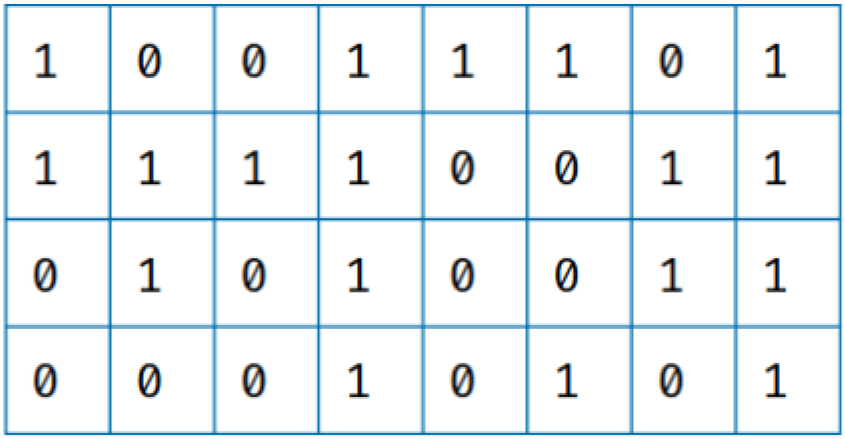
Fundamentals of computer science

Cheat Sheet

Understanding Binary

Binary Representation

A binary number is a number expressed in Ones (1) and Zeros (0).



1	0	0	1	1	1	0	1
1	1	1	1	0	0	1	1
0	1	0	1	0	0	1	1
0	0	0	1	0	1	0	1

How computer see information

Computers see information in binary.

Number System

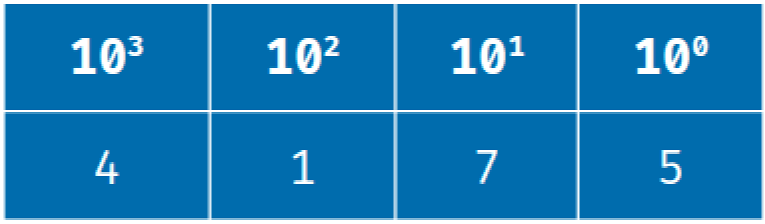
A writing system for expressing numbers.

Decimal Notation

Decimal is a notation in which we have learned most of the mathematics. The unique digits used to represent numbers in decimal notation are 0, 1, 2, 3, 4, 5, 6, 7, 8, 9.

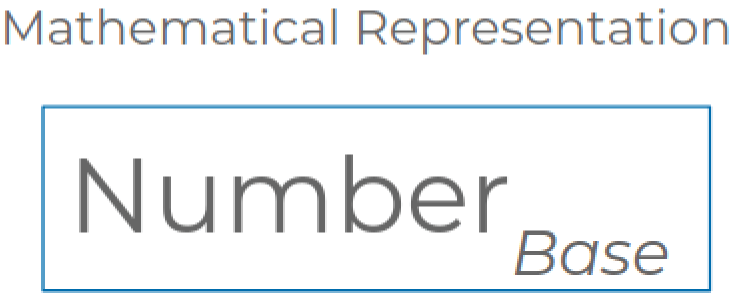
The decimal notation for 4175 is

$4 \text{ x } 10^3 + 1 \text{ x } 10^2 + 7 \text{ x } 10^1 + 5 \text{ x } 10^0 = 4175$



10^3	10^2	10^1	10^0
4	1	7	5

The decimal notation has a base of 10 which means the number of unique symbols or digits used to represent numbers in that notation is 10.



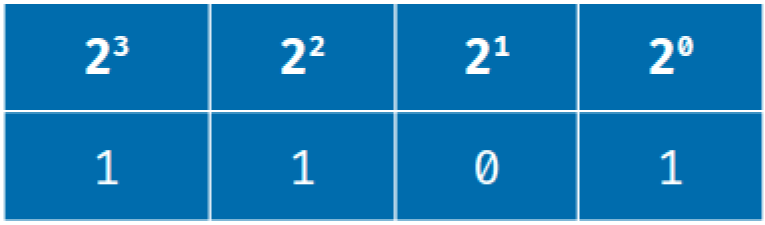
Mathematical Representation

$\text{Number}_{\text{Base}}$

Binary Notation

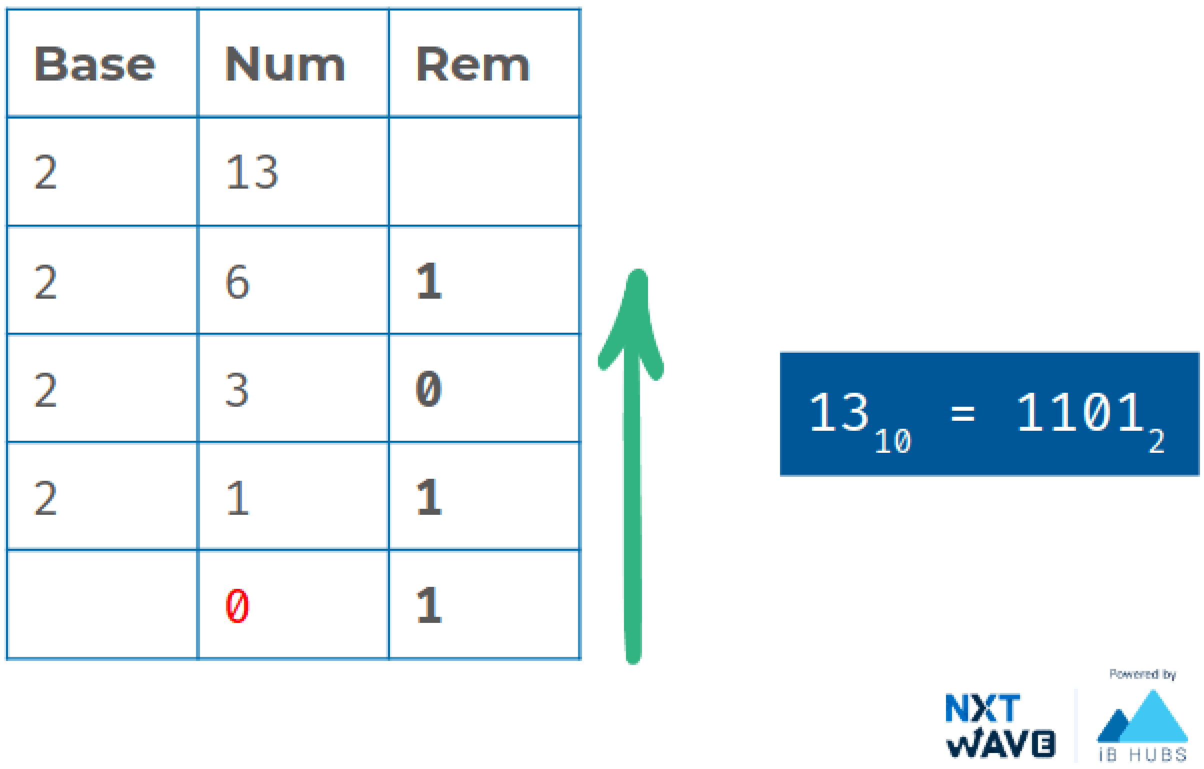
The only notation that computers understand is Binary Notation. The unique digits used to represent numbers in that notation are 0, 1. It has base 2.

$1101_2 = 1 \times 2^3 + 1 \times 2^2 + 0 \times 2^1 + 1 \times 2^0 \qquad = 8 + 4 + 1 \qquad = 13_{10}$



2^3	2^2	2^1	2^0
1	1	0	1

Decimal to Binary Conversion



Base	Num	Rem
2	13	
2	6	1
2	3	0
2	1	1
	0	1

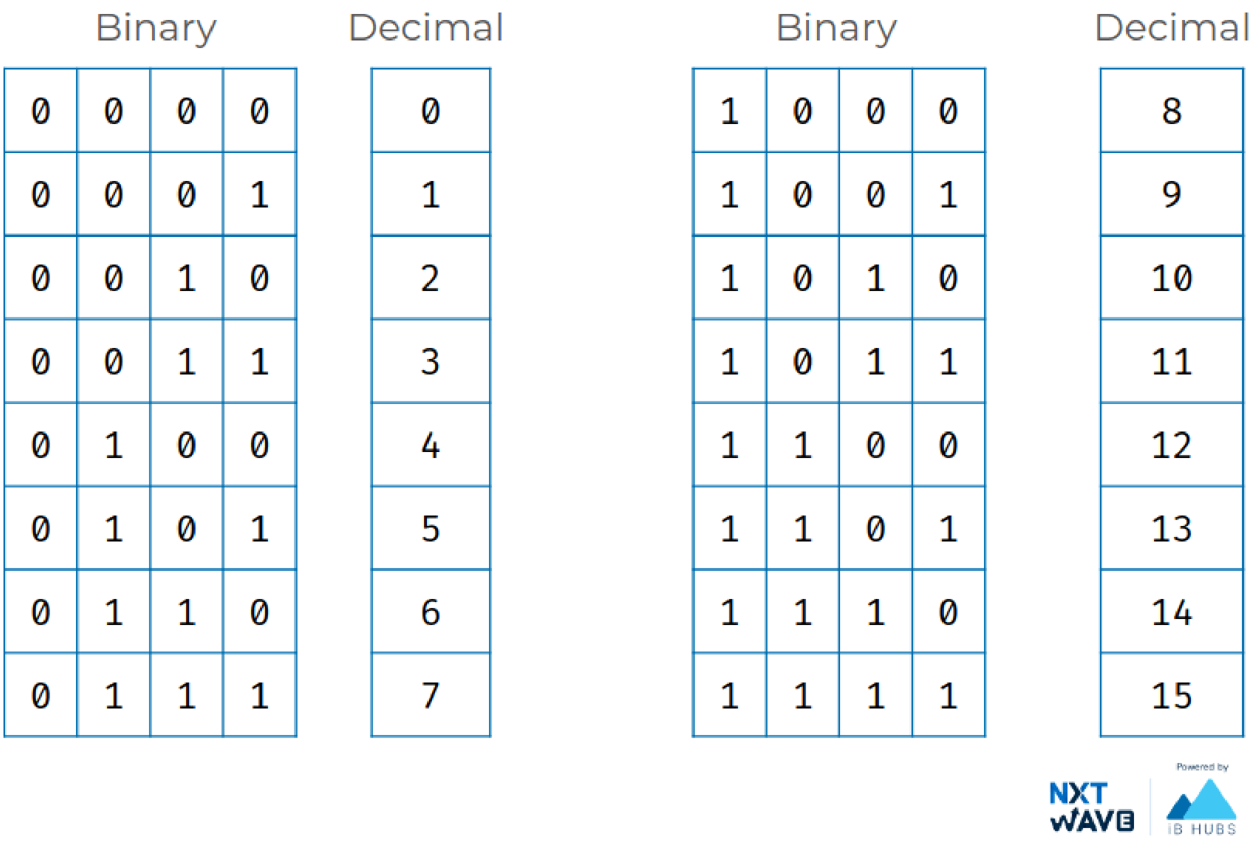
$13_{10} = 1101_{2}$

Binary Representation for Different Files

All types of content (text, images, videos, and everything) must be converted to binary so that it can be stored by the computer. Different types of content occupy different amounts of space in the computer because they are represented by different lengths of Binary numbers.

Storing Integers in 4 Bits

16 different types of information can be represented in 4 - bit. The numbers can be represented from 0 to 15.



Binary				Decimal	Binary				Decimal
0	0	0	0	0	1	0	0	0	8
0	0	0	1	1	1	0	0	1	9
0	0	1	0	2	1	0	1	0	10
0	0	1	1	3	1	0	1	1	11
0	1	0	0	4	1	1	0	0	12
0	1	0	1	5	1	1	0	1	13
0	1	1	0	6	1	1	1	0	14
0	1	1	1	7	1	1	1	1	15

NXT WAVE
Powered by
iB HUBS

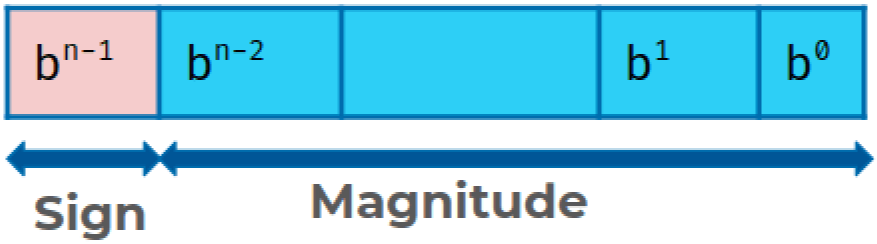
n-Bit Binary Number

- 2^n distinct combinations are possible.
- Can represent numbers from 0 to 2^n- 1.

if n = 8, the numbers can be represented from 0 to 255 (2^8 - 1).

Negative Numbers Representation

- Different representations and conventions are developed to store different types of information.
- One bit is represented for the sign, and the remaining bits represent numbers. If the bit contains zero, it represents a positive value, and if it has one, it represents a negative number.



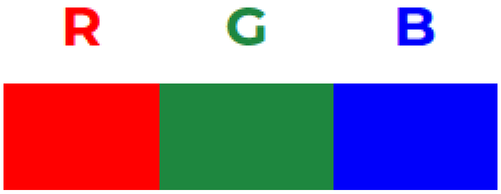
Representing Images

Image is made up of pixels. Pixel is the smallest unit in the picture. Each pixel has a specific colour.

Representing Colors

The colours are represented using different models. The basic one used in a lot of computers is the RGB or red, green, and blue model.

Based on the intensity of RGB, colours are represented as a number. Now, this number is represented as binary in computers.

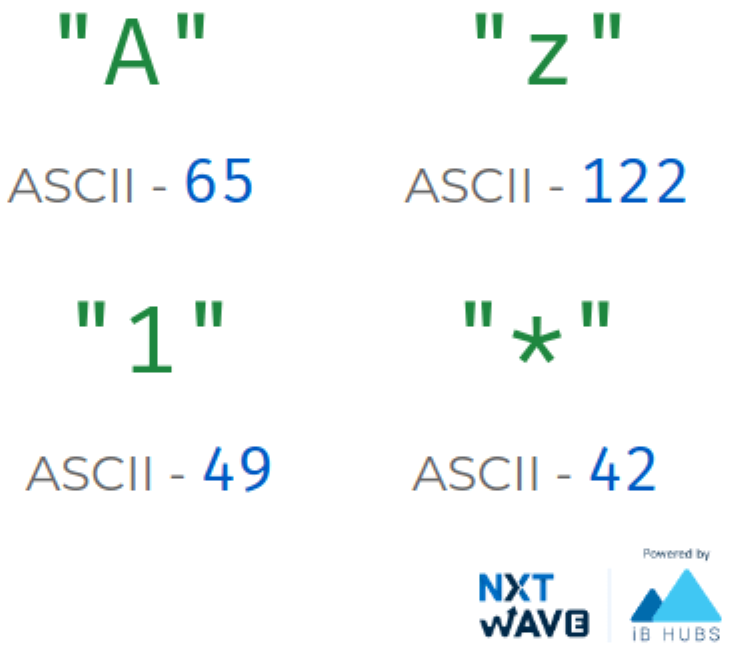


Representing Text

Each character is encoded as an integer which is represented by a binary value.

ASCII

American Standard Code for Information Interchange(ASCII) represents the English alphabet, digits, and punctuation marks and requires only one byte to store a character. It means it represents 256 different things. There is a limit on no. of characters that can be represented.



"A" ASCII - 65

"z" ASCII - 122

"1" ASCII - 49

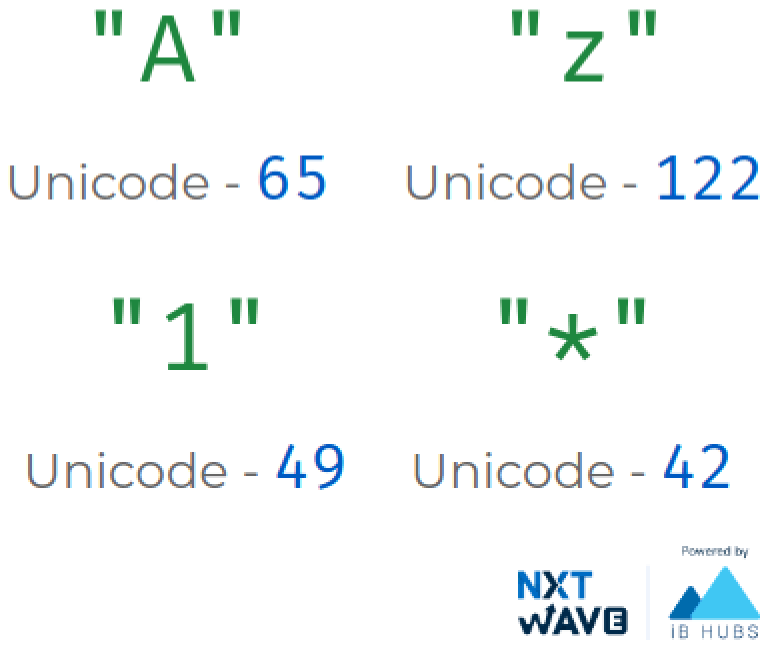
"*" ASCII - 42

UNICODE

Unicode allows us to represent characters using more than 1 byte.

UTF - 8

UTF-8 (Unicode Transformation Format - 8) is capable of storing 1,112,064 different characters and it has the same values for ASCII characters.

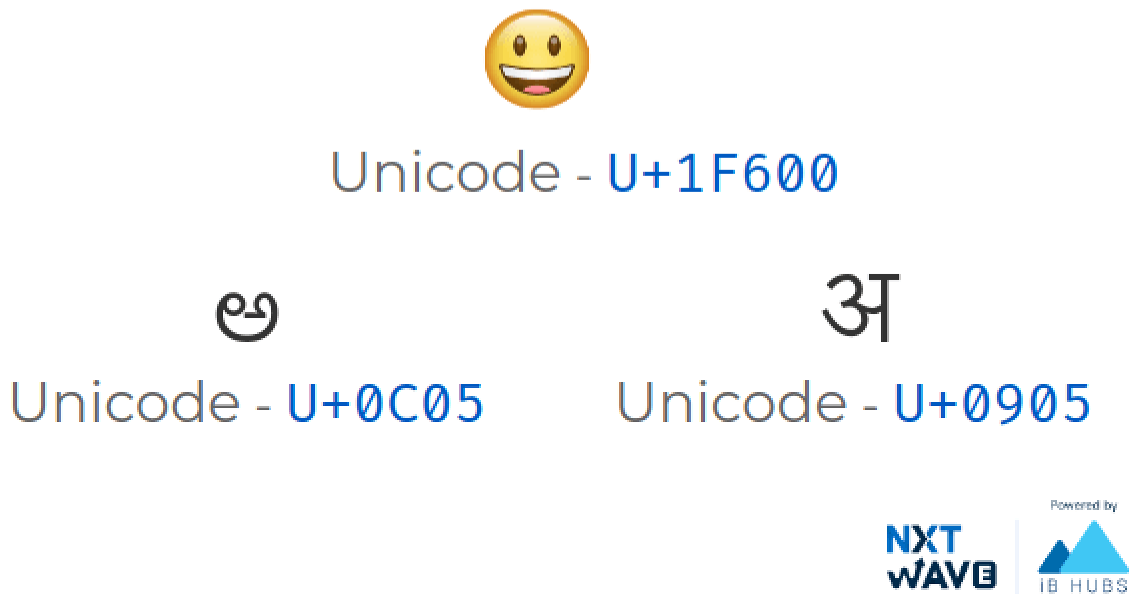


"A" Unicode - 65

"z" Unicode - 122

"1" Unicode - 49

"*" Unicode - 42



How Computer Gets Zero's and One's

Imagine we have a light bulb and a switch that turns the state of the light on or off. If we turn the light on, we can denote that the state is one. If the light bulb is off, we can represent the state is zero. In a similar way, the computer gets zero's and one's.

1
0
NXT WAVE
Powered by
iB HUBS

Computer Hardware

Computer hardware is the physical components that a computer system requires to function. It can be categorized as having either internal or external components.

Internal Components

Component parts that are found inside the computer's CPU case are

- Motherboard
- Processor & RAM
- Graphics Card (GPU)
- Fans
- Hard Disk etc.,

External Components

External hardware components or peripherals are usually input/output devices that are

- Keyboard & Mouse
- Monitors & Speakers
- USB Drives
- WebCam

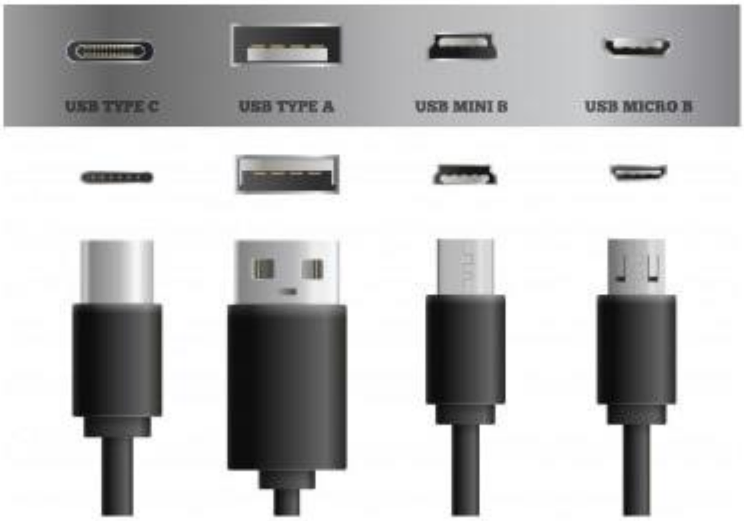
Ports

Connectors or Ports allow us to plug in different components to a computer to enhance its capabilities.

- The video display ports are DVI, VGA, and so on...
- The video and audio ports are HDMI, display port, and so on...

USB Port

Instead of having ports for different things, we can use USB (Universal Serial Bus) Port. USB is Designed to standardize many peripheral connections. In addition to audio and video, it allows data transfer and power. Recent mostly using USB is USB Type C which allows charging, data transfer, extend audio, video projection, etc.



Motherboard

The motherboard is the computer's central communication's backbone connectivity point through which all components and external components connect.

CPU

The CPU (Central Processing Unit or processor) is responsible for processing all information from programs. It processes many instructions from different programs every second.

CPU Instruction Set

Programs on the computer are broken down into very small and simple instructions.

For example,

- Add two numbers
- Load value from Memory Location
- Store Value to Memory Location

CPU will only perform small operations in binary but lakhs of these operations are done in a single second.

CPU Clock Speed

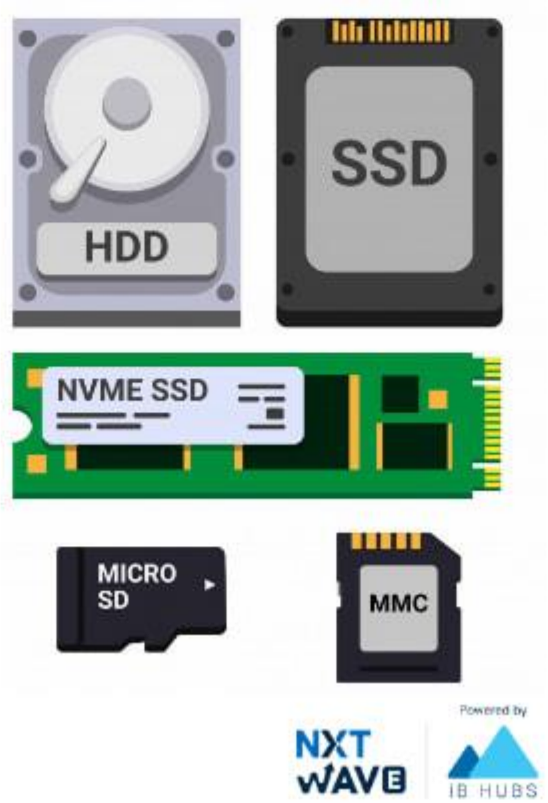
- One of the performance matrices is clock speed which measures the number of instructions that the CPU can execute per second.
- The CPU speed measured in gigahertz (GHz).
- 1 GHz = a Billion instructions per second
- A higher clock speed means a faster CPU.

CPU Cores

A CPU core is a CPU's processor. A core can work on one task. Modern-day devices have multiple cores. 8 core processor means 8 cores and similar to other cores. Each core of a CPU can perform operations separately from the others. They may work together to perform parallel operations.

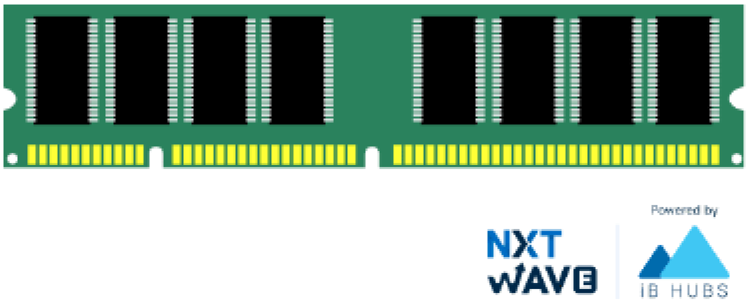
Hard Drive

- The hard drive is Non-volatile storage which is responsible for storing permanent and temporary data.
- SSD (Solid-state drives) is 25 times faster than regular HDDs (Hard-disk drives)



RAM

- RAM is also another type of storage. we can store photos, videos, etc. The RAM is measured in GB we might have heard a lot of times that 2GB RAM, 4GB RAM in Phones, Laptops, etc.
- RAM is volatile memory means the data remains in RAM if the computer is on, but it's lost when the computer is turned off.
- RAM is much faster to read and write than other types of storage.



RAM vs Hard Drive

RAM	HARD Drive
RAM is volatile memory	Hard drive is a nonvolatile memory
Relatively Expensive	Relatively Cheaper
Faster Data Access	Relatively Slower Data Access

Registers

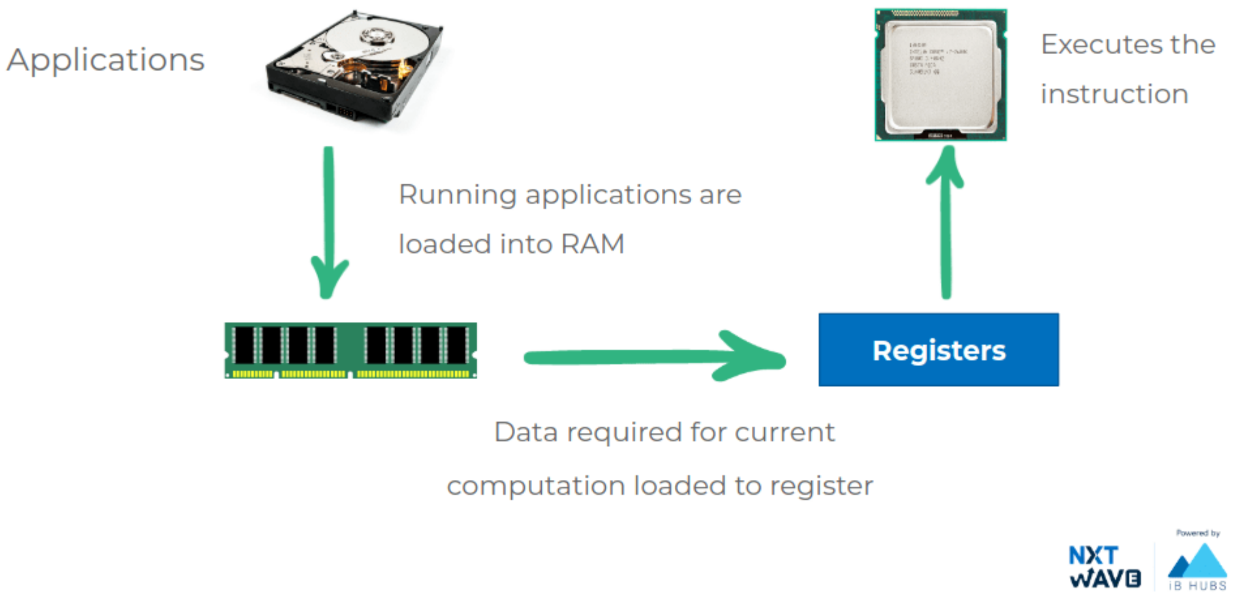
- Registers are another type of storage. Many times much faster than regular storage devices and even RAM which is placed very close to the CPU.
- They are extremely fast and extremely expensive.
- They are available in very little storage. Usually in MBs.
- They are used to quickly accept, store, and transfer data and instructions that are being used immediately by the CPU.

Data Access Times in RAM, Hard Disk and Registers

Storage Devices	Access Times	Data Transfer Speed
Registers	< 2 nano seconds	~ 100 GB/s
RAM	~ 100 nano seconds	~ 10 GB/s
Hard Disk	~ milli seconds	~ 1000 MB/s

Running Applications

All applications and programs are stored in a hard drive. As the hard drive is slow the running applications are loaded into RAM and as registers are faster than RAM the data required for current computation loaded to register are quickly accessible by the CPU. If the RAM is less, we can’t store much data while running multiple programs the system gets slow.

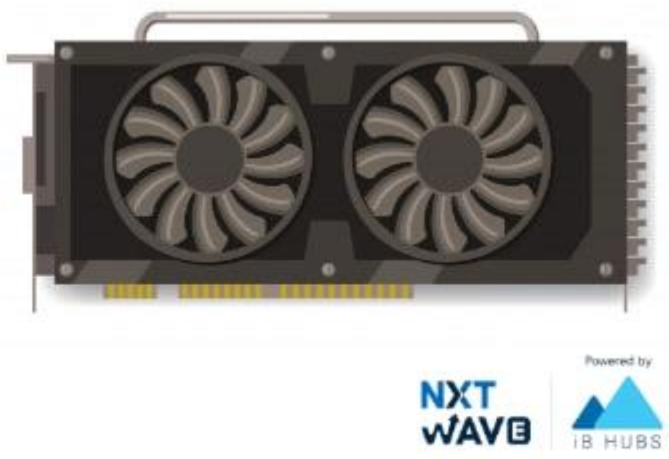

Applications
Running applications are loaded into RAM
Registers
Data required for current computation loaded to register
Executes the instruction
NXT WAVE
Powered by
IB HUBS

32-bit and 64-bit Systems

- A 32-bit processor includes a 32-bit register, which can store 2^{32} values.
    - Can support up to 4 GB RAM
- A 64-bit processor includes a 64-bit register, which can store 2^{64} values.

Graphics Card

Graphics Card is a Specialized processor designed to deal with specific types of instructions efficiently. It is used for 3D Rendering, Games, Deep Learning, and so on.

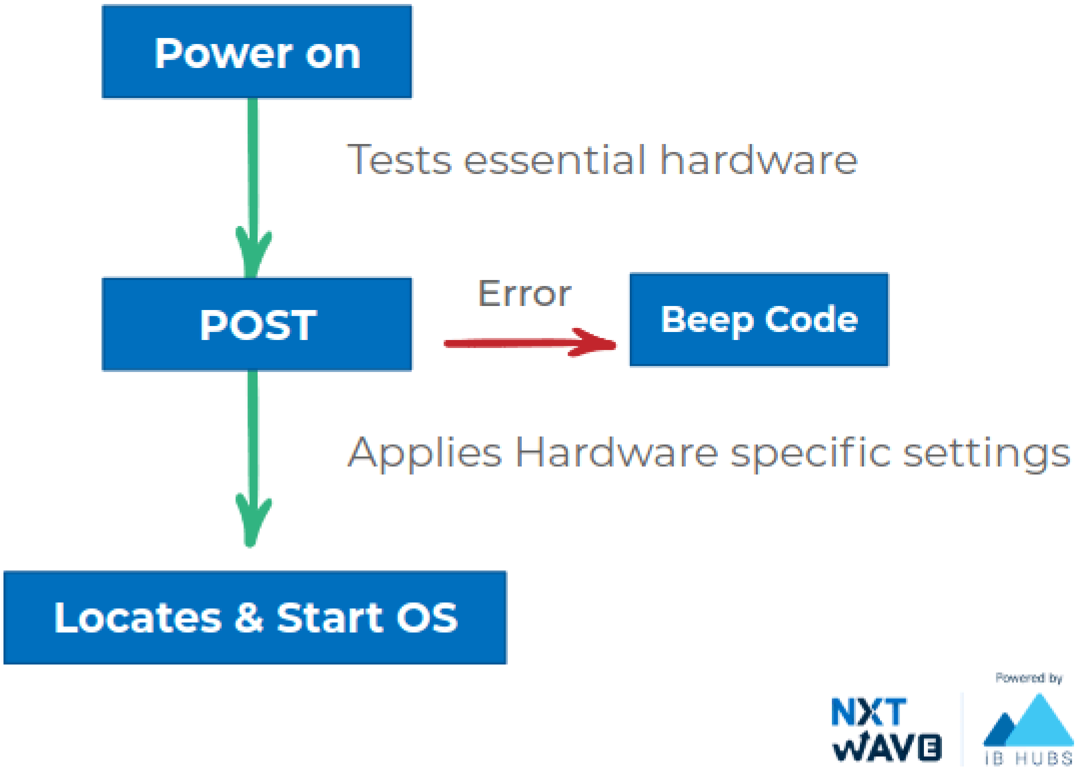


CPU vs GPU

CPU	GPU
A CPU can perform a variety of different instructions	GPU can perform a specialized set of instructions.
Has few cores (around 2 - 8)	Has 1000s of Cores. Good at parallelization

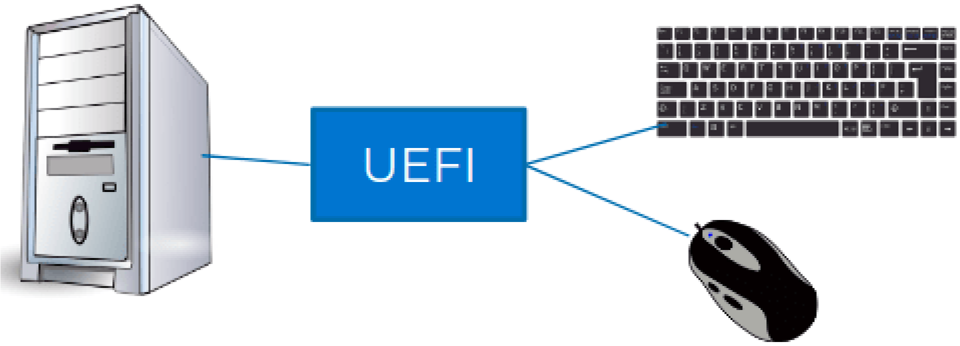
BIOS

- Basic Input Output Services (BIOS) initializes the hardware and gets our operating system up and running.
- After system power on the first BIOS tests, the hardware is working or not. you will hear a combination of beeps indicating what is the hardware status(Having any error or not). If you didn't hear beep sounds it means there are no built-in speakers in the system.
- If the test is a success then it handovers hardware to the operating system.



UEFI

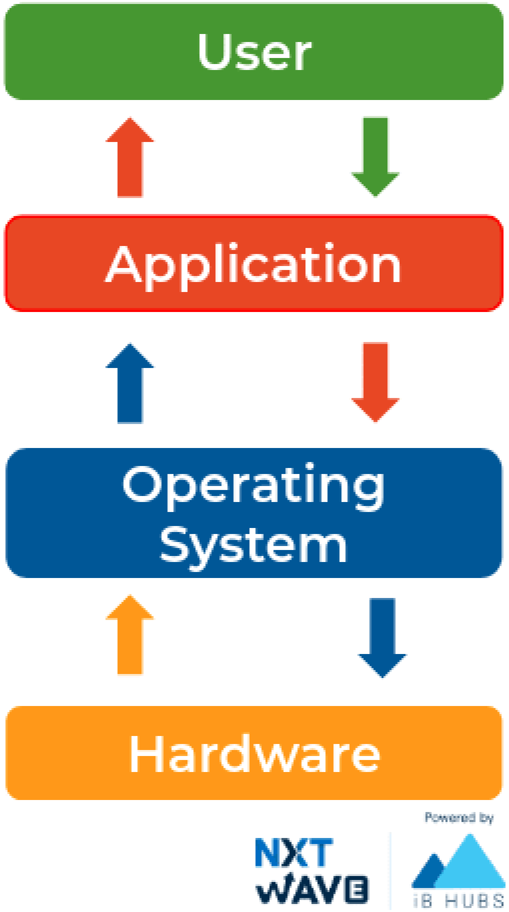
- UEFI (Unified Extensible Firmware Interface) is similar to BIOS, which supports the latest hardware
- It performs a secure boot
- It performs faster boot



Introduction to Operating System

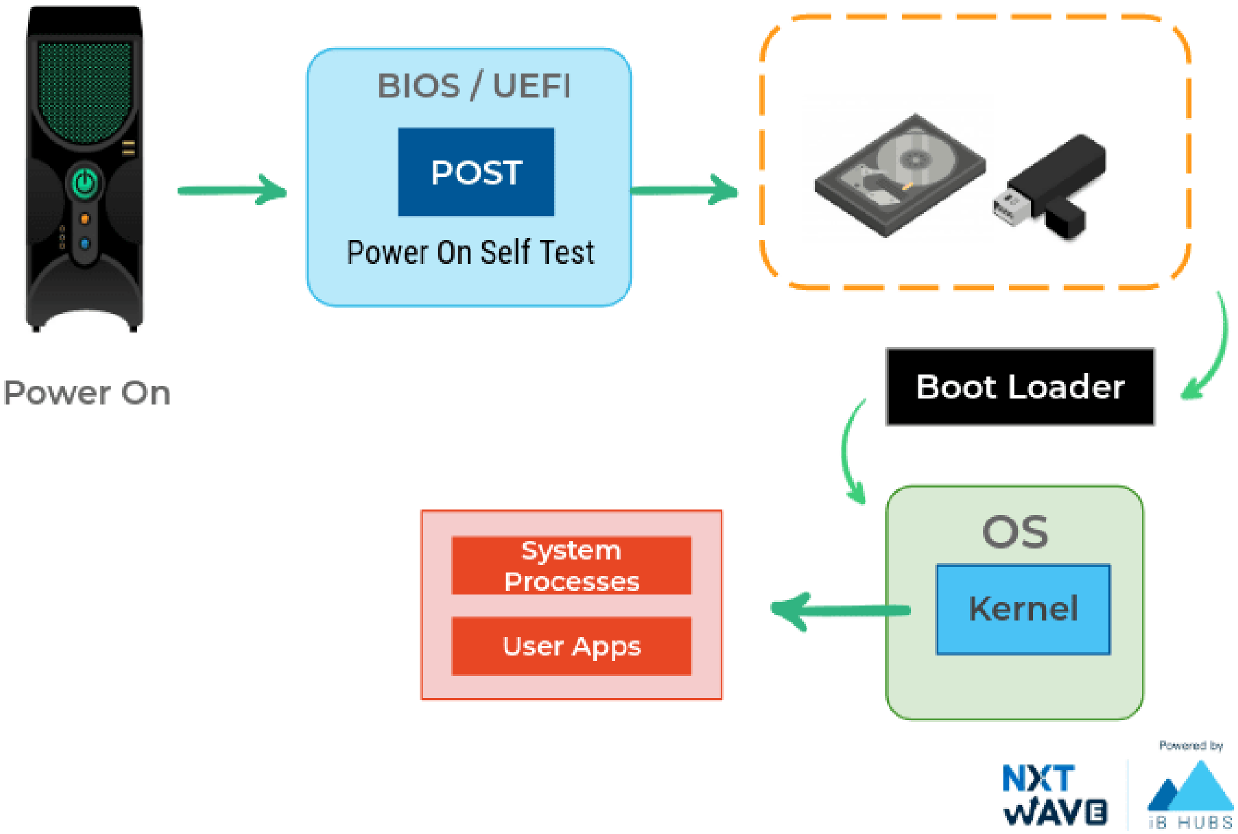
Operating System

Operating System is an interface between hardware and software.



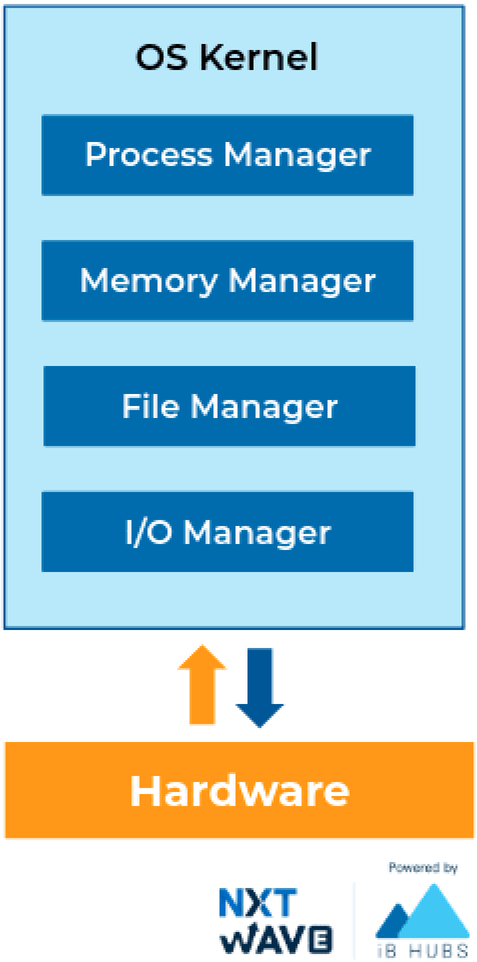
Boot Process

1. Power on
2. Computers run a test to make sure all the hardware is working correctly this is called a Power On Self Test(POST).
3. Based on the BIOS configuration, a boot device(HDD, CD, or USB) that contains OS is selected for a program known as a boot loader which then loads the operating system.
4. Once OS gets loaded the essential part of OS Kernel is started and appropriately set for us to use.



OS Kernel

- The Kernel is the core component of OS. As users, we don't interact with the kernel directly.
- It interacts directly with the hardware.
- It manages the hardware resources.



Linux Kernel

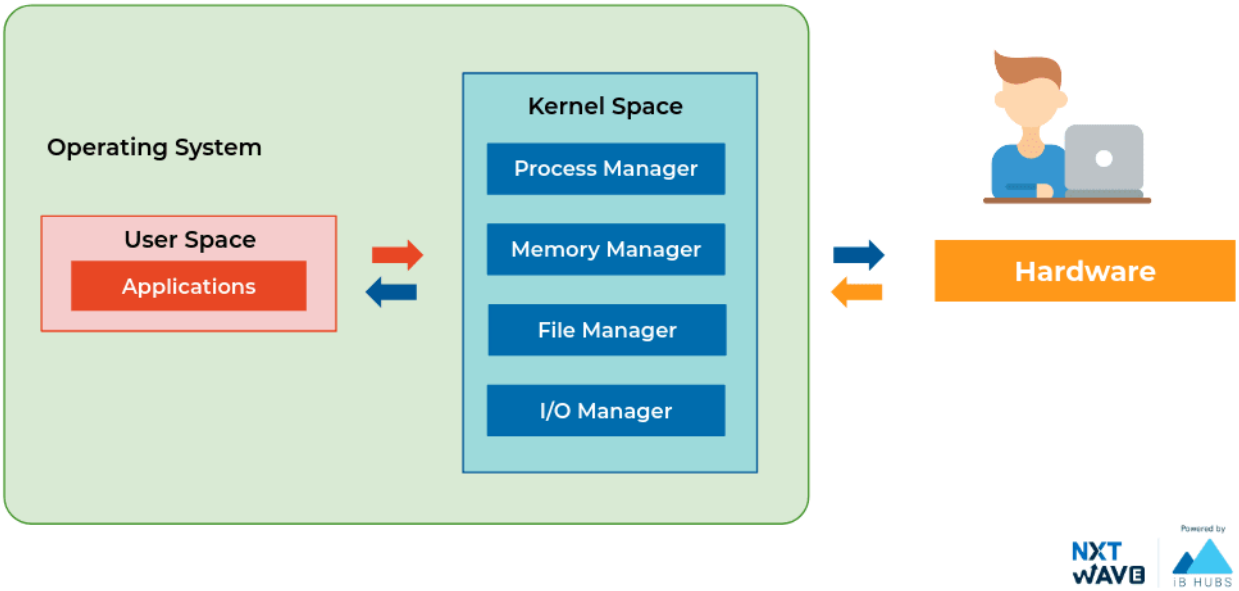
- Linux is a free and open-source OS Kernel developed by Linus Torvalds.
- Operating Systems made from software based on Linux Kernel are referred to as Linux Distributions (or Linux Distros).

Components of OS

Broadly we can categorize components of os are

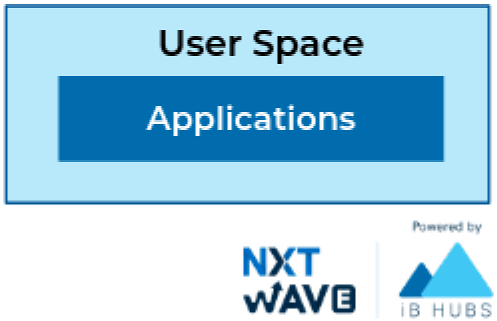
- Userspace
- Kernel

All apps interact with the kernel. The kernel will interact with hardware



Userspace

- Userspace is everything outside the kernel.
- Programs we interact with directly are part of Userspace which are Text Editors, Music Players, User Interfaces, etc.
- Userspace can include some OS programs like the GUI etc.



Responsibilities of OS

Abstraction

Abstraction means hiding the unnecessary complex inner details and showing only the relevant features. Once the operating system is up we no longer deal with the internals. For example, consider a car driver he does need to know how the engine works in the car. he/she simply needs to know few things like steering, gear, and so on.

- Operating System abstracts away the complex details of the underlying hardware.
- It provides a common API to applications and services.
- Thus, It simplifies development applications.

Resource Management

Resource management is one of the important tasks for OS. power, memory, CPU, and so on are resources. For example, Mobile/laptop use multiple applications at the same time like chrome, music app, chat app. These applications required CPU, ram, GPU resources.

- Operating System allocates hardware resources to applications in a fairway.
- It makes efficient use of the limited resources.
- It prevents improper use of computer resources.

Isolation and Protection

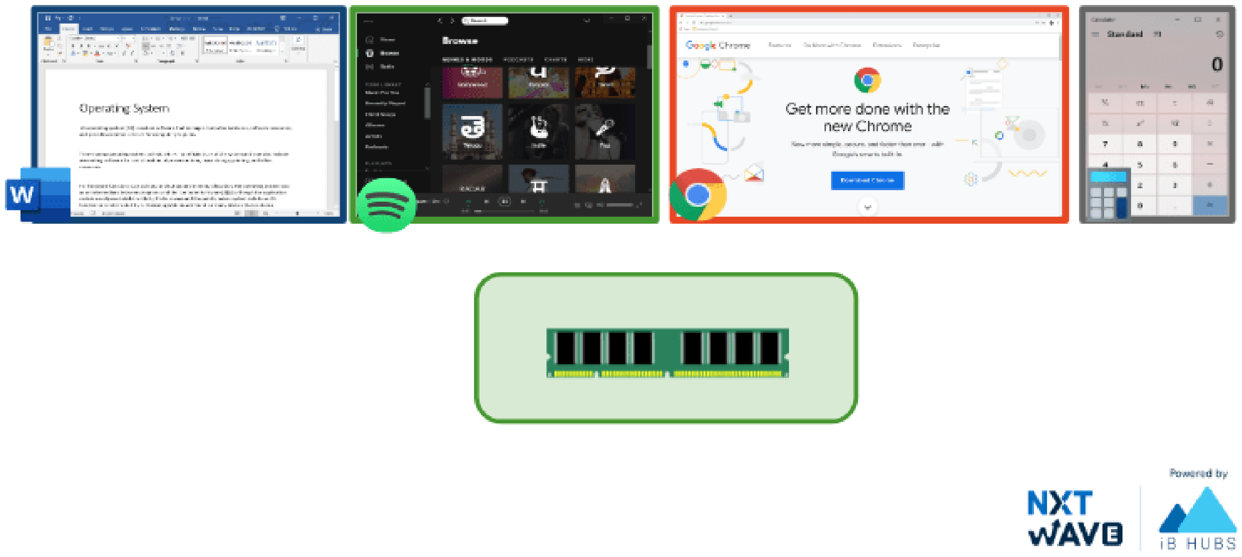
- Operating System Provides a safe way for multiple applications to simultaneously run without interfering with each other.
- It protects applications from each other. If one application fails Operating System ensures that it does not affect others.

Multi-User System

In Laptops, computers, multiple users can access at the same time. Resource management between the apps used by each user is the responsibility of the operating system.

Memory Management

We generally run multiple applications at a time like Whatsapp, Spotify. How much memory should be allotted to the applications in ram is taken care of by Operating System. CPU resources are also allocated by OS.

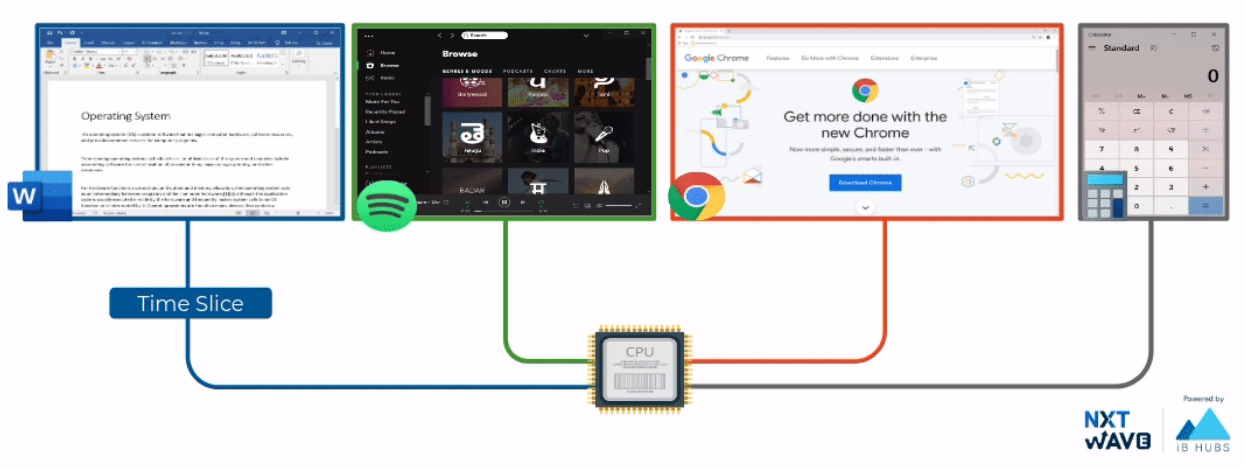


Process Management

Many times, we have more running applications than the number of CPU cores. even our Chrome Window has multiple tabs, and you can simultaneously play video in one tab, solve questions in another, Kernel continuously switches and gives different applications to use CPU. The switching is so fast that they all seem to run parallelly.

Similarly, all of the RAM is not being used by a single app. OS kernel allocates the required amount of RAM to each app as and when required.

Here, the time slicing concept is applied. A time slice is a very short interval of time that gets allocated to a process for CPU execution.



CPU runs billions of instructions per second. So in the same second, let's say, first few instructions for Word document next few instructions for Music, similarly Web browsing and Calculator. CPU will run billions of instructions per sec. We can't exactly see the difference this is called process management.

File and File Systems

File

A file is a collection of related information that is stored on the storage medium.

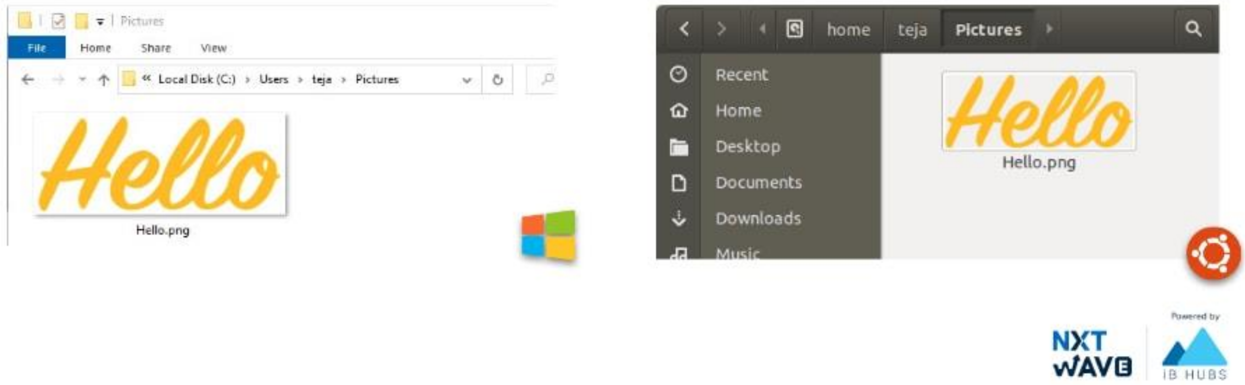
File Extensions

- Files may have an extension that is used to denote the type of file.
- Programs open and operate files based on the file extension.
- Image & Video Extensions: .jpeg, .png, .tiff, .mp4, .mov

- Docs: .html, .js, .py, .docx, .txt

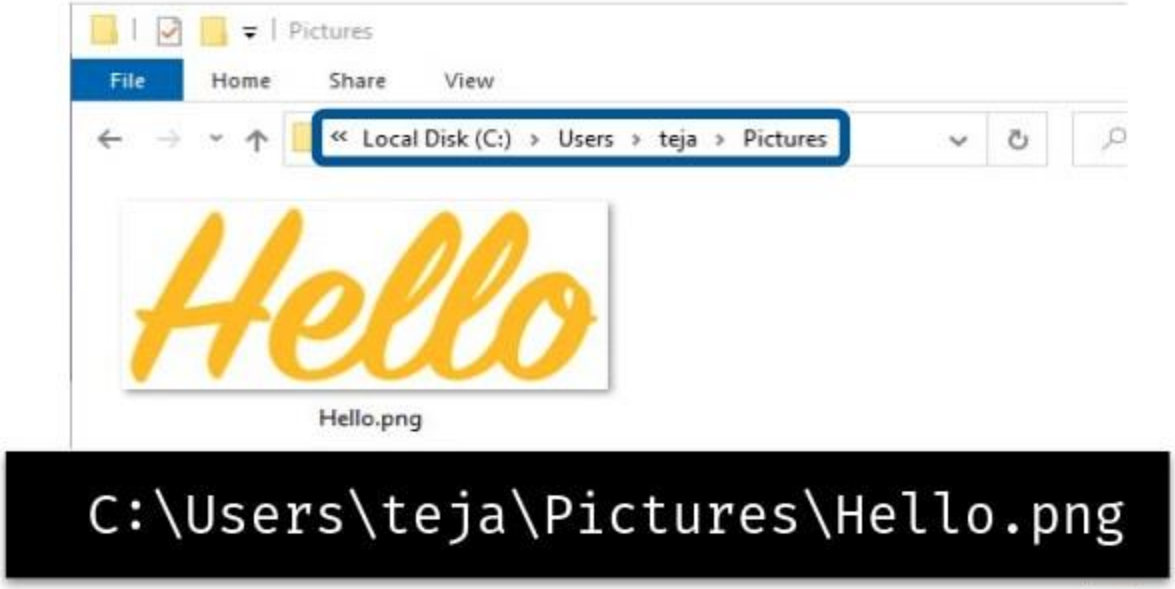
Files and Folders

Usually, we organize files in folders and these folders are stored in other folders and this continues until we arrive at C:\ drive or some other drive in Windows and Root directory (/)in Linux-based systems and macOS.

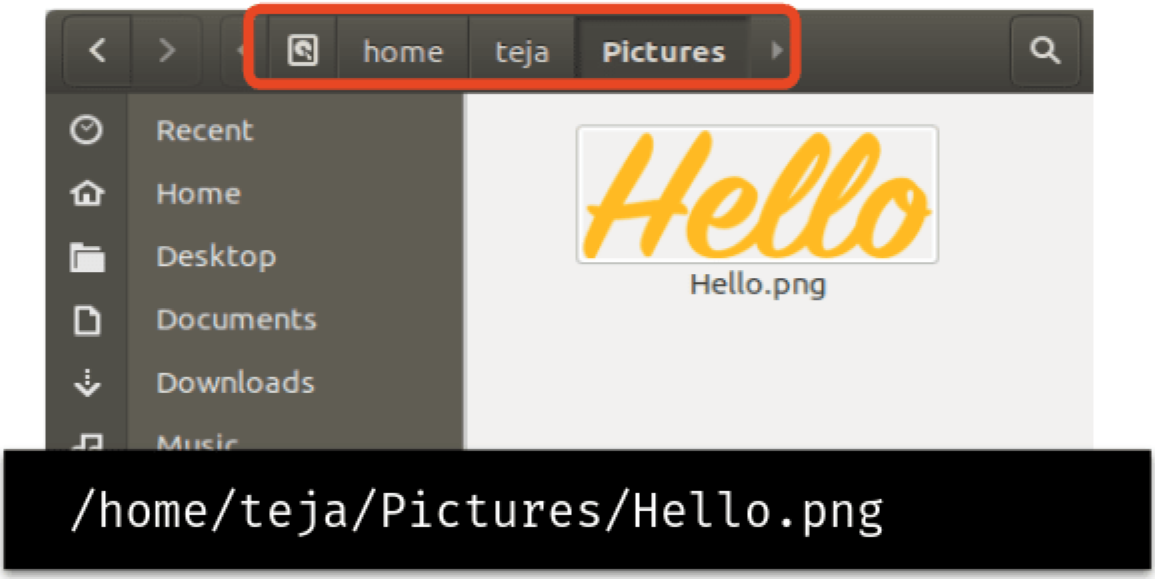


File Path

- The general form of the name of a file or directory specifies a unique location in a file system.
- File path usually is folder names separated by slashes.



```
C:\Users\teja\Pictures\Hello.png
```



```
/home/teja/Pictures/Hello.png
```

File and File Systems

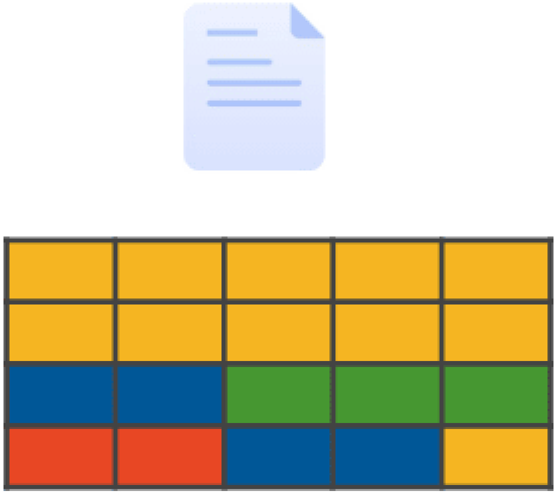
The kernel handles file storage on our machines. There are three main components to handling files on an OS.

- File System

- File Data
- File Metadata

File Data

- The data of a file is broken down into pieces of data blocks of a fixed size before storing in a disk.
- Default Block Size is 4 KB. So, the size occupied by a file on the disk will be in multiples of 4 KB.
- Block storage improves faster handling of data because the data isn't stored on one long piece and it can be accessed quicker. It's also better for utilizing storage space.



File Content

Though everything is stored in binary, files can be classified

- Text Files
- Non - Text Files

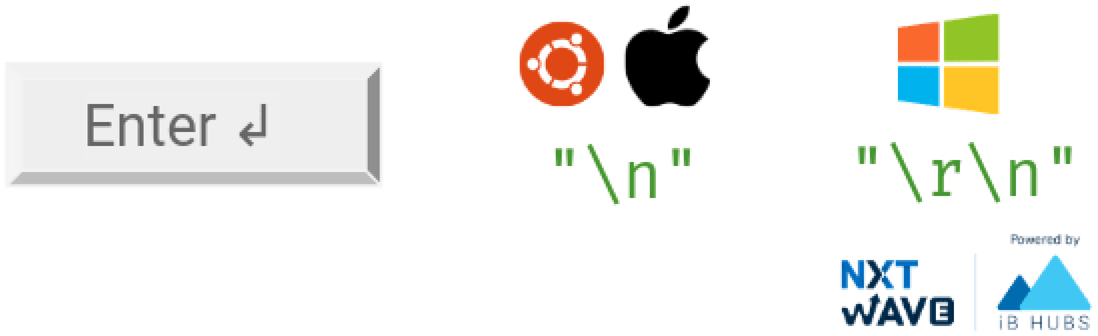
Text Files

Consists of plain, unformatted words, letters, and punctuation (Unicode Characters) intended to be readable by humans. The text files can be opened by any text editor.

- e.g. All Code files and plain text files

Line Endings

The line ending character when Enter ↲ is typed is different in different Operating Systems.



Non-Text Files

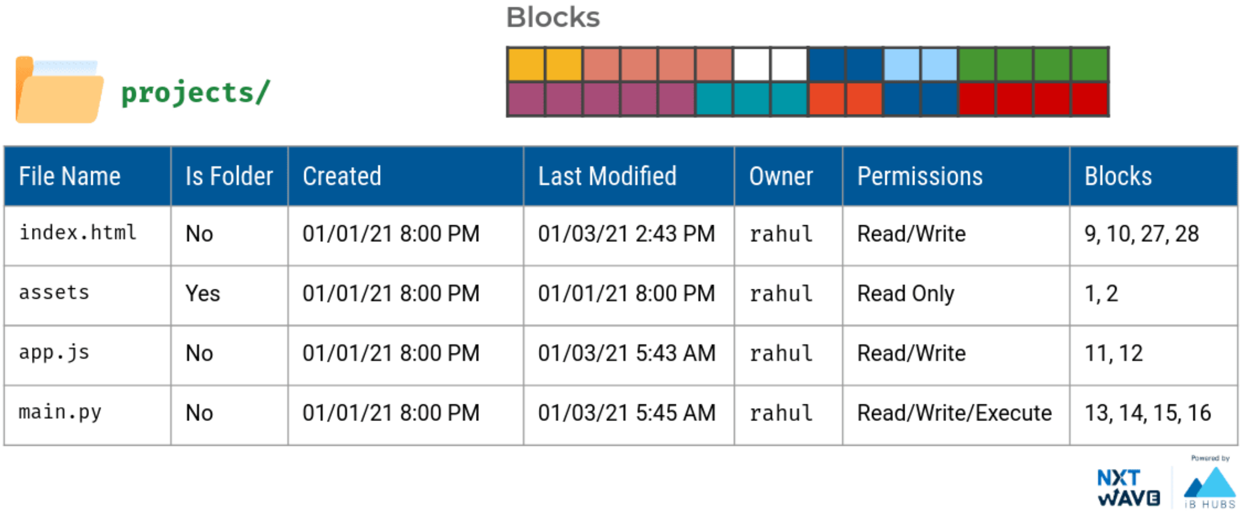
Consists of non-textual data meant primarily to be read by applications that translate those content into something useful by humans.

- Pictures, Audio, Video, Rich formatted text (Docs), etc.

File Metadata

The file metadata contains the information about our file.

- Owner of File
- Permissions (Read, Write, Execute)
- File Size
- Creation Date Time
- Last Modified Date Time


Blocks
projects/

File Name	Is Folder	Created	Last Modified	Owner	Permissions	Blocks
index.html	No	01/01/21 8:00 PM	01/03/21 2:43 PM	rahul	Read/Write	9, 10, 27, 28
assets	Yes	01/01/21 8:00 PM	01/01/21 8:00 PM	rahul	Read Only	1, 2
app.js	No	01/01/21 8:00 PM	01/03/21 5:43 AM	rahul	Read/Write	11, 12
main.py	No	01/01/21 8:00 PM	01/03/21 5:45 AM	rahul	Read/Write/Execute	13, 14, 15, 16

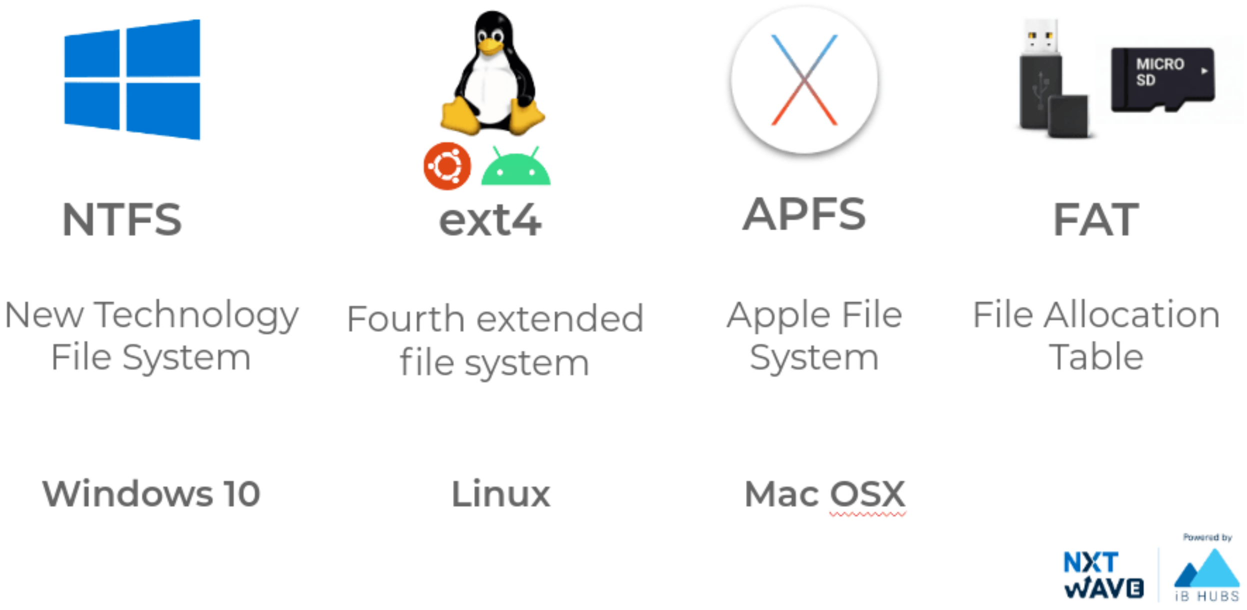
File Systems

- Responsible for managing files on the disk.

- Provides a mechanism to store the data and access the file contents including data and programs.
- Different file systems were developed for different use cases.

File System determine various aspects like

- Limitations on File Size and File Names
- Size occupied on the disk
- Encryption
- File Search
- Backup & Versions

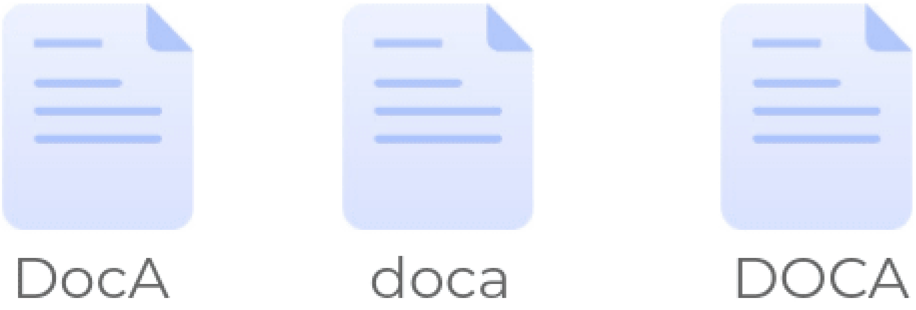


NTFS	ext4	APFS	FAT
New Technology File System	Fourth extended file system	Apple File System	File Allocation Table
Windows 10	Linux	Mac OSX	

Case Sensitivity

Each File System has its own rules, preferences, etc. Similarly, some File Systems are case sensitive. A case-sensitive file system distinguishes between uppercase and lowercase letters in file names and treats them as different files.

- Ext4 (Linux) is case-sensitive.
- NTFS (Windows), APFS (Mac), FAT are case-insensitive.

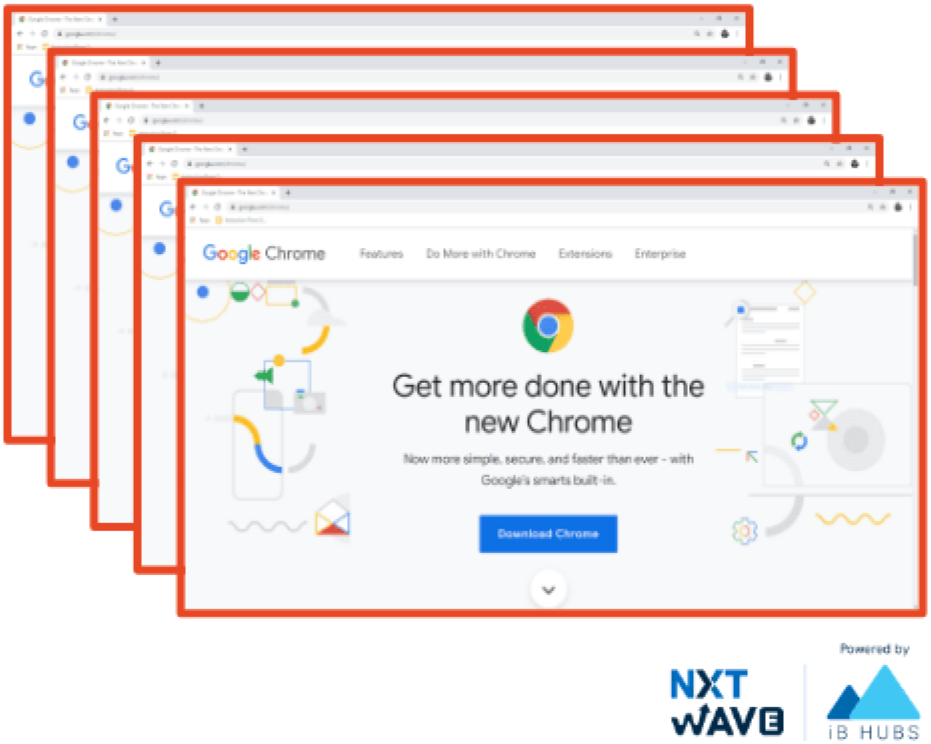


Overview of Process Management

Process

- The process is a program that's executing.
- When a program is loaded into the memory, it becomes a process.

- Each instance of the program is a different process.



Get more done with the new Chrome

Now more simple, secure, and faster than ever - with Google's smarts built-in.

Process Memory

Each process has process memory. Process Memory is divided as follows.

Stack:

It stores temporary data like function parameters, returns addresses, and local variables. The stack grows and shrink.

Heap:

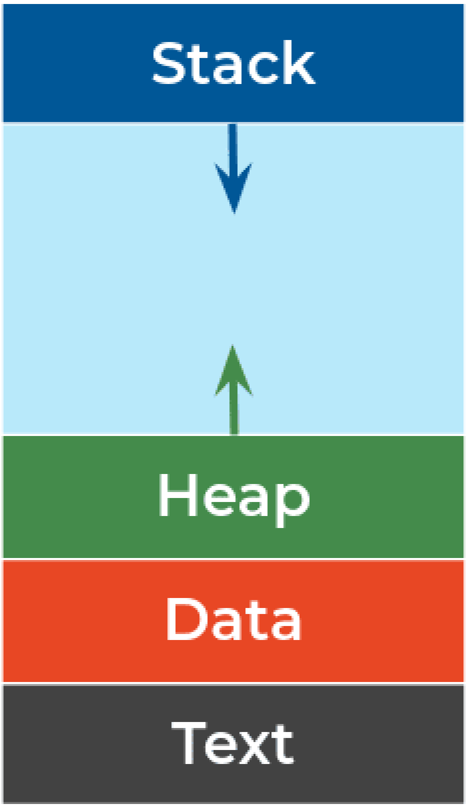
It allocates memory, which may be processed during its run time.

Data:

It contains constants.

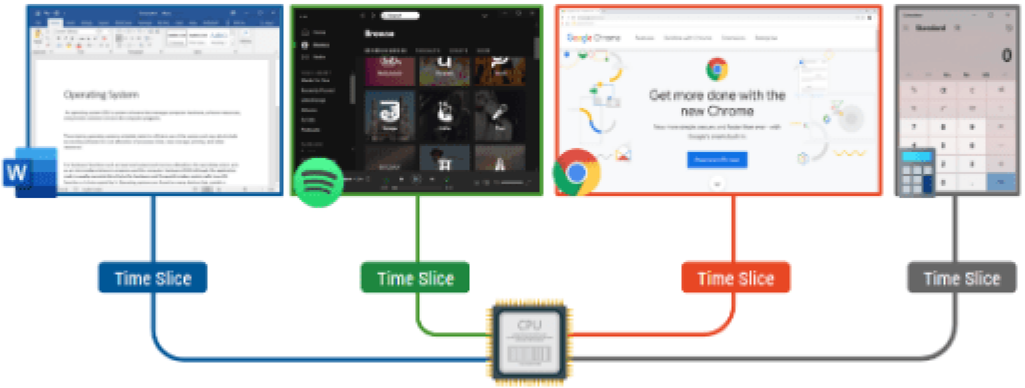
Text:

It includes the current activity, Contains Program Code.



Context Switching

Kernel switching from one process to another process is known as Context Switching.



Reasons for Context Switching

- A process with high priority arrives for execution.
- Running process requires I/O resources in the system and many more ...

Process State

When a process executes, it passes through different states. These stages may differ in different operating systems.

New:

The initial state when a process is first started/created.

Ready:

The process is waiting to be assigned to a processor.

Running:

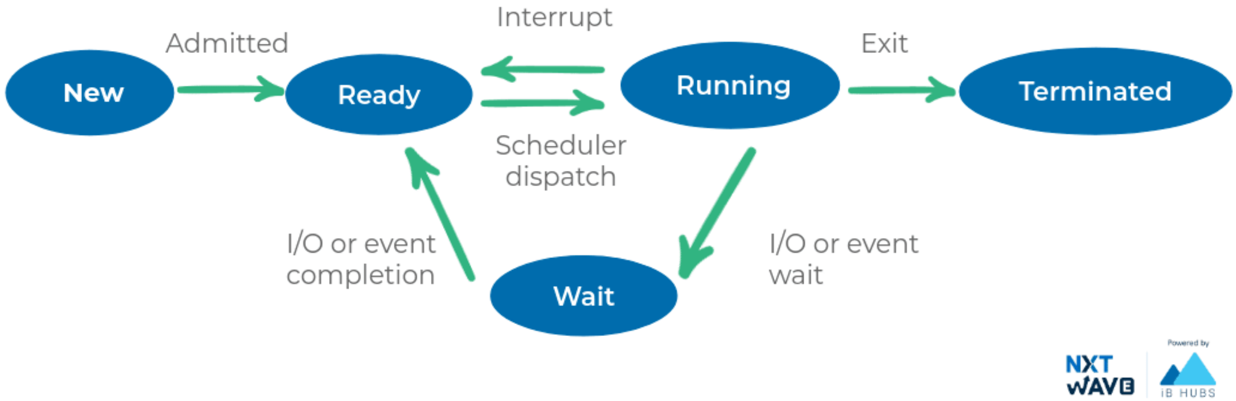
The process is chosen by the CPU for execution and the instructions within the process are executed.

Waiting:

Process wait for a resource, such as waiting for user input, or waiting for a file to become available.

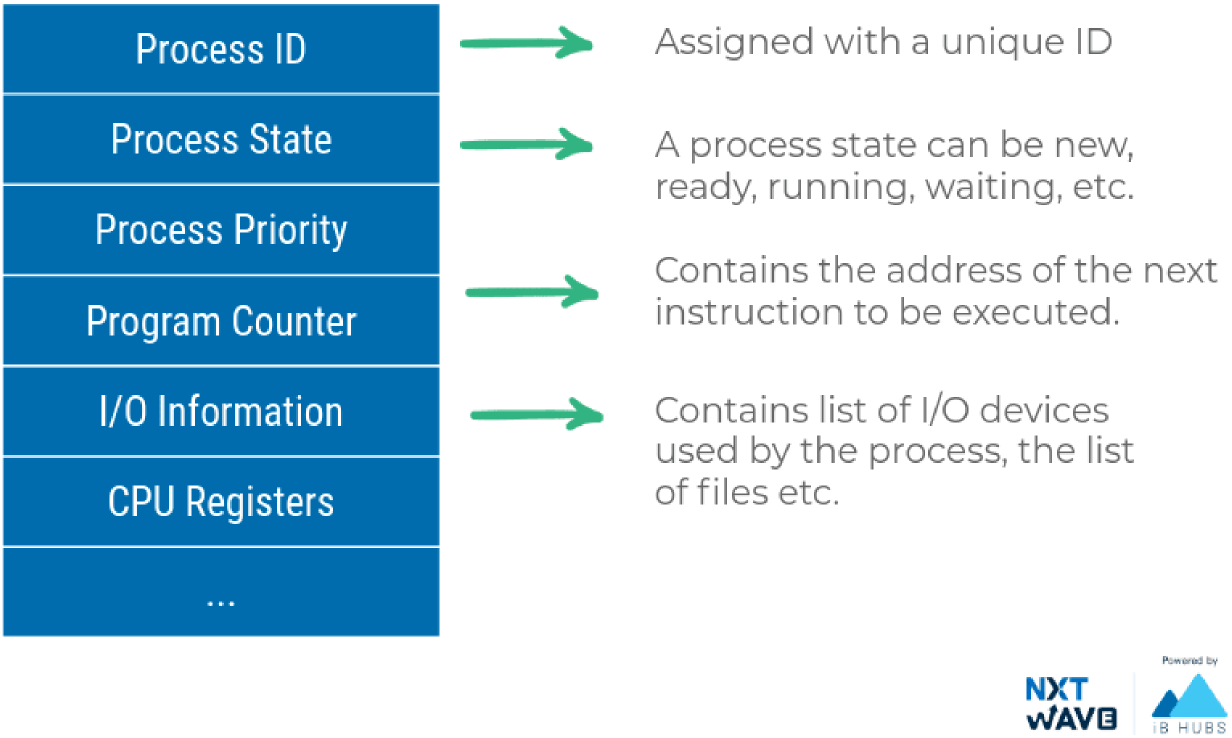
Terminated:

When a process finishes its execution, it comes in the termination state.



Process Control Block

A Process Control Block is a data structure maintained by the Operating System for every process. Context of a process represented in the PCB. The PCB is identified by an integer process identifier (PID).



Process ID → Assigned with a unique ID

Process State → A process state can be new, ready, running, waiting, etc.

Process Priority

Program Counter → Contains the address of the next instruction to be executed.

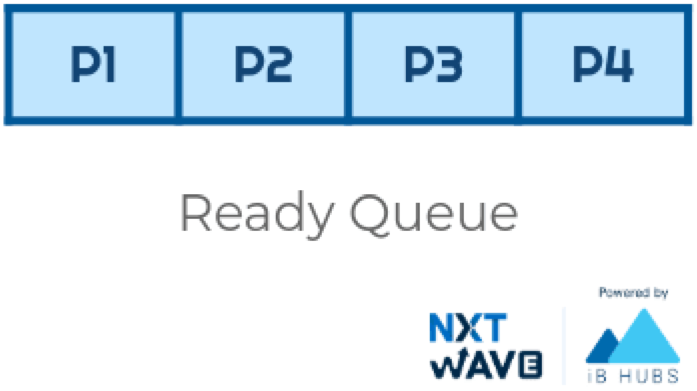
I/O Information → Contains list of I/O devices used by the process, the list of files etc.

CPU Registers

...

CPU Scheduling

- Process of determining which process will own CPU for execution while another process is on hold.
- The aim of CPU scheduling is to make the system efficient, fast, and fair.
- Whenever the CPU becomes idle, the OS must select one of the processes in the ready queue to be executed.



Ready Queue

CPU Scheduling Algorithms

The selection process is carried out by the CPU scheduler using different algorithms they are

- First Come First Serve (FCFS)

- Shortest-Job-First (SJF) Scheduling
- Priority Scheduling
- Round Robin Scheduling and many more ...

CPU Scheduling Terminology

Arrival Time:

Time at which the process arrives in the ready queue.

Completion Time:

Time at which process completes its execution.

Burst Time:

The time required by a process for CPU execution.

Turnaround Time:

The total amount of time spent by a process in the system.

Turnaround Time = Completion Time – Arrival Time

Waiting Time:

Amount of time spent by a process in the ready queue.

Waiting Time = Turnaround Time – Burst Time

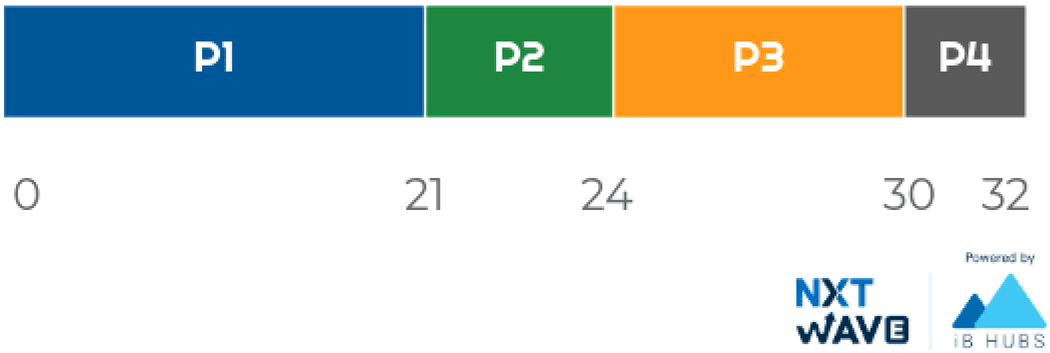
First Come First Serve

- A process that requests the CPU first gets the CPU allocated first.
- A process that arrived in the ready queue earlier will be executed earlier.

Example:

Process	Arrival Time	Burst Time
P1	0	21
P2	1	3
P3	2	6
P4	3	2

The process p1, p2, p3, and p4 are schedule with First Come First Serve are as follows.



So, the Completion time, Turnaround time, Waiting time of the process(p1, p2, p3 & p4) are as follows.

Process	Arrival Time	Burst Time	Completion Time	Turn Around Time	Waiting Time
P1	0	21	21	21	0
P2	1	3	24	23	20
P3	2	6	30	28	22
P4	3	2	32	29	27

Average Waiting Time = (0 + 20 + 22 + 27)/4 = 17.25

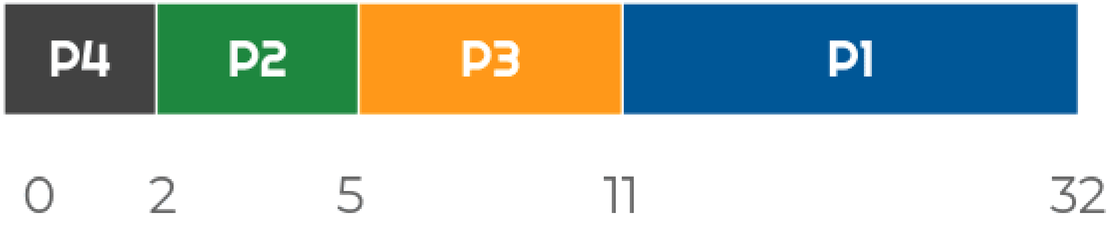
Average Turnaround Time = (21 + 23 + 28 + 29)/4 = 25.25

Shortest Job First

Shortest Job First(SJF) scheduling works on the process with the shortest burst time or duration first.

Process	Arrival Time	Burst Time
P1	0	21
P2	1	3
P3	2	6
P4	0	2

The shortest job first scheduling for process p1, p2, p3, and p4 is as follows.

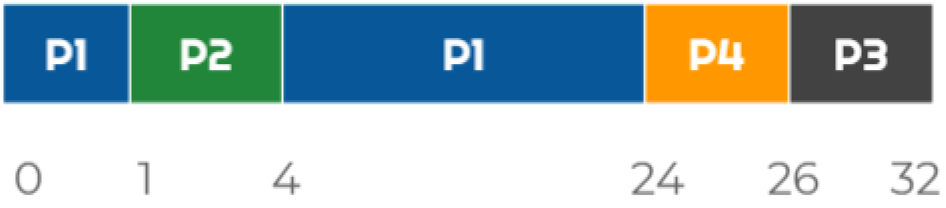


Priority Scheduling

Priority Scheduling Algorithm works on the process with the higher priority. Processes with the same priority are executed in an FCFS manner.

Process	Arrival Time	Burst Time	Priority
P1	0	21	2
P2	1	3	1
P3	2	6	4
P4	3	2	3

The processes(p1, p2, p3, and p4) scheduled based on Priority Scheduling are as follows.



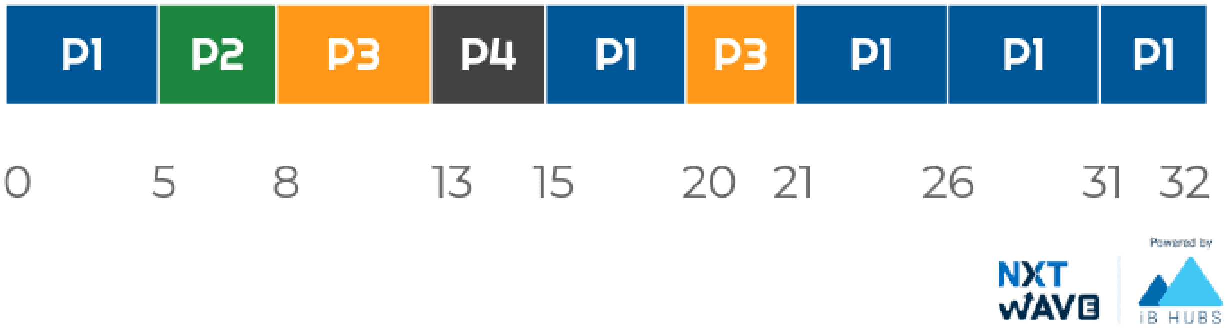
Round Robin Scheduling

In Round Robin Scheduling a fixed time is allotted to each process called Time Slice for execution. Once a process is executed for the given time period then another process executes for the given time period.

Example:

Process	Arrival Time	Burst Time
P1	0	21
P2	1	3
P3	2	6
P4	3	2

Suppose the Time Slice is 5 units and the process p1, p2, p3, and p4 are scheduled in Round Robin Scheduling are as follows.



Inter Process Communication

IPC is a mechanism, where the OS allows various processes to communicate with each other.

Communication can be done through:

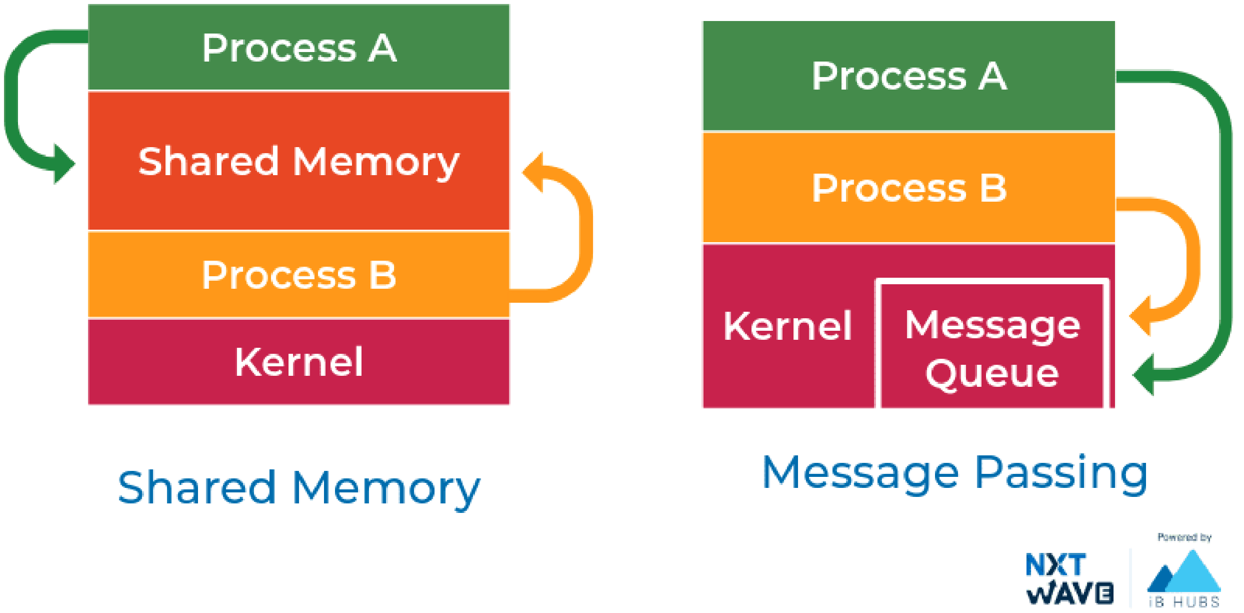
- Shared Memory
- Message passing

Shared Memory

- A particular region of memory is shared between the cooperating process.
- Exchanging information by reading and writing data to this shared region.

Message Passing

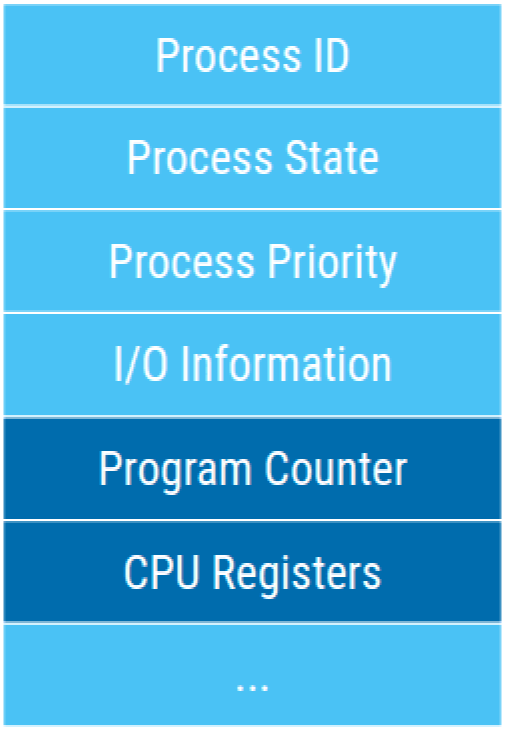
- Communication takes place by exchanging messages directly between the cooperating process.



Threads & Concurrency

Process Execution State

- Program Counter determines the next instruction to be executed.
- Results in a stream of instructions which are often referred to as a thread of execution.



Process ID

Process State

Process Priority

I/O Information

Program Counter

CPU Registers

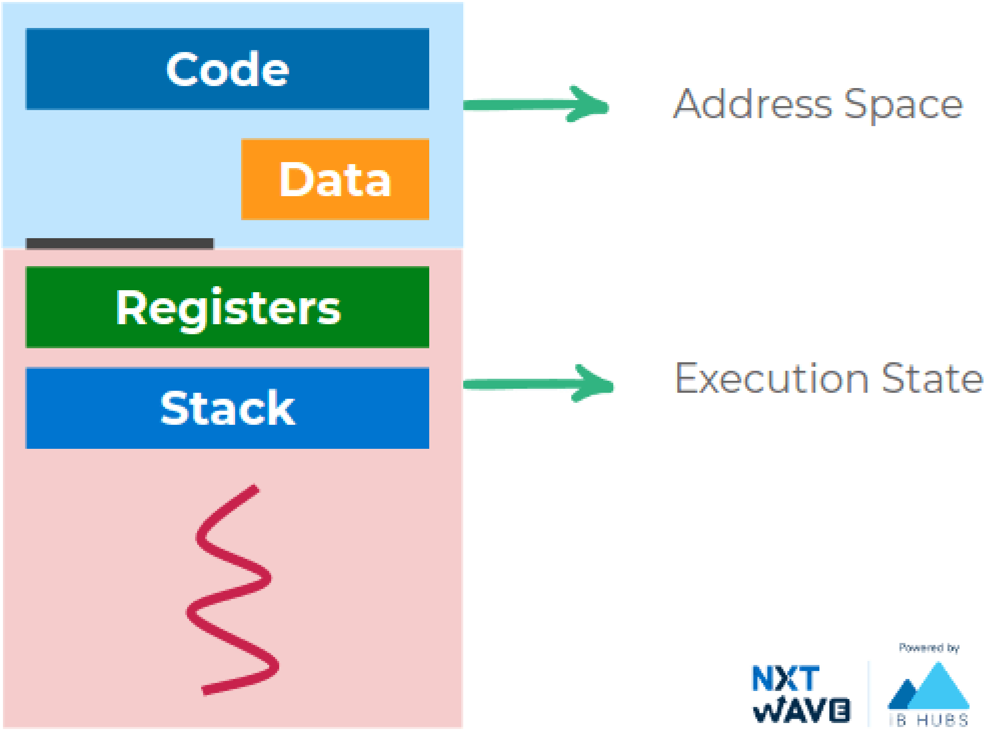
...

Threads

The sequential flow of control within a process is called as a thread.

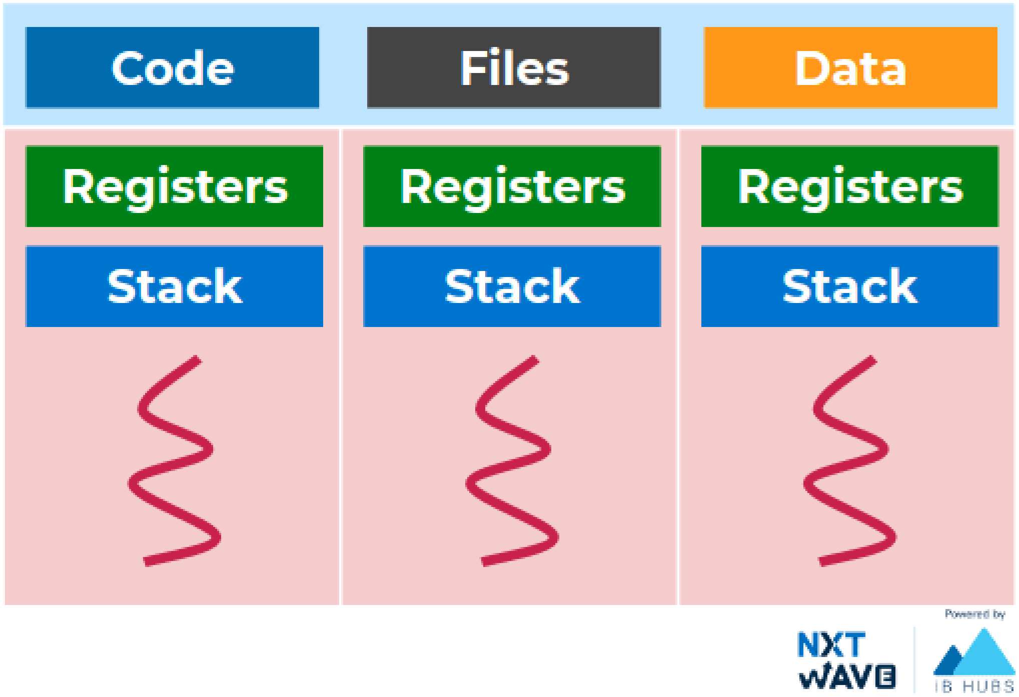
Each thread has

- A program counter
- A register set
- A stack space

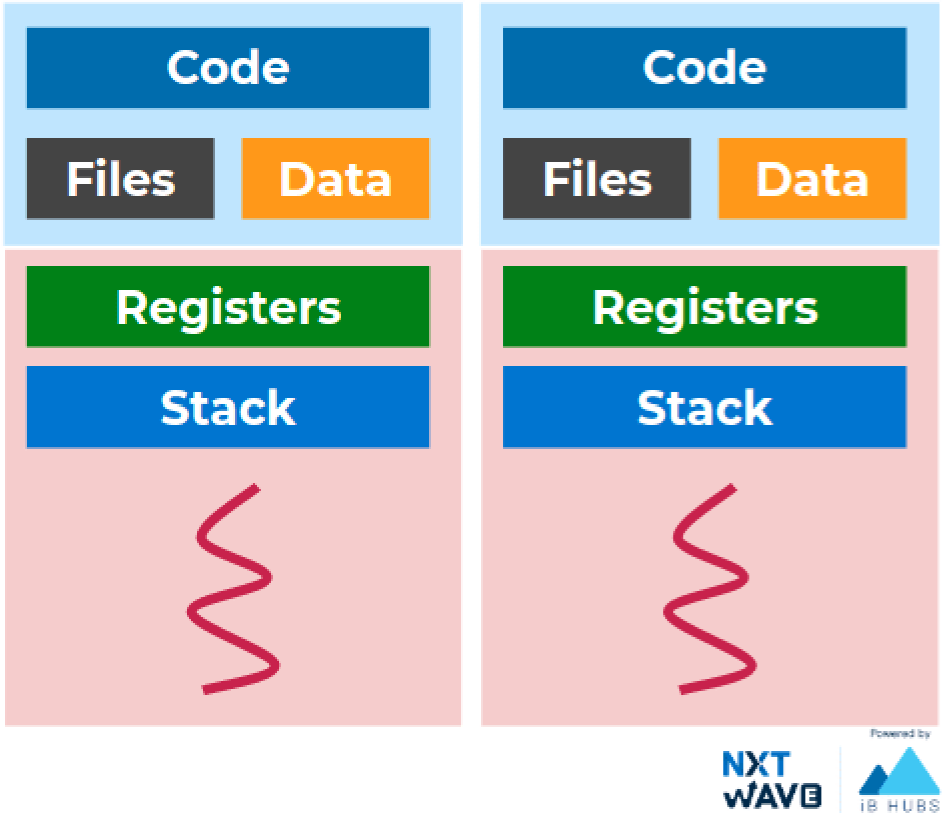


Multi-Threaded Process

- A process can have multiple threads.
- Each thread can run on a separate processor, thus utilizing multiple cores of the CPU



Single-Thread Process



Process vs Threads

Threads are lightweight.

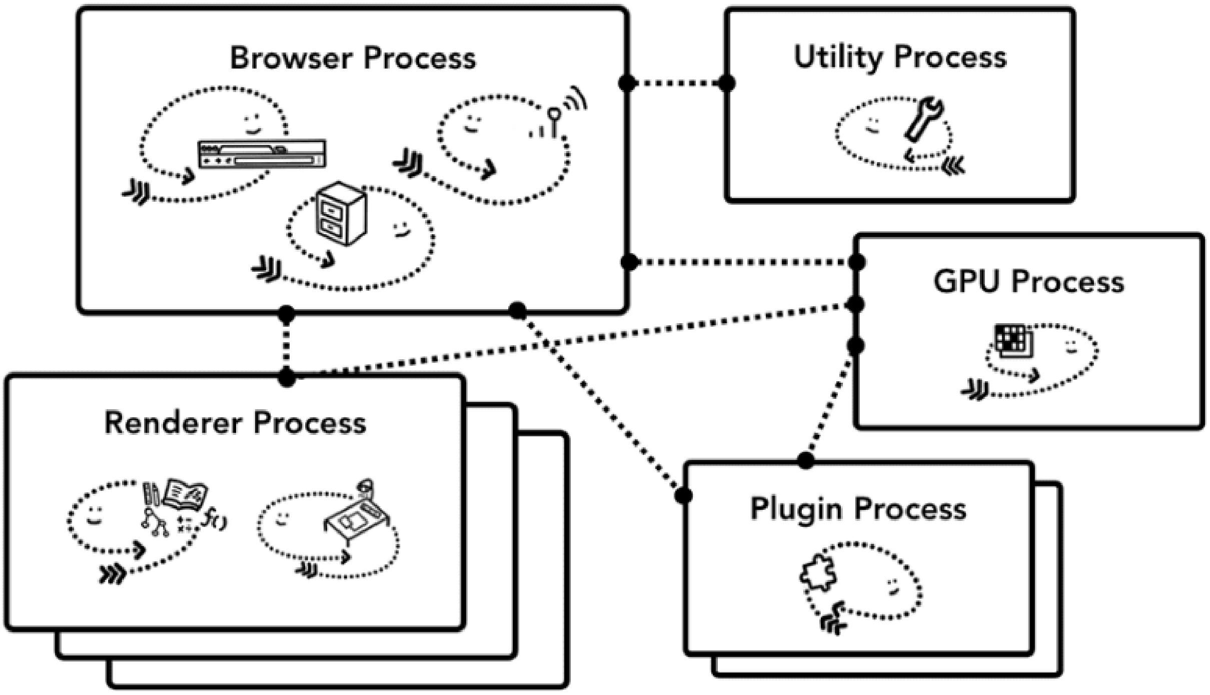
- Less time to create and terminate.
- Faster to context switch between threads.
- Quicker communication between threads.

Example

Google Chrome

- Core control part of the browser - browser process
- Rendering the web page and interaction - render process
- Plugins process
- Efficient running - GPU process

In Chrome, each and every tab you open gets its own content process. Each of those processes has its own memory. Maximizes performance Ten tabs, 10 processes. One hundred tabs, 100 processes.



Mozilla Firefox

- Four different processes. Multiple tabs run as threads within these processes.
- Relatively less memory utilization compared to Chrome.

Advantages of Multi-Thread

- Responsiveness
- Resource Sharing
- Economy
- Utilization of multiprocessor architectures

Thread Concurrency

- Concurrency is the execution of the multiple instruction sequences at the same time.
- Concurrency occurs when several process threads running in parallel.

Advantages

- Responsive applications
- Better resource utilization & performance.

Disadvantages

- Multi-threaded programs are harder to develop and can have bugs that are difficult to debug.

Example:

Let

- user1 balance = 2500
- user2 balance = 500

Thread A

```
def  transfer_amount(u1, u2, amt):
   b1 =  get_balance(u1)
   b2 =  get_balance(u2)
   b1 -= amt
   b2 += amt
   update_balance(u1, b1)
   update_balance(u2, b2)
transfer_amount(u1, u2, 1000)
```

PYTHON

If the above operation is executing in the Thread A, and after the execution of four line of code Thread B comes with priority. So context switching happens in middle of the bank operation.

Thread B

```
def  transfer_amount(u1, u2, amt):
    b1 =  get_balance(u1)
    b2 =  get_balance(u2)
    b1 -= amt
    b2 += amt
    update_balance(u1, b1)
    update_balance(u2, b2)
transfer_amount(u1, u2, 1600)
```

PYTHON

After the execution of Thread B

- user1 balance = 900
- user2 balance = 2100

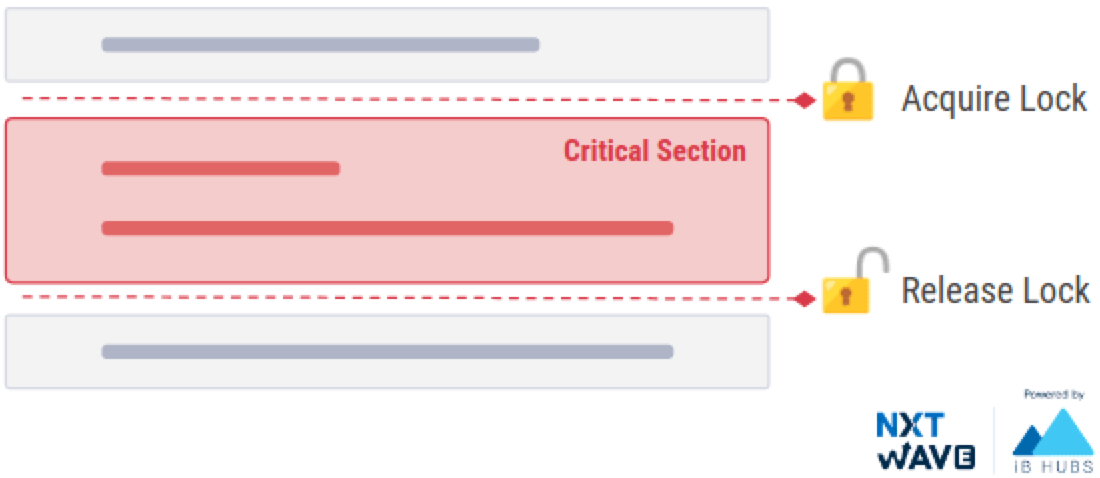
Thread A completes its remaining operation, after execution of Thread A

- user1 balance = 1500
- user2 balance = 1500

Which is not at all possible.

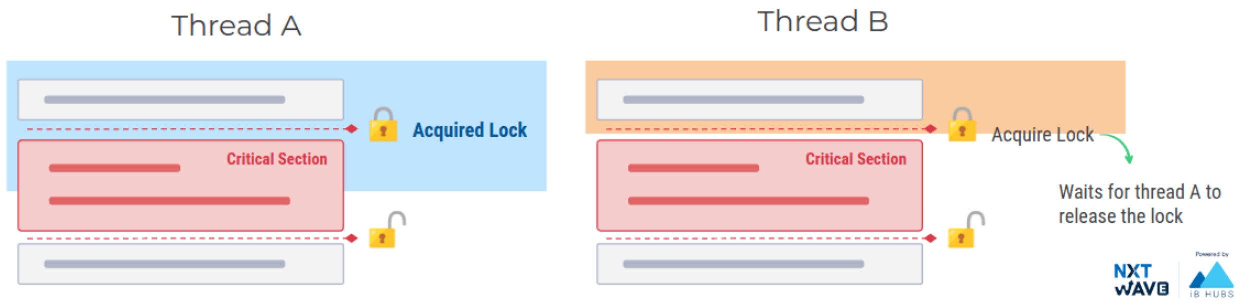
Solution

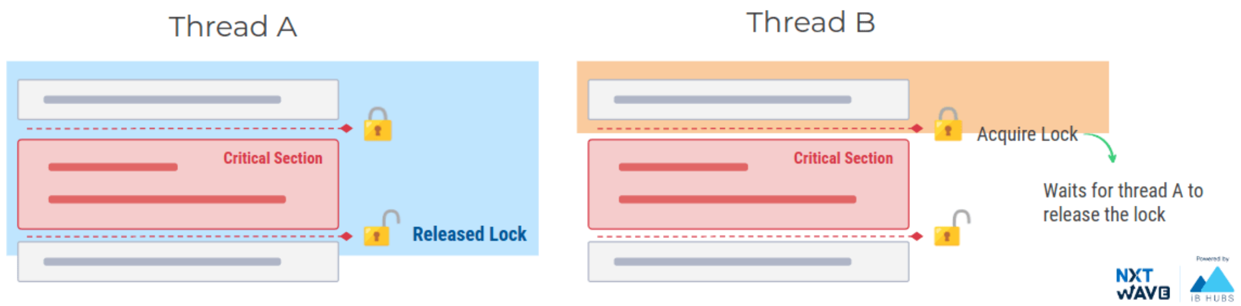
- The critical section is a code segment where the shared data is operated.
- Ensuring that only one process/thread can be in its critical section at a time eliminates concurrency issues.
- Use Locks to ensure this

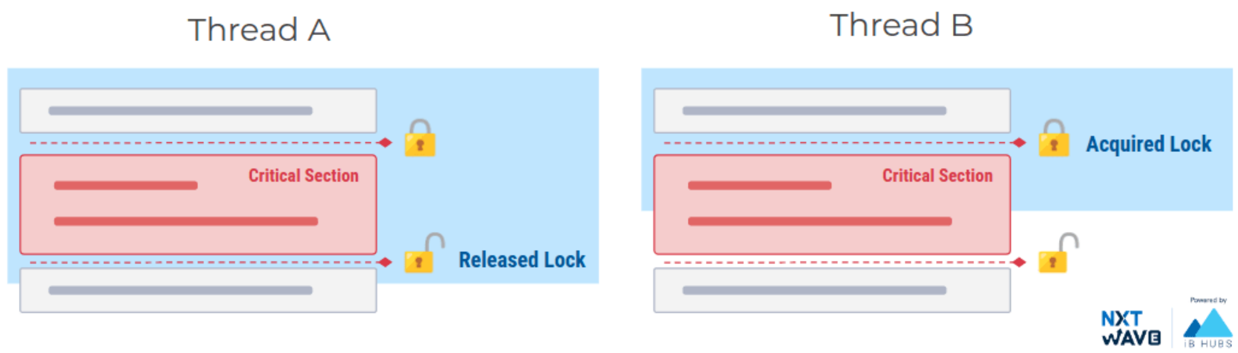


Locks allow us to limit number of threads running the critical section at a given time. Ensure all threads have to

- Acquire a lock before entering critical section
- Release after leaving critical section

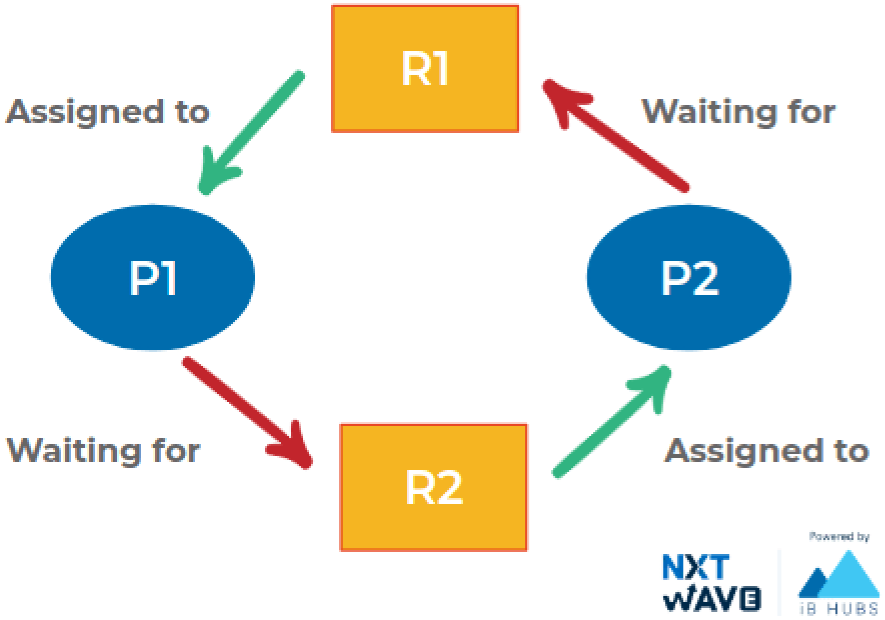






Deadlock

A situation where a set of processes are blocked because each process is holding a resource and waiting for another resource acquired by some other process.



Deadlock Conditions

Conditions for Deadlock:

- **Mutual exclusion**: The resource can be held by one process at a time
- **Hold and wait**: A process can hold multiple resources and still request more resources from other processes which are holding them.
- **No preemption**: A resource cannot be acquired from a process by force. A process can only release a resource voluntarily.
- **Circular wait**: If every process is waiting for each other
  to release the resource and no one is releasing their own resource. This is called a circular wait.

Dealing with Deadlock

- Avoid deadlocks by ensuring at least one of the above conditions is not met.
- Let deadlocks happen. Detect and break the deadlocks

Memory Management

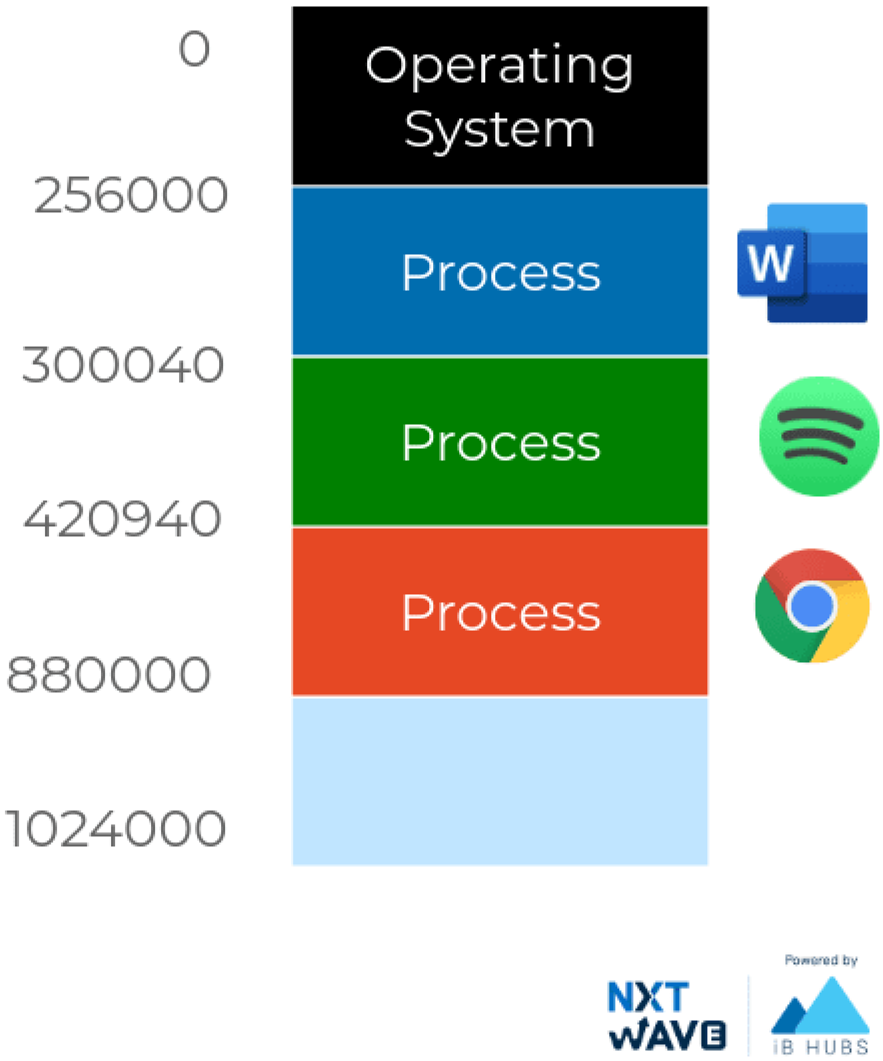
- Allocates physical memory to processes
- Tracks of allocated and available memory locations
- Manages Virtual Memory, etc

Terminology

- The terms Primary Memory, Main Memory & Physical Memory refer to RAM
- Secondary Memory refers to storage devices, such as hard drives and solid state drives.

Memory Allocation

- Ensure that each process can only access the address space allocated to it
- Restrict access to address space outside its allocated space



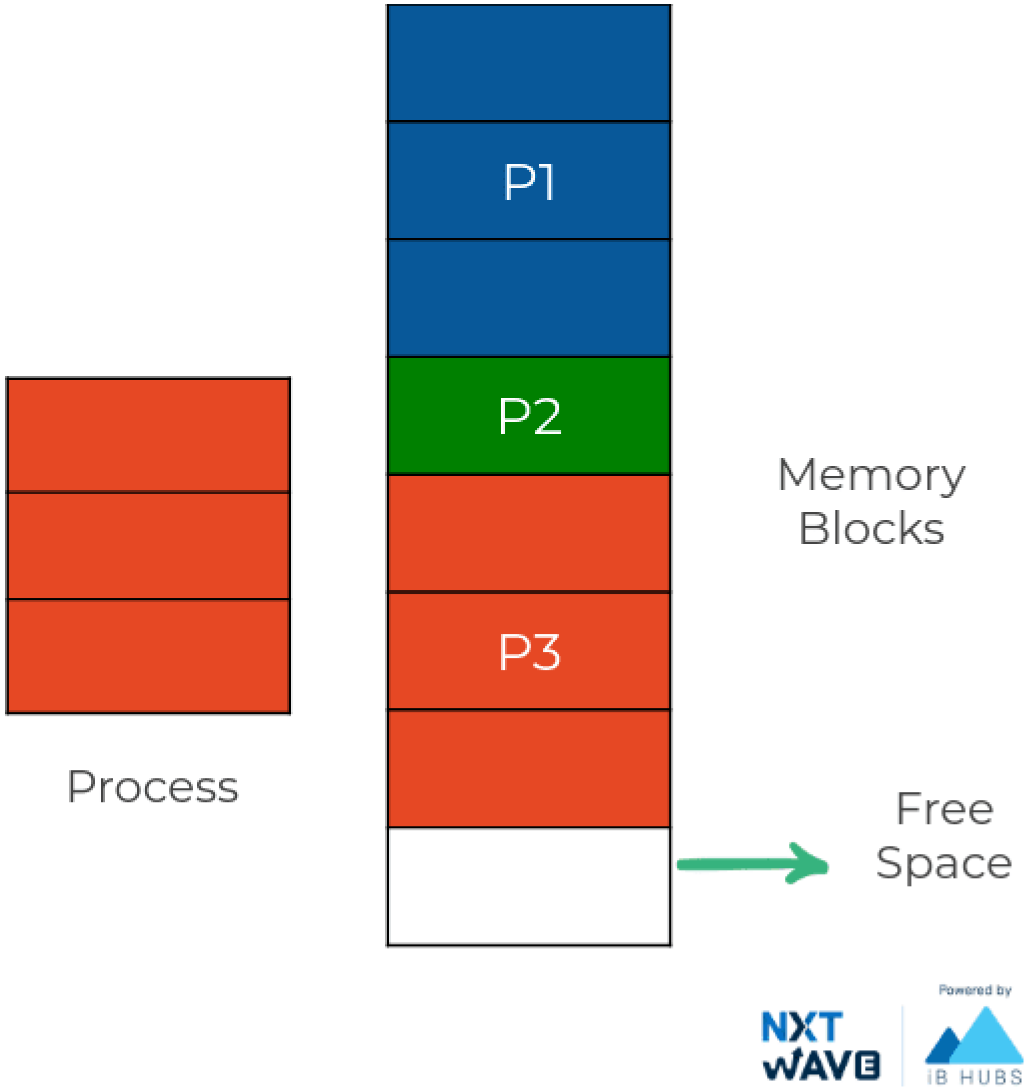
Memory allocation is a process by which computer programs are assigned memory.

- Contiguous Memory Allocation
- Non-Contiguous Memory Allocation

Contiguous Memory Allocation

A single contiguous section of memory blocks is assigned to the process according to its requirement.

Memory is also allocated as blocks of fixed sizes.



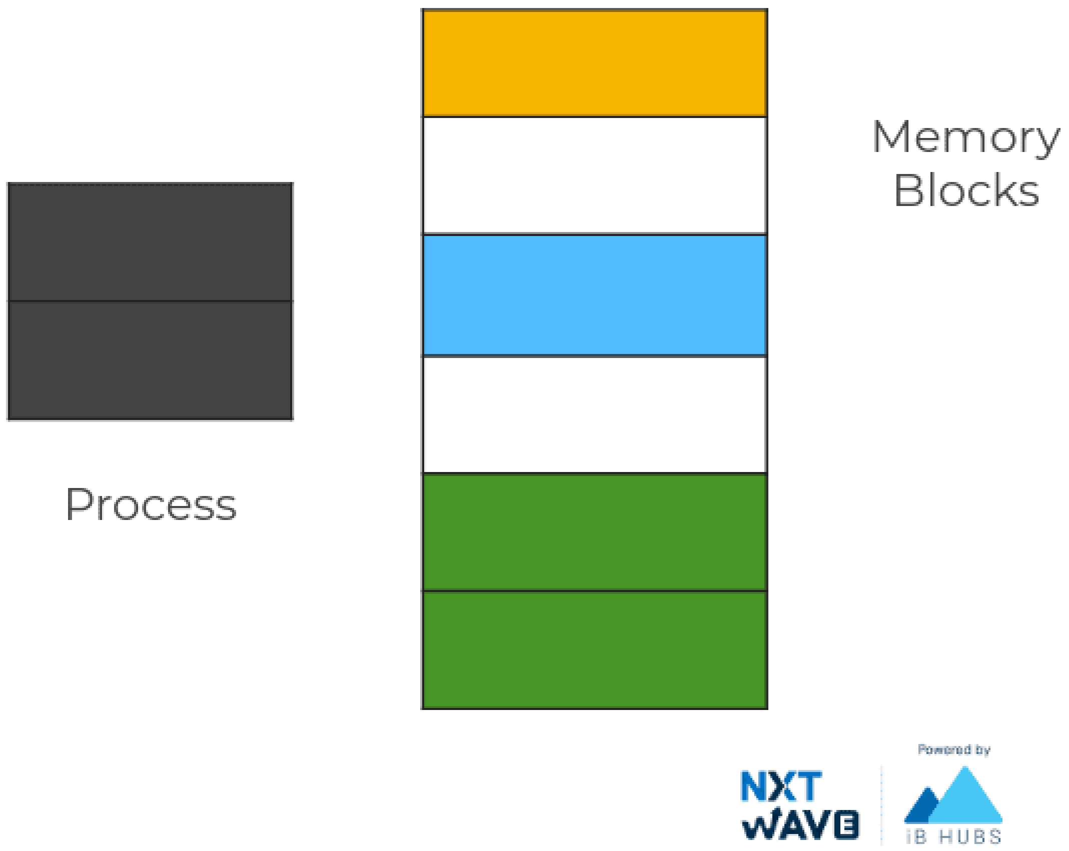
Fragmentation

Memory is used inefficiently, reducing capacity & performance.
Types of Fragmentation:

- External Fragmentation
- Internal Fragmentation

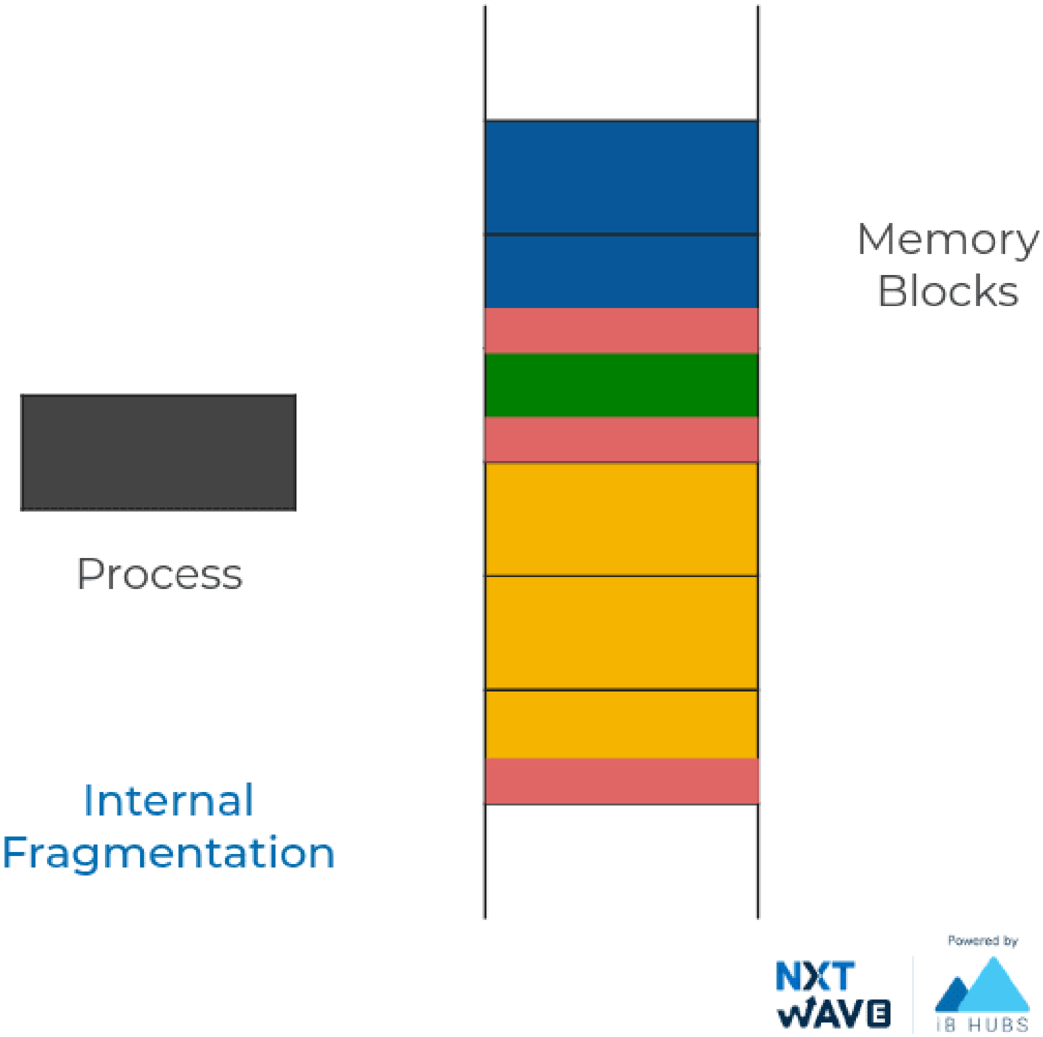
External Fragmentation

Total memory space is enough to satisfy a request, but cannot be used as it is not contiguous.



Internal Fragmentation

Process is allocated a memory block of size more than the size of that process. Due to this some part of the memory is left unused and this cause internal fragmentation.



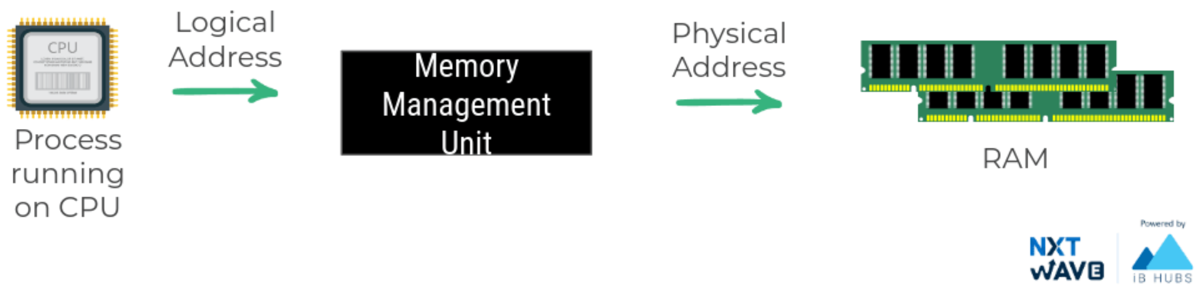
Non-Contigous Memory Allocation

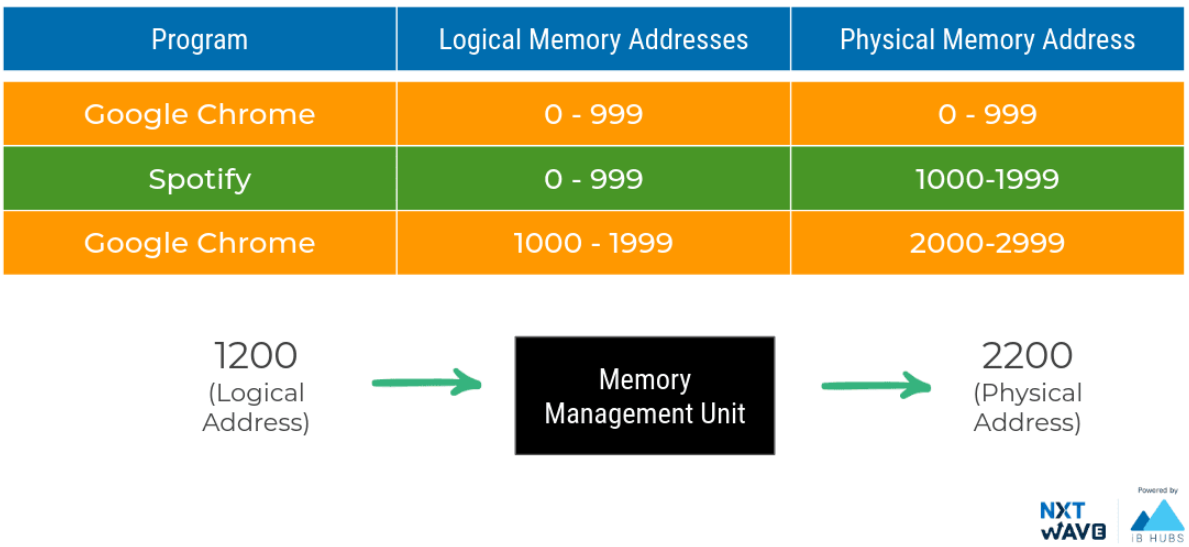
Assigns different blocks of memory in a non consecutive manner to a process.

- Paging
- Segmentation

Memory Management Unit

The run time mapping between Virtual address and Physical Address is done by hardware device known as MMU.



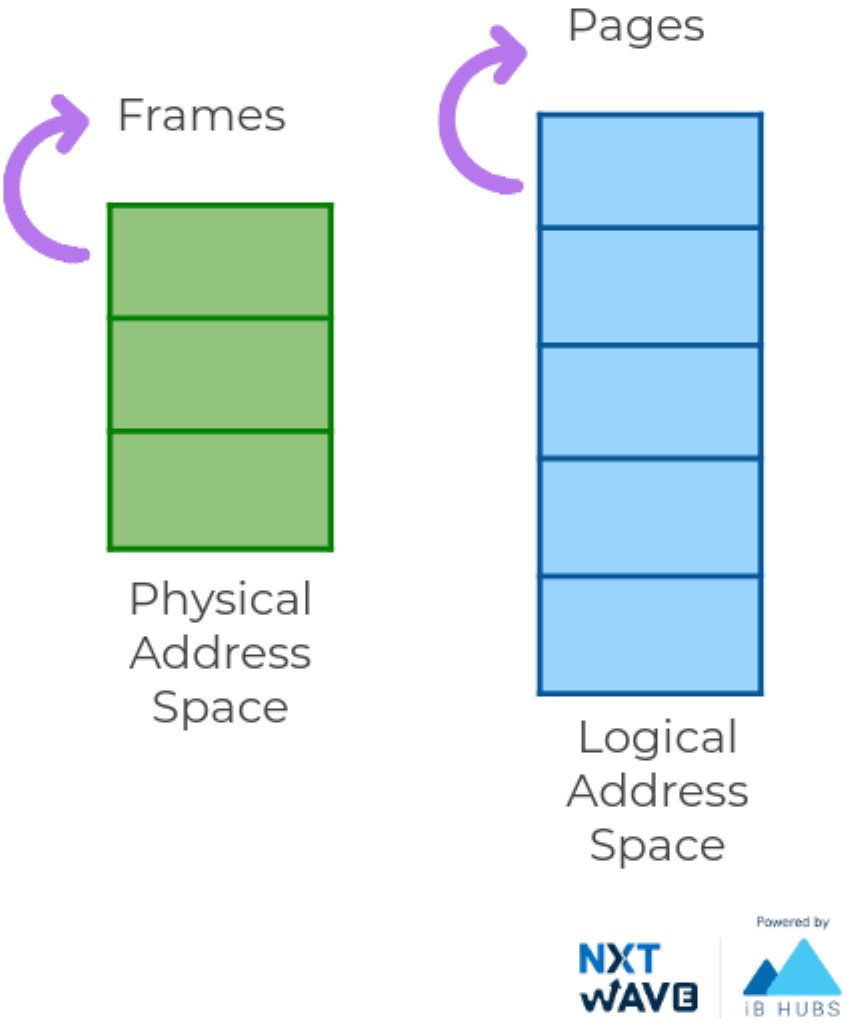


Program	Logical Memory Addresses	Physical Memory Address
Google Chrome	0 - 999	0 - 999
Spotify	0 - 999	1000-1999
Google Chrome	1000 - 1999	2000-2999

1200 (Logical Address)
Memory Management Unit
2200 (Physical Address)
NXT WAVE
Powered by
iB HUBS

Paging

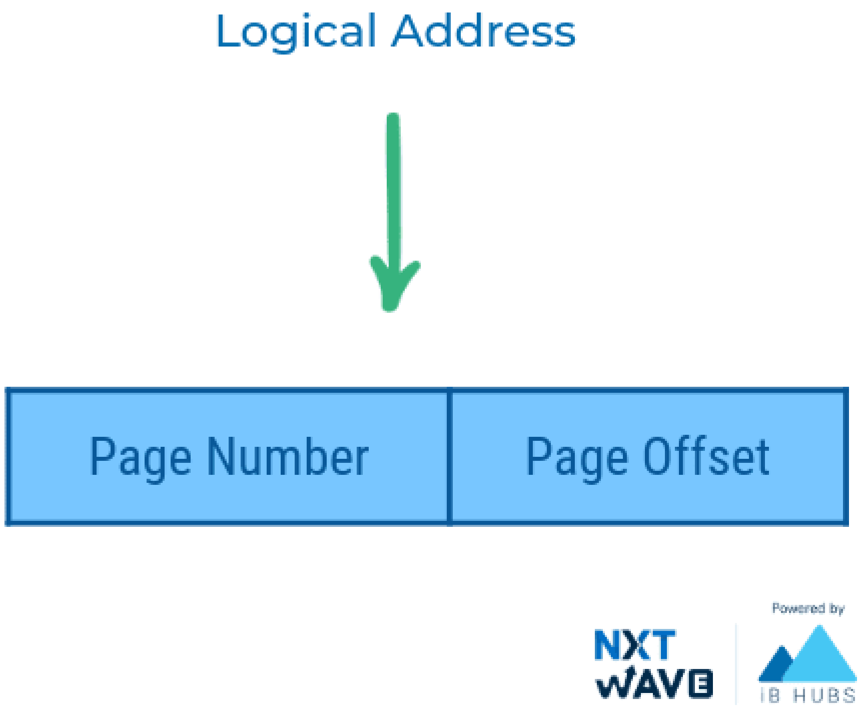
- Logical Memory is divided into equal fixed size partitions called pages.
- Each of this pages corresponds to a frame in physical memory.



Logical Address

The generated logical address can be broken into two parts

- Page Number
- Page Offset



Logical Address

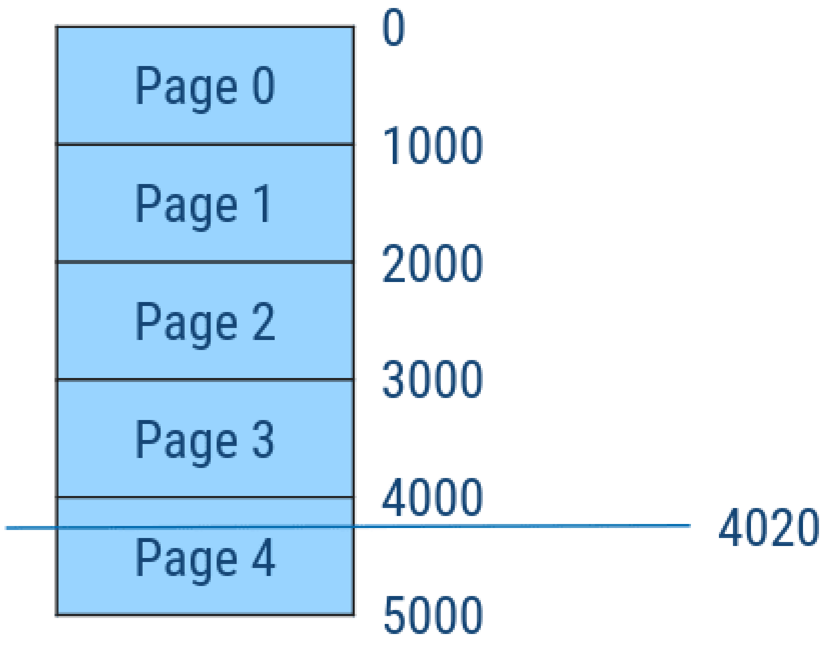
Page Number

Page Offset

Different parts of the same process can be stored at different places in the main memory.

Consider Page Size of 1000 bytes,

Logical Address Location 4020 = Page 4, Offset 20



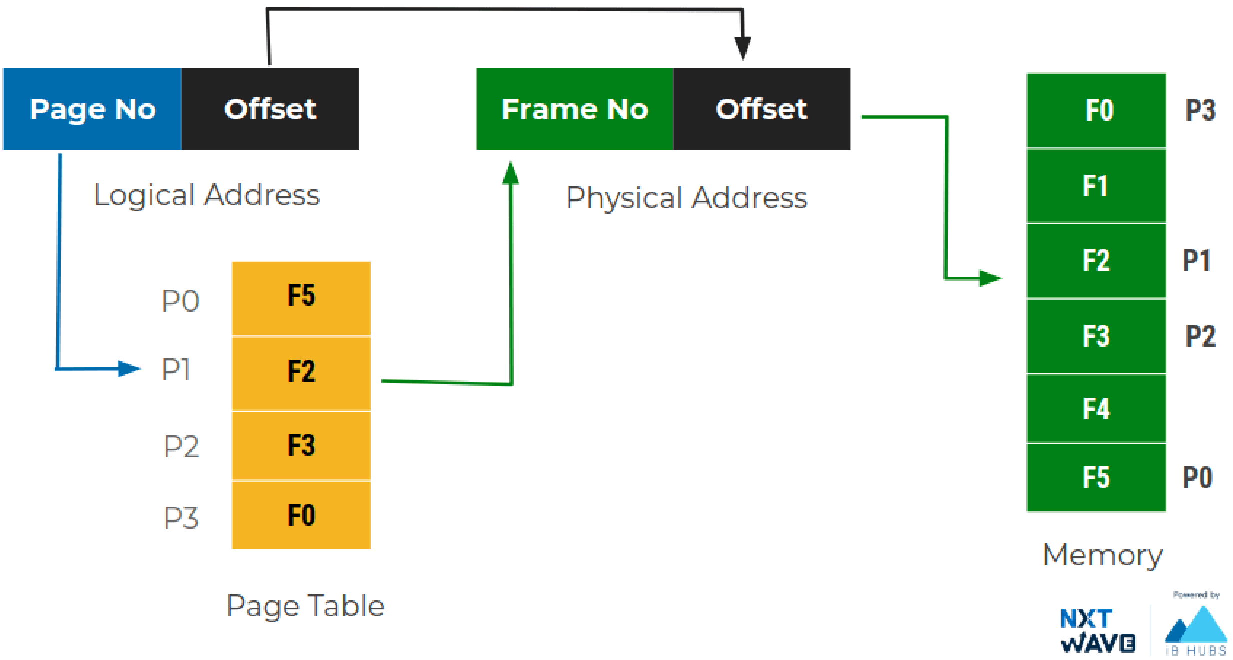
Page Table

Maps the page number referenced by the CPU to the frame number where that page is stored.

- Each process has its own independent page table.

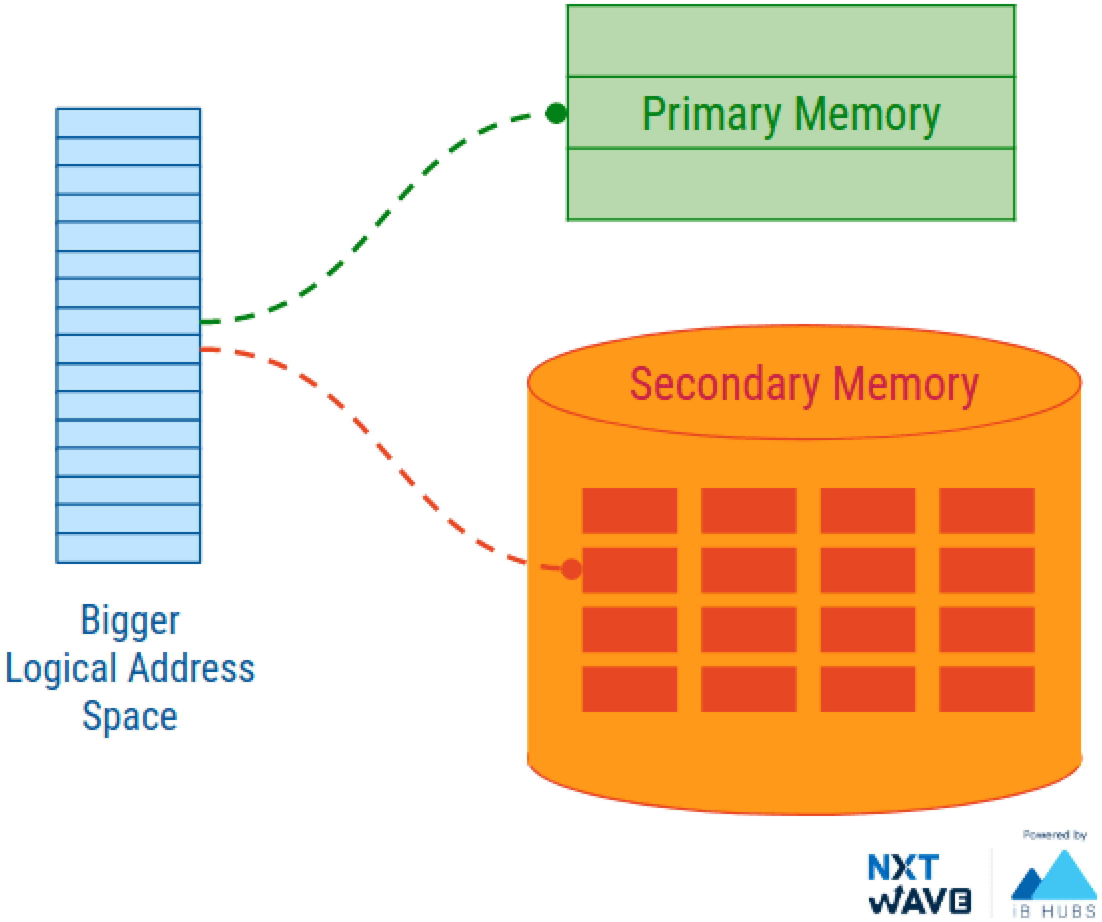
Paging

- Frame number specifies the specific frame where the required page is stored.
- Page Offset specifies the specific word that has to be read from that page



Virtual Memory

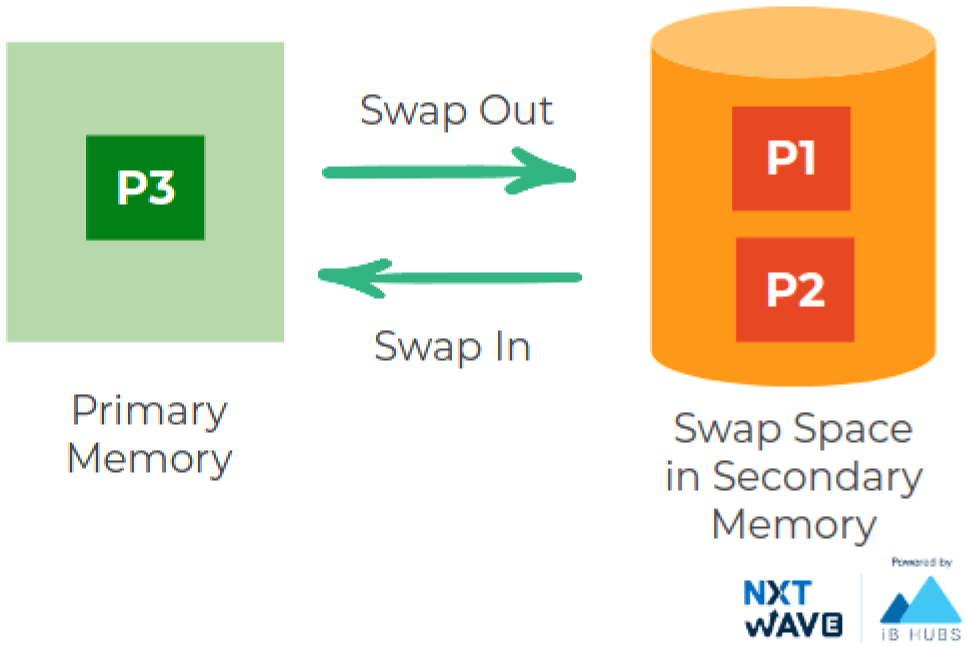
- Virtual memory is a memory management technique that creates an illusion to users of larger physical memory.
- Ability to execute partially-loaded program.
- When using virtual memory, logical address is also referred as virtual address.
- The secondary memory used is also referred to as swap space.



Swapping

Moving memory of a process between Primary Memory and Secondary Memory is called Swapping

- Swap In the currently executing process
- Swap Out the least required process



Demand Paging

- Pages are loaded only on demand, not in advance.
- Page referred for the first time in the main memory, then that page will be found in the secondary memory.

Page Replacement Algorithm

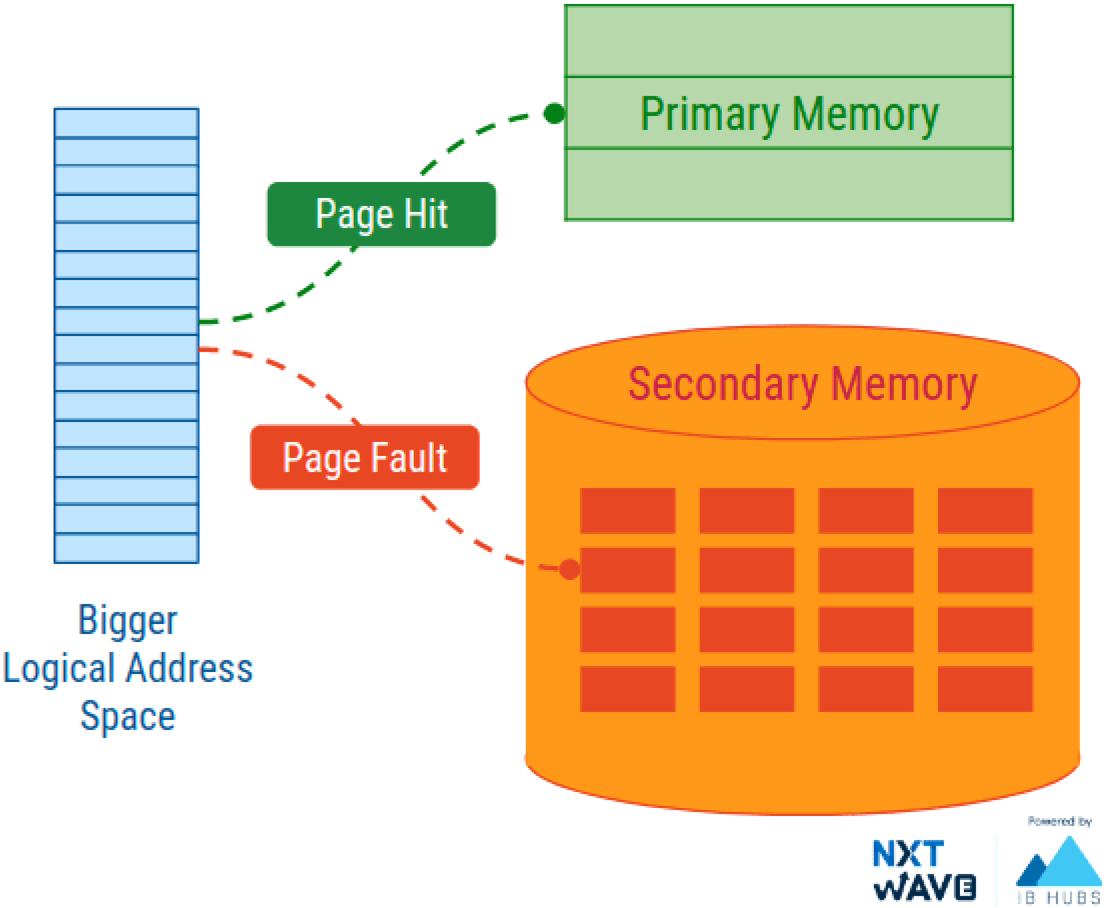
Page replacement algorithm decides which page to remove when a new page needs to be loaded into the primary memory.

Page Replacement Algorithms :

- Least Recently Used (LRU)
- Least Frequently Used (LFU)

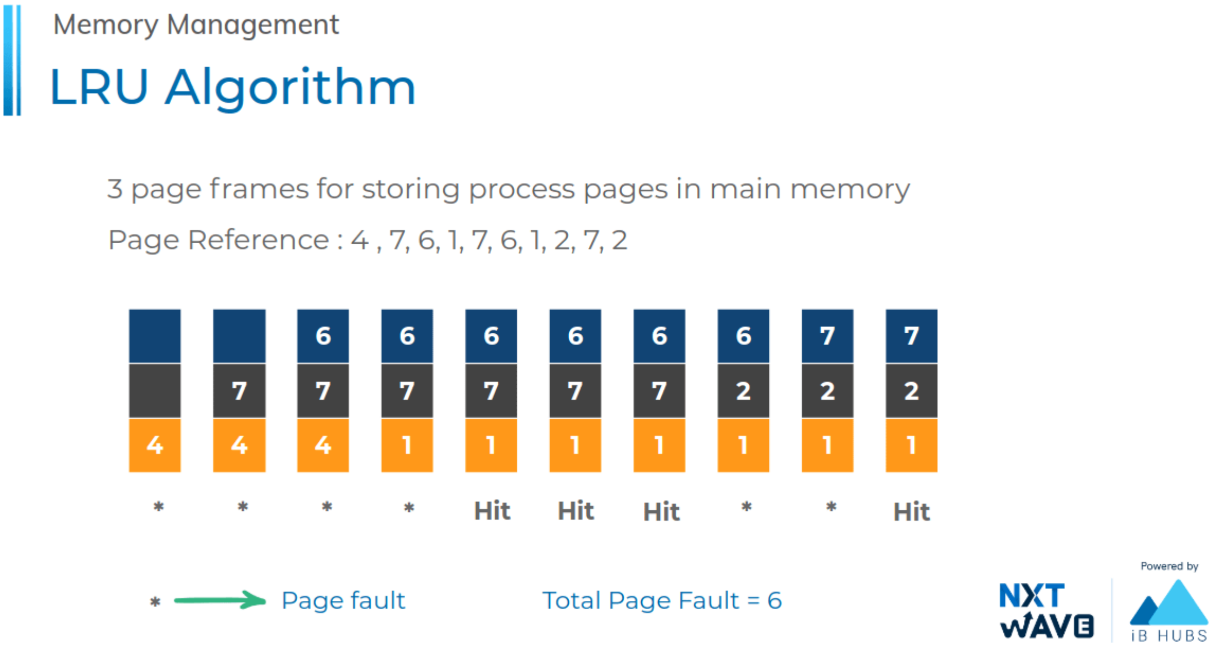
and many more...

- If a process requests for page and that page is found in the main memory then it is called page hit.
- If page is not found in memory then it is page miss or page fault.



LRU Algorithm

- Least Recently Used (LRU) page.
- It replaces the page that has not been used for the longest period of time.



Memory Management

LRU Algorithm

3 page frames for storing process pages in main memory

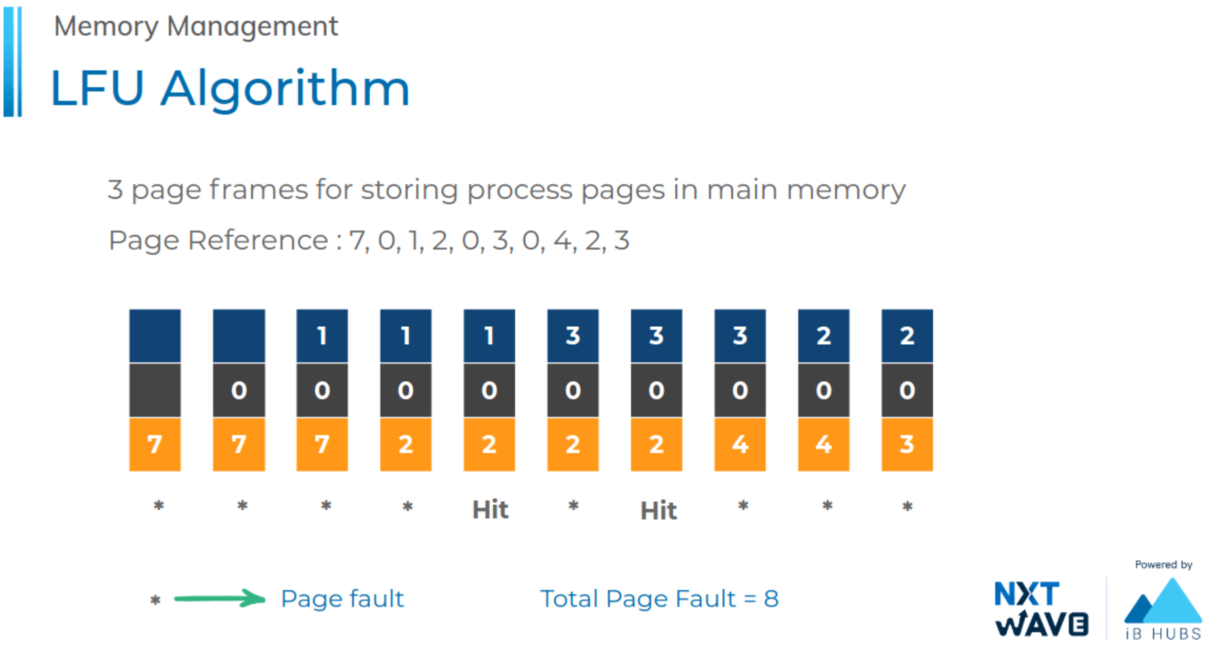
Page Reference : 4 , 7, 6, 1, 7, 6, 1, 2, 7, 2

		6	6	6	6	6	6	7	7
	7	7	7	7	7	7	2	2	2
4	4	4	1	1	1	1	1	1	1
*	*	*	*	Hit	Hit	Hit	*	*	Hit

* → Page fault   Total Page Fault = 6

LFU Algorithm

- Page with the minimum no. of previous hits is selected for replacement.
- If counts are equal than it applies FIFO.



Memory Management

LFU Algorithm

3 page frames for storing process pages in main memory

Page Reference : 7, 0, 1, 2, 0, 3, 0, 4, 2, 3

		1	1	1	3	3	3	2	2
	0	0	0	0	0	0	0	0	0
7	7	7	2	2	2	2	4	4	3
*	*	*	*	Hit	*	Hit	*	*	*

* → Page fault   Total Page Fault = 8

Understanding Computer Networks

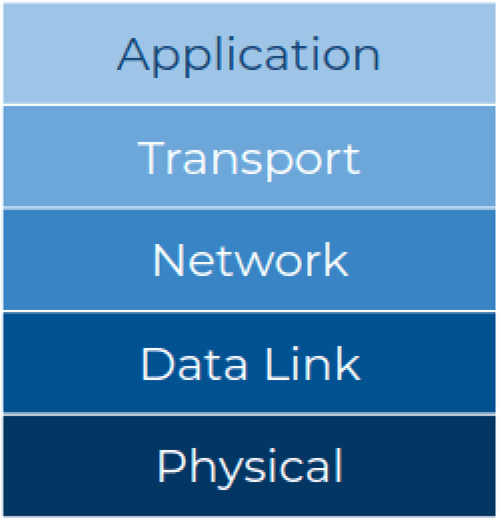
Internet

Computer network that interconnects billions of computing devices throughout the world

- Billions of Users
- Billions of Devices
- Millions of Applications

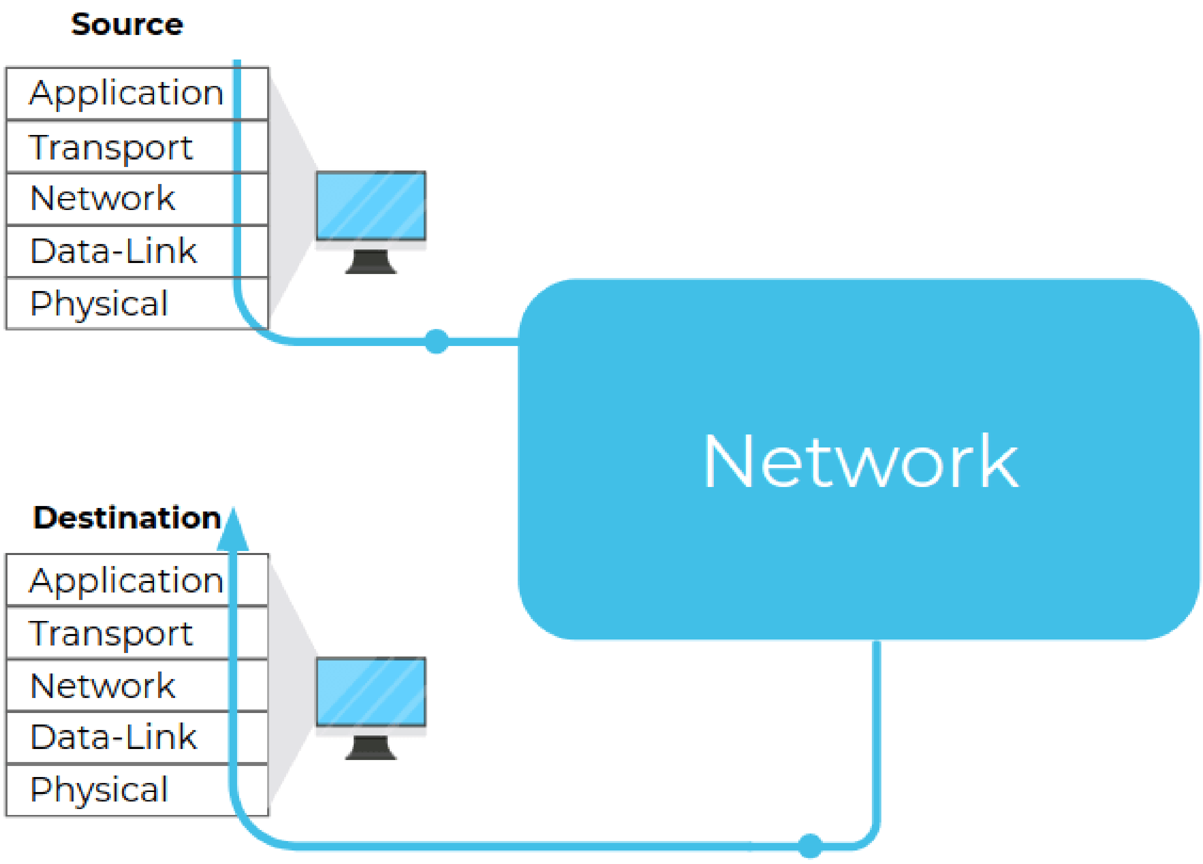
Network Model

- To reduce the complexity, networks are organized as a stack of layers.
- Each layer offers certain services to the layers above it and abstracts the layers below it.



TCP/IP Network Model

The header (and footer) and the data together form the PDU for the next layer. The process continues until reaching the lowest-level layer (physical layer or network access layer), from which the data is transmitted to the receiving device. The receiving device reverses the process, de-encapsulating the data at each layer with the header and footer information directing the operations. Then the application finally uses the data. The process is continued until all data is transmitted and received.



Physical Layer

It represents the physical devices that interconnect computers and the technologies that develop them

- Networking Cables
- Wi-Fi and Cellular Hardware

Signals are sent over these media.

Data Link Layer

Data link layer is responsible for defining a common way of interpreting signals.

Network Layer

Network Layer is responsible for getting data delivered across a collection of networks.

- Internet Protocol (IP)

Transport Layer

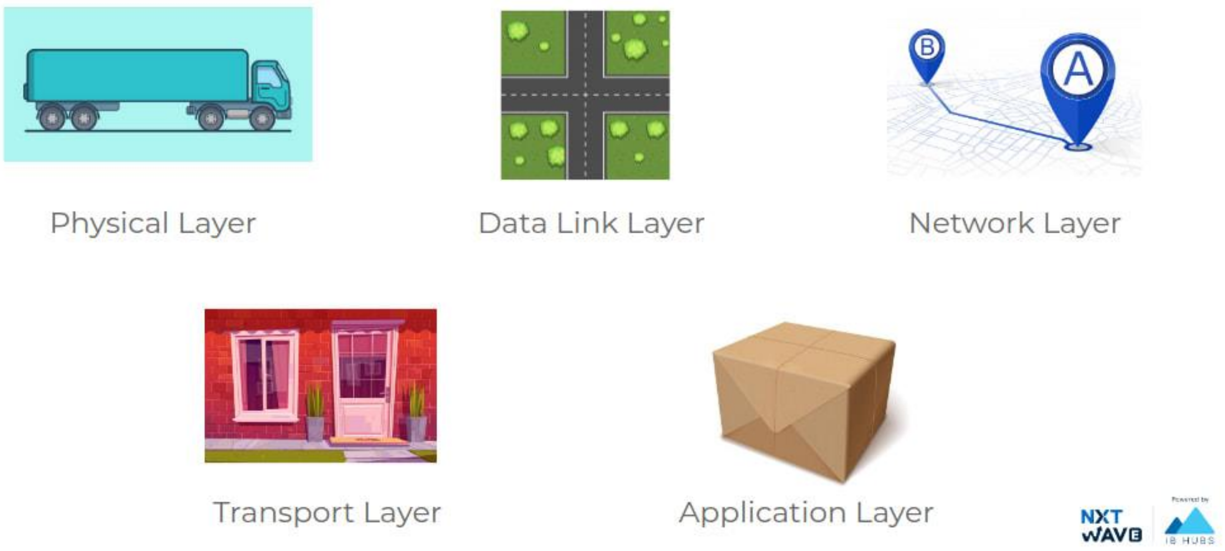
Transport Layer is responsible for ensuring that data gets to the right process (application).

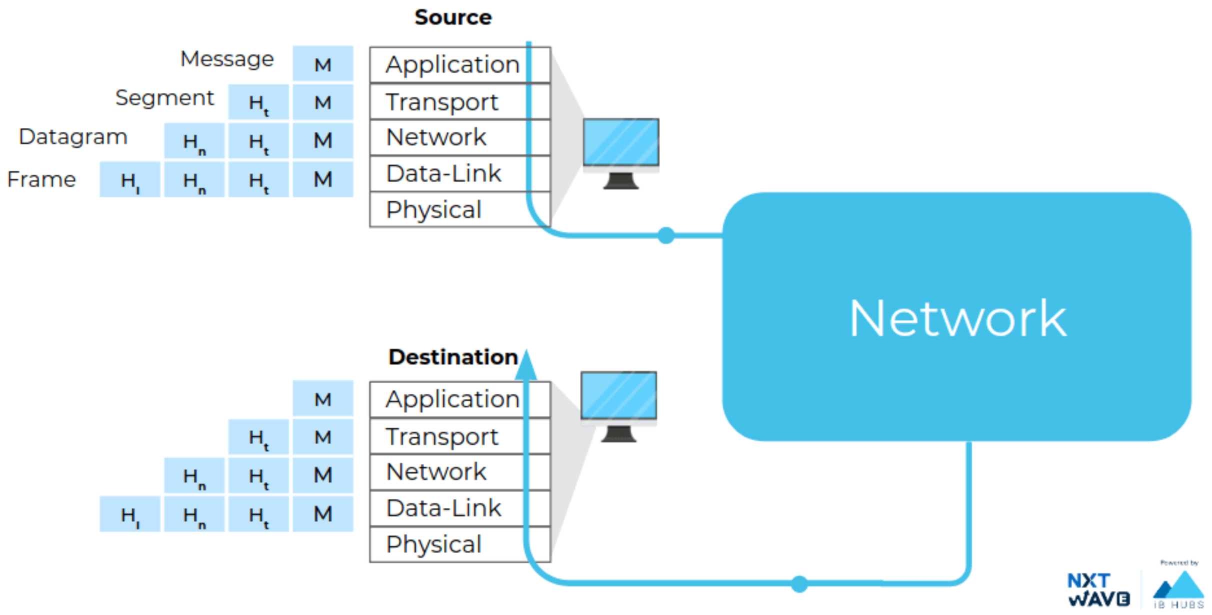
Application Layer

The protocols in Application Layer are used to allow you to browse the web or send receive email are some common ones.

TCP/IP Network Model

You can think of layers like different aspects of a package being delivered. The physical layer is the delivery truck and the roads. The data link layer is how the delivery trucks get from one intersection to the next over and over. The network layer identifies which roads need to be taken to get from address A to address B. The transport layer ensures that delivery driver knows how to knock on your door and deliver the package to appropriate person. And the application layer is the contents of the package itself.





Protocol

A protocol is a standard set of rules that allow electronic devices to communicate and understand one another.

- Format of message
- Order of messages

Layer Name	Protocol
Application	HTTP, SMTP,...
Transport	TCP, UDP
Network	IP
Data Link	Ethernet, Wi-Fi
Physical	10 Base T, 802.11

Wired Network Cables

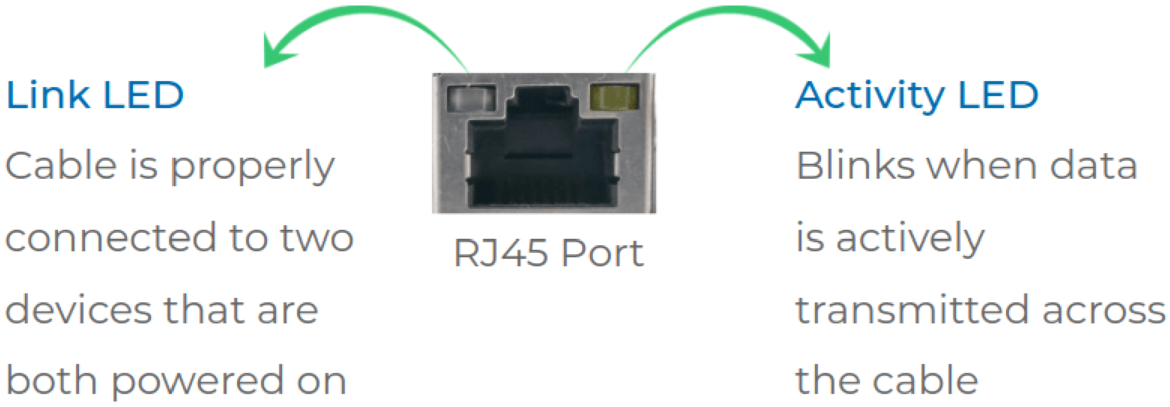
Cables are what connect different devices to each other, allowing data to be transmitted over them.

Network Port

Network ports are generally directly attached to the devices that make up a computer network.

RJ45 Port

RJ45 plug



Link LED

Cable is properly connected to two devices that are both powered on

RJ45 Port

Activity LED

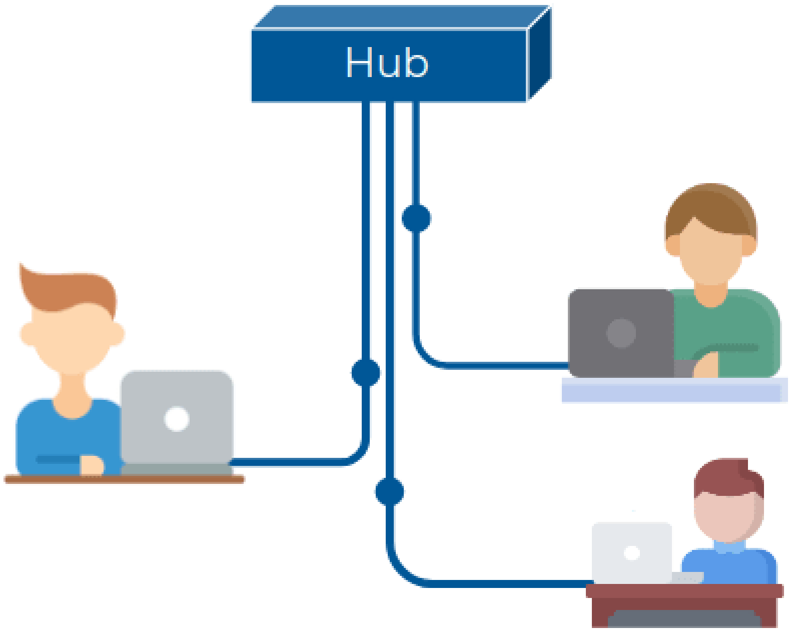
Blinks when data is actively transmitted across the cable

Two computers can communicate across the cable connected between them.

How data is transferred If they are more than two devices?

Hub

- Transfers data to every other port connected to the hub
- The System needs to determine if the incoming data was meant for them or not.



How does the system knows which data is meant for them?

MAC Address

Media Access Control (MAC) address is a globally unique identifier attached to an individual network interface (hardware).

- 48-bit number
- six groupings of two hexadecimal numbers(0-9, A, B, C, D, E, F)

00	1A	3F	F1	4C	C6

MAC address are same for your system no matter where you use it.

Data Collisions

When two computers were to send data across the wire at the same time, the data signals can interfere. Collision Domain is a network where collisions can occur.

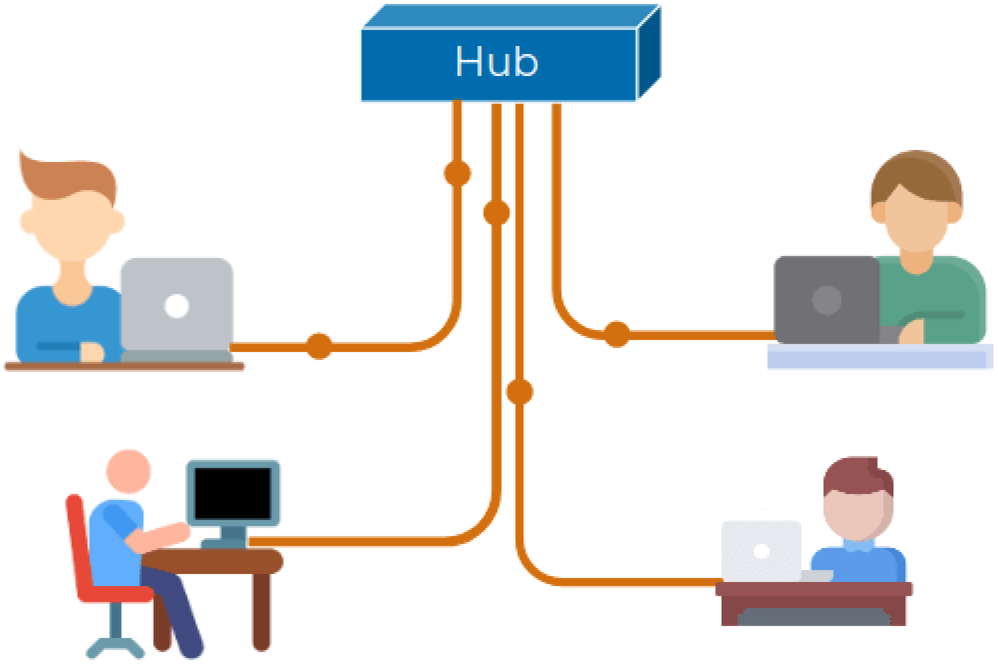
Carrier Sense Multiple Access

CSMA used to determine when the communications channels are clear and when the device is free to transmit data.

- Listen-Before-Talking

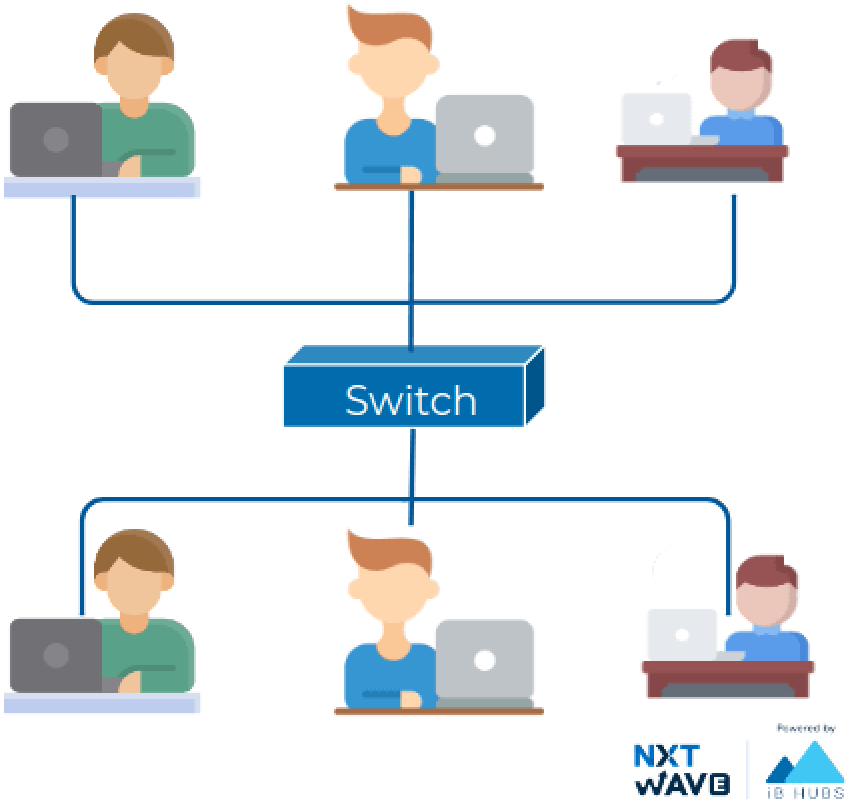
Collision domain

- The electrical pulses sent across the cable can interfere with each other.
- Systems to have to wait for long time before they try sending their data again.



Network Switch

- Switch determine which system the data is intended for and send that data to that one system.
- Transmissions in different small networks can happen at once
- It store MAC address



Hub Vs Switch

Hub	Switch
They operate in the physical layer	They operate in the data link layer
It sends message to all ports.	It sends message to selected destination ports.
Collisions occur mostly in setups using hubs.	Less collisions occur in switch.
A network hub can't store MAC addresses.	Switch stores the MAC addresses

Network Layer

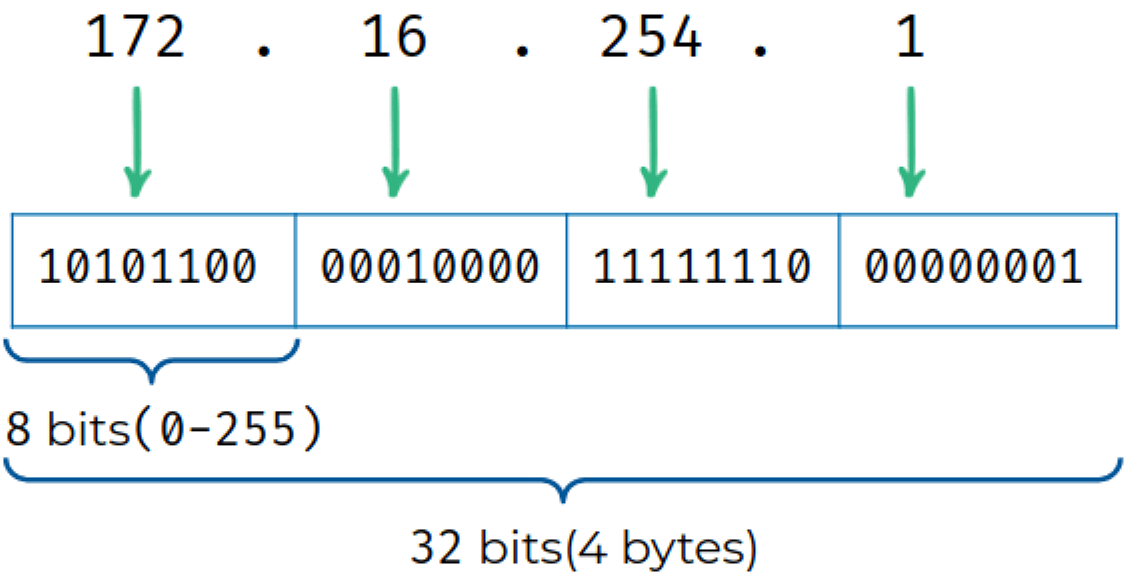
IP Address

IP addresses belong to the networks, not the devices attached to those networks.

Example: 172.16.254.1

IPv4 Addressing

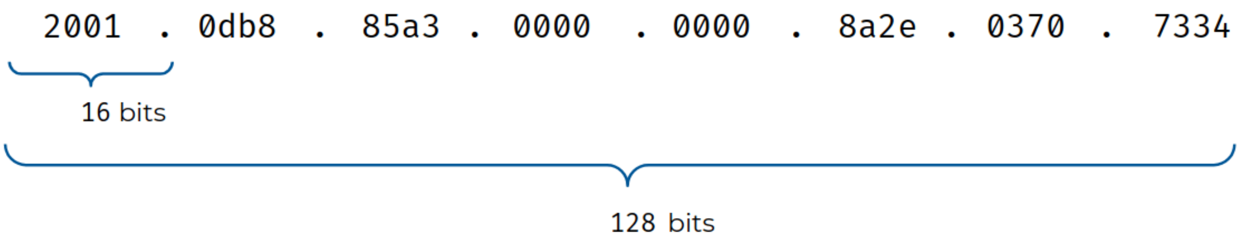
- IP addresses are a 32 bit long numbers made up of four octets,
- each octet is normally described in decimal numbers.
- 8 bits of data or a single octet can represent all decimal numbers from 0 to 255. This format is known as dotted decimal notation.



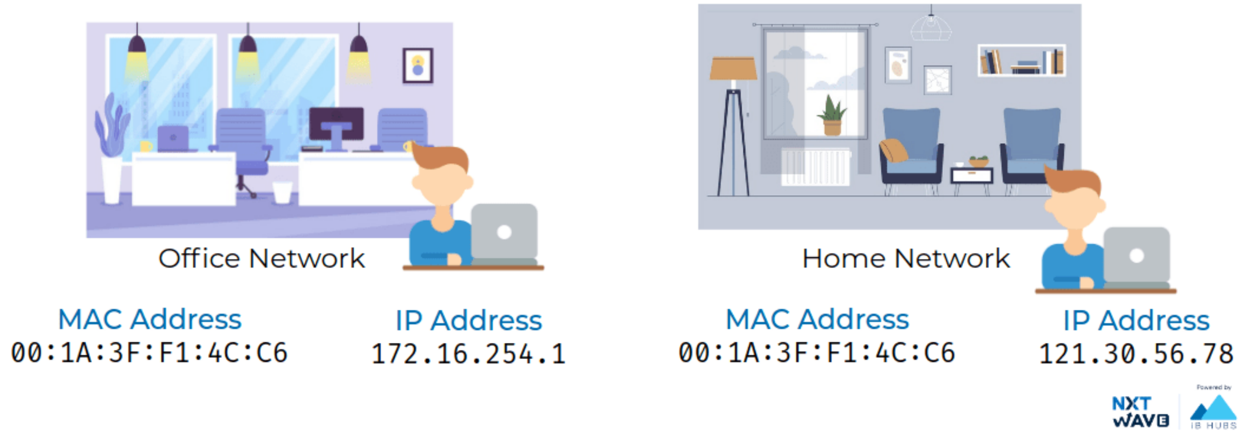
$2^{32} = 4 \times 10^9$ Possible IPs

IPv6 Addressing

- IPv6 IP addresses are a 128 bit long numbers made up of eight groups of four hexadecimal digits,
- each group representing 16 bits



IP addresses belong to the networks, not the devices attached to those networks.your laptop will always have the same MAC address no matter where you use it, but it will have a different IP address assigned to it at an office than it would when you're at home. The LAN at the office, or the LAN at your house would each be individually responsible for handing out an IP address to your laptop if you power it on there.



Router

- Forward data between independent networks.
- Store route information for different networks all over the world.
- Network Layer device

Wireless Router

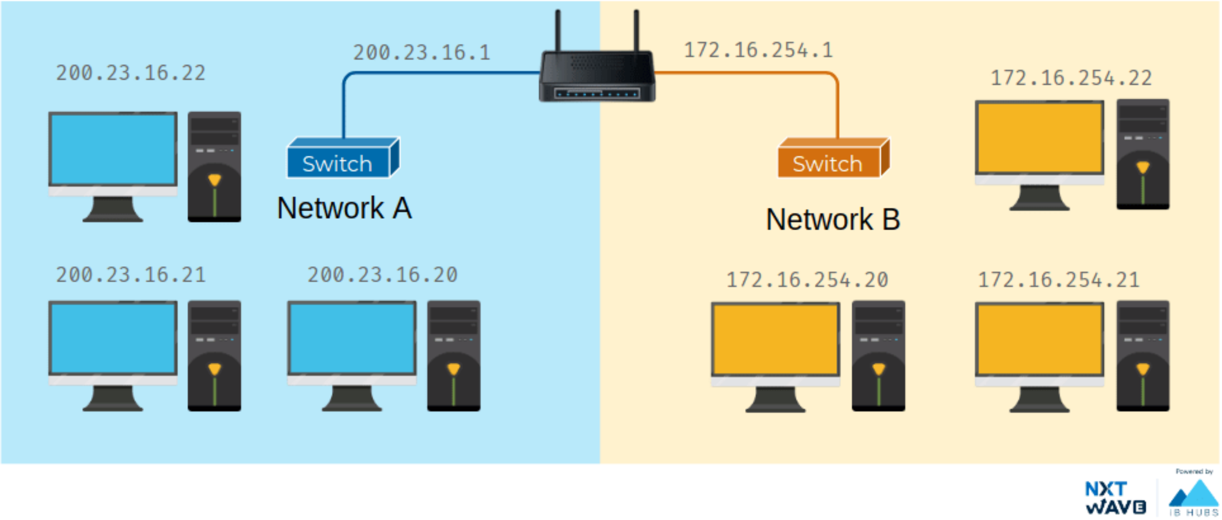
Wireless Routers are combination of

- Router
- Switch
- Wireless Access Point

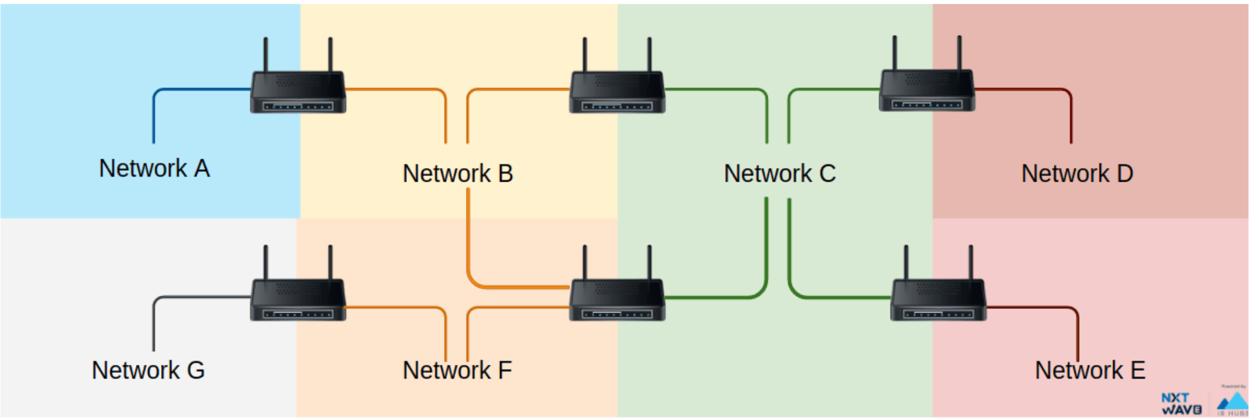
Wireless Access Point

Wireless access point (WAP): Allows other Wi-Fi devices to connect to a wired network.

Router

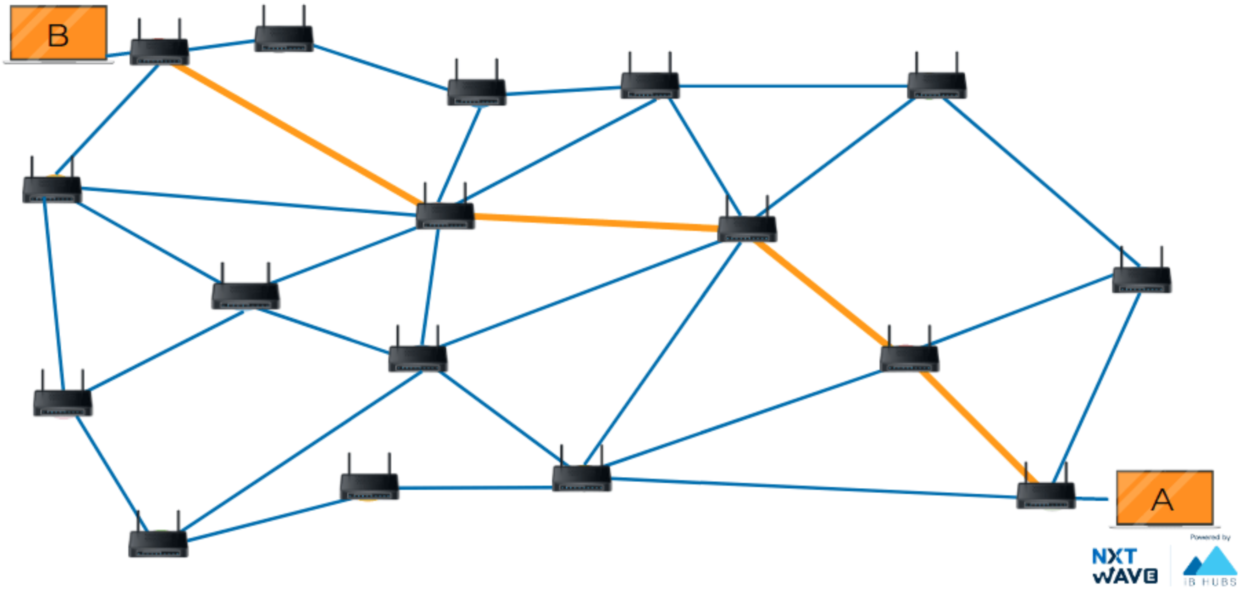


Network of Networks



Route

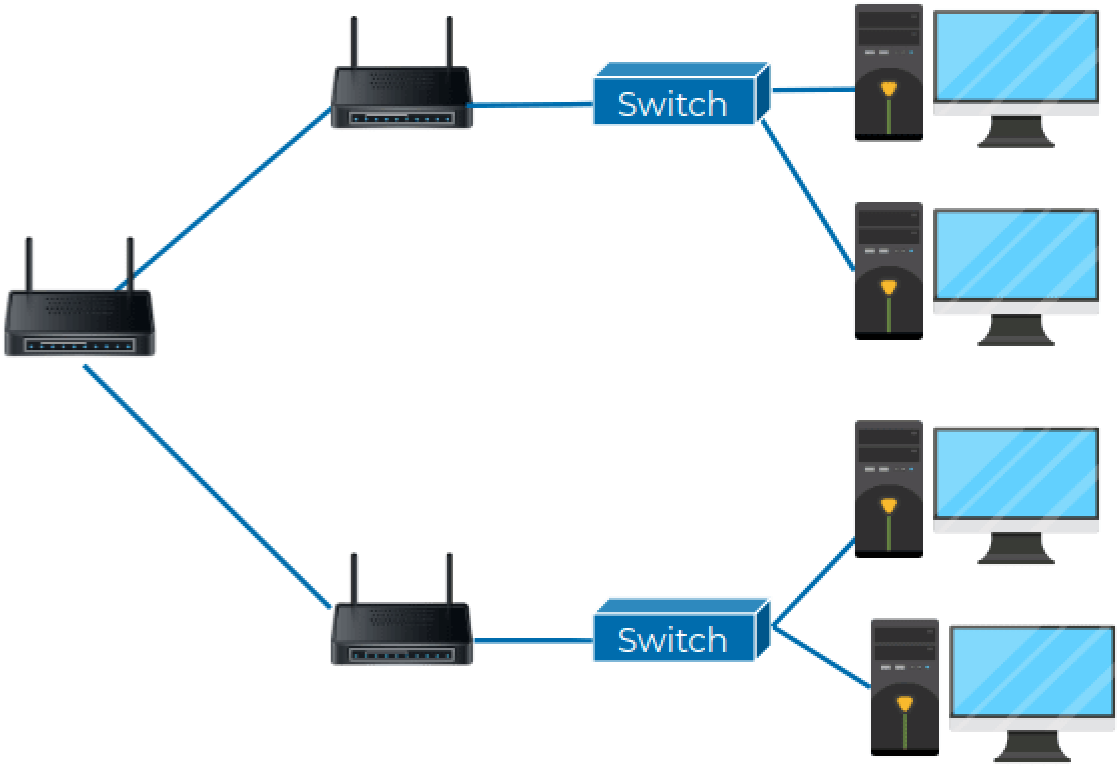
- The sequence of communication links and packet switches traversed by a packet from the sending end system to the receiving end system is known as a route or path.
- The routers are constantly trying to balance the load across whataevere routes they know to ensure speedy and reliable delivery which is conjunction control.



Subnetting

Subnetting is the process of splitting a large network into many smaller sub-networks or subnets

- Network Efficiency
- Security
- Easier Management

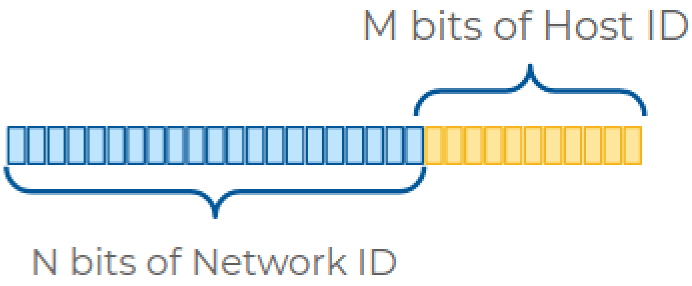


Subnetting IP Address

Network ID: Used to identify the network.

Host ID: Used to identify individual hosts in the network.

Subnet Mask is used to identify the network ID and Host ID



IP Address	223	1	1	2
IP Address (Binary)	11011111	00000001	00000001	00000010

IP Address	223	1	1	2
Subnet Mask (Binary)	11111111	11111111	11111111	00000000
Subnet Mask	255	255	255	0

CIDR Notation

The slash notation (a.b.c.d/x) is also known as CIDR notation. It can represent an IP and a subnet mask.

CIDR	223.1.1.2/23
IP Address	223.1.1.2
Subnet mask	255.255.254.0

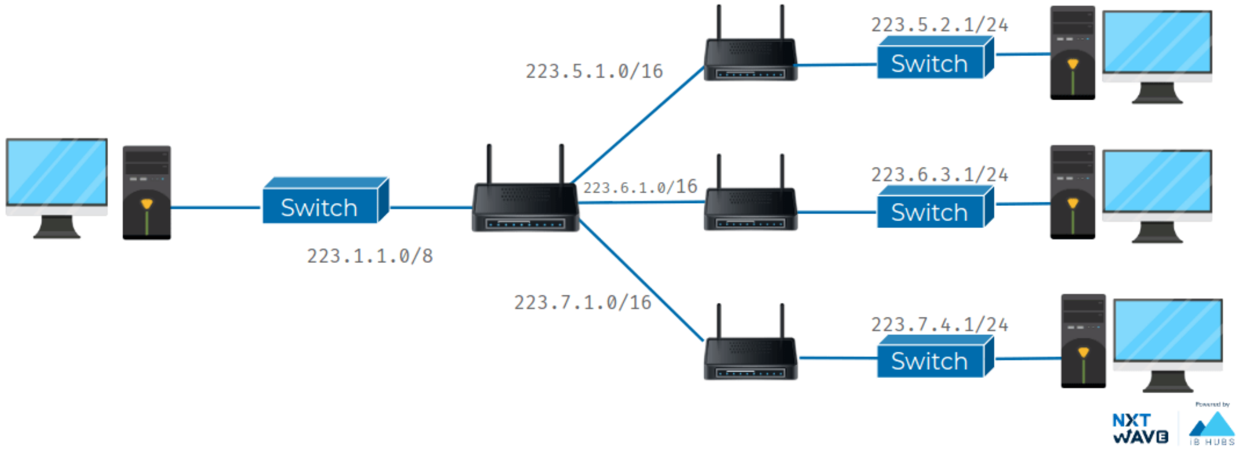
- The entire IP and subnet mask can be written now as 223.1.1.2/24
- /24 notation represents no of ones in subnet mask.
- We know that the octet available for host IDs can contain the numbers 0-255, but zero is generally not used and 255 is normally reserved as a broadcast address for the subnet. This means that, really, only the numbers 1-254 are available for assignment to a host.

CIDR	IP Address in Binary	IP Address in Binary	IP Address in Binary	IP Address in Binary	IP address Range	Possible Hosts
223.1.1.0/24	11011111	00000001	00000001	00000000	223.1.1.0 to223.1.1.255	254
223.2.1.0/23	11011111	00000010	00000001	00000000	223.2.0.0 to223.2.1.255	510
223.3.1.0/22	11011111	00000011	00000001	00000000	223.3.0.0 to223.3.3.255	1022

Note

*First address in the network is reserved for the network.

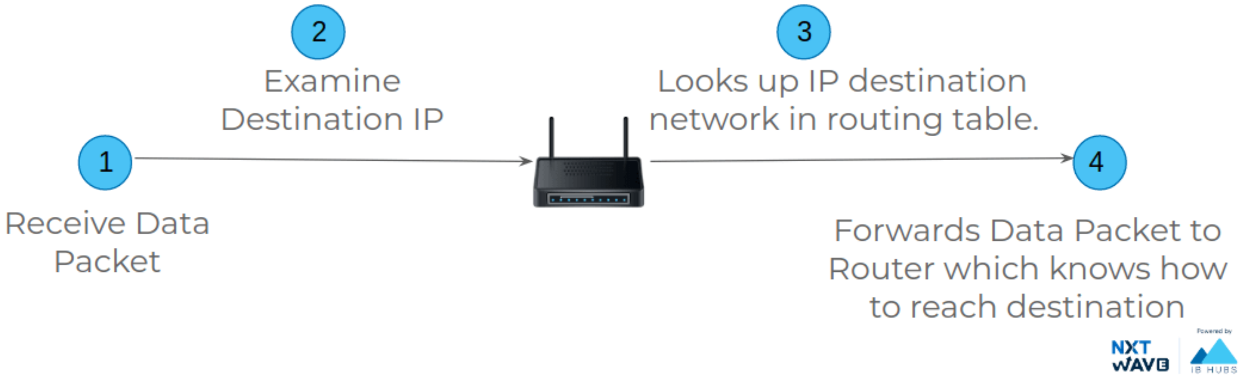
*Last address in the network would be reserved for broadcasting.



Forwarding

Basic routing has just a few steps

1. a router receives a packet of data on one of its interfaces
2. the router examines the destination IP of this packet.
3. the router then looks up the destination network of this IP in its routing table.
4. the router forwards that out though the interface that's closest to the remote network. As determined by additional info within the routing table.
5. These steps are repeated as often as needed until the traffic reaches its destination.

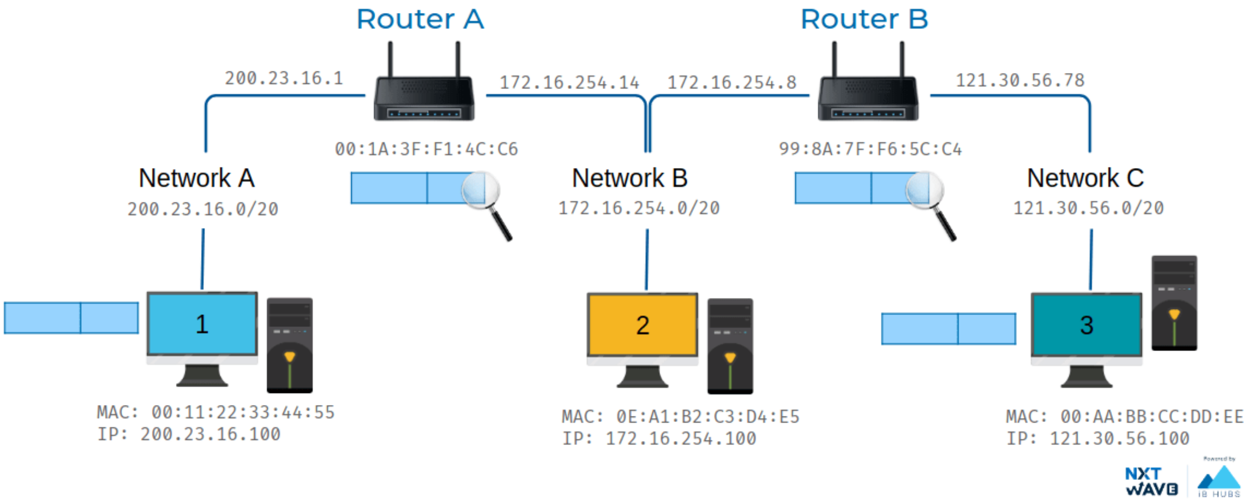

2
Examine Destination IP
3
Looks up IP destination network in routing table.
1
Receive Data Packet
4
Forwards Data Packet to Router which knows how to reach destination
NXT WAVE
IB HUBS

Routing Table

Router maintain a routing table which contain the information about where to forward the incoming packet.

Destination	Next Hop
9.1.1.0/24	172.16.5.10
223.1.1.0/24	172.16.5.11
64.0.10.0/24	172.16.5.12
*	172.16.5.10

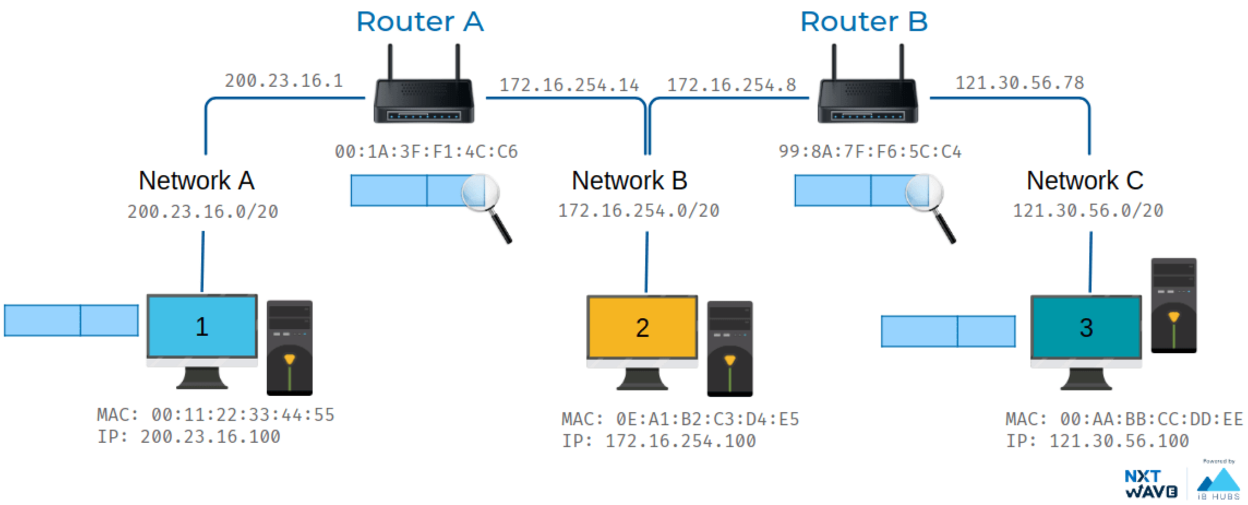
- **Destination Network**: Row for each network that the router knows about. Network ID and Net Mask.
- **Next Hop**: Next router that data should be forwarded.



Example consider three network that data need to be sent data from computer on network A to computer on network c.the third Network C has an address space of 172.16.1.0/23. There is a second router connecting network B and network C. It's interface on network B has an IP of 172.16.254.8 and its interface on Network C has an IP of 121.30.56.78.

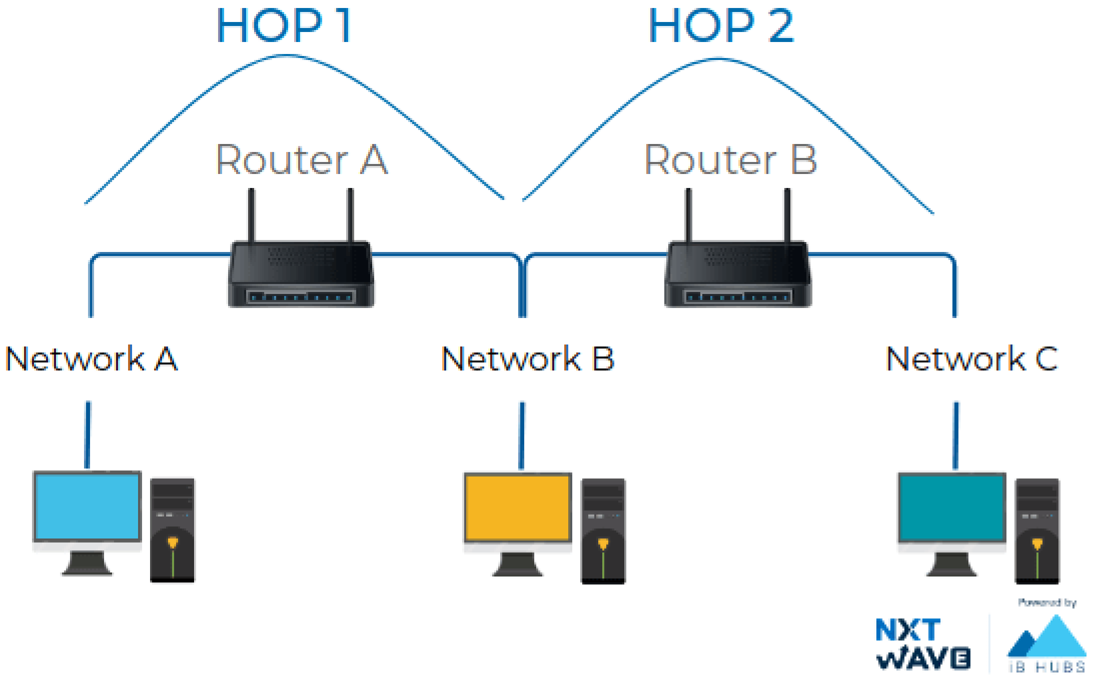
The computer at 200.23.16.100 knows that 121.30.56.100 is not on its local network, so it sends a packet to its gateway, the router between Network A and Network B. Again, the router inspects the content of this packet. It sees a destination address of 121.30.56.100 and through a lookup of its routing table, it knows that the quickest way to get to the 121.30.56.0/20 network is via another router. With an IP of 121.30.56.100 The router decrements the TTL field and sends it along to the router of 172.16.254.8. This router then goes through the motions, knows that the destination IP of 121.30.56.100 is directly connected and forwards the packet to its final destination. That's the basics of routing. The only difference between our examples and how things work on the Internet is scale. Routers are usually connected to many more than just two networks. Very often, your traffic may have to cross a dozen routers before it reaches its final destination. And finally, in order to protect against breakages, core

Internet routers are typically connected in a mesh, meaning that there might be many different paths for a packet to take.



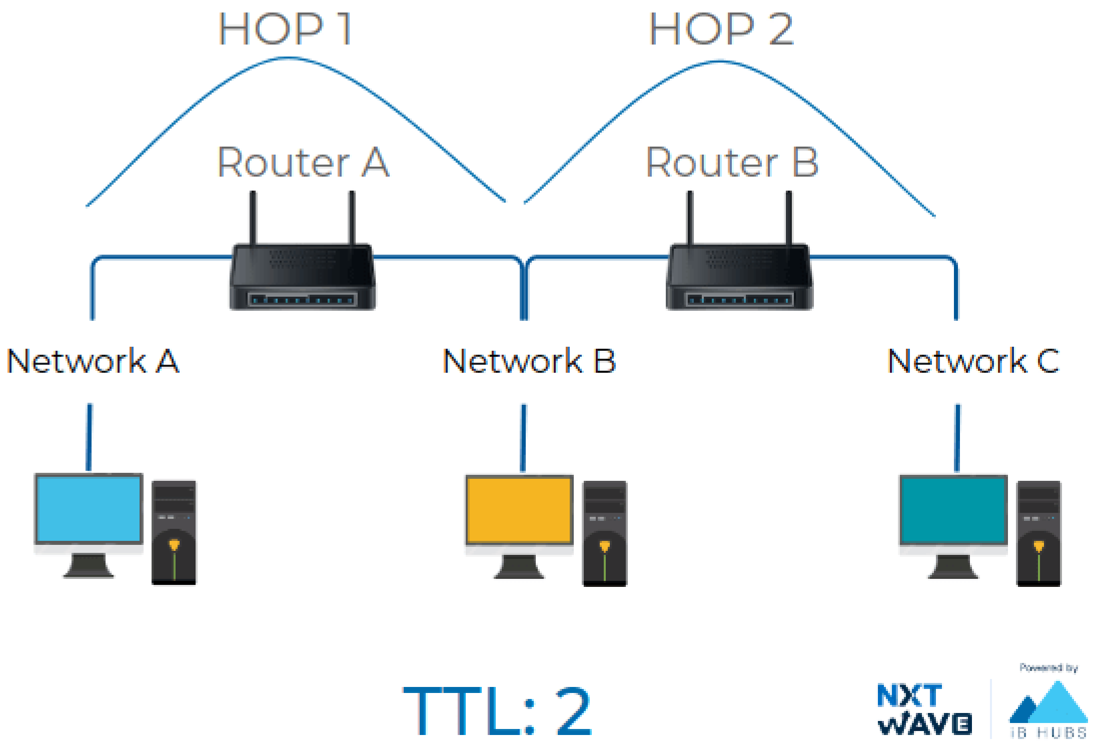
Hop

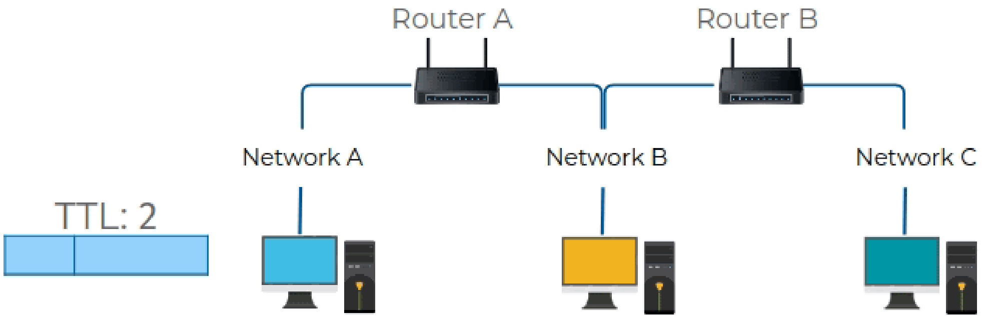
A hop occurs when a packet is passed from one network segment to the next.



Time to Live

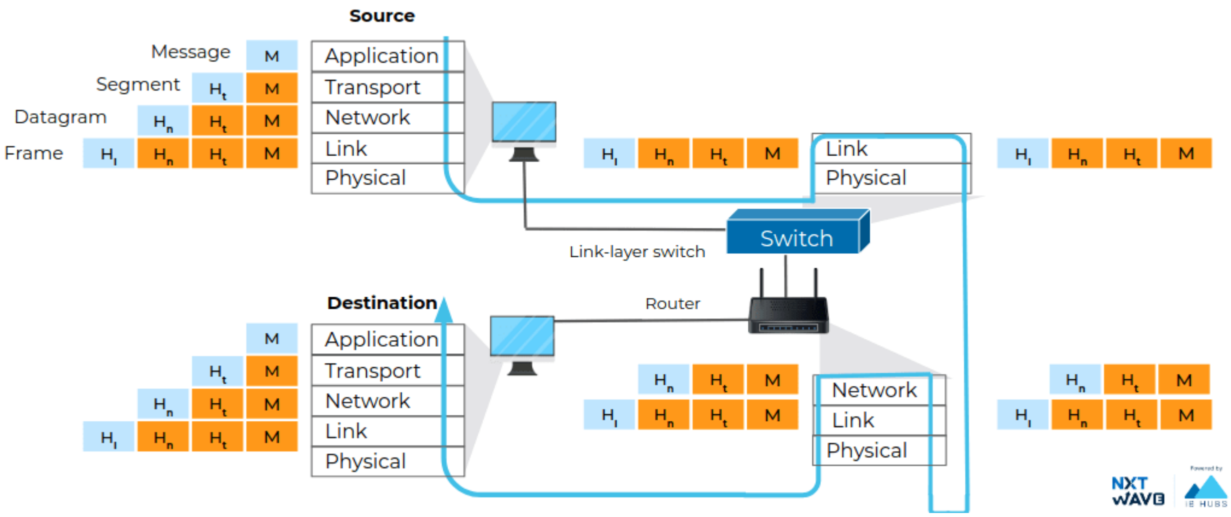
The Time to Live or TTL indicates how many router hops a datagram can traverse before it reaches destination.





Transferring Data Over Network

The header (and footer) and the data together form the PDU for the next layer. The process continues until reaching the lowest-level layer (physical layer or network access layer), from which the data is transmitted to the receiving device. The receiving device reverses the process, de-encapsulating the data at each layer with the header and footer information directing the operations. Then the application finally uses the data. The process is continued until all data is transmitted and received.



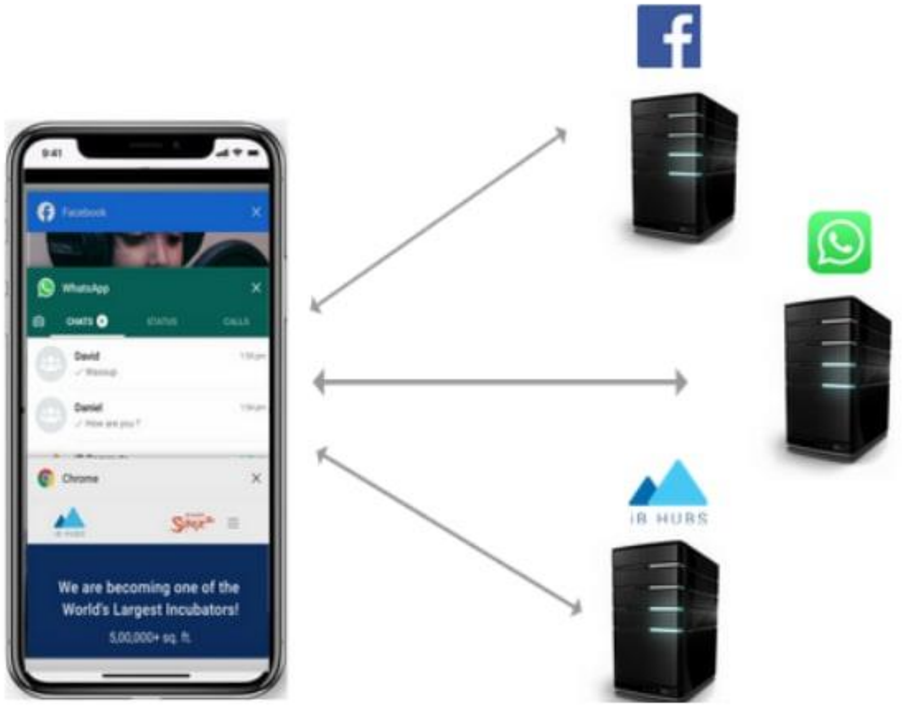
Transport Layer

Processes

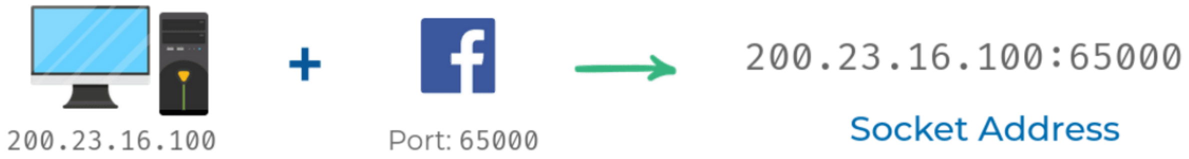
You might run an email program and a web browser, both client applications, on your PC at the same time, and your email and web server might both run on the same server. Even so, emails end up in your email application and web pages end up in your web browser.

Port Numbers

A port is a 16-bit number that's used to direct traffic to specific services running on a networked computer.



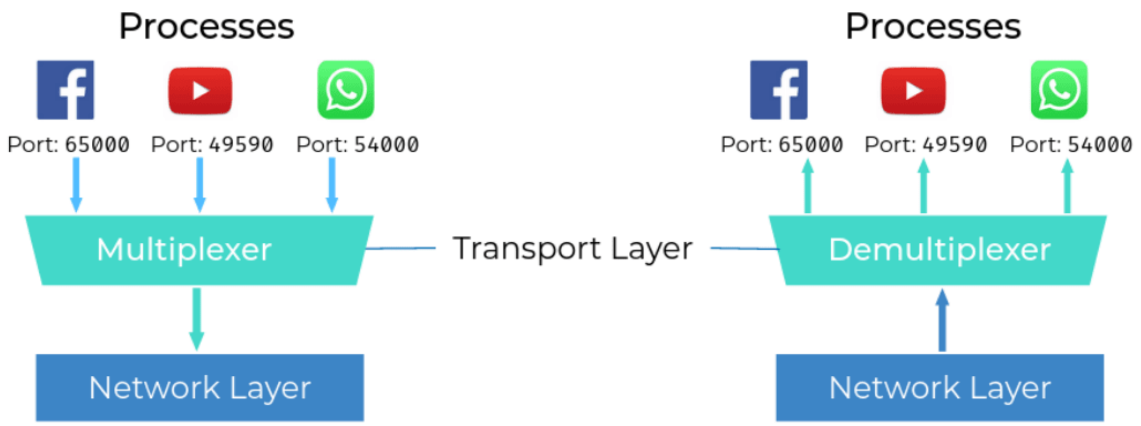
Socket Address



IP + Port = Socket Address

Multiplexing and Demultiplexing

- **Multiplexing** in the transport layer means that nodes on the network have the ability to direct traffic toward many different receiving services.
- **Demultiplexing** is the same concept, just at the receiving end, it's taking traffic that's all aimed at the same node and delivering it to the proper receiving service.



Headers & Payload

At every layer, the data packets consist

- Header
  - Header of data link layer frame contains MAC address
  - Header of Network layer datagram contains IP address
  - Header of Transport layer Segment contains Port number
- Payload

Protocols

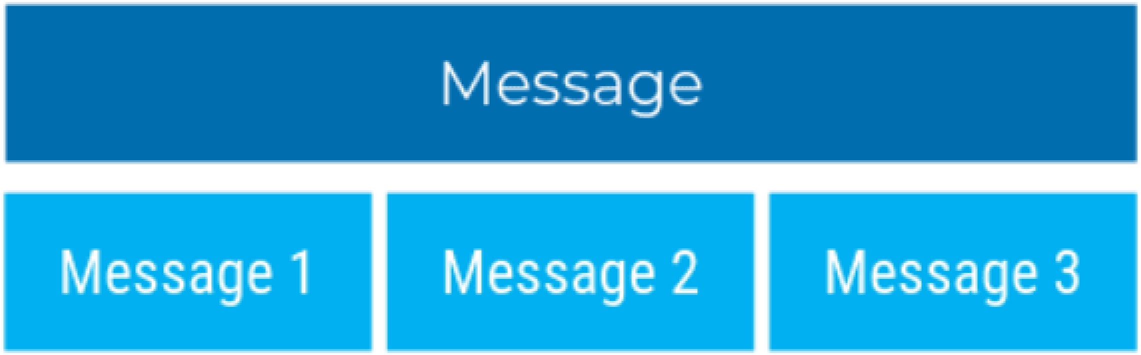
Protocols used in transport layer:

- Transmission Control Protocol (TCP)
- User Datagram Protocol (UDP)

Packets

- At each layer, a larger message is broken down to smaller messages.
- Sending smaller messages is more efficient than sending a large message over a network.

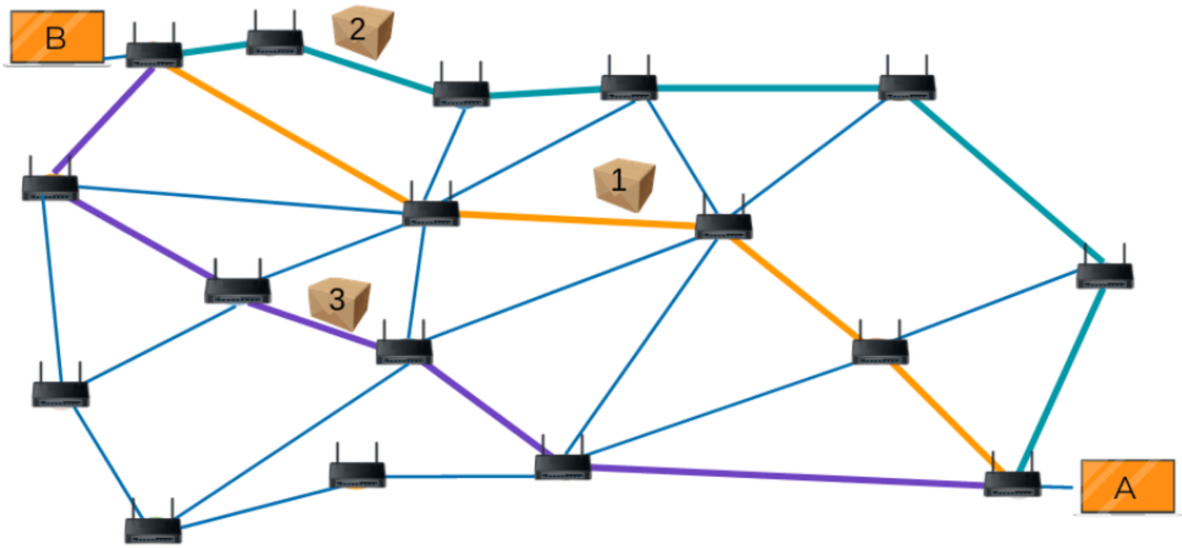
- The sending end system breaks the data into smaller parts known as packets



Route

ANALOGY: Transportation Network

- Buildings, Roads, Highways, Intersections, Trucks ..
- The sequence of communication links and packet switches traversed by a packet from the sending end system to the receiving end system is known as a route or path
- The routers are constantly trying to balance the load across whichever routes they know to ensure speedy and reliable delivery which is conjunction control



Sequence Numbers

TCP ensure the packets reassembled in right order.

- A 32-bit sequence number is used to keep track of the order of segments.

Packet Loss

Data Packets can be lost in transit over the network.

- TTL Expired
- Network Congestion
- Network Hardware Failures

Transmission Control Protocol

TCP provides mechanisms to ensure that data is reliably delivered.

Acknowledgement

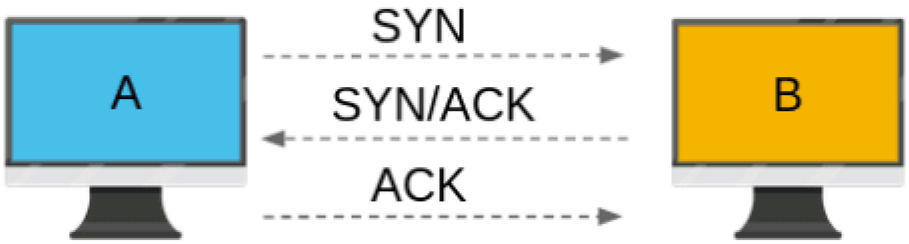
- TCP protocol ensure reliable data transfer by sending acknowledgement for each message received
- The acknowledgment includes the sequence number of the next expected segment.

TCP Connection

- SYN (Synchronize): Used to initiate and establish a connection. It also helps you to synchronize sequence numbers between devices.
- FIN (Finish): Used to terminate a connection when the transmitting computer doesn't have any more data to send.

Establishing TCP Connection: Three - Way HandShake

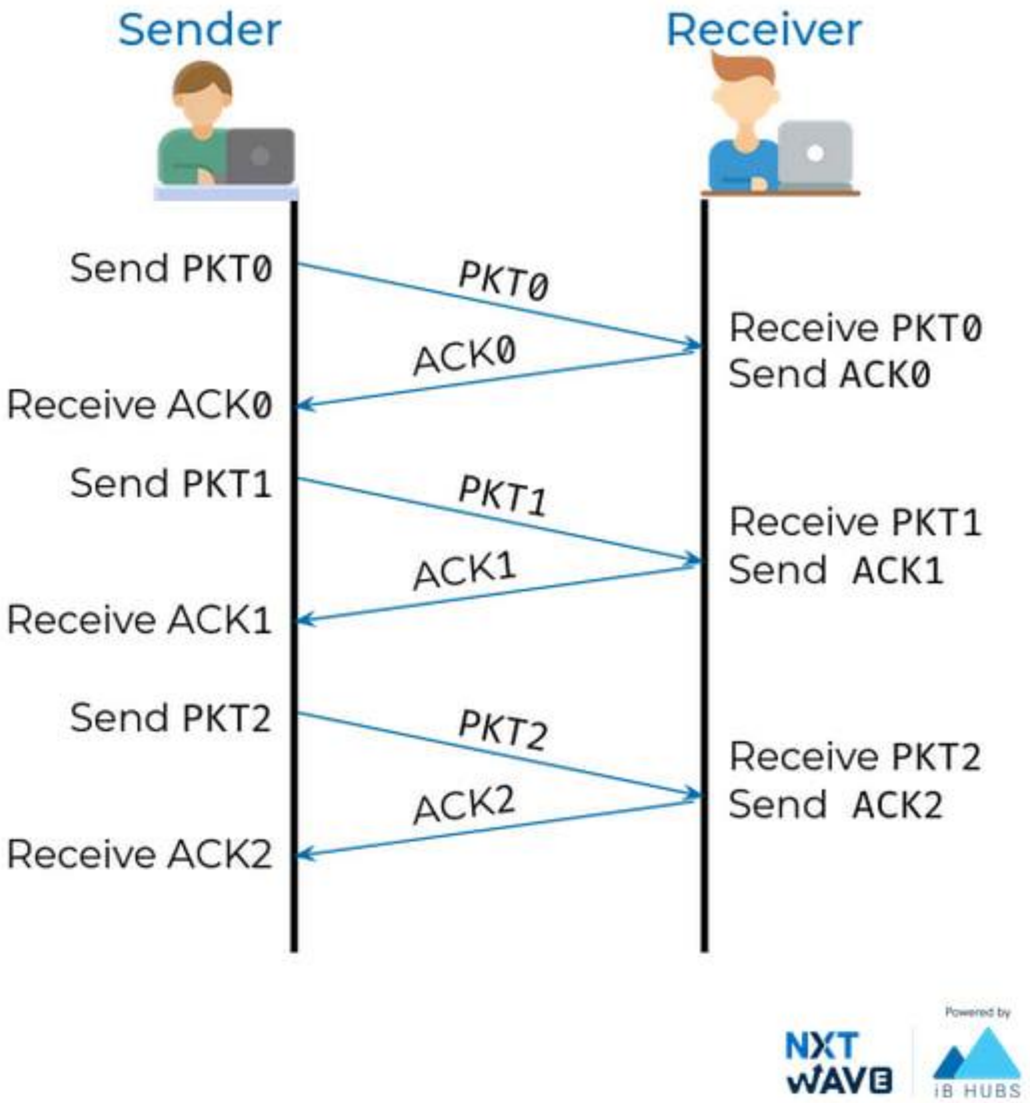
- A handshake is a way for two devices to ensure that they're speaking the same protocol and will be able to understand each other.
- After completion of three-way handshake actual data transfer starts.



Data Transmission

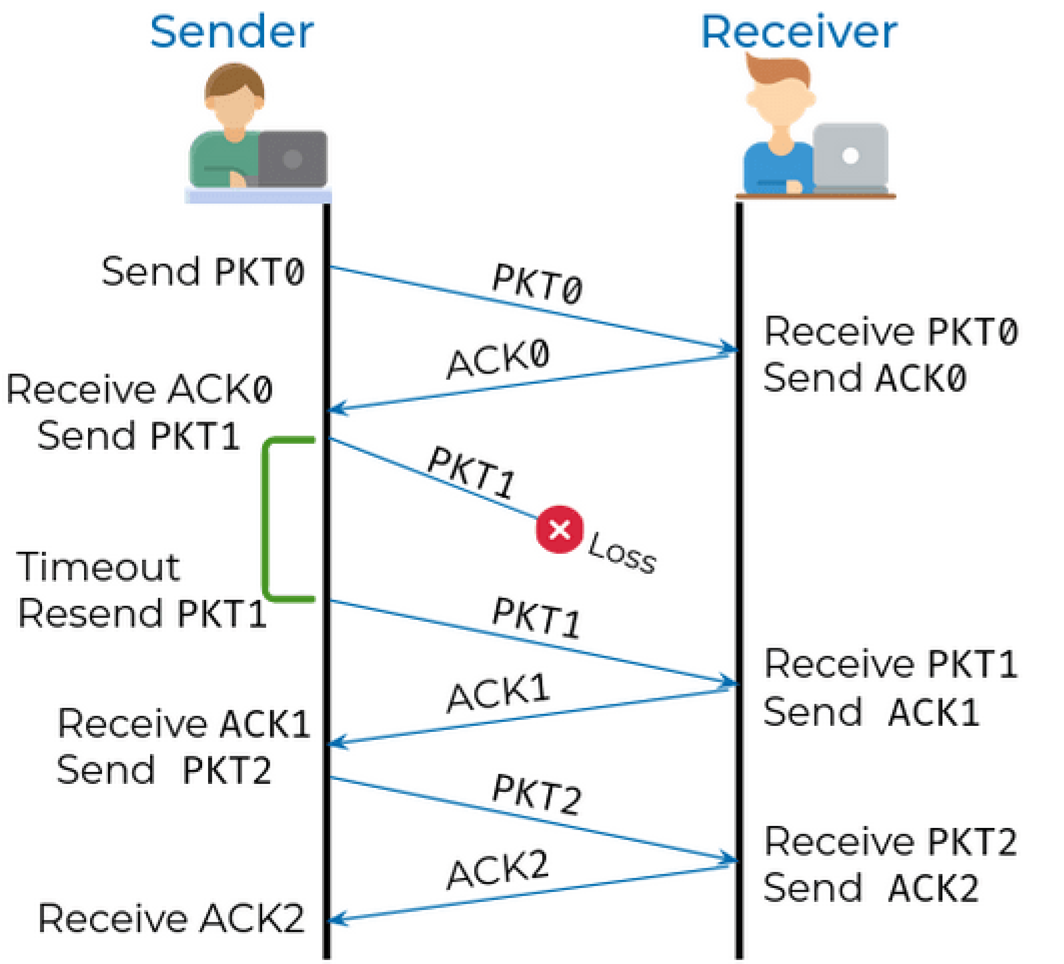
Operation With No Loss

- Once the three-way handshake is complete, the application data can begin to flow between the client and the server.
- The client can send a data packet immediately after the ACK packet
- When a packet of data is sent over TCP, the recipient must always acknowledge what they received.
- The first computer sends a packet with data and a sequence number.
- The second computer acknowledges it by setting the ACK bit and increasing the acknowledgement number by the length of the received data.
- Those two numbers help the computers to keep track of which data was successfully received, which data was lost, and which data was accidentally sent twice.

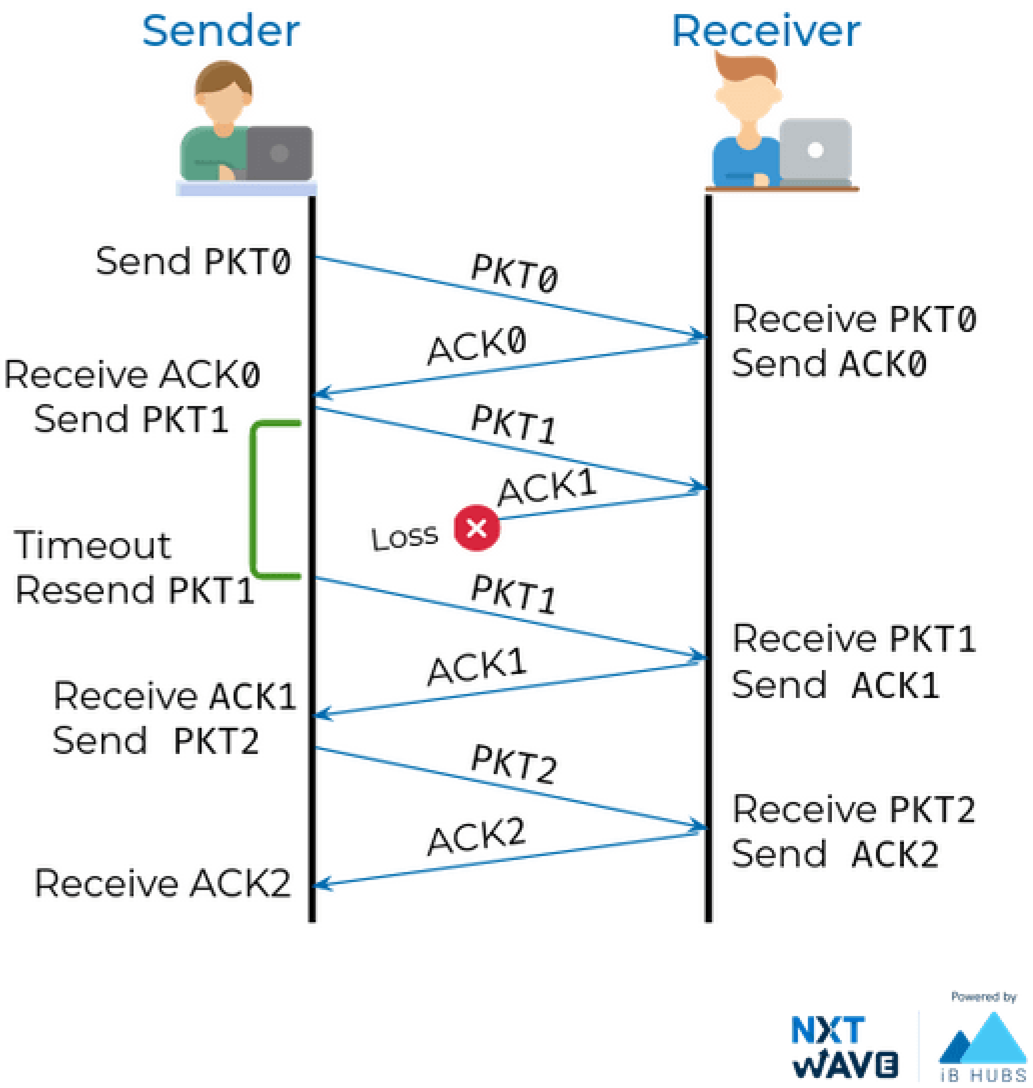


Lost Packet

- TCP connections can detect lost packets using a timeout.
- After sending off a packet, the sender starts a timer and puts the packet in a retransmission queue.
- If the timer runs out and the sender has not yet received an ACK from the recipient, it sends the packet again.

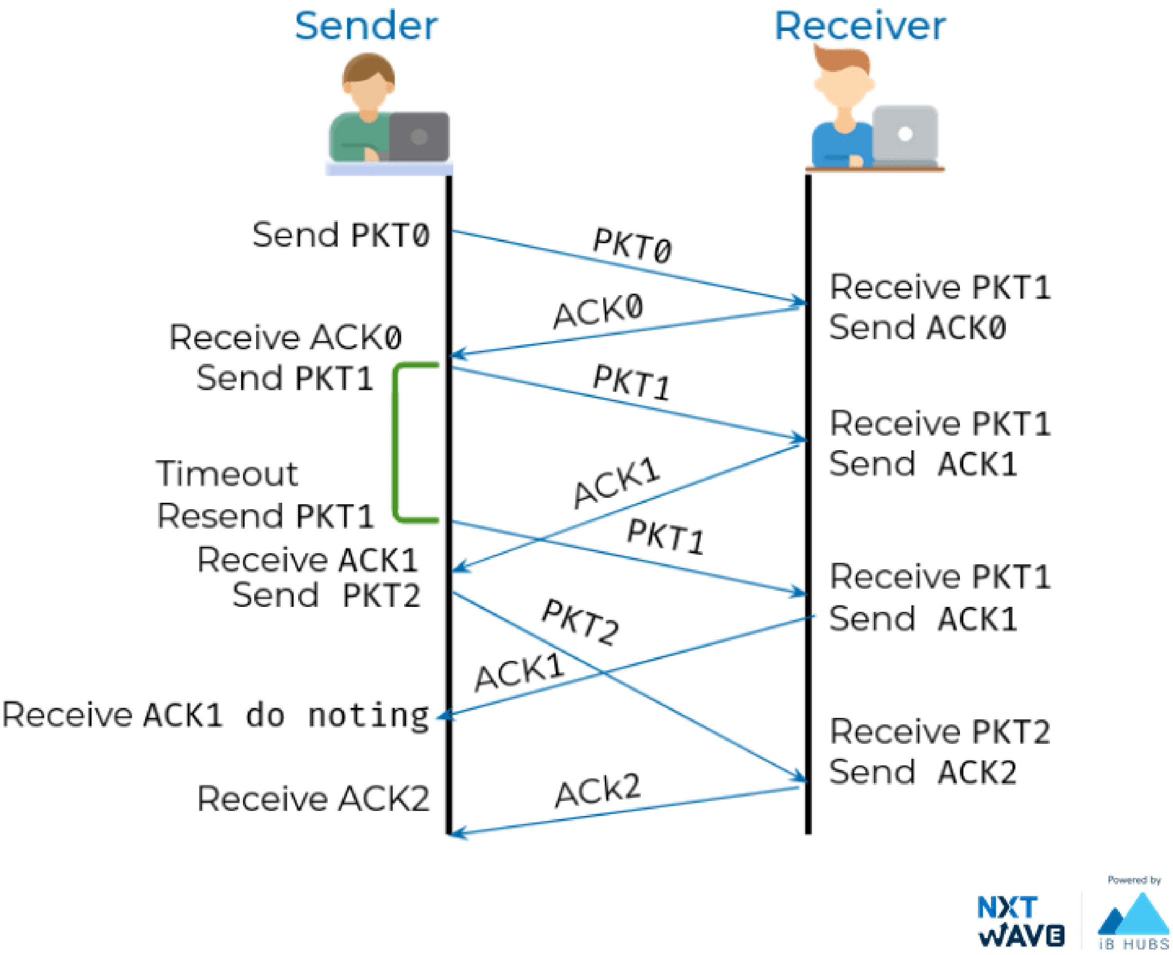

Sender
Receiver
Send PKT0
PKT0
Receive PKT0
Send ACK0
ACK0
Receive ACK0
Send PKT1
PKT1
Loss
Timeout
Resend PKT1
PKT1
Receive PKT1
Send ACK1
ACK1
Receive ACK1
Send PKT2
PKT2
Receive PKT2
Send ACK2
ACK2
Receive ACK2

Lost ACK



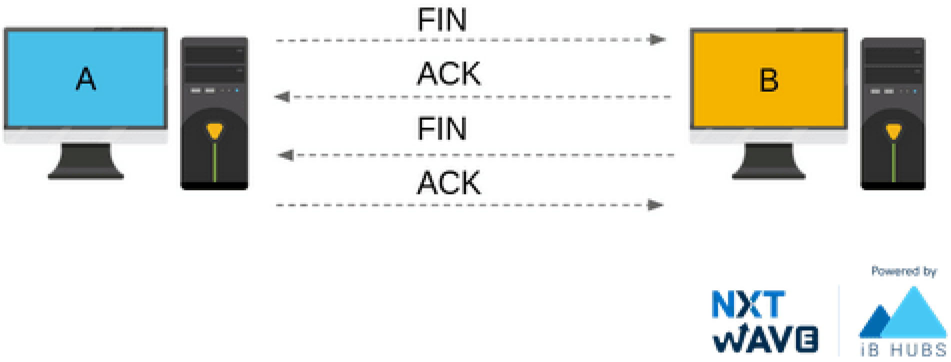
Premature timeout

- The retransmission may lead to the recipient receiving duplicate packets, if a packet was not actually lost but just very slow to arrive or be acknowledged.
- If so, AS recipient has sequence it can simply discard duplicate packets.
- It's better to have the data twice than not at all!



Terminating TCP Connection: Four - Way Handshake

To Close the Connection four way handshake happens



Internet traffic

TCP - Congestion Avoidance

Two phases in TCP Congestion Control Algorithms

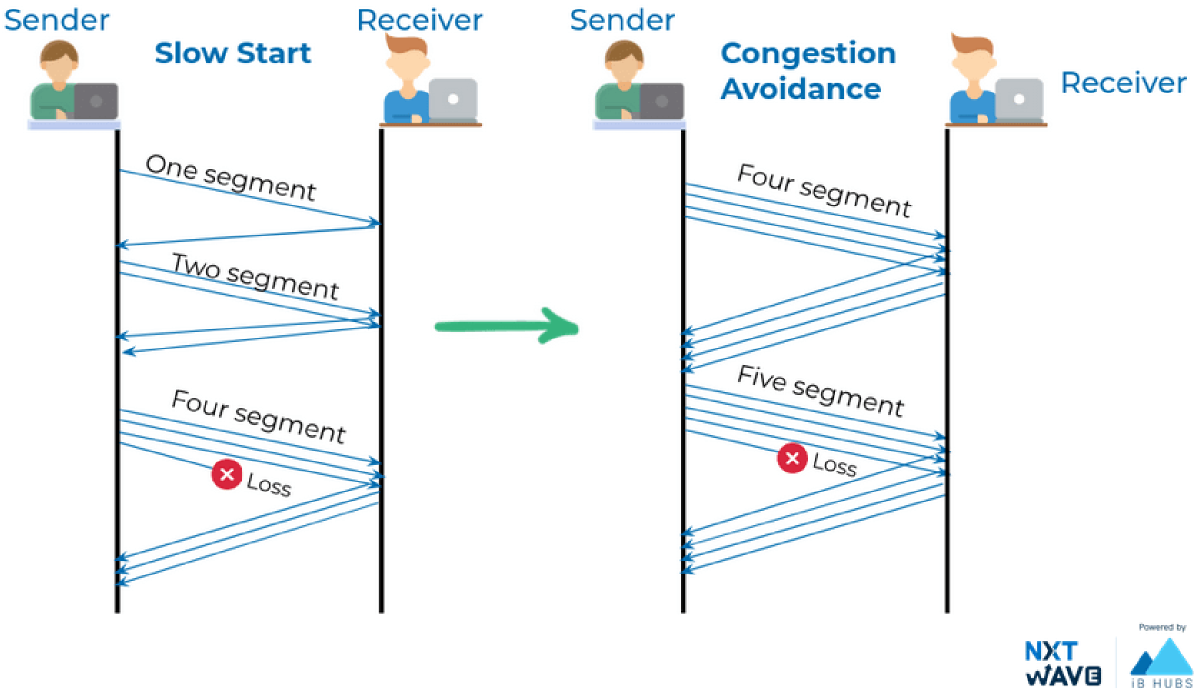
1. Slow Start
2. Congestion Avoidance

Slow Start Phase: When connection begins increases rate exponentially(1, 2, 4, 8, 16…) until first packet loss. due to overload or congestion that packets may be dropped. Packets lost will trigger a timeout at the sender. When this happens, the sender goes into congestion avoidance phase

Example : first 1 packet is sent and got 1 ack , next doubles 2 packets is sent and after successful two acks it now send four packets the process goes on until the packet got lost.

Congestion Avoidance Phase: In the Congestion Avoidance. It increase rate linearly. If congestion was indicated by a timeout, the congestion window is reset to one segment, which automatically puts the sender into Slow Start mode.

Example: the packet was lost at 4 segments in slow start so, in congestion avoidance phase it send again 4 segments after successful acks it increment by one and so the next it sends 5 segments . if the packet was lost. It reset with one segment it automatically puts the sender into Slow Start mode.



Transmission Control Protocol

- Establish the connection
- Send constant stream of acknowledgments
- Tear the connection down at the end

TCP vs UDP

- Let's imagine that each UDP datagram is a single frame of a video. For the best viewing experience, you might hope that every single frame makes it to the viewer, but it doesn't really matter if a few get lost along the way. A video will still be pretty watchable unless it's missing a lot of its frames.
- So, in Most video chats like skype and game streaming, video and music streaming, etc. uses UDP. That's if you are on bad internet connection skype get all gitchy. Not all UDP packets are delivered to your system.
- But this approach doesn't work for other application like it doesn't really work if you send an email and shows up with the middle part missing. The whole message really needs to get there correctly.when it absolutely, correctly needs to get there programs uses TCP (Transmission Control Protocol)

User Datagram Protocol

- Referred to as a "null protocol",
- Provides no guarantees about message delivery or notifications of failure
- Less Overhead

IP Addresses

Public & Private IPs

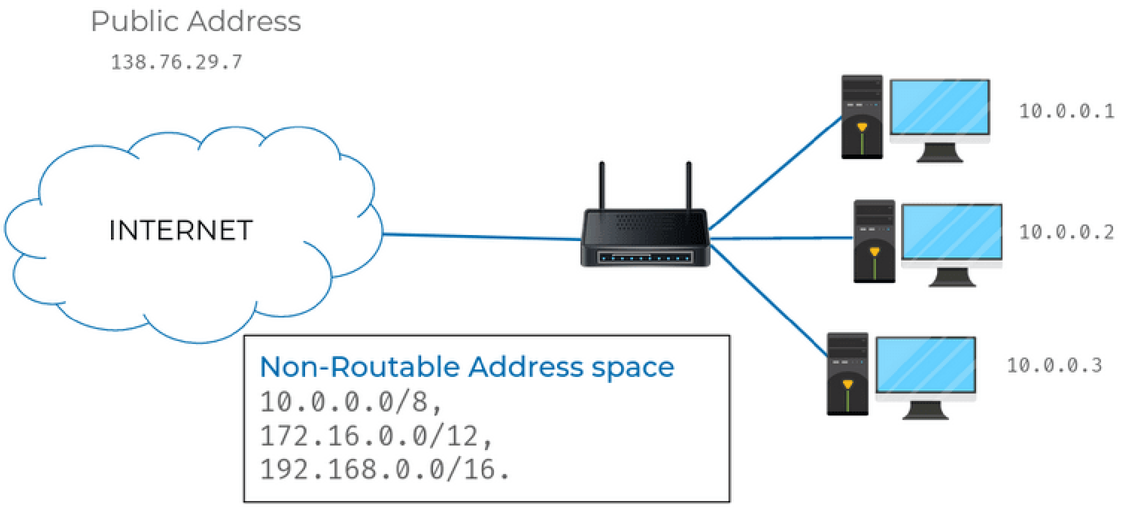
- Internet Assigned Numbers Authority (IANA) assigns different IP ranges to different organizations and Internet Service Providers
- Certain range of IPs are reserved for Private Usage

Private IP Address Ranges	Possible Hosts
10.0.0.0 to 10.255.255.255	16777216
172.16.0.0 to 172.31.255.255	1048576
192.168.0.0 to 192.168.255.255	65536

Network Layer

Non-Routable Address Space

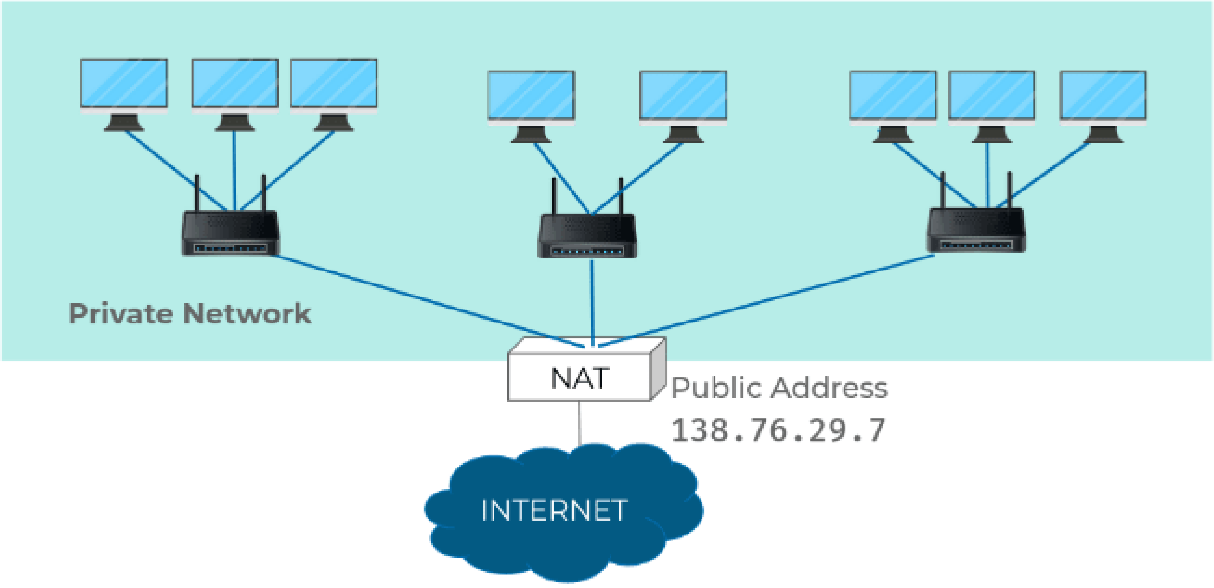
- RFC(Request for Comments) 1918 defined non-routable address space.
- Non-routable address space are ranges of IPs set aside for use by anyone that are not registered in the internet.
- RFC 1918 defined three ranges of IP addresses that will never be registered in the internet anywhere are 10.0.0.0/8, 172.16.0.0/12, and 192.168.0.0/16.
- This IP address are used only internally like home network or business



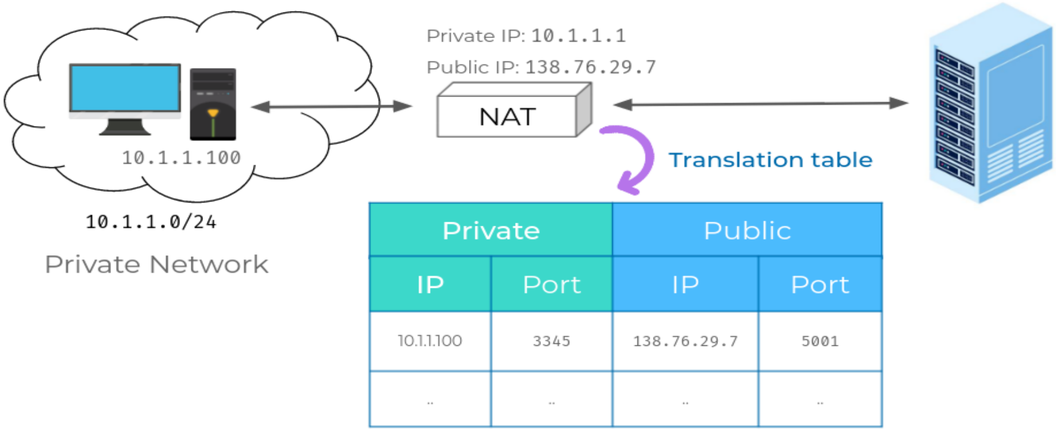
Network Address Translator

- With NAT, you can have hundreds even thousands of machines using non-routable address space.
- Yet, with just a single public IP, all those computers can still send traffic to and receive traffic from the internet.

- All you need is one single IPv4 address and via NAT, a router with that IP can represent lots and lots of computers behind it.



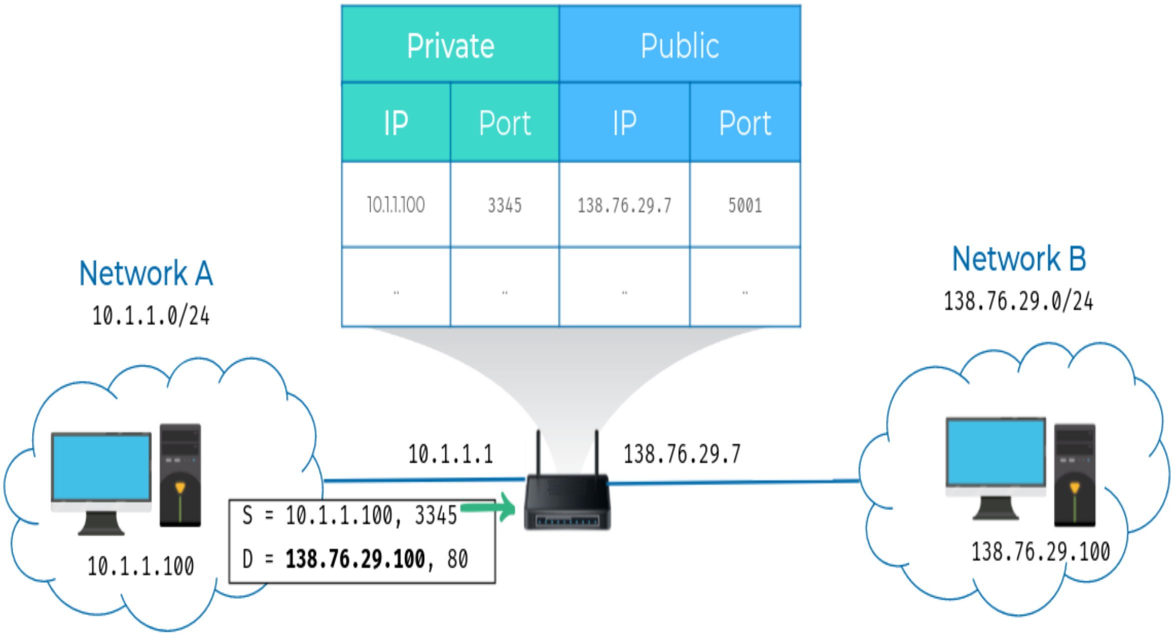
NAT is a technology that allows a gateway, usually a router, to rewrite the source IP of an outgoing IP datagram while retaining the original IP in order to rewrite it into the response.



Private		Public	
IP	Port	IP	Port
10.1.1.100	3345	138.76.29.7	5001
..	..	..	..

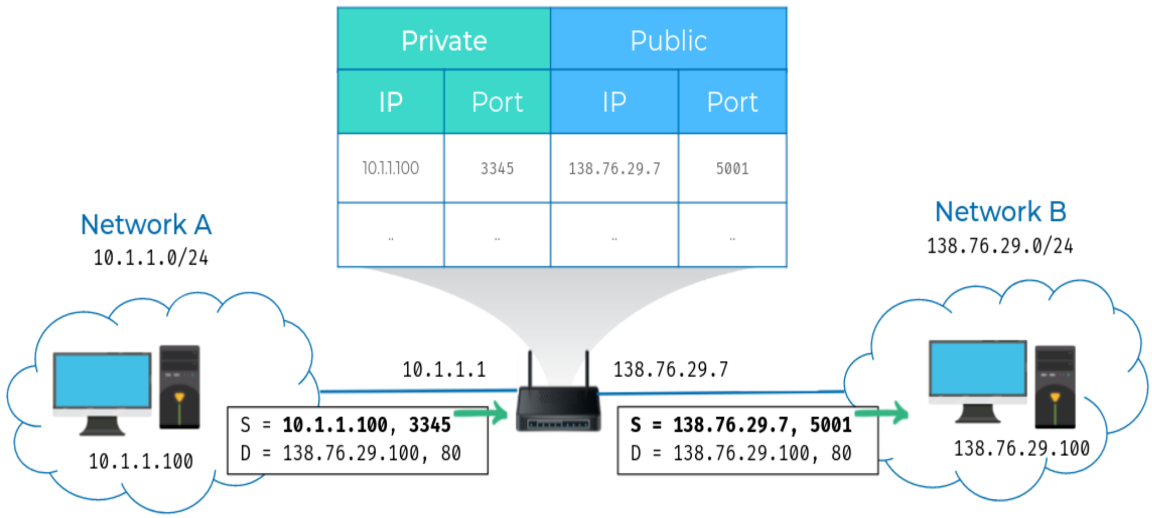
- Here, we have two networks. Network A consists of the 10.1.1.0/24 address space and network B consists of the 138.76.29.0/24 address space. Sitting between these networks is a router that has an interface on network A with an IP of 10.1.1.1 and an interface on network B of 138.76.29.7. Computer 1 is on network A and has an IP of 10.1.1.100. And computer 2 is on network B and has an IP of 138.76.29.100.

- Computer 1 wants to communicate with a web server on computer 2.
- So it crafts the appropriate packet at all layers and sends this to its primary gateway, the router sitting between the two networks.
- So far, this is a lot like many of our earlier examples, but in this instance,



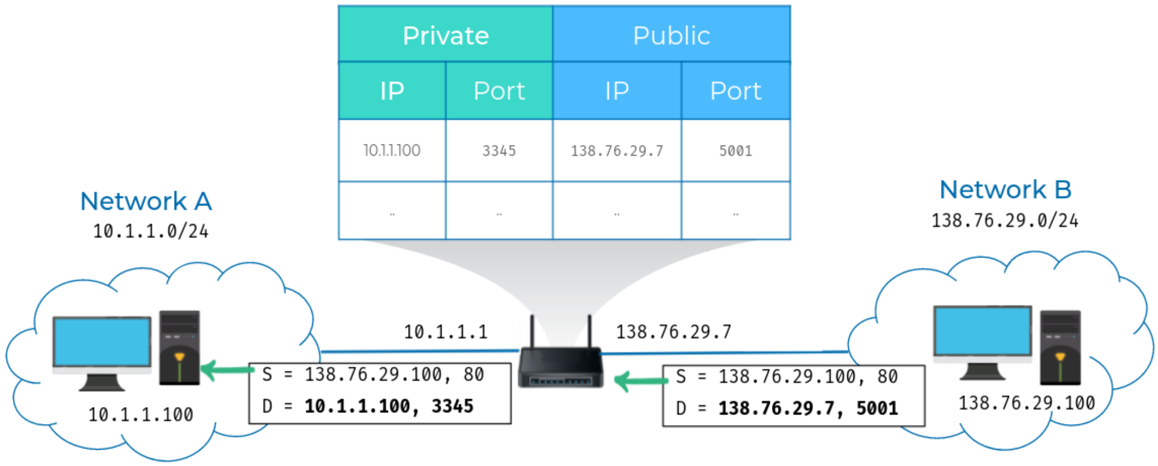
Private		Public	
IP	Port	IP	Port
10.1.1.100	3345	138.76.29.7	5001
..	..	..	..

- The router is configured to perform NAT for any outbound packets.
- Normally, a router will inspect the contents of an IP datagram, decrement the TTL by 1, and forward the rest of the data at the network layer without touching it.
- But with NAT, the router will also rewrite the source IP address, which in this instance, becomes the router's IP on network B or 138.76.29.7. When the datagram gets to computer 2, it'll look like it originated from the router, not from computer 1.



Private		Public	
IP	Port	IP	Port
10.1.1.100	3345	138.76.29.7	5001
..	..	..	..

Now, computer 2 crafts its response and sends it back to the router. The router, knowing that this traffic is actually intended for computer 1, rewrites the destination IP field before forwarding it along.



Private		Public	
IP	Port	IP	Port
10.1.1.100	3345	138.76.29.7	5001
..	..	..	..

Application Layer

Protocols

Various applications communicate using different protocols used in application layer:

- Web Apps: HTTP, HTTPS
- Mails: POP, SMTP, IMAP

- DNS, FTP, SSH, etc.

Layer	Application Layer
Protocols	DNS, HTTP, FTP, SSH, ..
Data Unit	Message

IP Address

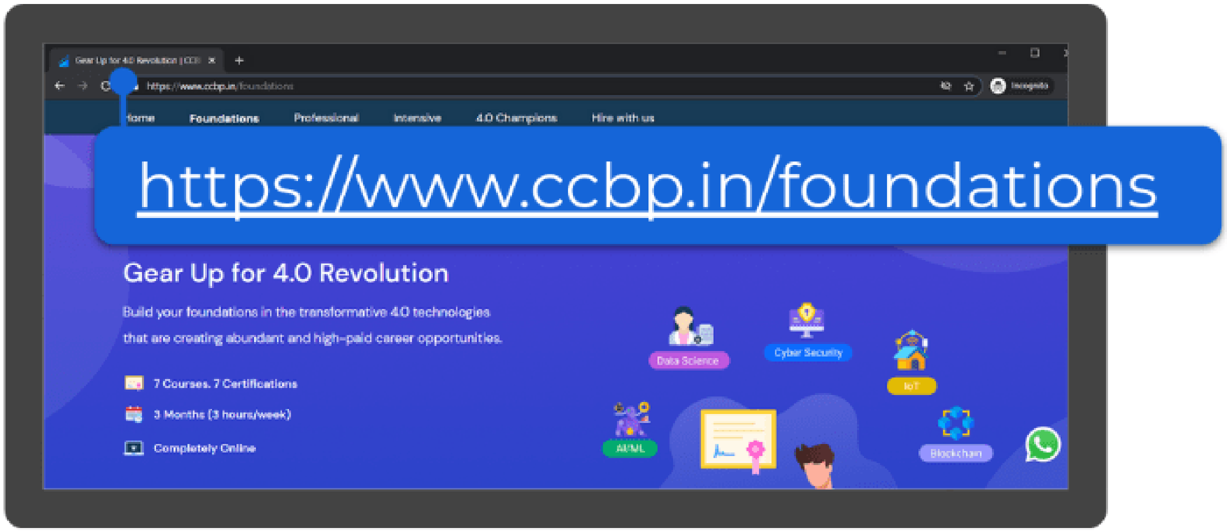
We learned previously that each server has an IP address.

- It is very difficult to memorize the IP address.
- Things get more complicated when, as users, we access various websites and services.
- It becomes quite a challenge when the addresses of those services are dynamic.

Accessing Websites

URL - Uniform Resource Locator

Every object (image, html document) has a unique path (URL) on the web server



https://www.ccbp.in/foundations

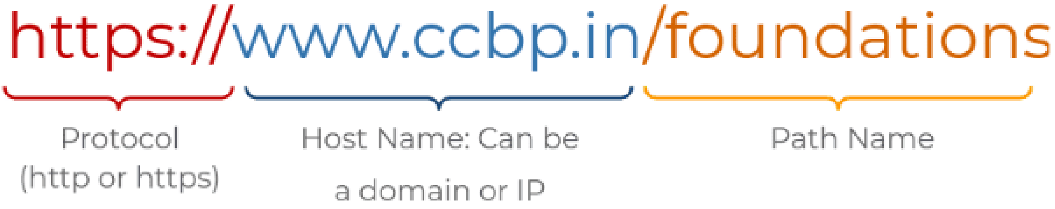
Gear Up for 4.0 Revolution

Build your foundations in the transformative 4.0 technologies that are creating abundant and high-paid career opportunities.

7 Courses, 7 Certifications

3 Months (3 hours/week)

Completely Online



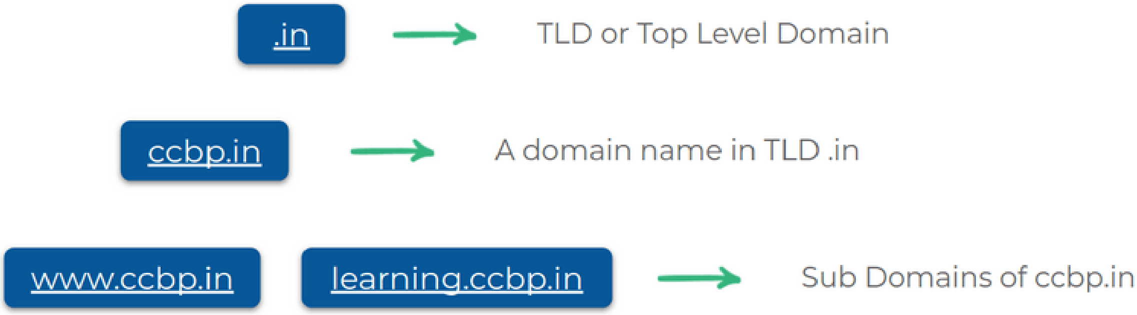
https://www.ccbp.in/foundations

Protocol (http or https)

Host Name: Can be a domain or IP

Path Name

Domain Name



.in → TLD or Top Level Domain

ccbp.in → A domain name in TLD .in

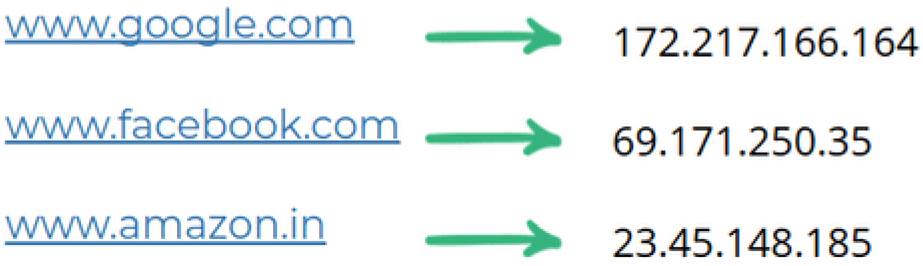
www.ccbp.in learning.ccbp.in → Sub Domains of ccbp.in

Domain Registrars

Domain Name Registrars manage the reservation of internet domain names.

DNS: Domain Name System

- DNS is a global and highly distributed network service.
- DNS is a translation service .
- It resolves strings of letters into IP address.
- IP Addresses of these servers may change, which will appropriately be updated.



www.google.com → 172.217.166.164

www.facebook.com → 69.171.250.35

www.amazon.in → 23.45.148.185

DNS Server

There are five primary types of DNS servers:

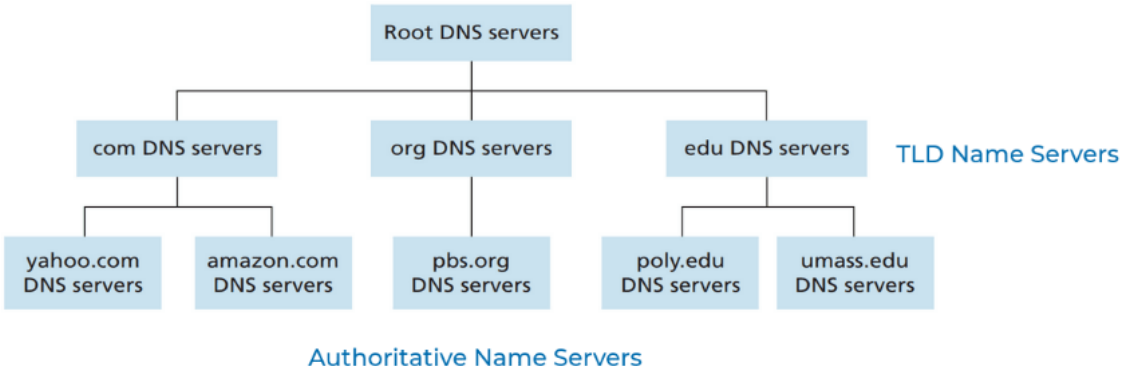
- Root Name Servers.
- TLD Name Servers.
- Authoritative Name Servers.
- Caching Name Servers.
- Recursive Name Servers.

1. **Root Name Servers** redirect to appropriate TLD Name Server.
2. **TLD Name Server** redirect to appropriate Authoritative Name Server for the domain.

3. **Authoritative Name Server** gives the appropriate IP for the domain.
4. **Caching** and **Recursive Name Servers** are provided by ISP. These reduce the load on the root, tld, authoritative name servers.

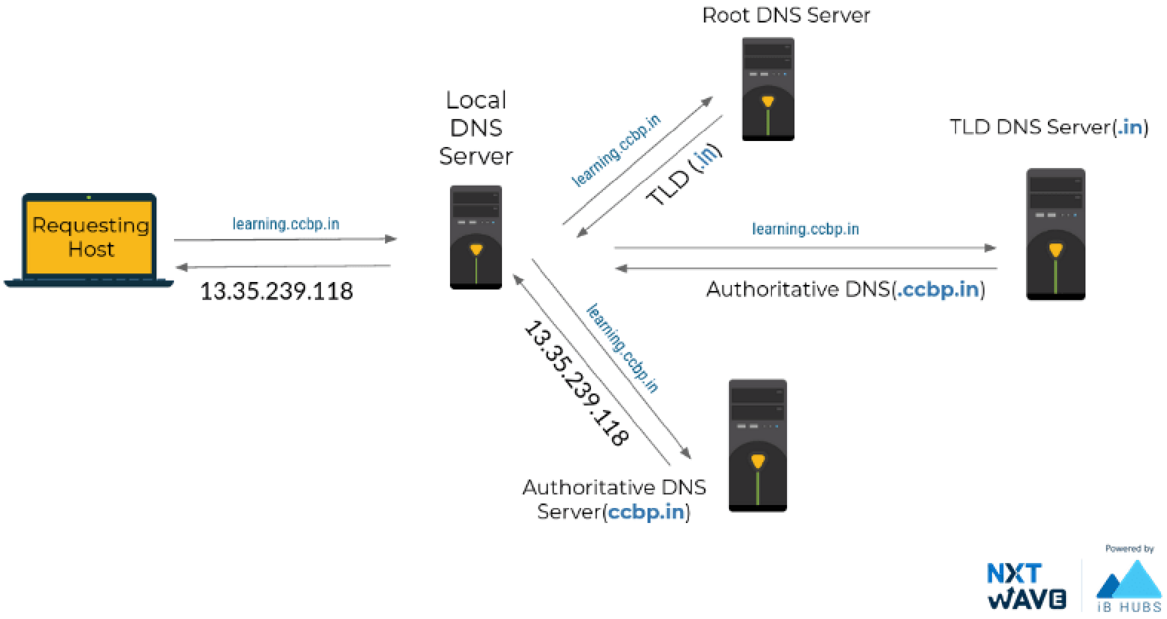
Root Server

- Any domain name registered in the DNS is a domain name.
- Domain names are organized in subordinate levels (subdomains) of the DNS root domain, which is nameless.



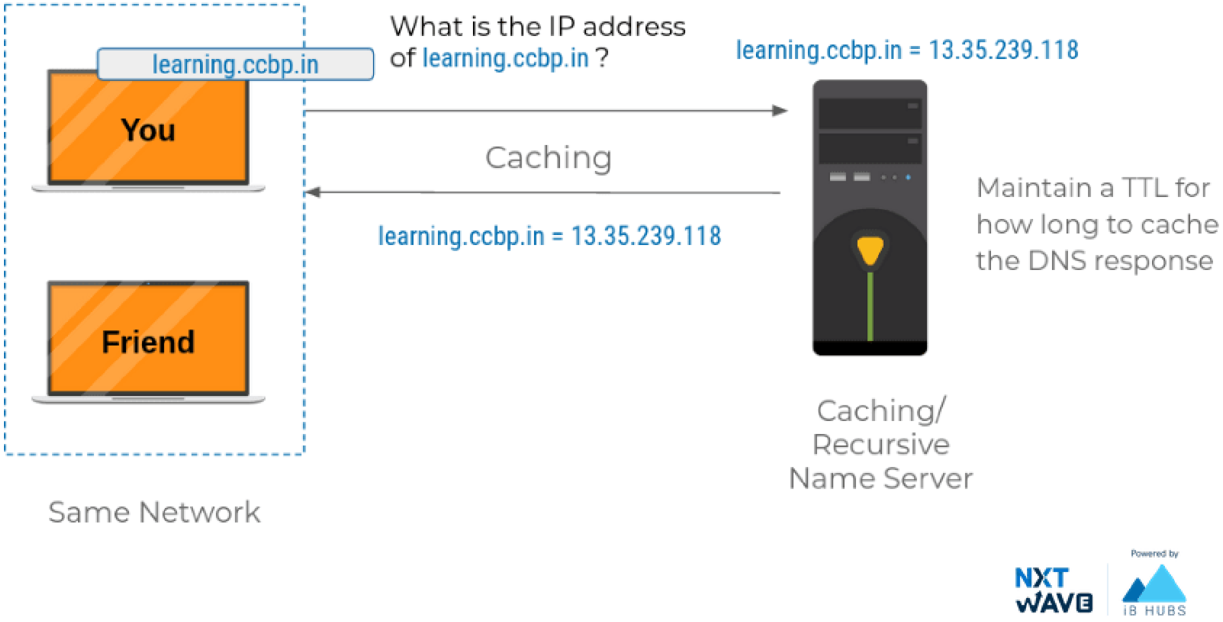
DNS Resolution

- The first step is always to contact a root named server.
- The root servers will respond to a local DNS server with the TLD name server that should be queried like (.in). For each TLD in existence, there is a TLD name server.
- The TLD name servers will respond again to the Local DNS server with what authoritative DNS server(.ccbp.in) to contact.
- The Local DNS server could be redirected at the authoritative server for ccbp.in which would finally provide the actual IP of the server in question.

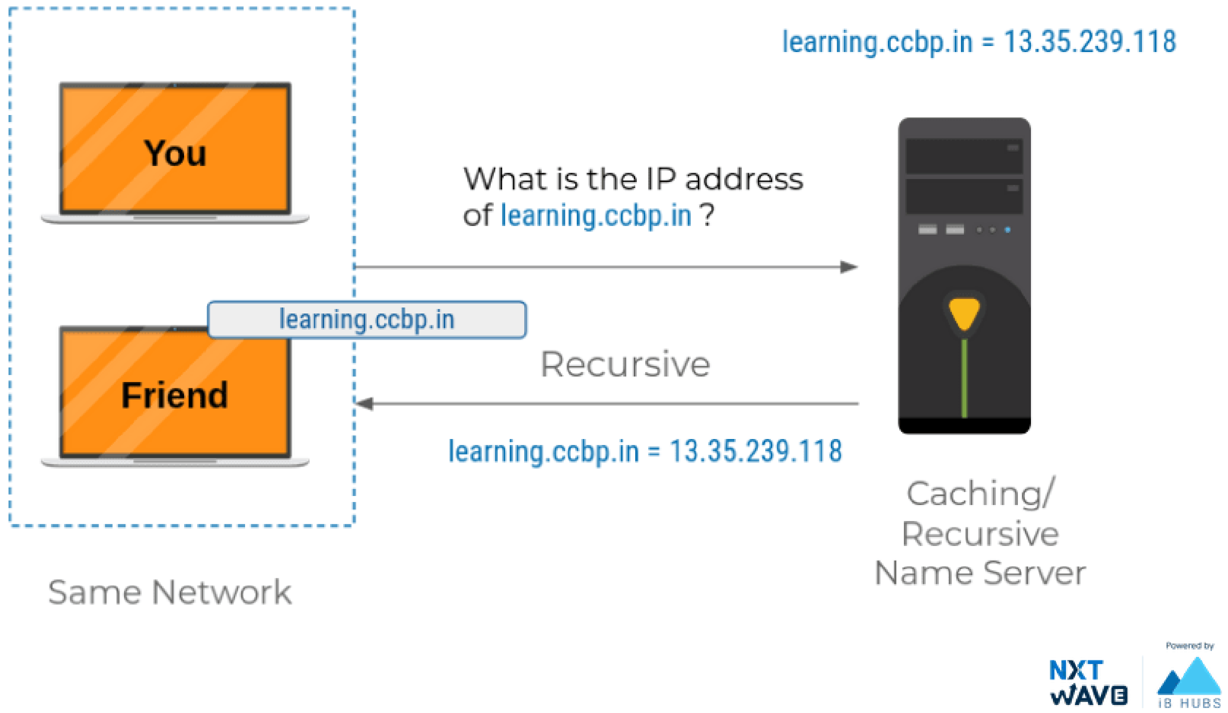


Recursive & Caching Name Servers

- Recursive name servers are ones that perform full DNS resolution requests.
- Performs a fully recursive resolution to discover the correct IP for www.facebook.com. This involves a bunch of steps we'll cover in just a moment.


learning.ccbp.in
You
Friend
Same Network
What is the IP address of learning.ccbp.in ?
learning.ccbp.in = 13.35.239.118
Caching
learning.ccbp.in = 13.35.239.118
Maintain a TTL for how long to cache the DNS response
Caching/ Recursive Name Server
NXT WAVE
Powered by iB HUBS

Since the domain name learning.ccbp.in had just been looked up, the local name server still has the IP that it resolved to stored and is able to deliver that back to your computer without having to perform a full lookup.


learning.ccbp.in = 13.35.239.118
You
What is the IP address of learning.ccbp.in ?
learning.ccbp.in
Friend
Recursive
learning.ccbp.in = 13.35.239.118
Caching/ Recursive Name Server
Same Network
NXT WAVE
Powered by
iB HUBS

DNS

- DNS uses UDP for the transport layer instead of TCP.
- UDP is connectionless.
- A single DNS request and its response can usually fit inside of a single UDP datagram.

DNS Record Types

- DNS servers create a DNS record.
- It provides important information about a domain or hostname, particularly its current IP address.
- DNS record type allows for different kinds of DNS resolutions to take place.

Record Type	Abbreviation	Description
A	Address Mapping record	A record is used to point a certain domain name at a certain IPv4 IP address.
AAAA	IP Version 6 Address record	A record is used to point a certain domain name at a certain IPv6 IP address.
CNAME	Canonical Name record	A CNAME record is used to redirect traffic from one domain to another

Record Type	Abbreviation	Description
MX	Mail exchanger record	Mx record is used in order to deliver email to the correct server.
NS	Name Server records	It allows you to delegate the DNS of one of your subdomains to a different nameserver.
TXT	Text Record	Typically carries machine-readable data

Note :A single A record is configured for a single domain name. But, a single domain name can have multiple A records, too.

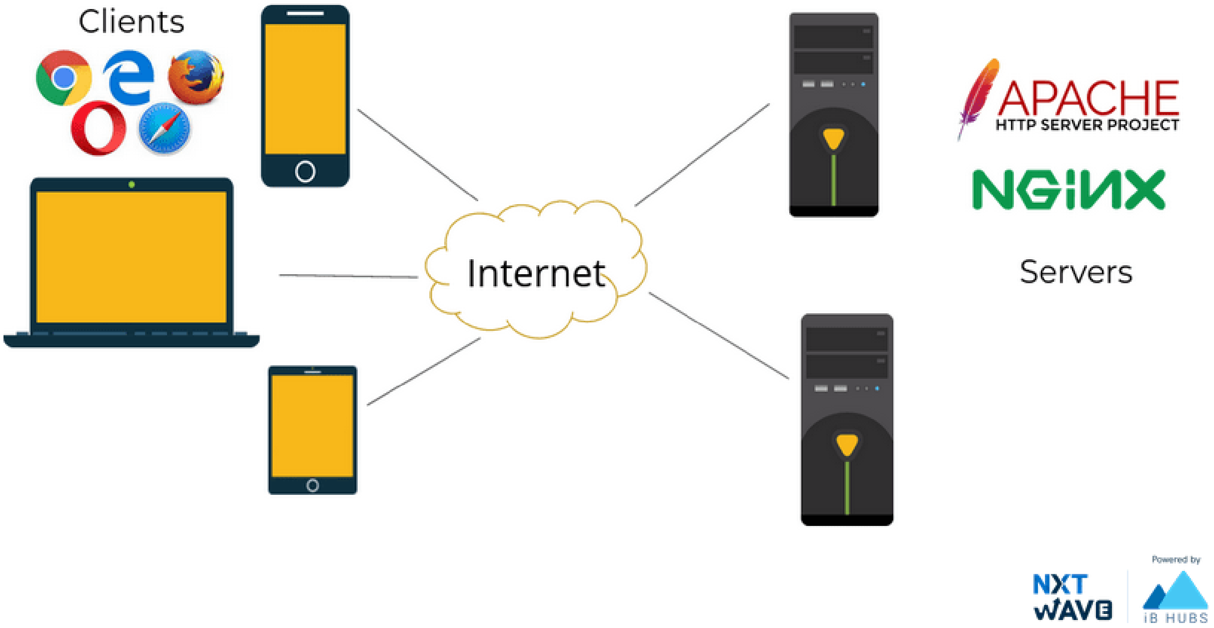
Web

Web is a "Client-Server Application".

- A standard for document formats (HTML).
- Web Browsers.
- Web Servers.
- Relies on HTTP, HTTPS application-layer protocols which rely on TCP.

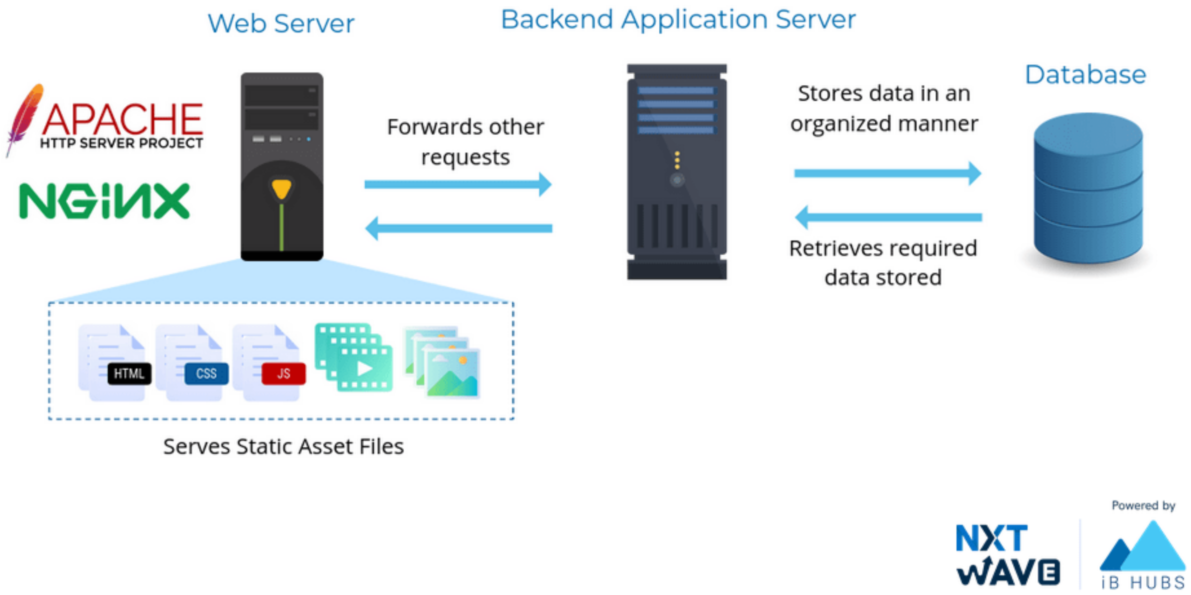
Client-Server Architecture

- In a client-server architecture, there is an always-on host, called the server, which services requests from many other hosts, called clients.
- A classic example is the “Web application” for which an always-on Web server services requests from browsers running on client hosts.



Web Servers

- They host Web objects, each addressable by a URL.
- A Web server is always ON, with a fixed IP address.
- 80 is default port used by a HTTP web server.


Web Server
Backend Application Server
Database
APACHE
HTTP SERVER PROJECT
NGINX
Forwards other requests
Stores data in an organized manner
Retrieves required data stored
HTML
CSS
JS
Serves Static Asset Files
Powered by
NXT WAVE
iB HUBS

HTTP

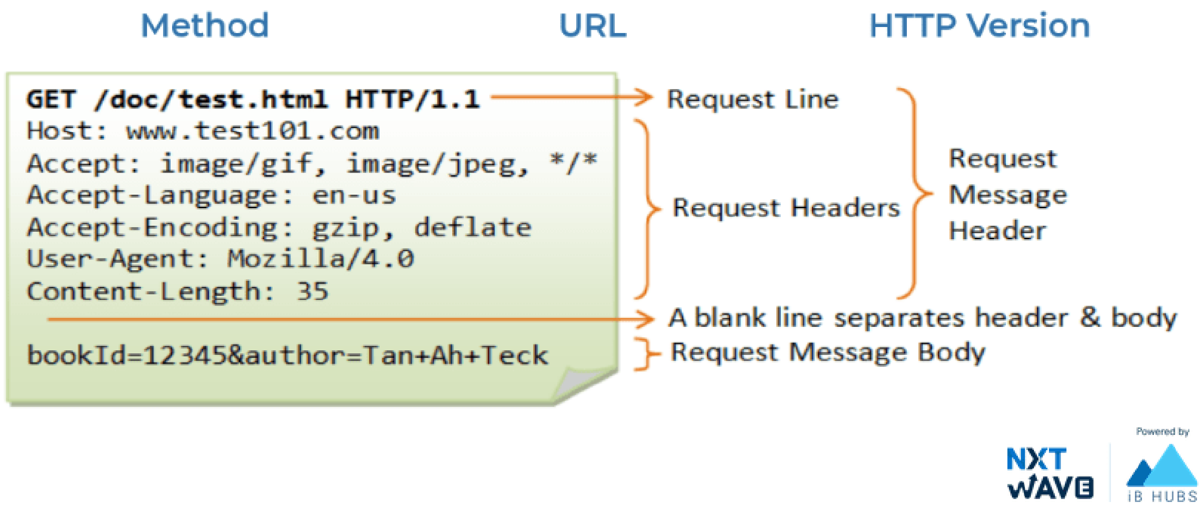
Request & Response

- The messages sent by the client usually a Web browser, are called requests.
- The messages sent by the server as an answer are called responses.

Request

- A HTTP request consists of a Request Header and a Request Body.
- A Request Header contains various info like:
    - Host : (Example: developer.mozilla.org)
    - Method : (Example: GET).
    - Path : (Example: /).
    - HTTP version : (Example: HTTP/1.1).

HTTP Message Format - Request



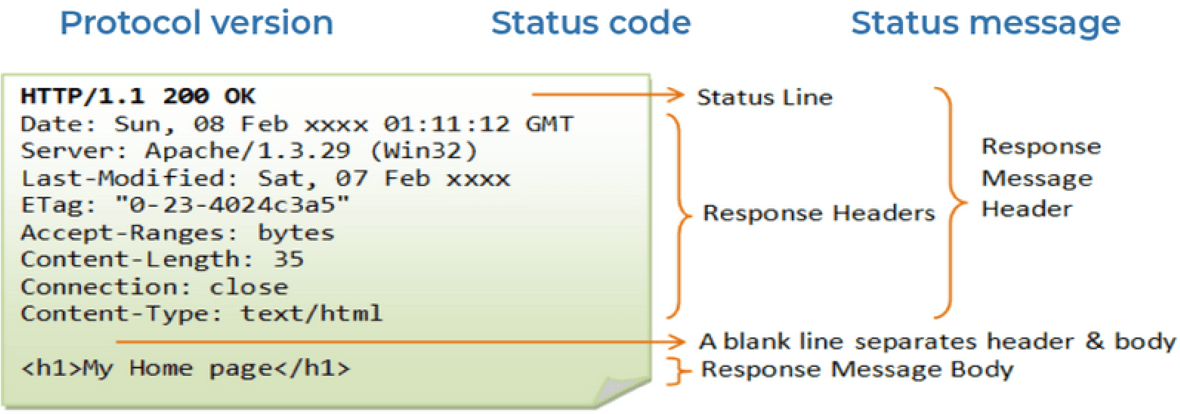
```
GET /doc/test.html HTTP/1.1
Host: www.test101.com
Accept: image/gif, image/jpeg, */*
Accept-Language: en-us
Accept-Encoding: gzip, deflate
User-Agent: Mozilla/4.0
Content-Length: 35

bookId=12345&author=Tan+Ah+Teck
```

HTTP Method

- GET : retrieves data, like a blog article.
- POST : creates data, like a new blog article.
- PUT : replaces data, like an existing blog article.
- DELETE : deletes data, like an existing blog article.

Response

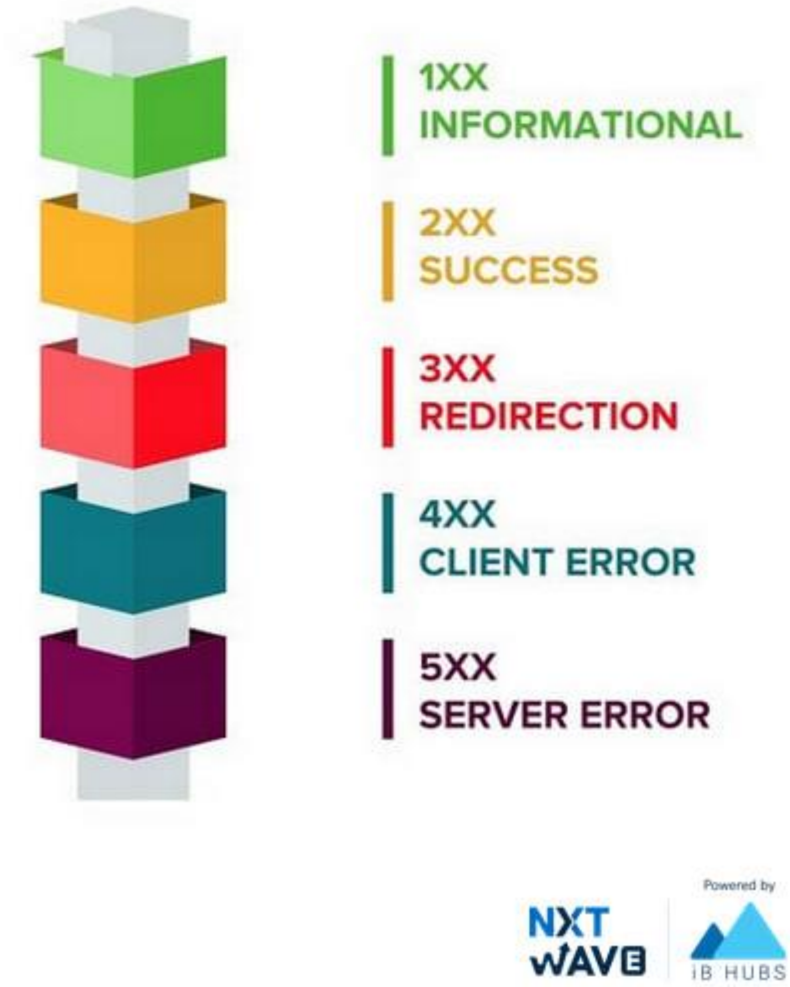
HTTP Message Format - Response



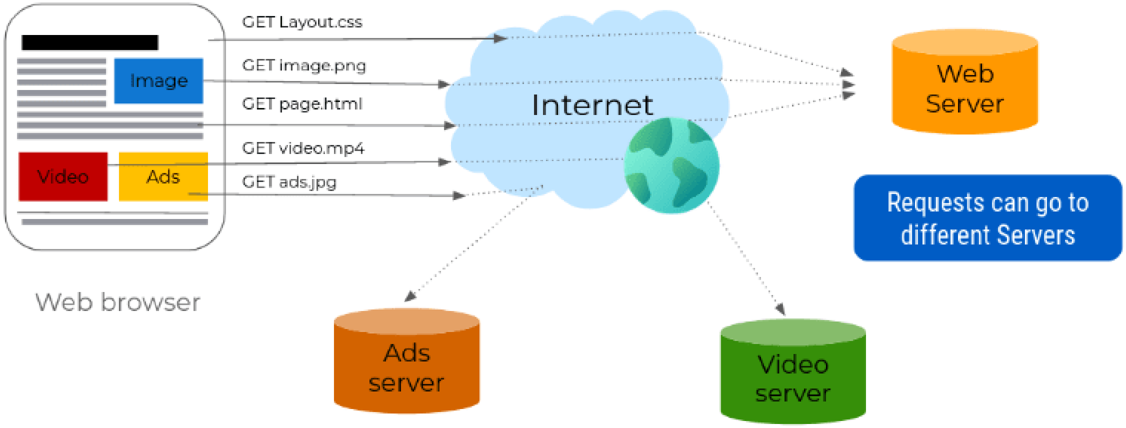
```
HTTP/1.1 200 OK
Date: Sun, 08 Feb xxxx 01:11:12 GMT
Server: Apache/1.3.29 (Win32)
Last-Modified: Sat, 07 Feb xxxx
ETag: "0-23-4024c3a5"
Accept-Ranges: bytes
Content-Length: 35
Connection: close
Content-Type: text/html

<h1>My Home page</h1>
```

HTTP Response Codes

- A HTTP response of a Response Header and a Response Body.
- A Response Header contains various info like:
  - Response Code : (1XX, 2XX, 3XX, 4XX, 5XX).
  - Content Type : (e.g.: text/html, image/png).
  - Content Length.



HyperText Transfer Protocol

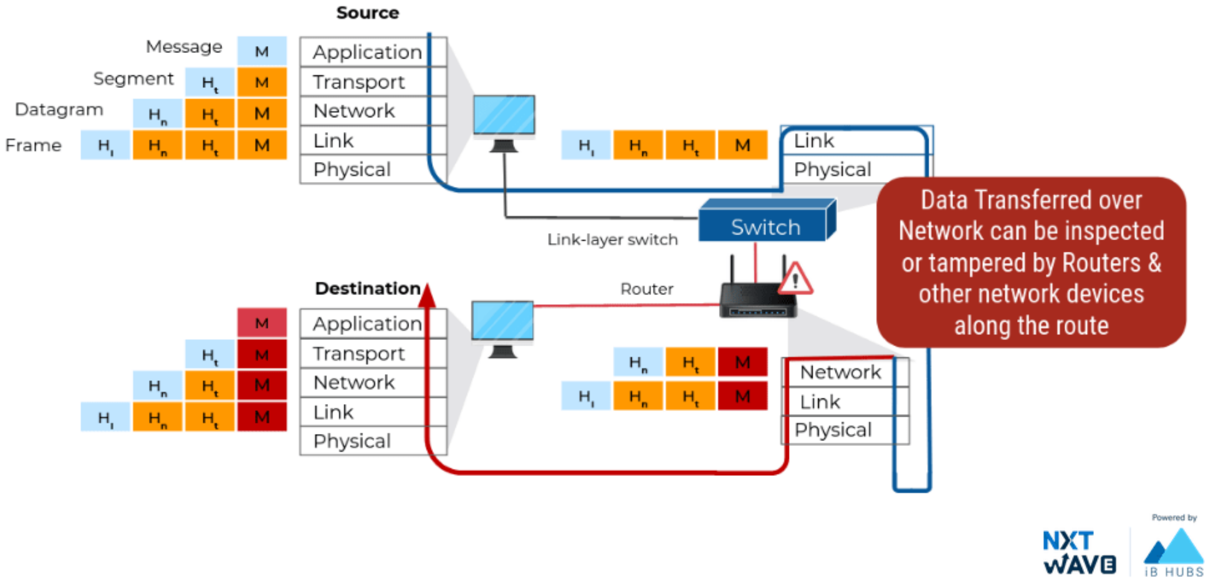


Computer Networks

Transferring Data Over Network

- The header (and footer) and the data together form the Protocol Data Unit (PDU) for the next layer.
- The process continues until reaching the lowest-level layer (physical layer or network access layer), from which the data is transmitted to the receiving device.
- The receiving device reverses the process, de-encapsulating the data at each layer with the header and footer information directing the operations.
- Then the application finally uses the data.

- The process is continued until all data is transmitted and received.



Data Transferred over Network can be inspected or tampered by Routers & other network devices along the route

HTTPS Protocol

HTTP Secure Protocol encrypts the data between the Browser and Server.

- Protects the privacy and security of the users.
- Protects the integrity of the website.
- Uses 443 port by default.