Linux Installation:

(Virtual Memory and Swap space Disk Partition (df, fdisk), Adding Swap Space)

hitbullseye

Table of Contents

What is virtual memory?

- **Virtual memory is a memory management approach that allows secondary memory to be utilized as if it were primary memory. Virtual memory is a method that is often utilized in computer operating systems (OS). Virtual memory enables a computer to compensate for physical memory shortages by temporarily shifting data from random access memory (RAM) to disk storage. Mapping memory chunks to disk files allows a computer to utilize secondary memory as if it were main memory. Most personal computers (PCs) nowadays have at least 8 GB (gigabytes) of RAM.**

- However, this is not always enough to execute many applications at the same time. This is where virtual memory enters the picture. Virtual memory frees up RAM by transferring data that hasn't been utilized in a while to a storage device like a hard disk or solid-state drive (SSD). Virtual memory is essential for increasing system speed, multitasking, and running huge applications. However, users should not depend too much on virtual memory since it is significantly slower than RAM. If the operating system needs to swap data between virtual memory and RAM too often, the machine will begin to slow down, a process known as thrashing.

- **Virtual memory was created at a period when actual memory, commonly known as RAM, was prohibitively costly. Because computers have a limited quantity of RAM, memory will ultimately run out if numerous applications are running at the same time. A virtual memory system emulates RAM by utilizing a part of the hard disk. A system may use virtual memory to load bigger or numerous applications at the same time, allowing each one to function as if it had more capacity without needing to buy extra RAM.**

How virtual memory works

- **Virtual memory operates using both hardware and software. When an application is running, data from that program is saved in RAM at a physical address. A memory management unit (MMU) transfers the address to RAM and translates addresses automatically. The MMU, for example, may link a logical address space to a physical address. If the RAM space is ever required for anything more important, data may be moved out of RAM and into virtual memory. The memory manager on the computer is in charge of keeping track of the transitions between physical and virtual memory. If the data is required again, the computer's MMU will continue execution through a context switch.**

- **The OS splits memory with a defined number of addresses into page files or swap files when moving virtual memory into physical memory. Each page is kept on a disk, and when a page is required, the operating system transfers it from the disk to the main memory and converts virtual addresses to actual addresses. However, the process of switching from virtual to physical memory is fairly sluggish. This implies that employing virtual memory typically results in a considerable decrease in performance. Because of swapping, computers with more RAM are thought to run better.**

Types of virtual memory

- The MMU of a computer controls virtual memory operation. The MMU hardware is often built into the central processor unit of modern computers (CPU). The virtual address space is also generated by the CPU. Virtual memory is often paged or split. Memory is divided into portions or paging files via paging. When a computer's available RAM is depleted, pages that are not in use are moved to the hard drive through a swap file. A swap file is a space on the hard drive reserved for usage as a virtual memory extension for the computer's RAM. When the swap file is required, it is returned to RAM through a process known as page swapping. This method guarantees that the operating system and apps on the computer do not run out of actual memory. The page file may have a maximum size of 12 to four times the computer's actual memory.

- **Page tables are used in the virtual memory paging process to transform the virtual addresses used by the OS and programs into the physical addresses used by the MMU. The page table entries show whether or not the page is in RAM. If the operating system or a program cannot locate what it requires in RAM, the MMU replies to the missing memory reference with a page fault exception, instructing the OS to transfer the page back to memory when it is required. When a page is loaded into RAM, its virtual address is recorded in the page table.**

- Virtual memory is also managed via segmentation. This method splits virtual memory into parts of varying lengths. Memory segments that are no longer in use may be relocated to virtual memory space on the hard disk. A segment table keeps track of whether or not a segment is existing in memory if it has been updated, and what its physical address is. Furthermore, file systems in segmentation are only made up of segments that are mapped into the possible address space of a process.
- In terms of how memory is partitioned, segmentation and paging vary as memory models; nonetheless, the operations may be coupled. Memory is organized into frames or pages in this situation. The segments occupy many pages, and the virtual address contains both the segment number and the page number.

- Other techniques of page replacement include first-in-first-out (FIFO), optimum algorithm, and least recently used (LRU). Memory selects the replacement for a page that has been in the virtual address for the longest period using the FIFO approach. The best algorithm technique chooses page replacements depending on which page is most likely to be replaced after the longest period of time; although complex to implement, this results in fewer page faults. The LRU page replacement technique replaces the page in the main memory that has not been utilized for the longest period.

How to manage virtual memory

- **Managing virtual memory inside an operating system is simple since there are default settings that define how much hard disk space to assign for virtual memory. These settings will work for most apps and processes, but there may be instances when you need to manually reset the amount of hard drive space allotted to virtual memory, such as when using applications that need rapid reaction times or when the computer has several hard disk drives (HDDs). The minimum and maximum amount of hard disk space to be utilized for virtual memory must be specified when manually resetting virtual memory.**

- **Allocating insufficient HDD space for virtual memory might cause a computer to run out of RAM. If a system often requires extra virtual memory space, it may be prudent to explore RAM expansion. Common operating systems may typically advise users not to expand virtual memory over 12 times the amount of RAM.**
- **The way virtual memory is managed varies depending on the operating system. As a result, IT workers must grasp the fundamentals of managing physical memory, virtual memory, and virtual addresses.**
- **RAM cells in SSDs have a finite lifetime as well. Because RAM cells have a finite amount of writes, utilizing them for virtual memory often lowers the drive's lifetime.**

What are the advantages of using virtual memory?

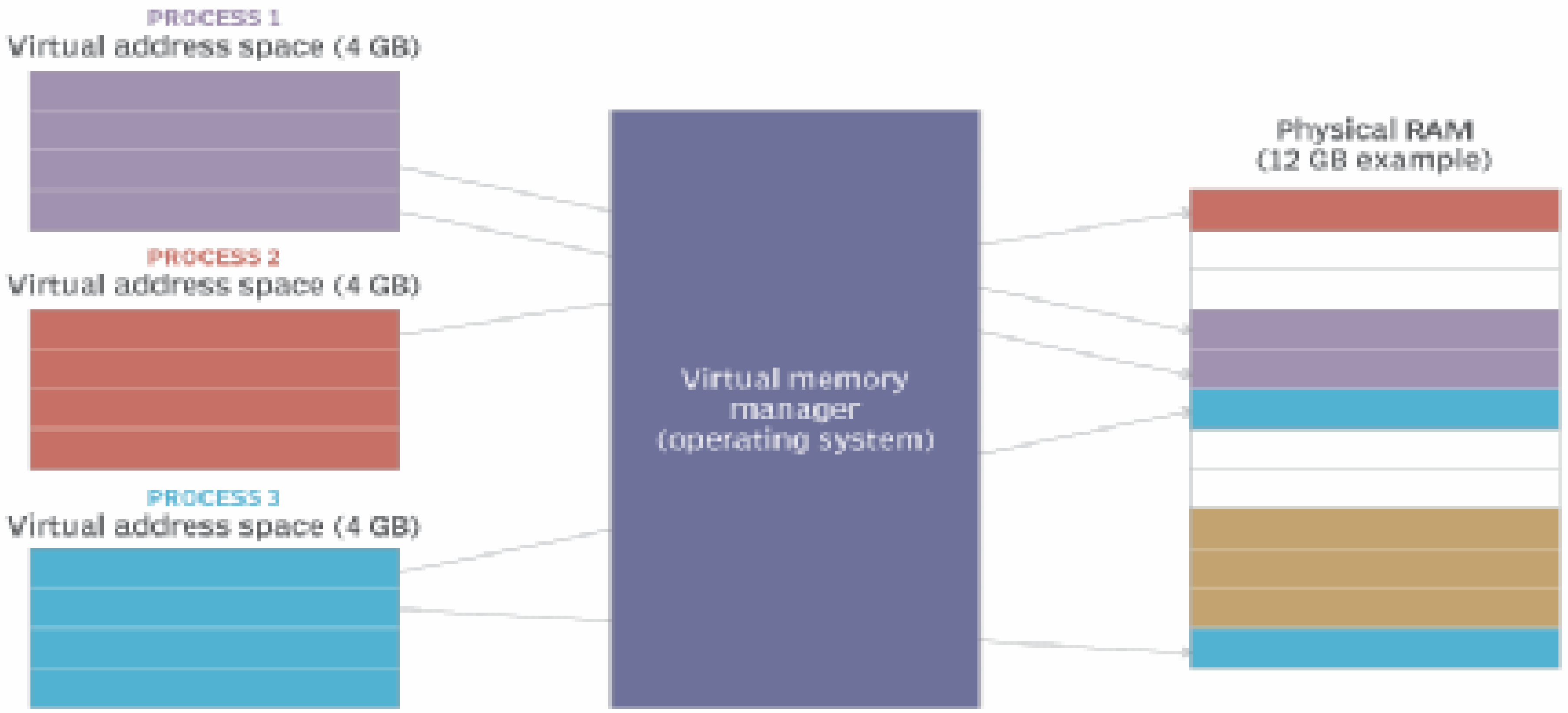
The benefits of utilizing virtual memory include:

- **it can handle twice as many addresses as the main memory**
- **it allows for more programs to be utilized at the same time.**
- **It relieves programs of the burden of maintaining shared memory and eliminates the need for users to add memory modules when RAM space runs out.**
- **It improves performance when just a portion of a program is required for execution;**
- **It improves security due to memory separation; and**
- **It allows numerous bigger programs to execute concurrently.**

- Allocating RAM is a low-cost operation.
- No external fragmentation is required.
- Using the CPU to manage logical partition workloads is beneficial.
- Data may be sent automatically.
- During a fork system call action that generates a clone of itself, pages in the original process may be shared.

In addition to these advantages, managers in a virtualized computing environment may employ virtual memory management methods to assign more memory to a virtual machine (VM) that has exhausted its resources. These virtualization management techniques help boost VM performance and administration flexibility.

The physical RAM after the OS organizes 3 processes



hitbullseye

What are the limitations of using virtual memory?

Although the use of virtual memory has its benefits, it also comes with some trade-offs worth considering, such as:

- **Applications run slower if they are running from virtual memory.**
- **Data must be mapped between virtual and physical memory, which requires extra hardware support for address translations, slowing down a computer further.**
- **The size of virtual storage is limited by the amount of secondary storage, as well as the addressing scheme with the computer system.**

- Thrashing can occur if there is not enough RAM, which will make the computer perform slower.
- It may take time to switch between applications using virtual memory.
- It lessens the amount of available hard drive space.

Virtual memory (virtual RAM) vs. Physical memory (RAM)

When talking about the differences between virtual and physical memory, the biggest distinction commonly made is to speed. RAM is considerably faster than virtual memory. RAM, however, tends to be more expensive.

When a computer requires storage, RAM is the first used. Virtual memory, which is slower, is used only when the RAM is filled.

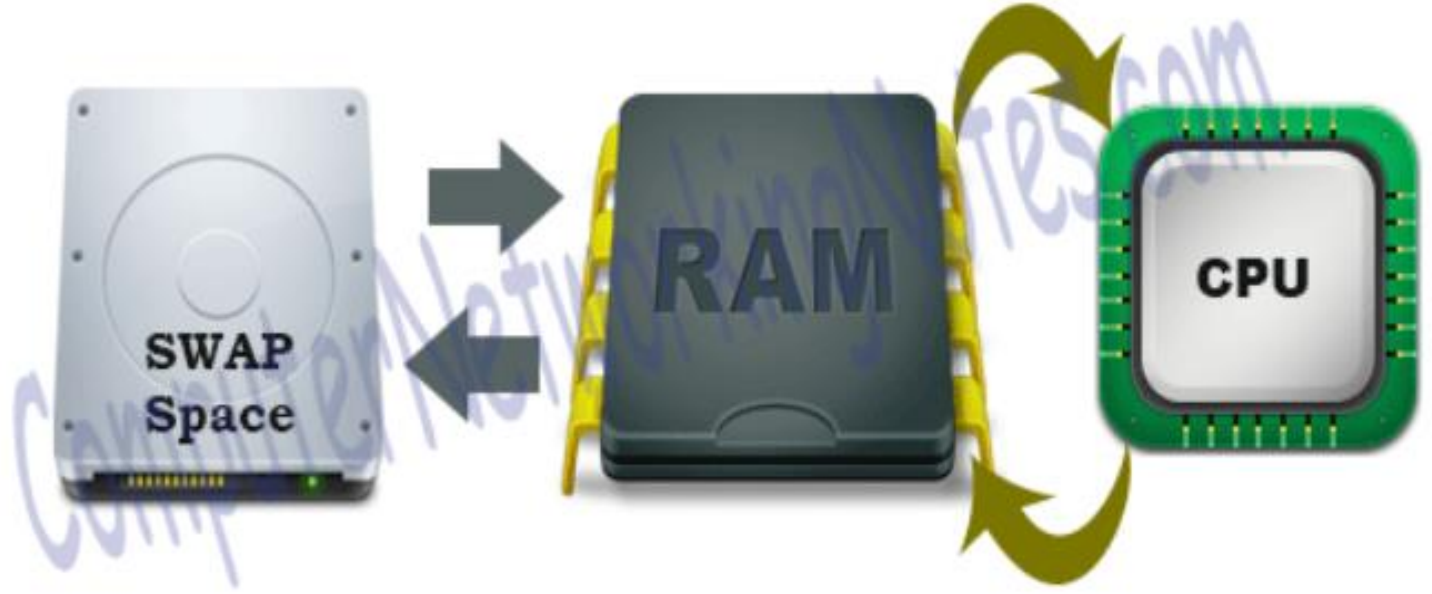
Virtual vs. Physical memory

Virtual Ram	Ram
Uses a segment of physical memory	Physical memory
Slower	Faster
Uses paging	Uses swapping techniques
Limited by the size of the physical memory	Limited to the size of the ram chip
Does not have direct access to the CPU	Can directly access the CPU
Limited by the size of the computer’s hard drive	Can add RAM by installing more RAM chips

- This graph compares virtual RAM (virtual memory) to RAM (physical memory). Users may actively increase RAM to their computers by purchasing and installing additional RAM chips. This is important if they are suffering slowdowns as a result of too many memory swaps. The quantity of RAM on a computer is determined by what is installed. The capacity of the computer's hard drive, on the other hand, limits virtual memory. Virtual memory settings are often regulated by the operating system.
- Furthermore, RAM employs swapping mechanisms, while virtual memory employs paging. Virtual memory is restricted by the capacity of the hard drive, while physical memory is limited by the size of the RAM chip. RAM has direct access to the CPU as well, but virtual RAM does not.

What is the swap space?

- Linux was traditionally designed for server systems. Server systems are designed to handle several tasks at the same time. Some processes may use more memory than planned. For example, a university website housed on a Linux server system may use more RAM on the day exam results are released. Memory spikes of this kind are fairly prevalent in server systems. During busy hours, any service may use more memory. To compensate for a lack of physical memory, Linux employs swap space. The swap space is hard drive space that may be utilized as memory.

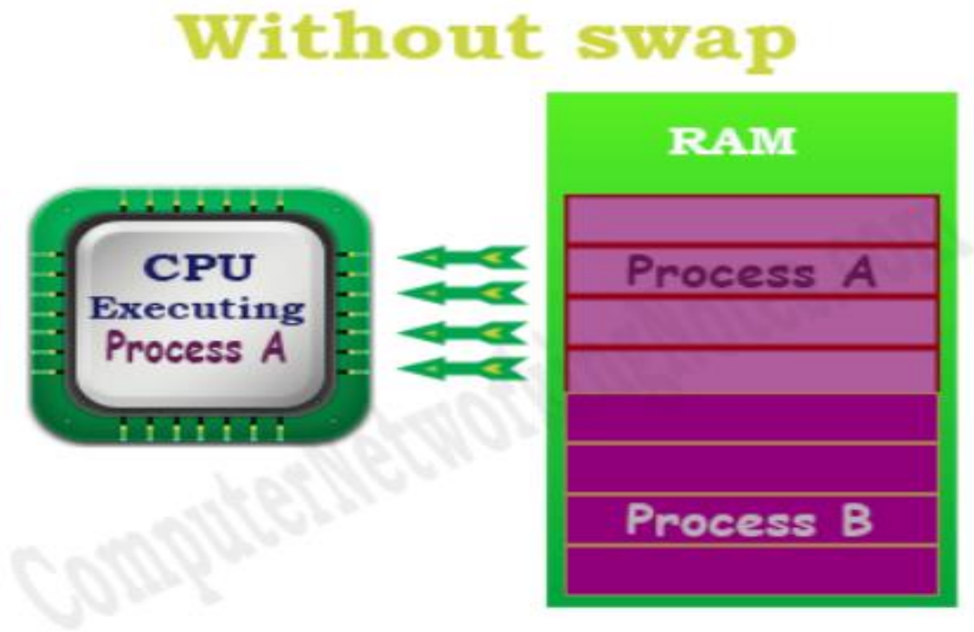


- Let's take a closer look at it. RAM is divided into smaller bits known as memory pages by the system. Memory pages are allotted to services or processes that are operating in the CPU. The CPU performs round-robin processing of services or applications.

Round-robin Fashion

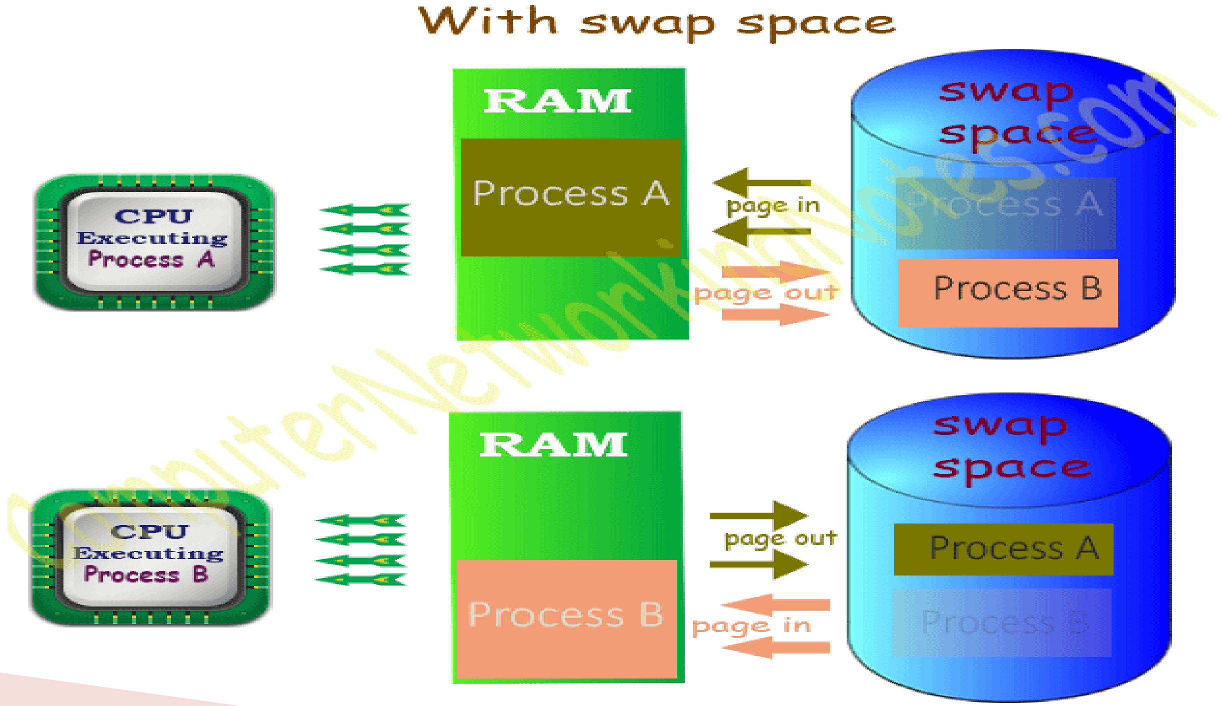
- All processes are scheduled in a round-robin method in a queue based on the time-sharing approach. Each process is also given a quantum (allowed CPU Time slot). If the process is not completed within the time limit, it will be canceled and restarted when the process's next time slot becomes available. For example, if the time slot is 100 milliseconds and job1 takes a total of 250 milliseconds to complete, the round-robin scheduler will suspend the job after 100 ms and allocate CPU time to other tasks.

- When all of the other tasks have received their fair share (100 ms each), job1 will be given another allotment of CPU time, and the cycle will begin again. This method is repeated until the work is completed and no more CPU time is required. Assume two processes are running. Memory pages are allotted to both processes. Process one is being processed by the CPU. This situation is shown in the image below.



- As seen in the diagram above, no memory pages are accessible for another process. The system will not begin a new process until any current processes are completed. This problem can be somewhat improved using swap space. Idle pages will be shifted into swap space if swap space is specified and RAM is full.

- When process A is running, memory pages occupied by process B are relocated to swap space. When process B executes, memory pages held by process A are relocated to swap space. This allows Linux to run more processes using swap space while utilizing the same physical memory. A page out is the process of shifting idle memory pages from RAM to swap. A page in is the process of returning necessary memory pages from swap to RAM. This procedure is shown in the diagram below.



Key points

- Swap space is hard drive space used to augment system RAM by storing idle memory pages.
- Because swap space is stored on the hard drive, it is much slower than RAM.
- Swap space is a critical component in Linux setup.
- If swap space is enabled, a physical memory threshold value is also enabled. When physical memory is utilized over the threshold, swap space is used.
- No swap space will be utilized as long as free memory stays below the threshold.
- Swap utilization should be regularly checked. If swap becomes heavily utilized, the system's performance will suffer.

THANK YOU