloHi
(d) HiHello (e) Compile error

33. Which of the following UNIX command is used to remove or delete files?
(a) rm (b) mv (c) cp
(d) lpr (e) man

34. When an operating system or computer hardware fails, then what type of failure occur in oracle database?
(a) Application Failure (b) Instance failure
(c) Media Failure (d) Rollback Failure
(e) None of these

35. Which keyword is used to prevent any changes in the variable within a C program?
(a) immutable (b) mutable (c) const
(d) volatile (e) double

36. What is the “c” in latest version of oracle 12c?
(a) Command (b) Cloud Computing
(c) Character (d) Change (e) Common

37. Which of the following condition follows, if a relation is in 1st NF?
(a) There is no duplication of data.
(b) There are no composite attributes in the relation.
(c) There are only a few composite attributes.
(d) All attributes are of uniform type.
(e) None of the above.

38. A data model is a collection of conceptual tools for describing
(a) data and data relationship.
(b) data semantics and consistency constraints
(c) data, data relationship, data semantics and consistency constraints.
(d) only (a) and (b)
(e) All of the above

39. In which database model the rule of Normalization can apply?
(a) Network Model (b) Hierarchical Model
(c) Relational Model (d) Document model
(e) Context model

40. Which of the following is a technique of improving the priority of process waiting in Queue for CPU allocation?
(a) Aging (b) Starvation (c) Revocation
(d) Relocation (e) None of these

41. Which is not a software life cycle model?
(a) Waterfall model
(b) Spiral model
(c) Prototyping model
(d) Capability maturity model
(e) V-model

42. Which of the following is the time interval between the submission and completion of job?
(a) Waiting time (b) Turnaround time
(c) Throughput (d) Response time
(e) None of these

43. Acceptance tests are normally conducted by the:
(a) developer (b) end users
(c) test team (d) systems engineers
(e) None of these

44. Which of the following operator in c is used to explicitly to convert the value from one type to another?
(a) Bitwise (b) Cast (c) Logical
(d) Arithmetic (e) Comma

45. What is OLTP?
(a) Operating Limit Transaction Processing
(b) Operation Line Transaction Processing
(c) On-line Transaction Processing
(d) Online Termination processing
(e) Operation Line Termination Processing
46. Which of the following is the basic unit of execution in operating system?
(a) Bit (b) Thread (c) Thrashing
(d) Data (e) Byte
47. In the Ethernet network, what method is used to access the media?
(a) Demand priority (b) CSMA/CD
(c) Polling (d) CSMA/CA
(e) None of these
48. Logical link Control (LLC) is a Sub-layer of which layer of OSI model.
(a) Network layer (b) Data link
(c) Transport (d) Presentation
(e) None of these
49. Which of the following register holds the contents of data or instruction read from, or written in memory?
(a) Memory Address Register
(b) Memory Buffer Register
(c) Program Counter
(d) Instruction Register
(e) None of these
50. Which of the following is/are a PHP code editor?
(a) Notepad (b) Notepad++
(c) Adobe Dreamweaver (d) PDT
(e) All of the mentioned.
51. Which arithmetic operator in java used for string concatenation?
(a) (+) (b) (-) (c) (*)
(d) (/) (e) (%)
52. Which of the following is a special method that is used to initialize a class object?
(a) Abstract Method (b) Constructor
(c) Static method (d) Overloaded method
(e) Dynamic method
53. Method overloading is determined at:
(a) Run time (b) Compile time
(c) Coding time (d) Execution time
(e) None of these
54. Which of the following is a way of converting real world objects in terms of class?
(a) Polymorphism (b) Encapsulation
(c) Abstraction (d) Inheritance
(e) Parameters
55. Which of the following is final phase of the compiler that coverts the intermediate code or the optimized code into the target language?
(a) Code optimizer (b) Code generator
(c) Intermediate code (d) Code Assembler
(e) Syntax Analyzer
56. Which data structure represents hierarchical relationship between various elements?
(a) Linked List (b) Tree (c) Array
(d) Stack (e) Queue
57. Which one is the most important feature of spiral model?
(a) Quality management
(b) Risk management
(c) Performance management
(d) Efficiency management
(e) Cost management
58. What is the complexity of Merge Sort?
(a) O(1) (b) O($n^2$) (c) O(log n)
(d) O(n log n) (e) O(n)
59. Software programs that close potential security breaches in an operating System are known as____
(a) Security breach fixes (b) Refresh Patches
(c) Security Repair (d) Security patches
(e) Integrity Repair
60. Which message denotes the request has been fulfilled and a new source is created?
(a) 202 (b) 201 (c) 403
(d) 404 (e) 302

## SOLUTIONS

1. (c); An entity set is a collection of similar types of entities. Entity set may contain entities with attribute sharing similar values. For example, Students set may contain all the student of a school; like-wise Teachers set may contain all the teachers of school from all faculties. Entities sets need not to be disjoint.
2. (d); To disallow a method from being overridden, specify final as a modifier at the start of its declaration. Methods declared as final cannot be overridden.
3. (a); A software bug is a problem causing a program to crash or produce invalid output. The problem is caused by insufficient or erroneous logic. A bug can be an error, mistake, defect or fault, which may cause failure or deviation from expected results. Most bugs are due to human errors in source code or its design.
4. (c); The structural testing is the testing of the structure of the system or component. Structural testing is often referred to as 'white box' or 'glass box' or 'clear-box testing' because in structural testing we are interested in what is happening 'inside the system or application'.
5. (b); Integration testing is a software testing methodology used to test individual software components or units of code to verify interaction between various software components and detect interface defects. Components are tested as a single group or organized in an iterative manner. After the integration testing has been performed on the components, they are readily available for system testing.
6. (d); SIM stands for set interrupt mask which is handling the interrupts in 8085 microprocessors.
7. (c); A proxy firewall is a network security system that protects network resources by filtering messages at the application layer. A proxy firewall may also be called an application firewall or gateway firewall.
8. (a); Public clouds are owned and operated by companies that offer rapid access over a public network to affordable computing resources. With public cloud services, users don't need to purchase hardware, software, or supporting infrastructure, which is owned and managed by providers.

9. (a); Asymmetric cryptography, also known as public key cryptography, uses public and private keys to encrypt and decrypt data. One key in the pair can be shared with everyone; it is called the public key. The other key in the pair is kept secret; it is called the private key.
10. (a); <a> tag defines a hyperlink, which is used to link from one page to another.
11. (b); The while loop evaluates the test expression. If the test expression is true (nonzero), codes inside the body of while loop are executed. The test expression is evaluated again. The process goes on until the test expression is false. When the test expression is false, the while loop is terminated.
12. (d); A metamorphic virus is one that can transform based on the ability to translate, edit and rewrite its own code. It is considered the most infectious computer virus, and it can do serious damage to a system if it isn't detected quickly.
13. (b); The LEFT JOIN keyword returns all records from the left table, and the matched records from the right table. The result is NULL from the right side, if there is no match.
    Syntax:
    SELECT column_name(s)
    FROM table1
    LEFT JOIN table2 ON table1.column_name
    = table2.column_name;
14. (a); insertion sort is the best sort when the list is already sorted as the no of comparison reduces its complexity is n.
15. (e); Asymmetric cryptography, also known as public key cryptography, uses public and private keys to encrypt and decrypt data. The keys are simply large numbers that have been paired together but are not identical (asymmetric). One key in the pair can be shared with everyone; it is called the public key.
16. (c); Broadcast storms can lead to network failure by consuming CPU and bandwidth resources.
17. (b); The number of usable host addresses on a subnet with mask 255.255.255.0 is 2^8 minus 2 equals 254.
18. (d); Layer 6 of the OSI model is named the presentation layer and is responsible for character code translation (i.e. ASCII vs. EBCDIC vs. Unicode), data conversion, compression, and encryption.
19. (b); Timing refers to two vital characteristics: when data should be sent and how fast they can be sent. For example, if a sender produces data at 100 Mbps but the receiver can process data at only 1 Mbps, the transmission will overload the receiver and data will be largely lost.
20. (b); An Encapsulating Security Payload (ESP) is a protocol within the IPSec for providing authentication, integrity and confidentially of network packets data/payload in IPv4 and IPv6 networks. ESP provides message/payload encryption and the authentication of a payload and its origin within the IPSec protocol suite.
21. (a); 503 shows when the server is currently unavailable (overloaded or down).
22. (b); A POST request is used to send data to the server, for example, customer information, file upload, etc. using HTML forms.
23. (d); After semantic analysis the compiler generates an intermediate code of the source code for the target machine. It represents a program for some abstract machine. It is in between the high-level language and the machine language. This intermediate code should be generated in such a way that it makes it easier to be translated into the target machine code.
24. (c); Network security audit helps to determine the effectiveness of network security to resolving underlying network security issues. Network security audits are critical to understanding how well your organization is protected against security threats, whether they are internal or external. A network security audit is part of an overall information systems audit framework and includes application software audit, operation system audit, and business audit.
25. (a); C++ is an object-oriented programming language specifically designed to provide a simple, comprehensive feature set for programming modern applications without loss in performance. C++ follows the principle of software engineering that suggests each component should be highly cohesive and the collection of components should be loosely coupled.
26. (b); Multicasting is a transmission method in which copies of a single packet is transmitted to the group of the host in the network interested in receiving the packet.
27. (e); Transaction is a unit of program execution that accesses and possibly updates various data items. There are five transaction states:
    - Active: This is the initial state, the transaction stays in this state while it is executing.
    - Partially committed: When a transaction executes its final operation, it is said to be in a partially committed state.
    - Failed: when the normal execution can no longer proceed.
    - Aborted: after the transaction has been rolled back and the database has been restored to its state prior to the start of the transaction.
    - Committed: If a transaction executes all its operations successfully, it is said to be committed. All its effects are now permanently established on the database system.
28. (a); Phreaker is a telecom network hacker who hacks a telephone system illegally to make calls without paying for them.
29. (c); Salami attacks are a series of minor data security attacks that together result in a larger attack. For example, deducting a very small amount of money from a bank account which is not noticeable. But when the deduct very small amounts from large number of accounts, it become a huge amount.
30. (c); The SQL UNION operator is used to combine the result sets of 2 or more SELECT statements. It removes duplicate rows between the various SELECT statements.
    Syntax:
    SELECT column_name(s) FROM table1
    UNION
    SELECT column_name(s) FROM table2;

31. (a); XaaS or 'anything as a service' is the delivery of IT as a Service through hybrid Cloud computing and is a reference to either one or a combination of Software as a Service (SaaS), Infrastructure as a Service (IaaS), Platform as a Service (PaaS). communications as a service (CaaS) or monitoring as a service (Maas). XaaS is quickly emerging as a term that is being readily recognized as services that were previously separated on either private or public Clouds are becoming transparent and integrated.
32. (a);
33. (a);
34. (b); A database instance is a set of memory structures that manage database files. Causes of instance Failure- a power outage, a server hardware failure, failure of an oracle background process, emergency shutdown procedures.
35. (c); const keyword can be applied to the declaration of any variable to specify that its value will not be changed.
36. (b); 'C' represent cloud in 12c oracle database. The Oracle Database 12c is a high-performance, enterprise-class database. According to Oracle, this is "the first database designed for the cloud."
37. (b); As per First Normal Form, no two Rows of data must contain repeating group of information i.e each set of columns must have a unique value, such that multiple columns cannot be used to fetch the same row.
38. (c); Data model is a collection of conceptual tools for describing data, data relationships, data semantics and consistency constraints. Provides a way to describe the design of a database at the physical, local and view levels.
39. (c); Normalization is the process of organizing the columns (attributes) and tables (relations) of a relational database to reduce data redundancy and improve data integrity.
40. (a); Ageing is a scheduling technique used to avoid starvation. Fixed priority scheduling is a scheduling discipline, in which tasks queued for utilizing a system resource are assigned a priority each.
41. (d); Capability Maturity Model because CMM is not a software life cycle model. Instead, it is a strategy for improving the software process.
42. (b); Turnaround time is the time interval from the time of submission of a process to the time of the completion of the process.
43. (b); Acceptance testing, a testing technique performed to determine whether or not the software system has met the requirement specifications. This is a type of testing done by users, customers, or other authorized entities to determine application/software needs and business processes.
44. (b); Cast operator can be used to explicitly convert the value of an expression to a different data type. To do so, the name of the data type to which the conversion is to be made is enclosed in parenthesis and placed directly to the left of the expression whose value is to be converted.
45. (c); OLTP (On-line Transaction Processing) is characterized by a large number of short on-line transactions (INSERT, UPDATE, DELETE). The main emphasis for OLTP systems is put on very fast query processing, maintaining data integrity in multi-access environments and an effectiveness measured by number of transactions per second.
46. (b); A thread is the smallest unit of processing that can be performed in an OS.
47. (b); All versions of Ethernet use a Media Access Control (MAC) protocol called CSMA/CD. CSMA/CD (Carrier Sense Multiple Access /Collision Detection) a set of rules determining how network devices respond when two devices attempt to use a data channel simultaneously.
48. (b); The Logical Link Control (LLC) layer is one of two sub layers that make up the Data Link Layer of the OSI model. The Logical Link Control layer controls frame synchronization, flow control and error checking.
49. (b); Memory Buffer Register holds the contents of data or instruction read from, or written in memory. It means that this register is used to store data/instruction coming from the memory or going to the memory.
50. (e); Any of the above editors can be used to type php code and run it.
51. (a); The (+) operator can also be used to concatenate two or more strings.
52. (b); A class contains constructors that are invoked to create objects from the class blueprint. Constructor declarations look like method declarations—except that they use the name of the class and have no return type.
53. (b); Method overloading is one of the way java supports static polymorphism. Here we have two definitions of the same method add () which add method would be called is determined by the parameter list at the compile time. That is the reason this is also known as compile time polymorphism.
54. (c); Abstraction is a way of converting real world objects in terms of class. It's a concept of defining an idea in terms of classes or interface.
55. (b); The code generator takes the optimized representation of the intermediate code and maps it to the target machine language. The code generator translates the intermediate code into a sequence of (generally) re-locatable machine code. Sequence of instructions of machine code performs the task as the intermediate code would do.
56. (b); Tree is a hierarchical data structure, with a root value and sub trees of children with a parent node, represented as a set of linked nodes.
57. (b); Spiral Model is a combination of a waterfall model and iterative model. Each phase in spiral model begins with a design goal and ends with the client reviewing the progress. Risk Management is the most important feature of spiral Model.
58. (d); Merge Sort is an efficient, general-purpose, comparison-based sorting algorithm. Complexity of merge sort is O (n log n).
59. (d); Security Patches is a program that eliminates a vulnerability exploited by malicious hackers.
60. (b); 201 created shows when the request has been fulfilled, resulting in the creation of a new resource.

# Professional Knowledge Practice Set:07

1. Which process is the examination and alteration of a system to reconstitute it in a new form?
(a) Software reengineering
(b) Software reverse engineering
(c) Software forward engineering
(d) Software testing
(e) Software costing
2. Which level of database architecture representing total view of database?
(a) Conceptual view (b) Internal view
(c) External view (d) Physical view
(e) None of these
3. Which SQL operator allows you to specify multiple values in a WHERE clause?
(a) Between (b) Select
(c) Joins (d) In
(e) Like
4. Which among the following is an advantage of linked list representation of binary trees over arrays?
(a) dynamic size
(b) ease of insertion/deletion
(c) ease in randomly accessing a node
(d) both dynamic size and ease in insertion/deletion
(e) None of the above
5. SELECT __________
FROM employee
WHERE dept name= 'Finance';
Which of the following should be used to find the mean of the salary?
(a) Count(salary) (b) Sum(salary)
(c) Mean(salary) (d) Avg(salary)
(e) In(salary)
6. How can you delete the records where the "FirstName" is "Shivam" in the Persons Table, in SQL?
(a) DELETE FROM Persons WHERE FirstName = 'Shivam';
(b) DELETE FirstName='Shivam' FROM Persons;
(c) DELETE ROW FirstName='Shivam' FROM Persons;
(d) DELETE FirstName='Shivam' INTO Persons;
(e) DELETE INTO Persons, FirstName='Shivam';
7. What is the output of the given program:
```
#include <iostream>
using namespace std;
int main() {
cout << "Friends forever";
return 0;    }
```
(a) Compiler Error (b) Friends forever
(c) Run time error (d) forever
(e) Friends
8. If different properties and functions of a real-world entity is grouped or embedded into a single element, what is it called in OOP language?
(a) Inheritance (b) Polymorphism
(c) Abstraction (d) Encapsulation
(e) Generalization
9. Which keyword is used to define a scope where identifiers like variables, functions, classes, etc are declared in C++?
(a) namespace (b) mutable
(c) long (d) while
(e) int
10. Which among the following is not a necessary condition for constructors?
(a) Its name must be same as that of class
(b) It must not have any return type
(c) It must contain a definition body
(d) It can contain arguments
(e) None of these
11. SELECT * FROM employee WHERE dept_name="CompSci";
In the SQL given above there is an error. Identify the error.
(a) Dept_name (b) Employee
(c) "CompSci" (d) From
(e) No Error
12. What is the output of this program?
```
class main_class {
public static void main(String args[])
{
int a = 6;
if (a == 6) {
int a = 5;
System.out.println(a);
}
}
}
```
(a) 6 (b) 5
(c) 11 (d) Compiler error
(e) Runtime error
13. Which of the following is not a DDL command?
(a) CREATE (b) ALTER
(c) DROP (d) UPDATE
(e) TRUNCATE
14. Which of the following is the correct syntax for java main method?
(a) public void main(String[] args)
(b) public static void main(string[] args)
(c) public static void(string args)
(d) public static void main(String[] args)
(e) public void main(String args)
15. Ethernet uses how many bit/byte physical address, that is imprinted on the network interface card?
(a) 64 bit (b) 64 byte
(c) 6 bit (d) 6 byte
(e) 32 bit
16. DNS is defining at which layer of TCP/IP model?
(a) Internet layer (b) Data link layer
(c) Transport layer (d) Application layer
(e) Both (a) and (c)

17. Three-way hand shake method used in which of the following?
(a) Connection establishment
(b) Connection termination
(c) Connection release
(d) Path finding
(e) packet forwarding
18. Which type of security attack occurs, when a malicious actor inserts himself as a relay/proxy into a communication session between people or systems?
(a) Compromised key attack
(b) Sniffer Attack
(c) Man-in-the-Middle attack
(d) Denial-of-service attack
(e) Eavesdropping
19. What is use of "touch" command in Linux?
(a) create a file (b) remove a file
(c) Show the manual page (d) Show home directory
(e) Copy files
20. What is SGML?
(a) Standard generalized making loop
(b) Special generalized markup language
(c) Standard general markup language
(d) Short generalized markup language
(e) Standard generalized markup language
21. HTML tags are used to describe document ___________.
(a) Content (b) Definition
(c) Language (d) Table
(e) Grammar
22. _______ is passed to a method by use of call-by-reference.
(a) Variables (b) Objects
(c) Methods (d) Operators
(e) Encapsulation
23. In Java, each thread has its own ________, in which it runs?
(a) main() method (b) JVM
(c) Call stack (d) Memory
(e) Shell
24. What is cfront?
(a) cfront is the front end of a C compiler
(b) cfront is the pre-processor of a C compiler
(c) cfront is a tool that translates a C++ code to its equivalent C code
(d) cfront used to print
(e) None of these
25. Which of the following functionality of transport layer over and above the network Layer?
(a) Recovery from packet losses
(b) Detection of duplicate packets
(c) Packet delivery in the correct order
(d) End to end connectivity
(e) All of the above
26. Which clause is used to specify condition while retrieving data from table in SQL?
(a)WHERE (b)LIKE
(c)HAVING (d)Group By
(e) Select
27. virtual memory is commonly implemented by which technique?
(a) Virtualization (b) Demand paging
(c) Thrashing (d) Processing
(e) None of these
28. What is indexed allocation of file system implementation?
(a) each file must occupy a set of contiguous blocks on the disk
(b) all the pointers to scattered blocks are placed together in one location
(c) each file is a linked list of disk blocks
(d) Both (a) and (b)
(e) none of the mentioned
29. Which of the following is used to extract the value of an XML element and add it to the output stream of the transformation?
(a) <xsl:value-of> (b) <xsl:val-of>
(c) <xl:value-of> (d) <xdf:value-of>
(e) <xdg: value-of>
30. A table that contains the base address of each page in physical memory is called…..
(a) Page table (b) Frame table
(c) Memory table (d) Base Table (d) Proto Table
31. __________ defines the syntax of a programming language.
(a) Token (b) Pattern
(c) Context-Free Grammar (d) Structure
(e) None of the above
32. Which of the following is Big Data Technology?
(a)NoSQL Database (b)Hadoop
(c)Spark (d)Kafka
(e)All of the above
33. If a relation has several candidate keys, one is chosen arbitrarily to be the __________ key?
(a) Primary key (b) Secondary key
(c) Foreign key (d) Candidate key
(e) Attribute key
34. _______ is the part of malware such as worms or viruses which performs the malicious action; deleting data, sending spam or encrypting data.
(a) Denial of Service (b) Exploits
(c) Scams (d) Payload
(e) Spamming
35. Which of the following is the IEEE standard specification for Token Ring?
(a) IEEE 802.4 (b) IEEE 802.5
(c) IEEE 802.1 (d) IEEE 802.3
(e) IEEE 802.2

36. Which of the following are used as the simplest form of error detecting code in networking?
(a) Parity bits (b) Network bits
(c) Data bits (d) security bit
(e) All of the Above

37. Which one of the following task is not done by transport layer?
(a) connection control
(b) error control
(c) flow control
(d) process-to-process message delivery
(e) channel coding

38. Which among the following backs up all the data that has changed?
(a) Copy backup (b) Incremental backup
(c) Differential backup (d) Full backup
(e) None of these

39. 128. 167. 152. 2 belongs to which address class?
(a) A (b) B
(c) C (d) D
(e) E

40. Which HTTP message shows "Request Timeout"?
(a) 408 (b) 409
(c) 302 (d) 202
(e) 103

41. The person who use his skills for security of organization is known as:
(a) Black Hat Hacker (b) White Hat Hacker
(c) Script Kiddie (d) Phreaker
(e) Attacker

42. Which of the following is commonly used between hosts on the Internet to exchange routing information?
(a) HTTP (b) EGP
(c) UDP (d) SMTP
(e) IGP

43. Which of the following is SMTP port number?
(a) 80 (b) 23
(c) 25 (d) 21
(e) 110

44. which of the following is used to maintain the accuracy and integrity of data inside table?
(a) trigger (b) joins
(c) constraints (d) Schema
(e) Cardinality

45. 802. 10 IEEE standard used for:
(a) Wireless networking (b) Ethernet
(c) Broadband technology (d) Network security
(e) Token ring

46. What is the use of ESP in IPSec?
(a) Provide encryption and authentication
(b) Negotiate Security Associations
(c) Provide data integrity
(d) Manage parameter
(e) Both (a) and (c)

47. Which unix command is used to look at file and one page at a time?
(a) man (b) more
(c) grep (d) lpr
(e) rm

48. Choose the odd one:
(a) Boolean (b) Catch
(c) Instanceof (d) New
(e) struct

49. Which privilege is required to create a database in Oracle?
(a) SYSDBA (b) DBA
(c) SYSOPER (d) RESOURCE
(e) CREATE

50. Which of the following is the reduction of a particular body of data to a simplified representation of the whole?
(a) Data Isolation (b) Data Migration
(c) Database Abstraction (d) Database Instance
(e) Database Schema

51. The processes that are residing in main memory and are ready and waiting to execute are kept on a list called:
(a) job queue (b) ready queue
(c) execution queue (d) process queue
(e) priority queue

52. In SQL server, indexes are organized as:
(a) List (b) Array
(c) Queue (d) B-tree
(e) Stack

53. Convert octal number $(255)_8$ to its binary equivalent?
(a) $(11001100)_2$ (b) $(10101101)_2$
(c) $(10101010)_2$ (d) $(11100011)_2$
(e) $(11000110)_2$

54. How is a server different from a workstation computer?
(a) The server works as a standalone computer.
(b) The server provides services to clients.
(c) The workstation has fewer applications installed.
(d) The workstation has more users who attach to it.
(e) There is no difference between server and workstation

55. Which OSI layer is responsible for establishing, managing, and terminating connections between applications at each end of the communication?
(a) Transport layer (b) Presentation layer
(c) Session layer (d) Application layer
(e) Physical layer

56. Which of the following is the process of aggregating two or more networks for which it generates single IP address?
(a) Subnetting (b) Supernetting
(c) Routing (d) Indexing
(e) Multinetting

57. IP address 135. 58. 24. 17 belongs to which class of IPv4 address?
(a) Class A
(b) Class B
(c) Class C
(d) Class D
(e) Class E
58. Which feature works with both Microsoft Explorer and Microsoft Edge and monitors websites, pages, and downloads for known issues?
(a) AgentX
(b) Blocker
(c) Armenia
(d) SmartScreen
(e) None of there
59. A free file storage and file sharing service provided by Microsoft when you sign up for a Microsoft account is:
(a) OneDrive
(b) Google Drive
(c) iCloud
(d) AWS
(e) D-cloud
60. Polymorphism reduces the effort required to extend an object system by_____.
(a) Coupling objects together more tightly
(b) Enabling a number of different operations to share the same name
(c) Making objects more dependent on one another
(d) Removing the barriers imposed by encapsulation.
(e) None of these

## SOLUTIONS

1. (a); Software Reengineering is the examination and alteration of an existing subject system to reconstitute it in a new form. This process encompasses a combination of sub-processes such as reverse engineering, restructuring, redocumentation, forward engineering, and retargeting.
2. (a); Conceptual level describes the structure of the whole database for a group of users. Conceptual schema is a representation of the entire content of the database. These schemas contain all the information to build relevant external records. It hides the internal details of physical storage.
3. (d); The IN operator allows you to specify multiple values in a WHERE clause.
Syntax:
SELECT column_name(s)
FROM table_name
WHERE column_name IN (value1, value2, ...);
4. (d)
5. (d); The AVG() function returns the average value of a numeric column.
Syntax:
SELECT AVG(column_name)
FROM table_name
WHERE condition;
6. (a); The DELETE statement is used to delete existing records in a table.
Syntax:
DELETE FROM table_name
WHERE condition;
7. (b)
8. (d); Encapsulation is one of the fundamental concepts in object-oriented programming (OOP). It describes the idea of bundling data and methods that work on that data within one unit, e.g., a class in Java. This concept is also often used to hide the internal representation, or state, of an object from the outside.
9. (a); A namespace is a declarative region that provides a scope to the identifiers (the names of types, functions, variables, etc) inside it. Namespaces are used to organize code into logical groups and to prevent name collisions that can occur especially when your code base includes multiple libraries.
10. (c); Constructors are predefined implicitly, even if the programmer doesn't define any of them. Even if the programmer declares a constructor, it's not necessary that it must contain some definition.
11. (c); For any string operations single quoted (') must be used to enclose.
12. (d); Two variables with the same name can't be created in a class. compiler error generates.
13. (d); DDL (Data Definition Language) commands are used for creating, modifying, and dropping the structure of database objects. The commands are CREATE, ALTER, DROP, RENAME, and TRUNCATE.
UPDATE is a DML (Data Manipulation Language) command.
14. (d)
15. (d) Ethernet uses a 6 -byte (48-bit) physical address that is imprinted on the network interface card (NIC). The physical address also known as the link address is the address of a node physical address, also known as the link address, is the address of a node as defined by its LAN or WAN. It is included in the frame used by the data link layer.
16. (d); One of the most important components of the application layer is the Domain Name System (DNS) server. DNS is a distributed hierarchical and global directory that translates machine or domain names to numerical IP addresses. DNS can run over either UDP or TCP.
17. (a) A three-way handshake is a method used in a TCP/IP network to create a connection between a local host/client and server. It is a three-step method that requires both the client and server to exchange SYN and ACK (acknowledgment) packets before actual data communication begin.
A three-way handshake is also known as a TCP handshake.

18. (c); A man-in-the-middle attack is a type of cyberattack where a malicious actor inserts him/herself into a conversation between two parties, impersonates both parties and gains access to information that the two parties were trying to send to each other. A man-in-the-middle attack allows a malicious actor to intercept, send and receive data meant for someone else, or not meant to be sent at all, without either outside party knowing until it is too late.
19. (a); The touch command is used to create a file. It can be anything, from an empty txt file to an empty zip file.
20. (e); SGML (Standard Generalized Markup Language) is a standard for how to specify a document markup language or tag set. Such a specification is itself a document type definition (DTD). SGML is not in itself a document language, but a description of how to specify one.
21. (a);
22. (b);
23. (c); Thread of execution means an individual 'lightweight' process that has its own call stack. In java each thread has its own call stack.
24. (c); cfront is a tool that translates a C++ code to its equivalent C code
25. (d); Transport Layer is responsible for delivering messages between hosts. In Transport layer, data travels in the form of segments. Transport layer is responsible for creating an end to end connection between source IP and the destination.
26. (a); The WHERE clause is used to extract only those records that fulfill a specified condition.
27. (b); Demand paging is the process of loading the page into memory on demand (whenever page fault occurs) is known as demand paging. Virtual memory is implemented using Demand Paging or Demand Segmentation.
28. (b); In index allocation a special block known as the Index block contains the pointers to all the blocks occupied by a file. Each file has its own index block. The ith entry in the index block contains the disk address of the ith file block.
29. (a);
30. (a);
31. (c);
32. (e); Big data management is the organization, administration and governance of large volumes of both structured and unstructured data. The goal of big data management is to ensure a high level of data quality and accessibility for business intelligence and big data analytics applications. All these technologies are Big Data Technology.
33. (a) A candidate key is a column, or set of columns, in a table that can uniquely identify any database record without referring to any other data. Each table may have one or more candidate keys, but one candidate key is unique, and it is called the primary key.
34. (d) In computer security, the payload is the part of malware such as worms or viruses which performs the malicious action; deleting data, sending spam or encrypting data. In addition to the payload, such malware also typically has overhead code aimed at simply spreading itself, or avoiding detection.
35. (b); Token ring is a network where all computers are connected in a circular fashion. The term token is used to describe a segment of information that is sent through that circle; when a computer on the network can decode that token, it receives data. The token ring is used by ARCNET (Attached Resource Computer network), token bus and FDDI (Fiber Distributed Data Interface). IEEE standard of token ring is IEEE 802.5.
36. (a); A parity bit is a bit, with a value of 0 or 1, that is added to a block of data for error detection purposes. It gives the data either an odd or even parity, which is used to validate the integrity of the data.
37. (e); Channel coding done on presentation layer.
38. (c); Differential backup which backs up all data that has changed since the last full backup. Differential backup will refer the "Archive Ready" bit to find the file has been changed, but after backup is performed, the "Archive Ready" is not changed. Differential backup will include the same backup as the previous differential backup plus any additional files that have changed after the last Differential backup.
39. Ans-(b) Range of Class B address is 128.0.0.0 to 191.255.255.255
40. (a); 408 shows Request Timeout means the server timed out waiting for the request.
41. (b); White Hat hacker has the skills to break into networks but he uses his skills to protect organizations. A White Hat hacker can conduct vulnerability assessments and penetration tests are also known as an Ethical Hacker. Often White Hat hackers are employed by companies and organizations to check the vulnerabilities of their network and make sure that no hole is available in their network for an intruder.
42. Ans-(b); Exterior Gateway Protocol (EGP) is a protocol for exchanging routing information between two neighbor gateway hosts (each with its own router) in a network of autonomous systems. EGP is commonly used between hosts on the Internet to exchange routing table information.

43. (c)
44. (c); Constraints enforce limits to the data or type of data that can be inserted/updated/deleted from a table. The whole purpose of constraints is to maintain the data integrity during an update/delete/insert into a table.
45. (d); IEEE 802.10 provides specifications for an interoperable data link layer security protocol and associated security services.
46. (e); IPSec uses ESP (Encapsulating Security Payload) to provide Data Integrity, Encryption, Authentication, and Anti-Replay functions for IPSec VPN. Cisco IPSec implementations use DES, 3DES and AES for Data Encryption. ESP authenticates the data within the VPN, ensuring Data Integrity and that it is coming from the correct source.
47. (b)
48. (e); Except 'struct' all other are java keywords. Struct is a keyword of C programming.
49. (a); The SYSDBA system privilege is for fully empowered database administrators.
50. (c); Data abstraction is the reduction of a particular body of data to a simplified representation of the whole
51. (b); Ready queue keeps a set of all processes residing in main memory, ready and waiting to execute. A new process is always put in this queue.
52. (d); In SQL Server, indexes are organized as B-trees. Each page in an index B-tree is called an index node. The top node of the B-tree is called the root node. The bottom level of nodes in the index is called the leaf nodes. Any index levels between the root and the leaf nodes are collectively known as intermediate levels. In a clustered index, the leaf nodes contain the data pages of the underlying table. The root and intermediate level nodes contain index pages holding index rows. Each index row contains a key value and a pointer to either an intermediate level page in the B-tree, or a data row in the leaf level of the index. The pages in each level of the index are linked in a doubly-linked list.
53. (b); Converting from octal to binary is as easy as converting from binary to octal. Simply look up each octal digit to obtain the equivalent group of three binary digits.

| OCTAL | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| BINARY | 000 | 001 | 010 | 011 | 100 | 101 | 110 | 111 |

Step 1 − Convert each octal digit to a 3-digit binary number (the octal digits may be treated as decimal for this conversion).
$(2)_{10} = (010)_2$
$(5)_{10} = (101)_2$
$(5)_{10} = (101)_2$
Step 2 − Combine all the resulting binary groups (of 3 digits each) into a single binary number.
$(255)_8 = (10101101)_2$
54. (b); Server and Workstation both are different from each other because they are designed for performing different tasks and functions. A server is a physical computer dedicated to host or run to one or more services to serve the needs of users and other computers on a network. Workstation is different and narrower term than server. It is a kind of computer/system, which is designed for performing personal use. Mostly they are manufactured for performing high-level task like scientific, technical and operational work.
55. (c); Session layer is the fifth layer of seven layered Open Systems Interconnection (OSI) Model. The session layer is responsible for establishing, managing, and terminating connections between applications at each end of the communication.
56. (b); Supernetting is the method for combining two or more contiguous network address spaces to simulate a single, larger, address space.
57. (b); In a Class B network, the first 16 bits are the network part of the address. All Class B networks have their first bit set to 1 and the second bit set to 0. In dotted decimal notation, that makes 128.0.0.0 to 191.255.0.0 as Class B networks. There are 16,384 possible Class B networks.
58. (d); SmartScreen Filter helps you identify reported phishing and malware websites and also helps you make informed decisions about downloads.
59. (a); OneDrive (formerly SkyDrive) is an online cloud storage service from Microsoft. OneDrive integrates with Microsoft Office so users can access Word, Excel and PowerPoint documents in OneDrive. The system allows users to simultaneously edit Office documents, edit documents in browsers, and create and share folders.
60. (b); Polymorphism reduces the effort required to extend an object system by enabling a number of different operations to share the same name.

# Professional Knowledge Practice Set:08

1. The Process of loading an operating system into the computer's main memory or Random-Access Memory (RAM) is known as:
(a) Throughput (b) Boot
(c) Waiting Time (d) Thrashing
(e) Running Time
2. The if..else statement can be replaced by which operator?
(a) Bitwise operator (b) Conditional operator
(c) Multiplicative operator (d) Arithmetic operator
(e) Comma operator
3. How much time will be required to search the minimum element in a binary max heap having n number of nodes?
(a) O(1) (b) O(nlogn)
(c) O(n) (d) O(logn)
(e) O(n^2)
4. Which key is used to verify the signature in digital signature technique of network security?
(a) Private key (b) Secret key
(c) Sender's public key (d) Sender's private key
(e) Symmetric key
5. Select name, course_id
from instructor, teaches
where instructor_ID= teaches_ID;
This Query can be replaced by which one of the following ?
(a) Select name,course_id from teaches,instructor where instructor_id=course_id; (b) Select name, course_id from instructor natural join teaches;
(c) Select name ,course_id from instructor; (d) Select course_id from instructor join teaches;
(e) None of these
6. Which clause compares data with an expression using wildcard operators?
(a) WHERE (b) USING
(c) LIKE (d) DISTINCT
(e) JOIN
7. Which UNIX command is used to copy and convert?
(a) df (b) vi
(c) mv (d) dd
(e) ping
8. What is the port number of DNS(Domain Name Server)?
(a) 20 (b) 21
(c) 53 (d) 50
(e) 51
9. Which SQL command is used to remove all the records from a table along with their allocated spaces?
(a) ALTER (b) SELECT
(c) ROLLBACK (d) COMMIT
(e) TRUNCATE
10. What is the use of scanf function in C?
(a) used to print (b) used to read character, string, numeric data from keyboard
(c) initialize the program (d) used to run program
(e) None of these
11. If condition is missing in for loop, then infinite looping can be avoided by which statement?
(a) continue statement (b) goto statement
(c) return statement (d) break statement
(e) only (b) , (c) and (d)
12. Suppose we have a system in which processes is in hold and wait condition then which of the following approach prevent the deadlock.
(a) Request all resources initially
(b) S33333pool everything
(c) Take resources away
(d) Order resources numerically
(e) None of these
13. Which of the following is a secret-key Algorithm?
(a) RSA (b) DES
(c) Public Key (d) Both (a) and (b)
(e) All of the above
14. which method is used to refresh a web page in javascript?
(a) window.reload (b) location.reload
(c) window.refresh (d) page.refresh
(e) page.reload
15. Which of the following are two types of basic adder circuits?
(a) Sum and Carry
(b) Asynchronous and Synchronous
(c) Half-adder and Full-adder
(d) One and Two's Complement
(e) None of these
16. Which of the following is the correct statement to select all record from table STUDENT where subject is COMPUTER?
(a) Select from STUDENT where subject='COMPUTER';
(b) Select from STUDENT where subject="COMPUTER";
(c) Select all from STUDENT where subject='computer';
(d) Select * from STUDENT where subject='Computer';
(e) Select * from STUDENT where subject=='COMPUTER';

17. Which of the following is consider as a super key, if attribute ID, CITY and attribute NAME is given?
(a) NAME (b) NAME and CITY
(c) CITY (d) ID
(e) None of these

18. Which of the following is not a built in aggregate function in SQL?
(a) avg (b) max
(c) total (d) count
(e) None of these

19. Which of the following is used to transfer data intended for a peripheral device into a disk so that it can be transmitted to the peripheral device at a more suitable time or in bulk?
(a) Virtual programming (b) Multi programming
(c) Multitasking (d) Spooling
(e) Paging

20. What is Multiple Inheritance?
(a) Deriving a class from Base class
(b) Deriving a derived class from two and more class
(c) Deriving two or more class from a base class
(d) two methods with the same method name and parameters
(e) None of these

21. Checkpoints are a part of which of the following measures?
(a) Recovery measure (b) Security measure
(c) Concurrency measure (d) Authorized measure
(e) Design measures

22. In which mechanism one class acquires the property of another class?
(a) Abstraction (b) Inheritance
(c) Polymorphism (d) Encapsulation
(e) Method

23. Which of the following is the correct syntax of UPDATE Command?
(a) UPDATE table_name where condition SET column_name = <value>
(b) UPDATE table_name SET column_name = <value> WHERE <condition>
(c) UPDATE table table_name set column_name = <value> WHERE <condition>
(d) UPDATE IN table_name WHERE condition set column_name = <value>
(e) UPDATE IN table_name LIKE condition set column_nam =<value>

24. What is Optimal Page Replacement algorithm?
(a) Replace the page that will not be used for a long time
(b) Replace the page that has been used for a long time
(c) Replace the Page that has not been used for a long time
(d) Replace Page not in memory
(e) Used to extract page in main memory

25. Which of the following concepts means wrapping up of data and functions together?
(a) Abstraction (b) Encapsulation
(c) Inheritance (d) Polymorphism
(e) Generalization

26. If requirements are easily understandable and defined, then which model is best suited?
(a) Spiral model (b) Waterfall model
(c) Prototyping model (d) Incremental model
(e) Evolution model

27. A server farm is an example of:
(a) Server clustering (b) Redundant servers
(c) Multiple servers
(d) Server fault tolerance
(e) None of these

28. The ability to modify the internal schema without causing any change to the external scheme:
(a) Physical data independence
(b) Logical data independence
(c) External data independence
(d) Both Physical and External data independence
(e) None of these

29. To commit a transaction means making the changes:
(a) Temporary (b) Permanent
(c) Rolled back (d) Parsed
(e) Failed

30. The files that contains settings that apply to the all of the ASP.NET application:
(a) Web.config
(b) ApplicationName.config
(c) Machine.config
(d) Host.config
(e) None of these

31. Which of the following is a form of software that automatically displays ads such as banners or pop-ups when a user is online.
(a) Adware (b) Spyware
(c) Scumware (d) Greyware
(e) Hostware

32. The primary key indexing technique does not allow:
(a) Duplicate data in a field
(b) Multiple attributes
(c) Sets of relation
(d) One to many relation
(e) None of these

33. A given relation is known to be in third normal form. Select the statement which can be inferred from this:
(a) All attributes contribute to the primary key
(b) Each non-key attribute determines the primary key
(c) Each non-key attribute is determining by the primary key
(d) Every determinant is a candidate key
(e) None of the above

34. Java servlets are an efficient and powerful solution for creating _______for the web.
(a) Dynamic Content (b) Static Content
(c) Hardware (d) Both a and b
(e) links
35. Which HTML tag is used to make a bulleted list with numbers?
(a) <dl> (b) <ol>
(c) <list> (d) <ul>
(e) <h1>
36. Which of the following is not a domain suffix?
(a) .edu (b) .org
(c) .com (d) .int
(e) .net
37. Which of the following term is similar to phishing, but refers to fraudulent messages sent over SMS rather than email?
(a) Flaming (b) Pharming
(c) Mashup (d) Smishing
(e) Spooling
38. Two methods are said to be overloaded if they have_____.
(a) Same name and same number of parameter but different return type.
(b) They have same name.
(c) They have different name but same number of argument.
(d) Have same name but different parameters.
(e) None of these
39. which java keyword is used to stop the execution of current iteration and start the execution of next iteration in a loop?
(a) Default (b) Continue
(c) Class (d) Try
(e) This
40. what is abstraction in object-oriented programming?
(a) Shows data binding
(b) Showing only the essential features of the application and hiding the details
(c) Reuse once written code again and again
(d) Allows to create function with same name
(e) Defining different forms of a method
41. Which of the following type of attack where the attackers attempt to prevent legitimate users from accessing the service?
(a) Virus attack (b) Worm attack
(c) Denial-of-service attack (d) Botnet process
(e) None of these
42. Which type of service provided by DNS servers?
(a) They run a spell check on host names to ensure accurate routing
(b) They convert domain names into IP address
(c) Given an IP address and they determine the name of the host that is sought
(d) They map individual hosts to their specific IP addresses
(e) Check error message
43. Circuit-level Gateway firewall works at which layer of OSI Model?
(a) Network layer (b) Transport layer
(c) Physical layer (d) Session layer
(e) Datalink layer
44. Which of the following is the practice of testing a computer system, network or web application to find vulnerabilities that an attacker could exploit?
(a) Session Testing (b) Penetration Testing
(c) Firewall testing (d) Packet Testing
(e) None of these
45. In which of the following oops property, different properties and functions of a real-world entity is grouped or embedded into a single element?
(a) Polymorphism (b) Encapsulation
(c) Abstraction (d) Inheritance
(e) Function overloading
46. A function call mechanism that passes arguments to a function by passing a copy of the values of the arguments is _________
(a) call by name (b) call by value
(c) call by reference (d) call by value result
(e) None of these
47. What kind of Protocol is used to provide Internet access from mobile?
(a) TCP/IP (b) ISD
(c) WAP (d) HTTP
(e) UDP
48. Which of the following refers to applications and services that run on a distributed network using virtualized resources.
(a) Distributed (b) Cloud
(c) Soft (d) Parallel
(e) Batch
49. In OSI network architecture, the dialog control and token management are responsibility of Which layer?
(a) Session Layer (b) Network Layer
(c) Transport Layer (d) Data Link Layer
(e) Physical layer
50. Which of the following is the smallest unit of storage in oracle database?
(a) Control file (b) Datafile
(c) Data Block (d) Segment
(e) Clustered
51. A circular list can be used to represents which type of data structure?
(a) A queue (b) B-tree
(c) B++ tree (d) Stack
(e) All of the above

52. Which of the following is/are a PHP code editor?
(a) Notepad
(b) Notepad++
(c) Adobe Dreamweaver
(d) PDT
(e) All of the mentioned.

53. Which of the following is not version of Android?
(a) Nougat
(b) Eclair
(c) Oreo
(d) Jelly Bean
(e) Pancake

54. Which of the following protocol is used to automatically assigns a unique IP address to each device that connects to a network?
(a) DNS
(b) IP
(c) DHCP
(d) HTTP
(e) LLC

55. Which of the following device is used to regenerate the signal, so that increases the signal's strength and it can be transmitted and received over a greater distance without a loss in quality?
(a) Bridge
(b) Router
(c) Repeater
(d) Switch
(e) Gateway

56. Which transaction property ensure that multiple transactions can occur concurrently without leading to inconsistency of database state?
(a) Consistency
(b) Isolation
(c) Durability
(d) Atomicity
(e) Not any property ensures the transaction independency

57. Which type of maintenance mainly deals with implementing new or changed user requirement?
(a) Corrective
(b) Preventive
(c) Adaptive
(d) Perfective
(e) None of these

58. Boundary Value Analysis is based on which design technique?
(a) White box testing
(b) Statement coverage
(c) Path testing
(d) Black box testing
(e) Clear Box testing

59. System threat mechanism by which a hacker can detects system vulnerabilities to make an attack on the system is known as:
(a) Worm
(b) Denial of Service
(c) Port Scanning
(d) Trojan horse
(e) Virus

60. Which of the following file accessing method contains a pointer to a various block?
(a) Sequential Access
(b) Direct Access
(c) Index Access
(d) Random Access
(e) None of these

## SOLUTIONS

1. (b); To boot (as a verb; also "to boot up") a computer is to load an operating system into the computer's main memory or random-access memory (RAM). Once the operating system is loaded (and, for example, on a PC, you see the initial Windows or Mac desktop screen), it's ready for users to run applications.
2. (b); We may replace the use of if..else statements by conditional operators. In the conditional operator, it will predicate the output using the given condition.
3. (c)
4. (c); A digital signature (not to be confused with a digital certificate) is a mathematical technique used to validate the authenticity and integrity of a message, software or digital document. Digital signature use asymmetric key cryptography. The signature can be verified by receiver using sender's public key.
5. (b) Join clause joins two tables by matching the common column .
6. (c)
7. (d); The dd command copies a file, converting the format of the data in the process, according to the operands specified.
8. (c)
9. (e); The SQL TRUNCATE command is used to delete all the rows from the table and free the space containing the table.
10. (b);
11. (e); An infinite loop (sometimes called an endless loop) is a piece of coding that lacks a functional exit so that it repeats indefinitely. In computer programming, a loop is a sequence of instruction s that is continually repeated until a certain condition is reached. In a for loop if condition is missing, then infinite looping can be avoided by goto, return, and break statement.
12. (a); There are two approaches to eliminate hold and wait condition first is allocate all required resources to the process before start of its execution, this way hold and wait condition is eliminated but it will lead to low device utilization and second is require processes to request resources only when it has none. Thus, first they must release all their currently held resources before requesting all the resources they will need from scratch.
13. (b); A secret key algorithm (sometimes called a symmetric algorithm) is a cryptographic algorithm that uses the same key to encrypt and decrypt data. The best-known algorithm is the U.S. Department of Defense's Data Encryption Standard (DES). DES is a symmetric encryption scheme in which both the sender and the receiver need to know the secret key in order to communicate securely.

14. (b)
15. (c)
16. (d); The SELECT statement is used to select data from a database, WHERE clause is used to filter records and (*) is used to select all records from table.
    Syntax:
    SELECT * FROM table_name WHERE condition;
17. (d); Here the id is the only attribute which can be taken as a key. Other attributes are not uniquely identified.
18. (c); SQL does not include total as a built in aggregate function. The avg is used to find average, max is used to find the maximum and the count is used to count the number of values.
19. (d); Spooling is a technique to hold data temporarily in the computer memory and provide to devices when they need it. The temporary holding area of the memory is sometimes referred to as the buffer. One common example of spooling is printing, where files are stored in the buffer and sent to the printer when it is ready.
20. (b); Multiple Inheritance is a feature of C++ where a class can inherit from more than one classes.
21. (a); Checkpoint is a mechanism where all the previous logs are removed from the system and stored permanently in a storage disk. Checkpoint is used in the recovery measure of database.
22. (b); The process by which one class acquires the properties (data members) and functionalities(methods) of another class is called inheritance. The aim of inheritance is to provide the reusability of code so that a class has to write only the unique features and rest of the common properties and functionalities can be extended from another class.
23. (b); The UPDATE statement is used to modify the existing records in a table.
    Syntax:
    UPDATE table_name
    SET column1 = value1, column2 = value2, ...
    WHERE condition;
24. (a); In optimal page replacement algorithm, pages are replaced which are not used for the longest duration of time in the future.
25. (b); Encapsulation is an Object-Oriented Programming concept that binds together the data and functions that manipulate the data, and that keeps both safe from outside interference and misuse. Data encapsulation led to the important OOP concept of data hiding.
26. (b); Waterfall Model is a linear sequential model life cycle model. It is very simple to understand and use. Waterfall model is easy to manage due to rigidity of the model – each phase has specific deliverable and a review process. It works well for smaller projects where requirements are very well understood.
27. (a); A server farm or server cluster is a collection of computer servers - usually maintained by an organization to supply server functionality far beyond the capability of a single machine. Server farms often consist of thousands of computers which require a large amount of power to run and to keep cool.
28. (a); Physical Data Independence the ability to modify the way database stored (data structures, file organization, etc. - internal view of a database) without affecting the next higher level conceptual schema is the physical data independence.
29. (b); Committing a transaction means making permanent the changes performed by the SQL statements within the transaction.
30. (a); The Web.config file for a specific website contains settings that apply to the website and inherit downwards through all the ASP.NET applications and subdirectories of the site.
31. (a); adware is a program that installs an additional component that feeds advertising, often by delivering pop-up ads or by installing a toolbar in your browser.
32. (a); primary key indexing technique does not allow duplication of data file in the data field because primary key is the unique key.
33. (c); A table is in a third normal follow two conditions− first, it is in second normal form and All non-primary fields are dependent on the primary key.
34. (a); java servlets offer a fast, powerful, portable environment for creating dynamic web content.
35. (b); The <ol> tag defines an ordered list. An ordered list can be numerical or alphabetical.
36. (d)
37. (d); SMiShing is a security attack in which the user is tricked into downloading a Trojan horse virus or other malware onto his cellular phone or other mobile device. SMiShing is short for "SMS phishing."
38. (d); If two or more method in a class has same name but different parameters, it is known as Method Overloading.
39. (b); A continue statement is used to end the current loop iteration and return control to the loop statement. continue skips the current executing loop and moves to the next loop.
40. (b); Abstraction refers to, providing only essential information to the outside world and hiding their background details, i.e., to represent the needed information in program without presenting the details.

41. (c); A denial-of-service attack is a security event that occurs when an attacker takes action that prevents legitimate users from accessing targeted computer systems, devices or other network resources.
42. (b); DNS (Domain Name Server) is an internet service that translates domain names into IP addresses.
43. (d); Circuit level gateways are deployed at the Session layer of the OSI model and they monitor sessions like TCP three-way handshake to see whether a requested connection is legitimate or not. Major Screening happens before the Connection is Established. Information sent to a Computer outside the network through a circuit level gateway appears to have originated from the Gateway. This helps in creating a stealth cover for the private network from outsiders.
44. (b); Penetration Testing is the name given to methodologies used to test networks, applications, and even people, for security holes. It is done by trained professionals who identify security vulnerabilities in the IT Systems of an organization, in order to eliminate them before hackers or unauthorized users exploit them.
45. (b); Data Encapsulation is an Object-Oriented Programming concept that bind a group of related properties, functions, and other members are treated as a single unit.
46. (b); The call by value method of passing arguments to a function copies the actual value of an argument into the formal parameter of the function. In this case, changes made to the parameter inside the function have no effect on the argument.
47. (c); Wireless Application Protocol (WAP) is a technical standard for accessing information over a mobile wireless network. A WAP browser is a web browser for mobile devices such as mobile phones that uses the protocol.
48. (b); Cloud Computing applications are accessed by common Internet protocols and networking standards.
49. (a); The session layer in a network model is the network dialog controller. It establishes, maintains, and synchronizes the interaction among communicating systems. Session layer also responsible for token management.
50. (c); The logical units of database space allocation are data blocks.
51. (a); A circular list can be used to represent a queue. Nodes can be accessed easily and deletion of nodes is easier.
52. (e); Any of the above editors can be used to type php code and run it.
53. (e)
54. (c); Dynamic Host Configuration Protocol (DHCP) is a client/server protocol that automatically provides an Internet Protocol (IP) host with its IP address and other related configuration information such as the subnet mask and default gateway.
55. (c); A repeater operates at the physical layer. Its job is to regenerate the signal over the same network before the signal becomes too weak or corrupted so as to extend the length to which the signal can be transmitted over the same network.
56. (b); If there are multiple transactions executing simultaneously, then all the transaction should be processed as if they are single transaction. But individual transaction in it should not alter or affect the other transaction. That means each transaction should be executed as if they are independent.
57. (d); Perfective maintenance includes modifications and updates done in order to keep the software usable over long period of time. It includes new features, new user requirements for refining the software and improve its reliability and performance.
58. (d); Boundary Value Analysis (BVA) is a black box test design technique based on test cases. This technique is applied to see if there are any bugs at the boundary of the input domain. Thus, with this method, there is no need of looking for these errors at the center of this input.
59. (c); A port scan is a series of messages sent by someone attempting to break into a computer to learn which computer network services, each associated with a "well-known" port number, the computer provides. Port scanning, a favorite approach of computer cracker, gives the assailant an idea where to probe for weaknesses.
60. (c); In this method an index is created which contains a key field and pointers to the various block. To find an entry in the file for a key value, we first search the index and then use the pointer to directly access a file and find the desired entry.

# Professional Knowledge Practice Set:09

1. Which of the following is used to create a new process?
(a) Thread (b) Signal
(c) Method (d) fork()
(e) exec()
2. Which of the following allows to create complex entities from existing entities and relationships?
(a) Specialization (b) Generalization
(c) Aggregation (d) Inheritance
(e) Encapsulation
3. Breadth first search uses which of the following method.
(a) Stack (b) Set
(c) List (d) Queue
(e) None of these
4. Which of the following system run more than one process concurrently at a time?
(a) batched system
(b) real-time system
(c) multi programming system
(d) Multiprocessing system
(e) None of these
5. "this" keyword in java is used to?
(a) refer to current class object.
(b) refer to static method of the class.
(c) refer to parent class object.
(d) refer to static variable of the class.
(e) None of these
6. Convert $(126)_8$ to decimal____.
(a) $(84)_{10}$ (b) $(83)_{10}$
(c) $(86)_{10}$ (d) $(87)_{10}$
(e) $(85)_{10}$
7. Which of the following is a form of software that automatically displays ads such as banners or pop-ups when a user is online?
(a) Adware (b) Spyware
(c) Scumware (d) Greyware
(e) Spamware
8. Which of the following allows making copies of the database periodically to help in the cases of crashes and disaster?
(a) Recovery Utility (b) Monitoring Utility
(c) Backup Utility (d) Data loading Utility
(e) Crash Utility
9. Which of the following is a set of Flip flops integrated together?
(a) Adder (b) Memory
(c) Counter (d) Register
(e) Clock
10. Which of the following is not a storage class supported by C++?
(a) Register (b)automatic
(c)mutable (d)dynamic
(e)External
11. The command javac is used to:
(a) Debug a java program
(b) Compile a java program
(c) Interpret a java program
(d) Execute a java program
(e) None of these
12. An administrator attempts a traceroute but receives a "Destination Unreachable" message. Which protocol is responsible for that message?
(a) RARP (b) RUDP
(c) ICMP (d) SNMP
(e) ARP
13. Which type of page replacement algorithms suffers from Belady's anomaly?
(a) Shortest job first (b) Round robin
(c) first-in first-out (d) Optimal Page replacement
(e) Least Recently Used
14. Which of the following are cloud computing deployment model?
(a) Public cloud (b) Private cloud
(c) Hybrid Cloud (d) All of the above
(e) Only (a) and (b)
15. Which SQL statement is used to return only different values?
(a)SELECT DISTINCT (b)SELECT DIFFERENT
(c)SELECT UNIQUE (d)SELECT JOIN
(e)SELECT RETURN
16. Which of the following scheduling algorithms gives minimum average waiting time?
(a) FCFS (b) SJF
(c) Round Robin (d) Priority
(e) None
17. Which of the following is the most commonly used HTTP methods?
(a) PRE and POST (b) GET and SET
(c) ASK and REPLY (d) GET and POST
(e) GET and PRE
18. Which of the following concepts is applicable with respect of 2 NF?
(a) Full functional dependency
(b) Partial dependency
(c) Transitive dependency
(d) None- transitive dependency
(e) None of these

19. Which keyword is used to prevent any changes in the variable within a C++ program?
(a) Immutable (b) Mutable
(c) Const (d) Volatile
(e) Bool
20. Which of the following is a style sheet language used for describing the presentation of a document written in a markup language.
(a) Creative Style Shots (b) Creative Style Sheets
(c) Cascading Style Sheets (d) Cascading Style Shots
(e) Calculative Sheet Style
21. Which of the following is the process of testing make sure that the older programming still works with the new changes?
(a) Unit testing (b) Performance testing
(c) Usability testing (d) Beta testing
(e) Regression testing
22. Which type of malware causes an immense problem to the functioning of a computer system by altering the browser settings?
(a) Adwares (b) Toolbars
(c) Hijackers (d) Cookies
(e) Bootstrap
23. Which of the following computer virus that actively attacks an antivirus program or programs to prevent detection?
(a) Micro virus (b) Stealth virus
(c) Retrovirus (d) Armored virus
(e) Polymorphic
24. Which of the following is not the major function of PKI?
(a) Confidentiality (b) Accessibility
(c) Authentication (d) Integrity
(e) Non-Repudiation
25. What is effect of thrashing in CPU utilization?
(a) Increase CPU utilization (b) Decrease CPU utilization
(c) Keeps constant (d) NO effect on CPU
(e) None of these
26. Which of the following is true about compaction in operating system?
(a) It is a technique for overcoming fatal error
(b) It is a page management technique
(c) It helps to solve the problem of fragmentation
(d) Allocate the address of physical memory
(e) None of these
27. Banker's Algorithm is used for:
(a) Deadlock Detection (b) Deadlock Prevention
(c) Deadlock Avoidance (d) Page Scheduling
(e) CPU Scheduling
28. Parsing is related to which phase of compiler design?
(a) Lexical Analysis
(b) Intermediate code generation
(c) semantic Analysis
(d) Syntax Analysis
(e) code optimization
29. Which of the following is the fifth phase of compiler design?
(a) Code optimization (b) Semantic analysis
(c) Lexical analysis (d) Code generation
(e) Syntax analysis
30. In queue operation elements are deleted which end?
(a) Front (b) Base
(c) Push (d) Rear
(e) Top
31. Which of the following is the complexity of selection sort?
(a) O(n) (b) O(n2)
(c) O(1) (d) O(n log n)
(e) O(n+1)
32. Which of the following is also known as fetch–decode–execute cycle in microprocessor?
(a) Machine cycle (b) Memory cycle
(c) Instruction cycle (d) Clock cycle
(e) Fetch cycle
33. which of the following is a trap set to detect, deflect or in some manner counteract attempts at unauthorized use of information systems?
(a) Firewall (b) Honey pot
(c) Cryptography (d) Bootnet
(e) None of these
34. Which protocol is used to look up an IP address from a known Ethernet address?
(a) IP (b) ARP
(c) RARP (d) TCP
(e) ICMP
35. Which protocol is used for Ping?
(a) TCP (b) ARP
(c) ICMP (d) BootIP
(e) HTTP
36. Which type of entity cannot contain primary key.
(a) Strong Entity (b) Weak Entity
(c) Both (a) and (b) (d) Composite Entity
(e) Recursive Entity
37. Which of the following is not the state of a process in operating system?
(a) New (b) Ready
(c) Waiting (d) Committed
(e) Terminated
38. Each node in a linked list must contain at least:
(a) Three fields (b) Five fields
(c) Four fields (d) One fields
(e) Two fields

39. Which of the following is the type of software that has self-replicating software that causes damage to files and system?
(a) Viruses (b) Trojan horses
(c) Bots (d) Worms
(e) Backdoors
40. You can add a row using SQL in a database with which of the following?
(a) ADD (b) CREATE
(c) INSERT (d) MAKE
(e) UPDATE
41. Default subnet mask of Class C IP address is:
(a) 255. 255. 0.0 (b) 255. 255. 255. 0
(c) 255. 255. 255. 255 (d) 255. 0.255. 0
(e) None of these
42. What is WiMAX?
(a) Wireless maximum communication
(b) Worldwide interoperability for microwave access
(c) Worldwide international standard for microwave access
(d) Wireless internet maximum access
(e) Wireless interoperability maximum communication
43. In which type of cryptography same key is used as encryption and decryption?
(a) Symmetric cryptography (b) Asymmetric cryptography
(c) Key cryptography (d) Encryption cryptography
(e) None of these
44. What is Kerberos in networking?
(a) Authentication Protocol (b) Routing Protocol
(c) Flow Control Protocol (d) Error Control Protocol
(e) Both (c) and (d)
45. Deadlocks are possible only when one of the transactions wants to obtain a(n) ____ lock on a data item.
(a) binary (b) exclusive
(c) shared (d) complete
(e) None of these
46. How can you change "Singh" into "Gupta" in the "LastName" column in the Persons table?
(a) MODIFY Persons SET LastName='Singh' INTO LastName='Gupta'
(b) MODIFY Persons SET LastName='Gupta' WHERE LastName='Singh'
(c) UPDATE Persons SET LastName='Singh' INTO LastName='Gupta'
(d) UPDATE Persons SET LastName='Gupta' WHERE LastName='Singh'
(e) CREATE Person SET LastName='Gupta' WHERE LastName='Singh'
47. Which of following is world fastest supercomputer?
(a) Sequoia (b) Titan
(c) Tianhe-2 (d) Sunway TaihuLight
(e) Piz Daint
48. Which of the following is not an antivirus?
(a) Bitdefender (b) Kaspersky
(c) Cobian (d) F-secure
(e) Avira
49. Which of the following is the combination of public and private cloud?
(a) Business cloud (b) Space cloud
(c) Rackspace cloud (d) Hybrid cloud
(e) Commodity cloud
50. In which debugging technique, when an error has occurred, one needs to start tracing the program backward one step at a time evaluating the values of all variables until the cause of error is found?
(a) Cause Elimination (b) Program Slicing
(c) Backtracking (d) Brute Force Method
(e) None of these
51. Which of the following maintenance strategies involves performing maintenance before problems occur to avoid breakdowns and disruptions?
(a) Perfective Maintenance (b) Corrective Maintenance
(c) Adaptive Maintenance (d) Preventive Maintenance
(e) Breakdown Maintenance
52. Data bus used in microprocessor is always:
(a) Unidirectional (b) Bi-directional
(c) Many directional (d) Both (a) and (b)
(e) Micro directional
53. Which of the following is a systematic evolution of the network security in an organization?
(a) Authorization (b) Evolution
(c) Security audit (d) Data management
(e) Authentication
54. In C++ "||" operator used for:
(a) Conjunction (b) Disjunction
(c) Inclusive (d) Exclusive
(e) Equality
55. Which of the following is a best data structure to check an arithmetic expression has balanced parenthesis?
(a) Queue (b) Stack
(c) Tree (d) List
(e) Graph
56. What is the default port number of HTTPS?
(a) 80 (b) 443
(c) 25 (d) 53
(e) 64
57. What is the use of mkdir command in unix?
(a) Rename files (b) Send file to printer
(c) Remove directory
(d) creates a single directories or multiple directories
(e) Copy file

58. Which of the following are the two mode of IP security?
(a) Certificate and tunnel
(b) Certificate and transport
(c) Transport and tunnel
(d) Security and phases
(e) Encryption and security
59. Which of the following is a unit of storage that can store one or more records in a hash file organization?
(a) Buckets (b) Disk pages
(c) Blocks (d) Nodes
(e) Table
60. Which of the following is true about SQL joins?
(a) The join condition is not separated from other search conditions in a query.
(b) The ON clause makes code difficult to understand.
(c) The join condition for natural join is basically an equijoin of all columns with same name.
(d) Both (a) and (b)
(e) None of the above.

## SOLUTIONS

1. (d); fork() creates a new process by duplicating the calling process, The new process, referred to as child, is an exact duplicate of the calling process, referred to as parent
2. (c); Aggregation allows creating complex entities from existing entities and relationship. Aggregation is the way used to abstracting a larger amount object.
3. (d); Breadth First Search (BFS) algorithm traverses a graph in a breadthward motion and uses a queue to remember to get the next vertex to start a search, when a dead end occurs in any iteration.
4. (d); Multiprocessing refers to a computer system's ability to support more than one process (program) at the same time. Multiprocessing operating systems enable several programs to run concurrently. UNIX is one of the most widely used multiprocessing systems.
5. (a)
6. (c); $(1x8^2) + (2x8^1) + (6 \times 8^0)$
$64 + 16 + 6 = (86)_{10}$
7. (a); adware is a program that installs an additional component that feeds advertising, often by delivering pop-up ads or by installing a toolbar in your browser.
8. (c); Backup Utility software is an application used to create a duplicate copy of data to safeguard it and enable recovery in the event it is lost, corrupted or infected by malware.
9. (d); The flip-flop is the basic unit of digital memory. A flip-flop can remember one bit of data. Sets of flip-flops are called registers, and can hold bytes of data. Sets of registers are called memories, and can hold many thousands of bits, or more.
10. (d); A storage class in the context of C++ variable declarations is a type specifier that governs the lifetime, linkage, and memory location of objects. There are following storage classes, which can be used in a C++ Program- automatic, register, static, external, mutable.
11. (b); The javac command in Java compiles a program from a command prompt. It reads a Java source program from a text file and creates a compiled Java class file.
12. (c); The ICMP destination unreachable message is generated by a router (which is reachable) to inform the source host that the destination unicast address is unreachable.
13. (c); Belady's anomaly is the phenomenon in which increasing the number of page frames results in an increase in the number of page faults for certain memory access patterns. This phenomenon is commonly experienced when using the first-in first-out (FIFO) page replacement algorithm.
14. (d)
15. (a); DISTINCT keyword can be used to return only distinct (different) values.
16. Ans-(b); Shortest Job First(SJF) is the best approach to minimize waiting time.
17. (d); Two most commonly used HTTP methods-
GET - Requests data from a specified resource
POST - Submits data to be processed to a specified resource
18. (a); A database is in second normal form if it satisfies the following conditions: It is in first normal form and all non-key attributes are fully functional dependent on the primary key.
19. (c); Const keyword is used to declare immutable data or functions that do not change data
20. (c); Cascading Style Sheets (CSS) is a style sheet language used for describing the presentation of a document written in a mark-up language.
21. (e); Regression testing is the testing after modification of a system, component, or a group of related units to ensure that the modification is working correctly and is not damaging or imposing other modules to produce unexpected results. It falls under the class of black box testing.
22. (c); Hijackers are the type of malware that take control of the behavior of your web browser like the home page, default search pages, toolbar etc. Hijackers redirect your browser to another URL if you mistype the URL of the website you want to visit. Hijackers can also prevent you from opening a particular web site. Hijackers are annoyance to the users who use the browser often.

23. (c); Retrovirus is type virus which tries to attack and disable the anti-virus application running on the computer. A retrovirus can be considered anti-antivirus. Some Retroviruses attack the anti-virus application and stop it from running or some other destroys the virus definition database.
24. (b); The Public Key Infrastructure (PKI) is a set of hardware, software, people, policies, and procedures needed to create, manage, store, distribute, and revoke Digital Certificates. A Public Key Infrastructure (PKI) enables users of a basically unsecure public network such as the Internet to securely and privately exchange data using a public and a private cryptographic key pair that is obtained and shared through a trusted authority. There are 4 major function of Public Key Infrastructure (PKI): Confidentiality, Authentication, Integrity, Non-Repudiation.
25. (b); Thrashing or disk thrashing describes when a hard drive is being overworked by moving information between the system memory and virtual memory excessively. Thrashing occurs when the system does not have enough memory, the system swap file is not properly configured, too much is running at the same time, or has low system resources. When thrashing occurs, you will notice the computer hard drive always working and a decrease in system performance.
26. (c); Compaction refers to combining all the empty spaces together and processes. Compaction helps to solve the problem of fragmentation, but it requires too much of CPU time.
27. (c); The banker's algorithm is a resource allocation and deadlock avoidance algorithm that tests for safety by simulating the allocation for predetermined maximum possible amounts of all resources. Banker's algorithm is a deadlock avoidance algorithm. It is named so because this algorithm is used in banking systems to determine whether a loan can be granted or not.
28. (d); Syntax analysis is the second phase of compiler which is also called as parsing. Parser converts the tokens produced by lexical analyzer into a tree like representation called parse tree. A parse tree describes the syntactic structure of the input.
29. (a); Code optimization is the fifth phase of compiler design. It transforms the code so that it consumes fewer resources and produces more speed. The meaning of the code being transformed is not altered. Optimization can be categorized into two types: machine dependent and machine independent.
30. (a); Queue is also an abstract data type or a linear data structure, in which the first element is inserted from one end called REAR(also called tail), and the deletion of existing element takes place from the other end called as FRONT(also called head). This makes queue as FIFO(First in First Out) data structure, which means that element inserted first will also be removed first.
31. (b); Complexity of an algorithm is a measure of the amount of time and/or space required by an algorithm for an input of a given size (n).
32. (c); An instruction cycle (sometimes called fetch-decode-execute cycle) is the basic operation cycle of a computer. It is the process by which a computer retrieves a program instruction from its memory, determines what actions the instruction requires, and carries out those actions.
33. (b); A honeypot is a security system designed to detect and counteract unauthorized access or use of a computer system. The name "honeypot" is used in reference to the way the system traps unauthorized users, such as hackers or spammers so they can be identified and prevented from causing further problems.
34. (c); Reverse Address Resolution Protocol, a TCP/IP protocol that permits a physical address, such as an Ethernet address, to be translated into an IP address.
35. (c); ICMP is the protocol at the Network layer that is used Ping command. Ping is a networking utility program or a tool to test if a particular host is reachable. Ping basically used to check connectivity.
36. Ans-(b); a weak entity is an entity that cannot be uniquely identified by its attributes alone. A weak entity has no primary key; it must use a foreign key in conjunction with its attributes to create a primary key.
37. (d); The state of a process is defined in part by the current activity of that process. There are 5 states of a process- New, Running, Waiting, Ready and Terminated.
38. (e); A linked list is a linear data structure where each element is a separate object. Each element (we will call it a node) of a list is comprising of two items - the data and a reference to the next node. The last node has a reference to null. The entry point into a linked list is called the head of the list.
39. (d); A worm is a type of virus that spreads through your computer by creating duplicates of itself on other drives, systems and networks.
40. (c); An SQL INSERT statement adds one or more records to any single table in a relational database.
41. (b); A Subnet mask is a 32-bit number that masks an IP address, and divides the IP address into network address and host address. Default subnet mask of Class C IP address is 255.255.255.0

42. (b); WiMAX (Worldwide Interoperability for Microwave Access) is a wireless industry coalition dedicated to the advancement of IEEE 802. 16 standards for broadband wireless access (BWA) networks.
43. (a); In symmetric key cryptography, same key is shared, i.e. the same key is used in both encryption and decryption.
44. (a); Kerberos is a network authentication protocol that is designed to provide strong authentication for client/server applications using secret-key cryptography.
45. (b); When a statement modifies data, its transaction holds an exclusive lock on data that prevents other transactions from accessing the data.
46. (d); The UPDATE statement is used to modify the existing records in a table.
Syntax:
UPDATE table_name
SET column1 = value1, column2 = value2, ...
WHERE condition;
47. (d); A supercomputer is a computer that performs at or near the currently highest operational rate for computers. The Sunway TaihuLight is a Chinese supercomputer which, as of November 2017, is ranked number one in the TOP500 list as the fastest supercomputer in the world.
48. (c); Cobian Backup is a freeware for Windows that enables the users to back up files and directories from the existing location and can be done on the same computer or within the same network.
49. (d); Hybrid clouds combine public and private clouds, bound together by technology that allows data and applications to be shared between them. By allowing data and applications to move between private and public clouds, hybrid cloud gives businesses greater flexibility and more deployment options.
50. (c); In Backtracking method, beginning from the statement at which an error symptom has been observed, the source code is traced backwards until the error is discovered. Unfortunately, as the number of source lines to be traced back increases, the number of potential backward paths increases and may become unmanageably large thus limiting the use of this approach.
51. (d); Preventive maintenance refers to regular, routine maintenance to help keep equipment up and running, preventing any unplanned downtime and expensive costs from unanticipated equipment failure. It requires careful planning and scheduling of maintenance on equipment before there is an actual problem as well as keeping accurate records of past inspections and servicing reports.
52. (b); A data can be moved from microprocessor to memory and also from memory to microprocessor. So, Data bus of any microprocessor is always Bi-directional. Data bus is 8 Bits long.
53. (c); Network security audit helps to determine the effectiveness of network security to resolving underlying network security issues. Network security audits are critical to understanding how well your organization is protected against security threats, whether they are internal or external. A network security audit is part of an overall information systems audit framework and includes application software audit, operation system audit, and business audit.
54. (b); 'Or' operator. Performs a logical disjunction on two expressions. (if either or both expressions evaluate to True, result is True)
55. (b); Stack, because stack is used in evolution of expression.
56. (b)
57. (d)
58. (c); IPSec (IP Security) supports transport and tunnel mode, both of which can use either ESP or AH packets. Transport mode secures packets between two endpoints, typically in a client-to-gateway scenario, and leaves the original IP header unchanged. Tunnel mode encapsulates the IP header and payload into a new IPSec packet for transfer between two endpoints, typically two IPSec gateway devices.
59. (a); A unit of storage that can store one or more records in a hash file organization is denoted as buckets.
60. (c);

# Professional Knowledge Practice Set:10

1. Which of the following is a special kind of a store procedure that automatically executes when an event occurs in the database server like insertion, deletion or updation of data?
(a) Procedures (b) Triggers
(c) Functions (d) Attribute
(e) Join
2. Which of the following is most suitable programming language for scientific purpose?
(a) COBOL (b) BASIC
(c) FORTRAN (d) C
(e) C++
3. To promote encapsulation, Java provides an access modifier of type:
(a) Public (b) Private
(c) Protected (d) Both (a) and (b)
(e) None of the above
4. Which of the following is the use of javac in java?
(a)Convert a java program into binary code
(b)Convert a java program into bytecode
(c)Convert a java program into machine language
(d)Convert a binary code into java program
(e)Convert machine language into high level language
5. By which statement you can force immediate termination of a loop, by passing the conditional expression and any remaining code in the body of the loop.
(a)Break (b)Continue
(c)Terminate (d)Loop Close
(e) Return
6. What is the file extension of compiled java program?
(a) .class (b) .java
(c) .css (d) .html
(e) .c
7. Which of the following tag is used to link the URL?
(a) <a> (b) <style>
(c) <link> (d) <hyperlink>
(e) <b>
8. Which of the following software programs gets activated after a pre-defined event occurs?
(a) Malware (b) Spyware
(c) Logic Bomb (d) Adware
(e) Worm
9. "SaaS" stand for in cloud computing:
(a) System as a service (b) Software as a service
(c) Software as a system (d) Scheme as a service
(e) System as a system
10. Which of the following is the Correct HTML tag for the largest heading?
(a) <head> (b) <h6>
(c) <heading> (d) <h1>
(e) <h2>
11. Queue is which type of data structure?
(a) Linear data structure
(b) Non-linear data structure
(c) Random access data structure
(d) Self-sorted data structure
(e) Tree Data structure
12. A graph with all vertices having equal degree is known as a:
(a) Multi Graph (b) Regular Graph
(c) Simple Graph (d) Complete Graph
(e) None of these
13. What is Code Red?
(a) Word Processing Software
(b) Antivirus
(c) Virus
(d) Photo Editing Software
(e) None of these
14. IP address 128.167.152.2 is belongs to which class?
(a) A (b) B
(c) C (d) D
(e) None
15. The parameter which gives the probability of the transport layer itself spontaneously terminating a connection due to internal problem?
(a) Protection (b) Resilience
(c) Option negotiation (d) Transfer failure
(e) None
16. Which of the following is a single communication channel that is shared by all the users on the network?
(a) Point-to-Point (b)Broadcast Network
(c)Protocol (d)PAN
(e)None of these
17. what is the use of 'AS' clause in sql?
(a) Selection operation (b) Rename operation
(c) Join operation (d) Projection operation
(e) None of these
18. In which method, records are physically stored in a particular order according to a Key field in each record?
(a) Hash (b) Direct
(c) Sequential (d) All of the above
(e) None of the above
19. In asymmetric key cryptography, the private key is kept by:
(a) sender (b) receiver
(c) sender and receiver (d) all the connected devices to the network
(e) None of these
20. Which of the following part of the Unix operating system that interacts with the hardware?
(a) GRUB (b) The kernal
(c) The shell (d) File system
(e) Unix epoch

21. Which of the following statement is true?
(a) Char has a variable length
(b) Varchar is faster than char
(c) Varchar use hybrid memory
(d) Char takes one byte per character
(e) All are true

22. In which type of switching technique whole message are transmitted at a time from source to destination?
(a) Packet switching (b) Circuit switching
(c) Message switching (d) Data switching
(e) Key switching

23. How to convert weak entity set to strong entity set?
(a) Using aggregation
(b) adding appropriate attributes
(c) Using generalization
(d) Normalization
(e) None of these

24. Which supercomputer developed by India?
(a) Piz Daint (b) Titan
(c) SahasraT (d) Sunway TaihuLight
(e) Tianhe-2

25. Databases that support OLTP are most often referred to as ______.
(a) operational databases (b) relational databases
(c) data warehouses (d) data dictionaries
(e) None of these

26. Alpha testing is done by:
(a) Customer (b) Tester
(c) Developer (d) All of the above
(e) None of these

27. In Java, each thread has its own ________, in which it runs?
(a) main() method (b) JVM
(c) Call stack (d) Memory
(e) Class

28. Which of the following protocol is used for remote terminal connection service?
(a) TELNET (b) FTP
(c) RARP (d) UDP
(e) SMTP

29. Circuit Switching takes place at which layer?
(a) Transport layer (b) Application layer
(c) Presentation layer (d) Physical layer
(e) Data link layer

30. Which of the following protocol uses both TCP and UDP?
(a) FTP (b) SMTP
(c) TELNET (d) DNS
(e) SMTP

31. Page fault occur in which technique in operating system?
(a) Paging (b) Segmentation
(c) Dynamic partition (d) Virtual Memory
(e) Fragmentation

32. What is Process Control Block?
(a) It is a Process type variable
(b) It is a data structure
(c) A secondary storage sections
(d) A block in memory
(e) A page management technique

33. Which of the following situation occurs when one or more threads in your program are blocked from gaining access to a resource and, as a result, cannot make progress?
(a) Aging (b) Fragmentation
(c) Inversion (d) Starvation
(e) Paging

34. Cortana is a feature of which operating system?
(a) Windows 8 (b) Windows 10
(c) windows 8.1 (d) MacOS 10.13.2
(e) MacOS 10.12

35. Which data structure is used in BFS?
(a) List (b) Array
(c) Queue (d) Stack
(e) Both (c) and (d)

36. Which of the following is the time complexity of insertion sort?
(a) O(n^2) (b) O(1)
(c) O(n) (d) O(nlogn)
(e) O(logn)

37. All java classes are derived from:
(a)java.lang.Class (b)java.util.Name
(c)java.lang.Object (d)java.awt.Window
(e)java.dfd.window

38. What is the act of willful changing data, using fraudulent input or removal of controls called______.
(a) Data diddling (b) Data contaminating
(c) Data capturing (d) Data trashing
(e) Data spoofing

39. Which layer provides for the syntax of the data?
(a) Application Layer (b) Session Layer
(c) Transport Layer (d) Presentation Layer
(e) Physical Layer

40. What are the 3 principles of information security?
(a) Authentication, Availability, Accessibility
(b) Confidentiality, Integrity, Availability,
(c) Integrity, Authentication, Confidentiality
(d) Availability, Security, Encryption
(e) Login, Availability, Encryption

41. Which of the following is correct order of Clauses for an SQL Query?
(a) WHERE - ORDER BY - HAVING - GROUP BY
(b) WHERE - HAVING - GROUP BY - ORDER BY
(c) WHERE - GROUP BY - HAVING - ORDER BY
(d) ORDER BY - HAVING - GROUP BY - WHERE
(e) can be arrange in any order

42. Which of the following statement is correct about HAVING clause?
(a) HAVING clause is used for rows rather than columns.
(b) HAVING clause is used for columns rather than groups.
(c) HAVING clause acts exactly like a WHERE clause.
(d) HAVING clause is used to mention condition in Group based SQL functions
(e) All are true about HAVING Clause
43. Default subnet mask of Class A IP address is
(a) 255.255.0.0 (b) 255. 255.255.0
(c) 255.255.255.255 (d) 255. 0.0.0
(e) None of these
44. Which of the following protocol is used for securing connections between network application clients and servers over an insecure network?
(a) SSL (b) EGP
(c) UDP (d) HTTP
(e) IGP
45. Which of the following is an example for Exterior Gateway Protocol (EGP)?
(a) Open Shortest Path First (OSPF) Protocol
(b) Routing Information Protocol (RIP)
(c) Border Gateway Protocol
(d) Both (A) and (B)
(e) TCP
46. which types of data is transfer in handshaking method.
(a)Synchronous (b)Asynchronous
(c)Sequential (d)Directional
(e)Bidirectional
48. In a(n) ____ backup of the database, only the last modifications to the database are copied.
(a) full (b) incomplete
(c) differential (d) transaction log
(e) Security
49. which type of data structure is used to represent internal storage of data in RDBMS efficiently?
(a) Stack (b) Queue
(c) B+ Tree (d) Linked list
(e) Dequeue
50. Which of the following refers to the ability of several servers to connect to a single database?
(a) Row (b) Index
(c) Table (d) Cluster
(e) Object
51. Which routing protocol uses the Dijkstra Shortest Path First Algorithm?
(a) OSPF (b) EGP
(c) IGP (d) BGP
(e) RIP
52. Derived attribute in E-R Diagram represented by:
(a) Rectangle (b) Diamond
(c) Ellipse (d) Double Ellipse
(e) Dashed Ellipse
53. Which of the following is a part of time Sharing system that reduces the degree of multiprogramming?
(a) Long term Scheduler (b) Short term Scheduler
(c) Thrashing (d) Medium term Scheduler
(e) Deadlock
54. Which term in network security refers to the ability to ensure that a party to a contract or a communication cannot deny the authenticity of their signature on a document or the sending of a message that they originated?
(a) Data Integrity (b) Authentication
(c) Confidentiality (d) Authorization
(e) Non-Repudiation
55. What is the size of block in DES algorithm?
(a) 64 (b) 56
(c) 6 (d) 8
(e) 32
56. Which of the following is an online cloud storage service from Microsoft?
(a) MicroDrive (b) OneDrive
(c) CloudDrive (d) WindowsDrive
(e) None of these
57. What is the purpose of the CRC in network communications?
(a) Request retransmission of lost packets (b) Reassemble packets into their original order
(c) Detect lost data packets (d) Slow down or speed up data transmission
(e) Detect data errors
58. What is BGP in networking?
(a) Balanced Gateway Protocol (b) Border Gateway Protocol
(c) Broadband Gateway Protocol (d) Balanced Goal Protocol
(e) Balanced Gate Protocol
59. In C++ "||" operator used for:
(a) Conjunction (b) Disjunction
(c) Inclusive (d) Exclusive
(e) Equality
60. In which of the following algorithm, pages are replaced which are not used for the longest duration of time in the future?
(a) First In First Out
(b) Optimal Page replacement
(c) Least Recently Used
(d) Belady's anomaly
(e) Both (b) and (c)

## SOLUTIONS

1. (b); Triggers can be defined as the database objects which perform some action for automatic execution whenever users try to do execute data modification commands (INSERT, DELETE and UPDATE) on the specified tables.
2. (c); FORTRAN (Formula Translation) is a general-purpose, imperative programming language that is especially suited to numeric computation and scientific computing.
3. (b); Encapsulate provides contract to other object to what to hide and what to expose which can be accessed other objects. In JAVA, we use private access modifier to hide method and variable to restrict from outer world.
4. (b); javac.exe is the Java compiler program. It compiles Java source files (.java) into bytecode class files (.class).
5. (a); break statement is encountered inside a loop, the loop is immediately terminated and the program control resumes at the next statement following the loop.
6. (a); (.class) is the Compiled java source code file.
7. (a)
8. (c); Logic bombs are programs or code snippets that are executed when a pre-defined event occurs. They are kept secretly into the code of existing software, where they lie inactive till a pre-defined event occurs.
9. (b); Software-as-a-service (SaaS) is a method for delivering software applications over the Internet, on demand and typically on a subscription basis. With SaaS, cloud providers host and manage the software application and underlying infrastructure and handle any maintenance, like software upgrades and security patching. Users connect to the application over the Internet, usually with a web browser on their phone, tablet or PC.
10. (d)
11. (a); Queue is a linear data structure used to represent a linear list. It allows insertion of an element to be done at one end and deletion of an element to be performed at the other end.
12. (b); A graph in which all vertices are of equal degree, is called a regular graph. If the degree of each vertex is r, then the graph is called a regular graph of degree r.
13. (c); Code Red is a worm that caused possible billions of dollars of damage in the summer of 2001. It contains the text string "Hacked by Chinese!", which is displayed on web pages that the worm defaces. It is also one of the few worms able to run entirely in memory, leaving no files on the hard drive or any other permanent storage.
14. (b); Range of Class B address is 128.0.0.0 to 191.255.255.255
15. (b); Resilience is the ability to provide and maintain an acceptable level of service in the face of faults and challenges to normal operation.
16. (b); Broadcast network are characterized by the presence of a single communication channel that is shared by all users. Messages, in some cases called packets, are sent by each user and received by all the others.
17. (b); "AS" clause is used for rename operation in SQL.
18. (a); A method in which records are physically stored in a specified order according to a key field in each record is hash. In hash method, a hash function is performed on the key value to determine the unique physical address of the record to store or retrieve.
19. (b); Asymmetric cryptography, also known as public key cryptography, uses public and private keys to encrypt and decrypt data. The keys are simply large numbers that have been paired together but are not identical (asymmetric). One key in the pair can be shared with everyone; it is called the public key.
20. (b); In Unix operating system the kernel is the bridge between the application packages and the hardware it is running on.
21. (d); Char is fixed length memory storage and takes one byte per character. Char is faster than verchar.
22. (c); Message switching does not establish a dedicated path between the two communicating devices i.e. no direct link is established between sender and receiver. Each message is treated as an independent unit. In message switching, each complete message is then transmitted from device to device through the internetwork i.e. message is transmitted from the source node to intermediate node.
23. (b); We can convert any weak entity set to a strong entity set by simply adding appropriate attributes.
24. (c); Supercomputers are using their high computational speeds and efficiency to help many sectors in our country, such as scientific research, weather forecasting, missile simulation, understanding the creation of the universe, creating life-saving pharmaceutical drugs and much more. SahasraT supercomputer is located at Supercomputer Education and Research Centre (SERC) facility at Indian Institute of Science in Bangalore. SERC is India's state-of-the-art research facility for high-performance computing in the field of science and engineering.

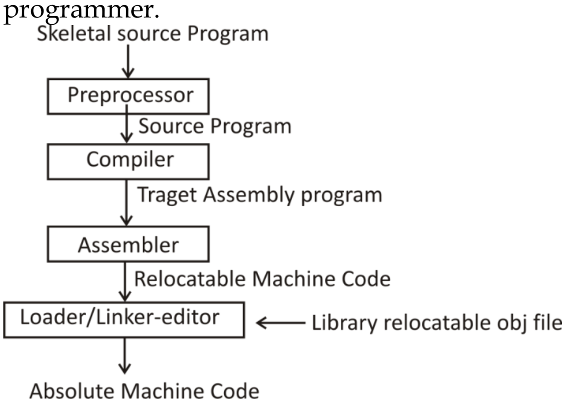
25. (a); Databases that support OLTP are most often referred to as operational databases.
26. (c); Alpha testing takes place at the developer's site by the internal teams, before release to external customers. Alpha testing is one of the most common software testing strategy used in software development. It is specially used by product development organizations.
27. (c); Thread of execution means an individual 'lightweight' process that has its own call stack. In java each thread has its own call stack.
28. (a); Telnet is a user command and an underlying TCP/IP protocol for accessing remote computers. Through Telnet, an administrator or another user can access someone else's computer remotely.
29. (d); Circuit switching is a switching method in which a dedicated communication path in physical form between two stations within a network is established, maintained and terminated for each communication session. It has basically three phases as circuit establishment, data transfer and circuit disconnect.
30. (d); DNS uses both TCP and UDP. DNS uses TCP for Zone transfer and UDP for name queries either regular (primary) or reverse. UDP can be used to exchange small information whereas TCP must be used to exchange information larger than 512 bytes.
31. (d); A page fault is a type of interrupt, raised by the hardware when a running program accesses a memory page that is mapped into the virtual address space (virtual memory), but not loaded in physical memory.
32. (b); Process Control block is used for storing the collection of information about the Processes and this is also called as the Data Structure which Stores the information about the process. The information of the Process is used by the CPU at the Run time.
33. (d); starvation occurs when a particular process needs to wait indefinitely, as it never gets a chance to proceed further. In this situation, the process or transaction either waits indefinitely or gets in restarted mode again and again. Deadlock and starvation differs from each other. Deadlock occurs when none of the processes in the set is able to move ahead due to occupancy of the required resources by some other process. On the other hand, starvation occurs when a process waits for an indefinite period of time to get the resource it requires.
34. (b); Cortana is a personalized digital assistant for windows 10. Cortana helps to do something like: Give you remainder based on time, place or people, create and manage list, open any app on your system, find facts, file, places and information etc.
35. (c); Breadth First Search (BFS) algorithm traverses a graph in a breadth ward motion and uses a queue to remember to get the next vertex to start a search, when a dead end occurs in any iteration.
36. (a)
37. (c); java.lang.Object class is the super base class of all Java classes. Every other Java classes descend from Object.
38. (a); Data diddling is a method adopted by computer criminals. Data diddling is the changing of data before or during entry into computer system or altering the raw data just before it processed by a computer and then changing it back after the processing is complete. Using this technique, the criminals can manipulate the output and it is not so easy to identify.
39. (d); The primary goal of presentation layer is to take care of the syntax and semantics of the information exchanged between two communicating systems. Presentation layer takes care that the data is sent in such a way that the receiver will understand the information (data) and will be able to use the data. Presentation layer also called syntax layer
40. (b); A principle which is a core requirement of information security for the safe utilization, flow, and storage of information is the CIA triad. CIA stands for confidentiality, integrity, and availability and these are the three main objectives of information security.
42. (c); Syntax of SQL SELECT Statement:
SELECT column_list FROM table-name
[WHERE Clause]
[GROUP BY clause]
[HAVING clause]
[ORDER BY clause];
43. (d); Having clause is used to filter data based on the group functions. This is similar to WHERE condition but is used with group functions. Group functions cannot be used in WHERE Clause but can be used in HAVING clause.
44. (d); In an IP network, every machine on the same physical network sees all the data packets sent out on the network. As the number of computers on a network grows, network traffic will grow many fold, bringing down performance drastically. In such a situation, you would divide your network into different subnetworks and minimize the traffic across the different subnetworks. Interconnectivity between the different subnets would be provided by routers, which will only transmit data meant for another subnet across itself. To divide a given network address into two or more subnets, you use subnet masks. The default subnet masks for class A networks is 255.0.0.0, for class B is 255.255.0.0, and for class C is 255.255.255.0.

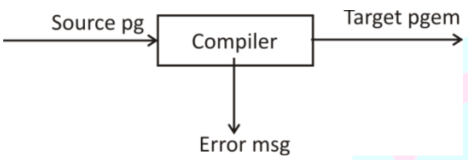
45. (a); Secure Sockets Layer (SSL) is a standard protocol used for the secure transmission of documents over a network. Developed by Netscape, SSL technology creates a secure link between a Web server and browser to ensure private and integral data transmission. SSL uses Transport Control Protocol (TCP) for communication.
46. (c); Exterior Gateway Protocol (EGP) is a protocol for exchanging routing information between two neighbor gateway hosts (each with its own router) in a network of autonomous systems. EGP is commonly used between hosts on the Internet to exchange routing table information.
47. (a); The synchronous transfer can be implemented with a method called handshaking. This method is used to accompany each data item being transfer with a control signal that indicates the presence of data in the bus .the unit receiving the data item respond with another control signal to acknowledge receipt of the data.
48. (c); A differential backup is a type of backup that copies all the data that has changed since the last full backup. For example, if a full backup is done on Sunday, Monday's differential backup backs up all the files changed or added since Sunday's full backup.
49. (c); B+ tree. Because in B+ tree, all the data is stored only in leaf nodes, that makes searching easier. This corresponds to the records that shall be stored in leaf nodes.
 B+ Tree is an advanced method of ISAM(Indexed Sequential Access Method) file organization. It uses the same concept of key-index, but in a tree like structure. B+ tree is similar to binary search tree, but it can have more than two leaf nodes. It stores all the records only at the leaf node. Intermediary nodes will have pointers to the leaf nodes. They do not contain any data/records.
50. (d); A cluster is a group of two or more independent servers operating as a single system. A database server cluster could help you achieve high availability or improve I/O throughput for your database service.
51. (a); Open Shortest Path First (OSPF) routing protocol is a Link State protocol and use Dijkstra Algorithm to find the shortest path.
52. (e); A derived attribute is based on another attribute. For example, an employee's monthly salary is based on the employee's annual salary.
53. (d)
54. (e); Nonrepudiation is the assurance that someone cannot deny something. Typically, nonrepudiation refers to the ability to ensure that a party to a contract or a communication cannot deny the authenticity of their signature on a document or the sending of a message that they originated.
55. (a)
56. (b); OneDrive integrates with Microsoft Office so users can access Word, Excel and PowerPoint documents in OneDrive. The system allows users to simultaneously edit Office documents, edit documents in browsers, and create and share folders.
57. (e); The Cyclical Redundancy Check (CRC) is a mathematical calculation added to each frame. Its purpose is to detect when a frame arrives that has been corrupted. The sending device calculates the CRC and adds it to the frame. The receiving device calculates the CRC when the frame is received. If the CRCs do not match, the frame has been corrupted or altered. The most common method of correcting errors is to request retransmission of the original frame.
58. (b); Border Gateway Protocol (BGP) is a routing protocol used to transfer data and information between different host gateways, the Internet or autonomous systems. BGP is a Path Vector Protocol (PVP), which maintains paths to different hosts, networks and gateway routers and determines the routing decision based on that.
59. (b); 'Or' operator. Performs a logical disjunction on two expressions. (if either or both expressions evaluate to True, result is True)
60. Ans. (b); In optimal page replacement algorithm, replace the page in the memory that will not be used for the longest period of time. This algorithm gives less number of page faults compared to any other algorithms. However, optimal algorithm is not possible to implement in practice.

# ANNEX. 1.
# Language Processing System

## 1. Compiler

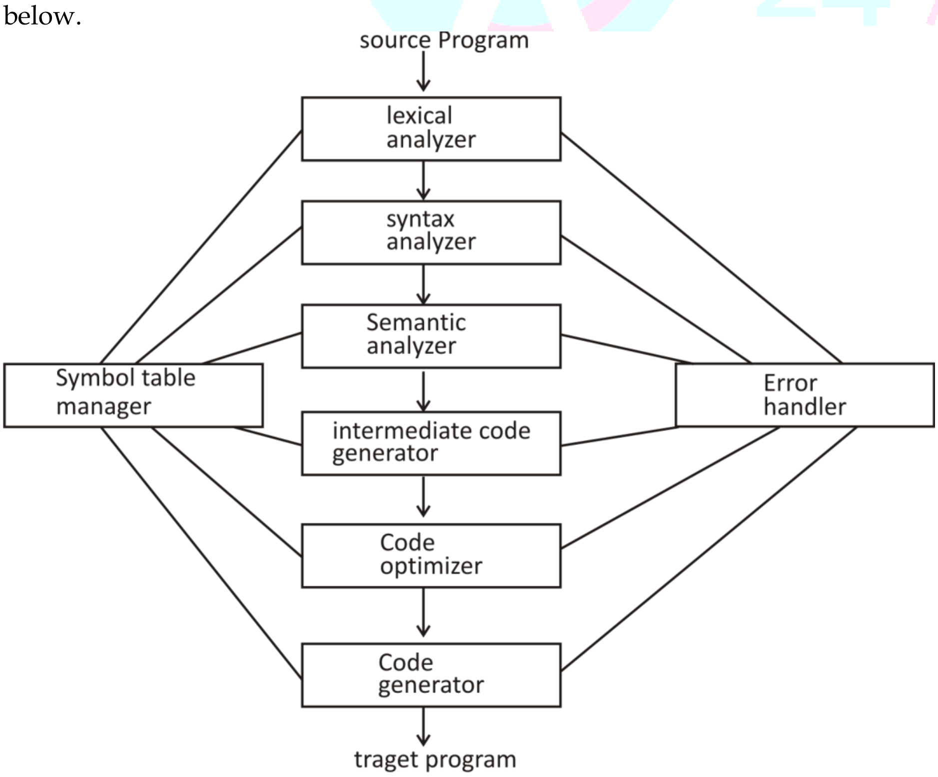
Compiler is a translator program that translates a program written in (HLL) the source program and translate it into an equivalent program in (MLL) the target program. As an important part of a compiler is error showing to the programmer.




Executing a program written n HLL programming language is basically of two parts. the source program must first be compiled translated into an object program. Then the resulted object program is loaded into a memory executed.

**Phases of a compiler:** A compiler operates in phases. A phase is a logically interrelated operation that takes source program in one representation and produces output in another representation. The phases of a compiler are shown in below.




Phases of a compiler

**Lexical Analysis:** LA or Scanners reads the source program one character at a time, carving the source program into a sequence of atomic units called tokens.

**Syntax Analysis:** The second stage of translation is called Syntax analysis or parsing. In this phase expressions, statements, declarations etc. are identified by using the results of lexical analysis. Syntax analysis is aided by using techniques based on formal grammar of the programming language.

**Intermediate Code Generations:** An intermediate representation of the final machine language code is produced. This phase bridges the analysis and synthesis phases of translation.

**Code Optimization:** This is optional phase described to improve the intermediate code so that the output runs faster and takes less space.

**Code Generation:** The last phase of translation is code generation. A number of optimizations to reduce the length of machine language program are carried out during this phase. The output of the code generator is the machine language program of the specified computer.

**Table Management:** This is the portion to keep the names used by the program and records essential information about each. The data structure used to record this information called a 'Symbol Table'.
**Error Handlers:** It is invoked when a flaw error in the source program is detected.

## 2. Interpreter

An interpreter is a program that appears to execute a source program as if it were machine language. Languages such as BASIC, SNOBOL, LISP can be translated using interpreters. JAVA also uses interpreter. The process of interpretation can be carried out in following phases.
1. Lexical analysis
2. Syntax analysis
3. Semantic analysis
4. Direct Execution

**Advantages:**
- Modification of user program can be easily made and implemented as execution proceeds.
- Type of object that denotes a various may change dynamically.
- Debugging a program and finding errors is simplified task for a program used for interpretation.
- The interpreter for the language makes it machine independent.

**Disadvantages:**
- The execution of the program is slower.
- Memory consumption is more.

## 3. Loader and Linker

Once the assembler procedures an object program, that program must be placed into memory and executed. The assembler could place the object program directly in memory and transfer control to it, thereby causing the machine language program to be execute. This would waste core by leaving the assembler in memory while the user's program was being executed. Also, the programmer would have to retranslate his program with each execution, thus wasting translation time. To overcome this problem of wasted translation time and memory. System programmers developed another component called loader.
A **loader** is a program that places programs into memory and prepares them for execution. It would be more efficient if subroutines could be translated into object form the loader could relocate directly behind the user's program.
**Linker** is a computer program that links and merges various object files together in order to make an executable file. In high level languages, some built in header files or libraries are stored. These libraries are predefined, and these contain basic functions which are essential for executing the program. These functions are linked to the libraries by a program called Linker. If linker does not find a library of a function, then it informs to compiler and then compiler generates an error. The compiler automatically invokes the linker as the last step in compiling a program.

## ANNEX. 2.
# Cloud Computing

*Cloud computing* refers to both the applications delivered as services over the Internet, and the hardware and system software in the datacenters that provide those services. The cloud makes it possible for you to access your information from anywhere at any time. While a traditional computer setup requires you to be in the same location as your data storage device, the cloud takes away that step. The cloud removes the need for you to be in the same physical location as the hardware that stores your data. Your cloud provider can both own and house the hardware and software necessary to run your home or business applications.

**Types of clouds**

There are different types of clouds that you can subscribe to depending on your needs. As a home user or small business owner, you will most likely use public cloud services.

**1. Public Cloud -** A public cloud can be accessed by any subscriber with an internet connection and access to the cloud space. The cloud infrastructure is made available to the general public or a large industry group, and is owned by an organization selling cloud services. Examples: Amazon Elastic-Compute-Cloud, IBM's BlueCloud, Sun Cloud, Google AppEngine.

**2. Private Cloud -** A private cloud is established for a specific group or organization and limits access to just that group.

- On-premise Private Cloud: On-premise private clouds, also known as internal clouds are hosted within one's own data center. This model provides a more standardized process and protection, but is limited in aspects of size and scalability.
- Externally hosted Private Cloud: This type of private cloud is hosted externally with a cloud provider, where the provider facilitates an exclusive cloud environment with full guarantee of privacy. This is best suited for enterprises that don't prefer a public cloud due to sharing of physical resources.

**3. Community Cloud -** A community cloud is shared among two or more organizations that have similar cloud requirements. With the costs spread over fewer users than a public cloud (but more than a single tenant) this option is more expensive but may offer a higher level of privacy, security and/or policy compliance. Examples of community cloud include Google's "Gov Cloud".

**4. Hybrid Cloud -** A hybrid cloud is essentially a combination of at least two clouds, where the clouds included are a mixture of public, private, or community. The term "Hybrid Cloud" has been used to mean either two separate clouds joined together (public, private, internal or external), or a combination of virtualized cloud server instances used together with real physical hardware.

There are many types of cloud providers that a user can subscribe to: Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS). These three types differ in the amount of control that a user have over his information, and conversely, how much the user can expect from the provider to do for him. Briefly, here is what you can expect from each type.

**1. Software as a Service -** A SaaS provider gives subscribers access to both resources and applications. SaaS makes it unnecessary for you to have a physical copy of software to install on your devices. SaaS also makes it easier to have the same software on all of your devices at once by accessing it on the cloud. In a SaaS agreement, you have the least control over the cloud. Examples are Google's Gmail and Apps, instant messaging from AOL, Yahoo and Google.

**2. Platform as a Service** - A PaaS system goes a level above the Software as a Service setup. A PaaS provider gives subscribers access to the components that they require to develop and operate applications over the internet. Examples include Microsoft's Azure and Salesforce's Force.com

**3. Infrastructure as a Service** - An IaaS agreement, as the name states, deals primarily with computational infrastructure. In an IaaS agreement, the subscriber completely outsources the storage and resources, such as hardware and software, that they need. The highest profile example is Amazon’s Elastic Compute Cloud (EC2) and Simple Storage Service, but IBM and other traditional IT vendors are also offering services, as is telecom-and-more provider Verizon Business.

**4. Communication-as-a-Service (CaaS):** A CaaS model allows a CaaS provider’s business customers to selectively deploy communications features and services throughout their company on a pay-as-you-go basis for service(s) used. CaaS is designed on a utility-like pricing model that provides users with comprehensive, flexible, and (usually) simple-to understand service plans.

**5. Anything-as-a-Service (XaaS):** or “everything as a service,” is an emerging business model that posits that everything from technology services through key business processes can be delivered as a service utility. XaaS is a collective term said to stand for a number of things including "X as a service," "anything as a service" or "everything as a service." As technology companies adopt XaaS, they realize the capabilities required to succeed with XaaS are vastly different from the attributes of traditional business models.

# ANNEX. 3.
# Oracle, PL/SQL and BIG Data

A database application is a software program that interacts with a database to access and manipulate data. An Oracle database is a collection of data treated as a unit. The purpose of a database is to store and retrieve related information. A database server is the key to solving the problems of information management. The latest version is Oracle Database 12c

## 1. Oracle Grid Architecture

Grid computing is a new IT architecture that produces more resilient and lower cost enterprise information systems. With grid computing, groups of independent, modular hardware and software components can be connected and rejoined on demand to meet the changing needs of businesses.

The grid style of computing treats collections of similar IT resources holistically as a single pool, while exploiting the distinct nature of individual resources within the pool. To address simultaneously the problems of monolithic systems and fragmented resources, grid computing achieves a balance between the benefits of holistic resource management and flexible independent resource control. IT resources managed in a grid include:

- **Infrastructure:** the hardware and software that create a data storage and program execution environment
- **Applications:** the program logic and flow that define specific business processes
- **Information:** the meanings inherent in all different types of data used to conduct business

## 2. PL/SQL

PL/SQL stands for Procedural Language/Structured Query Language. PL/SQL is a procedural language designed specifically to embrace SQL statements within its syntax. PL/SQL program units are compiled by the Oracle Database server and stored inside the database. And at run-time, both PL/SQL and SQL run within the same server process, bringing optimal efficiency.

PL/SQL is a **procedural extension** to Oracle SQL. PL/SQL is integrated with Oracle Database, enabling you to use all of the Oracle Database SQL statements, functions, and data types. You can use PL/SQL to control the flow of a SQL program, use variables, and write error-handling procedures. A primary benefit of PL/SQL is the ability to store application logic in the database itself. A PL/SQL procedure or function is a schema object that consists of a set of SQL statements and other PL/SQL constructs, grouped together, stored in the database, and run as a unit to solve a specific problem or to perform a set of related tasks. The principal benefit of server-side programming is that built-in functionality can be deployed anywhere.

**Building Blocks of PL/SQL Programs**

PL/SQL is a block-structured language. A PL/SQL block is defined by the keywords DECLARE, BEGIN, EXCEPTION, and END, which break up the block into three sections:

- **Declarative:** statements that declare variables, constants, and other code elements, which can then be used within that block
- **Executable:** statements that are run when the block is executed
- **Exception handling:** a specially structured section you can use to “catch,” or trap, any exceptions that are raised when the executable section runs

**PL/SQL has three categories of control statements:**

1. *Conditional selection statements:* which run different statements for different data values. The conditional selection statements are IF and CASE.
2. *Loop statements:* which run the same statements with a series of different data values. The loop statements are the basic LOOP, FOR LOOP, and WHILE LOOP. The EXIT statement transfers control to the end of a loop. The

CONTINUE statement exits the current iteration of a loop and transfers control to the next iteration. Both EXIT and CONTINUE have an optional WHEN clause, where you can specify a condition.

3. *Sequential control statements:* which are not crucial to PL/SQL programming. The sequential control statements are GOTO, which goes to a specified statement, and NULL, which does nothing.

**Static SQL** is a PL/SQL feature that allows SQL syntax directly in a PL/SQL statement.

- *Statements:* These are the PL/SQL static SQL statements, which have the same syntax as the corresponding SQL statements. A PL/SQL static SQL statement can have a PL/SQL identifier wherever its SQL counterpart can have a placeholder for a bind variable. The PL/SQL identifier must identify either a variable or a formal parameter.
- *Pseudocolumns:* A pseudocolumn behaves like a table column, but it is not stored in the table.

**What is a Cursor?**

A cursor is a temporary work area created in the system memory when a SQL statement is executed. A cursor contains information on a select statement and the rows of data accessed by it.

This temporary work area is used to store the data retrieved from the database, and manipulate this data. A cursor can hold more than one row, but can process only one row at a time. The set of rows the cursor holds is called the active set.

PL/SQL controls the context area through a cursor. **Implicit cursors** are automatically created by Oracle whenever an SQL statement is executed, when there is no explicit cursor for the statement. Programmers cannot control the implicit cursors and the information in it. **Explicit cursors** are programmer-defined cursors for gaining more control over the context area. An explicit cursor should be defined in the declaration section of the PL/SQL Block. It is created on a SELECT Statement which returns more than one row.

**Dynamic SQL** is a programming methodology for generating and running SQL statements at run time. It is useful when writing general-purpose and flexible programs like ad hoc query systems, when writing programs that must run database definition language (DDL) statements, or when you do not know at compilation time the full text of a SQL statement or the number or data types of its input and output variables.

*PL/SQL provides two ways to write dynamic SQL:*

- Native dynamic SQL, a PL/SQL language (that is, native) feature for building and running dynamic SQL statements
- DBMS_SQL package, an API for building, running, and describing dynamic SQL statements

Native dynamic SQL code is easier to read and write than equivalent code that uses the DBMS_SQL package, and runs noticeably faster (especially when it can be optimized by the compiler). However, to write native dynamic SQL code, you must know at compile time the number and data types of the input and output variables of the dynamic SQL statement. If you do not know this information at compile time, you must use the DBMS_SQL package.

## 3. PL/SQL Triggers

A trigger is like a stored procedure that Oracle Database invokes automatically whenever a specified event occurs. Like a stored procedure, a trigger is a named PL/SQL unit that is stored in the database and can be invoked repeatedly. Unlike a stored procedure, you can enable and disable a trigger, but you cannot explicitly invoke it. While a trigger is enabled, the database automatically invokes it—that is, the trigger fires—whenever its triggering event occurs. While a trigger is disabled, it does not fire.

You create a trigger with the CREATE TRIGGER statement. You specify the triggering event in terms of triggering statements and the item on which they act. If the trigger is created on a schema or the database, then the triggering event is composed of either DDL or database operation statements, and the trigger is called a system trigger. A conditional trigger has a WHEN clause that specifies a SQL condition that the database evaluates for each row that the triggering statement affects.

## 4. Big Data

Big data is a term that describes the large volume of data – both structured and unstructured – that inundates a business on a day-to-day basis. But it's not the amount of data that's important. It's what organizations do with the data that matters. Big data can be analyzed for insights that lead to better decisions and strategic business moves. Big data describes a holistic information management strategy that includes and integrates many new types of data and data management alongside traditional data.

**Big data has also been defined by the four Vs:**

1. **Volume:** The amount of data. While volume indicates more data, it is the granular nature of the data that is unique. Big data requires processing high volumes of low-density, unstructured Hadoop data—that is, data of unknown value, such as Twitter data feeds, click streams on a web page and a mobile app, network traffic, sensor-enabled equipment capturing data at the speed of light, and many more. It is the task of big data to convert such Hadoop data into valuable information.
2. **Velocity:** The fast rate at which data is received and perhaps acted upon. The highest velocity data normally streams directly into memory versus being written to disk. Some Internet of Things (IoT) applications have health and safety ramifications that require real-time evaluation and action. Other internet-enabled smart products operate in real time or near real time.
3. **Variety:** New unstructured data types. Unstructured and semi-structured data types, such as text, audio, and video require additional processing to both derive meaning and the supporting metadata. Once understood, unstructured data has many of the same requirements as structured data, such as summarization, lineage, auditability, and privacy.
4. **Value:** Data has intrinsic value—but it must be discovered. There are a range of quantitative and investigative techniques to derive value from data—from discovering a consumer preference or sentiment, to making a relevant offer by location, or for identifying a piece of equipment that is about to fail.