1. Basic Computer Science

Generations of Computers

A generation of computers refers to the specific improvements in computer technology with time. In 1946, electronic pathways called circuits were developed to perform the counting. It replaced the gears and other mechanical parts used for counting in previous computing machines.

In each new generation, the circuits became smaller and more advanced than the previous generation circuits. The miniaturization helped increase the speed, memory and power of computers. There are five generations of computers which are described below;

First Generation

The first generation (1946-1959) computers were slow, huge and expensive. In these computers, vacuum tubes were used as the basic components of CPU and memory. These computers were mainly depended on batch operating system and punch cards. Magnetic tape and paper tape were used as output and input devices in this generation;
Some of the popular first generation computers are;

- **ENIAC** (Electronic Numerical Integrator and Computer)
- **EDVAC** (Electronic Discrete Variable Automatic Computer)
- **UNIVACI**(Universal Automatic Computer)
- **IBM-701**
- **IBM-650**

Second Generation

The second generation (1959-1965) was the era of the transistor computers. These computers used transistors which were cheap, compact and consuming less power; it made transistor computers faster than the first generation computers.
In this generation, magnetic cores were used as the primary memory and magnetic disc and tapes were used as the secondary storage. Assembly language and programming languages like COBOL and FORTRAN, and Batch processing and multiprogramming operating systems were used in these computers.
Some of the popular second generation computers are;

- **IBM 1620**
- **IBM 7094**
- **CDC 1604**
- **CDC 3600**
- **UNIVAC 1108**

Third Generation

The third generation computers used integrated circuits (ICs) instead of transistors. A single IC can pack huge number of transistors which increased the power of a computer and reduced the cost. The computers also became more reliable, efficient and smaller in size. These generation computers used remote processing, time-sharing, multi programming as operating system. Also, the high-level programming languages like FORTRON-II TO IV, COBOL, PASCAL PL/1, ALGOL-68 were used in this generation.

Some of the popular third generation computers are;

- **IBM-360 series**
- **Honeywell-6000 series**
- **PDP(Personal Data Processor)**

- **IBM-370/168**
- **TDC-316**

Fourth Generation

The fourth generation (1971-1980) computers used very large scale integrated (VLSI) circuits; a chip containing millions of transistors and other circuit elements. These chips made this generation computers more compact, powerful, fast and affordable. These generation computers used real time, time sharing and distributed operating system. The programming languages like C, C++, DBASE were also used in this generation.
Some of the popular fourth generation computers are;

- **DEC 10**
- **STAR 1000**
- **PDP 11**
- **CRAY-1(Super Computer)**
- **CRAY-X-MP(Super Computer)**

Fifth Generation

In fifth generation (1980-till date) computers, the VLSI technology was replaced with ULSI (Ultra Large Scale Integration). It made possible the production of microprocessor chips with ten million electronic components. This generation computers used parallel processing hardware and AI (Artificial Intelligence) software. The programming languages used in this generation were C, C++, Java, .Net, etc.
Some of the popular fifth generation computers are;

- **Desktop**
- **Laptop**
- **NoteBook**
- **UltraBook**
- **ChromeBook**

Components of Computer system

Input

The keyboard of your computer is one of the most commonly used input devices. Other commonly used input devices are the mouse, floppy disk drive, hard disk drive and magnetic tape. Regardless of the type of input device used in a computer system, all input device perform the following functions.

- Accept data and instruction from the outside word
- Convert it to a form that the computer can understand.
- Supply the converted data to the computer system for further processing.

Output

The output unit of a computer provides the information and results of a computation to the outside world. Printer and Video Display Unit (VDU, also called display screen)are commonly used output devices. Other commonly used output devices are floppy disk drive, hard disk drive and magnetic tape drive in the earl ear generation computers, paper tape punch units and card punch units were also used as output devices.

Storage

The storage unit of the computer holds the data and instruction that you enter through the input unit before these are processed. It preserves the intermediate and final results before these are sent to the output devices.

It is also used to preserve the data for later use: e.g. you may like to save letter you type today for printing after one week. The various storage devices used in computer system are classified into two categories-primary and secondary.

Primary Storage

The primary storage also called the primary memory, store and provides information very fast. This is generally used to hold the program being currently executed in the computer, the data being received from the input unit and the intermediate and final results of the program. The primary generally loses its content when you switch off the computer. Therefore if you need to preserve the results or the input data, you have to transfer it to the secondary storage. The cost of primary storage is more compare to the secondary storage. Therefore, most computers have limited primary storage. Most of the computers use 'semiconductor memory' as primary storage.

Secondary Storage

On the other hand, the secondary storage (Memory) is used
Databases; etc .The program that you want to run on the computer is first Transferred to the primary memory before it can run. Similarly, after running the Program, if you need to save the result, you will transfer them to the secondary Storage. The secondary memory is slower and cheaper than the primary memory. Some of the commonly used secondary memory devices are floppy diskette, zip diskette, hard disk and magnetic tape.

Arithmetic-Logic Unit

All calculations are performed in the Arithmetic Logic Unit (ALU) of the computer. ALU also dose compressions and takes decisions. Whenever calculation has to be done, the control unit transfers the required data from the storage unit to ALU. The ALU can perform basic orations such as additions, subtractions, multiplications, division, etc. the ALU can also do logical operations: e.g. it can check if the number a is less than, equal to or greater than the number b. after the ALU has performed the calculation or the logical operation, the result is transfer to the storage unit.

Control Unit

The control unit controls all other units in the computer. The input unit does not know when to receive data and where to put the data in the storage unit after receiving it. It is the control unit that gives the necessary instructions to the input unit. Similarly, the control unit instructs the input unit where to store the data after receiving it from the user. In the same way, it controls the flow of data and instructions from the storage unit to ALU. It also controls the flow of the result from ALU to the storage unit. The control unit also controls what should be sent to the output unit and when. In brief, the control unit is the central nervous system of the computer that controls and synchronizes its working.

Central Processing Unit

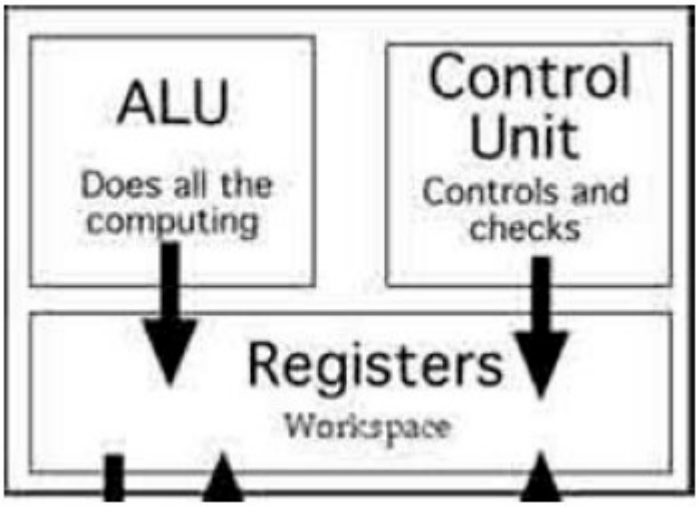
The control unit and ALU of the computer are together known as the central processing unit (CPU). In most modern computers, a single IC does the job of controlling all units of the computer. The same IC also contains the ALU. The CPU is like a computer's brain:
----- It performs all calculations.
----- It takes all Decisions.
---- It controls all units of the computer.

CPU(Central processing Unit .Alternately referred to as a processor, central processor, or microprocessor):is a

- Set of electronic circuitry that executes program instructions
- Converts data into information
- Acts as Control center i.e. it controls all the devices connected to system. because of this function it is called as Brain of Computer System

The three components of the CPU are following,

1. Arithmetic Logic Unit
2. Control Unit
3. Registers



ALU (arithmetic logic unit)

- Performs calculations , logical operations and comparisons (data changed)

Registers

- Small, *permanent* storage locations within the CPU used for a particular purpose
- Manipulated directly by the Control Unit
- Wired for specific function
- Size in bits or bytes (not MB like memory)
- Can hold data, an address or an instruction

Special-Purpose Registers

- *Program Count Register* (PC) Also called instruction pointer. it contains the memory address of instruction that is being executed by CPU ,after execution of instruction it points to address of memory location where next instruction to be executed is stored and these steps are repeated till all the instructions of the program are executed.

- *Instruction Register (IR)* Stores instruction fetched from memory

- *Memory Address Register (MAR)*
- *Memory Data Register (MDR)*
- *Status Registers* Status of CPU and currently executing program
  - *Flags* (one bit Boolean variable) to track condition like arithmetic carry and overflow, power failure, internal computer error

Register Operations

- Stores values from other locations (registers and memory)
- Addition and subtraction
- Shift or rotate data
- Test contents for conditions such as zero or positive

Control unit

- Part of the CPU that generates control signals and controls all operations of computer
- Moves data to and from CPU registers and other hardware components (no change in data)
- Accesses program instructions and issues commands to the ALU
- Directs the computer system to execute program instructions
- Communicates with other parts of the hardware through exchange of control signals

The picture below is an example of what the top and bottom of an Intel Pentium motherboardheat sink processor may look. The processor is placed and secured into a compatible CPU socket found on the motherboard. Processors produce heat, so they are covered with a fan to keep them cool and running smoothly.

As you can see in the above picture, the CPU chip is usually in the shape of a square or rectangle and has one notched corner to help place the chip properly into the CPU socket. On the bottom of the chip are hundreds of connector pins that plug into each of the corresponding holes in the socket. Today, most CPU's resemble the picture shown above. However, Intel AMD slot processors sockets and have also experimented with that were much larger and slid into a slot on the motherboard. Also, over the years, there have been dozens of different types of on motherboards. Each socket only supports specific types of processors and each has its own pin layout.

Bus

- The physical connection that makes it possible to transfer data from one location in the computer system to another
  - Group of electrical conductors for carrying signals from one location to another *Line*: each conductor in the bus

  - 4 kinds of signals Data (alphanumeric, numerical, instructions)
  - Addresses
  - Control signals
  - Power (sometimes)

- Connect CPU and Memory
- I/O peripherals: on same bus as CPU/memory or separate bus
  - Physical packaging commonly called *backplane* Also called *system bus* or *external bus*
  - Example of *broadcast bus*
  - Part of printed circuit board called *motherboard* that holds CPU and related components

Bus Characteristics

- Protocol Documented agreement for communication
- Specification that spells out the meaning of each line and each signal on each line

- Throughput, i.e., data transfer rate in bits per second
- Data width in bits carried simultaneously

Number Systems

The language we use to communicate with each other is comprised of words and characters. We understand numbers, characters and words. But this type of data is not suitable for computers. Computers only understand the numbers.

So, when we enter data, the data is converted into electronic pulse. Each pulse is identified as code and the code is converted into numeric format by ASCII. It gives each number, character and symbol a numeric value

(number) that a computer understands. So to understand the language of computers, one must be familiar with the number systems.

The Number Systems used in computers are:

- Binary number system
- Octal number system
- Decimal number system
- Hexadecimal number system

Binary number system

It has only two digits '0' and '1' so its base is 2. Accordingly, In this number system, there are only two types of electronic pulses; absence of electronic pulse which represents '0'and presence of electronic pulse which represents '1'. Each digit is called a bit. A group of four bits (1101) is called a nibble and group of eight bits (11001010) is called a byte. The position of each digit in a binary number represents a specific power of the base (2) of the number system.

Internet (Internet, Intranet, Extranet, Websites, Email)

- The **Internet** is a globally-connected network of computers that enables people to share information and communicate with each other.
- An **intranet**, on the other hand, is a local or restricted network that enables people to store, organize, and share information within an organization.

  Besides the spelling and pronunciation (which might sound forced and awkward at times), there are key differences between the two, one of the most important being that an intranet is a platform that can be bought and sold (or built, in some cases), while the Internet is the underlying technology that enables its connectivity.

  What is an extranet?

  There's one more type of 'net' to consider, and it relates to enterprise collaboration with external users:
- An **extranet** is a web portal that is accessible by an organization and its external vendors, partners, customers, or any other users that require access to restricted information.

  With an extranet, the host organization manages the site administration and content, and provides controlled access to internal and external members. Some example use cases for an extranet include a *partner or vendor portal*, a *customer community*, or a *franchise network.*

Website: Website a collection of contents / information in the form of pages that is a part of website which comprises links to other website is a content that is to be displayed on a website. The web page URL has an extension. Web page address depends on website address. Requires less time to develop as it is a part of a website.

Email: It is an application, server or email client which is used for the carrying of email through a web server. Outlook, Gmail and Yahoo are some common examples of webmail providers. The main advantage of webmail is that now you can send, receive, make folders, reply, forwards, filter and store the emails anytime from a web browser by having an internet connection.

Email address is nothing else but a website. In fact, it is a small part of a website as email is a service that is used to send and receive messages. A website is commonly perceived as a collection of pages containing information or used for shopping purposes. However, there are many more purposes of websites such as social networking (like Facebook, Twitter etc), sharing of video clips (like You tube), search engine (like Google, Yahoo, MSN etc). Email clients like Gmail, yahoo mail, AOL etc are also websites that are exclusively being used for sending and receiving emails. All one has to do is to become a member by opening an account with a mail client and add others having an account with any email client.

What is email client?

If you've used a program like Microsoft Outlook, Windows Live Mail, Mozilla Thunderbird or Apple Mail to manage your emails, then you've used an email client.
An email client is a piece of software that is installed on your computer. You then use this software to download your emails from the server to your computer and from there you can read and send emails. In other words, in order to use an email client to access your domain emails, you will first need to install the email client software on your computer.

What is computer security?

Computer security basically is the protection of computer systems and information from harm, theft, and unauthorized use. It is the process of preventing and detecting unauthorized use of your computer system. There are various types of computer security which is widely used to protect the valuable information of an organization.

What is Computer Security and its types?

One way to ascertain the similarities and differences among Computer Security is by asking what is being secured. For example,

- *Information security* is securing information from unauthorized access, modification & deletion
- *Application Security* is securing an application by building security features to prevent from Cyber Threats such as SQL injection, DoS attacks, data breaches and etc.
- C*omputer Security* means securing a standalone machine by keeping it updated and patched
- *Network Security* is by securing both the software and hardware technologies
- *Cybersecurity* is defined as protecting computer systems, which communicate over the computer networks

It's important to understand the distinction between these words, though there isn't necessarily a clear consensus on the meanings and the degree to which they overlap or are interchangeable.

So, **Computer security** can be defined as controls that are put in place to provide confidentiality, integrity, and availability for all components of computer systems. Let's elaborate the definition.

2. Introduction to OS and Hardware Basics

ANATOMY OF A COMPUTER

The internal design of computers differs from one model to another. But the basic components of computer remain the same for all models. To function properly, a computer needs both hardware and software. Hardware consists of the mechanical and electronic devices which we can see and touch. Key Board, Monitor, DVD are some examples for Computer Hardware. The software consists of programs, the operating systems and the data that reside in the memory and storage devices. JAVA, Microsoft Office, Open Office are some examples for Computer Software.

A computer mainly performs the following four functions.

1. **Receive input** – accept information from outside through various input devices like keyboard, mouse etc.
2. **Process information** – perform arithmetic or logical operations on the information.
3. **Produce output** – communicate information to the outside world through output devices like monitor, printer etc.
4. **Store information** – store the information in storage devices like hard disk, compact disk etc.

A computer has the following three main components.

a. Input/ Output Unit
b. Central Processing Unit
c. Memory Unit

a) Input/ Output Unit: Computer is a machine that processes the input data according to a given set of instructions and gives the output. The unit used for getting the data and instructions into the computer and

displaying or printing output is known as input/ output unit. Keyboard is the main input device while the monitor is the main output device.

b) Central Processing Unit: Central processing Unit (CPU) is the main component or 'brain' of the computer which performs all the processing of input data. In micro computers, the CPU is built on a single chip or Integrated Circuit (IC) and is called Microprocessor. The CPU consists of the following distinct parts:

i. Arithmetic Logic Unit (ALU)
ii. Control Unit (CU)
iii. Registers
iv. Buses
v. Clock

(i) **Arithmetic Logic Unit:** The arithmetic logic unit is responsible for all arithmetic operations like addition, subtraction, multiplication and divisions as well as logical operations such as less than, equal to and greater than.

(ii) **Control Unit:** The control unit is responsible for controlling the transfer of data and instructions among other units of a computer. It is considered as the 'Central Nervous System' of computer as it manages and coordinates all the units of the computer. It obtains the instructions from the memory, interprets them and directs the operation of the computer.

(iii) **Registers:** Registers are small high speed circuits which are used to store data, instructions and memory addresses, when ALU performs arithmetic and logical operations. Depending on the processor's capability, the number and type of registers vary from one CPU to another.

(iv) **Buses:** Data is stored as a unit of eight bits in a register. Each bit is transferred from one register to another by means of a separate wire. This group of eight wires which is used as a common way to transfer data between registers is known as a bus. Bus is a connection between two components to transmit signal between them. Bus is of three major types namely data bus, control bus and address bus.

(v) **Clock:** Clock is an important component of CPU which measures and allocates a fixed time slot for processing each and every micro-operation. CPU executes the instructions in synchronization with the clock pulse. The clock speed of CPU is measured in terms of Mega Hertz or millions of cycles per second. The clock speed of CPU varies from one model to another.

c) **Memory Unit:** Memory unit is used to store the data, instructions and information before, during and after the processing by ALU. It is actually a work area (physically a collection of integrated circuits) within the computer where the CPU stores the data and instructions. Memory is of two types:

i. Read Only Memory (ROM)
ii. Random Access Memory (RAM)

(i) **Read Only Memory:** Read Only Memory is an essential component of the memory unit. The memory which has essential instructions is known as Read Only Memory. This memory is permanent and is not erased when the system is switched off. The memory capacity of ROM varies from 64 KB to 256 KB depending on the model of computer.

(ii) **Random Access Memory:** Random Access Memory is used to store data and instructions during the execution of programs. Contrary to ROM, RAM is temporary and is erased when the computer is switched off. RAM is a read/ write type of memory and thus can be read and written by the user. As it is possible to randomly use any location of this memory, it is known as random access memory. The memory capacity of RAM varies from 640 KB to several mega bytes with different models of computer.

Hardware and software are two broad categories of computer components. Hardware refers to physical component while software to the programs required to operate computers.

INPUT DEVICES

An input device is any machine that feeds data, information and instructions into a computer. We may classify input devices into the following two broad categories.

i. Basic input devices
ii. Special input devices

Basic Input Devices: The input devices which are essential to operate a PC are called basic input devices. These devices are always required for basic input operations. These devices include keyboard and mouse.

Special Input Devices: The input devices which are not essential to operate a PC are called special input devices. These devices are used for various special purposes and are generally required for basic input operations. These devices include Trackball, Light Pen, Touch Screen, Joystick, Digitizer, Scanner, Optical Mark Reader (OMR), Bar Code Reader (BCR), Optical Character Reader (OCR), Magnetic Ink Character Recognition (MICR) and Voice-Input Devices.

Keyboard

Keyboard is the most common input device used for manual data entry. Computer keyboards are similar to electric-typewriter keyboards but contain additional keys. Keyboard has been standardized for use in all types of computers such as a PC, a workstation or a notebook computer. The keys on computer keyboards are classified as follows:

1. **Letter Keys:** These are the 26 letters of English alphabet arranged as in a typewriter.
2. **Digit Keys:** There are two sets of digit keys; one on the second row from the top of the keys just as in a typewriter and the other is a numeric key pad at the bottom right which allows quick entry of numbers with the fingers of one hand.
3. **Special character keys:** These are characters such as <, >, ?, /, {, }, [,], (,), ., “, @, #, $, %, &, *, etc
4. **Non-printable control keys:** These are used for backspacing, going to the next line, tabulation, moving the cursor up or down, insert, delete characters etc. There is also a space bar at the bottom for leaving a space.
5. **Function keys:** These are labeled F1, F2 up to F15 and when pressed invoke programs stored in the computer.

You can understand the function of each and every key actually by working on a PC. When any key is pressed, an electric signal is produced. This signal is detected by a keyboard encoder that sends a binary code corresponding to the key pressed to the CPU. There are many types of keyboards but 101 keys board is the most popular one.

Mouse

Mouse is a device that controls the movement of the cursor on the display screen. It is a small object you can roll along a hard, flat surface.

Input Devices

Trackball:

Trackball is an input device which is mostly used in notebook or laptop computer instead of a mouse. This is a ball which is half inserted and moving fingers on the ball, pointer can be moved. Trackball is considered better than mouse because it requires little arm movement and less desktop space.

Light Pen:

Light pen is a pointing device which is similar to a pen. It is used to select a displayed menu item or draw pictures on the monitor screen. It consists of a photocell and an optical system placed in a small tube. When the tip of a light pen is pressed, its photocell sensing element detects the screen location and sends the corresponding signal to the CPU.

Joystick:

Joystick is a pointing device which is used to move cursor position on a monitor screen. Joystick is a stick having a spherical ball at its both lower and upper ends. The lower spherical ball moves in a socket. Joystick can be moved in all four directions. The function of joystick is similar to that of a mouse. It is mainly used in Computer Aided Designing (CAD) and playing computer games.

Scanner:

Scanner works more like a photocopy machine. It is used when some information is available on a paper and it is to be transferred to the hard disk of the computer for further manipulation. Scanner captures images from the source which are then converted into the digital form that can be stored on the disc. These images can be edited before they are printed.

Optical Mark Reader (OMR):

OMR is a special type of optical scanner used to recognize the type of mark made by pen or pencil. It is used where one out of a few alternatives is to be selected and marked. It is specially used for checking the answer sheets of examinations having multiple choice questions.

Bar Code Reader (BCR):

BCR is an optical scanner used for reading bar-coded data (data in the form of light and dark lines) Bar coded data is generally used in labeling goods, numbering of books etc. Bar Code Reader scans a bar code image, converts it into an alphanumeric value which is then fed to the computer to which the Bar Code Reader is connected.

Optical Character Reader (OCR):

OCR is an optical scanner used to read a printed text. OCR scans text optically character by character, converts them into a machine readable code and stores the text on the system memory. It is used for reading of passenger tickets, computer printed bills of credit card companies and reading of ZIP codes in postal services.

Magnetic Ink Character Reader (MICR):

MICR is generally used in banks because of a large number of cheques to be processed every day. The bank's code number and cheque number are printed on the cheques with a special type of ink that contains particles of magnetic material that are machine readable. This reading process is called Magnetic Ink Character Recognition.

Voice-Input Devices:

Voice-input devices are the latest input devices that can recognize the human voice. Microphone is a voice input device to input sound which is then stored in digital form. It is used for various applications like adding sound to a multimedia presentation or for mixing music.

Memory Management in Operating System

- Difficulty Level : Easy
- Last Updated : 18 Aug, 2021

The term Memory can be defined as a collection of data in a specific format. It is used to store instructions and processed data. The memory comprises a large array or group of words or bytes, each with its own location. The primary motive of a computer system is to execute programs. These programs, along with the information they access, should be in the main memory during execution. The CPU fetches instructions from memory according to the value of the program counter.

To achieve a degree of multiprogramming and proper utilization of memory, memory management is important. Many memory management methods exist, reflecting various approaches, and the effectiveness of each algorithm depends on the situation.

Here, we will cover the following memory management topics:

- *What is Main Memory*
- *What is Memory Management*
- *Why memory Management is required*
- *Logical address space and Physical address space*
- *Static and dynamic loading*
- *Static and dynamic linking*
- *Swapping*
- *Contiguous Memory allocation*
    - *Memory Allocation*
        - *First Fit*
        - *Best Fit*
        - *Worst Fit*
    - *Fragmentation*
        - *Internal Fragmentation*
        - *External Fragmentation*
    - *Paging*

Now before, We start memory management let us known about what is main memory.

What is Main Memory:

The main memory is central to the operation of a modern computer. Main Memory is a large array of words or bytes, ranging in size from hundreds of thousands to billions. Main memory is a repository of rapidly available information shared by the CPU and I/O devices. Main memory is the place where programs and information are kept when the processor is effectively utilizing them. Main memory is associated with the processor, so moving instructions and information into and out of the processor is extremely fast. Main memory is also known as RAM(Random Access Memory). This memory is a volatile memory.RAM lost its data when a power interruption occurs.

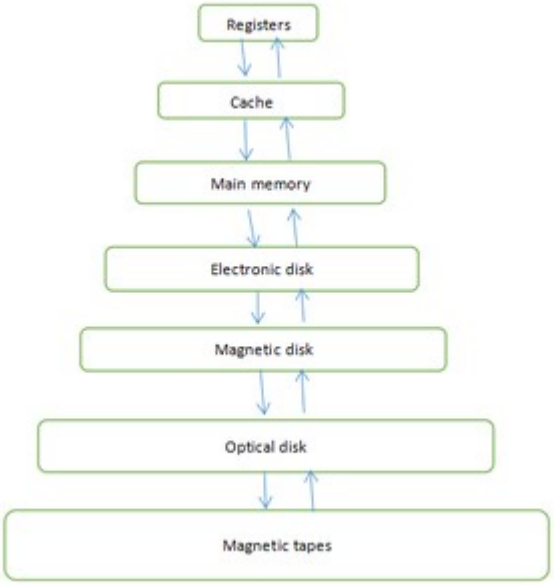


Figure 1: Memory hierarchy

What is Memory Management :

In a multiprogramming computer, the operating system resides in a part of memory and the rest is used by multiple processes. The task of subdividing the memory among different processes is called memory management. Memory management is a method in the operating system to manage operations between main memory and disk during process execution. The main aim of memory management is to achieve efficient utilization of memory.

Why Memory Management is required:

- Allocate and de-allocate memory before and after process execution.
- To keep track of used memory space by processes.
- To minimize fragmentation issues.
- To proper utilization of main memory.
- To maintain data integrity while executing of process.

Now we are discussing the concept of logical address space and Physical address space:

Logical and Physical Address Space:

Logical Address space: An address generated by the CPU is known as “Logical Address”. It is also known as a Virtual address. Logical address space can be defined as the size of the process. A logical address can be changed.

Physical Address space: An address seen by the memory unit (i.e the one loaded into the memory address register of the memory) is commonly known as a “Physical Address”. A Physical address is also known as a Real address. The set of all physical addresses corresponding to these logical addresses is known as Physical address

space. A physical address is computed by MMU. The run-time mapping from virtual to physical addresses is done by a hardware device Memory Management Unit(MMU). The physical address always remains constant.

Static and Dynamic Loading:

To load a process into the main memory is done by a loader. There are two different types of loading :

- **Static loading**:- In static loading load the entire program into a fixed address. It requires more memory space.
- **Dynamic loading**:- The entire program and all data of a process must be in physical memory for the process to execute. So, the size of a process is limited to the size of physical memory. To gain proper memory utilization, dynamic loading is used. In dynamic loading, a routine is not loaded until it is called. All routines are residing on disk in a relocatable load format. One of the advantages of dynamic loading is that unused routine is never loaded. This loading is useful when a large amount of code is needed to handle it efficiently.

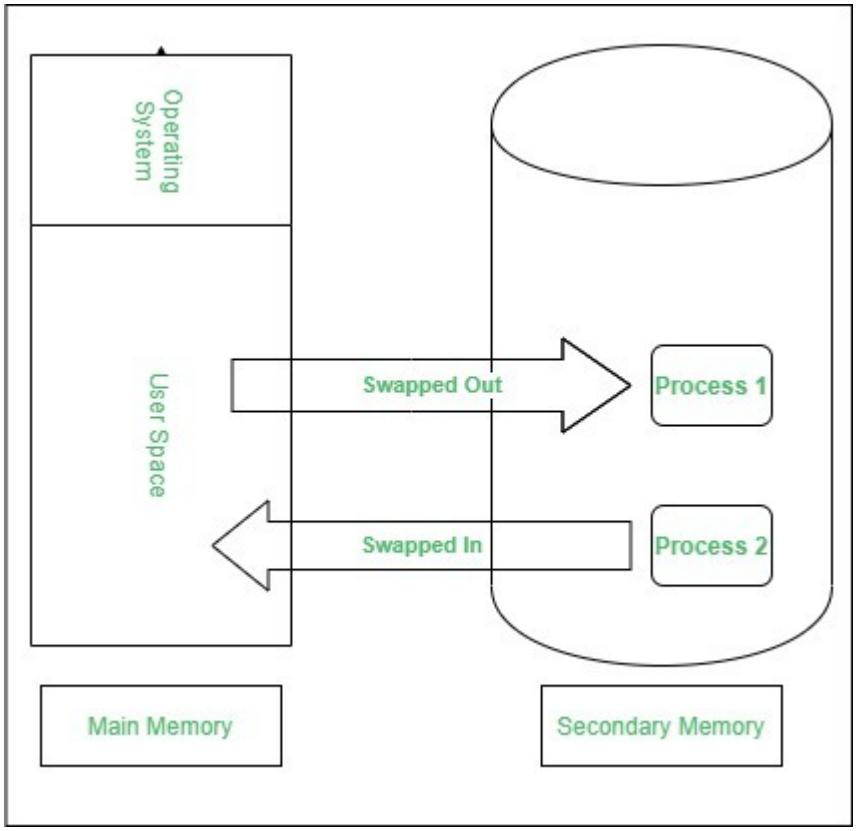
Static and Dynamic linking:

To perform a linking task a linker is used. A linker is a program that takes one or more object files generated by a compiler and combines them into a single executable file.

- **Static linking:** In static linking, the linker combines all necessary program modules into a single executable program. So there is no runtime dependency. Some operating systems support only static linking, in which system language libraries are treated like any other object module.
- **Dynamic linking:** The basic concept of dynamic linking is similar to dynamic loading. In dynamic linking, "Stub" is included for each appropriate library routine reference. A stub is a small piece of code. When the stub is executed, it checks whether the needed routine is already in memory or not. If not available then the program loads the routine into memory.

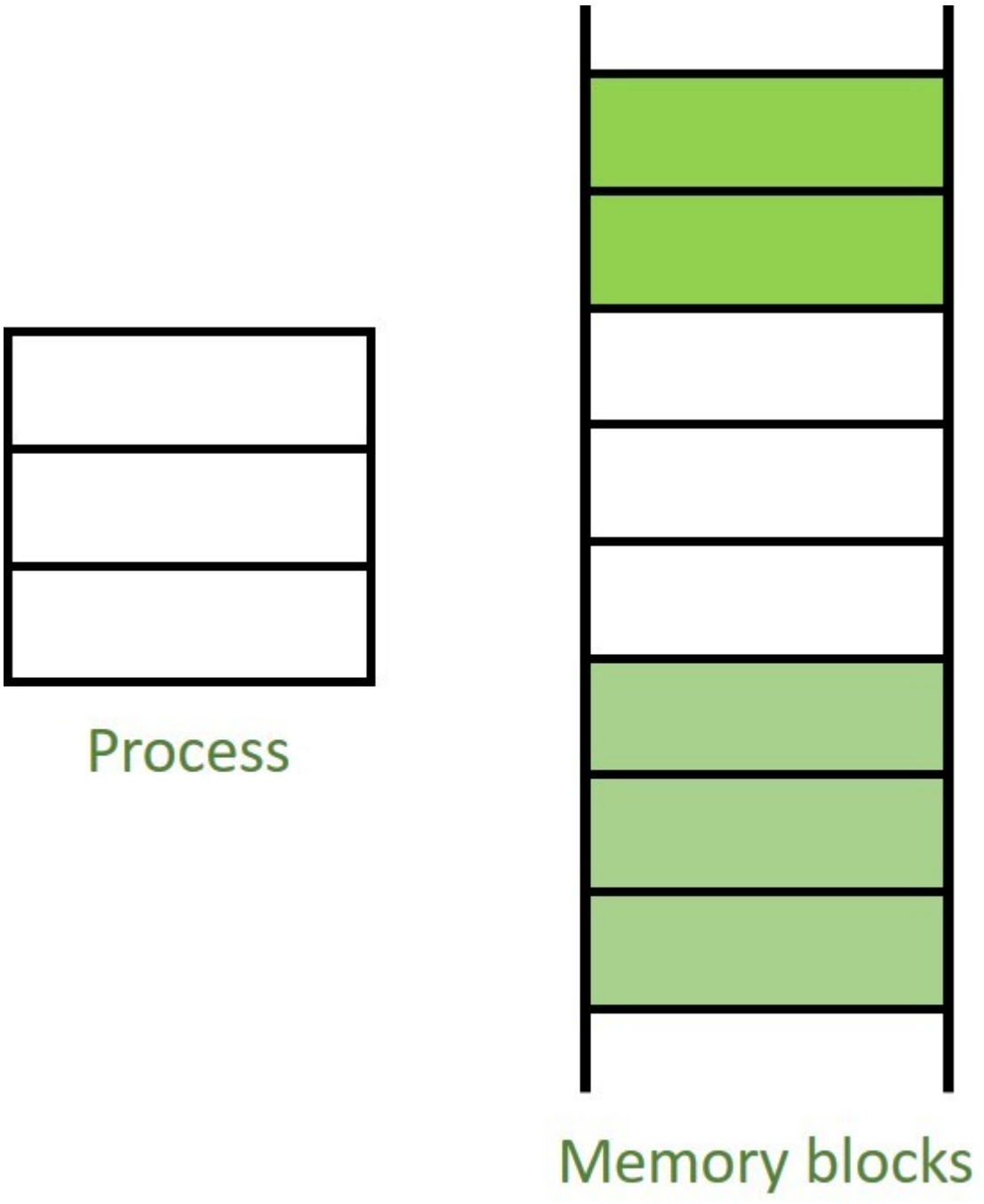
Swapping :

When a process is executed it must have resided in memory. Swapping is a process of swap a process temporarily into a secondary memory from the main memory, which is fast as compared to secondary memory. A swapping allows more processes to be run and can be fit into memory at one time. The main part of swapping is transferred time and the total time directly proportional to the amount of memory swapped. Swapping is also known as roll-out, roll in, because if a higher priority process arrives and wants service, the memory manager can swap out the lower priority process and then load and execute the higher priority process. After finishing higher priority work, the lower priority process swapped back in memory and continued to the execution process.



Contiguous Memory Allocation :

The main memory should oblige both the operating system and the different client processes. Therefore, the allocation of memory becomes an important task in the operating system. The memory is usually divided into two partitions: one for the resident operating system and one for the user processes. We normally need several user processes to reside in memory simultaneously. Therefore, we need to consider how to allocate available memory to the processes that are in the input queue waiting to be brought into memory. In adjacent memory allotment, each process is contained in a single contiguous segment of memory.



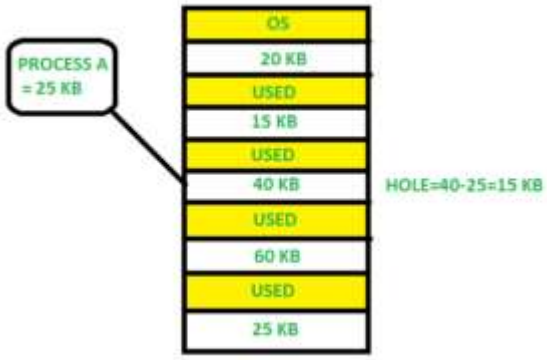
Contiguous Memory Allocation

Memory allocation:
To gain proper memory utilization, memory allocation must be allocated efficient manner. One of the simplest methods for allocating memory is to divide memory into several fixed-sized partitions and each partition contains exactly one process. Thus, the degree of multiprogramming is obtained by the number of partitions.
Multiple partition allocation: In this method, a process is selected from the input queue and loaded into the free partition. When the process terminates, the partition becomes available for other processes.

Fixed partition allocation: In this method, the operating system maintains a table that indicates which parts of memory are available and which are occupied by processes. Initially, all memory is available for user processes and is considered one large block of available memory. This available memory is known as “Hole”. When the process arrives and needs memory, we search for a hole that is large enough to store this process. If the requirement fulfills then we allocate memory to process, otherwise keeping the rest available to satisfy future requests. While allocating a memory sometimes dynamic storage allocation problems occur, which concerns how to satisfy a request of size n from a list of free holes. There are some solutions to this problem:

First fit:-

In the first fit, the first available free hole fulfills the requirement of the process allocated.



Here, in this diagram 40 KB memory block is the first available free hole that can store process A (size of 25 KB), because the first two blocks did not have sufficient memory space.

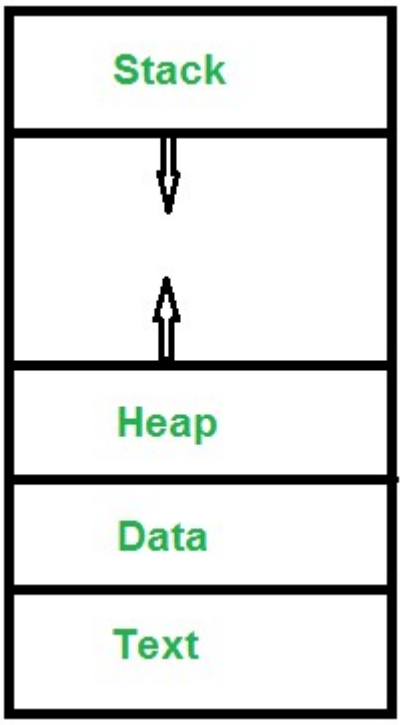
Process Management

Program vs Process

A process is a program in execution. For example, when we write a program in C or C++ and compile it, the compiler creates binary code. The original code and binary code are both programs. When we actually run the binary code, it becomes a process.

A process is an ‘active’ entity, instead of a program, which is considered a ‘passive’ entity. A single program can create many processes when run multiple times; for example, when we open a .exe or binary file multiple times, multiple instances begin (multiple processes are created).

What does a process look like in memory?



Text Section*:* A Process, sometimes known as the Text Section, also includes the current activity represented by the value of the ***Program Counter***.
Stack*:* The stack contains temporary data, such as function parameters, returns addresses, and local variables.
Data Section*:* Contains the global variable.
Heap Section*:* Dynamically allocated memory to process during its run time.
Refer to this for more details on sections.

Attributes or Characteristics of a Process
A process has the following attributes.
div block
1. Process Id: A unique identifier assigned by the operating system
2. Process State: Can be ready, running, etc.
3. CPU registers: Like the Program Counter (CPU registers must be saved and restored when a process is swapped in and out of CPU)
5. Accounts information:
6. I/O status information: For example, devices allocated to the process, open files, etc
8. CPU scheduling information: For example, Priority (Different processes may have different priorities, for example a shorter process assigned high priority in the shortest job first scheduling)

All of the above attributes of a process are also known as the ***context of the process***. Every process has its own process control block(PCB), i.e each process will have a unique PCB. All of the above attributes are part of the PCB.

States of Process:
A process is in one of the following states:

1. New: Newly Created Process (or) being-created process.

2. Ready: After creation process moves to Ready state, i.e. the process is ready for execution.

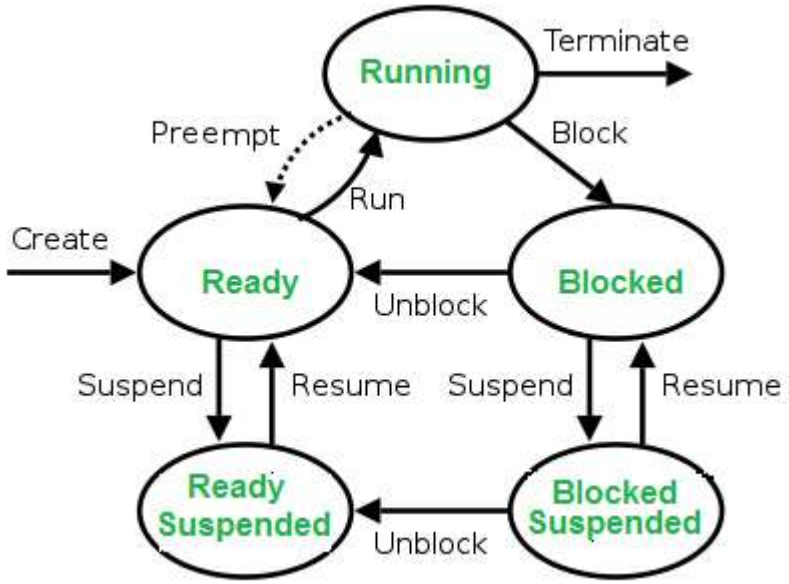
3. Run: Currently running process in CPU (only one process at a time can be under execution in a single processor).

4. Wait (or Block): When a process requests I/O access.

5. Complete (or Terminated): The process completed its execution.

6. Suspended Ready: When the ready queue becomes full, some processes are moved to suspended ready state

7. Suspended Block: When waiting queue becomes full.



Process Address Space

The process address space is the set of logical addresses that a process references in its code. For example, when 32-bit addressing is in use, addresses can range from 0 to 0x7fffffff; that is, 2^31 possible numbers, for a total theoretical size of 2 gigabytes.

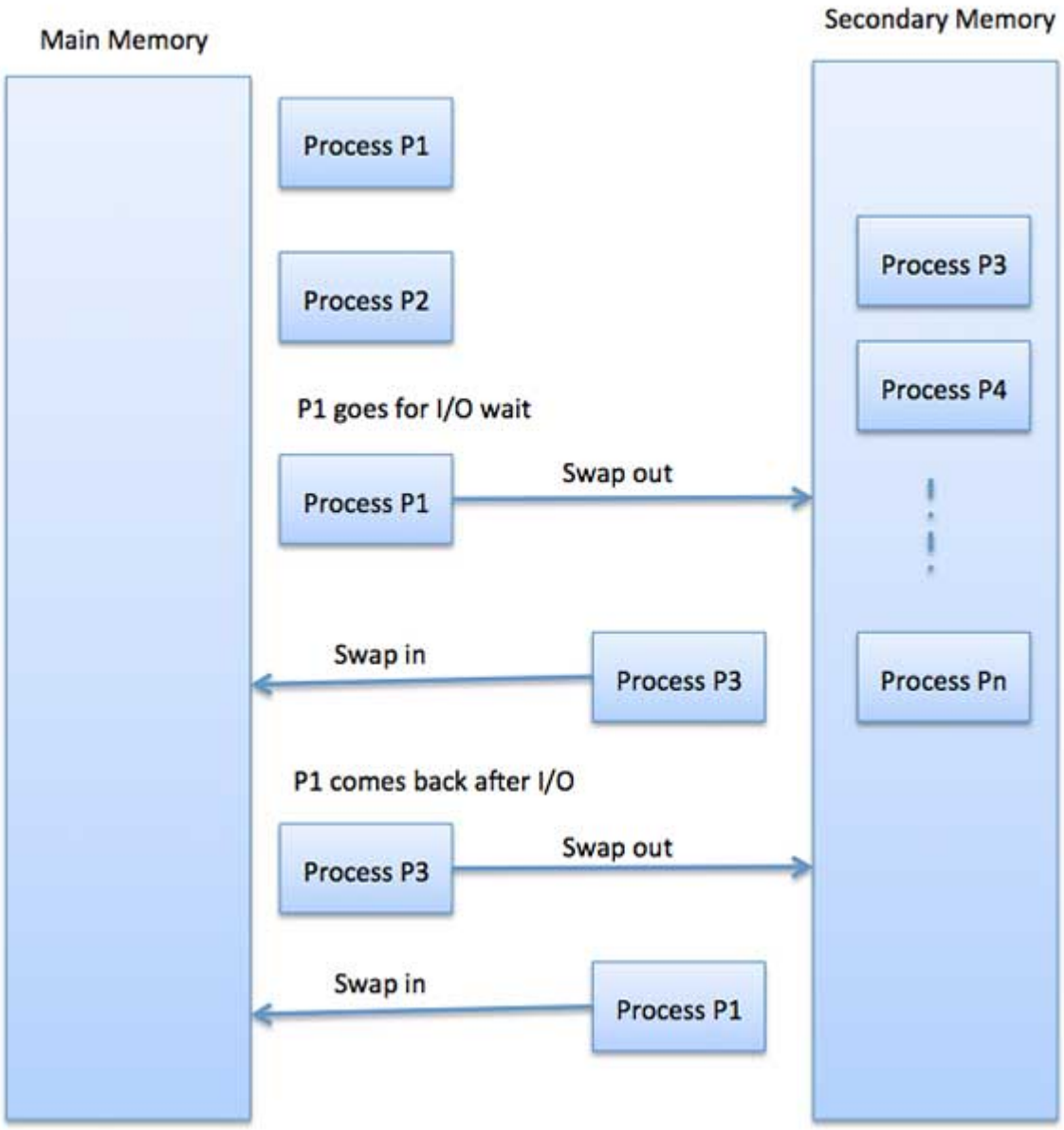
The operating system takes care of mapping the logical addresses to physical addresses at the time of memory allocation to the program. There are three types of addresses used in a program before and after memory is allocated −

S.N.	Memory Addresses & Description
1	**Symbolic addresses** The addresses used in a source code. The variable names, constants, and instruction labels are the basic elements of the symbolic address space.
2	**Relative addresses** At the time of compilation, a compiler converts symbolic addresses into relative addresses.
3	**Physical addresses** The loader generates these addresses at the time when a program is loaded into main memory.

Swapping

Swapping is a mechanism in which a process can be swapped temporarily out of main memory (or move) to secondary storage (disk) and make that memory available to other processes. At some later time, the system swaps back the process from the secondary storage to main memory.

Though performance is usually affected by swapping process but it helps in running multiple and big processes in parallel and that's the reason **Swapping is also known as a technique for memory compaction**.



The total time taken by swapping process includes the time it takes to move the entire process to a secondary disk and then to copy the process back to memory, as well as the time the process takes to regain main memory.

Let us assume that the user process is of size 2048KB and on a standard hard disk where swapping will take place has a data transfer rate around 1 MB per second. The actual transfer of the 1000K process to or from memory will take

2048KB / 1024KB per second

= 2 seconds

= 2000 milliseconds

Now considering in and out time, it will take complete 4000 milliseconds plus other overhead where the process competes to regain main memory.

Memory Allocation

Main memory usually has two partitions −

- **Low Memory** − Operating system resides in this memory.
- **High Memory** − User processes are held in high memory.

Operating system uses the following memory allocation mechanism.

S.N.	Memory Allocation & Description
1	**Single-partition allocation** In this type of allocation, relocation-register scheme is used to protect user processes from each other, and from changing operating-system code and data. Relocation register contains value of smallest physical address whereas limit register contains range of logical addresses. Each logical address must be less than the limit register.
2	**Multiple-partition allocation** In this type of allocation, main memory is divided into a number of fixed-sized partitions where each partition should contain only one process. 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.

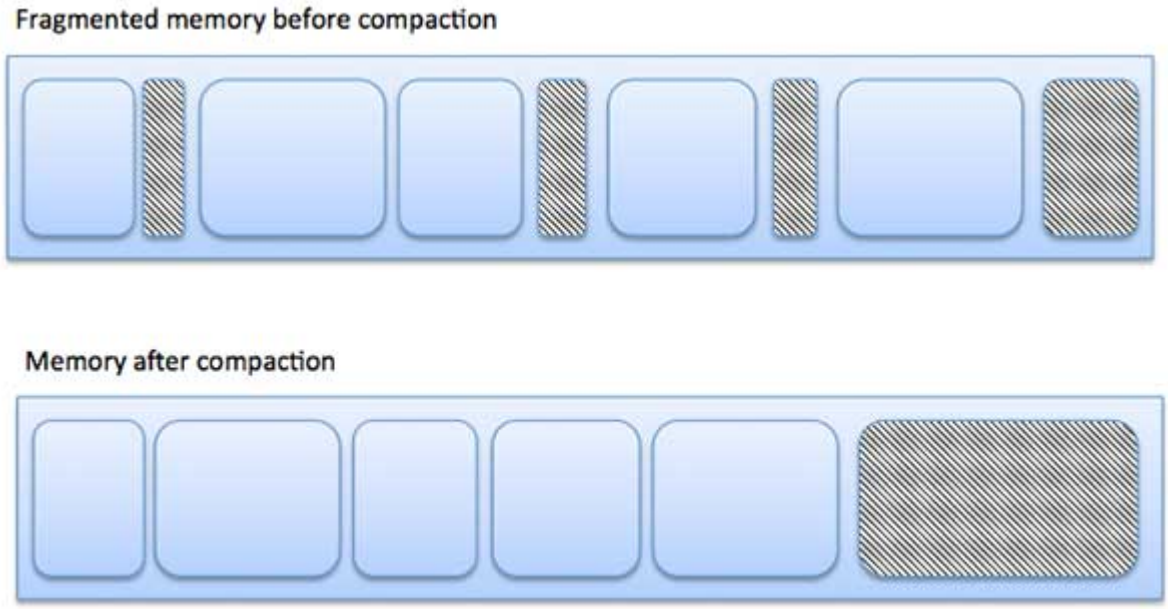
Fragmentation

As processes are loaded and removed from memory, the free memory space is broken into little pieces. It happens after sometimes that processes cannot be allocated to memory blocks considering their small size and memory blocks remains unused. This problem is known as Fragmentation.

Fragmentation is of two types −

S.N.	Fragmentation & Description
1	**External fragmentation** Total memory space is enough to satisfy a request or to reside a process in it, but it is not contiguous, so it cannot be used.
2	**Internal fragmentation** Memory block assigned to process is bigger. Some portion of memory is left unused, as it cannot be used by another process.

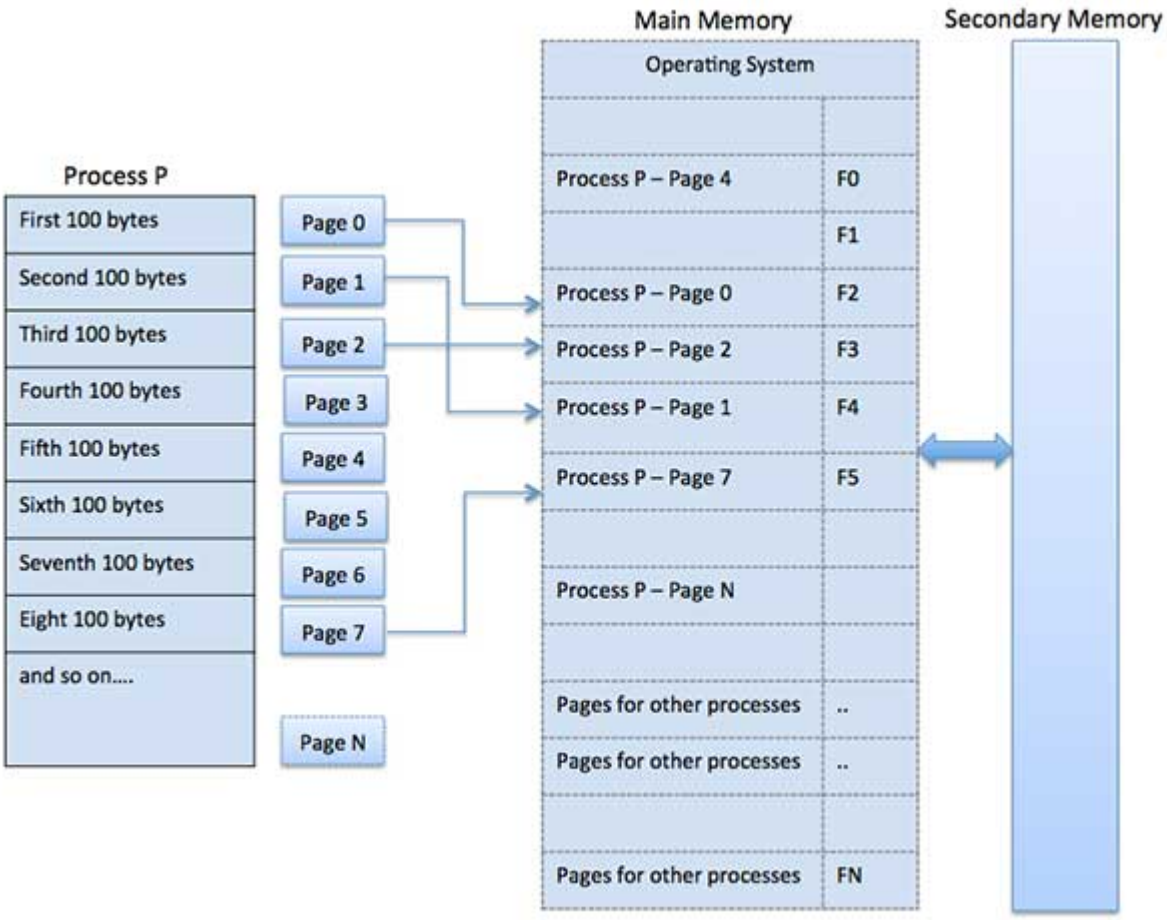
The following diagram shows how fragmentation can cause waste of memory and a compaction technique can be used to create more free memory out of fragmented memory −



External fragmentation can be reduced by compaction or shuffle memory contents to place all free memory together in one large block. To make compaction feasible, relocation should be dynamic.
The internal fragmentation can be reduced by effectively assigning the smallest partition but large enough for the process.

Paging

A computer can address more memory than the amount physically installed on the system. This extra memory is actually called virtual memory and it is a section of a hard disk that's set up to emulate the computer's RAM. Paging technique plays an important role in implementing virtual memory.
Paging is a memory management technique in which process address space is broken into blocks of the same size called **pages** (size is power of 2, between 512 bytes and 8192 bytes). The size of the process is measured in the number of pages.
Similarly, main memory is divided into small fixed-sized blocks of (physical) memory called **frames** and the size of a frame is kept the same as that of a page to have optimum utilization of the main memory and to avoid external fragmentation.



File Management

A file is a collection of related information that is recorded on secondary storage. Or file is a collection of logically related entities. From user's perspective a file is the smallest allotment of logical secondary storage.

Attributes Types Operations

Name Doc Create

Attributes	Types	Operations
Type	Exe	Open
Size	Jpg	Read
Creation Data	Xis	Write
Author	C	Append
Last Modified	Java	Truncate
protection	class	Delete
		Close

File type	Usual extension	Function
Executable	exe, com, bin	Read to run machine language program
Object	obj, o	Compiled, machine language not linked
Source Code	C, java, pas, asm, a	Source code in various languages
Batch	bat, sh	Commands to the command interpreter
Text	txt, doc	Textual data, documents
Word Processor	wp, tex, rrf, doc	Various word processor formats
Archive	arc, zip, tar	Related files grouped into one compressed file
Multimedia	mpeg, mov, rm	For containing audio/video information

FILE DIRECTORIES:

Collection of files is a file directory. The directory contains information about the files, including attributes, location and ownership. Much of this information, especially that is concerned with storage, is managed by the operating system. The directory is itself a file, accessible by various file management routines.

Information contained in a device directory are:

- Name
- Type
- Address
- Current length
- Maximum length
- Date last accessed
- Date last updated
- Owner id
- Protection information

Operation performed on directory are:

- Search for a file
- Create a file
- Delete a file
- List a directory
- Rename a file
- Traverse the file system

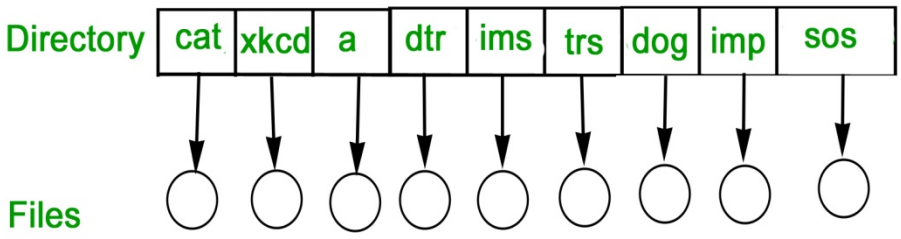
Advantages of maintaining directories are:

- **Efficiency:** A file can be located more quickly.
- **Naming:** It becomes convenient for users as two users can have same name for different files or may have different name for same file.
- **Grouping:** Logical grouping of files can be done by properties e.g. all java programs, all games etc.

SINGLE-LEVEL DIRECTORY

In this a single directory is maintained for all the users.

- **Naming problem:** Users cannot have same name for two files.
- **Grouping problem:** Users cannot group files according to their need.

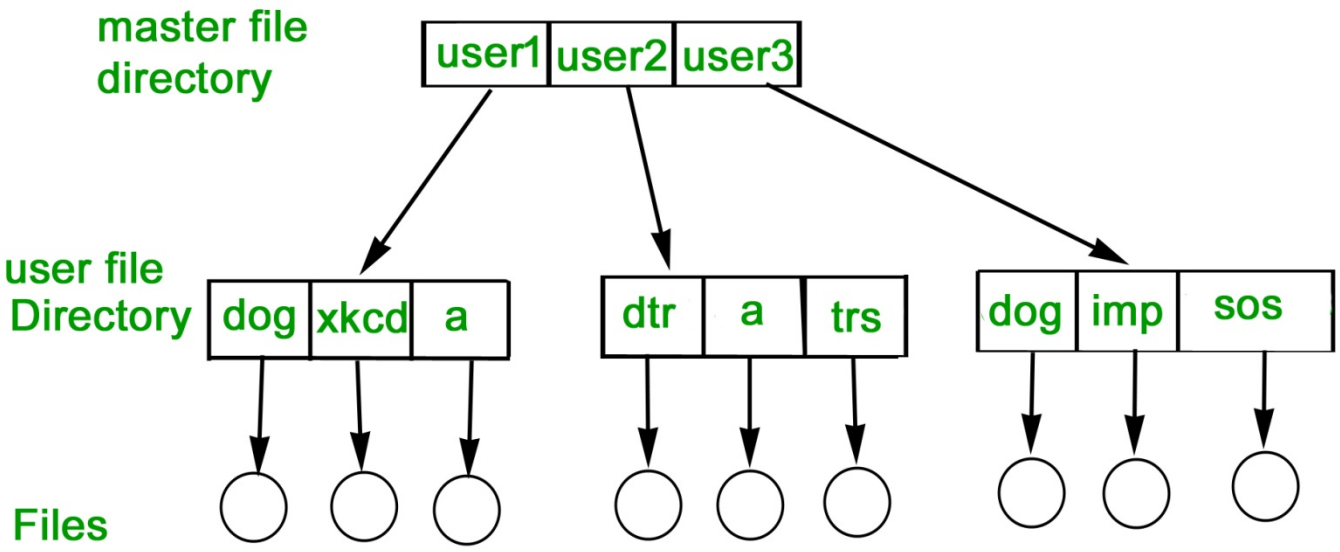


TWO-LEVEL DIRECTORY

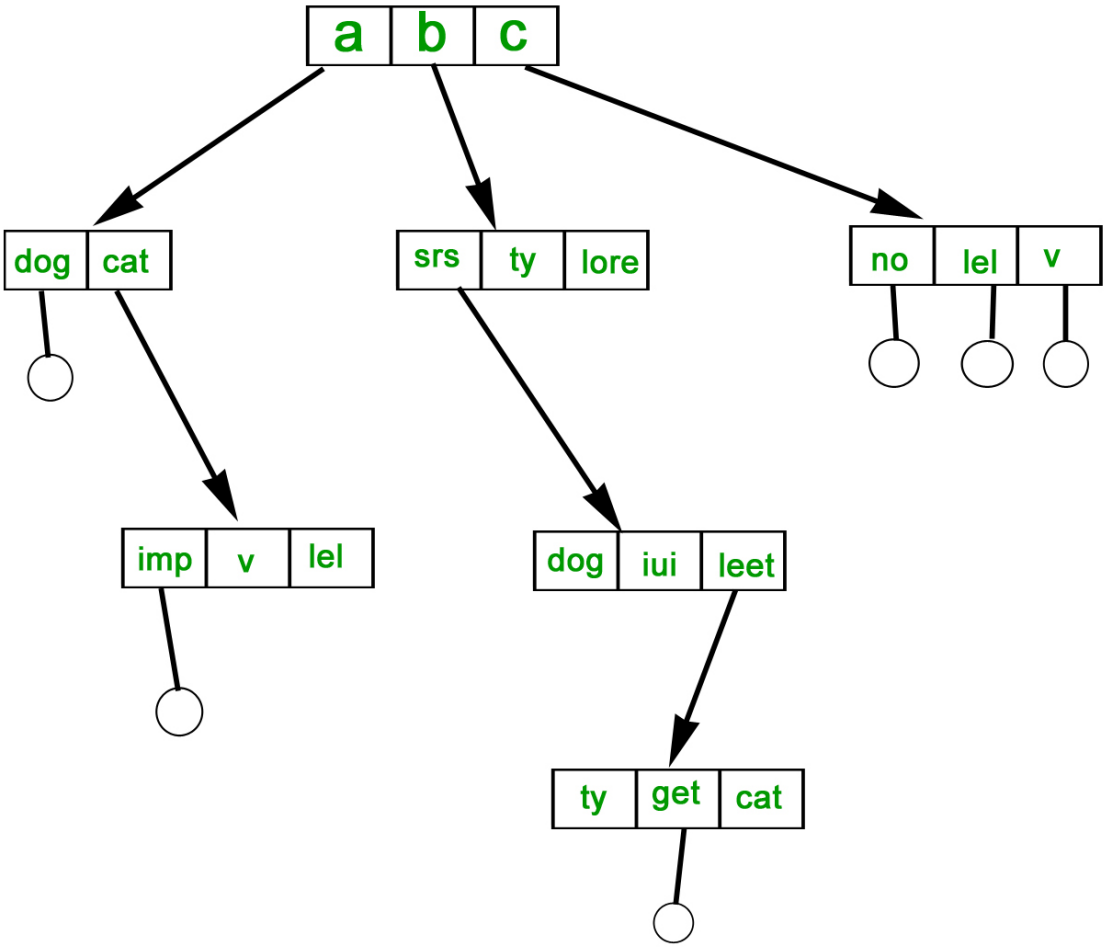
In this separate directories for each user is maintained.

- Path name:Due to two levels there is a path name for every file to locate that file.
- Now,we can have same file name for different user.

- Searching is efficient in this method.

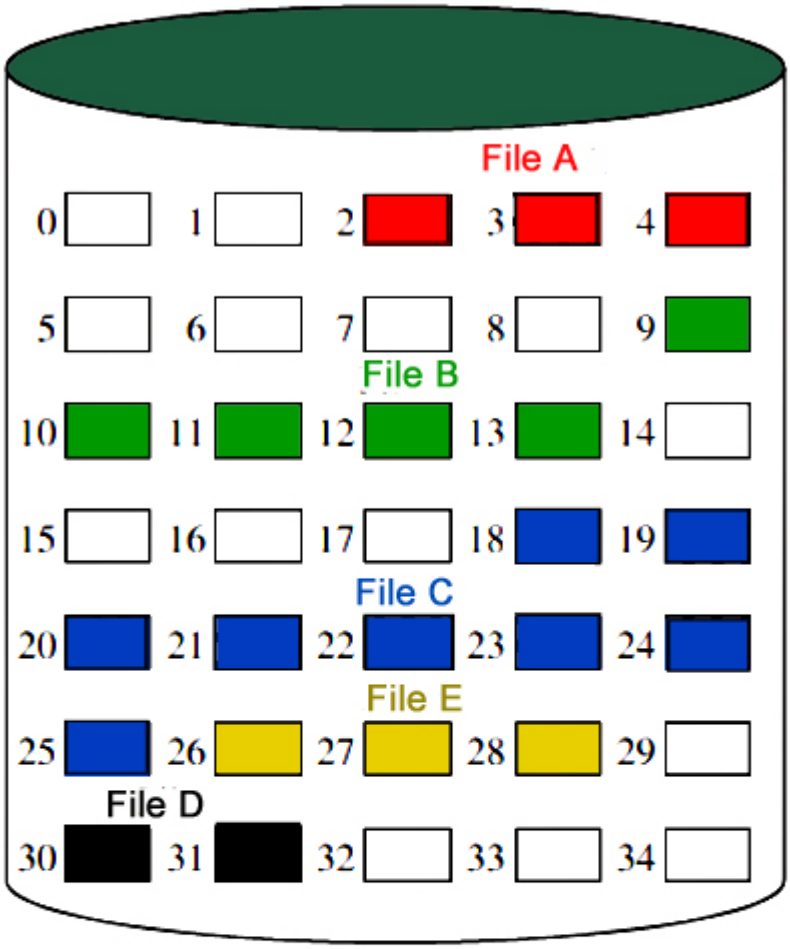


TREE-STRUCTURED DIRECTORY :
Directory is maintained in the form of a tree. Searching is efficient and also there is grouping capability. We have absolute or relative path name for a file.



FILE ALLOCATION METHODS

1. Continuous Allocation: A single continuous set of blocks is allocated to a file at the time of file creation. Thus, this is a pre-allocation strategy, using variable size portions. The file allocation table needs just a single entry for each file, showing the starting block and the length of the file. This method is best from the point of view of the individual sequential file. Multiple blocks can be read in at a time to improve I/O performance for sequential processing. It is also easy to retrieve a single block. For example, if a file starts at block b, and the ith block of the file is wanted, its location on secondary storage is simply b+i-1.



File allocation table

File name	Start block	Length
File A	2	3
File B	9	5
File C	18	8
File D	30	2
File E	26	3

Disadvantage

- External fragmentation will occur, making it difficult to find contiguous blocks of space of sufficient length. Compaction algorithm will be necessary to free up additional space on disk.
- Also, with pre-allocation, it is necessary to declare the size of the file at the time of creation.

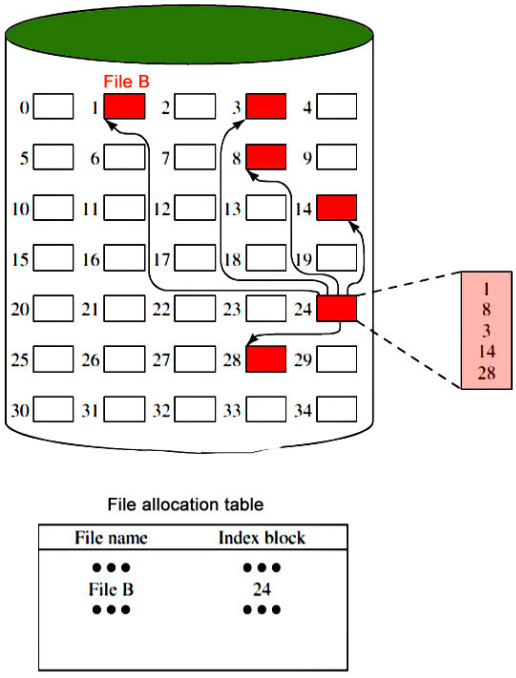
2. Linked Allocation(Non-contiguous allocation) : Allocation is on an individual block basis. Each block contains a pointer to the next block in the chain. Again the file table needs just a single entry for each file, showing the starting block and the length of the file. Although pre-allocation is possible, it is more common simply to allocate blocks as needed. Any free block can be added to the chain. The blocks need not be continuous. Increase in file size is always possible if free disk block is available. There is no external fragmentation because only one block at a time is needed but there can be internal fragmentation but it exists only in the last disk block of file.

Disadvantage:

- Internal fragmentation exists in last disk block of file.
- There is an overhead of maintaining the pointer in every disk block.
- If the pointer of any disk block is lost, the file will be truncated.
- It supports only the sequential access of files.

3. Indexed Allocation:

It addresses many of the problems of contiguous and chained allocation. In this case, the file allocation table contains a separate one-level index for each file: The index has one entry for each block allocated to the file. Allocation may be on the basis of fixed-size blocks or variable-sized blocks. Allocation by blocks eliminates external fragmentation, whereas allocation by variable-size blocks improves locality. This allocation technique supports both sequential and direct access to the file and thus is the most popular form of file allocation.



File allocation table

File name	Index block
•••	•••
File B	24
•••	•••

Disk Free Space Management

Just as the space that is allocated to files must be managed, so the space that is not currently allocated to any file must be managed. To perform any of the file allocation techniques, it is necessary to know what blocks on the disk are available. Thus we need a disk allocation table in addition to a file allocation table.

4. Unix / Linux

The **history of Unix** dates back to the mid-1960s when the Massachusetts Institute of Technology, AT&T Bell Labs, and General Electric were jointly developing an experimental time-sharing operating system called Multics for the GE-645 mainframe. Multics introduced many innovations, but had many problems.
Bell Labs, frustrated by the size and complexity of Multics but not the aims, slowly pulled out of the project. Their last researchers to leave Multics – Ken Thompson, Dennis Ritchie, Doug McIlroy, and Joe Ossanna among others[2] – decided to redo the work on a much smaller scale. In 1979, Dennis Ritchie described their vision for Unix.

- "Ken's new system" (→Unix) (1969)
- UNIX Time-Sharing System v1 (1971)
- UNIX Time-Sharing System v2 (1972)
- UNIX Time-Sharing System v3 (1973)
- UNIX Time-Sharing System v4 (1973)
- UNIX Time-Sharing System v5 (1974)
  - UNSW 01 (1978)
- UNIX Time-Sharing System v6 (1975)
  - Mini-UNIX (1977)
  - PWB/UNIX 1.0 (1977)
    - USG 1.0
      - CB UNIX 1
- UNIX Time-Sharing System v7 (1979)
  - UNIX System III (1981)
- UNIX Time-Sharing System v8 (1985)
- UNIX Time-Sharing System v9 (1986)
- UNIX Time-Sharing System v10 (1989)

Linux Directory Commands

1. pwd Command

The pwd command is used to display the location of the current working directory.

Syntax:

1. pwd

2. mkdir Command

The mkdir command is used to create a new directory under any directory.

Syntax:

mkdir **<directory** name**>**

3. rmdir Command

The rmdir command is used to delete a directory.

Syntax:

rmdir **<directory** name**>**

4. ls Command

The ls command is used to display a list of content of a directory.

Syntax:

ls

5. cd Command

The cd command is used to change the current directory.

cd **<directory** name**>**

Linux File commands

6. touch Command

The touch command is used to create empty files. We can create multiple empty files by executing it once.

Syntax:

touch **<file** name**>**

touch **<file1> <file2>** ....

7. cat Command

The cat command is a multi-purpose utility in the Linux system. It can be used to create a file, display content of the file, copy the content of one file to another file, and more.

Syntax:

cat [OPTION]... [FILE]..

To create a file, execute it as follows:

cat **> <file** name**>**

// Enter file content

Press "**CTRL+ D**" keys to save the file. To display the content of the file, execute it as follows:

cat **<file** name**>**

8. rm Command

The rm command is used to remove a file.

Syntax:

rm <file name>

9. cp Command

The cp command is used to copy a file or directory.

Syntax:

To copy in the same directory:

cp **<existing** file name**> <new** file name**>**

To copy in a different directory:

10. mv Command

The mv command is used to move a file or a directory form one location to another location.

Syntax:

mv **<file** name**> <directory** path**>**

Output:

11. rename Command

The rename command is used to rename files. It is useful for renaming a large group of files.

Syntax:

rename 's/old-name/new-name/' files

For example, to convert all the text files into pdf files, execute the below command:

rename 's/\.txt$/\.pdf/' *.txt

Linux File Content Commands

12. head Command

The head command is used to display the content of a file. It displays the first 10 lines of a file.

Syntax:

head **<file** name**>**

13. tail Command

The tail command is similar to the head command. The difference between both commands is that it displays the last ten lines of the file content. It is useful for reading the error message.

Syntax:

tail **<file** name**>**

14. tac Command

The tac command is the reverse of cat command, as its name specified. It displays the file content in reverse order (from the last line).

Syntax:

tac **<file** name**>**

Output:

15. more command

The more command is quite similar to the cat command, as it is used to display the file content in the same way that the cat command does. The only difference between both commands is that, in case of larger files, the more command displays screenful output at a time.

In more command, the following keys are used to scroll the page:

ENTER key: To scroll down page by line.

Space bar: To move to the next page.

b key: To move to the previous page.

/ key: To search the string.

Syntax:

more **<file** name**>**

16. less Command

The less command is similar to the more command. It also includes some extra features such as 'adjustment in width and height of the terminal.' Comparatively, the more command cuts the output in the width of the terminal.

Syntax:

less **<file** name**>**

Linux User Commands

17. su Command

The su command provides administrative access to another user. In other words, it allows access of the Linux shell to another user.

Syntax:

su **<user** name**>**

18. id Command

The id command is used to display the user ID (UID) and group ID (GID).

Syntax:

id

19. useradd Command

The useradd command is used to add or remove a user on a Linux server.

Syntax:

useradd username

20. passwd Command

The passwd command is used to create and change the password for a user.

Syntax:

passwd **<username>**

21. groupadd Command

The groupadd command is used to create a user group.

Syntax:

groupadd **<group** name**>**

Linux Filter Commands

22. cat Command

The cat command is also used as a filter. To filter a file, it is used inside pipes.

Syntax:

cat **<fileName>** | cat or tac | cat or tac |. . .

23. cut Command

The cut command is used to select a specific column of a file. The '-d' option is used as a delimiter, and it can be a space (' '), a slash (/), a hyphen (-), or anything else. And, the '-f' option is used to specify a column number.

Syntax:

cut -d(delimiter) -f(columnNumber) **<fileName>**

24. grep Command

The grep is the most powerful and used filter in a Linux system. The 'grep' stands for "**global regular expression print**." It is useful for searching the content from a file. Generally, it is used with the pipe.

Syntax:

command | grep **<searchWord>**

25. comm Command

The 'comm' command is used to compare two files or streams. By default, it displays three columns, first displays non-matching items of the first file, second indicates the non-matching item of the second file, and the third column displays the matching items of both files.

Syntax:

1. comm **<file1> <file2>**

26. sed command

The sed command is also known as **stream editor**. It is used to edit files using a regular expression. It does not permanently edit files; instead, the edited content remains only on display. It does not affect the actual file.

Syntax:

command | sed 's/**<oldWord>**/**<newWord>**/'

27. tee command

The tee command is quite similar to the cat command. The only difference between both filters is that it puts standard input on standard output and also write them into a file.

Syntax:

cat **<fileName>** | tee **<newFile>** | cat or tac |.....

28. tr Command

The tr command is used to translate the file content like from lower case to upper case.

Syntax:

command | tr **<**'old'**> <**'new'**>**

29. uniq Command

The uniq command is used to form a sorted list in which every word will occur only once.

Syntax:

command **<fileName>** | uniq

30. wc Command

The wc command is used to count the lines, words, and characters in a file.

Syntax:

wc **<file** name**>**

31. od Command

The od command is used to display the content of a file in different s, such as hexadecimal, octal, and ASCII characters.

Syntax:

1. od -b **<fileName>** // Octal format
2. od -t x1 **<fileName>** // Hexa decimal format
3. od -c **<fileName>** // ASCII character format

32. sort Command

The sort command is used to sort files in alphabetical order.

Syntax:

sort **<file** name**>**

33. gzip Command

The gzip command is used to truncate the file size. It is a compressing tool. It replaces the original file by the compressed file having '.gz' extension.

Syntax:

gzip **<file1> <file2> <file3>**...

34. gunzip Command

The gunzip command is used to decompress a file. It is a reverse operation of gzip command.

Syntax:

gunzip **<file1> <file2> <file3>**. .

Linux Utility Commands

35. find Command

The find command is used to find a particular file within a directory. It also supports various options to find a file such as byname, by type, by date, and more.

The following symbols are used after the find command:

(.) : For current directory name

(/) : For root

Syntax:

find . -name "*.pdf"

36. locate Command

The locate command is used to search a file by file name. It is quite similar to find command; the difference is that it is a background process. It searches the file in the database, whereas the find command searches in the file system. It is faster than the find command. To find the file with the locates command, keep your database updated.

Syntax:

locate **<file** name**>**

37. date Command

The date command is used to display date, time, time zone, and more.

Syntax:

date

38. cal Command

The cal command is used to display the current month's calendar with the current date highlighted.

Syntax:

cal**<**

39. sleep Command

The sleep command is used to hold the terminal by the specified amount of time. By default, it takes time in seconds.

Syntax:

sleep **<time>**

40. time Command

The time command is used to display the time to execute a command.

Syntax:

time

41. zcat Command

The zcat command is used to display the compressed files.

Syntax:

zcat **<file** name**>**

42. df Command

The df command is used to display the disk space used in the file system. It displays the output as in the number of used blocks, available blocks, and the mounted directory.

Syntax:

df

43. mount Command

The mount command is used to connect an external device file system to the system's file system.

Syntax:

mount -t type **<device> <directory>**

44. exit Command

Linux exit command is used to exit from the current shell. It takes a parameter as a number and exits the shell with a return of status number.

Syntax:

exit

After pressing the ENTER key, it will exit the terminal.

45. clear Command

Linux **clear** command is used to clear the terminal screen.

Syntax:

clear

After pressing the ENTER key, it will clear the terminal screen.

Linux Networking Commands

46. ip Command

Linux ip command is an updated version of the ipconfig command. It is used to assign an IP address, initialize an interface, disable an interface.

Syntax:

ip a or ip addr

47. ssh Command

Linux ssh command is used to create a remote connection through the ssh protocol.

Syntax:

ssh user_name@host(IP/Domain_name)**</p>**

48. ping Command

The ping command is used to check the connectivity between two nodes, that is whether the server is connected. It is a short form of "Packet Internet Groper."

Syntax:

ping **<destination>**

49. host Command

The host command is used to display the IP address for a given domain name and vice versa. It performs the DNS lookups for the DNS Query.

Syntax:

host **<domain** name**>** or **<ip** address**>**

A desktop computer system typically runs a user-friendly operating system and desktop applications to facilitate desktop-oriented tasks. In contrast, a server manages all network resources. Servers are often dedicated (meaning it performs no other task besides server tasks).

What is the difference between server OS and desktop OS?

It is an operating system that operates within desktop. It is used to obtain services from a server. It run on the client devices like laptop, computer and is very simple operating system.

Difference between Server OS and Client OS :

Server Operating System	**Client Operating System**
It can serve multiple client at a time.	It serves a single user at a time.

What is a server operating system?

A server operating system (OS) is a type of operating system that is designed to be installed and used on a server computer. It is an advanced version of an operating system, having features and capabilities required within a client-server architecture or similar enterprise computing environment.

Source: https://ostoday.org/other/what-is-difference-between-desktop-and-server-based-operating-system.html

Difference between Server OS and Client OS

- Last Updated : 02 Jul, 2020

1. Client OS :

It is an operating system that operates within desktop. It is used to obtain services from a server. It run on the client devices like laptop, computer and is very simple operating system.

2. Server OS :

It is an operating system that is designed to be used on server. It is used to provide services to multiple client. It can serve multiple client at a time and is very advanced operating system.

Difference between Server OS and Client OS :

Server Operating System	Client Operating System
It can be used to provide services to multiple client.	It can obtain services from a server.
It can serve multiple client at a time.	It serves a single user at a time.
It is complex operating system.	It is simple operating system.
It runs on the server.	It runs on the client devices like laptop, computer etc.
It is an operating system that is designed to be used on server.	It is an operating system that operates within desktop.
It provides more security.	It provides less security.

Server Operating System	Client Operating System
It has greater processing power.	It has less processing power.
It is more stable.	It is less stable.
It is highly efficient.	It is less efficient.
Examples: Red Hat, Linux.	Examples: Windows, Android.

Key Technical Differences Between a Windows Server and a Windows Desktop

Here are the key differences between a Windows server and a Windows desktop. Keep these in mind when choosing which system is best for your Alike installation. It's also important to keep this in mind should you ever need to recover, repair, or migrate your Alike installation.

1. Windows Server Supports More Memory

One of the main differences between a Windows server and desktop is the amount of memory each can support. A desktop running Windows 10 Enterprise has a 4 GB memory limit on an X86 and a 2TB limit on an X64. These numbers are greatly increased depending on the Windows server version.

2. Windows Server Uses CPUs More Efficiently

In general, a server OS is more efficient at using its hardware than a desktop OS, especially a CPU; therefore, if you install Alike on a server OS, you are taking full advantage of the hardware installed on your server, which also allows Alike to offer optimal performance.

3. More Network Connections Allowed on a Windows Server

With a desktop version of Windows, network connections are limited to 10-20. A Windows server is not locked down to just 20 network connections; therefore, a Windows server can support well beyond 20 network connections based on your hardware's capability.

Operating System	Network Connections
WinXP	5-10
Vista	2-25
Windows 7	10-20

4. The Server OS is Configured for Background Tasks

By default, server editions of the Windows OS are configured to give priority to background tasks and services, whereas the desktop editions focus on foreground. While priority can be changed, the desktop edition's management does not yield the same performance results.

How Alike Performs on a Server OS vs. a Desktop OS

A popular request for our tech support team is how to tune Alike for the highest performance and scalability that the customer's environment can handle. More often than not, customers experiencing performance issues are also running the Alike server on a desktop OS.

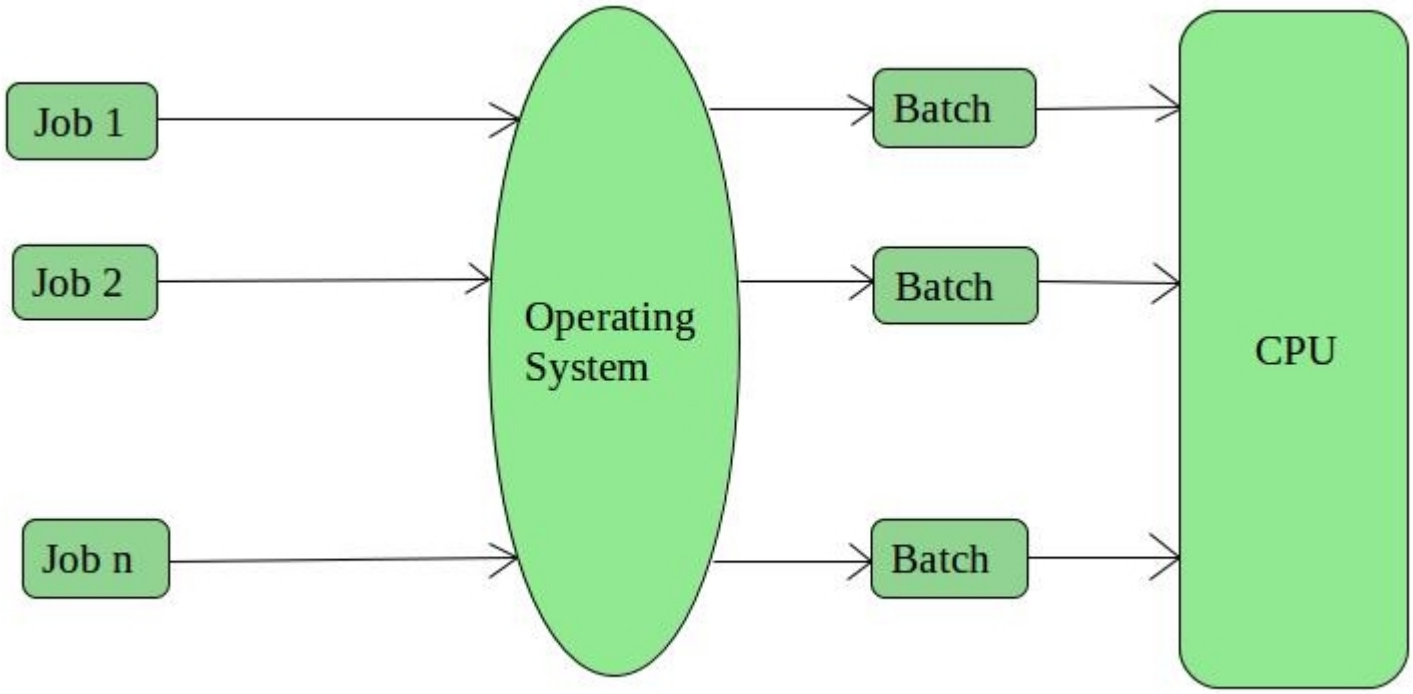
You may be wondering why running Alike on a desktop OS would cause any problems, especially since the Alike server itself doesn't do much work (the remote backup "agents," ABD and Q-Hybrid, do all the processing).

While this is true, installing Alike on a server grade OS is still preferable in most cases, and often necessary in medium to large environments.
Look at it this way: installing Alike on a Windows Desktop is like putting a jet engine in a go-kart. People have done it (no really, they have—check out this guy's rig), but while this design will let you go 60 mph in a go-kart, you will only get so far before the kart falls apart trying to keep up with a Boeing engine.
A more apropos way to think about this is to look at some of the limitations Microsoft has placed on its desktop operating systems. Understanding the key differences between a Windows server and a Windows desktop can help illustrate why—in most environments—you will see greater performance if you install Alike on a server OS.

Types of Operating Systems: Some widely used operating systems are as follows-

1. Batch Operating System –

This type of operating system does not interact with the computer directly. There is an operator which takes similar jobs having the same requirement and group them into batches. It is the responsibility of the operator to sort jobs with similar needs.



Advantages of Batch Operating System:

- It is very difficult to guess or know the time required for any job to complete. Processors of the batch systems know how long the job would be when it is in queue
- Multiple users can share the batch systems
- The idle time for the batch system is very less
- It is easy to manage large work repeatedly in batch systems

Disadvantages of Batch Operating System:

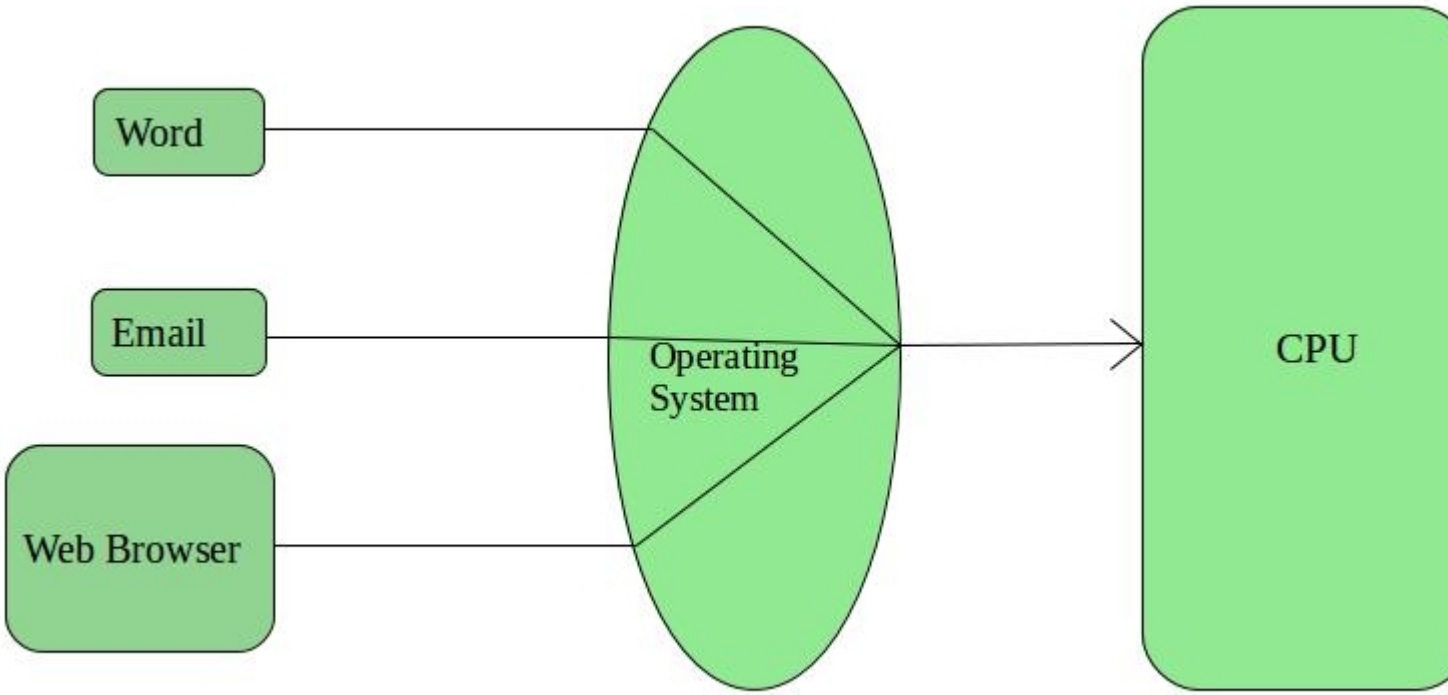
- The computer operators should be well known with batch systems
- Batch systems are hard to debug

- It is sometimes costly
- The other jobs will have to wait for an unknown time if any job fails

Examples of Batch based Operating System: Payroll System, Bank Statements, etc.

2. Time-Sharing Operating Systems –

Each task is given some time to execute so that all the tasks work smoothly. Each user gets the time of CPU as they use a single system. These systems are also known as Multitasking Systems. The task can be from a single user or different users also. The time that each task gets to execute is called quantum. After this time interval is over OS switches over to the next task.



Advantages of Time-Sharing OS:

- Each task gets an equal opportunity
- Fewer chances of duplication of software
- CPU idle time can be reduced

Disadvantages of Time-Sharing OS:

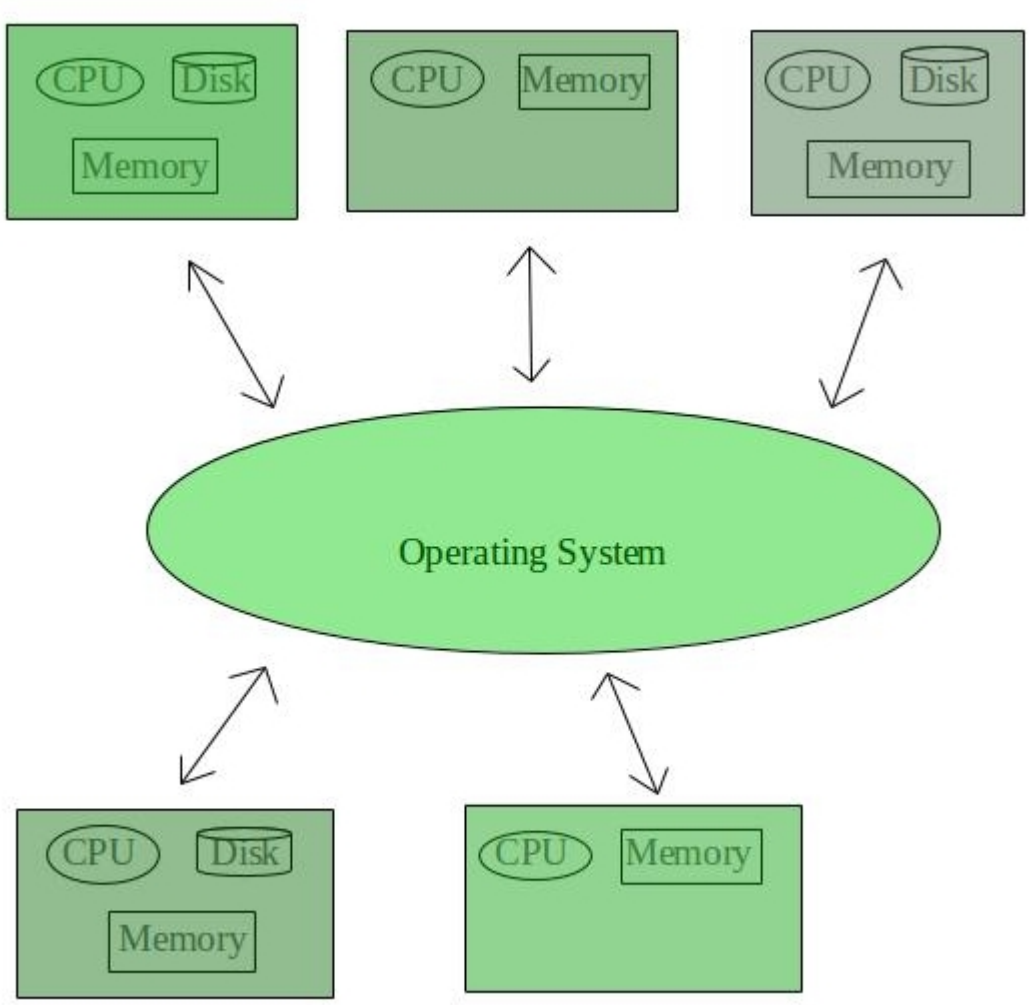
- Reliability problem
- One must have to take care of the security and integrity of user programs and data
- Data communication problem

Examples of Time-Sharing OSs are: Multics, Unix, etc.

3. Distributed Operating System –

These types of the operating system is a recent advancement in the world of computer technology and are being widely accepted all over the world and, that too, with a great pace. Various autonomous interconnected computers communicate with each other using a shared communication network. Independent systems possess their own memory unit and CPU. These are referred to as **loosely coupled systems** or distributed systems. These system's processors differ in size and function. The major benefit of working with these types of the operating system is that it is always possible that one user can access the files or software which are not actually present on his system but some other system connected within this

network i.e., remote access is enabled within the devices connected in that network.



Advantages of Distributed Operating System:

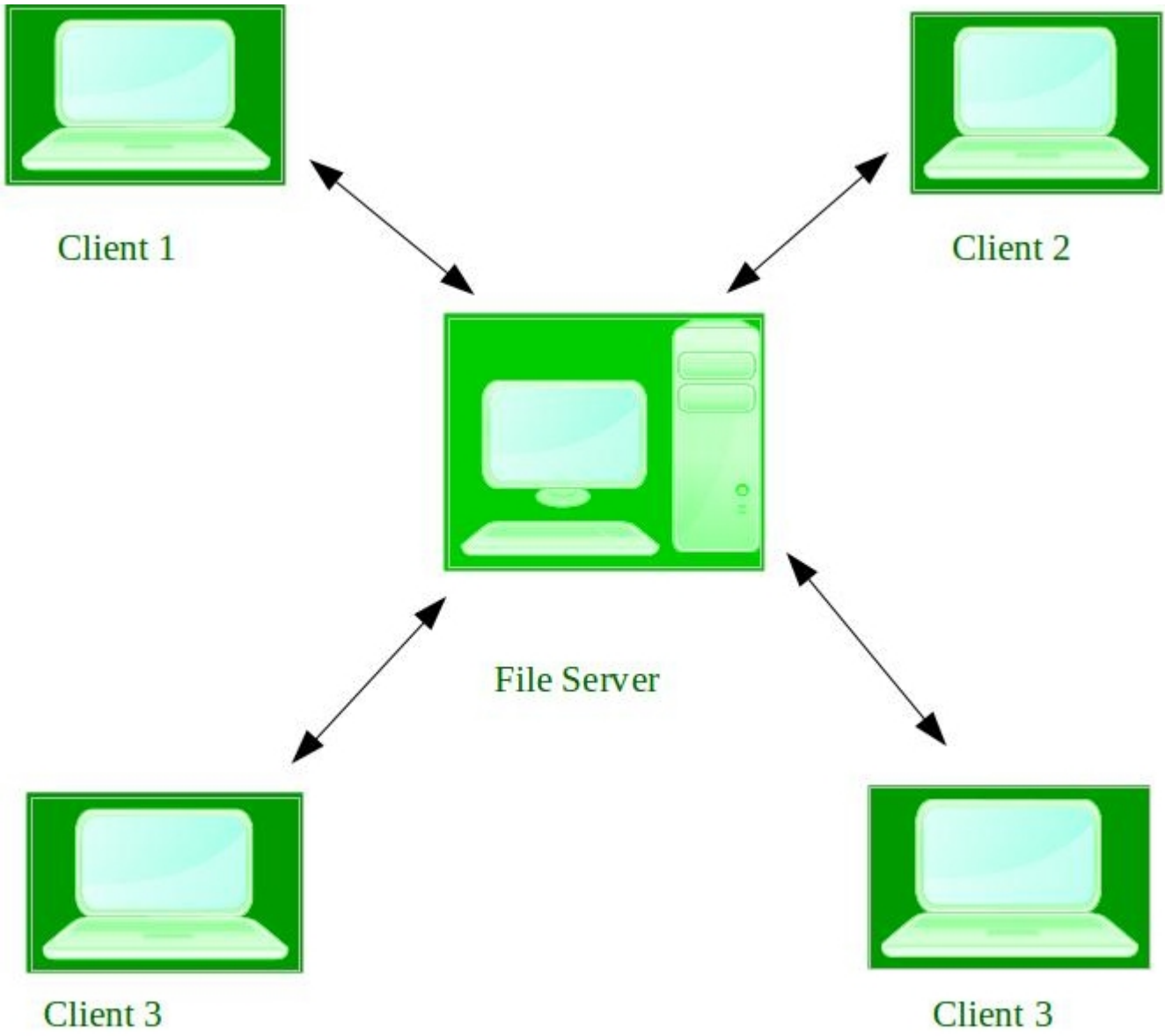
- Failure of one will not affect the other network communication, as all systems are independent from each other
- Electronic mail increases the data exchange speed
- Since resources are being shared, computation is highly fast and durable
- Load on host computer reduces
- These systems are easily scalable as many systems can be easily added to the network
- Delay in data processing reduces

Disadvantages of Distributed Operating System:

- Failure of the main network will stop the entire communication
- To establish distributed systems the language which is used are not well defined yet
- These types of systems are not readily available as they are very expensive. Not only that the underlying software is highly complex and not understood well yet

Examples of Distributed Operating System are- LOCUS, etc.

4. Network Operating System –
These systems run on a server and provide the capability to manage data, users, groups, security, applications, and other networking functions. These types of operating systems allow shared access of files, printers, security, applications, and other networking functions over a small private network. One more important aspect of Network Operating Systems is that all the users are well aware of the underlying configuration, of all other users within the network, their individual connections, etc. and that's why these computers are popularly known as **tightly coupled systems**.



Advantages of Network Operating System:
- Highly stable centralized servers
- Security concerns are handled through servers
- New technologies and hardware up-gradation are easily integrated into the system
- Server access is possible remotely from different locations and types of systems

Disadvantages of Network Operating System:
- Servers are costly
- User has to depend on a central location for most operations
- Maintenance and updates are required regularly

Examples of Network Operating System are: Microsoft Windows Server 2003, Microsoft Windows Server 2008, UNIX, Linux, Mac OS X, Novell NetWare, and BSD, etc.

Windows OS Editions

As in the past, we will offer different Windows editions that are tailored for various device families and uses. These different editions address specific needs of our various customers, from consumers to small businesses to the largest enterprises.

Windows 10 Home is the consumer-focused desktop edition. It offers a familiar and personal experience for PCs, tablets and 2-in-1s. Windows 10 Home will help people do great things, both big and small. With it, they will be more productive and have more fun thanks to a long list of new innovations: Cortana, the world's most personal digital assistant; the new Microsoft Edge web browser; Continuum tablet mode for touch-capable devices; Windows Hello face-recognition, iris and fingerprint login; and right out of the box, a broad range of universal Windows apps like Photos, Maps, Mail, Calendar, Music and Video*.

We are also bringing the Xbox gaming experience to Windows 10, giving games and gamers access to the Xbox Live gaming community, enabling the capture and share of gameplay and giving Xbox One owners the ability to play their Xbox One games from any Windows 10 PC in their home.

Windows 10 Mobile is designed to deliver the best user experience on smaller, mobile, touch-centric devices like smartphones and small tablets. It boasts the same, new universal Windows apps that are included in Windows 10 Home, as well as the new touch-optimized version of Office. Windows 10 Mobile offers great productivity, security and management capabilities for customers who use their personal devices at work. In addition, Windows 10 Mobile will enable some new devices to take advantage of Continuum for phone, so people can use their phone like a PC when connected to a larger screen.

Windows 10 Pro is a desktop edition for PCs, tablets and 2-in-1s. Building upon both the familiar and innovative features of Windows 10 Home, it has many extra features to meet the diverse needs of small businesses. Windows 10 Pro helps to effectively and efficiently manage their devices and apps, protect their sensitive business data, support remote and mobile productivity scenarios and take advantage of cloud technologies. Windows 10 Pro devices are a great choice for organizations supporting Choose Your Own Device (CYOD) programs and prosumer customers. Windows 10 Pro also lets customers take advantage of the new Windows Update for Business, which will reduce management costs, provide controls over update deployment, offer quicker access to security updates and provide access to the latest innovation from Microsoft on an ongoing basis.

Windows 10 Enterprise builds on Windows 10 Pro, adding advanced features designed to meet the demands of medium and large sized organizations. It provides advanced capabilities to help protect against the ever-growing range of modern security threats targeted at devices, identities, applications and sensitive company information. Windows 10 Enterprise also supports the broadest range of options for operating system deployment and comprehensive device and app management. It will be available to our Volume Licensing customers, so they can take advantage of the latest innovation and security updates on an ongoing basis. At the same time, they will be able to choose the pace at which they adopt new technology, including the option to use the new Windows Update for Business. With Windows 10, Enterprise customers will also have access to the Long Term Servicing Branch as a deployment option for their mission critical devices and environments. And as with prior versions of Windows, Active Software Assurance customers in Volume Licensing can upgrade to Windows 10 Enterprise as part of their existing Software Assurance benefits.

Windows 10 Education builds on Windows 10 Enterprise, and is designed to meet the needs of schools – staff, administrators, teachers and students. This edition will be available through academic Volume Licensing, and

there will be paths for schools and students using Windows 10 Home and Windows 10 Pro devices to upgrade to Windows 10 Education.

Windows 10 Mobile Enterprise is designed to deliver the best customer experience to business customers on smartphones and small tablets. It will be available to our Volume Licensing customers. It offers the great productivity, security and mobile device management capabilities that Windows 10 Mobile provides, and adds flexible ways for businesses to manage updates. In addition, Windows 10 Mobile Enterprise will incorporate the latest security and innovation features as soon as they are available.
There will also be versions of Windows 10 Enterprise and Windows 10 Mobile Enterprise for industry devices like ATMs, retail point of sale, handheld terminals and industrial robotics and **Windows 10 IoT Core** for small footprint, low cost devices like gateways.

5. Network Basics

Introduction to Networking - Concepts & Features

What is a computer network?

A computer network comprises two or more computers that are connected—either by cables (wired) or WiFi (wireless)—with the purpose of transmitting, exchanging, or sharing data and resources. You build a computer network using hardware (e.g., routers, switches, access points, and cables) and software (e.g., operating systems or business applications).

Geographic location often defines a computer network. For example, a LAN (local area network) connects computers in a defined physical space, like an office building, whereas a WAN (wide area network) can connect

computers across continents. The internet is the largest example of a WAN, connecting billions of computers worldwide.

You can further define a computer network by the protocols it uses to communicate, the physical arrangement of its components, how it controls traffic, and its purpose.

Computer networks enable communication for every business, entertainment, and research purpose. The internet, online search, email, audio and video sharing, online commerce, live-streaming, and social networks all exist because of computer networks.

Computer network types

As networking needs evolved, so did the computer network types that serve those needs. Here are the most common and widely used computer network types:

- **LAN (local area network):** A LAN connects computers over a relatively short distance, allowing them to share data, files, and resources. For example, a LAN may connect all the computers in an office building, school, or hospital. Typically, LANs are privately owned and managed.
- **WLAN (wireless local area network):** A WLAN is just like a LAN but connections between devices on the network are made wirelessly.
- **WAN (wide area network):** As the name implies, a WAN connects computers over a wide area, such as from region to region or even continent to continent. The internet is the largest WAN, connecting billions of computers worldwide. You will typically see collective or distributed ownership models for WAN management.
- **MAN (metropolitan area network):** MANs are typically larger than LANs but smaller than WANs. Cities and government entities typically own and manage MANs.
- **PAN (personal area network):** A PAN serves one person. For example, if you have an iPhone and a Mac, it's very likely you've set up a PAN that shares and syncs content—text messages, emails, photos, and more—across both devices.
- **SAN (storage area network):** A SAN is a specialized network that provides access to block-level storage—shared network or cloud storage that, to the user, looks and works like a storage drive that's physically attached to a computer. (For more information on how a SAN works with block storage, see Block Storage: A Complete Guide.)
- **CAN (campus area network):** A CAN is also known as a corporate area network. A CAN is larger than a LAN but smaller than a WAN. CANs serve sites such as colleges, universities, and business campuses.
- **VPN (virtual private network):** A VPN is a secure, point-to-point connection between two network end points (see 'Nodes' below). A VPN establishes an encrypted channel that keeps a user's identity and access credentials, as well as any data transferred, inaccessible to hackers.

Important terms and concepts

The following are some common terms to know when discussing computer networking:

- **IP address**: An IP address is a unique number assigned to every device connected to a network that uses the Internet Protocol for communication. Each IP address identifies the device's host network and the location of the device on the host network. When one device sends data to another, the data includes a 'header' that includes the IP address of the sending device and the IP address of the destination device.
- **Nodes**: A node is a connection point inside a network that can receive, send, create, or store data. Each node requires you to provide some form of identification to receive access, like an IP address. A few examples of nodes include computers, printers, modems, bridges, and switches. A node is essentially any network device that can recognize, process, and transmit information to any other network node.
- **Routers**: A router is a physical or virtual device that sends information contained in data packets between networks. Routers analyze data within the packets to determine the best way for the information to reach its ultimate destination. Routers forward data packets until they reach their destination node.

- **Switches**: A switch is a device that connects other devices and manages node-to-node communication within a network, ensuring data packets reach their ultimate destination. While a router sends information between networks, a switch sends information between nodes in a single network. When discussing computer networks, 'switching' refers to how data is transferred between devices in a network. The three main types of switching are as follows:
    - *Circuit switching*, which establishes a dedicated communication path between nodes in a network. This dedicated path assures the full bandwidth is available during the transmission, meaning no other traffic can travel along that path.
    - *Packet switching* involves breaking down data into independent components called packets which, because of their small size, make fewer demands on the network. The packets travel through the network to their end destination.
    - *Message switching* sends a message in its entirety from the source node, traveling from switch to switch until it reaches its destination node.
- **Ports**: A port identifies a specific connection between network devices. Each port is identified by a number. If you think of an IP address as comparable to the address of a hotel, then ports are the suites or room numbers within that hotel. Computers use port numbers to determine which application, service, or process should receive specific messages.
- **Network cable types**: The most common network cable types are Ethernet twisted pair, coaxial, and fiber optic. The choice of cable type depends on the size of the network, the arrangement of network elements, and the physical distance between devices.

Examples of computer networks

The wired or wireless connection of two or more computers for the purpose of sharing data and resources form a computer network. Today, nearly every digital device belongs to a computer network.

In an office setting, you and your colleagues may share access to a printer or to a group messaging system. The computing network that allows this is likely a LAN or local area network that permits your department to share resources.

A city government might manage a city-wide network of surveillance cameras that monitor traffic flow and incidents. This network would be part of a MAN or metropolitan area network that allows city emergency personnel to respond to traffic accidents, advise drivers of alternate travel routes, and even send traffic tickets to drivers who run red lights.

The Weather Company worked to create a peer-to-peer mesh network that allows mobile devices to communicate directly with other mobile devices without requiring WiFi or cellular connectivity. The Mesh Network Alerts project allows the delivery of life-saving weather information to billions of people, even without an internet connection.

Computer networks and the internet

The internet is actually a network of networks that connects billions of digital devices worldwide. Standard protocols allow communication between these devices. Those protocols include hypertext transfer protocol (the 'http' in front of all website addresses). Internet protocol (or IP addresses) are the unique identifying numbers required of every device that accesses the internet. IP addresses are comparable to your mailing address, providing unique location information so that information can be delivered correctly.

Internet Service Providers (ISPs) and Network Service Providers (NSPs) provide the infrastructure that allows the transmission of packets of data or information over the internet. Every bit of information sent over the internet doesn't go to every device connected to the internet. It's the combination of protocols and infrastructure that tells information exactly where to go.

How do they work?

Computer networks connect nodes like computers, routers, and switches using cables, fiber optics, or wireless signals. These connections allow devices in a network to communicate and share information and resources.

Networks follow protocols, which define how communications are sent and received. These protocols allow devices to communicate. Each device on a network uses an Internet Protocol or IP address, a string of numbers that uniquely identifies a device and allows other devices to recognize it.

Routers are virtual or physical devices that facilitate communications between different networks. Routers analyze information to determine the best way for data to reach its ultimate destination. Switches connect devices and manage node-to-node communication inside a network, ensuring that bundles of information traveling across the network reach their ultimate destination.

Architecture

Computer network architecture defines the physical and logical framework of a computer network. It outlines how computers are organized in the network and what tasks are assigned to those computers. Network architecture components include hardware, software, transmission media (wired or wireless), network topology, and communications protocols.

Main types of network architecture

There are two types of network architecture: *peer-to-peer (P2P)* and *client/server*. In P2P architecture, two or more computers are connected as "peers," meaning they have equal power and privileges on the network. A P2P network does not require a central server for coordination. Instead, each computer on the network acts as both a client (a computer that needs to access a service) and a server (a computer that serves the needs of the client accessing a service). Each peer makes some of its resources available to the network, sharing storage, memory, bandwidth, and processing power.

In a client/server network, a central server or group of servers manage resources and deliver services to client devices in the network. The clients in the network communicate with other clients through the server. Unlike the P2P model, clients in a client/server architecture don't share their resources. This architecture type is sometimes called a tiered model because it's designed with multiple levels or tiers.

Network topology

Network topology refers to how the nodes and links in a network are arranged. A network node is a device that can send, receive, store, or forward data. A network link connects nodes and may be either cabled or wireless links.

Understanding topology types provides the basis for building a successful network. There are a number of topologies but the most common are bus, ring, star, and mesh:

- A *bus network topology* is when every network node is directly connected to a main cable.
- In a *ring topology*, nodes are connected in a loop, so each device has exactly two neighbors. Adjacent pairs are connected directly; non-adjacent pairs are connected indirectly through multiple nodes.
- In a *star network topology*, all nodes are connected to a single, central hub and each node is indirectly connected through that hub.
- A *mesh topology* is defined by overlapping connections between nodes. You can create a full mesh topology, where every node in the network is connected to every other node. You can also create partial mesh topology in which only some nodes are connected to each other and some are connected to the nodes with which they exchange the most data. Full mesh topology can be expensive and time-consuming to execute, which is why it's often reserved for networks that require high redundancy. Partial mesh provides less redundancy but is more cost effective and simpler to execute.

Security

Computer network security protects the integrity of information contained by a network and controls who access that information. Network security policies balance the need to provide service to users with the need to control access to information.

There are many entry points to a network. These entry points include the hardware and software that comprise the network itself as well as the devices used to access the network, like computers, smartphones, and tablets. Because of these entry points, network security requires using several defense methods. Defenses may include firewalls—devices that monitor network traffic and prevent access to parts of the network based on security rules.

Processes for authenticating users with user IDs and passwords provide another layer of security. Security includes isolating network data so that proprietary or personal information is harder to access than less critical information. Other network security measures include ensuring hardware and software updates and patches

are performed regularly, educating network users about their role in security processes, and staying aware of external threats executed by hackers and other malicious actors. Network threats constantly evolve, which makes network security a never-ending process.

The use of public cloud also requires updates to security procedures to ensure continued safety and access. A secure cloud demands a secure underlying network.

Read about the top five considerations (PDF, 298 KB) for securing the public cloud.

Mesh networks

As noted above, a mesh network is a topology type in which the nodes of a computer network connect to as many other nodes as possible. In this topology, nodes cooperate to efficiently route data to its destination. This topology provides greater fault tolerance because if one node fails, there are many other nodes that can transmit data. Mesh networks self-configure and self-organize, searching for the fastest, most reliable path on which to send information.

Type of mesh networks

There are two types of mesh networks—full mesh and partial mesh:

- In a *full mesh topology*, every network node connects to every other network node, providing the highest level of fault tolerance. However, it costs more to execute. In a partial mesh topology, only some nodes connect, typically those that exchange data most frequently.
- A *wireless mesh network* may consist of tens to hundreds of nodes. This type of network connects to users over access points spread across a large area.

Load balancers and networks

Load balancers efficiently distribute tasks, workloads, and network traffic across available servers. Think of load balancers like air traffic control at an airport. The load balancer observes all traffic coming into a network and directs it toward the router or server best equipped to manage it. The objectives of load balancing are to avoid resource overload, optimize available resources, improve response times, and maximize throughput.

For a complete overview of load balancers, see Load Balancing: A Complete Guide.

Content delivery networks

A content delivery network (CDN) is a distributed server network that delivers temporarily stored, or cached, copies of website content to users based on the user's geographic location. A CDN stores this content in distributed locations and serves it to users as a way to reduce the distance between your website visitors and your website server. Having cached content closer to your end users allows you to serve content faster and helps websites better reach a global audience. CDNs protect against traffic surges, reduce latency, decrease bandwidth consumption, accelerate load times, and lessen the impact of hacks and attacks by introducing a layer between the end user and your website infrastructure.

Live-streaming media, on-demand media, gaming companies, application creators, e-commerce sites—as digital consumption increases, more content owners turn to CDNs to better serve content consumers.

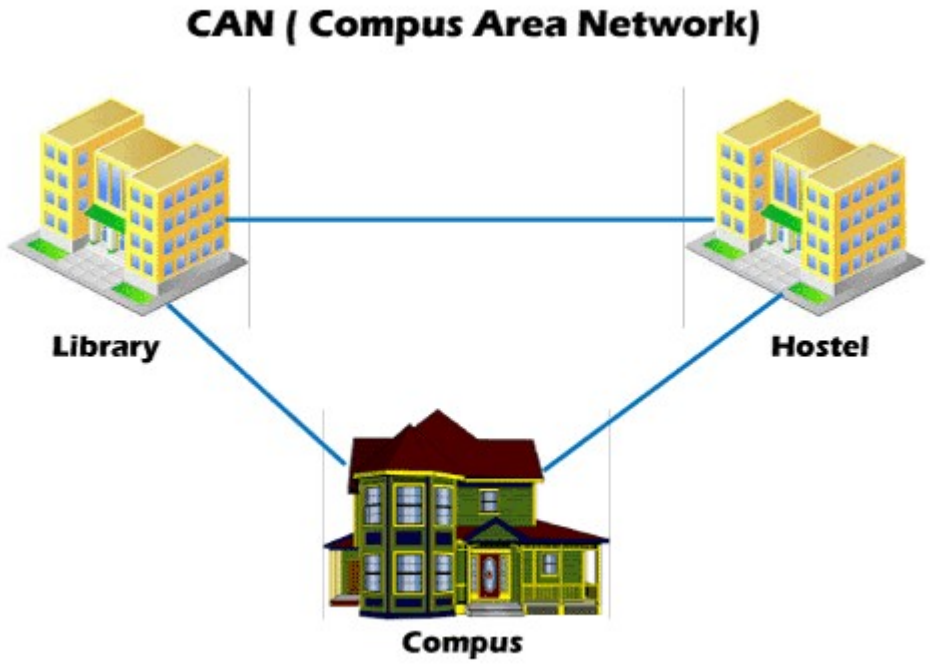
1. LAN

LAN or Local Area Network is a group of devices connecting the computers and other devices such as **switches, servers, printers,** etc., over a short distance such as office, home. The commonly used LAN is **Ethernet LAN.** This network is used as it allows the user to transfer or share data, files, and resources.

2. WLAN

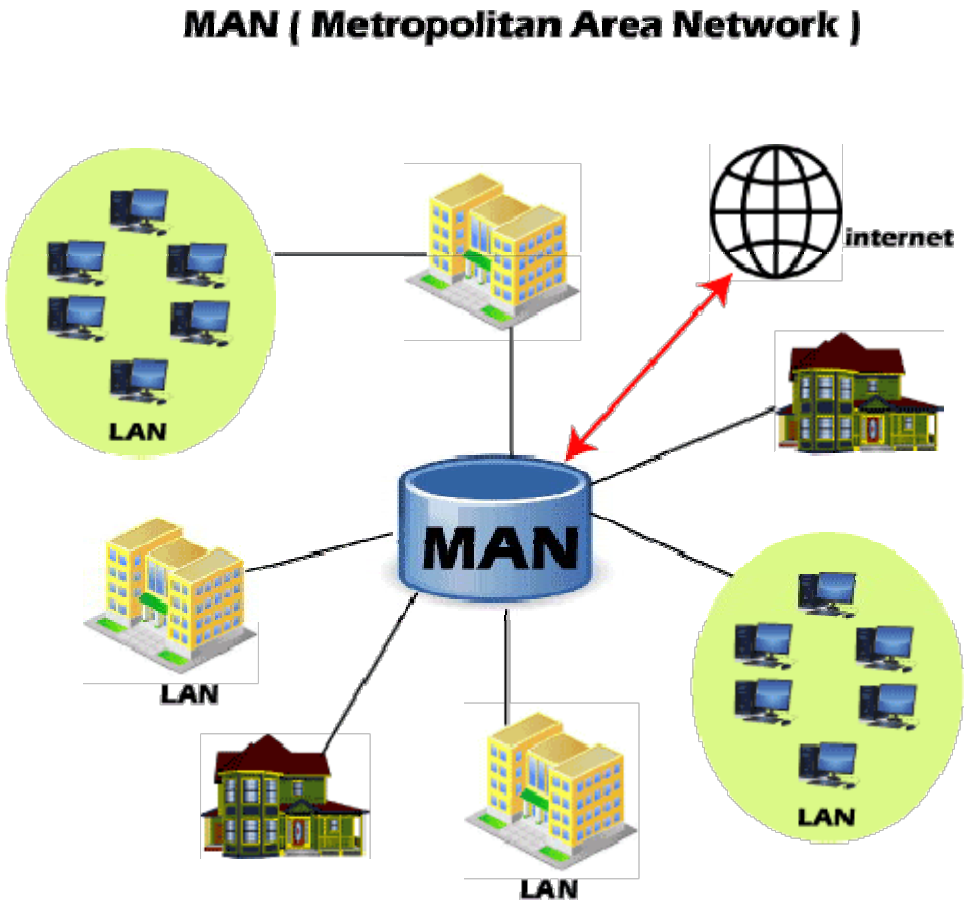
WLAN or Wireless local area network is similar to LAN with the difference that it uses wireless communication between devices instead of wired connections. WLAN typically involves a **Wi-Fi router or wireless access point for devices,** unlike **smartphones, laptops, desktops,** etc.

3. CAN



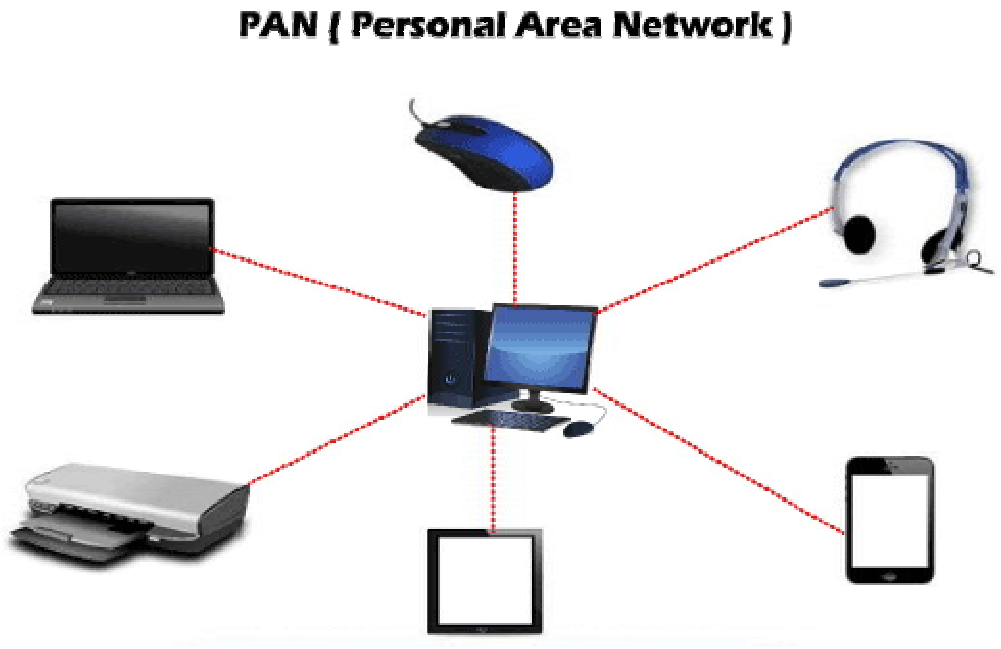
CAN or Campus Area Network is a closed corporate communication network. A CAN is a mobile network that may contain a private or public part. CANs are widely used **colleges, academies, and corporate sites.**

4. MAN



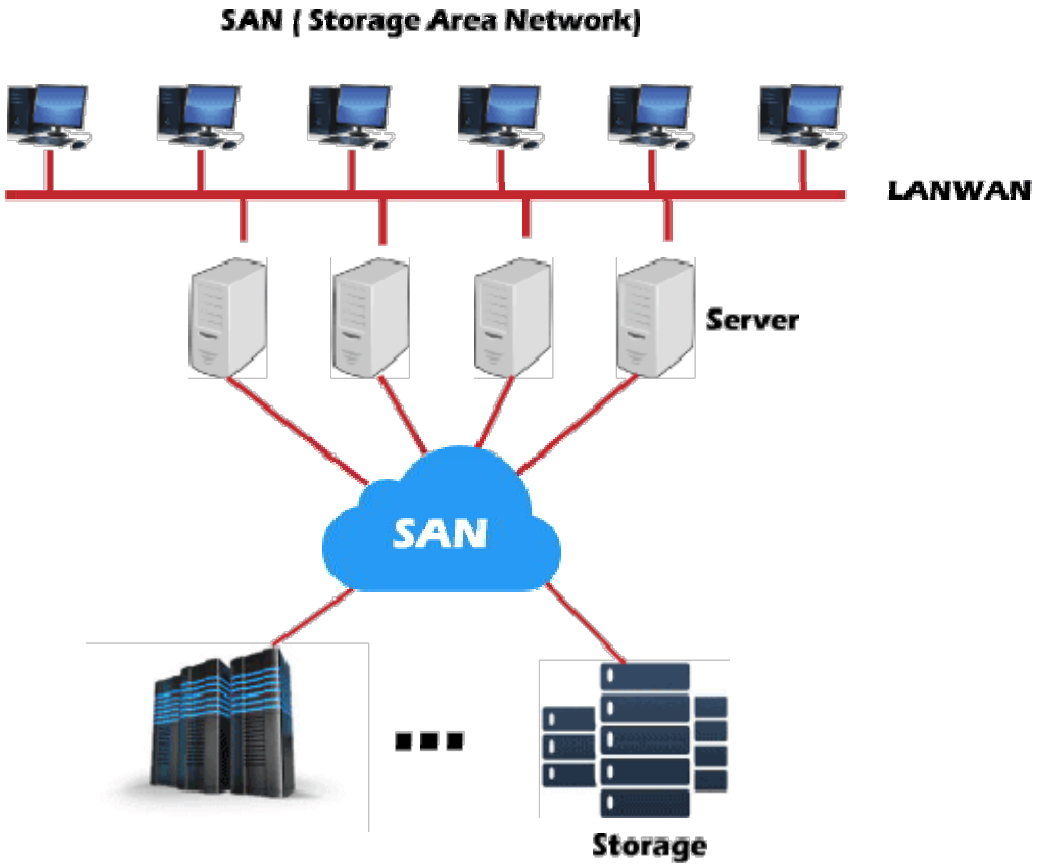
MAN or Metropolitan Area Network is typically a more extensive network when compared to LANs but is smaller than WANs. This network ranges between several buildings in the same city. Man networks are connected via fiber optic cable (usually high-speed connection). Cities and government bodies usually manage MANs.

5. PAN



PAN or Personal Area Network is a type of network used personally and usually serves one person. This network usually connects devices unlike your smartphones, laptop, or desktop to sync content and share small files, unlike songs, photos, videos, calendars, etc. These devices connect via **wireless networks such as Wi-Fi, Bluetooth, Infrared, etc.**

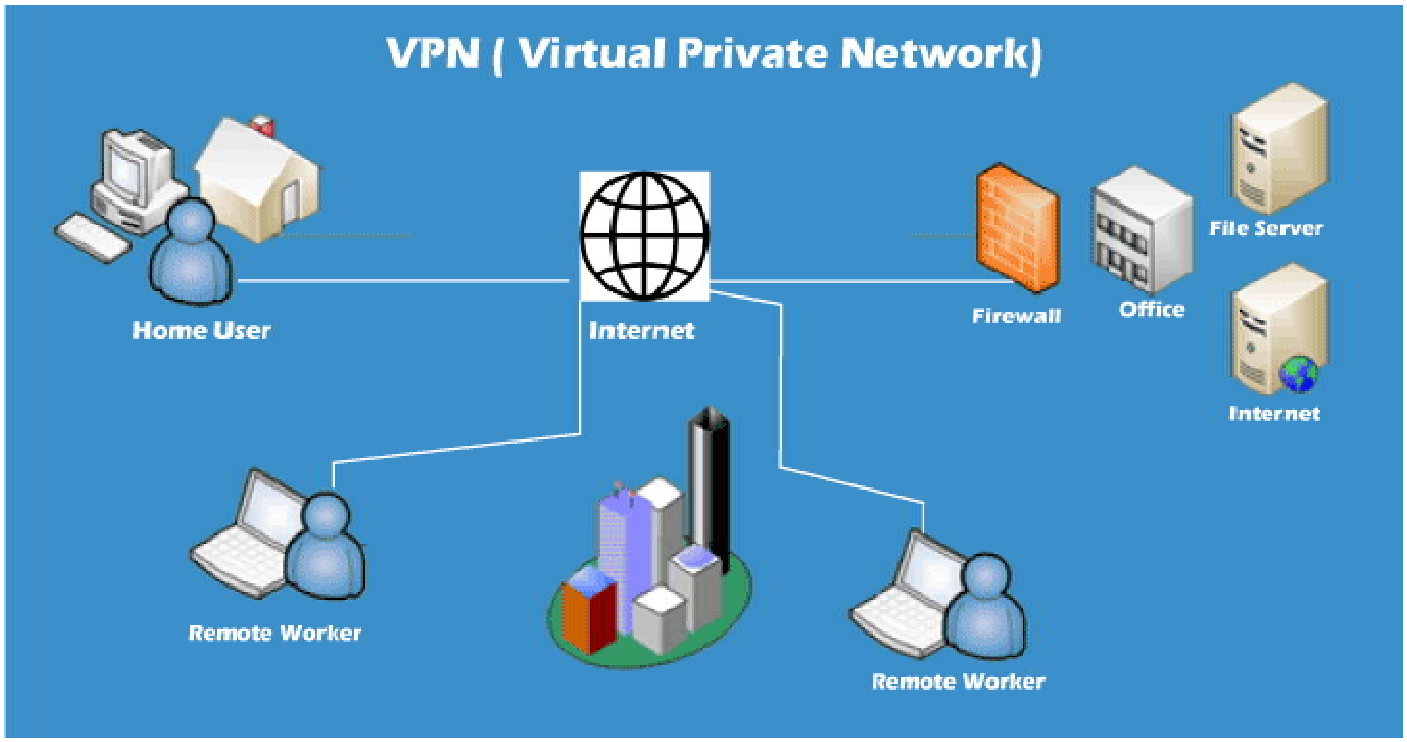
6. SAN



SAN or Storage Area Network is a specialized high-speed network that stores and provides access to block-level storage. It is a dedicated shared network that is used for cloud data storage that appears and works like a storage drive.

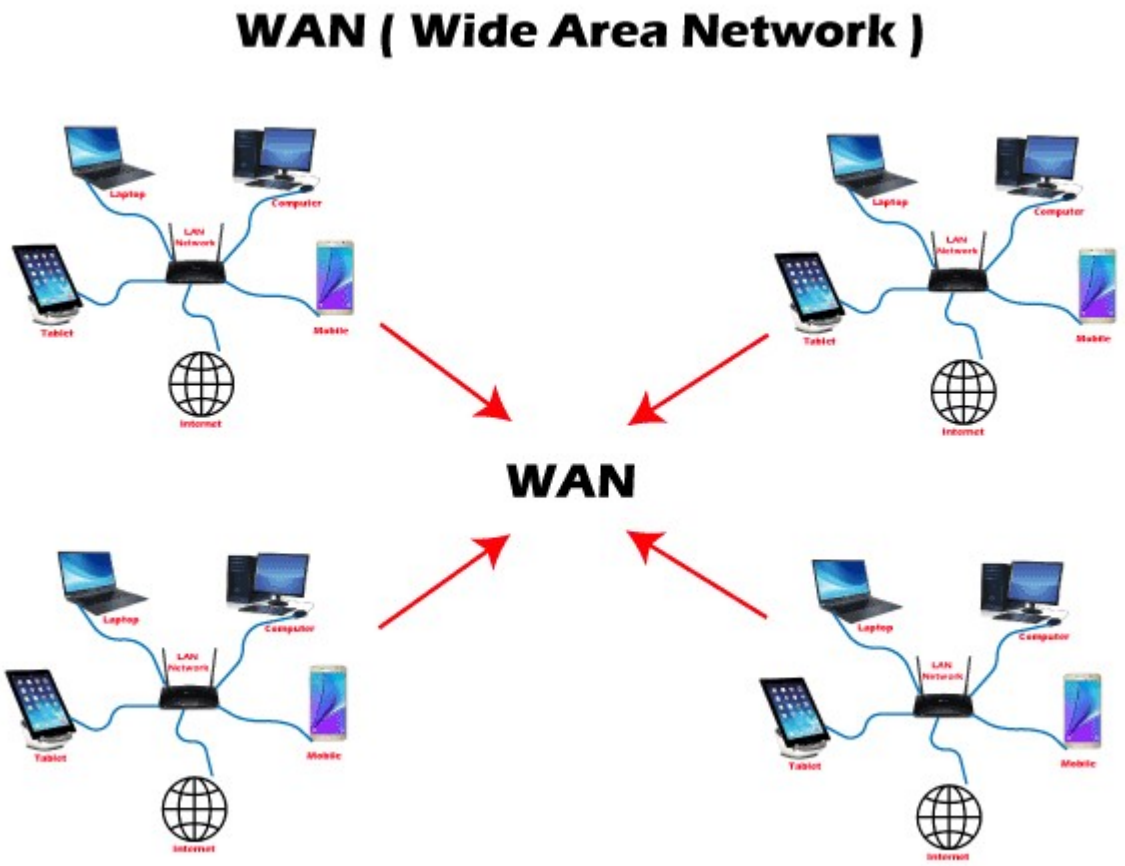
SAN consists of various **switches, servers, and disks array.** One of the advantages of SAN is that it is fault-tolerant, which means if any switch or server goes down, the data can still be accessed.

7. VPN



VPN or Virtual Private Network is a secure tool that encrypts point-to-point Internet connection and hides the user's IP address and virtual location. It determines an encrypted network to boost user's online privacy so as their identity and data are inaccessible to hackers.

8. WAN



WAN or Wide Area Network is the most significant network type connecting computers over a wide geographical area, such as a country, continent. WAN includes **several LANs, MANs, and CANs.** An example of WAN is the **Internet,** which connects billions of computers globally.

Networking terms and concepts

Some of the most commonly used terms in day-to-day networking life are as discussed below:

1. IP address

An IP address or ***Internet Protocol*** is a **unique number that represents the address where you live on the Internet.** Every device that is connected to the network has a string of numbers or IP addresses unlike house addresses.

You won't find two devices connected to a network with an identical IP address. When your computer sends data to another different, the sent data contains a 'header' that further contains the devices' IP address, i.e., the source computer and the destination device.

2. Nodes

A node refers to a networking **connection point where a connection occurs inside a network that further helps in receiving, transmitting, creating, or storing files or data.**

Multiple devices could be connected to the Internet or network using wired or wireless nodes. To form a network connection, one requires two or more nodes where each node carries its unique identification to obtain access, such as an IP **address.** Some examples of nodes are **computers, printers, modems, switches, etc.**

3. Routers

A router is a **physical networking device, which forwards data packets between networks.** Routers do the data analysis, perform the traffic directing functions on the network, and define the top route for the data packets to reach their destination node. A **data packet** may have to surpass multiple routers present within the network until it reaches its destination.

4. Switches

In a computer network, a switch is a device that **connects other devices and helps in node-to-node communication by deciding the best way of transmitting data within a network (usually if there are multiple routes in a more extensive network).**

Though a router also transmits information, it forwards the information only between networks, whereas a switches forwards data between nodes present in a single network.

Switching is further classified into three types, which are as follows:

- **Circuit Switching**
- **Packet Switching**
- **Message Switching**
- **Circuit Switching:** In this switching type, a secure communication path is established between nodes (or the sender and receiver) in a network. It establishes a dedicated connection path before transferring the data, and this path assures a good transmission bandwidth and prevents any other traffic from traveling on that path. For example, **the Telephone network.**
- **Packet Switching:** With this technique, a message is broken into independent components known as packets. Because of their small size, each packet is sent individually. The packets traveling through the network will have their source and destination IP address.
- **Message Switching:** This switching technique uses the store and forward mechanism. It sends the complete unit of the message from the source node, passing from multiple switches until it reaches its intermediary node. It is not suitable for real-time applications.

5. Ports

A port **allows the user to access multiple applications by identifying a connection between network devices.** Each port is allocated a set of string numbers. If you relate the IP address to a hotel's address, you can refer to ports as the hotel room number. Network devices use port numbers to decide which application, service, or method is used to forward the detailed information or the data.

6. Network cable types

Network cables are used as a **connection medium between different computers and other network devices.** Typical examples of network cable types are **Ethernet cables, coaxial, and fiber optic.** Though the selection of cable type usually depends on the size of the network, the organization of network components, and the distance between the network devices.

Computer Networks and the Internet

The Internet is the major example of a WAN, which connects billions of computers globally. Internet follows standard protocols that facilitate communication between these network devices. Those protocols include:

1. **HTTP (Hypertext Transfer Protocol)**
2. **IP (Internet protocol or IP addresses)**
3. **TCP (Transmission Control Protocol)**
4. **UDP (User Datagram Protocol)**
5. **FTP (File Transfer Protocol)**

ISPs (Internet Service Providers) NSPs (Network Service Providers) effectively support the internet infrastructure. The infrastructure allows the transportation of data packets to the recipient device over the Internet.

Internet is a giant hub of information, but this information is not sent to every computer connected to the Internet. The protocols and infrastructure are responsible for managing to share the precise information the user has requested.

How do they work?

1. The **Computer networks are formed by connecting multiple nodes** such as computers, desktops, routers, hubs, and switches with the help of either wired cables (Ethernet, data cables, fiber optics) or wireless networks (Bluetooth, Wi-Fi). This network connection enables the nodes to communicate and exchange data over the network.

2. Networks **follow communication protocols to send, receive, create or forward data.** Each note connected with a network is allocated a unique IP (Internet Protocol), the IP address used to identify a device and enables the other devices to identify it.
3. **Routers and Switches are the virtual or physical medium that supports and manages the communications** between networks. Routers examine the data packets to conclude the best route, following which the data can easily reach its destination node. In contrast, Switches connect the devices if there are multiple routes in a more extensive network and facilitate node-to-node communication, ensuring that the data packets traveling across the network reach their destination node.

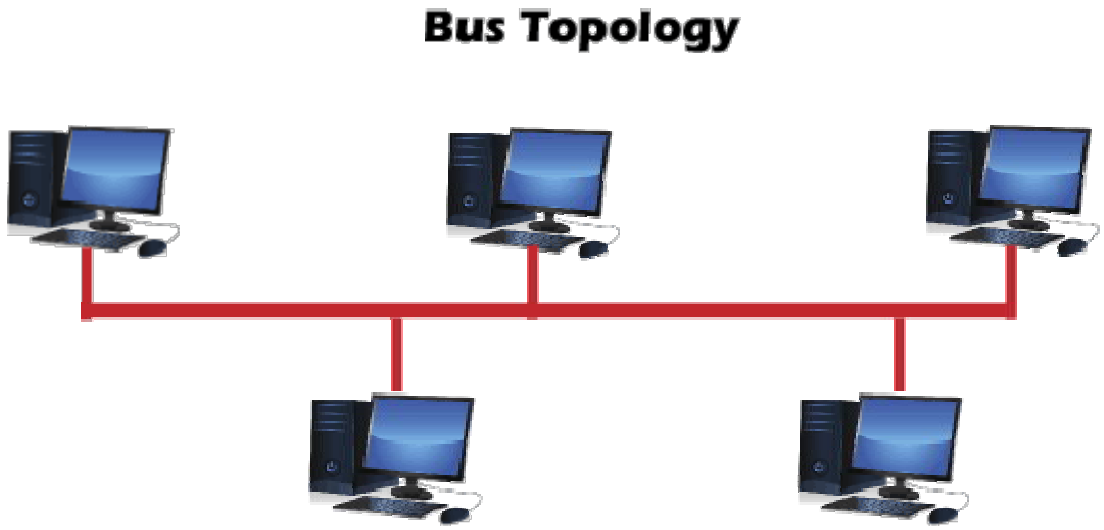
Network Topology

"Network topology is defined as the arrangement of computers or nodes of a computer network to establish communication among all."

A node refers to a device that can transmit, receive, create, or store information. The nodes are connected via a network link that could be either wired (cables, Ethernet) or wireless (Bluetooth, Wi-Fi).

To help build a successful network in different situations, topologies are further classified into several types. Though there are several topologies but in this tutorial, we will discuss the commonly used ones, which are as follows:

1. Bus Topology

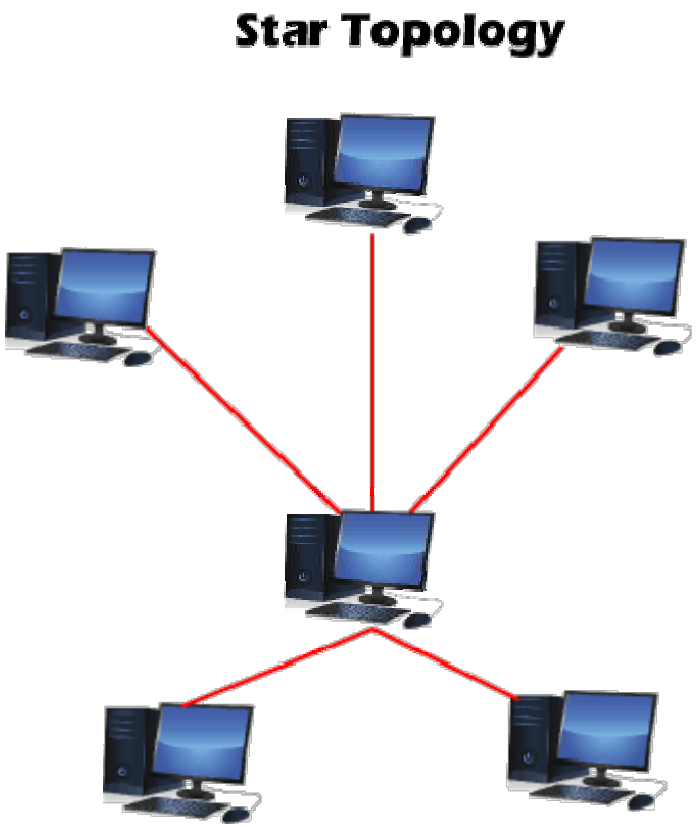


- A ***Bus network topology*** supports a common transmission medium where each node is directly connected with the main network cable.
- The data is transmitted through the main network cable and is received by all nodes simultaneously.
- A signal is generated through the source machine, which contains the address of the receiving machine. The signal travels in both the direction to all the nodes connected to the bus network until it reaches the destination node.
- Bus topology is not fault-tolerant and has a limited cable length.

2. Ring Topology

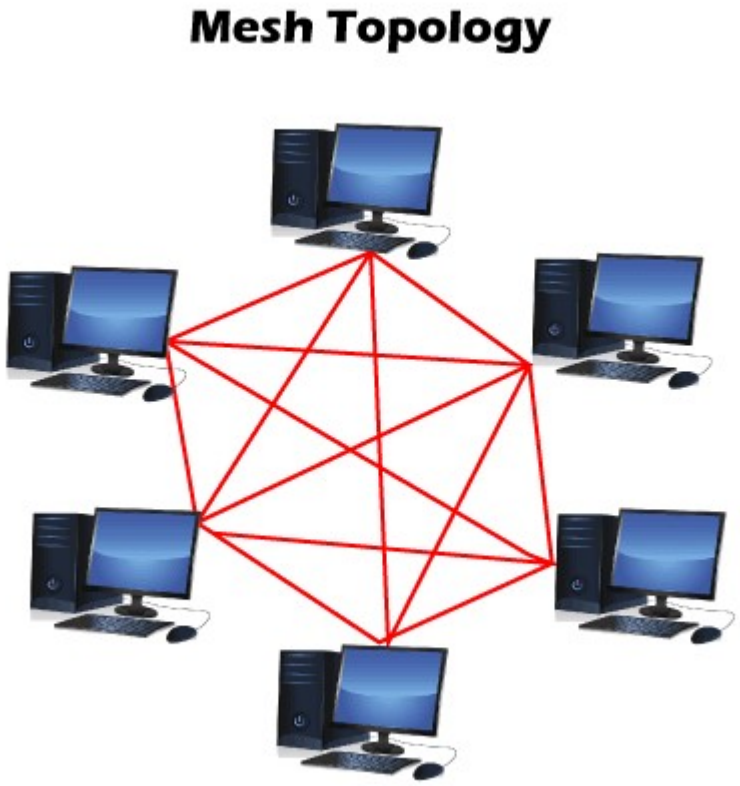
- A ***Ring topology*** is a modified version of bus topology where every node is connected in a closed-loop forming peer-to-peer LAN topology.
- Every node in a ring topology has precisely two connections. The Adjacent node pairs are connected directly, whereas the non-adjacent nodes are indirectly connected via various nodes.
- Ring topology supports a unidirectional communication pattern where sending and receiving of data occurs via **TOKEN.**

3. Star Topology



- In a ***Star network topology,*** every node is connected using a single central hub or switch.
- The hub or switch performs the entire centralized administration. Each node sends its data to the hub, and later hub shares the received information to the destination device.
- Two or more-star topologies can be connected to each other with the help of a repeater.

4. Mesh Topology



- In a ***Mesh topology,*** every node in the network connection is directly connected to one other forming overlapping connections between the nodes.
- This topology delivers better fault tolerance because if any network device fails, it won't affect the network, as other devices can transfer information.
- The Mesh networks self-configure and self-organize, finding the quickest, most secure way to transmit the data.
- One can form a full mesh topology by connecting every single node to another node in the network. **Full mesh** is expensive and is only used in the networks, which demands high data redundancy.
- Another type of mesh topology is **partial mesh topology,** where only a few devices are connected, and few are connected to the devices with which they share the most information. This mesh type is applicable in the networks, requiring less redundancy or a cost-effective network topology that is easy to execute.

Various Networking Devices at TCP/IP Layer

What are network devices?

Network devices, or networking hardware, are physical devices that are required for communication and interaction between hardware on a computer network.

Types of network devices

Here is the common network device list:

- Hub
- Switch
- Router
- Bridge

- Gateway
- Modem
- Repeater
- Access Point

Handpicked related content:

- [Free Download] Network Security Best Practices

Hub

Hubs connect multiple computer networking devices together. A hub also acts as a repeater in that it amplifies signals that deteriorate after traveling long distances over connecting cables. A hub is the simplest in the family of network connecting devices because it connects LAN components with identical protocols.

A hub can be used with both digital and analog data, provided its settings have been configured to prepare for the formatting of the incoming data. For example, if the incoming data is in digital format, the hub must pass it on as packets; however, if the incoming data is analog, then the hub passes it on in signal form.

Hubs do not perform packet filtering or addressing functions; they just send data packets to all connected devices. Hubs operate at the Physical layer of the Open Systems Interconnection (OSI) model. There are two types of hubs: simple and multiple port.

Switch

Switches generally have a more intelligent role than hubs. A switch is a multiport device that improves network efficiency. The switch maintains limited routing information about nodes in the internal network, and it allows connections to systems like hubs or routers. Strands of LANs are usually connected using switches. Generally, switches can read the hardware addresses of incoming packets to transmit them to the appropriate destination.

Using switches improves network efficiency over hubs or routers because of the virtual circuit capability. Switches also improve network security because the virtual circuits are more difficult to examine with network monitors. You can think of a switch as a device that has some of the best capabilities of routers and hubs combined. A switch can work at either the Data Link layer or the Network layer of the OSI model. A multilayer switch is one that can operate at both layers, which means that it can operate as both a switch and a router. A multilayer switch is a high-performance device that supports the same routing protocols as routers.

Switches can be subject to distributed denial of service (DDoS) attacks; flood guards are used to prevent malicious traffic from bringing the switch to a halt. Switch port security is important so be sure to secure switches: Disable all unused ports and use DHCP snooping, ARP inspection and MAC address filtering.

Handpicked related content:

- Why Native Network Device Auditing Is Not Enough

Router

Routers help transmit packets to their destinations by charting a path through the sea of interconnected networking devices using different network topologies. Routers are intelligent devices, and they store information about the networks they're connected to. Most routers can be configured to operate as packet-filtering firewalls and use access control lists (ACLs). Routers, in conjunction with a channel service unit/data service unit (CSU/DSU), are also used to translate from LAN framing to WAN framing. This is needed because LANs and WANs use different network protocols. Such routers are known as border routers. They serve as the outside connection of a LAN to a WAN, and they operate at the border of your network.

Router are also used to divide internal networks into two or more subnetworks. Routers can also be connected internally to other routers, creating zones that operate independently. Routers establish communication by maintaining tables about destinations and local connections. A router contains information about the systems connected to it and where to send requests if the destination isn't known. Routers usually communicate routing and other information using one of three standard protocols: Routing Information Protocol (RIP), Border Gateway Protocol (BGP) or Open Shortest Path First (OSPF).

Routers are your first line of defense, and they must be configured to pass only traffic that is authorized by network administrators. The routes themselves can be configured as static or dynamic. If they are static, they

can only be configured manually and stay that way until changed. If they are dynamic, they learn of other routers around them and use information about those routers to build their routing tables.

Routers are general-purpose devices that interconnect two or more heterogeneous networks. They are usually dedicated to special-purpose computers, with separate input and output network interfaces for each connected network. Because routers and gateways are the backbone of large computer networks like the internet, they have special features that give them the flexibility and the ability to cope with varying network addressing schemes and frame sizes through segmentation of big packets into smaller sizes that fit the new network components. Each router interface has its own Address Resolution Protocol (ARP) module, its own LAN address (network card address) and its own Internet Protocol (IP) address. The router, with the help of a routing table, has knowledge of routes a packet could take from its source to its destination. The routing table, like in the bridge and switch, grows dynamically. Upon receipt of a packet, the router removes the packet headers and trailers and analyzes the IP header by determining the source and destination addresses and data type, and noting the arrival time. It also updates the router table with new addresses not already in the table. The IP header and arrival time information is entered in the routing table. Routers normally work at the Network layer of the OSI model.

Handpicked related content:

- Why Monitoring of Network Devices Is Critical for Network Security

Bridge

Bridges are used to connect two or more hosts or network segments together. The basic role of bridges in network architecture is storing and forwarding frames between the different segments that the bridge connects. They use hardware Media Access Control (MAC) addresses for transferring frames. By looking at the MAC address of the devices connected to each segment, bridges can forward the data or block it from crossing. Bridges can also be used to connect two physical LANs into a larger logical LAN.

Bridges work only at the Physical and Data Link layers of the OSI model. Bridges are used to divide larger networks into smaller sections by sitting between two physical network segments and managing the flow of data between the two.

Bridges are like hubs in many respects, including the fact that they connect LAN components with identical protocols. However, bridges filter incoming data packets, known as frames, for addresses before they are forwarded. As it filters the data packets, the bridge makes no modifications to the format or content of the incoming data. The bridge filters and forwards frames on the network with the help of a dynamic bridge table. The bridge table, which is initially empty, maintains the LAN addresses for each computer in the LAN and the addresses of each bridge interface that connects the LAN to other LANs. Bridges, like hubs, can be either simple or multiple port.

Bridges have mostly fallen out of favor in recent years and have been replaced by switches, which offer more functionality. In fact, switches are sometimes referred to as “multiport bridges” because of how they operate.

Gateway

Gateways normally work at the Transport and Session layers of the OSI model. At the Transport layer and above, there are numerous protocols and standards from different vendors; gateways are used to deal with them. Gateways provide translation between networking technologies such as Open System Interconnection (OSI) and Transmission Control Protocol/Internet Protocol (TCP/IP). Because of this, gateways connect two or more autonomous networks, each with its own routing algorithms, protocols, topology, domain name service, and network administration procedures and policies.

Gateways perform all of the functions of routers and more. In fact, a router with added translation functionality is a gateway. The function that does the translation between different network technologies is called a protocol converter.

Modem

Modems (modulators-demodulators) are used to transmit digital signals over analog telephone lines. Thus, digital signals are converted by the modem into analog signals of different frequencies and transmitted to a modem at the receiving location. The receiving modem performs the reverse transformation and provides a digital output to a device connected to a modem, usually a computer. The digital data is usually transferred to

or from the modem over a serial line through an industry standard interface, RS-232. Many telephone companies offer DSL services, and many cable operators use modems as end terminals for identification and recognition of home and personal users. Modems work on both the Physical and Data Link layers.

Repeater

A repeater is an electronic device that amplifies the signal it receives. You can think of repeater as a device which receives a signal and retransmits it at a higher level or higher power so that the signal can cover longer distances, more than 100 meters for standard LAN cables. Repeaters work on the Physical layer.

Access Point

While an access point (AP) can technically involve either a wired or wireless connection, it commonly means a wireless device. An AP works at the second OSI layer, the Data Link layer, and it can operate either as a bridge connecting a standard wired network to wireless devices or as a router passing data transmissions from one access point to another.

Wireless access points (WAPs) consist of a transmitter and receiver (transceiver) device used to create a wireless LAN (WLAN). Access points typically are separate network devices with a built-in antenna, transmitter and adapter. APs use the wireless infrastructure network mode to provide a connection point between WLANs and a wired Ethernet LAN. They also have several ports, giving you a way to expand the network to support additional clients. Depending on the size of the network, one or more APs might be required to provide full coverage. Additional APs are used to allow access to more wireless clients and to expand the range of the wireless network. Each AP is limited by its transmission range — the distance a client can be from an AP and still obtain a usable signal and data process speed. The actual distance depends on the wireless standard, the obstructions and environmental conditions between the client and the AP. Higher end APs have high-powered antennas, enabling them to extend how far the wireless signal can travel.

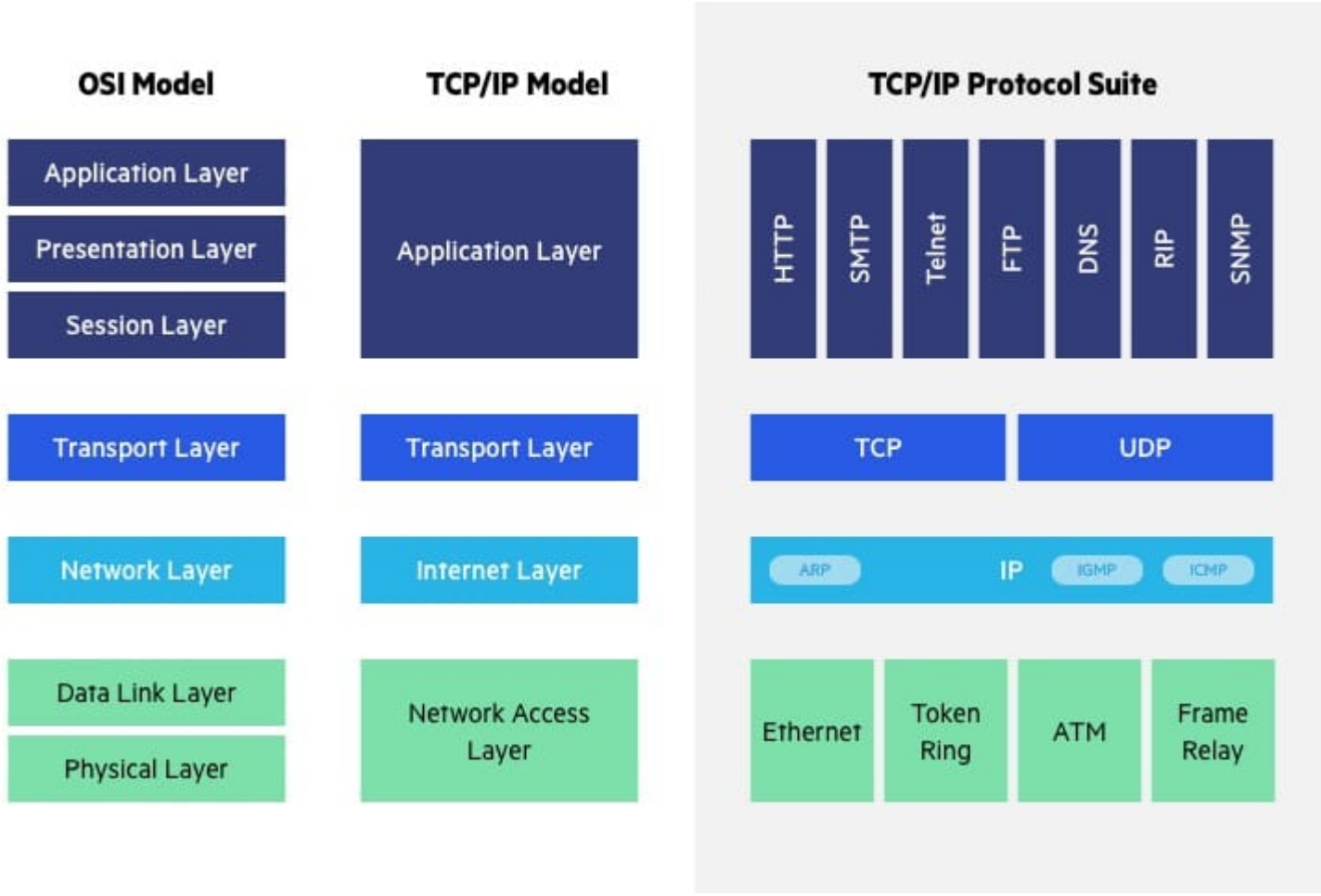
APs might also provide many ports that can be used to increase the network's size, firewall capabilities and Dynamic Host Configuration Protocol (DHCP) service. Therefore, we get APs that are a switch, DHCP server, router and firewall.

To connect to a wireless AP, you need a service set identifier (SSID) name. 802.11 wireless networks use the SSID to identify all systems belonging to the same network, and client stations must be configured with the SSID to be authenticated to the AP. The AP might broadcast the SSID, allowing all wireless clients in the area to see the AP's SSID. However, for security reasons, APs can be configured not to broadcast the SSID, which means that an administrator needs to give client systems the SSID instead of allowing it to be discovered automatically. Wireless devices ship with default SSIDs, security settings, channels, passwords and usernames. For security reasons, it is strongly recommended that you change these default settings as soon as possible because many internet sites list the default settings used by manufacturers.

Access points can be fat or thin. Fat APs, sometimes still referred to as autonomous APs, need to be manually configured with network and security settings; then they are essentially left alone to serve clients until they can no longer function. Thin APs allow remote configuration using a controller. Since thin clients do not need to be manually configured, they can be easily reconfigured and monitored. Access points can also be controller-based or stand-alone.

OSI Layer / Model

OSI vs. TCP/IP Model



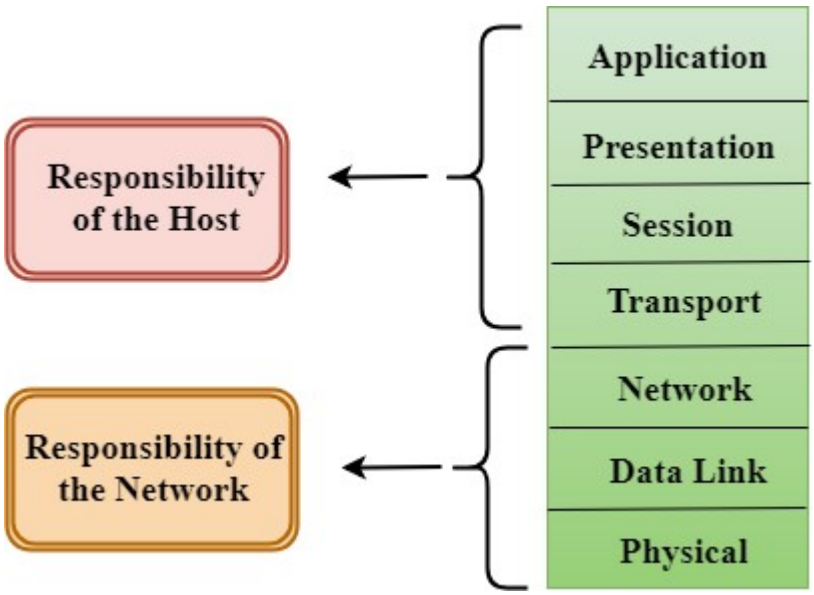
The Transfer Control Protocol/Internet Protocol (TCP/IP) is older than the OSI model and was created by the US Department of Defense (DoD). A key difference between the models is that TCP/IP is simpler, collapsing several OSI layers into one:

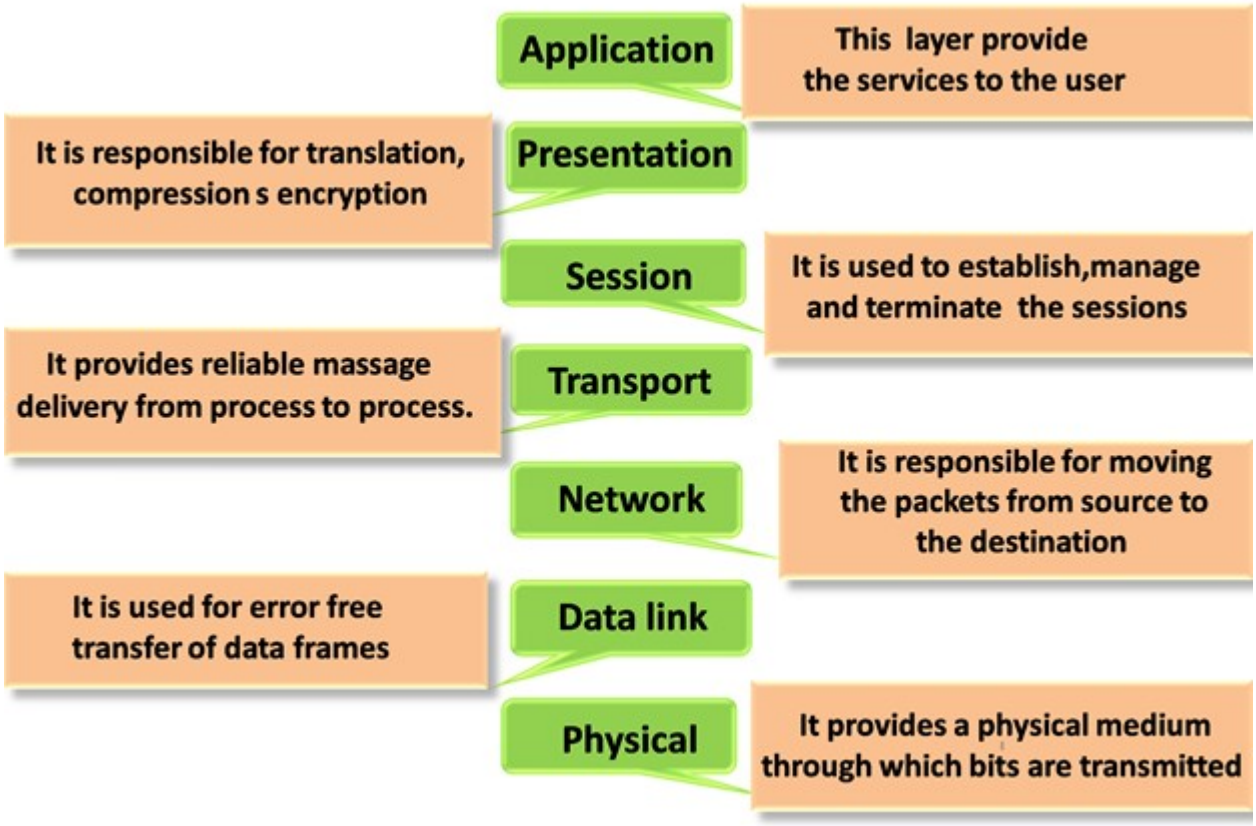
- OSI layers 5, 6, 7 are combined into one Application Layer in TCP/IP
- OSI layers 1, 2 are combined into one Network Access Layer in TCP/IP – however TCP/IP does not take responsibility for sequencing and acknowledgement functions, leaving these to the underlying transport layer.

Other important differences:

- TCP/IP is a functional model designed to solve specific communication problems, and which is based on specific, standard protocols. OSI is a generic, protocol-independent model intended to describe all forms of network communication.
- In TCP/IP, most applications use all the layers, while in OSI simple applications do not use all seven layers. Only layers 1, 2 and 3 are mandatory to enable any data communication.

Characteristics of OSI Model:





IP Address and Subnetting

Introduction

This document provides basic information needed in order to configure your router for routing IP, such as how addresses are broken down and how subnetting works. You learn how to assign each interface on the router an IP address with a unique subnet. There are examples included in order to help tie everything together.

Prerequisites

Requirements

Cisco recommends that you have a basic understanding of binary and decimal numbers.

Components Used

This document is not restricted to specific software and hardware versions.

The information in this document was created from the devices in a specific lab environment. All of the devices used in this document started with a cleared (default) configuration. If your network is live, make sure that you understand the potential impact of any command.

Additional Information

If definitions are helpful to you, use these vocabulary terms in order to get you started:

- **Address -** The unique number ID assigned to one host or interface in a network.
- **Subnet -** A portion of a network that shares a particular subnet address.
- **Subnet mask -** A 32-bit combination used to describe which portion of an address refers to the subnet and which part refers to the host.
- **Interface -** A network connection.

If you have already received your legitimate address(es) from the Internet Network Information Center (InterNIC), you are ready to begin. If you do not plan to connect to the Internet, Cisco strongly suggests that you use reserved addresses from RFC 1918.

Understand IP Addresses

An IP address is an address used in order to uniquely identify a device on an IP network. The address is made up of 32 binary bits, which can be divisible into a network portion and host portion with the help of a subnet mask. The 32 binary bits are broken into four octets (1 octet = 8 bits). Each octet is converted to decimal and separated by a period (dot). For this reason, an IP address is said to be expressed in dotted decimal format (for example, 172.16.81.100). The value in each octet ranges from 0 to 255 decimal, or 00000000 - 11111111 binary.

Here is how binary octets convert to decimal: The right most bit, or least significant bit, of an octet holds a value of 2^0. The bit just to the left of that holds a value of 2^1. This continues until the left-most bit, or most significant bit, which holds a value of 2^7. So if all binary bits are a one, the decimal equivalent would be 255 as shown here:

1 1 1 1 1 1 1 1

128 64 32 16 8 4 2 1 (128+64+32+16+8+4+2+1=255)

Here is a sample octet conversion when not all of the bits are set to 1.

0 1 0 0 0 0 0 1

0 64 0 0 0 0 0 1 (0+64+0+0+0+0+0+1=65)

And this sample shows an IP address represented in both binary and decimal.

10. 1. 23. 19 (decimal)

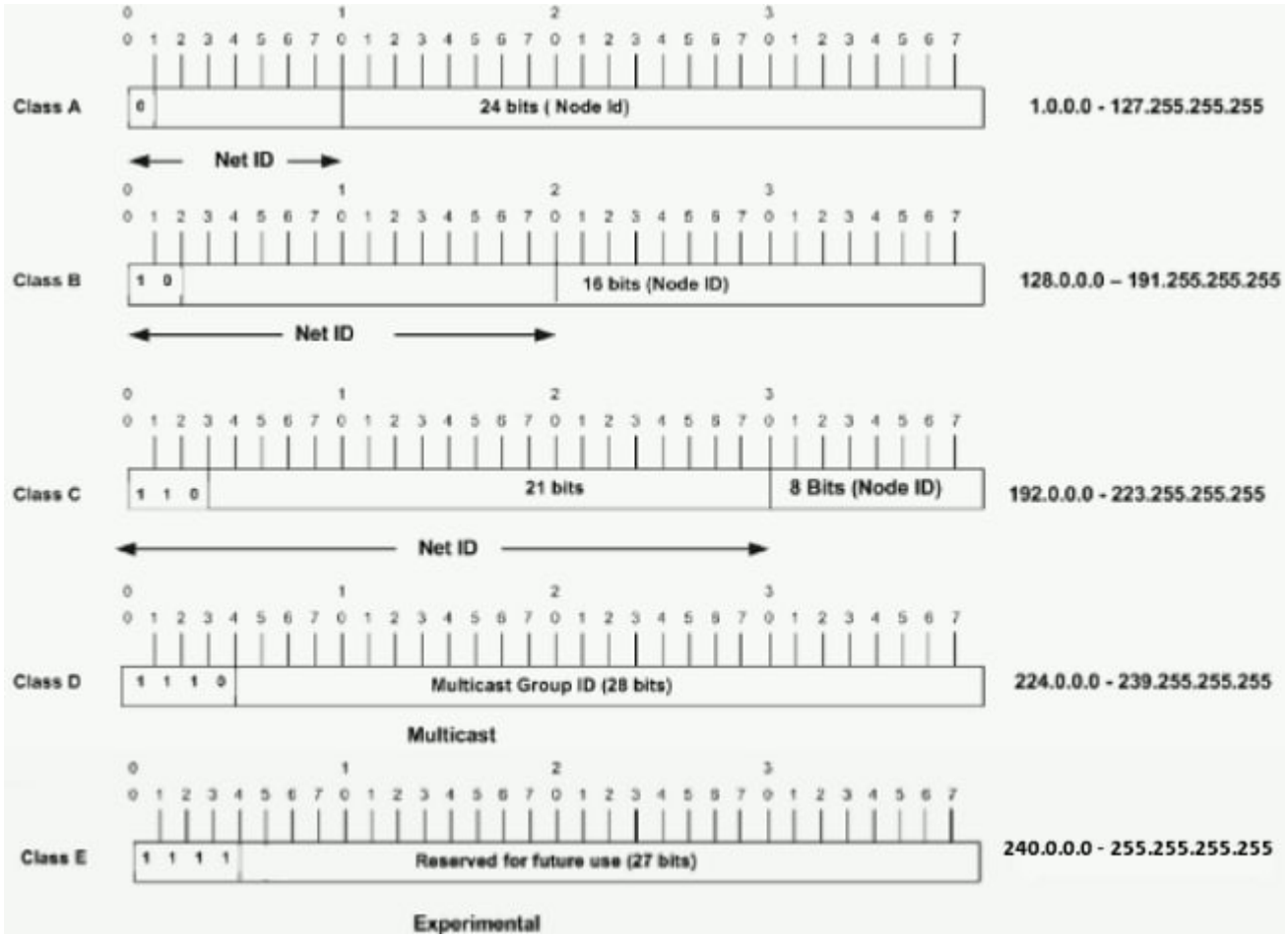
00001010.00000001.00010111.00010011 (binary)

These octets are broken down to provide an addressing scheme that can accommodate large and small networks. There are five different classes of networks, A to E. This document focuses on classes A to C, since classes D and E are reserved and discussion of them is beyond the scope of this document.

Note: Also note that the terms "Class A, Class B" and so on are used in this document in order to help facilitate the understanding of IP addressing and subnetting. These terms are rarely used in the industry anymore because of the introduction of classless interdomain routing (CIDR).

Given an IP address, its class can be determined from the three high-order bits (the three left-most bits in the first octet). Figure 1 shows the significance in the three high order bits and the range of addresses that fall into each class. For informational purposes, Class D and Class E addresses are also shown.

Figure 1



In a Class A address, the first octet is the network portion, so the Class A example in Figure 1 has a major network address of 1.0.0.0 - 127.255.255.255. Octets 2, 3, and 4 (the next 24 bits) are for the network manager to divide into subnets and hosts as he/she sees fit. Class A addresses are used for networks that have more than 65,536 hosts (actually, up to 16777214 hosts!).

In a Class B address, the first two octets are the network portion, so the Class B example in Figure 1 has a major network address of 128.0.0.0 - 191.255.255.255. Octets 3 and 4 (16 bits) are for local subnets and hosts. Class B addresses are used for networks that have between 256 and 65534 hosts.

In a Class C address, the first three octets are the network portion. The Class C example in Figure 1 has a major network address of 192.0.0.0 - 223.255.255.255. Octet 4 (8 bits) is for local subnets and hosts - perfect for networks with less than 254 hosts.

Network Masks

A network mask helps you know which portion of the address identifies the network and which portion of the address identifies the node. Class A, B, and C networks have default masks, also known as natural masks, as shown here:

Class A: 255.0.0.0

Class B: 255.255.0.0

Class C: 255.255.255.0

An IP address on a Class A network that has not been subnetted would have an address/mask pair similar to: 8.20.15.1 255.0.0.0. In order to see how the mask helps you identify the network and node parts of the address, convert the address and mask to binary numbers.

8.20.15.1 = 00001000.00010100.00001111.00000001

255.0.0.0 = 11111111.00000000.00000000.00000000

Once you have the address and the mask represented in binary, then identification of the network and host ID is easier. Any address bits which have corresponding mask bits set to 1 represent the network ID. Any address bits that have corresponding mask bits set to 0 represent the node ID.

8.20.15.1 = 00001000.00010100.00001111.00000001
255.0.0.0 = 11111111.00000000.00000000.00000000
---
net id | host id

netid = 00001000 = 8
hostid = 00010100.00001111.00000001 = 20.15.1

Understand Subnetting

Subnetting allows you to create multiple logical networks that exist within a single Class A, B, or C network. If you do not subnet, you are only able to use one network from your Class A, B, or C network, which is unrealistic.

Each data link on a network must have a unique network ID, with every node on that link being a member of the same network. If you break a major network (Class A, B, or C) into smaller subnetworks, it allows you to create a network of interconnecting subnetworks. Each data link on this network would then have a unique network/subnetwork ID. Any device, or gateway, that connects *n* networks/subnetworks has *n* distinct IP addresses, one for each network / subnetwork that it interconnects.

In order to subnet a network, extend the natural mask with some of the bits from the host ID portion of the address in order to create a subnetwork ID. For example, given a Class C network of 204.17.5.0 which has a natural mask of 255.255.255.0, you can create subnets in this manner:

204.17.5.0 - 11001100.00010001.00000101.00000000
255.255.255.224 - 11111111.11111111.11111111.11100000
---|sub|----

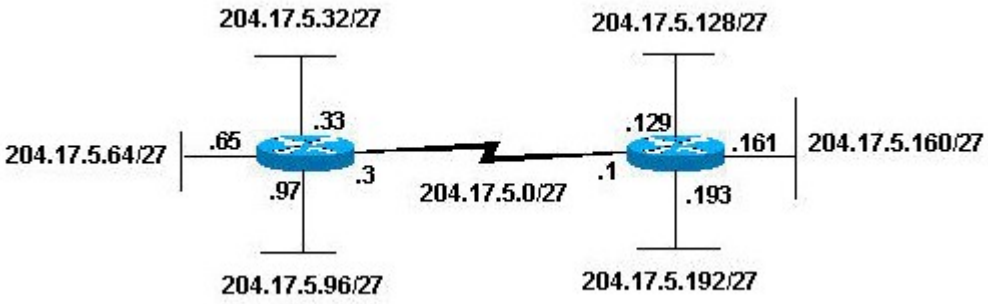
By extending the mask to be 255.255.255.224, you have taken three bits (indicated by "sub") from the original host portion of the address and used them to make subnets. With these three bits, it is possible to create eight subnets. With the remaining five host ID bits, each subnet can have up to 32 host addresses, 30 of which can actually be assigned to a device *since host ids of all zeros or all ones are not allowed* (it is very important to remember this). So, with this in mind, these subnets have been created.

204.17.5.0 255.255.255.224 host address range 1 to 30
204.17.5.32 255.255.255.224 host address range 33 to 62
204.17.5.64 255.255.255.224 host address range 65 to 94
204.17.5.96 255.255.255.224 host address range 97 to 126
204.17.5.128 255.255.255.224 host address range 129 to 158
204.17.5.160 255.255.255.224 host address range 161 to 190
204.17.5.192 255.255.255.224 host address range 193 to 222
204.17.5.224 255.255.255.224 host address range 225 to 254

Note: There are two ways to denote these masks. First, since you use three bits more than the "natural" Class C mask, you can denote these addresses as having a 3-bit subnet mask. Or, secondly, the mask of 255.255.255.224 can also be denoted as /27 as there are 27 bits that are set in the mask. This second method is used with CIDR. With this method, one of these networks can be described with the notation prefix/length. For example, 204.17.5.32/27 denotes the network 204.17.5.32 255.255.255.224. When appropriate, the prefix/length notation is used to denote the mask throughout the rest of this document.

The network subnetting scheme in this section allows for eight subnets, and the network might appear as:

Figure 2



Notice that each of the routers in Figure 2 is attached to four subnetworks, one subnetwork is common to both routers. Also, each router has an IP address for each subnetwork to which it is attached. Each subnetwork could potentially support up to 30 host addresses.

This brings up an interesting point. The more host bits you use for a subnet mask, the more subnets you have available. However, the more subnets available, the less host addresses available per subnet. For example, a Class C network of 204.17.5.0 and a mask of 255.255.255.224 (/27) allows you to have eight subnets, each with 32 host addresses (30 of which could be assigned to devices). If you use a mask of 255.255.255.240 (/28), the breakdown is:

204.17.5.0 - 11001100.00010001.00000101.00000000
255.255.255.240 - 11111111.11111111.11111111.11110000
----------|sub |---

Since you now have four bits to make subnets with, you only have four bits left for host addresses. So in this case you can have up to 16 subnets, each of which can have up to 16 host addresses (14 of which can be assigned to devices).

Take a look at how a Class B network might be subnetted. If you have network 172.16.0.0, then you know that its natural mask is 255.255.0.0 or 172.16.0.0/16. Extending the mask to anything beyond 255.255.0.0 means you are subnetting. You can quickly see that you have the ability to create a lot more subnets than with the Class C network. If you use a mask of 255.255.248.0 (/21), how many subnets and hosts per subnet does this allow for?

172.16.0.0 - 10101100.00010000.00000000.00000000
255.255.248.0 - 11111111.11111111.11111000.00000000
----------| sub |----------

You use five bits from the original host bits for subnets. This allows you to have 32 subnets (2^5). After using the five bits for subnetting, you are left with 11 bits for host addresses. This allows each subnet so have 2048 host addresses (2^{11}), 2046 of which could be assigned to devices.

Note: In the past, there were limitations to the use of a subnet 0 (all subnet bits are set to zero) and all ones subnet (all subnet bits set to one). Some devices would not allow the use of these subnets. Cisco Systems devices allow the use of these subnets when the **ip subnet zero** command is configured.

Examples

Sample Exercise 1

Now that you have an understanding of subnetting, put this knowledge to use. In this example, you are given two address / mask combinations, written with the prefix/length notation, which have been assigned to two devices. Your task is to determine if these devices are on the same subnet or different subnets. You can use the address and mask of each device in order to determine to which subnet each address belongs.

DeviceA: 172.16.17.30/20
DeviceB: 172.16.28.15/20

Determine the Subnet for DeviceA:

172.16.17.30 - 10101100.00010000.00010001.00011110

255.255.240.0 - 11111111.11111111.11110000.00000000
----------| sub|----------
subnet = 10101100.00010000.00010000.00000000 = 172.16.16.0
Looking at the address bits that have a corresponding mask bit set to one, and setting all the other address bits to zero (this is equivalent to performing a logical "AND" between the mask and address), shows you to which subnet this address belongs. In this case, DeviceA belongs to subnet 172.16.16.0.

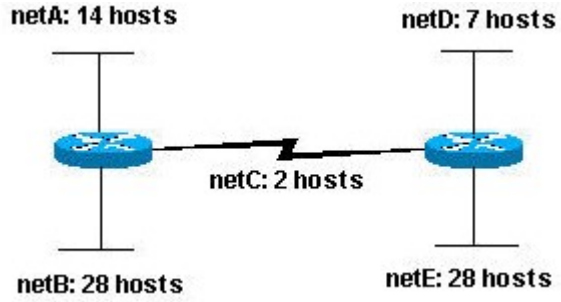
Determine the Subnet for DeviceB:

172.16.28.15 - 10101100.00010000.00011100.00001111
255.255.240.0 - 11111111.11111111.11110000.00000000
----------| sub|----------
subnet = 10101100.00010000.00010000.00000000 = 172.16.16.0
From these determinations, DeviceA and DeviceB have addresses that are part of the same subnet.

Sample Exercise 2

Given the Class C network of 204.15.5.0/24, subnet the network in order to create the network in Figure 3 with the host requirements shown.

Figure 3



Looking at the network shown in Figure 3, you can see that you are required to create five subnets. The largest subnet must support 28 host addresses. Is this possible with a Class C network? And if so, then how?

You can start by looking at the subnet requirement. In order to create the five needed subnets you would need to use three bits from the Class C host bits. Two bits would only allow you four subnets (2^2).

Since you need three subnet bits, that leaves you with five bits for the host portion of the address. How many hosts does this support? $2^5 = 32$ (30 usable). This meets the requirement.

Therefore you have determined that it is possible to create this network with a Class C network. An example of how you might assign the subnetworks is:

netA: 204.15.5.0/27 host address range 1 to 30
netB: 204.15.5.32/27 host address range 33 to 62
netC: 204.15.5.64/27 host address range 65 to 94
netD: 204.15.5.96/27 host address range 97 to 126
netE: 204.15.5.128/27 host address range 129 to 158

VLSM Example

In all of the previous examples of subnetting, notice that the same subnet mask was applied for all the subnets. This means that each subnet has the same number of available host addresses. You can need this in some cases, but, in most cases, having the same subnet mask for all subnets ends up wasting address space. For example, in the Sample Exercise 2 section, a class C network was split into eight equal-size subnets; however, each subnet did not utilize all available host addresses, which results in wasted address space. Figure 4 illustrates this wasted address space.

Figure 4

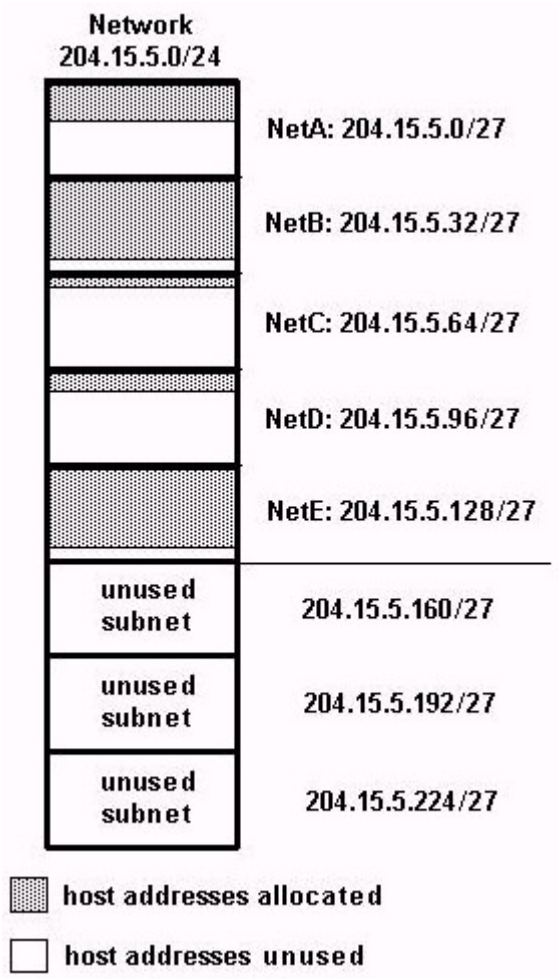


Figure 4 illustrates that of the subnets that are being used, NetA, NetC, and NetD have a lot of unused host address space. It is possible that this was a deliberate design accounting for future growth, but in many cases this is just wasted address space due to the fact that the same subnet mask is used for all the subnets.

Variable Length Subnet Masks (VLSM) allows you to use different masks for each subnet, thereby using address space efficiently.

VLSM Example

Given the same network and requirements as in Sample Exercise 2 develop a subnetting scheme with the use of VLSM, given:

netA: must support 14 hosts
netB: must support 28 hosts
netC: must support 2 hosts
netD: must support 7 hosts
netE: must support 28 host

Determine what mask allows the required number of hosts.

netA: requires a /28 (255.255.255.240) mask to support 14 hosts
netB: requires a /27 (255.255.255.224) mask to support 28 hosts
netC: requires a /30 (255.255.255.252) mask to support 2 hosts
netD*: requires a /28 (255.255.255.240) mask to support 7 hosts
netE: requires a /27 (255.255.255.224) mask to support 28 hosts

* a /29 (255.255.255.248) would only allow 6 usable host addresses
therefore netD requires a /28 mask.

The easiest way to assign the subnets is to assign the largest first. For example, you can assign in this manner:

netB: 204.15.5.0/27 host address range 1 to 30
netE: 204.15.5.32/27 host address range 33 to 62
netA: 204.15.5.64/28 host address range 65 to 78
netD: 204.15.5.80/28 host address range 81 to 94
netC: 204.15.5.96/30 host address range 97 to 98
This can be graphically represented as shown in Figure 5:

Figure 5

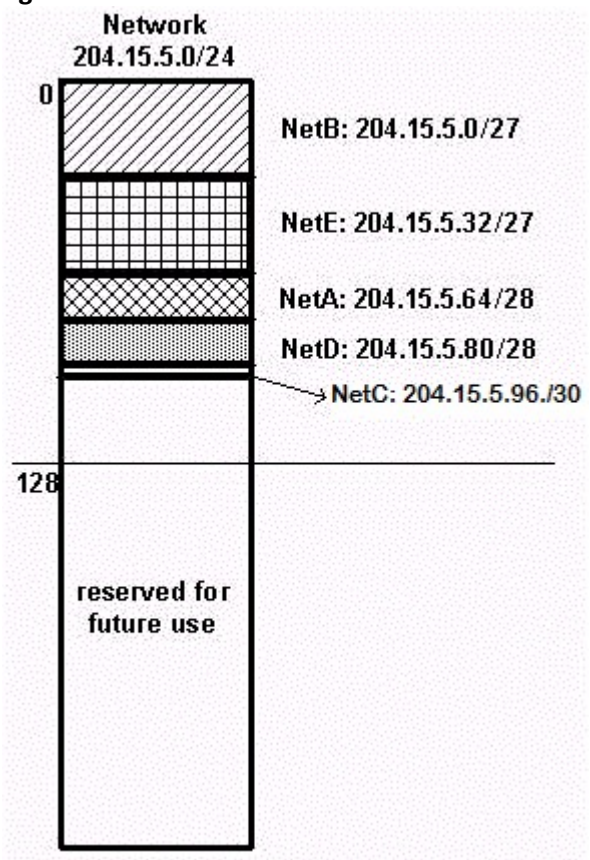


Figure 5 illustrates how using VLSM helped save more than half of the address space.

CIDR

Classless Interdomain Routing (CIDR) was introduced in order to improve both address space utilization and routing scalability in the Internet. It was needed because of the rapid growth of the Internet and growth of the IP routing tables held in the Internet routers.

CIDR moves away from the traditional IP classes (Class A, Class B, Class C, and so on). In CIDR , an IP network is represented by a prefix, which is an IP address and some indication of the length of the mask. Length means the number of left-most contiguous mask bits that are set to one. So network 172.16.0.0 255.255.0.0 can be represented as 172.16.0.0/16. CIDR also depicts a more hierarchical Internet architecture, where each domain takes its IP addresses from a higher level. This allows for the summarization of the domains to be done at the higher level. For example, if an ISP owns network 172.16.0.0/16, then the ISP can offer 172.16.1.0/24, 172.16.2.0/24, and so on to customers. Yet, when advertising to other providers, the ISP only needs to advertise 172.16.0.0/16.

For more information on CIDR, see RFC 1518 and RFC 1519.

Special Subnets

31-bit Subnets

A 30-bit subnet mask allows for four IPv4 addresses: two host addresses, one all-zeros network, and one all-ones broadcast address. A point-to-point link can only have two host addresses. There is no real need to have the broadcast and all-zeros addresses with point-to-point links. A 31-bit subnet mask will allow for exactly two

host addresses, and eliminates the broadcast and all-zeros addresses, thus conserving the use of IP addresses to the minimum for point-to-point links.
Refer to RFC 3021 - Using 31-Bit Prefixes on IPv4 Point-to-Point Links.
The mask is 255.255.255.254 or /31.

The /31 subnet can be used on true point-to-point links, such as serial or POS interfaces. However, they can also be used on broadcast interface types like ethernet interfaces. If that is the case, make sure there are only two IPv4 addresses needed on that ethernet segment.

Example

192.168.1.0 and 192.168.1.1 are on the subnet 192.168.1.0/31.

```
R1(config)#int gigabitEthernet 0/1
R1(config-if)#ip address 192.168.1.0 255.255.255.254
% Warning: use /31 mask on non point-to-point interface cautiously
```

The warning is printed because gigabitEthernet is a broadcast segment.

32-bit Subnets

A subnet mask of 255.255.255.255 (a /32 subnet) describes a subnet with only one IPv4 host address. These subnets cannot be used for assigning address to network links, because they always need more than one address per link. The use of /32 is strictly reserved for use on links that can have only one address. The example for Cisco routers is the loopback interface. These interfaces are internal interfaces and do not connect to other devices. As such, they can have a /32 subnet.

Example

```
interface Loopback0
ip address 192.168.2.1 255.255.255.255
```

Appendix

Sample Configuration

Routers A and B are connected via serial interface.

Router A

```
 hostname routera
 !
 ip routing
 !
 int e 0
 ip address 172.16.50.1 255.255.255.0
 !(subnet 50)
 int e 1 ip address 172.16.55.1 255.255.255.0
 !(subnet 55)
 int s 0 ip address 172.16.60.1 255.255.255.0
 !(subnet 60) int s 0
 ip address 172.16.65.1 255.255.255.0 (subnet 65)
 !S 0 connects to router B
 router rip
 network 172.16.0.0
```

Router B

```
 hostname routerb
 !
 ip routing
 !
 int e 0
 ip address 192.1.10.200 255.255.255.240
 !(subnet 192)
```

```
int e 1
ip address 192.1.10.66 255.255.255.240
!(subnet 64)
int s 0
ip address 172.16.65.2 (same subnet as router A's s 0)
!Int s 0 connects to router A
router rip
network 192.1.10.0
network 172.16.0.0
```

Host/Subnet Quantities Table

Class B # bits	Mask	Effective Subnets	Effective Hosts
1	255.255.128.0	2	32766
2	255.255.192.0	4	16382
3	255.255.224.0	8	8190
4	255.255.240.0	16	4094
5	255.255.248.0	32	2046
6	255.255.252.0	64	1022
7	255.255.254.0	128	510
8	255.255.255.0	256	254
9	255.255.255.128	512	126
10	255.255.255.192	1024	62
11	255.255.255.224	2048	30
12	255.255.255.240	4096	14
13	255.255.255.248	8192	6
14	255.255.255.252	16384	2

Class C # bits	Mask	Effective Subnets	Effective Hosts
1	255.255.255.128	2	126
2	255.255.255.192	4	62
3	255.255.255.224	8	30
4	255.255.255.240	16	14
5	255.255.255.248	32	6
6	255.255.255.252	64	2

*Subnet all zeroes and all ones included. These might not be supported on some legacy systems.
*Host all zeroes and all ones excluded.

Summary

When you configure the TCP/IP protocol on a Windows computer, the TCP/IP configuration settings require:

- An IP address
- A subnet mask
- A default gateway

To configure TCP/IP correctly, it's necessary to understand how TCP/IP networks are addressed and divided into networks and subnetworks.

The success of TCP/IP as the network protocol of the Internet is largely because of its ability to connect together networks of different sizes and systems of different types. These networks are arbitrarily defined into

three main classes (along with a few others) that have predefined sizes. Each of them can be divided into smaller subnetworks by system administrators. A subnet mask is used to divide an IP address into two parts. One part identifies the host (computer), the other part identifies the network to which it belongs. To better understand how IP addresses and subnet masks work, look at an IP address and see how it's organized.

IP addresses: Networks and hosts

An IP address is a 32-bit number. It uniquely identifies a host (computer or other device, such as a printer or router) on a TCP/IP network.

IP addresses are normally expressed in dotted-decimal format, with four numbers separated by periods, such as 192.168.123.132. To understand how subnet masks are used to distinguish between hosts, networks, and subnetworks, examine an IP address in binary notation.

For example, the dotted-decimal IP address 192.168.123.132 is (in binary notation) the 32-bit number 11000000010100001111011100000100. This number may be hard to make sense of, so divide it into four parts of eight binary digits.

These 8-bit sections are known as octets. The example IP address, then, becomes 11000000.10101000.01111011.10000100. This number only makes a little more sense, so for most uses, convert the binary address into dotted-decimal format (192.168.123.132). The decimal numbers separated by periods are the octets converted from binary to decimal notation.

For a TCP/IP wide area network (WAN) to work efficiently as a collection of networks, the routers that pass packets of data between networks don't know the exact location of a host for which a packet of information is destined. Routers only know what network the host is a member of and use information stored in their route table to determine how to get the packet to the destination host's network. After the packet is delivered to the destination's network, the packet is delivered to the appropriate host.

For this process to work, an IP address has two parts. The first part of an IP address is used as a network address, the last part as a host address. If you take the example 192.168.123.132 and divide it into these two parts, you get 192.168.123. Network .132 Host or 192.168.123.0 - network address. 0.0.0.132 - host address.

Subnet mask

The second item, which is required for TCP/IP to work, is the subnet mask. The subnet mask is used by the TCP/IP protocol to determine whether a host is on the local subnet or on a remote network.

In TCP/IP, the parts of the IP address that are used as the network and host addresses aren't fixed. Unless you have more information, the network and host addresses above can't be determined. This information is supplied in another 32-bit number called a subnet mask. The subnet mask is 255.255.255.0 in this example. It isn't obvious what this number means unless you know 255 in binary notation equals 11111111. So, the subnet mask is 11111111.11111111.11111111.00000000.

Lining up the IP address and the subnet mask together, the network, and host portions of the address can be separated:

11000000.10101000.01111011.10000100 - IP address (192.168.123.132)
11111111.11111111.11111111.00000000 - Subnet mask (255.255.255.0)

The first 24 bits (the number of ones in the subnet mask) are identified as the network address. The last 8 bits (the number of remaining zeros in the subnet mask) are identified as the host address. It gives you the following addresses:

11000000.10101000.01111011.00000000 - Network address (192.168.123.0)
00000000.00000000.00000000.10000100 - Host address (000.000.000.132)

So now you know, for this example using a 255.255.255.0 subnet mask, that the network ID is 192.168.123.0, and the host address is 0.0.0.132. When a packet arrives on the 192.168.123.0 subnet (from the local subnet or a remote network), and it has a destination address of 192.168.123.132, your computer will receive it from the network and process it.

Almost all decimal subnet masks convert to binary numbers that are all ones on the left and all zeros on the right. Some other common subnet masks are:

Decimal Binary 255.255.255.192 1111111.11111111.1111111.11000000 255.255.255.224 1111111.11111111.1111111.11100000

Internet RFC 1878 (available from InterNIC-Public Information Regarding Internet Domain Name Registration Services) describes the valid subnets and subnet masks that can be used on TCP/IP networks.

Network classes

Internet addresses are allocated by the InterNIC, the organization that administers the Internet. These IP addresses are divided into classes. The most common of them are classes A, B, and C. Classes D and E exist, but aren't used by end users. Each of the address classes has a different default subnet mask. You can identify the class of an IP address by looking at its first octet. Following are the ranges of Class A, B, and C Internet addresses, each with an example address:

- Class A networks use a default subnet mask of 255.0.0.0 and have 0-127 as their first octet. The address 10.52.36.11 is a class A address. Its first octet is 10, which is between 1 and 126, inclusive.
- Class B networks use a default subnet mask of 255.255.0.0 and have 128-191 as their first octet. The address 172.16.52.63 is a class B address. Its first octet is 172, which is between 128 and 191, inclusive.
- Class C networks use a default subnet mask of 255.255.255.0 and have 192-223 as their first octet. The address 192.168.123.132 is a class C address. Its first octet is 192, which is between 192 and 223, inclusive.

In some scenarios, the default subnet mask values don't fit the organization needs for one of the following reasons:

- The physical topology of the network
- The numbers of networks (or hosts) don't fit within the default subnet mask restrictions.

The next section explains how networks can be divided using subnet masks.

Subnetting

A Class A, B, or C TCP/IP network can be further divided, or subnetted, by a system administrator. It becomes necessary as you reconcile the logical address scheme of the Internet (the abstract world of IP addresses and subnets) with the physical networks in use by the real world.

A system administrator who is allocated a block of IP addresses may be administering networks that aren't organized in a way that easily fits these addresses. For example, you have a wide area network with 150 hosts on three networks (in different cities) that are connected by a TCP/IP router. Each of these three networks has 50 hosts. You are allocated the class C network 192.168.123.0. (For illustration, this address is actually from a range that isn't allocated on the Internet.) It means that you can use the addresses 192.168.123.1 to 192.168.123.254 for your 150 hosts.

Two addresses that can't be used in your example are 192.168.123.0 and 192.168.123.255 because binary addresses with a host portion of all ones and all zeros are invalid. The zero address is invalid because it's used to specify a network without specifying a host. The 255 address (in binary notation, a host address of all ones) is used to broadcast a message to every host on a network. Just remember that the first and last address in any network or subnet can't be assigned to any individual host.

You should now be able to give IP addresses to 254 hosts. It works fine if all 150 computers are on a single network. However, your 150 computers are on three separate physical networks. Instead of requesting more address blocks for each network, you divide your network into subnets that enable you to use one block of addresses on multiple physical networks.

In this case, you divide your network into four subnets by using a subnet mask that makes the network address larger and the possible range of host addresses smaller. In other words, you are 'borrowing' some of the bits used for the host address, and using them for the network portion of the address. The subnet mask 255.255.255.192 gives you four networks of 62 hosts each. It works because in binary notation, 255.255.255.192 is the same as 1111111.11111111.1111111.11000000. The first two digits of the last octet become network addresses, so you get the additional networks 00000000 (0), 01000000 (64), 10000000 (128) and 11000000 (192). (Some administrators will only use two of the subnetworks using 255.255.255.192 as a subnet mask. For more information on this topic, see RFC 1878.) In these four networks, the last six binary digits can be used for host addresses.

Using a subnet mask of 255.255.255.192, your 192.168.123.0 network then becomes the four networks 192.168.123.0, 192.168.123.64, 192.168.123.128 and 192.168.123.192. These four networks would have as valid host addresses:

192.168.123.1-62 192.168.123.65-126 192.168.123.129-190 192.168.123.193-254

Remember, again, that binary host addresses with all ones or all zeros are invalid, so you can't use addresses with the last octet of 0, 63, 64, 127, 128, 191, 192, or 255.

You can see how it works by looking at two host addresses, 192.168.123.71 and 192.168.123.133. If you used the default Class C subnet mask of 255.255.255.0, both addresses are on the 192.168.123.0 network. However, if you use the subnet mask of 255.255.255.192, they are on different networks; 192.168.123.71 is on the 192.168.123.64 network, 192.168.123.133 is on the 192.168.123.128 network.

Default gateways

If a TCP/IP computer needs to communicate with a host on another network, it will usually communicate through a device called a router. In TCP/IP terms, a router that is specified on a host, which links the host's subnet to other networks, is called a default gateway. This section explains how TCP/IP determines whether or not to send packets to its default gateway to reach another computer or device on the network.

When a host attempts to communicate with another device using TCP/IP, it performs a comparison process using the defined subnet mask and the destination IP address versus the subnet mask and its own IP address. The result of this comparison tells the computer whether the destination is a local host or a remote host.

If the result of this process determines the destination to be a local host, then the computer will send the packet on the local subnet. If the result of the comparison determines the destination to be a remote host, then the computer will forward the packet to the default gateway defined in its TCP/IP properties. It's then the responsibility of the router to forward the packet to the correct subnet.

Troubleshooting

TCP/IP network problems are often caused by incorrect configuration of the three main entries in a computer's TCP/IP properties. By understanding how errors in TCP/IP configuration affect network operations, you can solve many common TCP/IP problems.

Incorrect Subnet Mask: If a network uses a subnet mask other than the default mask for its address class, and a client is still configured with the default subnet mask for the address class, communication will fail to some nearby networks but not to distant ones. As an example, if you create four subnets (such as in the subnetting example) but use the incorrect subnet mask of 255.255.255.0 in your TCP/IP configuration, hosts won't be able to determine that some computers are on different subnets than their own. In this situation, packets destined for hosts on different physical networks that are part of the same Class C address won't be sent to a default gateway for delivery. A common symptom of this issue is when a computer can communicate with hosts that are on its local network and can talk to all remote networks except those networks that are nearby and have the same class A, B, or C address. To fix this problem, just enter the correct subnet mask in the TCP/IP configuration for that host.

Incorrect IP Address: If you put computers with IP addresses that should be on separate subnets on a local network with each other, they won't be able to communicate. They'll try to send packets to each other through a router that can't forward them correctly. A symptom of this problem is a computer that can talk to hosts on remote networks, but can't communicate with some or all computers on their local network. To correct this problem, make sure all computers on the same physical network have IP addresses on the same IP subnet. If you run out of IP addresses on a single network segment, there are solutions that go beyond the scope of this article.

Incorrect Default Gateway: A computer configured with an incorrect default gateway can communicate with hosts on its own network segment. But it will fail to communicate with hosts on some or all remote networks. A host can communicate with some remote networks but not others if the following conditions are true:

- A single physical network has more than one router.
- The wrong router is configured as a default gateway.

This problem is common if an organization has a router to an internal TCP/IP network and another router connected to the Internet.

6. Storage basics

Introduction to Storage

Objects, blocks and files are different storage architectures that store and present the data in different ways based on requirements. Each has its benefits and limitations. Given below is a brief on each type of storage and when the storage architecture is typically used.

What are storage devices?

Any device or medium which is either permanently attached or is movable/transportable, capable of storing information in an electronic form can be referred to as a storage device. Some examples are hard disk drive, CD ROM, flash media, DVD ROM, memory stick, etc. Devices such as iPod, PDA, mobile phones, etc. also contain storage devices as part of the hardware.

Classification of storage devices

Two types of storage devices are used with computers. **The first type is referred to as primary storage and comprises of the RAM or the internal storage.** The RAM stores data that is directly accessible by the computer's processor (CPU). RAM is commonly referred to as 'memory'. The primary memory is temporary and is used to store program instructions and intermediate results of procedures.

The secondary storage device may be internal, external or removable. An example of secondary storage is the hard disk drive. It is a device that is non-volatile. Though it may be located inside a computer, this type of storage is not considered primary because it cannot be directly accessed by the CPU. Data on a hard disk drive is organized in the form of files. This affords slower access and is cheaper.

Tertiary storage is typically not located inside a computer. These are usually high-capacity storage devices that are designed to hold large amounts of data. An optical disk is an example. This type of device includes a mechanism to locate specific data and transfer it to a drive when requested.

Different Types of Storage Devices

Any computer user ends up using many different types of secondary storage devices.

- The internal or external hard disk drive is usually connected to the computer for retrieval of stored data using an interface such as a USB cable.
- Computers of today have the disk drives as externally connected devices. These can be easily removed and stored elsewhere. These devices hold the stored data unless explicitly deleted or overwritten using instructions from the computer.
- Until a few years ago, floppy disks, magnetic tapes, and other magnetic media were used as secondary storage devices and were popular. These devices used the principle of a read/write head magnetizing material that was coated on the disk to store information.
- Once these became cumbersome to use, storage device manufacturers started looking at optical storage devices. Such a device is written onto and read from using a laser beam. The marked advantage that they have is the capability to store large amounts of data.
- The different types of optical memory are CD-ROM, DVD-ROM, CD-R/W, etc. With every new software or application that is being developed, the demand for storage is greater.
- Flash drives also referred to as thumb drives are the most popular portable storage devices of today. They are compact and connect with the help of a USB port.
- Memory cards are used in digital cameras or mobile phones are also secondary storage devices. These can be transferred to a computer using a reader that is connected through a USB port.
- Some computers have installed solid-state drives (SSD) in place of hard disks. An SSD is advantageous in that it has no movable parts like the hard disk. Their costs are now very competitive with the result that they are being used more in computers now.

- **Cloud storage refers to the technology that uses data stored in a remote location. Cloud storage uses devices located elsewhere and are connected using the Internet or other types of network connection.** The cloud storage provider manages the devices and provides data backup too. Many users appreciate the managed services provided by cloud operators.

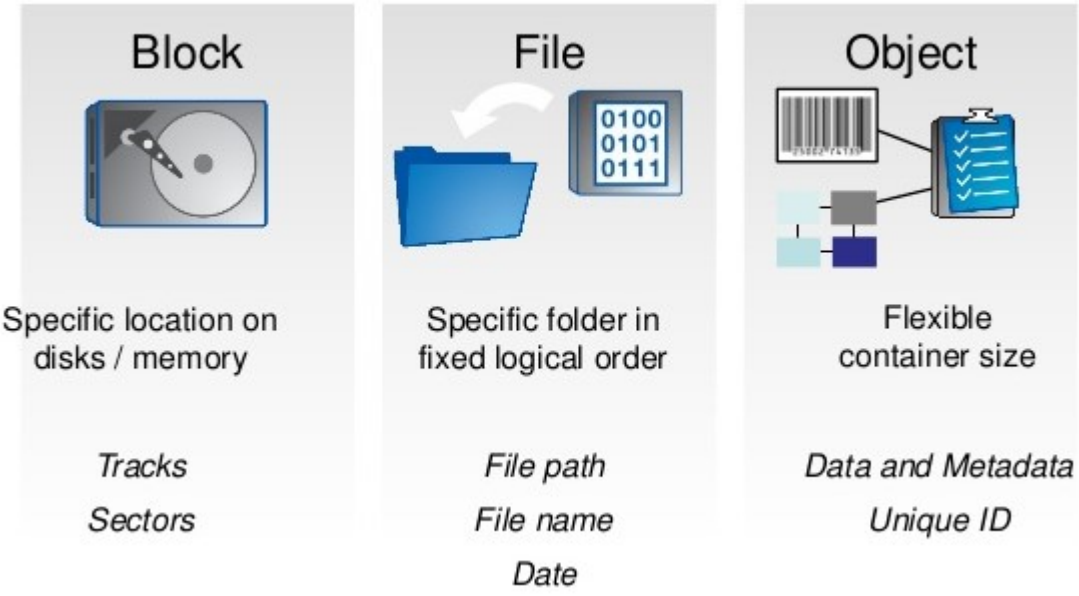


Image Credit: trigent.com

Object Storage

Object storage or object-based storage refers to the data storage architecture that manages stored data in the form of objects and not as files. The units of data storage called objects consist of:

1. **Data:** The stored data is organized into objects. These may be complete files or parts of a file (sub-file). They may also simply be a collection of bits and bytes that are related but nor part of any file.
2. **Metadata:** The metadata is created by the one who creates the object storage. The metadata is made up of information that is contextual, i.e., what the data is all about, what it should be used for, any other information about how it is used to be used, and its confidentiality.
3. **A unique identifier:** This is a unique piece of information (usually 128 bits) that is tagged along to be able to find out this data over a distributed system. The physical location of the data becomes irrelevant.

Object Data Storage Implementation

Once comprehensive metadata is added to the file, it is placed in a flat address space referred to as a storage pool. The comprehensive metadata is what makes object storage successful as it provides information about both the use and function of the data lying in the storage pool.

Benefits of Object Storage

Some of the benefits of object storage are as follows:

- **Data Analytics:** As object storage is driven by metadata, and because metadata for every piece of data is available, the opportunity for analyzing data becomes much greater.
- **Infinite Scalability:** You can keep on adding any amount of data to the pool.
- **Quicker Retrieval:** As there is no folder hierarchy, data can be retrieved quickly.
- **Less costly:** It is a cheaper method to store large amounts of data.
- **Resource Optimization:** Resources can be used at an optimal level with object storage of data.

Object storage works best for cases such as web content, archives and data backup. The flat address space facilitates ease of use. Objects that are protected are built into the object architecture by using multiple copies (at least three) of the data over a distributed system. If any node fails, the data can still be made available to the user by using the copies. Corruption of data is also overcome in much the same way.

Object storage always is not the answer. You, as a user, have to decide as to which storage architecture suits your needs the best.

Object Storage Providers

Storing a large amount of data in a traditional environment or by a single user is expensive. There are many object storage providers that offer cloud storage services. We list some of them here.

AWS S3: They are one of the pioneers in object storage technology. You can upload any amount of data and they will allow you to upload/download using API, the browser. They promise 99.99% durability of data. They offer the Standard, Standard IA and Glacier types of storage. You can move your data from one type of storage to another. They allow for storage of data on the basis of region and this means you can have faster content distribution when you keep your data near your customer. Any client can start with free storage offer of 5 GB.

Google Cloud Storage suits players in the small size to enterprise level of business. They offer four types of storage: **Multi-Regional, Regional, Nearline, and Coldline**. You can choose the one that suits you depending on the frequency of access to data and from which part of the world it is accessed more frequently. They have multiple data centres in the world and you can store your data in any city that you want. You also get to enjoy free tier storage of 5 GB.

Spaces by DigitalOcean, offers a two-month trial period. In New York and Amsterdam, where they operate data centres from, the cost is $5 monthly for 250 GB storage with an additional $0.20 per GB. For fixed-rate monthly fares, they are among the cheapest service providers.

Other players in this segment are Cloud Files by Rackspace, IBM Cloud Object Storage, Alibaba OSS, Microsoft Azure storage, and Oracle Storage. Whereas Azure has multiple data centres across different places in India, Oracle allows for different ways to connect with the data. Alibaba vouches for 99.9% availability guarantee and can handle over 50000 requests per second.

Object Storage Vendors

Object storage vendors deploy storage for regulatory archiving in the financial, insurance, legal, and healthcare sectors. Others use it for large-scale web files high-performance video, computation and entertainment purposes.

Some of the leading Object Storage Vendors are IBM Cloud Object Storage, Tarmin GridBank Enterprise, Dell EMC ECS, Quantum Lattus, NooBaa, NetApp, StorageGrid, Hitachi Content Platform, Caringo, and Cloudian HyperStore among others.

Object storage is continuously evolving in a manner similar to the way in which file accessed storage continues to evolve.

Object Storage API

A user can create, get and modify objects and metadata using Object Storage API. This is implemented as a set of Representational State Transfer (REST) web services. The HTTPS (SSL) protocol can be used to interact with Object Storage. Standard HTTP calls can be used to perform API operations. For using APIs that are language-specific, you can use RESTful API.

To change data, the user is required to authenticate oneself with a token. The token can be obtained from an authentication service after presenting the user's credentials. On clearance, the service returns a token and an account URL.

Object Storage API supports a non-serialized response format. The Object Storage system organizes data like accounts, containers and objects through the service provider. The service provider creates your account and you become the owner of the account. It is made up of all the containers. The container is the namespace for objects. You can make use of an access control list (ACL) to control access to the objects in a container. The object stores the data content such as images and documents.

Object Storage API allows a user to perform the following functions in addition to many others.

- Store as many objects as desired
- Upload and store objects of any size
- Compress files
- Manage object security
- Schedule deletion of objects

This list is not exhaustive.

Object Storage AWS

In the object storage arena, Amazon Web Services (AWS) has many product offerings. It is useful to have a basic idea to know which product to use for what purpose. S3, EBS, and EFS are three products that work differently.

Amazon S3 (Simple Storage Service) offers both high-level data scalability performance and security. Customers across different industries can use this service to store any amount of web content, mobile app, enterprise application, IoT device, big data, and archive data. The management features of S3 are simple and easy to use. This enables users to easily access stored data and tune according to their business needs. AWS S3 offers 99.99% data durability for millions of companies worldwide.

File Storage

Types of File Storage

Network storage devices are typified in the way they are interfaced on the client-side. Multiple clients are able to access a single shared folder in a traditional file-sharing system. The two popular protocols that enable this system are NFS and SMB/FICS.

The file system offers the simplest architecture for storage systems. When the amount of data grows larger, the resource demands grow and cannot be compensated by simply adding storage space.

The files are then organized into directories and sub-directories. Naming conventions make it easy for them to be organized. Most file storage systems allow for a centralized and easy retrievable system for the data. The cost is reasonable.

The file system works best when the amount of data is small and stored on personal computers and servers used in medium to large enterprises or workplaces. They can typically be used in conditions such as an office where you need to store and share files.

Locally archived files can use file types of storage systems with a NAS (Network Attached Storage) solution. The data centre in such a case is typically small. This type of storage architecture can also be used to protect data. This is supported by the use of standard protocols, different drive technologies, and native replication methods.

File Storage Containers

The data is stored in containers the form of files, given a name and tagged with metadata which consist of file creation date, modification date and the size of the file.

Data stored in files are retrieved using very little metadata. This metadata informs the computer where exactly the file is located. Every file is arranged following a specific hierarchy, by directories and sub-directories. This arrangement is well supported by NAS systems.

In cloud file storage, data is stored in the cloud. Other applications and servers can access this data through shared file systems. Users can create, edit, delete read and write files or even organize them in a logical fashion using directory trees. Cloud file-systems allowing shared access controls security with user and group permissions.

Block Storage

Block storage can be considered as the alternative to file storage. A block is a volume that is filled with files which are further split into blocks of equal size. Usually, such blocks can be well managed by a server-based operating system.

Each of these chunks of data can be managed as individual hard drives. There are many third-party applications used by organizations that help to manage data in block storage architectures.

Block storage is known to handle metadata very efficiently and the operating system allocates the storage for various applications and decides where the data is going to reside in the block. The control is efficient and accounts for the high performance that block storage architecture offers.

Block storage is used for a wide range of database applications that require high-performance levels, email servers that do not support file systems or network-based systems, virtual machines that use guest operating systems, etc.

Block-based architecture can be expanded as the volume of data grows; however, the integrity of such systems is likely to be tested when the volume grows into terabytes or petabytes of storage.

Block Storage Providers

Most of the major enterprise-level storage providers offer San products/block storage solutions. Top providers include Dell EMC, Hitachi Data Systems, IBM, HPE, NetApp, etc. In the cloud, AWS Elastic Block Store (AWS EBS) are providers of scalable block storage usable by EC2 (Elastic Cloud Compute) instances. Therefore, all those applications that run on SAN can place their databases, applications and workloads on Amazon's cloud.

Block Storage Vendors

Block storage vendors include Huawei, DataDirect Networks, Nutanix, Oracle, etc., in addition to the largest storage providers. These vendors have several block storage platforms. Some of them provide unified arrays for both block and file storage.

Block Storage in AWS

Amazon EBS (Elastic Block Store) is the block storage offered by Amazon Web Services. This is used to store persistent data. Highly available block storage volumes are provided for EC2 instances making it suitable for the same. The three types of volume offered are EBS General Purpose (SSD) suitable for small and medium workloads, Provisioned IOPS (SSD) which suits I/O intensive transactional workloads, and magnetic for infrequent access of data. The cost, performance and characteristics vary for the three types.

The benefits of using Amazon EBS include fine access control with encryption, reliable and secure storage, usage of SSD technology for higher performance, and easy data backup.

With the advancement of technology, the amount of data has grown exponentially and every major enterprise is looking for the best ways to store its data. Despite the fact that the IT applications of today are trending towards object-based storage, file and block systems are still used extensively. The point to start at is to identify individual needs appropriately and choosing the best-fit storage architecture.

ApacheBooster is a must-have plugin for those who need to increase their server's speed in a short span of time. It is cheaper and optimum in functionality compared to other plugins in the market.

One of your biggest concerns as an IT professional is determining what type of storage to use and for what types of data use cases – mobile applications, databases, websites, files, or backing up mission-critical data. Odds are that you will probably use a combination of data storage types to meet the needs of your users and the requirements of your data.

Data storage is used for a multitude of reasons. If you're developing an application, you may have users that upload documents, photos, videos or other files. You'll need somewhere to store user files. If you're a developer, you may use a content delivery network (CDN) and data storage to increase load speed, availability and reliability. If you're in charge of IT, your main concern may be storage and backup for disaster recovery and business continuity.

Understanding data storage is not hard but all of the different types and options can be confusing, especially if you're not an IT professional. In this article, we will discuss the different types of data storage along with the advantages and disadvantages of each plus use cases.

Direct Attached Storage (DAS)

Most people are familiar with direct attached storage (DAS) whether they know it or not. This is because most have already used it. In fact, if you have a laptop in front of you, there's a DAS hard drive within that laptop. It's called direct-attached storage because it is directly attached. DAS could also be an external hard drive or thumb drive to a computer, laptop or tablet.

Advantages of Direct Attached Storage (DAS)

What are the benefits of DAS? Direct attached storage is great because it's very cheap and very easy to use. In fact, you can purchase a 6TB external hard drive for as low as a few hundred dollars. DAS is extremely cheap for what you're getting. The price per GB is very low and pricing continues to trend downward for these types of storage devices.

Disadvantages of Direct Attached Storage (DAS)

What's the downside of direct attached storage? The main downside to DAS is that it is not very shareable. If you wanted to share your data with someone else, you would have to either upload it from your computer or

laptop to the cloud, send through email as an attachment or physically go over to that person's computer to share it.

DAS is definitely not useful in all business cases. This is one reason why direct attached storage is not used in cloud environments. Can you imagine your employees walking around with external hard drives and plugging them into virtual machines or the servers that they're running in their data centers?

Network Attached Storage (NAS)

The next data storage type we are going to discuss is network attached storage (NAS). Many descriptions of NAS make it seem like it is very complex. However, it's really not that hard to understand. There are three basic components associated with NAS.

First, NAS has to have connectivity to the internet and to a local area network (LAN). Second, you must have multiple hard drives attached to the NAS. Third, the hard drives have to be configured into what we call a RAID configuration. RAID is a redundant array of independent disks – hard drives are setup to replicate the data in various ways.

What's an example of RAID? Let's say that you have four hard drives and two of them are replicated amongst each other. The other two are replicated amongst each other. You set those hard drives up to store data independently. It may be partially replicated or fully replicated. With RAID, you have the assurance that if one of the drives fails, you're not going to lose all of your data.

Advantages of Network Attached Storage (NAS)

Although network attached storage NAS is more expensive than DAS, it's still pretty cheap. Further, network attached storage is great for collaboration. For example, let's say that your company has a lot of files and numerous employees working on those files, NAS could be the right data storage type for your business.

A use case would be a business with 10-50 employees that all need to access and edit files on the same hard drive. NAS is perfect for that scenario because when you see it on your computer, it'll show up as a single file on the shared drive.

NAS provides centralized control of all the files. You're able to set permissions on who can see what in the network. And finally, as we talked about earlier with the RAID configuration, you can replicate data and make sure that you have backups of that data.

Disadvantages of Network Attached Storage (NAS)

Performance can be a major issue for NAS. This means that if you have a lot of activity on your network, it's going to slow down the performance. Further, with low throughput and high latency, a NAS is not fast enough for high-performance applications.

NAS can also be limited from a scalability perspective. NAS is limited to its own resources and you can only scale by adding another NAS. This becomes even more complicated with NAS sprawl – too many devices.

Storage Area Network (SAN)

What is a storage area network (SAN)? Simply put, a SAN is a high-speed network that provides block-level network access to connect servers to their logical disk units (LUNs). LUNs consist of a range of blocks provisioned from a pool of shared storage and presented to the server as a logical disk. Server partitions and formats those blocks with a file system so it can store data on the LUN just as it would on local disk storage.

SANs make up about two-thirds of the total networked storage market. They are designed to remove single points of failure, making SANs highly available and resilient. A well-designed SAN can easily withstand multiple components or device failures.

However, SANs are fairly complex infrastructure with hosts, switches, storage elements, and storage devices that are interconnected using a variety of technologies, topologies, and protocols. Storage area networks may also span multiple sites or locations.

Advantages of Storage Area Network (SAN)

There are numerous benefits associated with storage area networks. SANs are often used to improve application availability with multiple data paths. SANs enhance application performance such as off-loading storage functions. They also increase storage utilization and effectiveness by consolidate storage resources and provide tiered storage.

Disadvantages of Storage Area Network (SAN)

There are a few disadvantages to the storage area network (SAN). First, SANs are very expensive. It's not ridiculously expensive if you're using it for cloud computing. However, if you try to set it up on your own, it's going to be very expensive. SANs are also complex and difficult to setup. That would be the downside. However, it's probably one of the more common types of storage that you'll see in cloud computing.

Block Storage

What is block storage? If you look at a hard drive, it's basically a block storage device. This means that the hard drive is broken up into partitions which is stores files on a file system in as little as 512-byte blocks. For example, you can run the EXT4 file system on partition one and three which is basically the Linux file system. You can run an Apple based file system on partition two and a Windows NTFS file system on partition four. That's block storage in a nutshell.

If you're running Windows, you could access this partition for all of your Windows files. For example, you have a Microsoft Excel file that is 200 kilobytes or 200,000 bytes. If each block is 512 bytes, this particular Excel file requires about 62 ½ blocks.

One of the best things about block storage is that when you edit your Excel file, maybe you make changes to cells 1-84. It's going to affect a few blocks within that entire file. If the file is 62 ½ blocks, you only have to edit four of them. When you save the file, it's going to only find the four blocks that it needs to edit. You don't have to replace it all at once so this makes block storage very efficient in that way.

Advantages of Block Storage

There are many benefits associated with block storage. First, their numerous programming languages can easily read and write files on block storage. Also, permissions and access controls for block storage are familiar and well-understood. Lastly, block storage provides low latency IO so they can be used with databases and dynamic data.

Disadvantages of Block Storage

There are also several disadvantages of block storage. Block storage is limited to one server at a time which impacts scalability. Further, blocks and filesystems have limited metadata about the information they're storing such as creation date, owner, size and more. Another major disadvantage of block storage is the cost structure. With block storage, you must pay for all of the block storage space you have allocated even if you are not using it.

Object Storage

The last data storage option we will be discussing in this article is called object storage. Object storage is very different from other storage types we've talked about. In fact, it's in a whole different realm. It's also a newer type of file storage and therefore it works differently.

First, you have objects rather than files or blocks. Essentially, you have unstructured data objects that have three parts. They have an ID. They have metadata such as the authors of the file. They have the date that the file was created, permissions on the file, so on and so forth.

Imagine you have a large quantity of unstructured data. It could be a picture or a large video file. Every time you update that file, you have to add the entire file. If you want to make any changes to your video you would have to create an entirely new object and you could have different versions but completely different files or objects. With object storage, you can't do piecemeal operations and it is best for very specific use cases – including storing lots of unstructured data.

With object storage, you write once and read many times. An example of this would be YouTube videos. Once an author uploads it, they really can't edit it and they can't change it much. That's a perfect use case for object storage. If you want to edit it, the user must upload a new video and delete the old one. You could also keep different versions of the same video file with slight variations which is versioning.

Advantages of Object Storage

There are many benefits associated with object storage including scalability. It's widely known for its compatibility with the cloud and that's because it has unlimited scalability. Because of its flat structure, object storage doesn't have the same limitations as file or block storage.

Object storage has faster data retrieval and better recovery than other types of data storage. With object storage, there's no need sift through file structures which means faster retrievals. The metadata allows for quick access and fewer limitations.
Lastly, object storage is known for being cost-effective. Because object storage scales out much easier than other storage types, it's less costly to store all your data.
Disadvantages of Object Storage
Are there disadvantages to object storage? Yes, object storage isn't right for every use case when it comes to data storage. In fact, you can't use object storage for traditional databases. It's only great for static data.
Another disadvantage of object storage that we discussed was that it doesn't allow you to alter just a piece of data. You must read and write the entire object at once.
Types of Storage devices

Local Storage Options
1. External Hard Drive
These are hard drives similar to the type that is installed within a desktop computer or laptop computer. The difference being that they can be plugged in to the computer or removed and kept separate from the main computer. They typically come in two sizes :
Desktop External Hard drive : Uses a 3.5 inch hard drive similar to that used in desktop computers.
Portable External Hard drive : Uses a 2.5 inch hard drive similar to that used in laptops.
Desktop External Hard Drives are generally cheaper than Portable External Hard Drives for the same storage space. Desktop External Hard Drives and usually faster and more robust.
Capacity : 160GB to 3TB (approx 3000GB)
Connection : Most common connections to the computer are through a USB 2.0 or USB3.0 connection. May also be available in a SATA or eSATA connector
Advantages :
- Very good option for local backups of large amounts of data.
- The cheapest storage option in terms of dollars per GB. Very reliable when handled with care

Disadvantages :
- Can be very delicate. May be damaged if dropped or through electrical surge

---

2. Solid State Drive (SSD)
Solid State Drives look and function similar to traditional mechanical/ magnetic hard drives but the similarities stop there. Internally, they are completely different. They have no moving parts or rotating platers. They rely solely on semiconductors and electronics for data storage making it a more reliable and robust than traditional magnetic. No moving parts also means that they use less power than traditional hard drives and are much faster too.
With the prices of Solid State Drives coming down and is lower power usage, SSD's are used extensively on laptops and mobile devices. External SSD's are also a viable option for data backups.
Capacity : 64GB to 256GB
Connections : USB 2.0/3.0 and SATA
Advantages :
- Faster read and write performance
- More robust and reliable than traditional magnetic hard drives
- Highly portable. Can be easily taken offsite

Disadvantages :
- Still relatively expensive when compared to traditional hard drives
- Storage space is typically less than that of traditional magnetic hard drives.

---

3. Network Attached Storage (NAS)

NAS are simply one or more regular IDE or SATA hard drives plugged in an array storage enclosure and connected to a network Router or Hub through a Ethernet port. Some of these NAS enclosures have ventilating fans to protect the hard drives from overheating.

Advantages :

- Very good option for local backups especially for networks and small businesses.
- As several hard drives can be plugged in, NAS can hold very large amounts of data
- Can be setup with Redundancy (RAID) increasing the reliability and/ or read and write performance. Depending on the type of RAID level used, the NAS can still function even if one hard drive in the RAID set fails. Or two hard drives can be setup to double the read and write speed of single hard drive.
- The drive is always connected and available to the network making the NAS a good option for implementing automated scheduled backups.

Disadvantages :

- Significantly more expensive than using single External Hard Drives
- Difficult to bring offsite making it very much a local backup hence still susceptible to some events like theft and floods, fire etc

---

4. USB Thumb Drive Or Flash Drive

These are similar to Solid State Drives except that it is much smaller in size and capacity. They have no moving parts making them quite robust. They are extremely portable and can fit on a keychain. They are Ideal for backing up a small amount of data that need to be brought with you on the go.

Capacity : 4GB to 64GB

Advantages :

- The most portable storage option. Can fit on a keychain making it an offsite backup when you bring it with you.
- Much more robust then traditional magnetic hard drives

Disadvantages :

Relatively expensive per GB so can only be used for backing up a small amount of data

---

5. Optical Drive (CD/ DVD)

CD's and DVD's are ideal for storing a list of songs, movies, media or software for distribution or for giving to a friend due to the very low cost per disk. They do not make good storage options for backups due to their shorter lifespan, small storage space and slower read and write speeds.

Capacity CD : 650MB to 900MB

Capacity DVD : 4.7GB to 17.08GB

Advantages :

- Low cost per disk

Disadvantages :

- Relatively shorter life span than other storage options
- Not as reliable as other storage options like external hard disk and SSD. One damaged disk in a backup set can make the whole backup unusuable.

Remote Storage Options

6. Cloud Storage

Cloud storage is storage space on commercial data center accessible from any computer with Internet access. It is usually provided by a service provider. A limited storage space may be provided free with more space available for a subscription fee. Examples of service providers are Amazon S3, Google Drive, Sky Drive etc.

Advantages :

- A very good offsite backup. Not affected by events and disasters such as theft, floods, fire etc

Disadvantages :

- More expensive then traditional external hard drives. Often requires an ongoing subscription.

- Requires an Internet connection to access the cloud storage.
- Much slower than other local backups

IDE and **SATA** are different types of interfaces to connect storage devices (like hard drives) to a computer's system bus. **SATA** stands for **Serial Advanced Technology Attachment** (or **Serial ATA**) and IDE is also called **Parallel ATA** or **PATA**. SATA is the newer standard and SATA drives are faster than PATA (IDE) drives. For many years ATA provided the most common and the least expensive interface for this application. But by the beginning of 2007, SATA had largely replaced IDE in all new systems.
The ATA interface (ATA stands for AT attachment where "AT" refers to IBM's PC/AT for which it was originally built) evolved in stages from Western Digital's original Integrated Drive Electronics (IDE) interface. After the introduction of Serial ATA in 2003, the original ATA was retroactively renamed Parallel ATA.

Comparison chart

IDE versus SATA comparison chart

	IDE	**SATA**
Advantages	Maximum compatibility	Inexpensive, large storage capacity.
Disadvantages	Lacks support for new technology such as native command queuing and hot-plugging hard drives	Lower MTBF than SAS (700,000 hours to 1.2 million hours of use at 25 °C), less suited for servers.
Hot plugging (add/remove component while the computer is running)	IDE interface does not support hot plugging	SATA interface supports hot plugging
Speed	data transfers at the rate of up to 133MB/s	Data transfers at the rate of up to 6 Gb/s
Data cable	Ribbon-like, wide, can be up to 18 inches long	Narrow, can be up to a meter (roughly 3ft) long. Power and data split into two connections.
Lineage	Superseded by SATA	Supersedes Parallel ATA (PATA) aka IDE
Year Created	1986	2003
Jumpers	In a computer system, it's possible to have more than one harddrive. To connect multiple IDE drives, you need to chain the ribbon cables from one to the next. The computer system has no idea which is the main drive, from which to load the OS.	SATA drives don't use jumpers. Each drive connects directly to the motherboard. To set the primary drive, you can access the settings from the computers BIOS (special software that runs when you start the computer).

Differences in Cables and Connectors

SATA (right) and IDE (left) hard drives. The SATA hard drive has the data cable on the right and power cable on the left. The IDE data cable is ribbon-like (on the left)

Hard drives need a cable/connection for data and one for power. Parallel ATA only allows data cable lengths up to 18 in (457 mm) while SATA allows cable lengths up to 1 m (3.28 ft). eSATA cables can be 2 m in length.

IDE consists of a 40-pin connector attached to a ribbon cable. 80-pin connectors were also introduced later. The connectors are black in a 40-pin connector while in an 80-pin connector, they come in 3 colors: blue - controller, gray - slave drive, and black - master drive. Each cable has two or three connectors, one connector is attached to the interface that connects to the computer system (mother board) and the others are connected to the drives.

SATA consists of an 8 mm wide wafer connector on each end and the cable has a 7-pin connector, 3 grounds and 4 active data lines in two pairs. It has the facility to attach only one drive and so Serial ATA does away with Master/Slave configurations.

What is a SCSI Hard Drive?

SCSI is an acronym for Small Computer System Interface, pronounced “scuzzy”. SCSI hard drives have been the backbone of enterprise computing for nearly 20 years. Though they typically don’t possess much in the way of capacity (the last generation of SCSI drives consisted mostly of 36GB, 73GB, and 146GB models), SCSI drives make up for it with speed.

SCSI drives come in 10,000 or 15,000 rotations per minute (RPM) versions, meaning it will access data much faster than your desktop will (desktop drives are generally 5400 or 7200 RPM). If you have mission-critical applications that aren’t too big but need to be accessed quickly, SCSI is a great choice.

SAS Hard Drive – A Better Choice for Mission Critical Applications

SAS, which stands for Serial Attached SCSI, is basically a beefed-up version of a SCSI drive. For mission-critical applications, an SAS hard drive is the better choice.

SAS drives have higher transfer speeds (3 or 6Gbit/s, as opposed to a maximum of 5120 Mbit/s for SCSI), thinner cables, and are more easily linkable with SATA drives. They also come in more form factors – all SCSI drives are 3.5”, but SAS drives can be 2.5”, allowing for their use in more compact systems.

SAS drives also come in larger capacities (they go up to 600GBs and beyond, whereas SCSI stops at 300GB), while maintaining the 10K and 15K RPM speeds. Naturally, though, the tradeoff is that SAS drives cost more than SCSI ones. Still, for important applications that require real-time access, SAS is the new SCSI.

The SATA Hard Drive

Then there are SATA drives. SATA (or Serial ATA, which stands for Serial Advanced Technology Attachment) is the interface used by most desktop and laptops on the market today.

That doesn’t mean that you can take an HDD out of your old desktop and slot it into a server, though – servers use special Enterprise-class SATA drives that are faster and more reliable. Even so, Enterprise SATA drives are going to be slower than a SCSI or SAS drive, only going up to 7200 RPM.

They make up for this in capacity, however – the current generation of Enterprise SATA drives don’t go much lower than 250GB and can go as high as 2TB.

If you're looking for drives with a lot of room and don't want to pay a huge premium for SAS-level speed, an **Enterprise SATA hard drive** is the way to go.

SATA vs SCSI vs SAS: The Takeaways

It's no surprise, but different interfaces are good for different things:

- If you need **speed** and **transfer rate**, SCSI is a good choice, and SAS even better.
- If **capacity** is your main concern, SATA is a better option.
- For **price**, SATA is once again king based on a pure specs-to-cost ratio, though for a mix of speed and cost, SCSI drives can be very affordable as well.

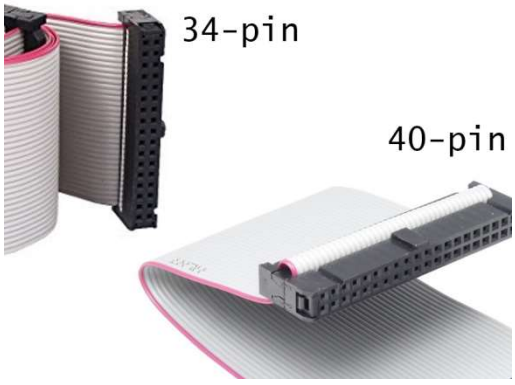
The bottom line though is that hard drive interface is just one choice to make when **picking a server**. Memory, CPU, number of drive bays, RAID, remote access control, etc. … all these must be factored in, and oftentimes your choices there will dictate your **choice of hard drives**.

The biggest example of this is **compatibility** – a server that will take SCSI drives will only take SCSI drives, whereas a server that can take SAS will be able to take SATA drives (with a few exceptions).

Topic	ATA/IDE	SCSI
Cost	Overall, IDE is a much cheaper solution.	Compared with IDE, SCSI is often more expensive t
Expansion	IDE/EIDE allows 2 two devices per channel. Most computers have 2 channels.	SCSI is capable of supporting up to 7 or 15 devices.
Ease	IDE is commonly an easier product to set up than SCSI.	Configuring SCSI can be more difficult for most use
Faster	Today, the latest IDE and SCSI drives running at the same RPM are very close. However, 10,000+ RPM drives are often only available for SCSI.	All the fastest drives are often available for SCSI RPM hard drives are only available as SCSI drives.
Resources	All motherboards today have an ATA/IDE interface and unless additional drives are needed no additional resources need to be taken.	Unlike IDE, SCSI requires an interface expansion motherboard already has it). Adding any new resources are going to be required.

SCSI PINOUTs

IDE / PATA PINOUTs



Listed below is a quick information about the total number of SCSI ID's allowed on a given SCSI chain:

The list below subtracts the address of the SCSI card.
The list below is specific to the maxiumum device ID addressing only.
(See below for additional information on cable lengths).

SCSI-1		=	7	Devices
SCSI-2	Fast	Narrow	=	7 Devices
SCSI-2	Fast	Wide	=	15 Devices
SCSI-2	Differential	Narrow	=	7 Devices
SCSI-2	Differential	Wide	=	15 Devices
SCSI-3	Ultra	Narrow	=	7 Devices
SCSI-3	Ultra	Wide	=	15 Devices
SCSI-3	Ultra2	=	15	Devices
SCSI-3	Ultra3	=	15	Devices

To answer the question completely it will depend on the specific SCSI controller used and the type of SCSI device you will be using.

There are some types of SCSI devices that can address up to 15 ID numbers with jumpers, but this is ultimately limited to the cable distance which can make it impossible to achieve total of 15 devices on the SCSI chain of a controller.

At the time of this article there are three types of pin connection on the SCSI bus Adaptec supports - the "25-pin", "50-pin" and "68-pin" connections.

- 25-pin or 50-pin (also known as "narrow" or the 8-bit SCSI bus) controllers and devices can address an ID value from 0 to 7. The controller requires an ID address which defaults to ID # 7, leaving an ID address space from 0 to 6. This means a total of 7 SCSI device on the chain can be addressed.

- 68-pin (also known as "wide" or the 16-bit SCSI Bus) controllers and devices can address an ID value from 0 to 15. The controller also requires an ID address which usually defaults to ID # 7, leaving an ID address space from 0 to 6 and from 8 to 15. This means a total of 15 SCSI devices on the chain can be addressed

Cables and cable length Standards:

Currently there are three types of SCSI electrical signals that Adaptec supports:

1. "SE" (Single Ended). Also called "Legacy / SE" can be a 25-pin, 50-pin or 68-pin connector.

The Legacy / SE bus segment supports the following cable lengths:

SCSI-1	Synchronous or Asynchronous,	25 or 50-pin	-	6 Meters
SCSI-2	Fast	Narrow, 50-pin	-	3 Meters
SCSI-2	Fast	Wide, 68-pin	-	3 Meters

SCSI-2 Ultra Narrow, 50-pin - 1.5 Meters with four or more devices on the SCSI chain or 3 Meters for three or less devices on the SCSI chain.
SCSI-2 Ultra Wide, 68-pin - 1.5 Meters with four or more devices on the SCSI chain or 3 Meters for three or less

devices on the SCSI chain.

2. "LVD" (Low-Voltage Differential). Also known as "Ultra2" or "Ultra160" or "Ultra320".

- Can only communicate in LVD mode with a 68-pin connection.

The Ultra2/Ultra160 LVD bus segment supports the following cable lengths:

(Applies for LVD Ultra2/Ultra160 and 68-pin only)
- 25 Meters point-to-point cable length with one device only.
- 12 Meters with 2 devices or more.

Note:
Avoid attaching an LVD device and SE device on the same SCSI chain as it will slow down the LVD devices.

3. "HVD" (High Voltage Differential).

- Can communicate in HVD mode with either a 50-pin or 68-pin connection.
- Can not be mixed with SE or LVD devices (see .

(The following applies for wide or 68-pin only)
- 25 Meters point-to-point cable length with one device only.
- 12 Meters with 2 devices or more.

Note:
Don't attach an HVD device or controller on a LVD/SE controller or device as irreparable damage can occur to either non-HVD devices or the HVD controller.

Following the explanation above about SE, LVD and HVD, it is good to know that a minimum distance between SCSI devices (counting the controller) is 30 cm (or 11.7 inches).

Question:

How many drives can be used if a "Single Ended Ultra Wide" hard drive is running on an Adaptec ASC-29160 which is an Ultra160 LVD / SE SCSI controller?

Answer:

Knowing above information and since it is an Ultra Wide drive:

- Fast Wide or Ultra Wide can address up to 15 devices.
- Ultra Narrow or Ultra Wide is limited to 1.5 Meters in cable length with four or more devices.
- Minimum distance between devices is 30 cm.

This gives us:

- 1.5 Meters total cable distance on the Ultra side of the bus divided by the 30 cm. minimum distance between devices. The answer is five [5].

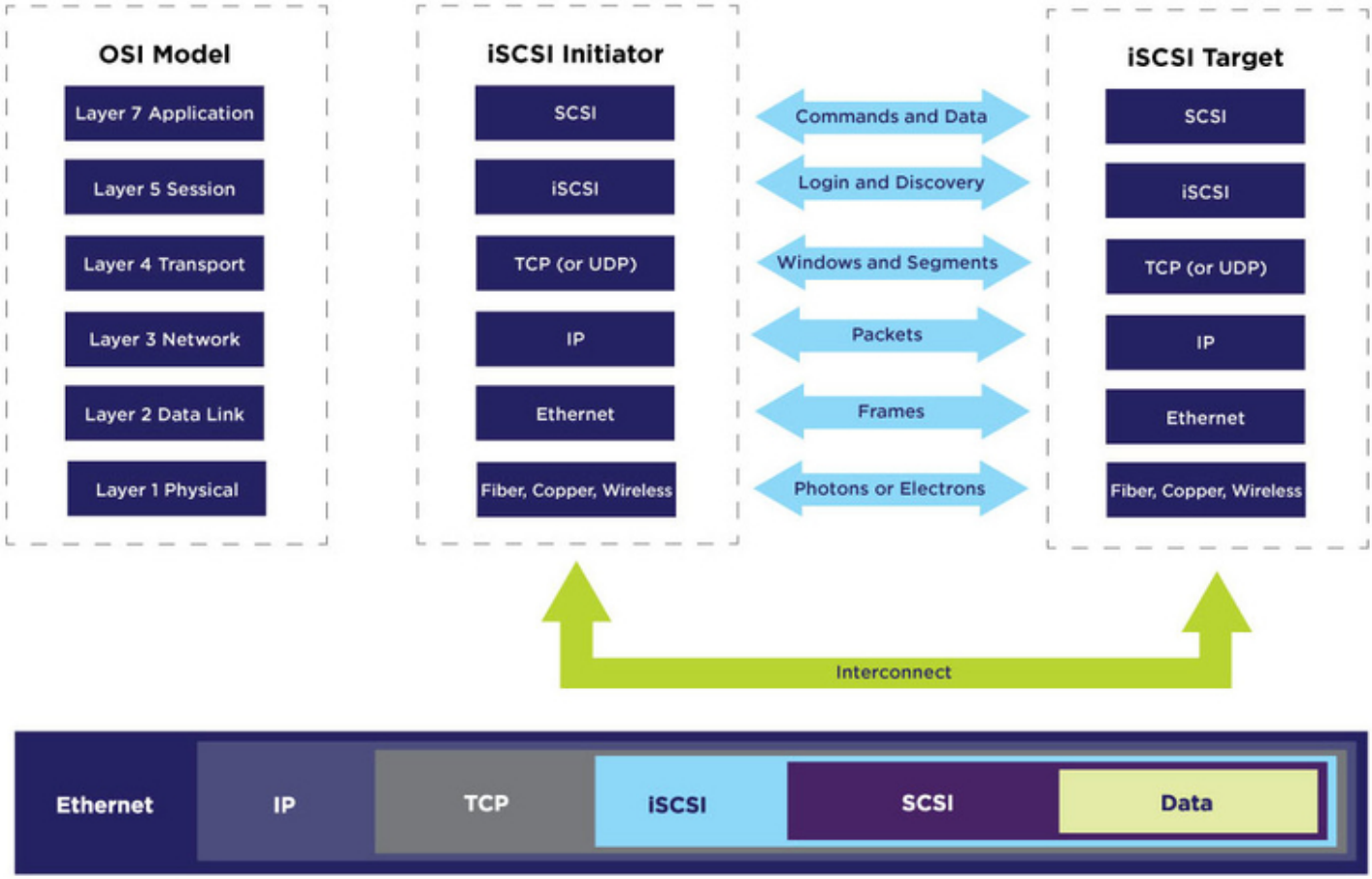
- Up to four drives can be used and one address for the controller.

Note:
No further information will be given on SCA SCSI devices or 80-pin devices that are designed for hot-swap cabinets or enclosures.

SCA (Single Connector Assembly) drives are usually designed for Hot-Swap bays or drive cages with SCA bay support that have its own specification. Although SCA drives have been known to work with Adaptec products with the proper cable converter and termination from the drive manufacturer, this configuration is not formally supported due to a technical issue known as the "stub length" that can cause data integrity issues.

SAS allows up to **65,535 devices** through the use of expanders, while Parallel SCSI has a limit of 8 or 16 devices on a single channel. SAS allows a higher transfer speed (3, 6 or 12 Gbit/s) than most parallel SCSI standards.
Speed: : SAS-1: Full-duplex 3 Gbit/s (2004); S...
No. of devices: 65,535

What is iSCSI?
The SNIA dictionary defines Internet Small Computer Systems Interface (iSCSI) as a transport protocol that provides for the SCSI protocol to be carried over a TCP-based IP network, standardized by the Internet Engineering Task Force and described in RFC 3720.
iSCSI is a block protocol for storage networking and runs the very common SCSI storage protocol across a network connection which is usually Ethernet.
iSCSI, like Fibre Channel, can be used to create a Storage Area Network (SAN). iSCSI traffic can be run over a shared network or a dedicated storage network. However, iSCSI does not support file access Network Attached Storage (NAS) or object storage access (they use different transport protocols).
There are multiple transports that can be used for iSCSI. The most common is TCP/IP over Ethernet, but Remote Direct Memory Access (RDMA) can also be used with iSER, which is iSCSI Extensions for RDMA. If using iSER, the transport is RoCE or InfiniBand and the underlying network is Ethernet (for RoCE) or InfiniBand (for InfiniBand transport).
iSCSI offers good block storage performance along with low cost. It is also widely supported by all major operating systems and hypervisors and can run on standard network cards or specialized Host Bus Adapters (HBAs). It is also supported by almost all enterprise storage arrays. For these reasons it has been popular for so-called “Tier 2” applications that require good, but not the best, block storage performance, and for storage that is shared by many hosts. It also is very popular among hyperscalers and large cloud service providers when they need a block storage solution that runs over Ethernet.



iSCSI can also be accelerated by using network adapters with an iSCSI hardware offload and/or a TCP Offload Engine (TOE). In the former, the hardware adapter (or HBA) offloads the iSCSI initiator function from the server CPU. In the latter case, the adapter offloads the TCP processing from the server kernel and CPU.
The rapid growth of faster Ethernet speeds such as 25G, 50G and 100G, along with increasing support for congestion management and traffic quality of service (QoS) on Ethernet switches, has greatly improved the potential performance, reliability and predictability of iSCSI as a storage protocol.

iSCSI

This reference provides cmdlet descriptions and syntax for all iSCSI Initiator-specific cmdlets. It lists the cmdlets in alphabetical order based on the verb at the beginning of the cmdlet.

iSCSI

ISCSI

Connect-IscsiTarget	Establishes a connection between the local iSCSI initiator and an iSCSI target device.
Disconnect-IscsiTarget	Disconnects sessions to the specified iSCSI target object.
Get-IscsiConnection	Gets information about connected iSCSI initiator connections.
Get-IscsiSession	Retrieves information about established iSCSI sessions.
Get-IscsiTarget	Returns an iSCSI target object for each iSCSI target that is registered with the iSCSI initiator.
Get-IscsiTargetPortal	Gets iSCSI target portals.
New-IscsiTargetPortal	Configures an iSCSI target portal.

Register-IscsiSession	Registers an active iSCSI session to be persistent using the session identifier as input.
Remove-IscsiTargetPortal	Removes the specified iSCSI target portal.
Set-IscsiChapSecret	Sets a CHAP secret key for use with iSCSI initiator connections.
Unregister-IscsiSession	Removes an active iSCSI session from being persistent using the session identifier as input.
Update-IscsiTarget	Refreshes the information about connected iSCSI target objects.
Update-IscsiTargetPortal	Updates information about the specified iSCSI target portal.

7. Virtualization

What is virtualization?

Virtualization uses software to create an abstraction layer over computer hardware that allows the hardware elements of a single computer—processors, memory, storage and more—to be divided into multiple virtual computers, commonly called virtual machines (VMs). Each VM runs its own operating system (OS) and behaves like an independent computer, even though it is running on just a portion of the actual underlying computer hardware.

It follows that virtualization enables more efficient utilization of physical computer hardware and allows a greater return on an organization's hardware investment.

Today, virtualization is a standard practice in enterprise IT architecture. It is also the technology that drives cloud computing economics. Virtualization enables cloud providers to serve users with their existing physical computer hardware; it enables cloud users to purchase only the computing resources they need when they need it, and to scale those resources cost-effectively as their workloads grow.

Benefits of virtualization

Virtualization brings several benefits to data center operators and service providers:

- **Resource efficiency:** Before virtualization, each application server required its own dedicated physical CPU—IT staff would purchase and configure a separate server for each application they wanted to run. (IT preferred one application and one operating system (OS) per computer for reliability reasons.) Invariably, each physical server would be underused. In contrast, server virtualization lets you run several applications—each on its own VM with its own OS—on a single physical computer (typically an x86 server) without sacrificing reliability. This enables maximum utilization of the physical hardware's computing capacity.
- **Easier management:** Replacing physical computers with software-defined VMs makes it easier to use and manage policies written in software. This allows you to create automated IT service management workflows. For example, automated deployment and configuration tools enable administrators to define collections of virtual machines and applications as services, in software templates. This means that they can install those services repeatedly and consistently without cumbersome, time-consuming. and error-prone manual setup. Admins can use virtualization security policies to mandate certain security configurations based on the role of the virtual machine. Policies can even increase resource efficiency by retiring unused virtual machines to save on space and computing power.
- **Minimal downtime:** OS and application crashes can cause downtime and disrupt user productivity. Admins can run multiple redundant virtual machines alongside each other and failover between them when problems arise. Running multiple redundant physical servers is more expensive.
- **Faster provisioning:** Buying, installing, and configuring hardware for each application is time-consuming. Provided that the hardware is already in place, provisioning virtual machines to run all

your applications is significantly faster. You can even automate it using management software and build it into existing workflows.

What is a virtual machine (VM)?

A virtual machine is a virtual representation, or emulation, of a physical computer. They are often referred to as a guest while the physical machine they run on is referred to as the host.

Virtualization makes it possible to create multiple virtual machines, each with their own operating system (OS) and applications, on a single physical machine. A VM cannot interact directly with a physical computer. Instead, it needs a lightweight software layer called a hypervisor to coordinate between it and the underlying physical hardware. The hypervisor allocates physical computing resources—such as processors, memory, and storage—to each VM. It keeps each VM separate from others so they don't interfere with each other.

While this technology can go by many names, including virtual server, virtual server instance (VSI) and virtual private server (VPS), this article will simply refer to them as virtual machines.

How virtualization works

When a hypervisor is used on a physical computer or server, (also known as bare metal server), it allows the physical computer to separate its operating system and applications from its hardware. Then, it can divide itself into several independent "virtual machines."

Each of these new virtual machines can then run their own operating systems and applications independently while still sharing the original resources from the bare metal server, which the hypervisor manages. Those resources include memory, RAM, storage, etc.

Hypervisors

A hypervisor is the software layer that coordinates VMs. It serves as an interface between the VM and the underlying physical hardware, ensuring that each has access to the physical resources it needs to execute. It also ensures that the VMs don't interfere with each other by impinging on each other's memory space or compute cycles.

There are two types of hypervisors:

- **Type 1 or "bare-metal" hypervisors** interact with the underlying physical resources, replacing the traditional operating system altogether. They most commonly appear in virtual server scenarios.
- **Type 2 hypervisors** run as an application on an existing OS. Most commonly used on endpoint devices to run alternative operating systems, they carry a performance overhead because they must use the host OS to access and coordinate the underlying hardware resources.
- **Type 1 hypervisors** run directly on the physical hardware (usually a server), taking the place of the OS. Typically, you use a separate software product to create and manipulate VMs on the hypervisor. Some management tools, like VMware's vSphere, let you select a guest OS to install in the VM.
- You can use one VM as a template for others, duplicating it to create new ones. Depending on your needs, you might create multiple VM templates for different purposes, such as software testing, production databases, and development environments.
- **Type 2 hypervisors** run as an application within a host OS and usually target single-user desktop or notebook platforms. With a Type 2 hypervisor, you manually create a VM and then install a guest OS in it. You can use the hypervisor to allocate physical resources to your VM, manually setting the amount of processor cores and memory it can use. Depending on the hypervisor's capabilities, you can also set options like 3D acceleration for graphics.

Types of virtualization

To this point we've discussed server virtualization, but many other IT infrastructure elements can be virtualized to deliver significant advantages to IT managers (in particular) and the enterprise as a whole. In this section, we'll cover the following types of virtualization:

- Desktop virtualization
- Network virtualization

- Storage virtualization
- Data virtualization
- Application virtualization
- Data center virtualization
- CPU virtualization
- GPU virtualization
- Linux virtualization
- Cloud virtualization

Desktop virtualization

Desktop virtualization lets you run multiple desktop operating systems, each in its own VM on the same computer.

There are two types of desktop virtualization:

- **Virtual desktop infrastructure (VDI)** runs multiple desktops in VMs on a central server and streams them to users who log in on thin client devices. In this way, VDI lets an organization provide its users access to variety of OS's from any device, without installing OS's on any device. See "What is Virtual Desktop Infrastructure (VDI)?" for a more in-depth explanation.
- **Local desktop virtualization** runs a hypervisor on a local computer, enabling the user to run one or more additional OSs on that computer and switch from one OS to another as needed without changing anything about the primary OS.

For more information on virtual desktops, see "Desktop-as-a-Service (DaaS)."

Network virtualization

Network virtualization uses software to create a "view" of the network that an administrator can use to manage the network from a single console. It abstracts hardware elements and functions (e.g., connections, switches, routers, etc.) and abstracts them into software running on a hypervisor. The network administrator can modify and control these elements without touching the underlying physical components, which dramatically simplifies network management.

Types of network virtualization include **software-defined networking (SDN)**, which virtualizes hardware that controls network traffic routing (called the "control plane"), and **network function virtualization (NFV)**, which virtualizes one or more hardware appliances that provide a specific network function (e.g., a firewall, load balancer, or traffic analyzer), making those appliances easier to configure, provision, and manage.

Storage virtualization

Storage virtualization enables all the storage devices on the network— whether they're installed on individual servers or standalone storage units—to be accessed and managed as a single storage device. Specifically, storage virtualization masses all blocks of storage into a single shared pool from which they can be assigned to any VM on the network as needed. Storage virtualization makes it easier to provision storage for VMs and makes maximum use of all available storage on the network.

For a closer look at storage virtualization, check out "What is Cloud Storage?"

Data virtualization

Modern enterprises store data from multiple applications, using multiple file formats, in multiple locations, ranging from the cloud to on-premise hardware and software systems. Data virtualization lets any application access all of that data—irrespective of source, format, or location.

Data virtualization tools create a software layer between the applications accessing the data and the systems storing it. The layer translates an application's data request or query as needed and returns results that can span multiple systems. Data virtualization can help break down data silos when other types of integration aren't feasible, desirable, or affordable.

Application virtualization

Application virtualization runs application software without installing it directly on the user's OS. This differs from complete desktop virtualization (mentioned above) because only the application runs in a virtual environment—the OS on the end user's device runs as usual. There are three types of application virtualization:

- **Local application virtualization:** The entire application runs on the endpoint device but runs in a runtime environment instead of on the native hardware.
- **Application streaming:** The application lives on a server which sends small components of the software to run on the end user's device when needed.
- **Server-based application virtualization** The application runs entirely on a server that sends only its user interface to the client device.

Data center virtualization

Data center virtualization abstracts most of a data center's hardware into software, effectively enabling an administrator to divide a single physical data center into multiple virtual data centers for different clients.

Each client can access its own infrastructure as a service (IaaS), which would run on the same underlying physical hardware. Virtual data centers offer an easy on-ramp into cloud-based computing, letting a company quickly set up a complete data center environment without purchasing infrastructure hardware.

CPU virtualization

CPU (central processing unit) virtualization is the fundamental technology that makes hypervisors, virtual machines, and operating systems possible. It allows a single CPU to be divided into multiple virtual CPUs for use by multiple VMs.

At first, CPU virtualization was entirely software-defined, but many of today's processors include extended instruction sets that support CPU virtualization, which improves VM performance.

GPU virtualization

A GPU (graphical processing unit) is a special multi-core processor that improves overall computing performance by taking over heavy-duty graphic or mathematical processing. GPU virtualization lets multiple VMs use all or some of a single GPU's processing power for faster video, artificial intelligence (AI), and other graphic- or math-intensive applications.

- **Pass-through GPUs** make the entire GPU available to a single guest OS.
- **Shared vGPUs** divide physical GPU cores among several virtual GPUs (vGPUs) for use by server-based VMs.

Linux virtualization

Linux includes its own hypervisor, called the kernel-based virtual machine (KVM), which supports Intel and AMD's virtualization processor extensions so you can create x86-based VMs from within a Linux host OS.

As an open source OS, Linux is highly customizable. You can create VMs running versions of Linux tailored for specific workloads or security-hardened versions for more sensitive applications.

Cloud virtualization

As noted above, the cloud computing model depends on virtualization. By virtualizing servers, storage, and other physical data center resources, cloud computing providers can offer a range of services to customers, including the following:

- **Infrastructure as a service (IaaS):** Virtualized server, storage, and network resources you can configure based on their requirements.
- **Platform as a service (PaaS):** Virtualized development tools, databases, and other cloud-based services you can use to build you own cloud-based applications and solutions.
- **Software as a service (SaaS)**: Software applications you use on the cloud. SaaS is the cloud-based service most abstracted from the hardware.

If you'd like to learn more about these cloud service models, see our guide: "IaaS vs. PaaS vs. SaaS."

Virtualization vs. containerization

Server virtualization reproduces an entire computer in hardware, which then runs an entire OS. The OS runs one application. That's more efficient than no virtualization at all, but it still duplicates unnecessary code and services for each application you want to run.

Containers take an alternative approach. They share an underlying OS kernel, only running the application and the things it depends on, like software libraries and environment variables. This makes containers smaller and faster to deploy.

8. Database Fundamentals

What is DBMS?

Database Management Systems (DBMS) are software systems used to store, retrieve, and run queries on data. A DBMS serves as an interface between an end-user and a database, allowing users to create, read, update, and delete data in the database.

DBMS manage the data, the database engine, and the database schema, allowing for data to be manipulated or extracted by users and other programs. This helps provide data security, data integrity, concurrency, and uniform data administration procedures.

DBMS optimizes the organization of data by following a database schema design technique called normalization, which splits a large table into smaller tables when any of its attributes have redundancy in values. DBMS offer many benefits over traditional file systems, including flexibility and a more complex backup system.

Database management systems can be classified based on a variety of criteria such as the data model, the database distribution, or user numbers. The most widely used types of DBMS software are relational, distributed, hierarchical, object-oriented, and network.

Distributed database management system

A distributed DBMS is a set of logically interrelated databases distributed over a network that is managed by a centralized database application. This type of DBMS synchronizes data periodically and ensures that any change to data is universally updated in the database.

Hierarchical database management system

Hierarchical databases organize model data in a tree-like structure. Data storage is either a top-down or bottom-up format and is represented using a parent-child relationship.

Network database management system

The network database model addresses the need for more complex relationships by allowing each child to have multiple parents. Entities are organized in a graph that can be accessed through several paths.

Relational database management system

Relational database management systems (RDBMS) are the most popular data model because of its user-friendly interface. It is based on normalizing data in the rows and columns of the tables. This is a viable option when you need a data storage system that is scalable, flexible, and able to manage lots of information.

Object-oriented database management system

Object-oriented models store data in objects instead of rows and columns. It is based on object-oriented programming (OOP) that allows objects to have members such as fields, properties, and methods.

Examples of DBMS

There is a wide range of database software solutions, including both enterprise and open source solutions, available for database management.

Here are some of the most popular database management systems:

Oracle

Oracle Database is a commercial relational database management system. It utilizes enterprise-scale database technology with a robust set of features right out of the box. It can be stored in the cloud or on-premises.

Learn how AppDynamics helps with Oracle monitoring

MySQL

MySQL is a relational database management system that is commonly used with open-source content management systems and large platforms like Facebook, Twitter, and Youtube.

Learn how AppDynamics helps with MySQL monitoring

SQL Server

Developed by Microsoft, SQL Server is a relational database management system built on top of structured query language (SQL), a standardized programming language that allows database administrators to manage databases and query data.

RDBMS

What is a Relational Database (RDBMS)?

A relational database is a type of database that stores and provides access to data points that are related to one another. Relational databases are based on the relational model, an intuitive, straightforward way of representing data in tables. In a relational database, each row in the table is a record with a unique ID called the key. The columns of the table hold attributes of the data, and each record usually has a value for each attribute, making it easy to establish the relationships among data points.

Industry's best RBDMS

A relational database example

Here's a simple example of two tables a small business might use to process orders for its products. The first table is a customer info table, so each record includes a customer's name, address, shipping and billing information, phone number, and other contact information. Each bit of information (each attribute) is in its own column, and the database assigns a unique ID (a key) to each row. In the second table—a customer order table—each record includes the ID of the customer that placed the order, the product ordered, the quantity, the selected size and color, and so on—but not the customer's name or contact information.

These two tables have only one thing in common: the ID column (the key). But because of that common column, the relational database can create a relationship between the two tables. Then, when the company's order processing application submits an order to the database, the database can go to the customer order table, pull the correct information about the product order, and use the customer ID from that table to look up the customer's billing and shipping information in the customer info table. The warehouse can then pull the correct product, the customer can receive timely delivery of the order, and the company can get paid.

How relational databases are structured

The relational model means that the logical data structures—the data tables, views, and indexes—are separate from the physical storage structures. This separation means that database administrators can manage physical data storage without affecting access to that data as a logical structure. For example, renaming a database file does not rename the tables stored within it.

The distinction between logical and physical also applies to database operations, which are clearly defined actions that enable applications to manipulate the data and structures of the database. Logical operations allow an application to specify the content it needs, and physical operations determine how that data should be accessed and then carries out the task.

To ensure that data is always accurate and accessible, relational databases follow certain integrity rules. For example, an integrity rule can specify that duplicate rows are not allowed in a table in order to eliminate the potential for erroneous information entering the database.

The relational model

In the early years of databases, every application stored data in its own unique structure. When developers wanted to build applications to use that data, they had to know a lot about the particular data structure to find the data they needed. These data structures were inefficient, hard to maintain, and hard to optimize for delivering good application performance. The relational database model was designed to solve the problem of multiple arbitrary data structures.

The relational data model provided a standard way of representing and querying data that could be used by any application. From the beginning, developers recognized that the chief strength of the relational database model was in its use of tables, which were an intuitive, efficient, and flexible way to store and access structured information.

Over time, another strength of the relational model emerged as developers began to use structured query language (SQL) to write and query data in a database. For many years, SQL has been widely used as the language for database queries. Based on relational algebra, SQL provides an internally consistent mathematical language that makes it easier to improve the performance of all database queries. In comparison, other approaches must define individual queries.

Benefits of relational database management system

The simple yet powerful relational model is used by organizations of all types and sizes for a broad variety of information needs. Relational databases are used to track inventories, process ecommerce transactions, manage huge amounts of mission-critical customer information, and much more. A relational database can be considered for any information need in which data points relate to each other and must be managed in a secure, rules-based, consistent way.

Relational databases have been around since the 1970s. Today, the advantages of the relational model continue to make it the most widely accepted model for databases.

Relational model and data consistency

The relational model is the best at maintaining data consistency across applications and database copies (called instances). For example, when a customer deposits money at an ATM and then looks at the account balance on a mobile phone, the customer expects to see that deposit reflected immediately in an updated account balance. Relational databases excel at this kind of data consistency, ensuring that multiple instances of a database have the same data all the time.

It's difficult for other types of databases to maintain this level of timely consistency with large amounts of data. Some recent databases, such as NoSQL, can supply only "eventual consistency." Under this principle, when the database is scaled or when multiple users access the same data at the same time, the data needs some time to "catch up." Eventual consistency is acceptable for some uses, such as to maintain listings in a product catalog, but for critical business operations such as shopping cart transactions, the relational database is still the gold standard.

Commitment and atomicity

Relational databases handle business rules and policies at a very granular level, with strict policies about commitment (that is, making a change to the database permanent). For example, consider an inventory database that tracks three parts that are always used together. When one part is pulled from inventory, the other two must also be pulled. If one of the three parts isn't available, none of the parts should be pulled—all three parts must be available before the database makes any commitment. A relational database won't commit for one part until it knows it can commit for all three. This multifaceted commitment capability is called atomicity. Atomicity is the key to keeping data accurate in the database and ensuring that it is compliant with the rules, regulations, and policies of the business.

ACID properties and RDBMS

Four crucial properties define relational database transactions: atomicity, consistency, isolation, and durability—typically referred to as ACID.

- **Atomicity** defines all the elements that make up a complete database transaction.
- **Consistency** defines the rules for maintaining data points in a correct state after a transaction.
- **Isolation** keeps the effect of a transaction invisible to others until it is committed, to avoid confusion.
- **Durability** ensures that data changes become permanent once the transaction is committed.

Stored procedures and relational databases

Data access involves many repetitive actions. For example, a simple query to get information from a data table may need to be repeated hundreds or thousands of times to produce the desired result. These data access functions require some type of code to access the database. Application developers don't want to write new code for these functions in each new application. Luckily, relational databases allow stored procedures, which are blocks of code that can be accessed with a simple application call. For example, a single stored procedure can provide consistent record tagging for users of multiple applications. Stored procedures can also help developers ensure that certain data functions in the application are implemented in a specific way.

Database locking and concurrency

Conflicts can arise in a database when multiple users or applications attempt to change the same data at the same time. Locking and concurrency techniques reduce the potential for conflicts while maintaining the integrity of the data.

Locking prevents other users and applications from accessing data while it is being updated. In some databases, locking applies to the entire table, which creates a negative impact on application performance. Other databases, such as Oracle relational databases, apply locks at the record level, leaving the other records within the table available, helping ensure better application performance.

Concurrency manages the activity when multiple users or applications invoke queries at the same time on the same database. This capability provides the right access to users and applications according to policies defined for data control.

What to look for when selecting a relational database

The software used to store, manage, query, and retrieve data stored in a relational database is called a relational database management system (RDBMS). The RDBMS provides an interface between users and applications and the database, as well as administrative functions for managing data storage, access, and performance.

Several factors can guide your decision when choosing among database types and relational database products. The RDBMS you choose will depend on your business needs. Ask yourself the following questions:

- What are our data accuracy requirements? Will data storage and accuracy rely on business logic? Does our data have stringent requirements for accuracy (for example, financial data and government reports)?
- Do we need scalability? What is the scale of the data to be managed, and what is its anticipated growth? Will the database model need to support mirrored database copies (as separate instances) for scalability? If so, can it maintain data consistency across those instances?
- How important is concurrency? Will multiple users and applications need simultaneous data access? Does the database software support concurrency while protecting the data?
- What are our performance and reliability needs? Do we need a high-performance, high-reliability product? What are the requirements for query-response performance? What are the vendor's commitments for service level agreements (SLAs) or unplanned downtime?

SQL Server

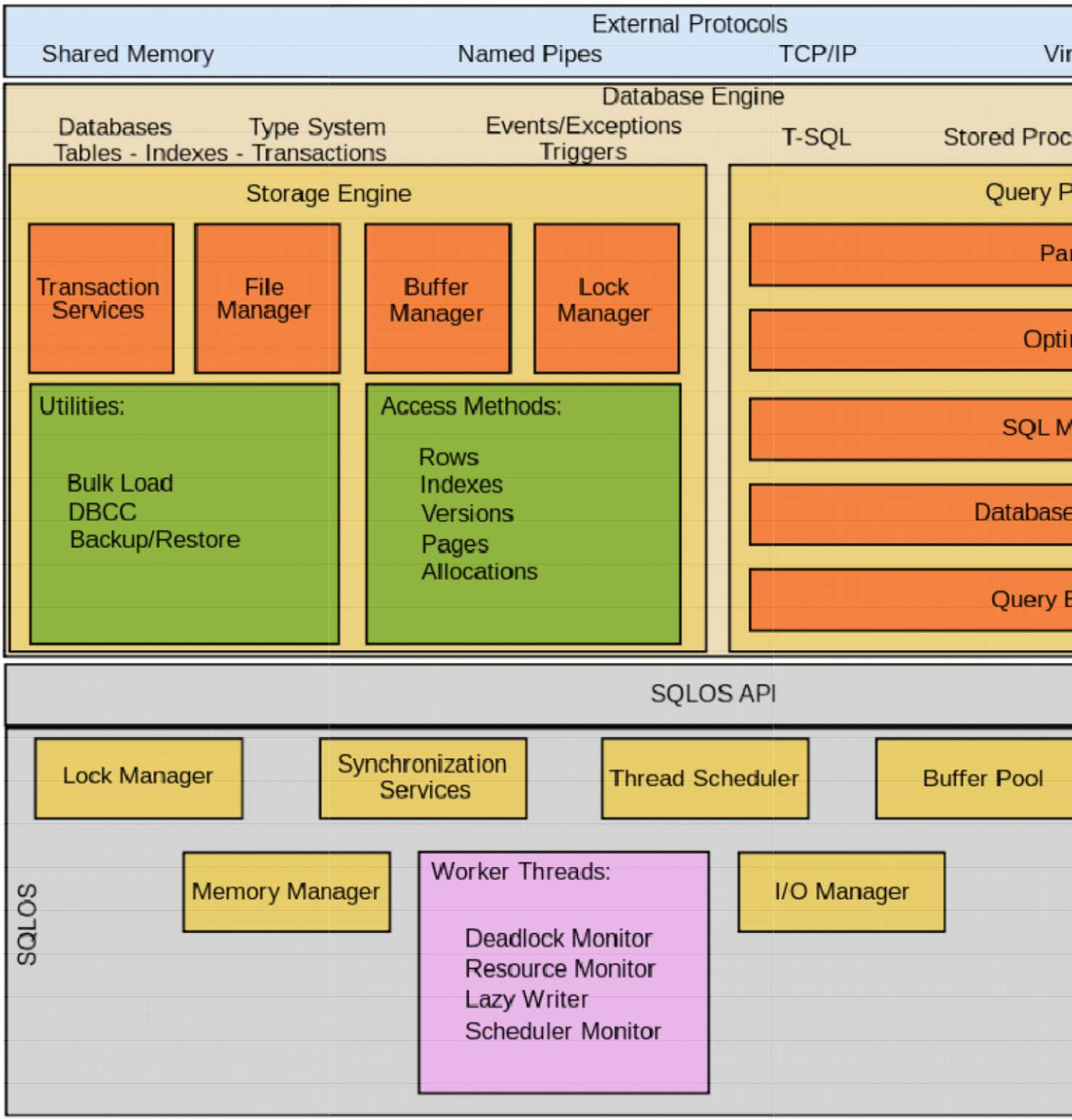
SQL Server is a relational database management system, or RDBMS, developed and marketed by Microsoft.

Similar to other RDBMS software, SQL Server is built on top of SQL, a standard programming language for interacting with the relational databases. SQL server is tied to Transact-SQL, or T-SQL, the Microsoft's implementation of SQL that adds a set of proprietary programming constructs.

SQL Server works exclusively on Windows environment for more than 20 years. In 2016, Microsoft made it available on Linux. SQL Server 2017 became generally available in October 2016 that ran on both Windows and Linux.

SQL Server Architecture

The following diagram illustrates the architecture of the SQL Server:



SQL Server consists of two main components:

1. Database Engine
2. SQLOS

Database Engine

The core component of the SQL Server is the Database Engine. The Database Engine consists of a relational engine that processes queries and a storage engine that manages database files, pages, pages, index, etc. The

database objects such as stored procedures, views, and triggers are also created and executed by the Database Engine.

Relational Engine

The Relational Engine contains the components that determine the best way to execute a query. The relational engine is also known as the query processor.

The relational engine requests data from the storage engine based on the input query and processed the results.

Some tasks of the relational engine include querying processing, memory management, thread and task management, buffer management, and distributed query processing.

Storage Engine

The storage engine is in charge of storage and retrieval of data from the storage systems such as disks and SAN.

SQLOS

Under the relational engine and storage engine is the SQL Server Operating System or SQLOS.

SQLOS provides many operating system services such as memory and I/O management. Other services include exception handling and synchronization services.

SQL Server Services and Tools

Microsoft provides both data management and business intelligence (BI) tools and services together with SQL Server.

For data management, SQL Server includes SQL Server Integration Services (SSIS), SQL Server Data Quality Services, and SQL Server Master Data Services. To develop databases, SQL Server provides SQL Server Data tools; and to manage, deploy, and monitor databases SQL Server has SQL Server Management Studio (SSMS).

For data analysis, SQL Server offers SQL Server Analysis Services (SSAS). SQL Server Reporting Services (SSRS) provides reports and visualization of data. The Machine Learning Services technology appeared first in SQL Server 2016 which was renamed from the R Services.

SQL Server Editions

SQL Server has four primary editions that have different bundled services and tools. Two editions are available free of charge:

SQL Server Developer edition for use in database development and testing.

SQL Server Expression for small databases with the size up to 10 GB of disk storage capacity.

For larger and more critical applications, SQL Server offers the Enterprise edition that includes all SQL server's features.

SQL Server Standard Edition has partial feature sets of the Enterprise Edition and limits on the Server regarding the numbers of processor core and memory that can be configured.

NoSQL

NoSQL databases (aka "not only SQL") are non-tabular databases and store data differently than relational tables. NoSQL databases come in a variety of types based on their data model. The main types are document, key-value, wide-column, and graph. They provide flexible schemas and scale easily with large amounts of data and high user loads.

In this article, you'll learn *what* a NoSQL database is, *why* (and when!) you should use one, and *how* to get started.

Overview

This article will cover:

- What is a NoSQL Database?
    - Brief History of NoSQL Databases
    - NoSQL Database Features
    - Types of NoSQL Database
    - Difference between RDBMS and NoSQL
    - Why NoSQL?
    - When should NoSQL be Used?

What is a NoSQL database?

When people use the term “NoSQL database,” they typically use it to refer to any non-relational database. Some say the term “NoSQL” stands for “non SQL” while others say it stands for “not only SQL.” Either way, most agree that NoSQL databases are databases that store data in a format other than relational tables.

Brief history of NoSQL databases

NoSQL databases emerged in the late 2000s as the cost of storage dramatically decreased. Gone were the days of needing to create a complex, difficult-to-manage data model in order to avoid data duplication. Developers (rather than storage) were becoming the primary cost of software development, so NoSQL databases optimized for developer productivity.

As storage costs rapidly decreased, the amount of data that applications needed to store and query increased. This data came in all shapes and sizes — structured, semi-structured, and polymorphic — and defining the schema in advance became nearly impossible. NoSQL databases allow developers to store huge amounts of unstructured data, giving them a lot of flexibility.

Additionally, the Agile Manifesto was rising in popularity, and software engineers were rethinking the way they developed software. They were recognizing the need to rapidly adapt to changing requirements. They needed the ability to iterate quickly and make changes throughout their software stack — all the way down to the database. NoSQL databases gave them this flexibility.

Cloud computing also rose in popularity, and developers began using public clouds to host their applications and data. They wanted the ability to distribute data across multiple servers and regions to make their applications resilient, to scale out instead of scale up, and to intelligently geo-place their data. Some NoSQL databases like MongoDB provide these capabilities.

NoSQL database features

Each NoSQL database has its own unique features. At a high level, many NoSQL databases have the following features:

- Flexible schemas
- Horizontal scaling
- Fast queries due to the data model
- Ease of use for developers

Check out What are the Benefits of NoSQL Databases? to learn more about each of the features listed above.

Types of NoSQL databases

Over time, four major types of NoSQL databases emerged: document databases, key-value databases, wide-column stores, and graph databases.

- Document databases store data in documents similar to JSON (JavaScript Object Notation) objects. Each document contains pairs of fields and values. The values can typically be a variety of types including things like strings, numbers, booleans, arrays, or objects.
- Key-value databases are a simpler type of database where each item contains keys and values.
- Wide-column stores store data in tables, rows, and dynamic columns.
- Graph databases store data in nodes and edges. Nodes typically store information about people, places, and things, while edges store information about the relationships between the nodes.

To learn more, visit Understanding the Different Types of NoSQL Databases.

Difference between RDBMS and NoSQL databases

While a variety of differences exist between relational database management systems (RDBMS) and NoSQL databases, one of the key differences is the way the data is modeled in the database. In this section, we'll work through an example of modeling the same data in a relational database and a NoSQL database. Then, we'll highlight some of the other key differences between relational databases and NoSQL databases.

RDBMS vs NoSQL: Data Modeling Example

Let's consider an example of storing information about a user and their hobbies. We need to store a user's first name, last name, cell phone number, city, and hobbies.

In a relational database, we'd likely create two tables: one for Users and one for Hobbies.

Users

ID	first_name	last_name	cell	city
1	Leslie	Yepp	8125552344	Pawnee

Hobbies

ID	user_id	hobby
10	1	scrapbooking
11	1	eating waffles
12	1	working

In order to retrieve all of the information about a user and their hobbies, information from the Users table and Hobbies table will need to be joined together.

The data model we design for a NoSQL database will depend on the type of NoSQL database we choose. Let's consider how to store the same information about a user and their hobbies in a document database like MongoDB.

```
{
  "_id": 1,
  "first_name": "Leslie",
  "last_name": "Yepp",
  "cell": "8125552344",
  "city": "Pawnee",
  "hobbies": ["scrapbooking", "eating waffles", "working"]
}
```

In order to retrieve all of the information about a user and their hobbies, a single document can be retrieved from the database. No joins are required, resulting in faster queries.

To see a more detailed version of this data modeling example, read Mapping Terms and Concepts from SQL to MongoDB.

Other differences between RDBMS and relational databases

While the example above highlights the differences in data models between relational databases and NoSQL databases, many other important differences exist, including:

- Flexibility of the schema
- Scaling technique
- Support for transactions
- Reliance on data to object mapping

Why NoSQL?

NoSQL databases are used in nearly every industry. Use cases range from the highly critical (e.g., storing financial data and healthcare records) to the more fun and frivolous (e.g., storing IoT readings from a smart kitty litter box).
In the following sections, we'll explore when you should choose to use a NoSQL database and common misconceptions about NoSQL databases.

When should NoSQL be used?

When deciding which database to use, decision-makers typically find one or more of the following factors lead them to selecting a NoSQL database:

- Fast-paced Agile development
- Storage of structured and semi-structured data
- Huge volumes of data
- Requirements for scale-out architecture
- Modern application paradigms like microservices and real-time streaming

See When to Use NoSQL Databases and Exploring NoSQL Database Examples for more detailed information on the reasons listed above.

NoSQL database misconceptions

Over the years, many misconceptions about NoSQL databases have spread throughout the developer community. In this section, we'll discuss two of the most common misconceptions:

- Relationship data is best suited for relational databases.
- NoSQL databases don't support ACID transactions.

To learn more about common misconceptions, read Everything You Know About MongoDB is Wrong.
Misconception: relationship data is best suited for relational databases
A common misconception is that NoSQL databases or non-relational databases don't store relationship data well. NoSQL databases can store relationship data — they just store it differently than relational databases do. In fact, when compared with relational databases, many find modeling relationship data in NoSQL databases to be easier than in relational databases, because related data doesn't have to be split between tables. NoSQL data models allow related data to be nested within a single data structure.
Misconception: NoSQL databases don't support ACID transactions

Another common misconception is that NoSQL databases don't support ACID transactions. Some NoSQL databases like MongoDB do, in fact, support ACID transactions.

Note that the way data is modeled in NoSQL databases can eliminate the need for multi-record transactions in many use cases. Consider the earlier example where we stored information about a user and their hobbies in both a relational database and a document database. In order to ensure information about a user and their hobbies was updated together in a relational database, we'd need to use a transaction to update records in two tables. In order to do the same in a document database, we could update a single document — no multi-record transaction required.

MongoDB Atlas PostgreSQL

	MongoDB Atlas	PostgreSQL
Full Featured Document Database	MongoDB Atlas is a full featured document database, with ACID transactions and a rich and expressive query language.	PostgreSQL is a tabular database with legacy overhead for developers. Its primitive JSON columns lack rich data types.
Flexible, Developer Controlled Schemas	Developers can dynamically evolve the database on demand, and lock schemas down if needed.	Data model changes necessitate complex schema migrations, slowing down the pace of development.
Distributed vs Monolithic	MongoDB is a distributed database by design with replication, self-healing recovery, native sharding, and geo-pinning.	PostgreSQL is a scale-up-only database, requiring third party tools and custom engineering for failover and scaling.
True JSON Support	MongoDB Atlas was designed from the start to use JSON documents, extended into BSON to support more data types.	PostgreSQL JSON support is retrofitted by stuffing JSON data into a single column which incurs significant overhead.

1. What is a NoSQL Database?
2. Types of NoSQL Databases
    1. Document-Based Database
    2. Key-Value Database
    3. Wide Column Based Database
    4. Graph-Based Database
3. Different NoSQL Databases
    1. MongoDB
    2. Cassandra
    3. ElasticSearch
    4. Amazon DynamoDB
    5. HBase

What is a NoSQL Database?
So what is a NoSQL database?
You might have heard people saying that a NoSQL Database is any non-relational database that doesn't have any relationship between the data. Well, that's not completely true. They can also store the relationship between the data but in a different way.
We can say that "NoSQL" stands for "Not Only SQL". Here, data is not split into multiple tables, as it allows all the data that is related in any way possible, in a single data structure. When you work with a huge amount of data, you don't need to worry about the performance lags when you query a NoSQL database. No need to run the expensive joins! They are highly scalable and reliable and designed to work in a distributed environment.

Types of NoSQL Databases

Now that we know what a NoSQL database is, let's explore the different types of NoSQL databases in this section.

1. Document-Based NoSQL Databases

Document-based databases store the data in JSON objects. Each document has key-value pairs like structures:


```
{
  "index": NumberInt(0),
  "name": "Aurelia Gonzales",
  "isActive": false,
  "registered": ISODate("2015-02-11T04:22:39+0000"),
  "age": NumberInt(20),
  "gender": "female",
  "eyeColor": "green",
  "favoriteFruit": "banana",
  "company": {
    "title": "YURTURE",
    "email": "[redacted]",
    "phone": "+1 [redacted]",
    "location": {
      "country": "USA",
      "address": "[redacted]"
    }
  },
  "tags": [
    "enim",
    "id",
    "velit",
    "ad",
    "consequat"
  ]
},
{
  "index": NumberInt(1),
  "name": "Kitty Snow",
  "isActive": false,
  "registered": ISODate("2018-01-23T04:46:15+0000"),
```


The document-based databases are easy for developers as the document directly maps to the objects as JSON is a very common data format used by web developers. They are very flexible and allow us to modify the structure at any time.

P_ID	NAME	AGE	DOB	CONTACT NO	EMAIL
2092	AKSHAT	22	01-01-2002	9090-0202	abc@gmail.com

P_ID	ADDRESS	TYPE
2092	342, Block A, DLF Phase 3, Gurgaon 122022	Office
2092	10-A, Sector 4, Gurgaon 122005	Home

P_ID	PRODUCT	ADDRESS TYPE	DELIVERY STATUS
2092	JBx Earphones 100	Home	Delivered
2092	Wall Stickers	Office	Shipped

Some examples of document-based databases are MongoDB, Orient DB, and BaseX.

2. Key-Value Databases

As the name suggests, it stores the data as key-value pairs. Here, keys and values can be anything like strings, integers, or even complex objects. They are highly partitionable and are the best in horizontal scaling. They can be really useful in session oriented applications where we try to capture the behavior of the customer in a particular session.

Some of the examples are DynamoDB, Redis, and Aerospike.

3. Wide Column-Based Databases

This database stores the data in records similar to any relational database but it has the ability to store very large numbers of dynamic columns. It groups the columns logically into column families.

For example, in a relational database, you have multiple tables but in a wide-column based database, instead of having multiple tables, we have multiple column families.

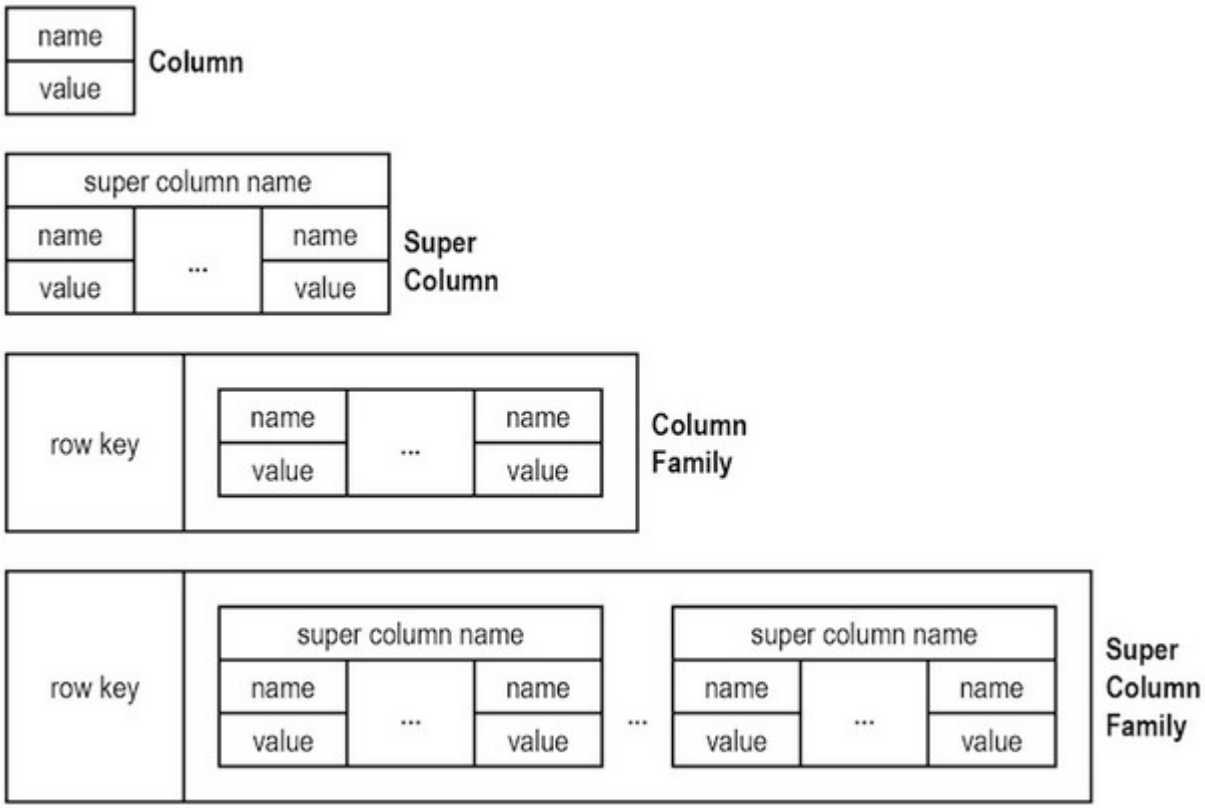


Image Source

Here is a good resource to learn more about column-based databases:
Popular examples of these types of databases are Cassandra and HBase.

4. Graph-Based Databases

They store the data in the form of nodes and edges. The node part of the database stores information about the main entities like people, places, products, etc., and the edges part stores the relationships between them. These work best when you need to find out the relationship or pattern among your data points like a social network, recommendation engines, etc.
Some of the examples are Neo4j, Amazon Neptune, etc.
Now, let's have a look at some of the NoSQL databases and their features.

List of the Different NoSQL Databases

1. MongoDB

MongoDB is the most widely used document-based database. It stores the documents in JSON objects.
According to the website stackshare.io, more than 3400 companies are using MongoDB in their tech stack. Uber, Google, eBay, Nokia, Coinbase are some of them.
When to use MongoDB?

1. In case you are planning to integrate hundreds of different data sources, the document-based model of MongoDB will be a great fit as it will provide a single unified view of the data
2. When you are expecting a lot of reads and write operations from your application but you do not care much about some of the data being lost in the server crash
3. You can use it to store clickstream data and use it for the customer behavioral analysis

2. Cassandra

Cassandra is an open-source, distributed database system that was initially built by Facebook (and motivated by Google's Big Table). It is widely available and quite scalable. It can handle petabytes of information and thousands of concurrent requests per second.

Again, according to stackshare.io, more than 400 companies are using Cassandra in their tech stack. Facebook, Instagram, Netflix, Spotify, Coursera are some of them.

When to use Cassandra?

1. When your use case requires more writing operations than reading ones
2. In situations where you need more availability than consistency. For example, you can use it for social network websites but cannot use it for banking purposes
3. You require less number of joins and aggregations in your queries to the database
4. Health trackers, weather data, tracking of orders, and time series data are some good use cases where you can use Cassandra databases

3. ElasticSearch

This is also an open-source, distributed NoSQL database system. It is highly scalable and consistent. You can also call it as an **Analytics Engine**. It can easily analyze, store, and search huge volumes of data.

If the full-text search is a part of your use case, ElasticSearch will be the best fit for your tech stack. It even allows search with fuzzy matching.

More than 3000 companies are using Elasticsearch in their tech stack, including Slack, Udemy, Medium, and Stackoverflow.

When to use ElasticSearch?

1. If your use case requires a full-text search, Elasticsearch will be the best fit
2. If your use case involves chatbots where these bots resolve most of the queries, such as when a person types something there are high chances of spelling mistakes. You can make use of the in-built fuzzy matching practices of the ElasticSearch
3. Also, ElasticSearch is useful in storing logs data and analyzing it

4. Amazon DynamoDB

It is a key-value pair based distributed database system created by Amazon and is highly scalable. But unfortunately, it is not open-source. It can easily handle 10 trillion requests per day so you can see why!

More than 700 companies are using DynamoDB in their tech stack including Snapchat, Lyft, and Samsung.

When to use DynamoDB?

1.
    1. In case you are looking for a database that can handle simple key-value queries but those queries are very large in number
    2. In case you are working with OLTP workload like online ticket booking or banking where the data needs to be highly consistent

5. HBase

It is also an open-source highly scalable distributive database system. HBase was written in JAVA and runs on top of the Hadoop Distributed File System (HDFS).

More than 70 companies are using Hbase in their tech stack, such as Hike, Pinterest, and HubSpot.

When to use HBase?

1. You should have at least petabytes of data to be processed. If your data volume is small, then you will not get the desired results
2. If your use case requires random and real-time access to the data, then HBase will be the appropriate option
3. If you want to easily store real-time messages for billions of people

9. Backup Fundamentals

Introduction to Backup Architecture

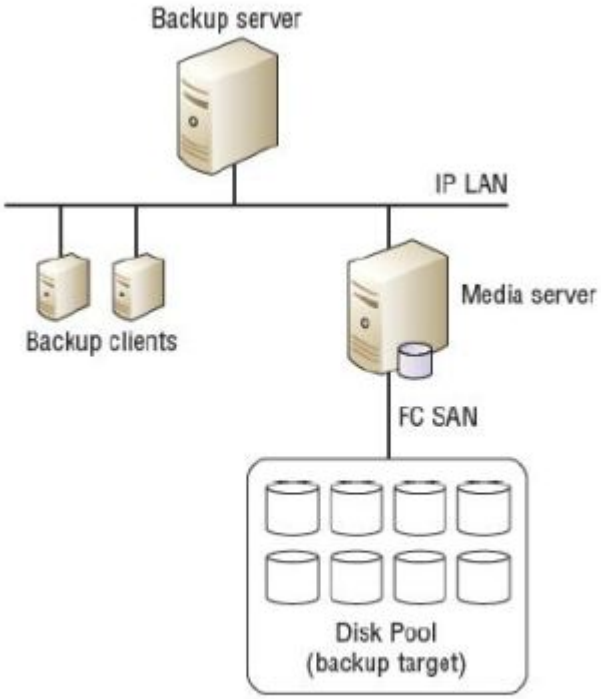
The common and widely used Backup Architecture is based on the Server-Client model. Any backup architecture is composed of the following four components.

- Backup Servers
- Backup Clients
- Media Servers
- Backup Destinations/Targets

The **backup server** manages the backup operations and maintains the backup database, which contains information about the backup configuration and backup metadata. The backup configuration contains information about when to run backups, which client data to be backed up, and so on. The backup metadata contains information about the backed up data.

The role of a **backup client** is to gather the data that is to be backed up and send it to the backup server. The backup client can be installed on application servers, mobile clients, and desktops. It also sends the tracking information to the backup server.

Media Servers connect to the backup destinations and make it available to backup clients so that they can send data to the backup target. In IBM TSM terminology, media servers are referred as Primary Library Manager and other TSM servers as Library Clients. The media servers controls one or more backup devices. Backup devices may be attached directly or through a network to the Media Servers. The Media Servers sends the tracking information about the data written to the backup device to the backup server. Typically this information is used for recoveries. For example, a media server might be connected to a pool of storage over an FC network and make that storage available to backup clients over an SAN.



A wide range of **backup destinations/targets** are currently available such as tape, disk, and virtual tape library. Traditional backup solutions primarily used tape as a backup destination and modern backup

approaches tend to use disk based pools which are shared over SAN or LAN. Disk arrays can also be used as virtual tape libraries to combine the benefits of Disk and Tape. Now, organizations can also back up their data to the cloud storage. Many service providers offer backup as a service that enables an organization to reduce its backup management overhead.

The backup window is another term that relates to backup and business continuity. A backup window is a period of time during which data can be backed up. For example, most organizations consider the time between 6 PM to 6 AM as the ideal backup window as there will be less activity on the servers. However, many businesses have different backup windows for different applications, depending on core business hours for those applications. A major reason for the backup window is that backups often negatively impact the performance of the system being backed up, so backing up a critical system during the core business day is usually not a good idea.

Different Types of Backup Destinations/Targets

Tape Library and Tape Drives

A tape library contains one or more tape drives that records and retrieves data on a magnetic tape. Tape is portable, and one of the primary reasons for the use of tape is long-term, off-site storage. Backups implemented using tape devices involve several hidden costs. Tapes must be stored in locations with a controlled environment to ensure preservation of the media and to prevent data corruption. The advantage of using tape drives as backup target is that modern tape drives and tape libraries have high capacity, high sequential performance and low power. The common disadvantages of tapes are they are subject to degradation over time, especially if they aren't stored in optimal conditions. Even though you have stored the tapes in optimal conditions, it is possible that you no longer have any tape drives or software that can read them. In addition, tapes are great at restoring data from full backups, but if you have to perform restores based on differential backups, they can take huge time.

Physical transportation of the tapes to offsite locations also adds management overhead and increases the possibility of loss of tapes during offsite shipment. Due to its sequential data access, both backing up of data and restoring it take more time with tape, and this may impact the backup window and RTO. Data integrity and recoverability are also major issues with tape-based backup media. Because of these concerns, as well as the drop in the cost of spinning disk media, many people are starting to move away from tape for backups.

Disk Drives

Another backup target which is widely used in backup infrastructure is the Disk Drives. The commonly used disk drives are Hard Disk Drives (HDD) and Solid State Drives (SSD). Hard Disk Drives gives better performance than Solid State Drives. Disk density has increased dramatically over the past few years, lowering the cost per gigabyte to the point where it became a viable backup target for organizations. When used in a highly available configuration in a storage array, disks offer a reliable and fast backup target medium. One way to implement a backup-to-disk system is by using it as a staging area, offloading backup data to a secondary backup target such as tape after a period of time (generally referred as migration). Some vendors offer a purpose-built, disk-based backup appliances that are emerged as the optimal backup target solution. These systems are optimized for backup and recovery operations, offering extensive integration with popular backup management applications. The built-in features such as replication, compression, encryption, and data deduplication increase the value of purpose-built backup appliances.

Virtual Tape Libraries

Virtual tape libraries use disks as backup media. Virtual tapes are disk drives that are emulated and presented as tapes to the backup software. Compared to physical tapes, virtual tapes offer better performance, better reliability, and random disk access. A virtual tape drive does not require the usual maintenance tasks associated with a physical tape drive, such as periodic cleaning and drive calibration. Compared to the disk

library, a virtual tape library offers easy installation and administration because it is preconfigured by the manufacturer. A key feature that is usually available on virtual tape library appliances is replication.

Taking backup to the Cloud

Since everything is going to cloud now, we can use cloud to send the backups as well. A common approach is to deploy a disk-to-disk-to-cloud (D2D2C) architecture. This is similar to disk-to-disk-to-tape (D2D2T), with the tape being replaced by the cloud. In this architecture, the disk-to-disk portion still occurs within your data center, utilizes deduplication technology, and transfers only deduplicated data to the cloud for long-term retention, where it's not expected to be required for restore operations as often as the data that is on premise on the intermediary disk backup platform. Deduplication technology and synthetic full backups, or other incremental forever approaches, also help here. It should also be noted in this D2D2C architecture that only a portion of the backup estate exists in the cloud.

Data Backup in Depth: Concepts, Techniques, and Storage Technologies

In an increasingly digitized business landscape, data backup is vital for the survival of an organization. You can get hacked or ransomed, and lose your data to thieves who'll sell your trade secrets to the highest bidder. Injected malware can corrupt your hard-earned information. Disgruntled employees or other insider threats can delete your valuable digital assets. Can you recover from data loss?

Data backup is a practice that combines techniques and solutions for efficient and cost-effective backup. Your data is copied to one or more locations, at pre-determined frequencies, and at different capacities. You can set up a flexible data backup operation, using your own architecture, or make use of available Backup as a Service (BaaS) solutions, mixing them up with local storage. Today, there are plenty of corporate storage TCO solutions to help you calculate costs, avoid data loss, and prevent data breaches.

In this article:

What Is a Data Backup?

Data backup is the practice of copying data from a primary to a secondary location, to protect it in case of a disaster, accident or malicious action. Data is the lifeblood of modern organizations, and losing data can cause massive damage and disrupt business operations. This is why backing up your data is critical for all businesses, large and small.

What does backup data mean?

Typically backup data means all necessary data for the workloads your server is running. This can include documents, media files, configuration files, machine images, operating systems, and registry files. Essentially, any data that you want to preserve can be stored as backup data.

Data backup includes several important concepts:

- **Backup solutions and tools**—while it is possible to back up data manually, to ensure systems are backed up regularly and consistently, most organizations use a technology solution to back up their data.
- **Backup administrator**—every organization should designate an employee responsible for backups. That employee should ensure backup systems are set up correctly, test them periodically and ensure that critical data is actually backed up.
- **Backup scope and schedule**—an organization must decide on a backup policy, specifying which files and systems are important enough to be backed up, and how frequently data should be backed up.
- **Recovery Point Objective (RPO)**—RPO is the amount of data an organization is willing to lose if a disaster occurs, and is determined by the frequency of backup. If systems are backed up once per day, the RPO is 24 hours. The lower the RPO, the more data storage, compute and network resources are required to achieve frequent backups.

- **Recovery Time Objective (RTO)**—RTO is the time it takes for an organization to restore data or systems from backup and resume normal operations. For large data volumes and/or backups stored off-premises, copying data and restoring systems can take time, and robust technical solutions are needed to ensure a low RTO.

The Importance of a Disaster Recovery Plan: Alarming Statistics

To understand the potential impact of disasters on businesses, and the importance of having a data backup strategy as part of a complete disaster recovery plan, consider the following statistics:

- **Cost of downtime**—according to Gartner, the average cost of downtime to a business is $5,600 per minute.
- **Survival rate**—another Gartner study found only 6% of companies affected by a disaster that did not have disaster recovery in place survived and continued to operate more than two years after the disaster.
- **Causes of data loss**—the most common causes of data loss are hardware/system failure (31%), human error (29%) and viruses, and malware of ransomware (29%).

6 Data Backup Options

There are many ways to backup your file. Choosing the right option can help ensure that you are creating the best data backup plan for your needs. Below are six of the most common techniques or technologies:

1. Removable media
2. Redundancy
3. External hard drive
4. Hardware appliances
5. Backup software
6. Cloud backup services

1. **Removable Media**

A simple option is to backup files on removable media such as CDs, DVDs, newer Blu-Ray disks, or USB flash drives. This can be practical for smaller environments, but for larger data volumes, you'll need to back up to multiple disks, which can complicate recovery. Also, you need to make sure you store your backups in a separate location, otherwise they may also be lost in a disaster. Tape backups also fall into this category.

2. **Redundancy**

You can set up an additional hard drive that is a replica of a sensitive system's drive at a specific point in time, or an entire redundant system. For example, another email server that is on standby, backing up your main email server. Redundancy is a powerful technique but is complex to manage. It requires frequent replication between cloned systems, and it's only useful against the failure of a specific system unless the redundant systems are in a remote site.

3. **External Hard Drive**

You can deploy a high-volume external hard drive in your network, and use archive software to save changes to local files to that hard drive. Archive software allows you to restore files from the external hardware with an RPO of only a few minutes. However, as your data volumes grow, one external drive will not be enough, or the RPO will substantially grow. Using an external drive necessitates having it deployed on the local network, which is risky.

4. **Hardware Appliances**

Many vendors provide complete backup appliances, typically deployed as a 19" rack-mounted device. Backup appliances come with large storage capacity and pre-integrated backup software. You install backup agents on the systems you need to back up, define your backup schedule and policy, and the data starts streaming to the backup device. As with other options, try to place the backup device isolated from the local network and if possible, in a remote site.

5. **Backup Software**

Software-based backup solutions are more complex to deploy and configure than hardware appliances, but offer greater flexibility. They allow you to define which systems and data you'd like to back up, allocate backups to the storage device of your choice, and automatically manage the backup process.

6. **Cloud Backup Services**

Many vendors and cloud providers offer Backup as a Service (BaaS) solutions, where you can push local data to a public or private cloud and in case of disaster, recover data back from the cloud. BaaS solutions are easy to use and have the strong advantage that data is saved in a remote location. However, if using a public cloud, you need to ensure compliance with relevant regulations and standards, and consider that over time, data storage costs in the cloud will be much higher than the cost of deploying similar storage on-premises.

What Is a 3-2-1 Backup Strategy?

A 3-2-1 backup strategy is a method for ensuring that your data is adequately duplicated and reliably recoverable. In this strategy, three copies of your data are created on at least two different storage media and at least one copy is stored remotely:

- **Three copies of data**—your three copies include your original data and two duplicates. This ensures that a lost backup or corrupted media do not affect recoverability.
- **Two different storage types**—reduces the risk of failures related to a specific medium by using two different technologies. Common choices include internal and external hard drives, removable media, or cloud storage.
- **One copy off-site**—eliminates the risk associated with a single point of failure. Offsite duplicates are needed for robust disaster and data backup recovery strategies and can allow for failover during local outages.

This strategy is considered a best practice by most information security experts and government authorities. It protects against both accidents and malicious threats, such as ransomware, and ensures reliable data backup and restoration.

Server Backup: Backing Up Critical Business Systems

The easiest way to backup a server is with a server backup solution. These solutions can come in the form of software or appliances.

Server backup solutions are typically designed to help you backup server data to another local server, a cloud server, or a hybrid system. In particular, backup to hybrid systems is becoming more popular. This is because hybrid systems enable you to optimize resources, support easy multi-region duplication, and can enable faster recovery and failover.

In general, server backup solutions should include the following features:

- **Support for diverse file types**—should not include any file types. In particular, solutions should support documents, spreadsheets, media, and configuration files.
- **Backup location**—you should be able to specify backup locations. The solution should support backup to a variety of locations and media, including on and off-site resources.
- **Scheduling and automation**—in addition to enabling manual backups, solutions should support backup automation through scheduling. This helps ensure that you always have a recent backup and that backups are created in a consistent manner.
- **Backup management**—you should be able to manage the lifecycle of backups, including number stored and length of time kept. Ideally, solutions also enable easy export of backups for transfer to external resources or for use in migration.
- **Partition selection**—partitions are isolated segments of a storage resource and are often used to separate data within a system. Solutions should enable you to independently backup data and restore partitions.
- **Data compression**—to minimize the storage needed for numerous backups, solutions should compress backup data. This compression needs to be lossless and maintain the integrity of all data.
- **Backup type selection**—you should be able to create a variety of backup types, including full, differential, and incremental backups. Differential backups create a backup of changes since the last full backup while incremental records the changes since the last incremental backup. These types can help you reduce the size of your backups and speed backup time.

Scaling—backup abilities should not be limited by the volume of data on your servers. Solutions should scale as your data does and support backups of any size.

Backup Storage Technology

Whichever technique you use to backup, at the end of the day, data must be stored somewhere. The storage technology used to hold your backup data is very significant:

The more cost-effective it is, the more data it is able to store, and the faster the storage and retrieval over a network, the lower your RPO and RTO will be.

The more reliable the storage technology, the safer your backups will be.

Below, you'll find a review of backup storage technologies and their unique advantages.

Network Shares and NAS

You can set up centralized storage such as Network Attached Storage (NAS), Storage Area Network (SAN), or regular hard disks mounted as a network share using Network File System (NFS) protocol. This is a convenient option for making large storage available to local devices for backup. However, it is susceptible to disasters affecting your entire data center, such as natural disasters or cyberattacks.

Tape Backup

Modern tape technology such as Linear Tape-Open 8 (LTO-8) can store up to 9 TB of data on a single tape. You can then ship the tape to a distant location, preferably at least 100 miles away from your primary location. Tape backups have been used for decades, but their obvious downside is the extremely high RTO and RPO due to the need to physically ship the tapes to and from a backup location. They also require a tape drive and an autoloader to perform backup and recovery, and this equipment is expensive.

Cloud-Based Object Storage

When using cloud providers, you have access to a variety of storage services. Cloud providers charge a flat price per Gigabyte, but costs can start to add up for frequent access. There are multiple tools that let you backup data to S3 automatically, both from within the cloud and from on-premise machines.

Top 10 Backup Vendors in 2021

Backup and recovery system vendors review.

Today, there are quite a lot of different backup and recovery vendors, each with their own list of features and shortcomings. Sometimes choosing a correct vendor for your company might be difficult, even without the fact that each of the existing vendors is constantly changing, adding and removing features from their products. While this helps them to stay relevant and up-to-date, it makes choosing the right vendor even more difficult for businesses.

Backup and recovery vendors can offer an overwhelming amount of different features, including encryption, one-click failovers, mobile data security, remote control capabilities and more. Even so, each of the backup vendors has their own strengths and weaknesses, and most of the time businesses have to do a lot of research to make a competent choice based on their own wants and needs. One way to go about it is to look for the features your business needs the most at the lowest possible cost.

Few of the Backup vendors listed below :

Rubrik
Veeam
Cohesity
IBM
DELL Backup
Vembu
Veritas
NetApp
Commvault
Bacula Enterprise

Now, let's talk about each of them in more detail.

Rubrik

Rubrik is a backup and recovery vendor that specializes in working with data management and protection within hybrid IT environments. One of the most notable features of Rubrik is RCDM (Rubrik Cloud Data Management) - their own high-class data protection solution that specializes in cloud integration. There's also Polaris - Rubrik's data management platform that works under the SaaS model (Software as a Service) and consists of mainly two parts - Polaris GPS and Polaris Radar. Polaris GPS's main purpose is reporting and policy management, while Polaris Radar is famous for its ransomware detection and rehabilitation functions. It's also worth noting that Rubrik can be deployed via both physical hardware and with the usage of virtual appliances.
However, nothing is perfect, and the same goes for Rubrik – there are some shortcomings of the service that were reported by current users, like somewhat cost-heavy VM and app protection for smaller companies or not enough granular control when it comes to working with Active Directory.

Veeam Backup & Replication

Virtual environments of any size can be easily protected by Veeam Backup & Replication. It's a fast, flexible and reliable way of recovering your data and apps. There's a single solution that includes both backup and recovery, and also protects the data of your VMware VSphere and Hyper-V virtual environments. The solution itself is able to scale well and support everything up to entire virtual infrastructures with features like instant file-level recovery, built-in deduplication capabilities and such.
Veeam Backup & Replication as a software has several different editions – from the free community version to more complex enterprise and enterprise-plus editions. Of course, they are priced differently, and offer a different number of features. Veeam is one of the top backup and recovery system vendors on the market.

Cohesity

One specific feature that separates Cohesity from its competitors is its clusterlike design with nodes when it comes to managing enterprise data. The solution's scaling is managed through adding more nodes to the cluster (the base number is 3 nodes). Cohesity keeps backups in app-native formats and also uses NAS (Network-attached storage) protocols to export all kinds of data. Its primary use case is the backup of both VM and standard enterprise-grade apps' infrastructures.
One of the key features of the solution is the speed of data restoration – it is capable of restoring several VMs worth of data in a small amount of time, making for a very fast RTOs. At the same time, if you want granular recovery for MS Exchange or SharePoint – they're both covered by specific add-ons and are charged over the base product cost.

IBM Spectrum Protect

IBM Spectrum Protect is designed to simplify data protection no matter the storage type - be it physical storage, virtual storage or cloud-based environments, and irrespective of system type, from virtual machines and file servers to databases, mainframes and desktops. It also provides multiple other capabilities - from backup and recovery to disaster recovery and bare metal recovery. Its base is an agentless, easy to use virtual environment with a low cost and automation capabilities, and it can be installed in both Hyper-V and VMware environments.
Licensing costs of this product are significantly lower than some other backup vendors since they charge per-backend-TB that is consumed no matter the data type or application type. The most popular Spectrum Protect use cases are VMware, Hyper-V, Oracle, SQL Server, Db2 and more.

Dell EMC Data Protection Suite

Dell EMC provides a full-fledged data protection solution to meet the needs of any organization. Data protection suite includes variable data protection levels depending on the current business needs. Comprehensive UI allows for easy visualisation of data protection even across multiple websites or systems,

and in-built continuous data protection technology allows for fast recovery times in VM environments, even in the case of a disaster. You can choose the required data protection level yourself depending on current situation.
The software itself is packed with a variety of applications, allowing for a number of data storage types like NAS, tape, hardware snapshots and more. That also includes a separate backup into a cloud storage of your choosing as a means of preventing data loss in case of a disaster. At the same time, Dell EMC offers a high level of resilience via automatization of data isolation, data analytics and data recovery.

Vembu BDR Suite
Vembu BDR suite is a complex backup and disaster recovery solution that specializes in data protection across both physical and virtual environments. Data protection concerns of company of any size can be handled by them – from small businesses to entire enterprises and data centers. It also provides a number of features like VMware backup, Hyper-V backup, Windows Image backup and basic file/application backup. There are also several data replication capabilities – you can store another copy of your data either within Vembu's cloud (Cloud DR) or on Vembu's offsite server (Offsite DR).
Vembu also takes pride in their easy configuration and overall responsiveness of the system as a whole. At the same time, some users have been reporting several hiccups or problems within the system, such as the ability to easily corrupt your own database unless you really know what you're doing. Its reporting feature could also use some work.

Veritas Backup Exec
One of the Veritas's key points is their significant portfolio – it's a relatively old company that has been around for several decades and works within a lot of fields, including backup and recovery services, information governance, multi cloud data management and so on. Veritas's backup and recovery solution is able to work as a deployable software that is installed on customer's hardware, as well as an integrated highly capable appliance that specializes in backup storage and backup process control.
Due to its long history, Veritas is favored by a lot of larger companies with some legacy and traditional traits in them. One of their key points is their ability to work on a significantly large scale, supporting over one thousand clients at once within a single environment. However, while Veritas scales well in regards to storage capacity - there's little to no scaling efficiency in regards to the computing power capacity, which means that the clients would need additional computing appliances to work properly with larger data masses.

NetApp SnapCenter
If you're looking for a prominent data backup and recovery vendor with more than 150 offices around the world - NetApp SnapCenter is a solution for you. All of the typical backup and protection features are available for files, images, videos, etc. There's also the ability to have constant access to your data either via laptop or using your own mobile phone. Specifically SnapCenter has a centralized interface to have all of your monitoring, backup scheduling, logging and other tasks performed and viewed in one convenient place. It applies to enterprise databases and applications, MS Exchange servers, virtual machines and more.
Users also praise support's response time and overall system centralization. However, while the overall GUI centralization is highly regarded, it still needs some work when it comes to interface speed and overall customer satisfaction. Some users would also prefer for the documentation to be somewhat more descriptive to have the ability to work out some of the problems without the support's help.

Commvault
Commvault applies exclusive technologies in the field of data backup and recovery software to cover various data sources, multiple file types, backup types and storage types. For example, it allows you to backup your VM, database and endpoints with pinpoint accuracy with the best approach for the exact data type and recovery profile of your choosing. Commvault can recover VMs, backup both unstructured as well as structured

data and transfer data from cloud to cloud. It offers an unprecedented level of VM management and protection using a variety of its tools.
Commvault supports over twenty different cloud storage providers, including the most popular ones – VMware, Azure, AWS and so on. There is some significant room for improvement though, including UI friendliness, according to clients reviews.

Bacula Enterprise
Bacula Enterprise is a highly scalable, flexible backup and recovery software that allows for easy data backup, protection and recovery. It is a high grade enterprise solution for medium to large enterprises. As a backup solution, Bacula offers a vast number of different specialist features, such as an especially wide range of different backup storage types (including a high number of tape types and many Cloud storages), easy setup, low deployment costs and native SQL database support (MySQL, PostgreSQL and more). It also supports practically all of the popular Linux distributions, like Debian, Ubuntu, SUSE, RHEL and more, and backs up and recovery data from many other operating systems - such as Microsoft, MacOS X, Android and Solaris. It also natively integrates with an unusually wide range of Hypervisors - such as VMware, Red Hat Virtualization, Xen, Proxmox, Hyper-V and KVM. The software's overall performance and advanced deduplication and protection levels are also something to consider.
Security features are a special strength of Bacula: enterprise levels of security are available on every level and stage of operating with the data. Overall there is a very high availability. One of the prime points of Bacula's service is the ability to escape price traps per data volume. This is because Bacula does not charge by data volume, and it therefore frees up organizations - not only to grow with much less restriction, but to also plan ahead with far more clarity.

Backup Monitoring Tools

Today, server and network problems have to be identified immediately and as soon as they occur — and luckily there are tools to help. With thousands to hundreds of thousands of dollars hinging on access every hour, real money can be lost with a downed network. As a result, an administrator needs top notch server monitoring tools that also track multiple features of a network — from response time to memory usage, to uptime to storage all-in-one. Fortunately, there are some really good server monitoring tools available now to manage the above. Here are three worth considering:
1. Happy Apps
Whether you want to monitor apps, databases, servers, or your entire IT system, Happy Apps allows you to do just that, across private, public and **hybrid clouds**. With it, you can monitor things like performance issues, outages, stored data, and more. Happy Apps allows you to view both individual and overall statuses, so you know exactly what needs attention — you can even set up your own custom queries if there is a particular area you want to monitor more closely. Best of all, this server monitoring tool allows you to set up rules for alerts so that you can be notified via email or SMS text when there is an issue or outage. Additionally, you can achieve aggregate performance tracking using this tool, as reports are saved. Their easy to read dashboard makes it easy to find the information you need to track issues, incidents, and reports.
2. Performance Co-Pilot
Used by Netflix, Performance Co-Pilot is a system performance and analysis framework which can collect metrics from a variety of operating systems, from databases to web servers and Mail systems, in real-time or via historical data. It works with all the major OS platforms, from Linux to Mac to Windows once a Performance Metrics Domain Agent (PMDA), necessary for collecting domain performance metrics, has been installed. With its distributed system, Performance Co-Pilot enables a single desktop to monitor remote server systems with different operating systems and varying architectures. That's a bonus for network administrators on the go and away from their desks, and businesses with diverse, spread-out IT infrastructures.

3. Nagios

If you're looking for a tool that is highly customizable and allows you to build your own dashboard and alerts, Nagios Core is the tool for you. Nagios was built and designed on open source code and provides an administrator a desktop management view of the entire network and **IT infrastructure**. Nagios Core handles check scheduling, check processing, check execution, event handling, and alerting. However, you'll need to integrate other Nagios projects in order to gain additional features like performance graphing, auto-discovery, distributed monitoring, and processing performance data. With over 800 independently developed add-ons, over 4,000 plugins and plenty of documentation all located within the **Nagios Exchange**, you can build the exact server monitoring tool you need.

While server monitoring is important to stay on-top of performance issues and threats, it is also critical that you monitor your backup data. If something slips through the cracks, despite your best server and data monitoring practices, you'll need to rely on data backup to restore the data that's been lost or compromised.

Use a **cloud backup service** to continually and automatically backup all of your critical data as it is created and changed. Ensure that you use a cloud backup provider that offers unlimited previous file version histories so that if you need to revert to a specific previous file version, whether due to an accidental change or a ransomware virus, you will be able to. A provider like Nordic Backup, which offers military grade security, unlimited previous file versions histories, continuous, automatic backups and more can help you better monitor your backup data. With backup reports that will notify you of successful backups, as well as unusually high numbers of recent file changes, you can stay alert to backup failures, the presence of ransomware viruses, and more.

Windows Backup Lab

The following subcommands for **wbadmin** provide backup and recovery functionality from a command prompt. To configure a backup schedule, you must be a member of the **Administrators** group. To perform all other tasks with this command, you must be a member of the **Backup Operators** or the **Administrators** group, or you must have been delegated the appropriate permissions.

You must run **wbadmin** from an elevated command prompt. (To open an elevated command prompt, click **Start**, right-click **Command Prompt**, and then click **Run as administrator**.)

TABLE 1

Subcommand	Description
Wbadmin enable backup	Configures and enables a daily backup schedule.
Wbadmin disable backup	Disables your daily backups.
Wbadmin start backup	Runs a one-time backup. If used with no parameters, uses the settings from the daily backup schedule.
Wbadmin stop job	Stops the currently running backup or recovery operation.
Wbadmin get versions	Lists details of backups recoverable from the local computer or, if another location is specified, from another computer.
Wbadmin get items	Lists the items included in a specific backup.
Wbadmin start recovery	Runs a recovery of the volumes, applications, files, or folders specified.
Wbadmin get status	Shows the status of the currently running backup or recovery operation.
Wbadmin get disks	Lists disks that are currently online.
Wbadmin start systemstaterecovery	Runs a system state recovery.
Wbadmin start systemstatebackup	Runs a system state backup.

TABLE 1

Subcommand	Description
Wbadmin delete systemstatebackup	Deletes one or more system state backups.
Wbadmin start sysrecovery	Runs a recovery of the full system (at least all the volumes that contain the operating system's state). This subcommand is only available if you are using the Windows Recovery Environment.
Wbadmin restore catalog	Recovers a backup catalog from a specified storage location in the case where the backup catalog on the local computer has been corrupted.
Wbadmin delete catalog	Deletes the backup catalog on the local computer. Use this command only if the backup catalog on this computer is corrupted and you have no backups stored at another location that you can use to restore the catalog.

To create a one-time backup using this command, you must be a member of the **Backup Operators** group or the **Administrators** group, or you must have been delegated the appropriate permissions. In addition, you must run **wbadmin** from an elevated command prompt, by right-clicking **Command Prompt**, and then selecting **Run as administrator**.

Syntax

wbadmin start backup [-backupTarget:{<BackupTargetLocation> | <TargetNetworkShare>}] [-include:<ItemsToInclude>] [-nonRecurseInclude:<ItemsToInclude>] [-exclude:<ItemsToExclude>] [-nonRecurseExclude:<ItemsToExclude>] [-allCritical] [-systemState] [-noVerify] [-user:<UserName>] [-password:<Password>] [-noInheritAcl] [-vssFull | -vssCopy] [-quiet]

Parameters

PARAMETERS

Parameter	Description
-backupTarget	Specifies the storage location for this backup. Requires a hard disk drive letter (f:), a volume GUID-based path in the format of \\?\Volume{GUID}, or a Universal Naming Convention (UNC) path to a remote shared folder (\\<servername>\<sharename>\). By default, the backup will be saved at: \\<servername>\<sharename>\WindowsImageBackup\<ComputerBackedUp>\.
-include	Specifies the comma-delimited list of items to include in the backup. You can include multiple files, folders, or volumes. Volume paths can be specified using volume drive letters, volume mount points, or GUID-based volume names. If you use a GUID-based volume name, it should be terminated with a backslash (\). You can use the wildcard character (*) in the file name when specifying a path to a file. The **-include** parameter should only be used in conjunction with the **-backupTarget** parameter.
-exclude	Specifies the comma-delimited list of items to exclude from the backup. You can exclude files, folders, or volumes. Volume paths can be specified using volume drive letters, volume mount points, or GUID-based volume names. If you use a GUID-based volume name, it should be terminated with a backslash (\). You can use the wildcard character (*) in the file name when specifying a path to a file. The **-exclude** parameter should only be used in conjunction with the **-backupTarget** parameter.
-nonRecurseInclude	Specifies the non-recursive, comma-delimited list of items to include in the backup. You can include multiple files, folders, or volumes. Volume paths can be specified using volume drive letters, volume mount points, or GUID-based volume names. If you use a GUID-based volume

PARAMETERS

Parameter	Description
	name, it should be terminated with a backslash (\). You can use the wildcard character (*) in the file name when specifying a path to a file. The **-nonRecurseInclude** parameter should only be used in conjunction with the **-backupTarget** parameter.
-nonRecurseExclude	Specifies the non-recursive, comma-delimited list of items to exclude from the backup. You can exclude files, folders, or volumes. Volume paths can be specified using volume drive letters, volume mount points, or GUID-based volume names. If you use a GUID-based volume name, it should be terminated with a backslash (\). You can use the wildcard character (*) in the file name when specifying a path to a file. The **-nonRecurseExclude** parameter should only be used in conjunction with the **-backupTarget** parameter.
-allCritical	Specifies that all critical volumes (volumes that contain operating system's state) be included in the backups. This parameter is useful if you're creating a backup for bare metal recovery. It should be used only when **-backupTarget** is specified, otherwise the command fails. Can be used with the **-include** option. **Tip:** The target volume for a critical-volume backup can be a local drive, but it Can't be any of the volumes that are included in the backup.
-systemState	Creates a backup that includes the system state in addition to any other items that you specified with the **-include** parameter. The system state contains boot files (Boot.ini, NDTLDR, NTDetect.com), the Windows Registry including COM settings, the SYSVOL (Group Policies and Logon Scripts), the Active Directory and NTDS.DIT on Domain Controllers and, if the certificates service is installed, the Certificate Store. If your server has the Web server role installed, the IIS Metadirectory will be included. If the server is part of a cluster, Cluster Service information will also be included.
-noVerify	Specifies that backups saved to removable media (such as a DVD) are not verified for errors. If you do not use this parameter, backups saved to removable media are verified for errors.
-user	If the backup is saved to a remote shared folder, specifies the user name with write permission to the folder.
-password	Specifies the password for the user name that is provided by the parameter **-user**.
-noInheritAcl	Applies the access control list (ACL) permissions that correspond to the credentials provided by the **-user** and **-password** parameters to \\<servername>\<sharename>\WindowsImageBackup\<ComputerBackedUp>\ (the folder that contains the backup). To access the backup later, you must use these credentials or be a member of the Administrators group or the Backup Operators group on the computer with the shared folder. If **-noInheritAcl** is not used, the ACL permissions from the remote shared folder are applied to the \<ComputerBackedUp> folder by default so that anyone with access to the remote shared folder can access the backup.
-vssFull	Performs a full back up using the Volume Shadow Copy Service (VSS). All files are backed up, each file's history is updated to reflect that it was backed up, and the logs of previous backups may be truncated. If this parameter isn't used, **wbadmin start backup** makes a copy backup, but the history of files being backed up is not updated. **Caution:** Don't use this parameter if you are using a product other than Windows Server Backup to back up apps that are on the volumes included in the current backup. Doing so can potentially break the incremental, differential, or other type of backups that the other backup product is creating because the history that they are relying on to determine how much data to backup might be missing and they might perform a full backup unnecessarily.

PARAMETERS

Parameter	Description
-vssCopy	Performs a copy backup using VSS. All files are backed up but the history of the files being backup up is not updated so you preserve the all the information on which files where changed, deleted, and so on, as well as any application log files. Using this type of backup does not affect the sequence of incremental and differential backups that might happen independent of this copy backup. This is the default value. **Warning:** A copy backup can't be used for incremental or differential backups or restores.
-quiet	Runs the command without prompts to the user.

Remarks

- If you save your backup to a remote shared folder, and then perform another backup to the same computer and the same remote shared folder, you will overwrite your previous backup.
- If your backup operation fails, you can end up without a backup because the older backup is overwritten, but the newer backup isn't usable. To avoid this, we recommend creating subfolders in the remote shared folder to organize your backups. However, because of this organization, you must have twice the space available as the parent folder.

Examples

To create a backup of volumes *e:*, *d:\mountpoint*, and \\?\Volume{cc566d14-4410-11d9-9d93-806e6f6e6963}\ to volume *f:*, type:

Copy

wbadmin start backup -backupTarget:f: -include:e:,d:\mountpoint,\\?\Volume{cc566d14-44a0-11d9-9d93-806e6f6e6963}\

To perform a one-time backup of *f:\folder1* and *h:\folder2* to volume *d:*, to backup the system state, and to make a copy backup so the normally scheduled differential backup isn't impacted, type:

Copy

wbadmin start backup –backupTarget:d: -include:g\folder1,h:\folder2 –systemstate -vsscopy

To perform a one-time, non-recursive backup of *d:\folder1* to the \\backupshare\backup1* network location, and to restrict access to members of the **Administrators** or **Backup Operators** group, type:

Copy

wbadmin start backup –backupTarget: \\backupshare\backup1 -noinheritacl -nonrecurseinclude:d:\folder

Microsoft has brought us the Windows Server Backup since Windows Server 2008, as a replacement of NTBackup. Windows Server Backup is at the block level, which is more advanced than the file and folder level of NTBackup. As a great improvement, it allows you to perform backups or restores in a faster, simpler, and more reliable way. You can configure a backup or recovery job via the graphic user's interface (GUI) or the command line. This article will show you how to use the Windows Server Backup command line tool WBadmin.exe in a few ways.

However, this backup utility is not pre-installed by default on Windows server 2019, 2016, 2012 or 2008. To be able to use it, you need to add both the Windows Server Backup and Command-line Tools in Server Manager. With the WBadmin.exe utility, you can do even more things from command line than that you do from the GUI.

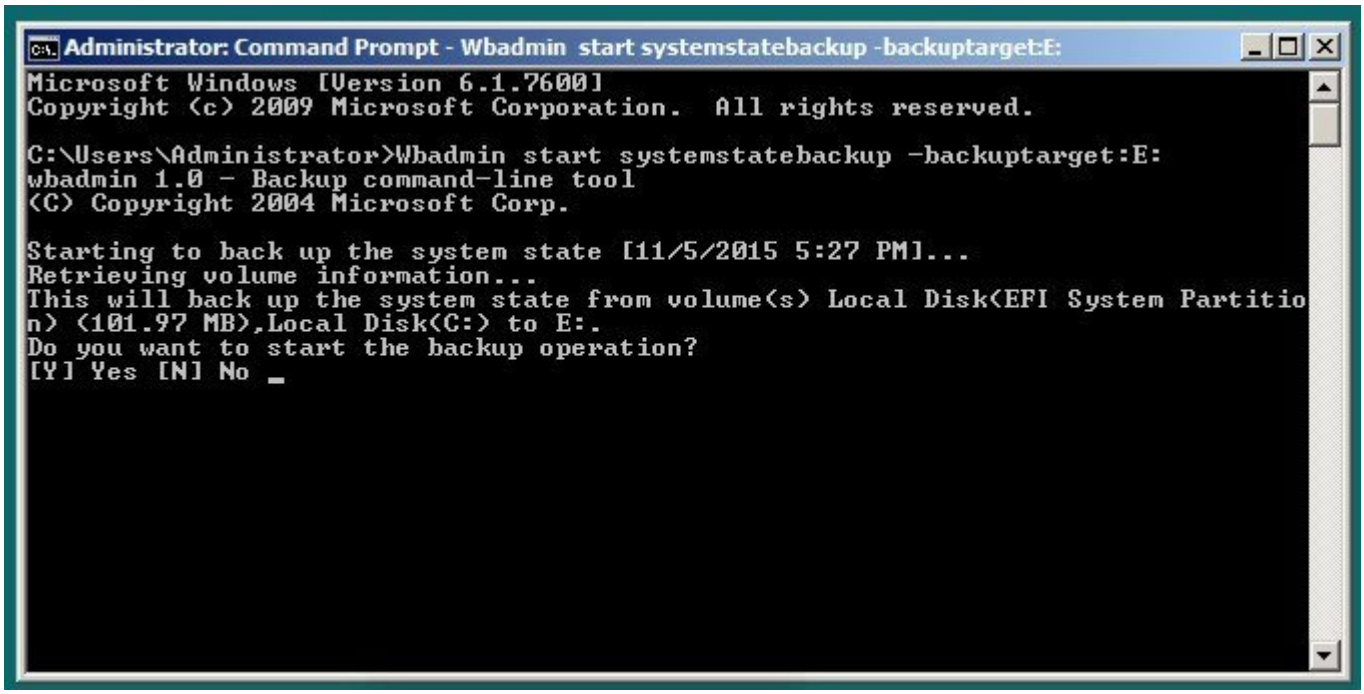
Actually, besides the Windows Server backup command line tool, you could also use the Windows Server Backup alternative to create backups with cmd. What's more, it offers more functions that WSB (Windows Server Backup) does not support, for example, it lets you set multiple scheduled backups (daily/weekly/monthly/Event triggers/USB plug in) while the WSB only supports one daily backup task. Refer to Part 2 to get it.

Part 1. Windows Server backup command line examples

To use Wbadmin.exe, you need to open an elevated command prompt with administrator permission, and then run a backup command based on your needs. To access an elevated command prompt, you can enter "CMD" in a Run box or right-click Start and select Command Prompt (Admin). With the command prompt, you can write a command to run a backup or recovery job referring to following WBadmin examples:

1. To create a system state backup to E: drive:

Wbadmin start systemstatebackup -backuptarget:E:



2. To create wbadmin backup to network share including all critcal volumes, system state, and "chrun" folder in D:drive to network shared folder:
Wbadmin start backup -allcritical -systemstate -include:D:\chrun -backuptarget:\\networkshare\backup -quiet

3. To create a scheduled backup job at 00:00 daily to the shared folder that can only be accessed with username and password:
wbadmin enable backup -addtarget:\\192.168.0.225\Public\schedule -include:D: -systemstate -user:admin -password:1234 -schedule:00:00

4. To restore a full server backup to dissimilar hardware with version:11/11/2015-00:00 to another server with machine name Varlar:
Wbadmin start sysrecovery –version: 11/11/2015-00:00 –backuptarget:E: -machine:Varlar

5. To restore system state with version: 01/11/2016-01:09:
wbadmin start systemstaterecovery -version:01/11/2016-01:09 -backupTarget: F: -quiet
Note:

- The switch: -quiet means to run the task without being prompted to enter "Y" to confirm.
- Beside backup and restore, you can still delete old backup with Windows Server Backup.

Although the Wbadmin command line tool is convenient for those who know how to use it, it does have some restrictions, such as:

- Only partitions formatted with NTFS can be used as backup source or backup target,
- You cannot configure a scheduled backup less frequent than daily.

How system state backup works

1. When a system state backup runs, DPM communicates with WSB request a backup of the server's system state. By default DPM and WSB will use the drive with the most available free space, and information about this drive is saved in the PSDataSourceConfig.XML file. This is the drive WSB will use to do backups to.
2. You can customize the drive that DPM uses for the system state backup. To do this on the protected server, go to *drive*:\Program Files\Microsoft Data Protection Manager\DPM\Datasources. Open the PSDataSourceConfig.XML file for editing. Change the <FilesToProtect> value for the drive letter. Save and close the file. If a protection group protects the computer's system state, run a consistency check. If the consistency check generates an alert, click **Modify protection group** link in the alert, and then step through the wizard. After finishing, run another consistency check.
3. Note that if the protection server is in a cluster it's possible that a cluster drive will be selected as the drive with the most free space. It's important to be aware of this because if that drive ownership has been switched to another node and a system state backup runs, the drive won't be available and the backup will fail. In this situation, you'll need to modify the PSDataSourceConfig.XML to point it to a local drive.
4. Windows Server Backup (WSB) creates a folder called WindowsImageBackup on the root of the. As it creates the backup, all data is placed in this folder. When the backup completes the file will then be transferred over to the DPM server. Note that:
    - This folder and its contents do not get cleaned up after the backup or transfer is done. The best way to think of this is that the space is being reserved for the next time a backup is done.
    - The folder gets created every time a backup is done. The time/date stamp will reflect the time of your last system state backup..

BMR backup

1. For BMR (including a system state backup) the backup job is performed directly to a share on the DPM server and not to a folder on the protected server.
2. DPM server calls WSB and shares out the replica volume for that BMR backup. In this case it doesn't tell WSB to use the drive with the most free space, but instead to use the share created for the job.
3. When the backup finishes the file is transferred to the DPM server. Logs are stored in C:\Windows\Logs\WindowsServerBackup.

Prerequisites and limitations

- BMR isn't supported for computers running Windows Server 2003 or for computers running client operating systems.
- You can't protect BMR and system state for the same computer in different protection groups.
- A DPM server can't protect itself for BMR.
- Short-term protection to tape (D2T) isn't supported for BMR. Long-term storage to tape (D2D2T) is supported.
- Windows Server Backup must be installed on the protected computer for BMR.
- For BMR protection (unlike system state protection) DPM doesn't have any space requirements on the protected computer. WSB directly transfers the backups to the DPM server. Note that the job for this doesn't appear in the DPM Jobs view.
- If you use Modern Backup Storage and want to increase the BMR default replica size > 30 GB, use the registry key: HKLM\Software\Microsoft\Microsoft Data Protection Manager\Configuration ReplicaSizeInGBForSystemProtectionWithBMR (DWORD).

- If you use Modern Backup Storage, SystemState and BMR backups consume more storage (than legacy storage) due to ReFS cloning. Each SystemState or BMR backup is a full recovery point. To mitigate this storage consumption, you may want to:
  - schedule fewer System State or BMR recovery points,
  - use a smaller retention period for the recovery points,
  - increase the available storage for System State or BMR backups.
- DPM reserves 30 GB of space on the replica volume for BMR. You can change this on the Disk Allocation page in the Modify Protection Group Wizard or using the Get-DatasourceDiskAllocation and Set-DatasourceDiskAllocation PowerShell cmdlets. On the recovery point volume, BMR protection requires about 6 GB for retention of five days. Note that you can't reduce the replica volume size to less than 15 GB. DPM doesn't calculate the size of BMR data source, but assumes 30 GB for all servers. Admins should change the value as per the size of BMR backups expected on their environments. The size of a BMR backup can be roughly calculated sum of used space on all critical volumes: Critical volumes = Boot Volume + System Volume + Volume hosting system state data such as AD. Process System state backup
- If you move from system state protection to BMR protection, BMR protection will require less space on the **recovery point volume.** However, the extra space on the volume is not reclaimed. You can shrink the volume size manually from the **Modify Disk Allocation** page of the **Modify Protection Group Wizard** or using the Get-DatasourceDiskAllocation and Set-DatasourceDiskAllocation cmdlets.
  If you move from system state protection to BMR protection , BMR protection will require more space on the **replica volume**. The volume will be extended automatically. If you want to change the default space allocations you can use Modify-DiskAllocation.
- If you move from BMR protection to system state protection you'll need more space on the recovery point volume. DPM might try to automatically grow the volume. If there is insufficient space in the storage pool, an error will be issued.
  If you move from BMR protection to system state protection you'll need space on the protected computer because system state protection first writes the replica to the local computer and then transfers it to the DPM server

Before you start

1. **Deploy DPM**: Verify DPM is deployed correctly. If you haven't see:
   - System requirements for DPM
   - What can DPM back up?
   - What's supported and what isn't for DPM?
   - Get DPM installed
2. **Set up storage**-You can store backed up data on disk, on tape, and in the cloud with Azure. Read more in Prepare data storage.
3. **Set up the DPM protection agent**-You'll need to install the DPM protection agent on machine you want to back up. Read Deploy the DPM protection agent

Back up system state and bare metal

Set up a protection group as described in Deploy protection groups. Note that you can't protect BMR and system state for the same machine in different groups, and that when you select BMR system state is automatically enabled.

1. Click **Protection** > **Actions** > **Create Protection Group** to open the **Create New Protection Group** wizard in the DPM console.
2. In **Select protection group** type click **Servers**.
3. In **Select Group Members** expand the machine and select **BMR** or **system state**
   Remember that you can't protect BMR and system state for the same machine in different groups, and that when you select BMR system state is automatically enabled. Learn more in Deploy protection groups.
4. In **Select data protection method** specify how you want to handle short and long-term backup. Short-term backup is always to disk first, with the option of backing up from the disk to the Azure cloud with

Azure backup (for short or long-term). As an alternative to long-term backup to the cloud you can also configure long-term back up to a standalone tape device or tape library connected to the DPM server.

5. In **Select short-term goals** specify how you want to back up to short-term storage on disk. In Retention range you specify how long you want to keep the data on disk. In Synchronization frequency you specify how often you want to run an incremental backup to disk. If you don't want to set a back-up interval, you can check, just before a recovery point so that DPM will run an express full backup just before each recovery point is scheduled.
6. If you want to store data on tape for long-term storage in **Specify long-term goals** indicate how long you want to keep tape data (1-99 years). In Frequency of backup specify how often backups to tape should run. The frequency is based on the retention range you've specified:
    - When the retention range is 1-99 years, you can select backups to occur daily, weekly, bi-weekly, monthly, quarterly, half-yearly, or yearly.
    - When the retention range is 1-11 months, you can select backups to occur daily, weekly, bi-weekly, or monthly.
    - When the retention range is 1-4 weeks, you can select backups to occur daily or weekly.

    On a stand-alone tape drive, for a single protection group, DPM uses the same tape for daily backups until there is insufficient space on the tape. You can also colocate data from different protection groups on tape.

    On the **Select Tape and Library Details** page specify the tape/library to use, and whether data should be compressed and encrypted on tape.
7. In the **Review disk allocation** page review the storage pool disk space allocated for the protection group. **Total Data size** is the size of the data you want to back up, and **Disk space to be provisioned on DPM** is the space that DPM recommends for the protection group. DPM chooses the ideal backup volume, based on the settings. However, you can edit the backup volume choices in the **Disk allocation details**. For the workloads, select the preferred storage in the dropdown menu. Your edits change the values for **Total Storage** and **Free Storage** in the **Available Disk Storage** pane. Underprovisioned space is the amount of storage DPM suggests you add to the volume, to continue with backups smoothly in the future.
8. In **Choose replica creation method** select how you want to handle the initial full data replication. If you select to replicate over the network we recommended you choose an off-peak time. For large amounts of data or less than optimal network conditions, consider replicating the data offline using removable media.
9. In **Choose consistency check options**, select how you want to automate consistency checks. You can enable a check to run only when replica data becomes inconsistent, or according to a schedule. If you don't want to configure automatic consistency checking, you can run a manual check at any time by right-clicking the protection group in the **Protection** area of the DPM console, and selecting **Perform Consistency Check**.
10. If you've selected to back up to the cloud with Azure Backup, on the **Specify online protection data** page make sure the workloads you want to back up to Azure are selected.
11. In **Specify online backup schedule** specify how often incremental backups to Azure should occur. You can schedule backups to run every day/week/month/year and the time/date at which they should run. Backups can occur up to twice a day. Each time a back up runs a data recovery point is created in Azure from the copy of the backed-up data stored on the DPM disk.
12. In **Specify online retention policy** you can specify how the recovery points created from the daily/weekly/monthly/yearly backups are retained in Azure.
13. In **Choose online replication** specify how the initial full replication of data will occur. You can replicate over the network, or do an offline backup (offline seeding). Offline backup uses the Azure Import feature. Read more.

14. On the **Summary** page review your settings. After you click **Create Group** initial replication of the data occurs. When it finishes the protection group status will show as **OK** on the **Status** page. Backup then takes place in line with the protection group settings.

Recover system state or BMR

You can recover BMR or system state to a network location. If you've backed up BMR use the WIndows Recovery Environment (WinRE) to start up your system and connect it to the network. Then use Windows Server Backup to recover from the network location. If you've backed up system state just use Windows Server Backup to recover from the network location.

Restore BMR

Run recovery on the DPM server:

1. In the Recovery pane find the machine you want to recovery > Bare Metal Recovery.
2. Available recovery points are indicated in bold on the calendar. Select the date and time for the recovery point you want to use.
3. In **Select Recovery Type** select **Copy to a network folder.**
4. In **Specify Destination** select where you want to copy the data to. Remember that the selected destination will need enough room. We recommend a new folder.
5. In **Specify Recovery Options** select the security settings to apply and select whether you want to use SAN-based hardware snapshots for quicker recovery (only an option if you have a SAN with this functionality enabled and the ability to create and split a clone to make it writable. In addition the protected machine and DPM server must be connected to the same network).
6. Set up notification options and click **Recover** on the **Summary** page.

Set up the share location:

1. In the restore location navigate to the folder that contains the backup.
2. Share the folder above WindowsImageBackup so that the root of the shared folder is the WindowsImageBackup folder. If it isn't restore won't find the backup. To connect using WinRE you'll need a share that you can access in WinRE with the correct IP address and credentials.

Restore the system:

1. Start the machine for which you want to restore the image to using the Windows DVD to match the system you are restoring.
2. On the first screen verify language/locale settings. .On the **Install** screen select **Repair your computer**.
3. On the **System Recovery Options** page select **Restore your computer using a system image that you created earlier**
4. On the **Select a system image backup** page select **Select a system image** > **Advanced** > **Search for a system image on the network**. Select **Yes** if a warning appears. Navigate to the share path, input the credentials, and select the recovery point. This scans for specific backups available in that recovery point. Select the recovery point.
5. In **Choose how to restore the backup** select **Format and repartition disks**. In the next screen verify settings and click **Finish** to begin the restore. Restart as required.

Restore system state

Run recovery on the DPM server:

1. In the Recovery pane find the machine you want to recovery > Bare Metal Recovery.
2. Available recovery points are indicated in bold on the calendar. Select the date and time for the recovery point you want to use.
3. In **Select Recovery Type** select **Copy to a network folder.**
4. In **Specify Destination** select where you want to copy the data to. Remember that the selected destination will need enough room. We recommend a new folder.
5. In **Specify Recovery Options** select the security settings to apply and select whether you want to use SAN-based hardware snapshots for quicker recovery (only an option if you have a SAN with this functionality enabled and the ability to create and split a clone to make it writable. In addition the protected machine and DPM server must be connected to the same network).

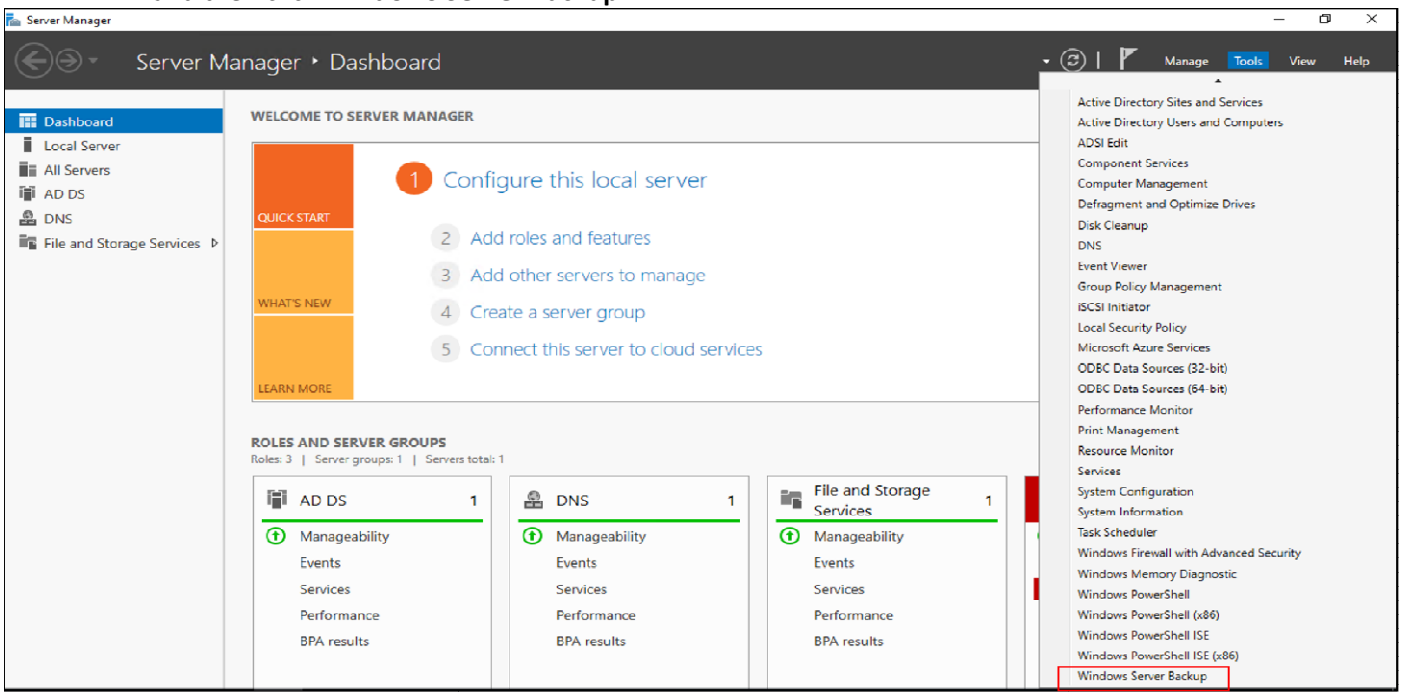
6. Set up notification options and click **Recover** on the **Summary** page.

Run Windows Server Backup

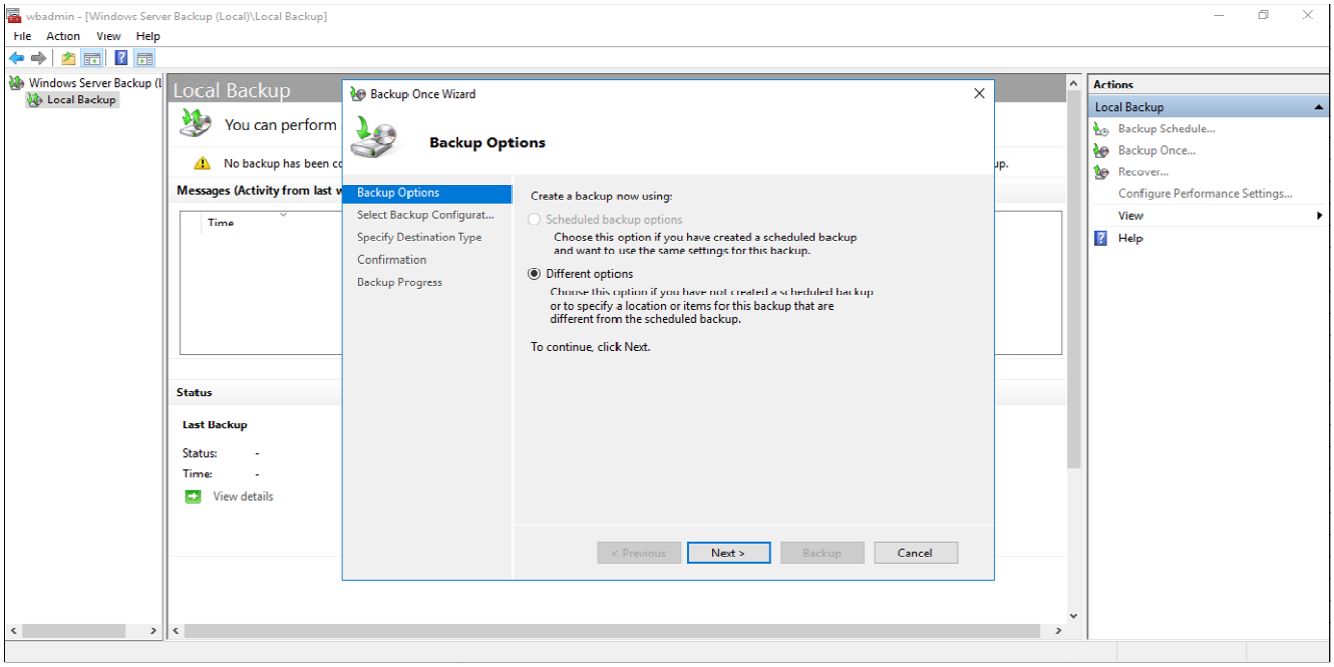
1. Click **Actions** > **Recover** > **This Server** > **Next.**
2. Click **Another Server** > **Specify Location Type** page > **Remote shared folder**. Provide the path to the folder that contains the recovery point.
3. In **Select Recovery Type** click **System state**. In **Select Location for System State Recovery** click **Original Location**
4. In **Confirmation** click **Recover**. You'll need to restart the server after the restore.
5. You can also run a system state restore from the command-line. To do this start Windows Server Backup on the machine you want to recover. From a command prompt type: **wbadmin get versions -backuptarget <servername\sharename>** to get the version identifier.
   Use the version identifier to start system state restore. At the command line type: **wbadmin start systemstaterecovery -version:<versionidentified> -backuptarget:<servername\sharename>** Confirm that you want to start the recovery. You can see the process in the command window. A restore log is created. You'll need restart the server after the restore.

To perform a system state backup using Windows Server Backup

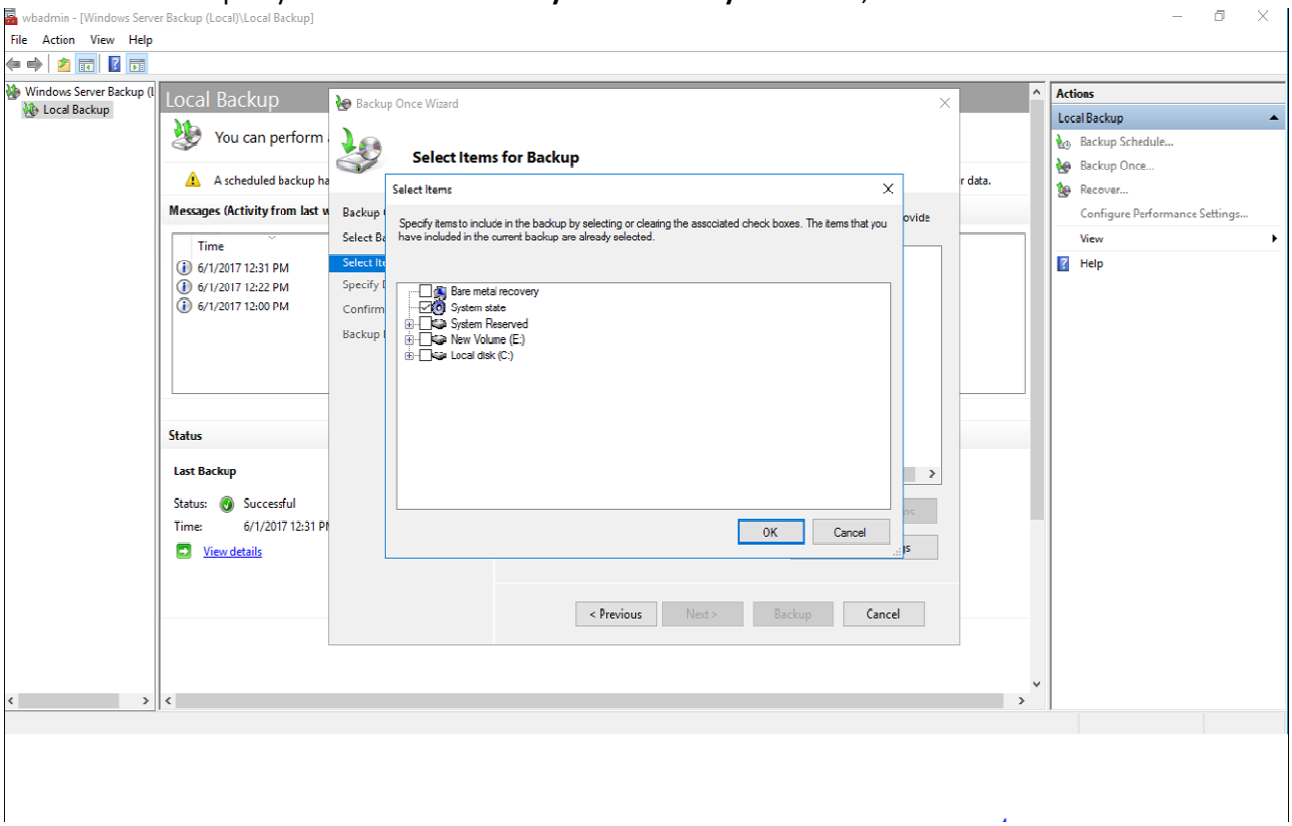
1. Open **Server Manager**, click **Tools**, and then click **Windows Server Backup**.
   - In Windows Server 2008 R2 and Windows Server 2008, click **Start**, point to **Administrative Tools**, and then click **Windows Server Backup**.



2. If you are prompted, in the **User Account Control** dialog box, provide Backup Operator credentials, and then click **OK**.
3. Click **Local Backup**.
4. On the **Action** menu, click **Backup once**.
5. In the Backup Once Wizard, on the **Backup options** page, click **Different options**, and then click **Next**.



6. On the **Select backup configuration** page, click **Custom)**, and then click **Next**.
7. On the **Select Items for Backup** screen, click **Add Items** and select **System State** and click **Ok**.
    - In Windows Server 2008 R2 and Windows Server 2008, select the volumes to include in the backup. If you select the **Enable system recovery** check box, all critical volumes are selected.



8. On the **Specify destination type** page, click **Local drives** or **Remote shared folder**, and then click **Next**. If you are backing up to a remote shared folder, do the following:
    - Type the path to the shared folder.
    - Under **Access Control**, select **Do not inherit** or **Inherit** to determine access to the backup, and then click **Next**.

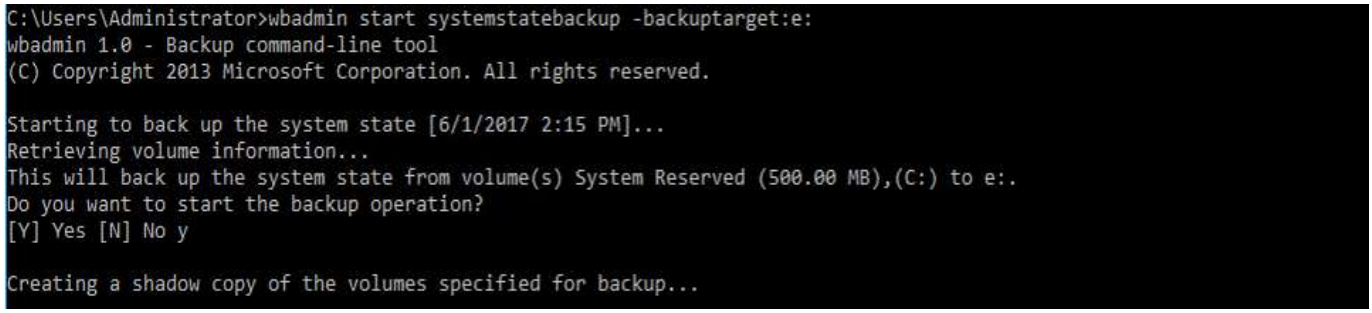
- In the **Provide user credentials for Backup** dialog box, provide the user name and password for a user who has write access to the shared folder, and then click **OK**.

9. For Windows Server 2008 R2 and Windows Server 2008, on the **Specify advanced option** page, select **VSS copy backup** and then click **Next**.
10. On the **Select Backup Destination** page, choose the backup location. If you selected local drive choose a local drive or if you selected remote share choose a network share.
11. On the confirmation screen, click **Backup**.
12. Once this has completed click **Close**.
13. Close Windows Server Backup.

To perform a system state backup using Wbadmin.exe

Open an elevated command prompt, type the following command and press ENTER:

wbadmin start systemstatebackup -backuptarget:<targetDrive>:



10. Cloud

Introduction to Cloud Computing

What is cloud computing?

To put it very simply, cloud computing is the delivery of computing resources as a service. Moving to the cloud basically means that the resources are owned and managed by a third-party provider, instead of the end-user.

This means that you don't need to worry about hard drives, main-frames, or where any of this hardware and software is located. As far as you, the user, is concerned, it's floating up there in a metaphorical 'cloud' – which you're able to access via the internet.

This shift from software and hardware that was on-premises to a networked, remote resource has meant that companies no longer have to worry about investing in labour, expertise, or capital for the maintenance of these resources. It has spawned a plethora of cloud computing companies, including key players like AWS and Microsoft Azure.

Types of cloud computing services

Cloud computing services are delivered in three main models, each of which offers customers different levels of support and flexibility. There's also some overlap between all three of them, so it can get a little confusing when trying to get your head around what they all mean.

These services are occasionally known as the cloud computing 'stack' as they are often built on top of one another. Knowing what each one of them is, and how they work, will give you a clearer sign about which service might be best suited to your needs and requirements.

Infrastructure as a Service (IaaS)

Also known as utility computing, this is the on-demand delivery of computing infrastructure. That means everything – from operating systems and storage to networking and components – is outsourced to a cloud computing company or service. As the individual or the company, you'll buy what you need on a pay-as-you-go model.

The simplest example of IaaS cloud computing is ordinary web-hosting. This is where you pay a monthly fee or by megabyte/gigabyte to have a company host your files from their servers. IaaS is an extremely flexible option, as it permits the user to customise the infrastructure of the computing environment. From web-hosting to big data analytics, IaaS covers the whole spectrum.

Software as a Service (SaaS)

This is when you use a complete application on a third-party server or system. Users can access these applications on-demand via the internet, without having to download or maintain any software. SaaS cloud tech is really popular with businesses and general users as it's usually easy to adopt. It can also be accessible from any device, and there is often a range of paid or free options to choose from.

Examples of SaaS applications include any web-based mail services. The different services supplied by Google such as Google Docs and Google Sheets are also examples of SaaS. Adobe Creative Cloud services is also another example of SaaS in action. With this kind of model, the user is only exposed to the interface that they choose to interact with.

Platform as a Service (PaaS)

This form of cloud computing is often used by software developers who are looking to focus on development rather than DevOps and administration. It's effectively an option to develop an application without having to worry about installing, configuring, and maintaining an infrastructure. This is supplied by the server as a standardised environment.

PaaS cuts down on the complexity of setting up and properly maintaining an infrastructure, while also allowing for supported collaboration between teams. An example of this is if you develop your own commerce site, but basically have the entire process running on a separate server. Like with SaaS, you're only exposed to the interface you interact with.

Types of cloud environments

On top of the different cloud systems, you can also get cloud environments. Not all clouds are the same, and the different types of cloud will suit different domains and how the cloud service itself is 'deployed'. These different types have been developed to try and suit as many different singular needs as possible.

You may also come across personal clouds and peer-to-peer clouds. These are smaller and more bespoke – below, we have a look at the main types of cloud environment. Let's have a look at the different types available to you.

Public cloud

Public cloud environments are operated by third-party providers. They provide computing resources such as servers and storage options using the Internet. While this type of cloud service isn't necessarily best suited for regulated industries like the healthcare sector, they could suit smaller businesses.

One of the biggest public cloud servers is Microsoft Azure, which owns and manages huge hardware and software infrastructure which you, as the user, can access online.

Private cloud

This type of cloud environment is owned and managed by one client. This means that only the client's employees can have access to this cloud system. A private cloud permits you to have much greater control over your computing environment and data and is commonly implemented in regulated industries like finance.

Private clouds are usually physically located in an organization's office building, but sometimes third-party services are also employed. They're much more secure than public clouds.

Hybrid cloud

Sometimes known as multi-clouds, hybrid clouds are basically a combination of private and public clouds. These clouds basically allow you to move information and data between the private and public clouds. This can give your organisation much greater flexibility and can optimise your infrastructure.

Amazon Web Services, or AWS, are one of the biggest companies that offer hybrid cloud solutions.

Uses of cloud computing

What you may not realise is that you're probably using a form of cloud computing right now. If you have been using an online email server, or you use Google docs, or if you even watch TV and listen to music on the internet, you have inadvertently been using forms of cloud computing.

This fact is made all the more remarkable because even the earliest cloud computing systems are barely 10 years old. Yet already, organisations big and small have been migrating to the cloud owing to the many different things that cloud computing offers the user. Let's take a look at the different ways people use cloud computing.

Data storage

This is arguably the most common use for cloud computing. Large organisations will end up amassing huge quantities of data which all need to be stored somewhere. Acquiring the necessary main-frames to store this quantity of data would end up being very expensive. Cloud computing offers a more cost-efficient storage solution.

Audio and video streaming

Connecting with an audience has been made really straightforward by the use of cloud computing. Some people will remember when Netflix was delivered through your letterbox – but thanks to cloud computing, your favourite movie can be beamed through to you via just about any device. The same goes for Spotify as well – no more trips to the library to borrow CDs!

Data analysis

Cloud computing allows you to unify all your teams and all their data, wherever they might be in the world. You can then use machine learning or AI to analyse all this data using Python and uncover all sorts of different insights which will lead you to more informed decision making and problem-solving.

App development

By using pre-made cloud computing infrastructures, developers can drastically reduce the time and the cost of application development. You can quickly build, deploy, and scale applications using cloud-native technologies and approaches for web, mobile, and API, as well as using Python for programming too.

Advantages and disadvantages of cloud computing

So now we've taken a look at the different things you can get from cloud computing, let's have a look at some of the advantages and disadvantages. As with anything in the digital domain, there are good bits and bad bits. And as cloud computing is still quite a new phenomenon, we're still learning the ins and outs of it.

Advantages

Lower upfront costs & reduced infrastructure costs

Running and maintaining a mainframe storage system is an extremely costly venture. By delegating this responsibility to a third party, who has all the technology and expertise at their fingertips, you'll end up saving money, especially in your upfront costs. Plus, not having to maintain your own infrastructure in-house will also save you time and money.

Lower carbon emissions

Companies the world over are looking for ways to reduce their carbon footprint. By having people share a centralised cloud computing system that is efficiently run, you'll end up using a lot less energy than you would if they all had their own system. AWS even claimed that cloud computing is capable of achieving carbon emissions savings of 88%. But the more people who use cloud computing, the more power inevitably ends up getting used.

Easy to scale up or down

As it can be difficult to predict the growth rates and success rates of apps as they're created, it can be really useful to be able to adjust your cloud computing capabilities accordingly. Cloud computing has been developed to scale quickly and handle unexpected growth, with more storage options being available at the click of the mouse.

Only pay for what you use

Many cloud computing services are based on a pay-as-you-go model. This means that you have an upfront cost of how much you think you use, which you can then adjust as and when you need to. If you haven't used as much storage as you thought you needed, it's very easy to downscale. More often than not, your service provider will alert you to this fact too.

Disadvantages

Ongoing operating costs

While you'll have access to computing services at a fraction of what it might cost you to have your own, over time, these expenses can rack up. If you have experienced unexpected growth, and your storage usage goes over capacity, your service provider could well charge you a premium for this usage.

Security

Owing to the use of API's and cloud-based credentials, there are more security vulnerabilities that come with cloud computing. These risks can come from both potential attackers as well as the fact that you are trusting a third party in an unknown location with potentially sensitive and private information. Learn about why cyber security is important, and set yourself up with the foundations of cyber security.

Dependency on Internet connection

Access to the cloud is only possible via an internet connection. This fact alone can end up being quite limiting, especially if you find yourself without the opportunity to connect. If your cloud system ends up losing connection, your entire organisation will end up being crippled. And without a physical backup system in place, chaos can ensue.

Vendor lock-in

This is the condition where it becomes difficult, or occasionally impossible, to change cloud computing service vendors due to the computing systems you have in place being closed and proprietary. Migrating to a different cloud computing server is tricky in itself, but having too much of a structure in your current system can make it even harder.

Cloud Computing Concepts

Cloud Computing

Let's start our discussion of cloud computing concepts with the most important term, "Cloud computing" itself. Often referred to simply as 'cloud,' this is web-based computing that provides on-demand access to various

compute resources. These resources include things like data centers, servers, application software and more. Most cloud computing service providers adopt a pay-as-you-go model. This allows companies to avoid the heavy infrastructure setup costs that were inevitable before the advent of the cloud.

XaaS (Anything-As-A-Service)
XaaS is another term you'll often come across and it relates to what is known as cloud service models. This is a general category of services related to remote access and cloud computing. It includes a wide selection of technologies, tools and products that are offered to users as a service over the web.

Basically, any IT function can be changed into a service for business consumption. The service is not paid as a license or upfront purchase, but in a flexible pay-for-what-you-use consumption model.

XaaS comes with several benefits, including speeding up business processes and application development, improving the expense model (CAPEX vs OPEX), and streamlining operations to free up resources for innovation.

SaaS (Software-As-A-Service)
SaaS is a software distribution model where a third-party provider hosts a variety of applications and makes them available to customers on the internet.

In this model, the provider gives clients network-based access to one copy of an application created for SaaS distribution. The source code of the application is similar for all customers. When an update is rolled out, it is distributed to all customers.

Depending on the implementation, customers can store data in the cloud, locally, or both. Companies can use application programming interfaces (APIs) to integrate SaaS applications with other software to achieve optimal benefits.

SaaS has several benefits, including flexibility in payments thanks to its pay-as-you-go model, scalable usage, easy accessibility and persistence and automatic updates.

PaaS (Platform-As-A-Service)
PaaS is one of the major cloud computing models where a third-party provider delivers software and hardware tools over the web. More often than not, these tools are used for app development. The provider hosts these tools on their own infrastructure. Therefore, PaaS allows developers to work without investing in in-house hardware and software, which saves costs!

Something worth noting is that PaaS doesn't replace your IT infrastructure for software development. Rather, it provides you with a simple and convenient way to get the job done in a timely fashion. It is provided through a cloud service and users access offerings through their web browsers. PaaS can be offered through hybrid, private, mobile or public clouds to deliver a host of services, including java development and application hosting.

Some examples of PaaS services include:

AWS Elastic Beanstalk (managed web application)
Microsoft Azure Web Apps (managed web application)
Amazon Relational Database Service (managed database service)
OpenShift (container platform)

Normally, you will pay for this service on a pay-as-you-go model, although some providers charge a monthly fee.

PaaS comes with several benefits, including eliminating the capital expenses companies used to incur for in-house hardware and software. It can also speed up application deployment.

IaaS (Infrastructure-As-A-Service)
IaaS delivers virtual computing infrastructure that is delivered and managed over the web. This model scales with demand and allows companies to only spend on what they use. It helps businesses avoid the complexity and expense of purchasing and managing in-house data center infrastructure.

There are lots of things businesses can do with IaaS, including:

Test and deployment – teams can swiftly dismantle development and test (dev-test) environments and create new applications faster.
Web apps – IaaS provides a host of resources and infrastructure needed for the development of web apps such as networking resources, application servers and storage.
Big data analysis – Big data is one of the most popular terms you will come across. IaaS provides the infrastructure and processing power needed to mine big data.
High-performance computing – Commonly referred to as HPC, helps solve complex problems involving millions of calculations or variables. This includes evaluating product designs, financial modeling, climate and weather predictions, and many more. IaaS provides the infrastructure to accomplish this.
Web hosting – Running websites using IaaS is easier, more convenient and less costly than doing on traditional platforms.
IaaS increases reliability, supportability, and stability. Additionally, it enables users to scale up resources quickly and respond to shifting business conditions.

Serverless
Serverless is a cloud computing model that allows developers to build modern applications without thinking about servers. It allows them to shift most of their operational responsibilities to the provider. By using this model, you can concentrate on writing application code while the platform takes care of resource allocation, run times, security, scaling and other server solutions.

Serverless eliminates a range of infrastructure management tasks like patching, operating system maintenance, server provisioning and more. This goes a long way in saving costs in all aspects of application development.

Examples of serverless cloud computing services include AWS Lambda and Microsoft Azure Functions.

Public Cloud
Public cloud is a type of cloud deployment. It refers to computing services delivered by third-party providers over the public internet in a multi-tenant model.

Public cloud services are made available to anyone who wants to buy or use them. The services are often sold on-demand, therefore allowing clients to pay per storage, the bandwidth they use, or the usage of CPU cycles.

Private Cloud
Also known as corporate or internal cloud, private cloud refers to computing services provided over a private network. These services are consumed by internal company departments rather than the general public.

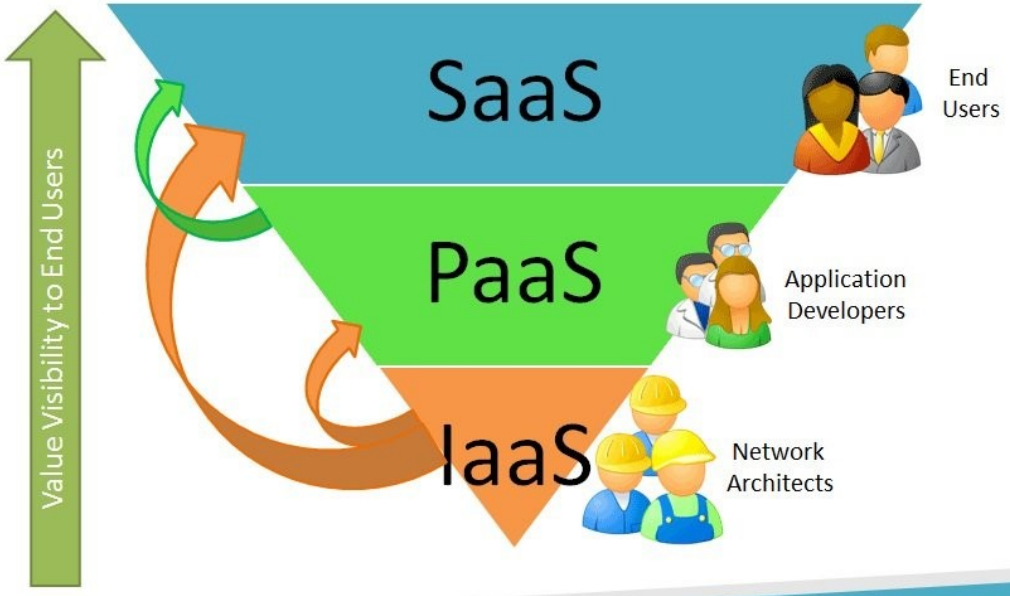
Private cloud provides businesses with a vast array of benefits, including scalability, elasticity and affordability. Perhaps one of the most important benefits of private cloud is that it provides high security and privacy levels through internal hosting and firewalls to ensure that hackers and snoopers do not access sensitive business data.

Hybrid Cloud

Hybrid cloud refers to an environment that incorporates private cloud and public cloud services. This kind of ecosystem allows workloads to move between public and private clouds, thus giving businesses more data deployment options and exceptional flexibility. Think of it as getting the best of both worlds.

Cloud - The technology of distributed data processing in which some scalable information resources and capacities are provided as a service to multiple external customers through Internet technology.

Cloud computing basics concept includes all of the following concepts:



SaaS (Software as a service) – Is the business model of software license, which

Cloud – The technology of distributed data processing in which some scalable information resources and capacities are provided as a service to multiple external customers through Internet technology.

Cloud computing basics concept includes all of the following concepts:

IaaS (Infrastructure as a Service) – A computer infrastructure, typically presented in the form of virtualization. Is a service within the concept of cloud hosting.

PaaS (Platform as a Service) – An integrated platform for the development, deployment, testing, and support of web-applications. Presented as a service on the basis of the concept of cloud hosting.

SaaS (Software as a service) – This is the business model of the software license, which involves the development and support of the software vendor. Customers also have the opportunity of paid use of it, usually through the Internet.

DaaS (Desktop as a Service) – Another business model license the software, which is a slightly improved model of SaaS, mostly involving the use of multiple services at the same time necessary to complete the work. Was first introduced in the early 2000s.

In addition to the above within the concept of cloud hosting technology, there are also common notions of Data as a service and Everything as a service respectively. Both concepts show that, through the World Wide Web using Cloud Hosting, can meet any requirements in the processing of information. This is the main advantage of **cloud computing** hosting in IT-based business solutions.

Cloud Computing Model

Cloud computing is offered in three different service models which each satisfy a unique set of business requirements. These three models are known as Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS).

SaaS

Software as a Service offers applications that are accessed over the web and are not managed by your company, but by the software provider. This relieves your organization from the constant pressure of software maintenance, infrastructure management, network security, data availability, and all the other operational issues involved with keeping applications up and running. SaaS billing is typically based on factors such as number of users, usage time, amount of data stored, and number of transactions processed. This service model has the largest market share in cloud computing; according to Gartner, its sales will reach 117 billion USD by the year 2021**(2)**. Current applications for SaaS include Field Service solutions, system monitoring solutions, schedulers and more.

PaaS

Platform as a Service is halfway between Infrastructure as a Service (IaaS) and Software as a Service (SaaS). It offers access to a cloud-based environment in which users can build and deliver applications without the need of installing and working with IDEs (Integrated Development Environments, which are often very expensive. Additionally, users can often customize the features they want included with their subscription. According to Gartner, PaaS has the smallest market share of the three service models, with a projected revenue of 27 billion USD by the year 2021[2]. In today's market, PaaS providers offer applications such as Microsoft Azure (also IaaS), Google App Engine, and Apache Stratos.

IaaS

Infrastructure as a service offers a standardized way of acquiring computing capabilities on demand and over the web. Such resources include storage facilities, networks, processing power, and virtual private servers. These are charged under a "pay as you go" model where you are billed by factors such as how much storage

you use or the amount of processing power you consume over a certain timespan. In this service model, customers do not need to manage infrastructure, it is up to the provider to guarantee the contracted amount of resources and availability. According to Gartner, this service model is forecasted to grow by 35.9% in 2018[2]. IaaS services offered today, include Google Cloud Platform and Amazon EC2.
Cloud computing has been around for quite some time now; however, it will continue to evolve as faster and more reliable networks offer increased benefits to service providers and consumers alike. With these advancements, there are growing opportunities to develop business models in an increasingly-connected economy.

Cloud computing in market

Source

https://www.globenewswire.com/news-release/2021/08/11/2278451/0/en/Cloud-Computing-Market-to-Hit-USD-791-48-Billion-by-2028-Rising-Demand-for-Improved-Virtual-Access-to-Information-among-Industries-to-Foster-Steady-Growth-Fortune-Business-Insights.html

Pune, India, Aug. 11, 2021 (GLOBE NEWSWIRE) -- The global cloud computing market size is projected to reach USD 791.48 billion by 2028, exhibiting a CAGR of 17.9% during the forecast period. Rising preference for omni-cloud systems will prove highly beneficial for the growth of this market, states Fortune Business Insights™ in its report, titled "**Cloud Computing Market Size, Share & COVID-19 Impact Analysis, By Type (Public Cloud, Private Cloud, Hybrid Cloud), By Service (Infrastructure as a Service (IaaS), Platform as a Service (PaaS), Software as a Service (SaaS)), By Industry (Banking, Financial Services, and Insurance (BFSI), IT and Telecommunications, Government, Consumer Goods, and Retail, Healthcare, Manufacturing, Others), and Regional Forecast, 2021-2028"**. According to this market research report, the value of the market stood at USD 219.00 billion in 2020.

List of Key Players Profiled in the Cloud Computing Market:

- Amazon Web Services (AWS) (Washington, United States)
- Oracle Corporation (California, United States)
- IBM Corporation (New York, United States)
- Alibaba Group Holding Limited (Hangzhou, China)
- Microsoft Corporation (New Mexico, United States)
- VMware, Inc. (California, United States)
- Alphabet Inc. (Google LLC) (California, United States)
- Rackspace Technology, Inc. (Texas, United States)
- SAP SE (Walldorf, Germany)
- Apple Inc. (California, United States)
- Salesforce.com, Inc. (California, United States)
- HCL Technologies Limited (Noida, India)
- Hewlett-Packard Company (HPE) (California, United States)
- Sprint Corporation (Kansas, United States)
- Verizon Wireless (New York, United States)
- Red Hat, Inc. (North Carolina, United States)
- Ooma Inc. (California, United States)
- Paytm (Uttar Pradesh, India)
- Adobe, Inc. (California, United States)

REPORT SCOPE & SEGMENTATION:

Report Coverage	**Details**
Forecast Period	2021 to 2028
Forecast Period 2021 to 2028 CAGR	17.9%
2028 Value Projection	USD 791.48 Billion
Base Year	2020
Market Size in 2020	USD 219.00 Billion
Historical Data for	2017 to 2019
No. of Pages	140
Tables, Charts & Figures	55
Segments covered	Type; Service; Industry; and Region
Growth Drivers	Integration of Big Data, AI, and ML with Cloud Will Provide Impetus to Market Proliferating Cloud-based Solutions amid COVID-19 Pandemic to Aid Growth Substantial Adoption of Omni-cloud over Multi-cloud to Boost Cloud Computing Market Growth
Pitfalls & Challenges	Data Privacy and Information Security Concerns Associated with Cloud Solutions to Impede Growth

Omni-cloud computing is a cloud solution that allows multiple cloud services to smoothly integrate and streamline their data on a single platform. The omni-cloud system is being increasingly preferred over the multi-cloud system owing to its multiple advantages and leading the cloud computing market trends. For example, an omni-cloud tool makes it possible to access real-time information from any location. In a departmental store, for instance, whenever there is an inventory shortfall, the cloud will send notification to the authorities, who will then take the necessary action. Similarly, storage of data on a unified platform also enables efficient analysis, enhances productivity, and elevates the quality of services. These, along with a few other benefits, are widening the applicability of omni-cloud computing across a variety of industries.

COVID-19 Impact:

Acceptance of Cloud-based Solutions to Favor Market During COVID-19 Pandemic

The acceptance of work from culture across public and private sectors has resulted in high demand for SaaS-based collaboration solutions. This has created opportunities for cloud companies to expand their customer base by introducing innovative solutions. For instance, in May 2020, Microsoft Corporation introduced "Hospital Emergency Response", a cloud-based solution for the healthcare sector. The solution is enabled by Azure platform. The Centres for Disease Control and Prevention (CDC) and other healthcare institutions are permitted to control the solution to develop COVID-19 calculation tools and reduce the work-stress on front-line workforces.