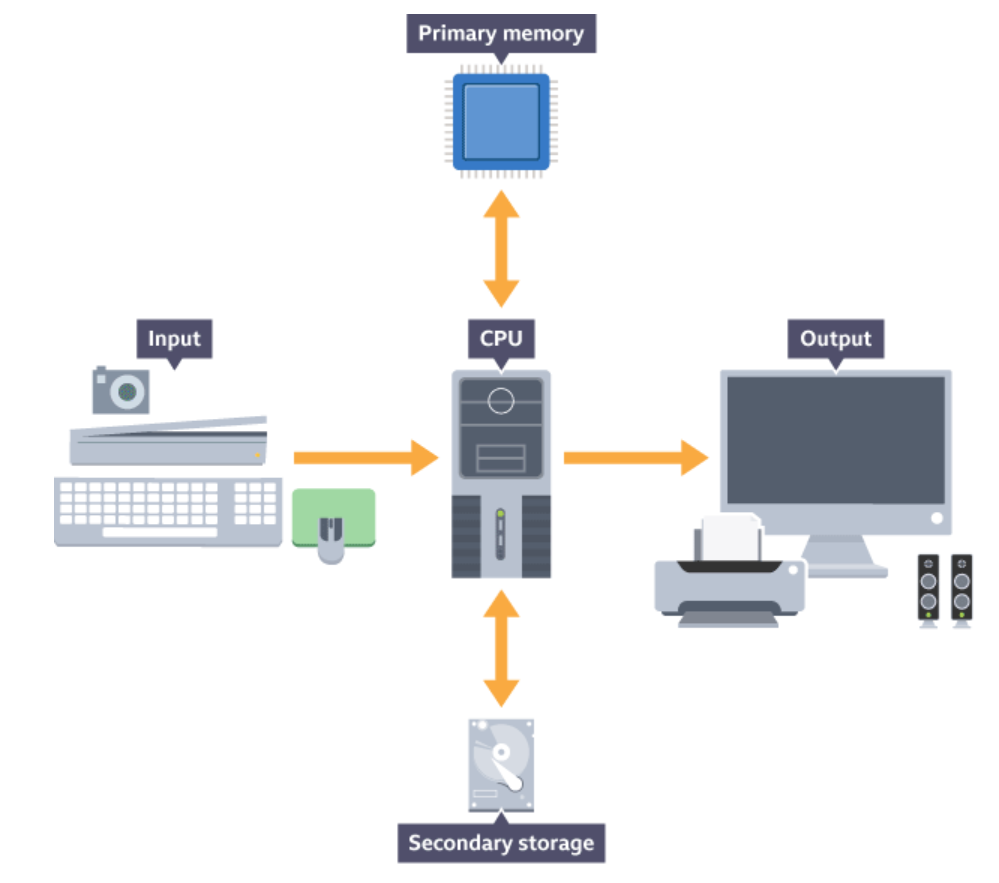
Hardware is the physical components of the computer, such as the central processing unit (CPU), hard disk, monitor, keyboard and mouse. Software is the programs that run on a computer.

The General Purpose Computer Model

The diagram illustrates the flow of data within a computer. Data is input, processed and then either output or sent to secondary storage. It is held in primary memory while it is being processed. The CPU receives instructions and data from an input or memory. The instructions and data are processed by the CPU and the results are either sent to an output or transferred to secondary storage.



**CPU – Central Processing Unit**

* CPU processes data, instructions and controls the rest of the computer system. All programs and data processing are run in the CPU and all hardware components are, to some extent, controlled by it.
* CPU has 6 key components, one of which is the cache. The cache is a small amount of high speed random access memory (RAM) built directly within the processor. It temporarily holds data and instructions that the processor is likely to use. This facilitates faster processing as processor does not have to wait for the data and instructions to be fetched from the RAM
* The clock is another CPU component. Coordinates each of the CPU components 🡪 clock sends out a regular electrical pulse which synchronises (keeps in time) all the components. The frequency of the pulses is known as the clock speed. Clock speed is measured in hertz. The higher the frequency, the more instructions can be performed in any given moment of time. In 1980s, processors commonly ran at a rate of between 3 megahertz (MHz) to 5 MHz 🡪today they commonly run at 3 gigahertz (GHz) - 5 GHz

Von Neumann architecture is the design upon which many general purpose computers are based. The key elements of Von Neumann architecture are:

* data and instructions are both stored as binary digits
* data and instructions are both stored in primary memory
* instructions are fetched from memory one at a time and in order (serially)
* the processor decodes and executes an instruction, before cycling around to fetch the next instruction
* the cycle continues until no more instructions are available

# Factors affecting CPU performance

* clock speed (hertz) – number of pulses the CPU’s clock generates per second. Higher pulses = more instructions processed in a time unit. N.B. User can speed this up – called overclocking – but can damage CPU as you cause it to overheat
* **cache size** – high speed RAM built within the processor. Bigger cache means less time a processor has to wait for instructions to be fetched
* number of cores - A processing unit within a CPU is known as a core. Each core is capable of fetching, decoding and executing its own instructions. More cores = greater the number of instructions that can be processed in a given space of time. Many modern CPUs are dual (two) or quad (four) core processors

**Embedded systems**

**Unlike general purpose systems like our windows laptop which have multiple applications.**

A small computer that forms part of a larger system, device or machine

Aims to control the device and to allow a user to interact with it

Tend to have one, or a limited number of tasks that they can perform.

Examples - central heating systems; TV, dishwasher,digital watch, electronic calculators, GPS systems, fitness trackers

Usually not programmable by a user –programming is usually done beforehand by the manufacturer. However, often possible to upgrade the software on an embedded device- software can often be upgraded by connecting the device to a PC and installing the new software. Advantages of embedded systems are:

* Limited number of functions means they are cheaper to design and build.
* They tend to require less power. Some devices run from batteries.
* They do not need much processing power. They can be built using cheaper, less powerful processors.

**Memory**

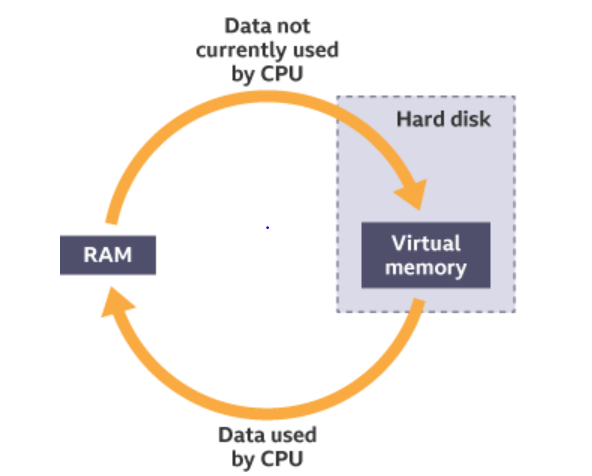
**Primary memory:**

* + built inside the computer, meaning data can be read from and written to primary memory extremely quickly.
  + Gives the processor fast access to the data and instructions that the primary memory holds.
  + Comparatively limited in size, especially when compared with secondary storage. In a modern personal computer, primary memory is often around 4 GB in size
  + 2 types of primary memory :

1. Read only memory (ROM)
   * non-volatile i.e. keeps contents even computer is switched off N.B cannot write to ROM
   * ideal for storing instructions and data needed for the computer to run
   * Such instructions are usually programmed by the computer's manufacturer and cannot be overwritten
   * The BIOS (Basic Input Output System) is an example of a program that is stored in ROM. Ensures hardware is functioning correctly, then runs a second program known as the bootup or bootstrap program that loads the computer's operating system from the hard drive into the RAM. The BIOS is always needed so it is stored in ROM.
2. random access memory (RAM)
   * volatile i.e. loses contents when unpowered
   * When RAM full, data is moved to virtual memory; a temporary storage place on the hard drive 🡪 using virtual memory slows down computer’s performance. This is because data has to be moved back and forth between RAM and virtual memory as needed
   * RAM is used to hold the operating system and any open documents and programs that are running.
   * The more RAM a computer has, the more data and programs it can hold simultaneously. RAM can also be upgraded easily, unlike other types of primary memory.

Virtual memory

* + The use of secondary storage as additional primary memory
  + Used when RAM needed to hold all running programs and data is greater than the amount of RAM available to the computer
  + computer's secondary storage e.g. hard disk stores this data temporarily
  + Virtual memory enables data that is in RAM and not currently being used to be transferred to the hard disk. This frees up room in RAM for other programs and data. When the data on the hard disk is needed again, any other unused data is transferred to the hard disk before the original data is transferred back to RAM. This process is known as swapping.
  + Using virtual memory slows your computer down, as the processor has to wait while data is swapped between hard disk and RAM.
  + secondary storage devices have slower access times than RAM, meaning the computer's processing performance can be severely impaired.
  + Processing performance can be improved by increasing the size of the RAM, thereby reducing the need for virtual memory.

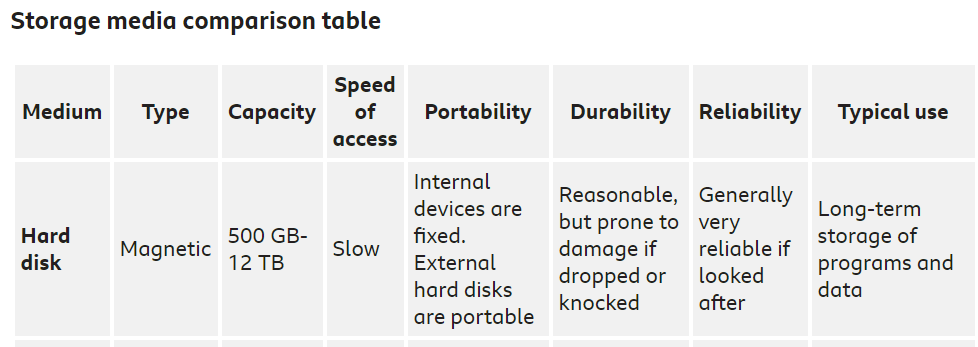


Cache Memory

* + Type of high-speed RAM
  + Built into the processor
  + Data can be transferred to and from cache memory more quickly than from RAM 🡪 thus used to temporarily hold data and instructions that the processor is likely to reuse 🡪 faster processing as computer doesn’t have to wait for data to be fetched from RAM
  + The more cache memory a computer has, the faster it runs
  + Its high-speed performance means cache memory is more expensive to build than RAM. Therefore, cache memory tends to be very small in size.
  + 2 types of cache: L1 cache (extremely fast transfer rates, but is very small in size) & L2 cache (slightly slower in speed, but bigger in size) – holds less frequently used data than L1 cache

Storage

* + Computers use primary memory such as RAM and cache to hold data that is being processed 🡪 but this type of memory is volatile, which means it loses its contents when the computer is switched off
  + General purpose computers (laptops etc) need to store programs and data for later use 🡪 since secondary storage is non-volatile, long-term storage, it is used to keep programs and data indefinitely. A hard disk drive has a high capacity and reasonable access speed, making it suitable for everyday storage of programs and data
  + Secondary storage devices are either:
* **magnetic** storage devices, such as hard disk drives
* **optical** storage devices, such as CD, DVD and Blu-ray discs (users lasers to read data - When the laser shines on the disc surface, lands reflect the light back, whereas pits (hollows) scatter the laser beam. A sensor looks for the reflected light. Reflected light (lands) represents a binary '1', and no reflection (pits) represents a binary '0'.)
* **solid state** storage devices, such as solid state drives and USB memory sticks



## Magnetic devices

Magnetic devices such as hard disk drives use magnetic fields to magnetise tiny individual sections of a metal spinning disk. Each tiny section represents one bit. A magnetised section represents a binary '1' and a demagnetised section represents a binary '0'. These sections are so tiny that disks can contain terabytes (TB) of data.

As the disk is spinning, a read/write head moves across its surface. To write data, the head magnetises or demagnetises a section of the disk that is spinning under it. To read data, the head makes a note of whether the section is magnetised or not.

Magnetic devices are fairly cheap, high in capacity and durable. However, they are susceptible to damage if dropped. They are also vulnerable to magnetic fields. A strong magnet might possibly erase the data the device holds.

# Networks

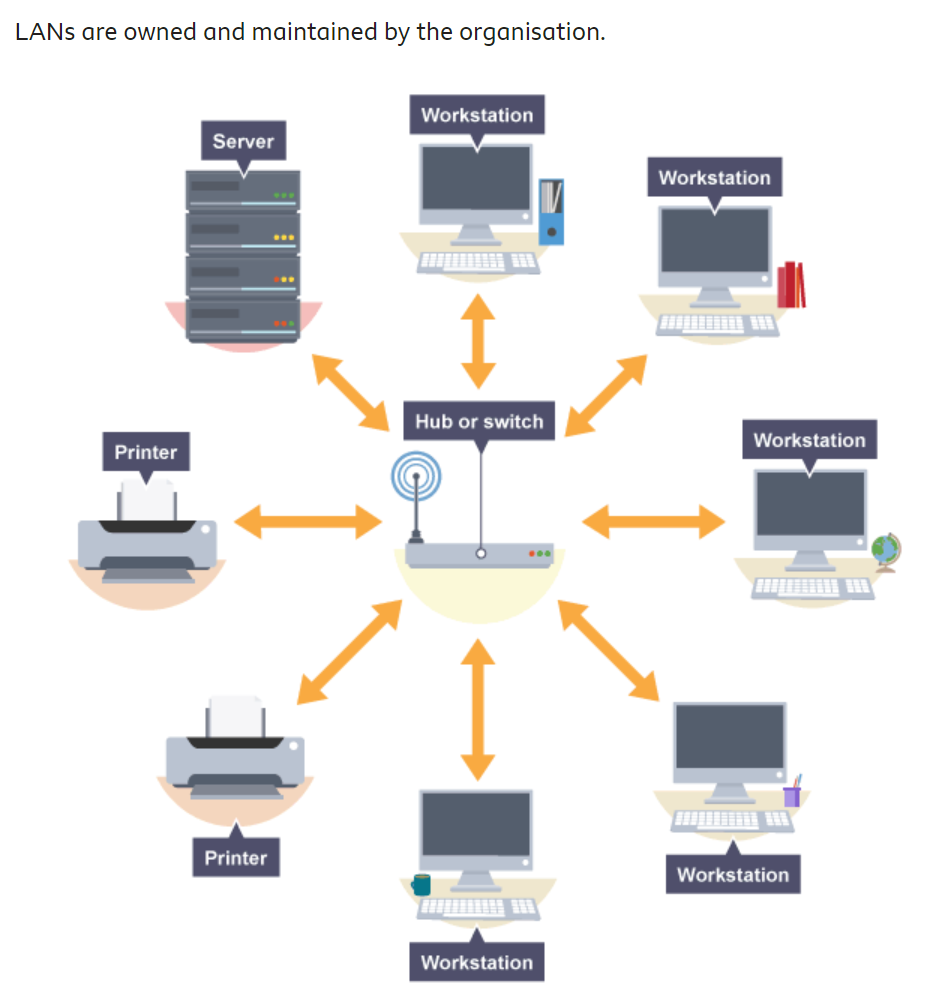
A network is two or more computers (or other electronic devices) that are connected together for the purpose of communication. They are connected by a wired medium such as cables, or by a wireless medium such as Wi-Fi.

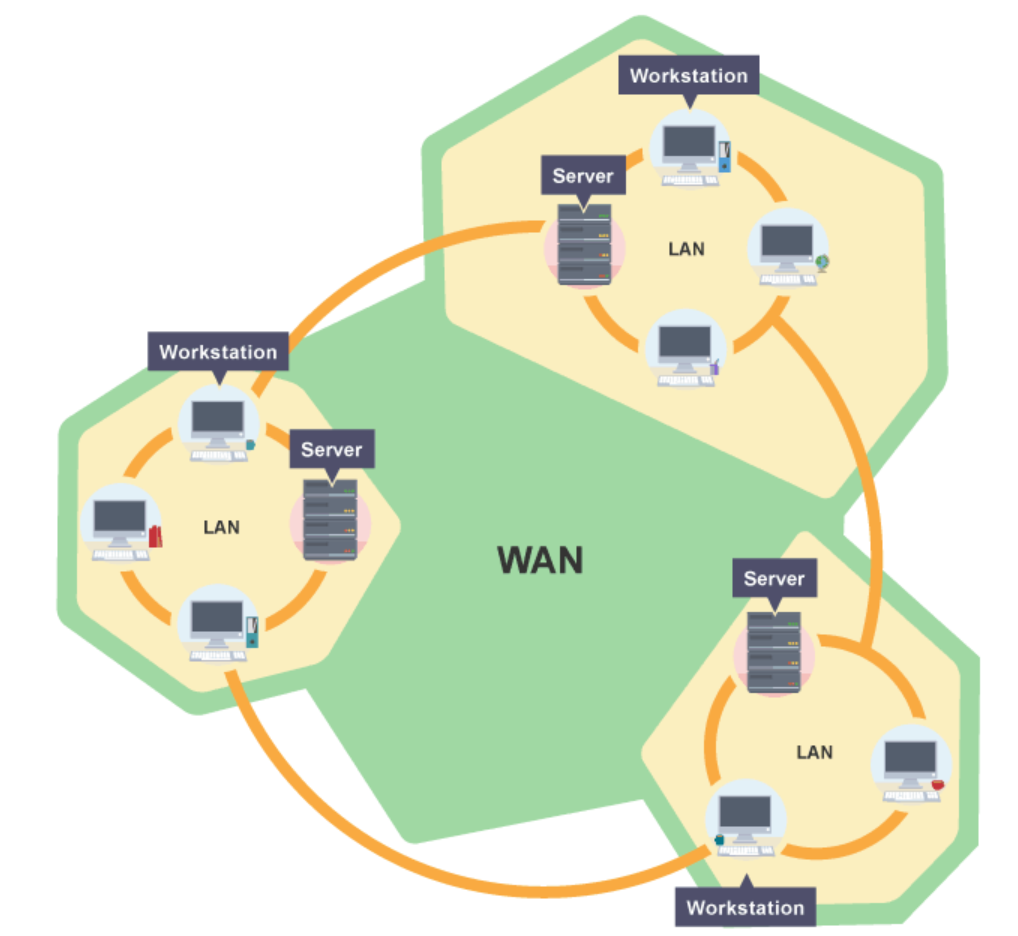
A computer or device that is not connected to a network is called a stand-alone.

Larger networks such as those used by sky enables :

* centralised maintenance and updates - network managers can apply software updates across a network, removing the need for a user to worry about having to do so
* centralised security - anti-virus software and firewalls can be implemented across a network, helping to protect user files from risks
* user monitoring - network managers can monitor what users do on a network
* levels of access - different users can be given different access rights. This gives network managers the ability to generally restrict user access to certain files, while granting permission to specific users

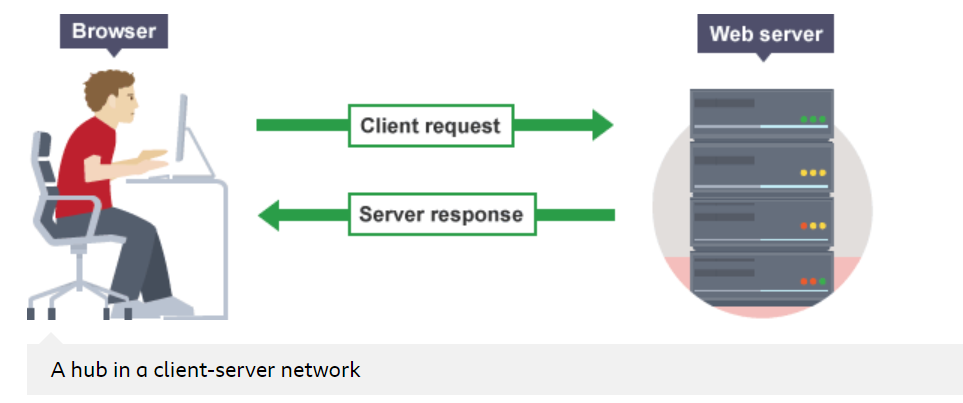
Two classifications of network:

* + LANs (Local Area Networks) - geographically confined to one building or site
  + WANs (wide area network) - a network that is spread over a wide geographical area.Sky will have a WAN as multiple UK offices - allows the head office to communicate and share data with the sub-offices and branches. Each office or branch has its own LAN that is connected together using the WAN. The internet is essentially a huge, international WAN.
  + 



How to connect computers over a network:

1. client-server networks
   * A computer is classified as either a server or client
   * Servers have high processing power as many computers connect to them
   * A server provides services to other computers e.g.s include:
     + file servers - hold and maintain user files
     + applications servers - allow programs to be run over a network
     + web servers - hold and share web pages
     + print servers - manage printing across a network
     + mail servers - handle emails between users
   * A client is a computer that relies on other computers (servers) to provide and manage data. The computer a person uses on a network is a client.
   * Client-server networks are best suited to organisations with many computers, or to situations where many computers need access to the same information. Many schools use this type of model.



1. peer-to-peer networks (P2P)
   * All computers have equal status
   * Peers store their own files, which can be accessed by other peers on the network. Therefore, a peer is both a client and a server.
   * P2P networks are best suited to smaller organisations that have fewer computers, or where fewer computers need access to the same data.

## Wireless access points (WAP)

## Routers - most commonly used connection devices. Send data signals across the internet. Used in homes, consisting of a hub and a WAP, enabling a small peer-to-peer network to be formed. Additionally, they contain a modem, which allows users to connect to the internet.

# The internet

* + The worldwide collection of computer networks
  + The term ‘internet’ refers to the network infrastructure, not the content stored on it.
  + The purpose of the internet is to enable communication on an international scale. Any device which connects to the internet - whether a PC, smartphone or internet fridge - becomes part of the internet, no matter where it is based.
  + The internet largely works on the client-server model. Web servers store and maintain web content, which is downloaded upon request to client computers; mail servers handle web-based email; media servers allow clients to stream music and videos

## Hosting

## Hosting is the storing of files and data on a web server. The web server is referred to as a host.

## The URL (uniform resource locator) for a website includes the host name. For example, the host name bbc.co.uk is included in the URL [www.bbc.co.uk](http://www.bbc.co.uk).

## The host name consists of three hierarchical levels, which work in reverse order of the name:

## 

## Domain name servers (DNS)

Every website on the internet has a network address, consisting of four sets of three digits. For example, the BBC website's address is 151.101.128.81.

Addresses like this are hard for users to remember. Instead, websites are accessed using domain names, for example, ‘bbc.co.uk’. Domain names are much easier for users to remember.

When a user visits a website, they enter the domain name as a URL. For example, the BBC website’s address is www.bbc.co.uk. A domain name server takes the domain name and looks up its equivalent network address. The user’s request is then forwarded to the server that resides at that network address.

## 

# The cloud

* + The cloud is a generic term for remotely accessed storage.
  + This storage is accessed through the internet.
  + Users do not actually know where their data is stored - the geographical location is unimportant. Users only need to know that their data is stored on a server connected to the internet.

Using cloud storage brings several advantages:

* the ability to access files from any location or any device, so long as an internet connection exists
* similarly, access can be granted to another user so they can remotely access your data
* reduced need to make backups - cloud storage services back up the data for you

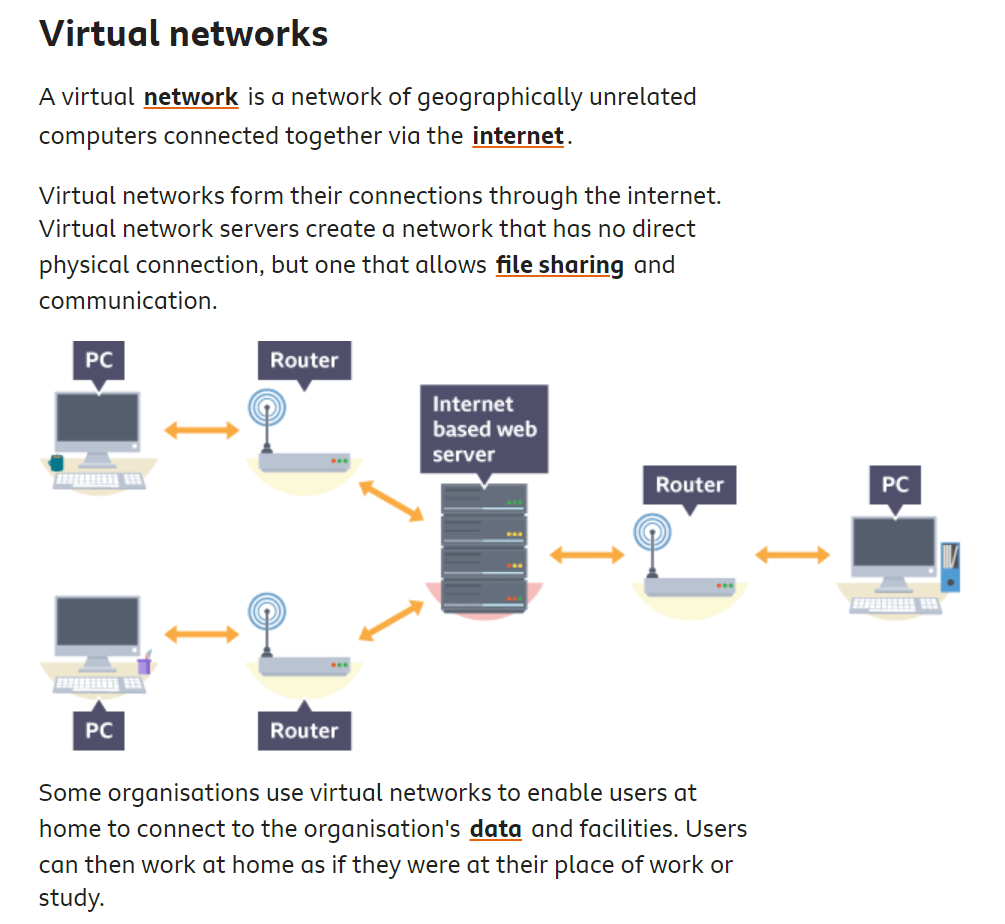
There are disadvantages to using the cloud:

* there is no guarantee that someone else is not accessing your data
* there is no guarantee that your data is being backed-up
* access to data is only possible with an internet connection - no connection means no access to data

In recent years, the cloud has also become a term for internet-based software facilities. These facilities usually offer limited versions of popular software such as word processors, spreadsheets and presentation software.

Cloud software facilities are ideal for low-powered devices, or for users who spend a lot of time travelling.

Virtual (Private) Networks:



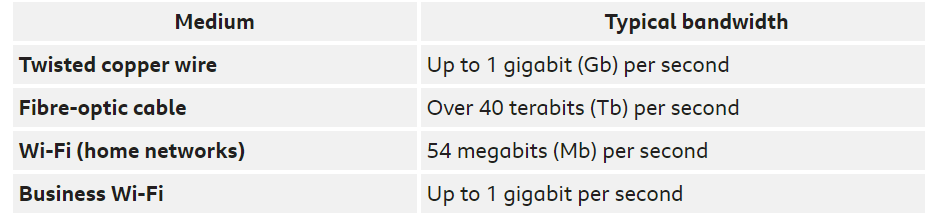
# Factors that affect the performance of networks

Network performance is about response time - how fast a message can be sent or how quickly a document can be retrieved. The performance of a network can be affected by various factors:

* the number of devices on the network
* the bandwidth of the transmission medium
* the type of network traffic
* network latency
* the number of transmission errors

Any network can be affected by one or a combination of these factors.

Bandwidth is a measure of the amount of data that the medium can transfer over a given period of time. Each transmission medium has a different bandwidth:



Each connected device requires bandwidth to be able to communicate. A home Wi-Fi network with one device would allocate 54 Mb per second to that device. If a second device joins the network, the bandwidth would be split between the two, giving 27 Mb per second to each, and so on. If ten devices were connected, the bandwidth allocated to each device would drop to 5.4 Mb per second, thereby reducing the rate at which data can be sent to any particular device.

In reality, however, things are more complicated. Different types of network traffic usually have different bandwidth requirements. For example, streaming a high definition video requires more bandwidth than streaming a low definition video. Some network switches are capable of determining the type of traffic and adjusting the bandwidth allocated to a particular device to accommodate the traffic's requirements.

**Latency**

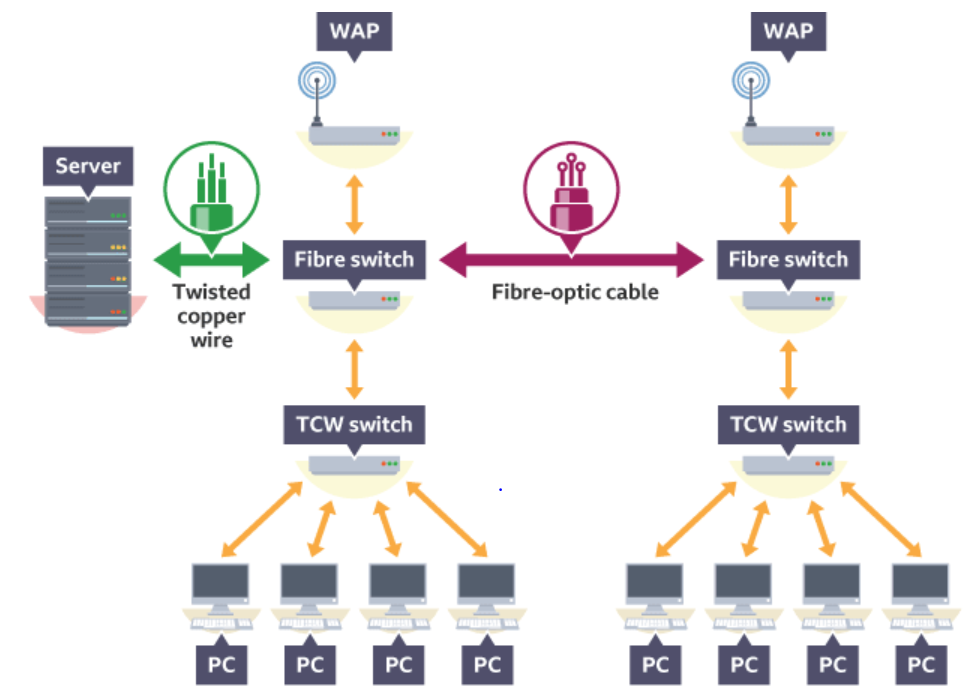
Network latency is a measure of how long it takes a message to travel from one device to another across a network. A network with low latency experiences few delays in transmission, whereas a high latency network experiences many delays. The more delays there are, the longer it takes to transmit data across a network.

Latency is affected by the number of devices on the network and the type of connection device.

A hub-based network will usually experience higher latency than a switch-based network because hubs broadcast all messages to all devices. Switch-based networks transmit messages only to the intended recipient.

The greater the number of devices connected to a network, the more important the choice of transmission medium becomes. Wi-Fi generally handles less traffic than twisted copper wire (TCW), which in turn handles less traffic than fibre-optic cable. Many networks include a combination of all three media:

* fibre-optic cables allow high data transmission between different buildings
* TCW runs from switches within buildings to individual devices
* Wi-Fi allows guest devices to connect to the network



## Transmission errors

Inevitably there will be times when devices try to communicate with each other at the same time. Their signals collide with each other and the transmission fails. It is similar to when two people speak to each other simultaneously - neither person is able to clearly hear what the other person is saying.

The greater the number of devices on a network, the more chance of a collision occurring, and the longer it takes to transmit a message.

**What is an OS?**

**Examples are: Windows/ Mac OS X/iOS/Android(based on linux)/ Linux / Ubuntu/**

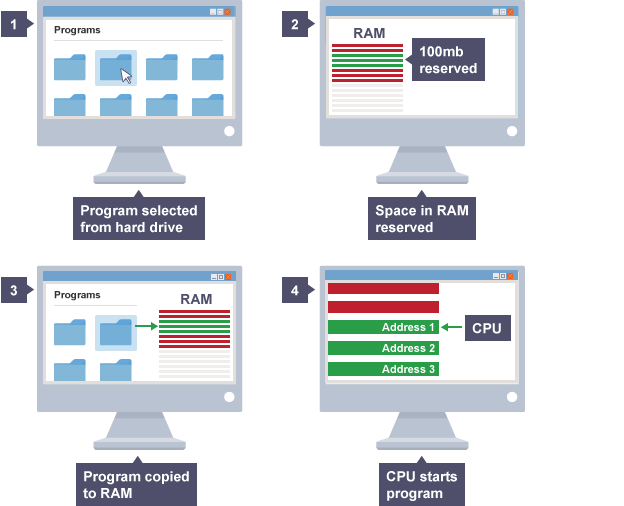
**Functions include:**

* **interface - provides a user interface so it is easy to interact with the computer**
* **manages the CPU - runs applications and executes and cancels processes**
* **multi-tasks - allows multiple applications to run at the same time**
* **manages memory - transfers programs into and out of memory, allocates free space between programs, and keeps track of memory usage. Different** processes running at the same time must not interfere with one another. This means they have to use different parts of the computer’s memory.
* **manages peripherals - opens, closes and writes to peripheral devices such as storage attached to the computer**
* **organises - creates a file system to organise files and directories**
* **security - provides security through user accounts and passwords**
* **utilities - provides tools for managing and organising hardware**

**The OS provides 2 types of UI: GUI (graphical) or CLI (text-based)**

When the OS runs a piece of software it:

1. Finds the program files and code on the storage drive
2. Loads them into main memory - a section of RAM is reserved for the program and space is allocated for the program's data
3. Instructs the CPU to start executing the program from the beginning. For this the CPU program counter is set to the memory location of the first instruction in the program, and execution begins



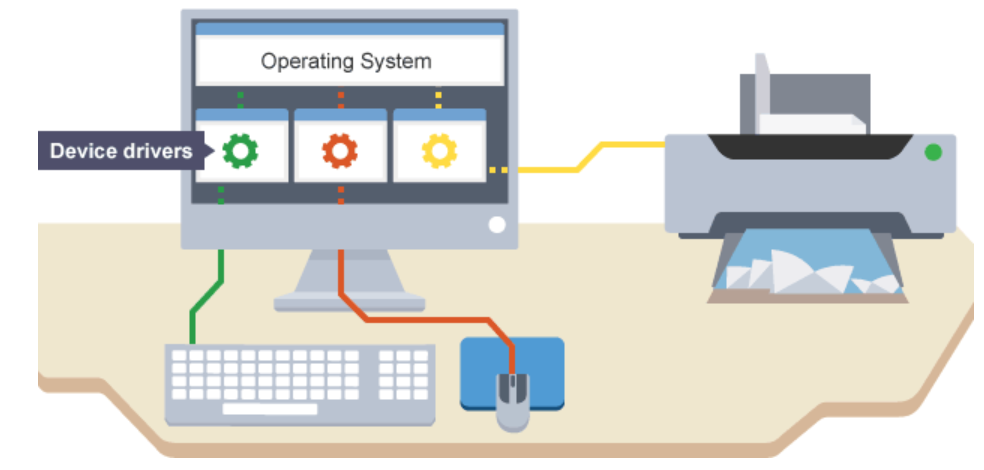
Several programs can be stored in RAM at the same time, however only one program at a time is processed by the CPU. Programs can be in one of three states:

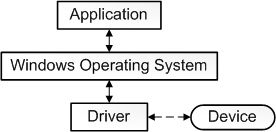
* running
* waiting
* runnable

Only one process can be running at any one time. CPUs are extremely fast, so if a program is processed for even a short time it can do quite a lot. The OS decides the best way to swap between running, runnable and waiting processes. It controls which process is being executed by the CPU at any point in time, and shares access to the CPU between processes. The job of working out when to swap processes is known as **scheduling**.

Swapping happens so fast that it appears that all processes are running at the same time. When there are too many processes, or some of them are making the CPU work especially hard, it can look as though some or all of them have stopped.

**What is a (device) driver?**

* The OS uses programs called **device drivers** to manage connections with peripherals e.g. mouse/keypad/printer
* It handles the translation of requests between a device and the computer
* wakes up the device when it is needed and put it back to sleep when it is not
* An OS will have generic device drivers to enable it to connect to most common peripherals. Some peripherals, however, will have their own drivers that need to be installed before use
* Peripherals that use the same protocol may be controlled by the same driver. If a number of identical game controllers are plugged in, each device will store its data in a different place so they do not interfere with each other.
* Each peripheral is programmed with its own machine code. Each has its own rules that dictate how it transmits data values between the computer and the device. These rules make up a protocol for controlling and communicating with the device
* 
* A software component that lets the OS and a device communicate with each other
* the code that allows us to use our video cards, keyboards, mice, printers, and so forth.
* An application might need to read some data from a device, so the application calls a function implemented by the operating system, and the operating system calls a function implemented by the driver. The driver knows how to communicate with the device hardware to get the data. After the driver gets the data from the device, it returns the data to the operating system, which returns it to the application.



* **The driver provides a software interface to hardware devices, enabling operating systems and other computer programs to access hardware functions without needing to know precise details about the hardware being used**
* Drivers are hardware dependent and operating-system-specific
* **The main purpose of device drivers is to provide abstraction by acting as a translator between a hardware device and the applications or operating systems that use it**

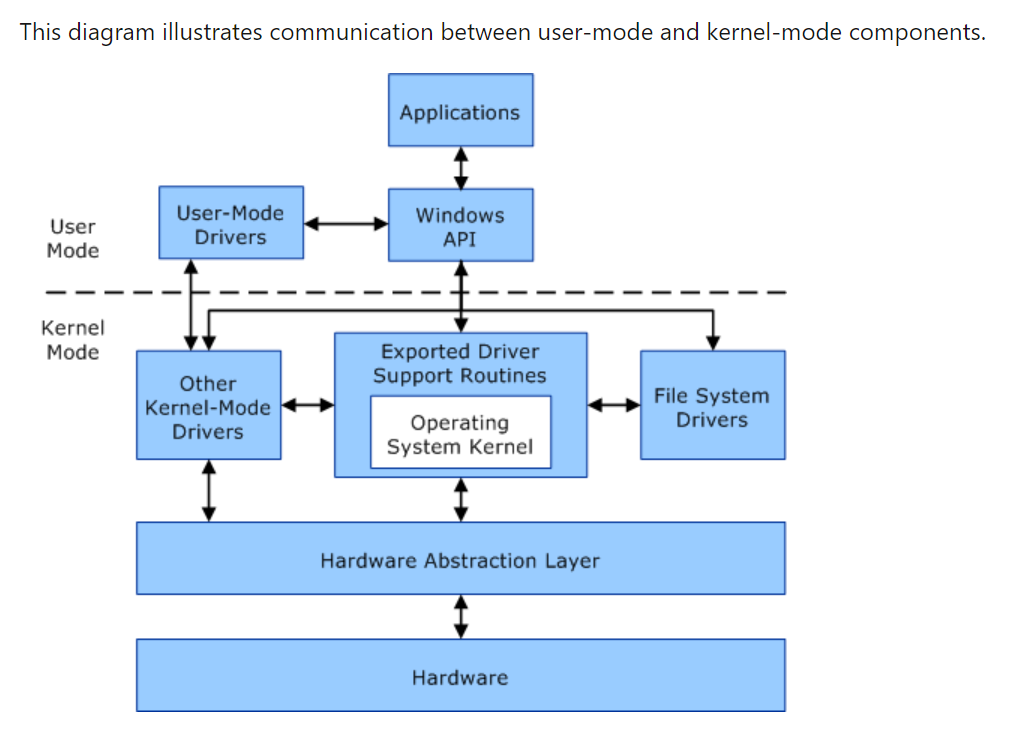
**Virtual Device Drivers**

* **Used to emulate a hardware device, particularly in virtualization environments, for example when a DOS program is run on a Microsoft Windows computer or when a guest operating system is run on, for example, a Xen host**
* **They emulates a piece of hardware, so that the guest OS and its drivers running inside a virtual machine can have the illusion of accessing real hardware**
* **Attempts by the guest operating system to access the hardware are routed to the virtual device driver in the host operating system as e.g., function calls. The virtual device driver can also send simulated processor-level events like interrupts into the virtual machine.**

**Kernel vs User Mode**

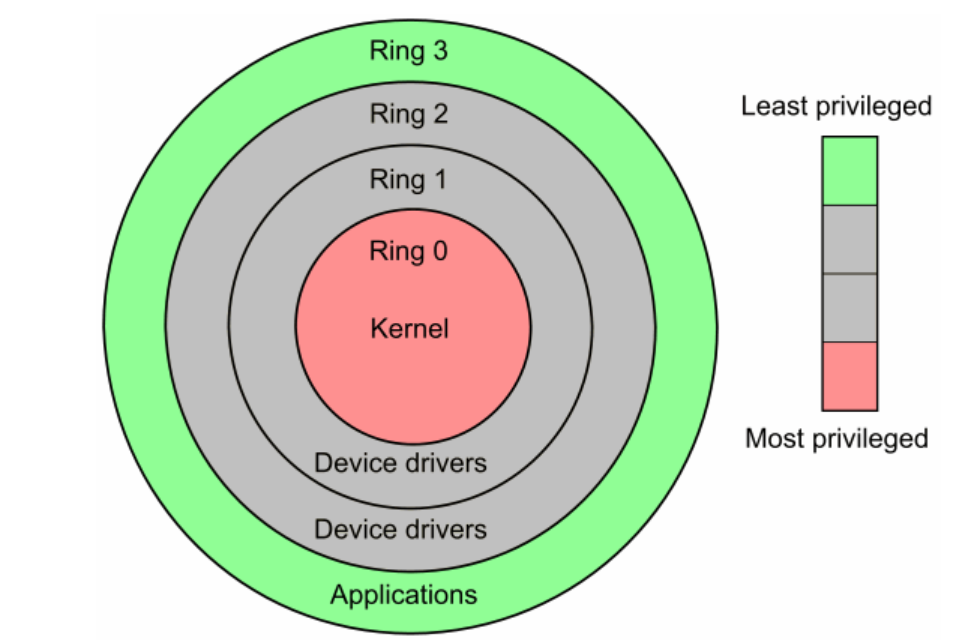
* A processor in a computer running Windows has two different modes: user mode and kernel mode
* The processor switches between the two modes depending on what type of code is running on the processor.
* Applications run in user mode, and core operating system components run in kernel mode
* While many drivers run in kernel mode, some drivers may run in user mode.
* User mode ensures each application runs in isolation, and if an application crashes, the crash is limited to that one application. Other applications and the operating system are not affected by the crash.
* A processor running in user mode cannot access virtual addresses that are reserved for the operating system. Limiting the virtual address space of a user-mode application prevents the application from altering, and possibly damaging, critical operating system data.
* All code that runs in kernel mode shares a single virtual address space. This means that a kernel-mode driver is not isolated from other drivers and the operating system itself. If a kernel-mode driver accidentally writes to the wrong virtual address, data that belongs to the operating system or another driver could be compromised. If a kernel-mode driver crashes, the entire operating system crashes.

|  |  |
| --- | --- |
| **Kernel Mode** | **User Mode** |
| **The executing code has complete and unrestricted access to the underlying hardware. It can execute any CPU instruction and reference any memory address. Kernel mode is generally reserved for the lowest-level, most trusted functions of the operating system. Crashes in kernel mode are catastrophic; they will halt the entire PC.** | **In User mode, the executing code has no ability to directly access hardware or reference memory. Code running in user mode must delegate to system APIs to access hardware or memory. Due to the protection afforded by this sort of isolation, crashes in user mode are always recoverable. Most of the code running on your computer will execute in user mode.** |
| **Running something in kernel mode ensures maximum performance** | **Running something in user mode ensures maximum stability** |



**The CPU's strict segregation of code between User and Kernel mode ensures that if an app crashes the whole computer does not catastrophically crash all the time**

**x86 CPU hardware actually provides four protection rings: 0, 1, 2, and 3. Only rings 0 (Kernel) and 3 (User) are typically used.**



**Environmental Variables**

Path is an environmental variable. Specify here a set of directories where executable programs are located. N.B. Many programs do not appear in the path as they are not designed to be executed from a command window, but rather from a Graphical User Interface.

**Custom Hotkeys & Running Macros**

Assign custom keyboard shortcuts / hotkeys – need third party software

* [Microsoft Mouse and Keyboard Center](https://www.microsoft.com/accessories/downloads/mouse-keyboard-center)
* WinHotKey
* Or can create your own macro within a word doc and assign hotkeys to it – can then apply to all word docs or just the current one. N.B. Cannot right click so must access everything using the ribbon etc.
* NOTE: Cannot record a macro within ppt 2016 onwards, so would need to know how to write VBA code. Can however record macros in excel

**The role of an interpreter**

Computers cannot understand code in the way humans write it and hence, you need an interpreter between the computer and the human written code. An interpreter like python will convert the code into a format that computers can then understand and process. A python interpreter processes code like so:

* Processes python script sequentially
* Compiles the code into a byte code format which is a lower level language understood by computers
* A python virtual machine (PVM) iterates over the instructions of your low-level bytecode to run them one-by-one

**What is a .exe file?**

- An executable file is one which contains a program - that is, a particular kind of file that is capable of being executed or run as a program in the computer

- COM and BAT are other types of executable file types in Windows

- A file whose name ends in ".exe" is really a program that when opened causes the operating system to run the program. Users who receive an .exe file as an e-mail attachment should always be sure that the file comes from a trusted source and does not contain a virus.

**.bat file ?**

* A script file in windows
* Consists of a series of commands to be executed by the CLI, stored in a plain text file
* Allows a user to create a script to automate multiple regular tasks
* Can comment out using rem
* Can set variables e.g. set foo = ‘this is foo’
* Can print to CLI using echo Hello World

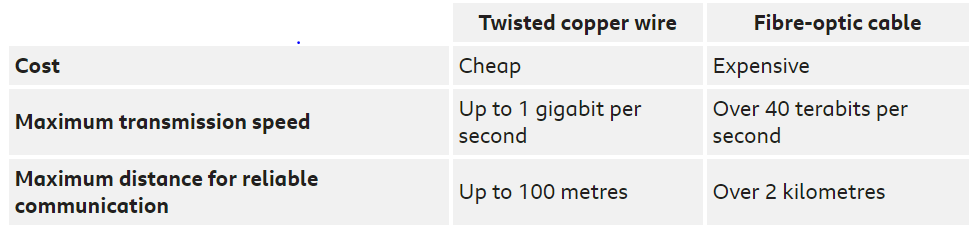
Bandwidth = maximum rate of data transfer across a given path. Bandwidth may be characterized as network bandwidth, data bandwidth, or digital bandwidth. Fibre has higher bandwidth than copper. N.B. **Bandwidth** refers to the maximum capacity of an internet connection, not the actual speed. if you order a 100 Mbps package from your Internet Service Provider (ISP), your bandwidth would be 100 Mbps. That means the most data your connection can download at one time is 100 Mbps. Your actual internet speed, on the other hand, is likely to be less than your bandwidth most of the time because of network congestion and other external factors.

Transmission media carry data signals from one computer to another. Media are either wired or wireless.

In modern networks, wired media use either twisted copper wiring, or fibre-optic cable. Each wired medium has differing characteristics.

Wired connections are reliable and not usually subject to interference. However, they limit mobility - a wired device usually has to stay in place.

Wireless connections use radio waves to carry signals. These signals are limited in range (usually up to 50 metres), but are subject to magnetic interference - they can also be blocked by walls. However, they are ideal for mobile devices as a device can connect to a network as long as it is in the range of a wireless access point.



### **What is latency? (aka ping)**

**Lower latency is better.**Latency refers to how much time it takes for a signal to travel to its destination and back. When referring to internet connections, the destination is usually your ISP’s servers.

Lower latency is better because latency is essentially a delay between when you take an action and when you see the result—high latency is when it takes longer to see the results. The less delay, the better.

Every time you put in a request to your internet connection (search for something on Google, check social media, etc.), it sends a signal to the server to retrieve the information and then bring it back to you. Since this usually happens pretty quickly, latency is measured in milliseconds.

**Gaming -** Latency is key for online gaming – otherwise get lags and could be dead. Bandwidth less important as normally have the game loaded onto computer already.

### **Video chat -** Video chatting, like FaceTime or Skype, can be negatively impacted by both low **bandwidth** and high **latency**. Low bandwidth will affect the quality of your chat, making things hard to see. Latency will cause sync issues and freezing.

**Streaming** - involves downloading content from a server, **bandwidth** tends to be the major factor in both video and audio streaming. That’s because streaming happens with little input on your end: you just click and wait. Low bandwidth will usually show itself in two ways. It will either manifest as a painful amount of buffering as your connection tries to keep up with the size of the content. Or it will show up as terrible video quality because your streaming service is attempting to compensate for the slow download speed

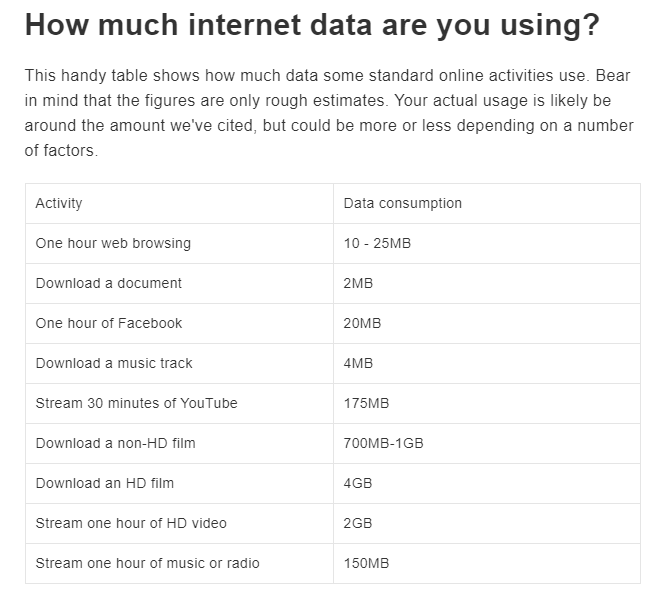
### **Browsing -** Not even basic, everyday web browsing is immune to the effects of a poor internet connection. Low **bandwidth** will cause pages to load sloooowly and in segments (like in the old dial-up days).

And with high **latency**, pages may load superfast but there will be a maddening delay at the beginning where it seems like nothing’s happening.

* **Satellite Internet** – e.g. Sky – can get BB / satellite TV / etc is not very widespread anymore but perfect for getting internet to places that are difficult to reach with traditional cables or mobile masts - like rural areas, mountains, islands, forests, and jungles. If you've ever used [Wi-Fi](https://www.broadbandchoices.co.uk/guides/wifi/wifi-broadband) on an aeroplane or a boat, chances are you've used satellite broadband.It’s also good for rural areas. Speeds are comparable to standard [ADSL broadband](https://www.broadbandchoices.co.uk/broadband/adsl-broadband-deals) - and within a few years they'll be closer to [fibre optic broadband](https://www.broadbandchoices.co.uk/broadband/fibre-optic). Latency (also known as lag) is very high, because of the distance the signal has to travel - i.e. 22,200 miles into space and back. You'll probably find it a bit difficult to [play online games](https://www.broadbandchoices.co.uk/broadband/gaming) or access a remote desktop. It's often expensive. Monthly bills tend to be higher than those for fixed line broadband, and set-up costs can be in the hundreds. It's prone to interference, thanks to the weather and all the other wireless signals floating around the sky. A bad storm may make your connection suffer. Unless you're in a very remote area, we'd generally recommend fixed line broadband - it's a lot cheaper, and more ideal for tasks that need a low latency.

Fibre vs Copper

|  |  |
| --- | --- |
| **Copper** | **Fibre** |
| Uses electrons for data transmission | **Uses photons (i.e. light) for data transmission** |
| Lower bandwidth | Light is faster than electrical pulses, so fiber can transmit more bits of data per second and offer higher bandwidth. |
|  | All data signals degrade over a range, but fiber offers significantly better signal durability. Fiber only loses 3% of the signal over distances greater than 100 meters, compared to copper's 94% loss of signal. |
| Copper cables are designed to conduct electricity, making copper Internet vulnerable to power lines, lightning, and deliberate signal-scrambling. | Fiber optic bundles do not conduct electrical currents, making fiber data connections fully-resistant to fire, electromagnetic interference, lightning, or radio signals. |
| Copper cable can easily be broken during an installation or by accident. Despite its large size, copper has a low tolerance for tension | Fiber is smaller, lighter, and more durable than copper cabling and can generally only be damaged through deliberate vandalism, though you do have to be careful with fiber as it is made out of glass. Typically, it's sheathed in a protective coat to make it more durable. |
|  |  |



Task Manager – to see how much RAM etc each program is using

Indexing Options – can search files faster as computer knows where to look. Currently have changed advanced settings to enable search within file’s contents too (but only of R files)- By default, all the properties of your files are indexed, including file names and full file paths. For files with text, their contents are indexed to allow you to search for words within the files. **ISSUE** index less than 10 percent of the size of the indexed files. For example, if you have 100 MB of text files, the index for those files will be less than 10 MB.

**Parsing**

* To split a file or other input into its (more meaningful/understandable) component parts so that it can be easily stored/manipulated/ processed by a computer.
* a string of commands – usually a program – is separated into more easily processed components, which are analyzed for correct syntax and then attached to tags that **define** each component. The **computer** can then process each program chunk and transform it into machine language
* Parsing an argument in python – is done in CLI

Onedrive VS SharePoint VS Teams

* If you’re working on a file by yourself, [save it to OneDrive](https://support.office.com/en-gb/article/video-create-files-and-folders-in-onedrive-work-or-school-e1f59717-2f02-494d-93c6-8ef9613e82ba). Your OneDrive files are private unless you [share](https://support.office.com/en-gb/article/share-onedrive-files-and-folders-9fcc2f7d-de0c-4cec-93b0-a82024800c07) them with others,
* If you’re already working as a team — in Microsoft Teams, SharePoint, or Outlook—you should save your files where your team works, because OneDrive connects you to all your shared libraries, too.

What's the difference between OneDrive and SharePoint?

OneDrive is the files experience for Microsoft 365 and SharePoint Server, giving you access to all your work files, including the files that people share with you directly or through the teams you're working on. OneDrive provides a consistent, intuitive files experience across all your devices, including web, mobile, and the desktop of your Windows PC or Mac.

Behind the scenes, SharePoint provides the content services for all files in Microsoft 365, including files you work with in Teams, Yammer, and Outlook. SharePoint is always there, helping manage and protect your files, and powering content collaboration across Microsoft 365. And beyond files, SharePoint enables portals, news, pages, lists, and a platform for business apps.

With both OneDrive and SharePoint, your files are stored in the cloud. You can sync either OneDrive or SharePoint files to your computer. See [Sync OneDrive files](https://support.office.com/en-gb/article/sync-files-with-onedrive-in-windows-615391c4-2bd3-4aae-a42a-858262e42a49) or [sync SharePoint files](https://support.office.com/en-gb/article/sync-sharepoint-files-with-the-new-onedrive-sync-app-6de9ede8-5b6e-4503-80b2-6190f3354a88) for more info.

## Moving documents from OneDrive to SharePoint

Sometimes documents that you start by yourself grow in importance and become relevant to a project. When that happens, it may make sense to [copy](https://support.office.com/en-gb/article/copy-files-and-folders-between-onedrive-and-sharepoint-sites-67a6323e-7fd4-4254-99a8-35613492a82f) or [move](https://support.office.com/en-gb/article/move-files-and-folders-between-onedrive-and-sharepoint-5916f90d-f58a-4bf9-b135-10853f516d0b) files from OneDrive to SharePoint.

**Indexing Options (windows search) -** helps you get faster results when you're searching it for files and other things. Indexing is the process of looking at files, email messages, and other content on your PC and cataloging their information, such as the words and metadata in them. BUT All data gathered from indexing is stored locally on your PC. A rule of thumb is that the index will be less than 10 percent of the size of the indexed files. For example, if you have 100 MB of text files, the index for those files will be less than 10 MB. So if huge files, be careful. **I changed it so it searches within file as well**

PARALLEL PROCESSING:

* Different to multithreading
* a process is split into parts, which are executed simultaneously on different processors attached to the same computer.

If you’re concerned with the speed of your python code, then changing the CPU/RAM will be fruitless 🡪 INSTEAD, use parallel processing

Agnostic -  refers to the ability of something – such as software or hardware – to work with various systems, rather than being customized for a single system e.g. code is agnostic if can be applied across various sky products for various years etc. Database agnostic means software can run in many DBMS. Can be device-agnostic i.e. software works across multiple devices.

Web page and web app are two ways you interact with the web.

|  |  |
| --- | --- |
| Web page | Web app |
| Static  Click on a link (on a website), which causes the website to generates a git request from the web browser to the web server for a new html document which is then rendered.  Causes the browser bar to spin | Telltale sign is a loading bar shown on screen requiring a longer start up time e.g. google music / google drive – does this because its downloading a bunch of java script/code to provide an application like experience  There’s a loading spinner in the middle of the page, but the page isn’t getting reloaded…if you look at URL…would appear to change except that anything after the ‘#’ isn’t recognised by the server…so as far as it’s concerned you’re still on the same page  Will often run on javascript  The changes to the display are being handled by javascript and not the server  ​Will often be coded in browser supported language e.g. html/javascript as these languages rely on the browser to render the program executable  computer program that performs a specific function by using a web browser as its client  Benefits: A web application relieves the developer of the responsibility of building a client for a specific type of computer or a specific operating system, so anyone can use the application along as they have internet access. Since the client runs in a web browser, the user could be using an IBM-compatible or a Mac. They can be running Windows XP or Windows Vista. They can even be using Internet Explorer or Firefox, though some applications require a specific [Web browser](https://www.lifewire.com/what-is-a-browser-446234).  ​ |

The "client" is used in client-server environment to refer to the program the person uses to run the application. A client-server environment is one in which multiple computers share information such as entering information into a database. The "client" is the application used to enter the information, and the 'server' is the application used to store the information.

Operating Systems – the Software which can control the hardware e.g. Windows/Linux/ioS/Android/MacOS/

Version control repositories = CVS, Subversion or Git - contain every change to the source code including the date (the "when"), the responsible developer (the "who"), as well as little message that describes the intention (the "what") of a change.

Kernel – the central module of the OS. The kernel code is responsible for memory management, process/task management/disk management. Connects the system hardware to the application software.

Airflow ( consensus is that it’s better than the GCP dataflow equivalent) - *a platform to programmatically author, schedule and monitor workflows.* Basically, it helps to automate scripts in order to perform tasks. Airflow is Python-based but you can execute a program irrespective of the language. For instance, the first stage of your workflow has to execute a C++ based program to perform image analysis and then a Python-based program to transfer that information to S3. Possibilities are endless.

Jira – used for sprints/project management

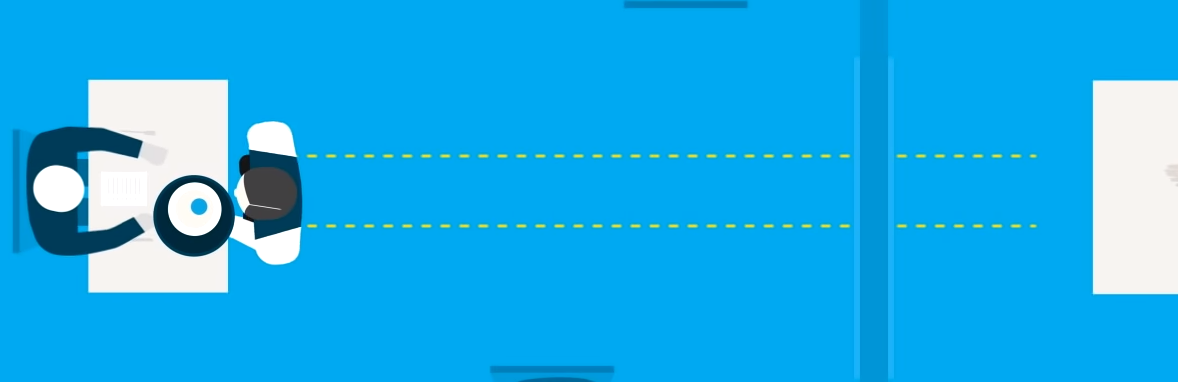
Multithreading - a technique by which a single set of code can be used by several processors at different stages of execution.

Virtual Desktop - Virtual desktops often rely on virtualization software, which abstracts operating systems, applications and data from a computing device's underlying hardware. User’s desktop environment is stored remotely on a server, rather than a local PC. It separates the desktop OS, applications and data from the hardware client, storing this virtua desktop on a remote server. In a virtual desktop environment users access their personal desktop remotely, over the internet froma a client device. Means can access business desktop from smartphone, other PC etc.Provide greater security for companies as employees not carrying around confidential company info on a laptop that could be lost/stolen. Also, you can setup multiple independent workspaces to house different sets of windows. When you create a new virtual desktop (press Ctrl+Win+D), you’re given a blank canvas to open a new set of apps and windows. The apps you had open on your first desktop aren’t visible on the new one and won’t appear on the taskbar. Likewise, any apps you open on the new desktop will be invisible on the original.. You can switch between virtual desktops using the Ctrl+Win+Left and Ctrl+Win+Right keyboard shortcuts. You can also visualise all your open desktops using Task View – either click the icon on the taskbar, or press Win+Tab. USEFUL when working on say 3 independent projects at same time and can keep them separate

VM – a CPU with memory,storage etc. but it’s all virtual, not physical. Several VMs can live within a server- they don’t have a 1-2-1 relationship with hardware. Can spin up a VM on a desktop/ on a server / or in the cloud 🡪 use them for: creating high availability clusters and minimise downtimes.

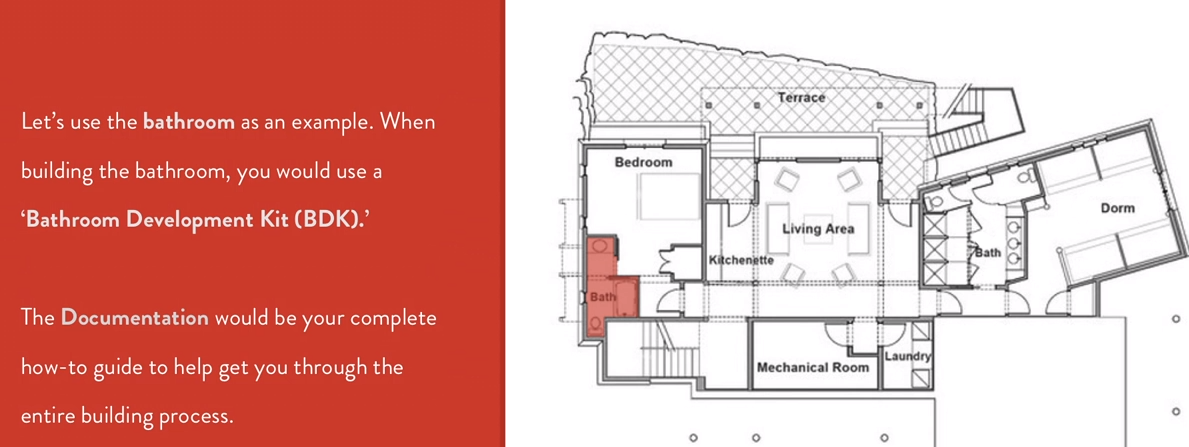
API – Application Programming Interface

* The messenger OR
* ‘the waiter’- you’re in a restaurant and looking at the info on the menu. The kitchen is the part of the system that prepares your order. How do you communicate your order to the kitchen and how does it get delivered back as food to your table.



* When choosing flights on BA/QATAR etc - you essentially have the menu- you select data/departure city/class etc 🡪 this goes directly to the database and gives the results back 🡪 BUT if using skyscanner (an aggregator) – it interacts with the airline’s APIs from BA to Qatar. i.e APIs allow these travelsites to exist!
* – takes requests- tells a system what you want to do, then returns the response back to you
* pieces of software that one program uses to interface with another
* helps standardise the coding process to allow two different interfaces to interact.
* Allows our different devices and software to work well together
* MAC OS/Windows/Linux all have APIs with a GUI, that means you can just click start menu/open chrome, rather than have to tediously type code to open up these applications. Provides a more reliable/ consistent UX e.g. close button the same across all applications.
* A simpler way for developers to interact with other kinds of software – Twitter/FB APIs exist – makes it easier for other users to embed tweets/facebook statuses into their own website/blog 🡪 otherwise would have to share code directly with owners of other sites-tedious/difficult/might give away trade secrets.
* Windows API either win 32 or win 64, dependent on whether running a 32-/64-bit application.

Software Development Kit (SDK)

* Analogy – don’t build a house from scratch (takes a lot of time/money) – you build off of the current house and make renovations, room by room. Use Bathroom Development Kit (BDK) for bathroom. Documentation teaches you how to attach the new bathroom fixtures to the the current piping. The libraries = the existing bathroom infrastructure e.g. bath/toilet/sink/floorplan. Can alter these or leave them how they are. The API connects the bathroom to the rest of the house’s water supplies and to the city’s water supplies/ pipes etc.
* 
* SDKs are building blocks
* A virtual toolbox (to build the house / grow the grain to make the toast etc)- a set of tools used for developing applications for specific operating systems
* Rather than tools e.g. hammer, contains **libraries**, which makes it play nicely on whatever platform it’s on. Contains **Visual Editors (**to design and layout graphical elements).
* E.g. the library within Windows SDK (Microsoft word) makes calls to the Operating System to get it to perform certain functions easily
* Rather than build an app from scratch, can springboard using these SDKs e.g. chat service like whatsapp.
* SDK contains:

1. Documentation (describes how to integrate the SDK)
2. Libraries (the code a developer can incorporate)
3. API (allows you to request services from another application)

Kubernetes - "platform for automating deployment, scaling, and operations of application containers across clusters of hosts

* a container orchestration solution

Worker – i.e. a node – a A node may be a VM or physical machine, depending on the cluster

Computers interact with each other by:

* **Unicast** – direct communication, one-to-one
* **Multicasting** – one device talking to a specific group e.g. subscribers
* **Broadcasting** – one-to-all, any device can pick up the info

Computers have several different **identifiers**:

* NETBIOS- old school 15 character max name
* hostname (cmd prompt)
* MAC address (Network card)
* IP address (Network card)

Websites have domain names e.g .net .com . org .io (British Indian Ocean Territory- mainly startups)

* 2 letter domain codes reserved for country codes e.g .uk .fr .de
* ICANN (Internet Corporation for Assigned Names and Numbers) are in charge of keeping track and assigning domain nam
* Whois- lookup who owns domain name
* Reverse whois – looks up names and find their registered domains

**IP (**Internet Protocol **) Addresses**

* A 32 bit address (Version 4- currently use this). This means there can be a 2^32 unique addresses.
* IP version 6 (Next Generation) - 128 bit code, so can have 2^128 unique addresses.
* The IP address is split into 4 octets w.y.y.z
* First octet (high order) of the host name (not network component) tells you type:
  + 1 🡪 233 = unicast
  + 224 🡪 240 = multicast
  + 255 = broadcast
* IP address split into network and host. The number of digits within each side of the split is variable so long as they add up to 32 (can put split anywhere)
* Subnet mask hides network number
* For computers to unicast/interact directly they must share the same network ID
* Subnet mask always continuous string of 1s followed by all 0s. e.g. 255.255.254.0

=111111111 11111111 11111110 00000000

BINARY

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **128** | **64** | **32** | **16** | **8** | **4** | **2** | **1** |
| **10=** | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 |

(but don’t need first few 0s in general so 10 = 1010 in binary)

* When trying to run cd [\\skyshare\\......](file:///\\skyshare\\......)
* Get met with the following error: CMD does not support UNC paths as current directories.
* UNC = Universal Naming Convention
* You cannot change to a remote network/directory using command prompt 🡪 probably could troubleshoot though ………

**GPU vs CPU**

|  |  |
| --- | --- |
| **Graphics Processing Unit** | **Central Processing Unit (essentially the brain)** |
| designed to quickly render high-resolution images **and** video concurrently. | **CPU** is designed to handle a wide-range of tasks quickly  limited in the concurrency of tasks that can be running |
| Requires less memory than CPU | **Needs more memory than GPU** |
| Faster than CPU | **Lower speed than GPU** |
| contains many but weaker cores | Contains fewer, but more complex, powerful cores |
| Suitable for parallel instruction processing – thousands of threads computing at a time . Better at simpler but multiple concurrent calculations. More logical cores whose basic design is to process a set of simpler and more identical computations in parallel | Not suitable for parallel instruction – processing is more sequential with few threads at any one time  Handles single, more complex calculations |
| GPU emphasis on high throughput | CPU emphasis on low latency. |
| Bandwidth optimised | Latency (memory access time) optimised |

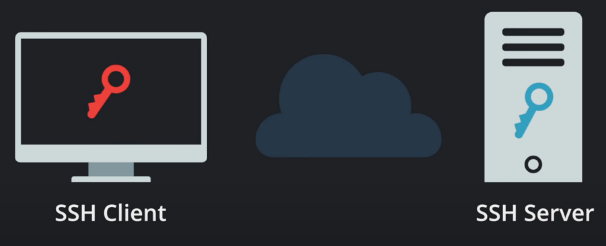
**So which to use – CPU or GPU ?**

* GPU compute instances will typically cost 2-3x that of CPU compute instances, so unless you’re seeing 2-3x performance gains in your GPU-based training models, I would suggest going with CPUs.
* For deep learning (specifically for training your model), use GPUs due to significant speed increase as can embrace parallel processing required in these massive calculations involved in finding the optimum parameters 🡪 model training is composed of simple matrix math calculations, the speed of which can be greatly enhanced if the computations can be carried out in parallel
* Inference stage is less resource heavy (compared to training) so can use CPUs 🡪 this is also cost saving

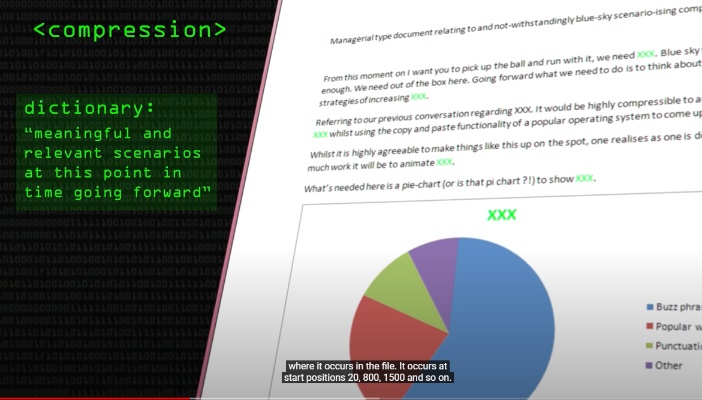
How to use GPUs?

* Kubernetes
* Use CUDA package (parallel computing architecture)
* Use tensorflow GPU enabled

SSH (Secure Shell)

* Allows one computer to securely connect to another computer over unsecure networks like the internet
* Without encryption, data travels through the web in plain text 🡪 easy to intercept password/user data etc.
* But SSH encrypts data through a tunnel so can securely login to remote machine/transfer files etc
* Commonly implemented using client-server model 🡪 have SHH client (i.e. your local computer) & SSH server 🡪 is set up with a public key (stored on SSH Server) + Private Key (locally stored on SSH client)
* The SSH client provides the ID of the key pair it wants to use to prove its identity 🡪 SSH server creates a challenge which is encrypted by the public key 🡪 you decrypt this with your private key and send the original challenge back to SSH Sever
* Connection should then be established
* N.B. To **use SSH on Windows**, **you** have to download an **SSH** client. **One** of the best and freely available clients is called PuTTY. Will also need **PuTTYgen to**  generate SSH key pairs, which are special files you can use for encryption, authentication, and so on

Data Compression

* Reducing the amount of information but without in principle losing any of it – you could essentially recover the full information that you started with
* But will cost you less in terms of transmitting over a network as pay be bit/byte
* Drawback 🡪 decompression can take time at the other end 🡪
* You compress data by looking for repeated patterns and predictability e.g. for a text file, if continually referring to a source ‘The Introduction to Statistical Learning for Application in ML’. This phrase might be used 30 times, where each usage represents 50 bytes. You take one copy and a note of where it appears in the file (see image). These numbers then take 3 bytes each time

E.G.2 Image compression – windows screensaver with desert island tree – a lot of blue pixels which can just be represented by one pixel. Can represent this pattern pretty easily. Just need to give coordinates of where computer places each pixel. In contrast, the picture of a race car in movement/ andy pollock picture would be far less easy to compress as far less predictable patterns and more random 🡪 would have to transfer the whole pic



E.G. 3 JPEG Compression – sites that host pics don’t want each of your pics taking up 20Mb 🡪 will want it compressed down to Kb 🡪 issue is that scenery is random enough that not easy to compress 🡪 JPEG does some clever tricks that still allows you to compress but at the expense of causing blurriness if you magnify the image e.g. will split pic up into areas of roughly similar colours and approximate a colour for the area etc 🡪 result is never fully recoverable