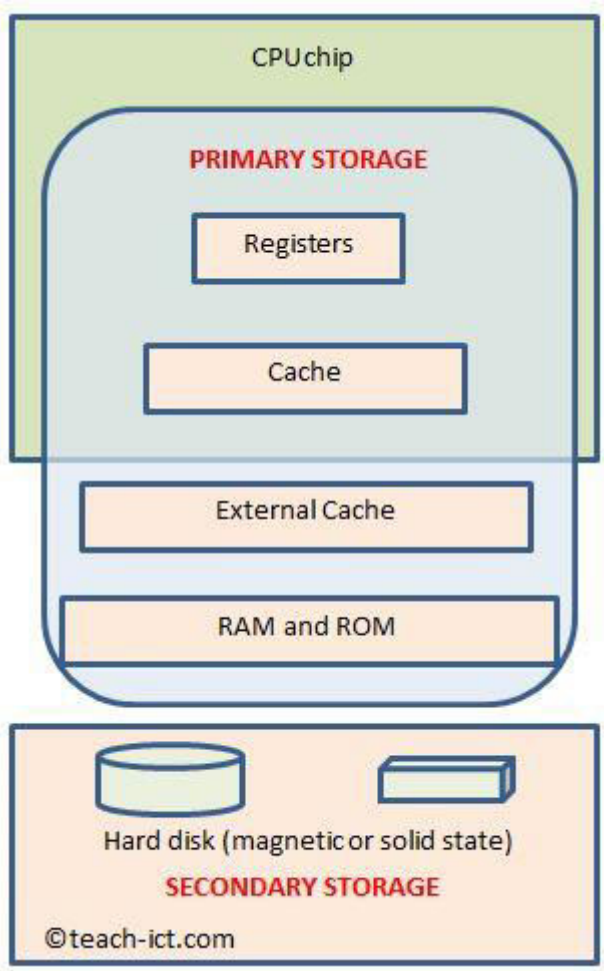
Primary memory

1. Primary Storage (Main memory).

The need and purpose of primary storage is to hold both data and programs that are in current use by the CPU.



Primary storage devices have a direct connection to the CPU core and some of them are built into the CPU chip itself.

Primary storage devices are extremely fast and so the CPU can run at top speed. Trying to run a program directly from secondary storage such as a hard disk would not allow the CPU to run at anywhere near of what it is capable of.

Primary storage devices are about 200 times faster than secondary storage devices.

There are other names given to primary storage such as **main memory, main storage, primary memory.**

Examples of main memory include

- Random Access Memory (RAM)
- ROM chips
- Cache
- CPU registers

Main memory can also hold more than one program at a time, which makes it quicker for the CPU to swap between tasks.

Primary memory

2. Volatile Memory

How many times have you worked for ages on something and just before you were about to save it the computer crashed or your mate 'accidentally' switched it off? And as most of you have probably experienced, when you reboot your computer, all of that work is gone forever.

The reason for this is that anything you are currently working on is stored in main memory. Main memory is almost always **volatile**, with the exception of ROM chips, which will be discussed later.

Volatile memory is cleared whenever the supply of power is lost. In contrast, non-volatile memory will continue to hold anything it was storing even if it is not receiving any power.

Secondary storage is always non-volatile, as the programs and data that are held there are meant to be kept long-term. This is why, while you might lose any files you are working on if power is cut, if you save them first those files are safe.

Volatile memory includes RAM, CPU registers and the cache.

KEY TERM: Volatile memory is a type of memory that loses its data once power to the memory chip is switched off or interrupted.

Primary memory

3. RAM

RAM is used as main memory. It acts as a temporary store for program instructions and data. It can only store things temporarily because it is volatile - it must constantly be powered or it will lose any data it is holding.
It is useful as a temporary store because any data or instructions in RAM can be accessed in any order, i.e. it can be accessed non sequentially. This is why it is called *Random Access* Memory.

Because the CPU can access any memory location at any time, programs held in RAM can run very quickly.

In summary:

- RAM is used as main memory. It is accessed directly by the CPU
- RAM is volatile memory. If it loses power, the data it holds is lost
- RAM holds billions of storage locations, each with its own memory address. These can be accessed in any order. This is why it is Random Access Memory

The photo below shows a typical RAM module. It is slotted into a socket on the motherboard. This is why it is fairly simple to change or upgrade RAM in a computer as you just swap out the memory module.

Primary memory

4. DRAM and SRAM

There are two types of RAM, namely DRAM and SRAM.

Dynamic Random Access Memory (**DRAM**) is what you will find in a RAM module such as the one below

DRAM is commonly used for main memory because it is relatively inexpensive. However, it needs to constantly receive a "refresh signal" from the memory controller system to keep its data intact otherwise it will lose the data it is holding.

SRAM does not require a refresh signal - it will retain its data as long as the power is on.

Static Random Access Memory (**SRAM**) is commonly used for the cache. This is because it is much faster to access than DRAM however SRAM is more expensive than DRAM and so it is seldom used as bulk volatile memory (perhaps an embedded computer might use it).

Comparison

Volatile memory	Volatile memory
Used in RAM modules	Used in the cache
Less expensive than SRAM	More expensive than DRAM
Slower to access than SRAM	Faster to access than DRAM
Requires a refresh signal to retain its data intact	No refresh signal required

Primary memory

5. How much RAM?

Memory is measured in the number of bytes it can store. For convenience, the number of bytes are combined into larger sized groups. Shown below.

Sizes

Bit	b	one bit	1 bit
Nibble		half a byte	4 bits
Byte	B	1 byte	8 bits
Kilobyte	KB	1024 bytes	1000 bytes
Megabyte	MB	1000 KB	1,000,000 bytes
Gigabyte	GB	1000 MB	1,000,000,000 bytes
Terrabyte	TB	1000 GB	1,000,000,000,000 bytes
Petabyte	PB	1000 TB	1,000,000,000,000,000 bytes

RAM these days is sized in gigabytes.

How much RAM does a modern personal computer need? The answer varies, it depends on what you plan on doing with it.

For example, Windows 10 needs a minimum of 2GB of RAM just to run. Then you need to specify more RAM for running applications. A budget laptop typically includes 4GB of RAM. This is enough for running some basic email, a browser and perhaps a word processor or movie player.

If you intend to run memory hungry applications such as computer games or image processing software you will more than likely need a minimum of 8GB.

One of the most memory hungry applications is professional video editing and 32GB of RAM is not unusual. Of course you could run video editing with much less RAM, but the application will run a lot slower.

Primary memory

6. ROM

Read Only Memory (**ROM**) is a type of primary storage because it is directly accessed by the CPU. Just like RAM, memory locations in ROM can be accessed in any order.

There are two key differences between RAM and ROM.

ROM is Read Only Memory which means that its data content cannot be changed, overwritten or removed by a running program, whereas RAM is read/write memory which means that the contents can be altered.

Secondly, unlike RAM, ROM is non volatile, meaning it keeps its data when there is no power supply.

These two features mean that ROM is ideal for storing key data that must not be lost or overwritten, such as

- Hardware settings that must not be changed
- Initial boot up instructions for the computer
- Data that should not be changed - e.g. MAC address in a network card

Comparison

ROM - Read Only Memory	RAM - Random Access Memory
You cannot write to ROM, you can only read from it.	RAM allows you to both read and write data.
ROM is a type of non-volatile memory.	RAM is a type of volatile memory.
ROM chips are usually located on the motherboard or printed circuit board such as a graphics card. They are not removable as they are soldered into the motherboard.	RAM chips are located in removable memory modules that are slotted into sockets on the motherboard. This means they can be easily removed and updated.
ROM is used to hold basic computer hardware settings and in the past it held the BIOS to boot up the computer.	RAM is used as main memory to hold both data and programs
There may be just a few megabytes of ROM in a computer.	There are usually gigabytes of RAM in a computer.

Primary memory

7. Virtual memory

RAM has billions of memory locations but sometimes even that is not enough room for all the data the CPU needs to handle. When RAM gets too full, the computer operating system can help out by temporarily marking sections of secondary storage for the CPU to store data on.

These sections are called **virtual memory**.

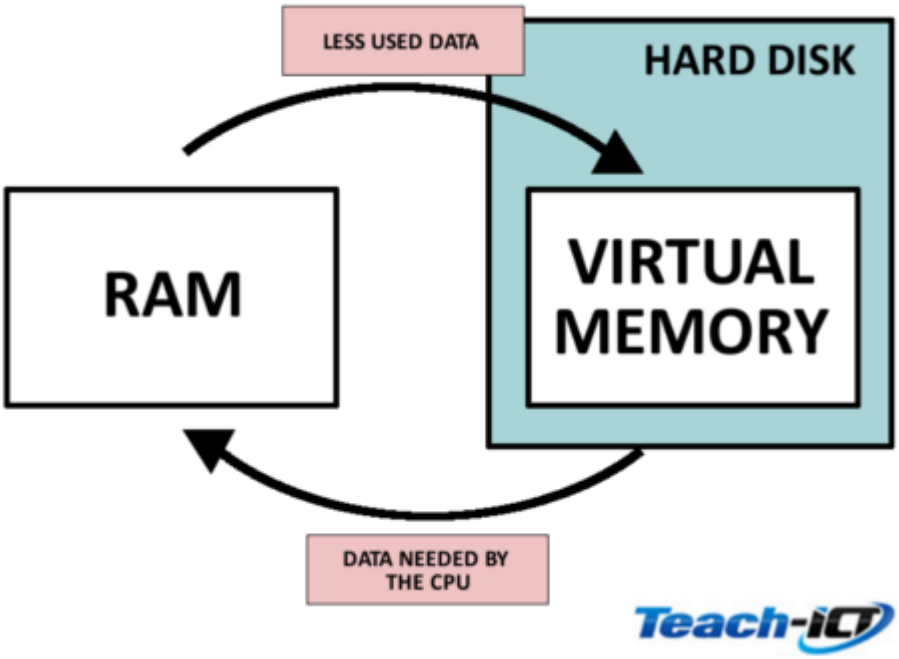
Virtual memory appears to exist as main memory (RAM) but it is in fact implemented with secondary storage devices by the operating system. Data moves automatically between main memory and virtual memory as needed.

Virtual memory is effectively volatile even though the data is stored on hard disk. If the computer is turned off, the operating system loses track of the content of virtual memory and so starts with a clean slate once more.

Even though the CPU can use virtual memory - it is very slow compared to RAM.

Let's say the CPU needs to use a block of data but it is sitting in virtual memory. First of all the operating system works out which data is the least-used in RAM at that instant. It then stores that chunk of data in virtual memory located on the hard disk and clears it from RAM. This gives enough room to re-load the needed data from virtual memory.

This moving back-and-forth between RAM and virtual memory continues as long as there isn't enough RAM to store all the data needed by the set of running applications.



Disk Thrashing is a problem that may occur when virtual memory is being used too heavily. As main memory fills up, then more and more pages need to be swapped in and out of virtual memory. This swapping leads to a very high rate of hard disk access leading to wear and tear on the disk and a slow-down in overall performance. The solution is to either fit more RAM or reduce the amount of applications being used at the same time.