**Virtual memory**:

The concept of virtual memory in computer organisation is allocating memory from the hard disk and making that part of the hard disk as a temporary RAM. In the earlier days, when the concept of virtual memory was not introduced, there was a big troubleshooting that when RAM is already full but program execution needs more space in RAM. The computers became unresponsive in such type of situation since processor forced the program to be in RAM but RAM can’t hold them because it is already running out of space. To deal with this type problem the concept of virtual memory was introduced.

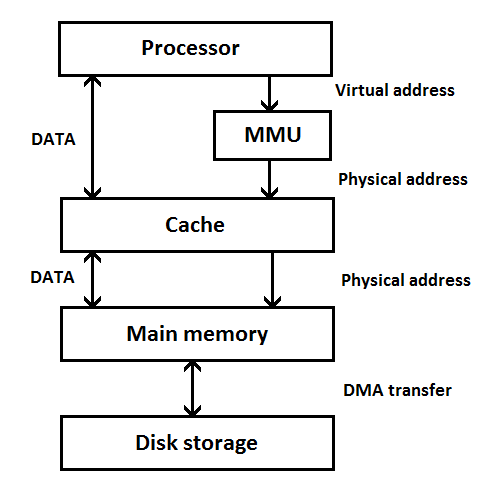
**The concept of virtual memory in computer organisation:**

Basically, virtual memory provides an illusion to the users that the PC has enough primary memory left to run the programs. Though the size of programs i.e. to be executed may sometimes very bigger than the size of primary memory left, the user never feels that he needs a bigger primary storage to run that program. When the RAM is full, but program execution needs more space in RAM, then the operating system occupies a portion of the hard disk and uses it as a RAM. In that part of the secondary storage, the part of the program which not currently being executed is stored and all the parts of the program that are eventually executed are first brought into the main memory.

**What are address space and memory space in virtual memory?**

So, more precisely, we can define virtual memory as the technique which automatically moves programs and data blocks into the main memory when they are required for execution. The processor generates a virtual or logical address for accessing the data from secondary memory and then it is converted into the physical address with some paging algorithm. Set of all such virtual address is called address space and the set of all physical address is called memory space.

If a virtual address pointing to a specific part of a program, which is currently in the main memory, then it is accessed immediately. But if the address pointing is not in the main memory i.e. when the referenced address is in the secondary memory, then the data from this address must be brought into the main memory before using that. In this situation processor actually performs a swapping operation between primary and secondary memory. The program or the part of the program which has the highest priority of execution are moved on to the main memory and lower priority programs are moved to the secondary storage.



**Fig -Virtual memory system**

The conversion of the virtual or logical address into the physical address is done by a special hardware unit known as MMU (Memory Management Unit). When the required data are found in the main memory, then they are fetched to the cache memory for further processing. If required data are not found in the main memory, the MMU with the help of operating system brought the data from secondary storage to main memory. This transfer of data from secondary memory to primary memory is performed using DMA (Direct Memory Access) scheme, which is a feature of a computer system that allows accessing the main memory by certain hardware system. This whole process is independent of the central processing unit.

Working **principle of virtual memory**:

Now, let see the whole concept of virtual memory with a small example. Suppose the capacity of the main memory of a computer system is 32K while the secondary memory can store 1024K. That means the secondary memory can store data equal to the capacity of 32 main memory. If we denote address space by N and memory space by M, then we can have N = 1024K and M = 32K.

Here, 15 bits are required to specify each physical address since 32K = 215 and 20 bits are required for each virtual address because, 1024K = 220. So, here CPU is referring an address of 20-bit long for accessing the data from main memory but in reality, our main memory has the only 15-bit long address for accessing data. So, in this case, to access the data from main memory CPU has to perform a mapping of 20-bit virtual address to a 15-bit physical address.

When the concept of virtual memory was not introduced, the programmers and software developers were needed to be aware of limitations imposed by the available main memory space. But now, in modern generations of computers, the operating system takes this responsibility of memory management with the help of hardware. It moves the programs and data automatically between main memory and secondary memory. So, programmers and developers do not need to be aware of the problem that may occur for main memory limitation at the time of developing a program.

Unlike primary memory and secondary memory i.e. RAM and Harddisk, we can’t see or touch virtual memory. As its name suggests, the whole concept is virtual or logical. But our operating system provides a way to see how much virtual memory is currently being used. Below is a screenshot captured from a windows 7 pc which clearly shows how much virtual memory is being used by that particular machine.

**Advantages and disadvantages of virtual memory:**

The big advantage of virtual memory is that only a part of the program needs to be in memory for execution. So, more programs can run simultaneously at the same time interval. Virtual memory can increase CPU utilization and overall power of a system since large programmes can be run with real less primary memory.

The major disadvantage of virtual memory is when a system uses virtual memory, the application may run very slowly. So, CPU takes more times to switch between the applications.