

Memory Management

Operating System Fundamentals

- It Is The Functionality Of The Operating System Which Manages The 'Main Memory' And Keep Track Of Process Moving From 'Secondary Memory' To 'CPU' And Vice Versa.
- **Functionality :**
 - Keep Track Of What Memory Is In Use And What Memory Is Free
 - Allocate Memory To A Process When Memory Is Free And DeAllocate Memory When They Do Not Required.
 - Managing The Transfer Of Memory - Main Memory And Secondary Memory
 - Prohibiting process Program To Enter In Others process Program Area And Operating System Area.
 - Keep Updating Status Of Process
 -

GOALS :

- Maximize CPU Utilization
- Minimize Response Time
- Maximize Memory Management
- Prioritize Important Process
- Protection To process Program And OS

- **Logical Address:**

- Logical Address is generated by CPU while a program is running.
- The logical address is virtual address as it does not exist physically, therefore, it is also known as Virtual Address.
- This address is used as a reference to access the physical memory location by CPU.
- The term Logical Address Space is used for the set of all logical addresses generated by a program's perspective.
- The hardware device called Memory-Management Unit is used for mapping logical address to its corresponding physical address.

- **Physical Address:**

- Physical Address identifies a physical location of required data in a memory.
- The process never directly deals with the physical address but can access by its corresponding logical address.
- The process program generates the logical address and thinks that the program is running in this logical address, but the program needs physical memory for its execution, therefore, the logical address must be mapped to the physical address by MMU before they are used.
- The term Physical Address Space is used for all physical addresses corresponding to the logical addresses in a Logical address space.

Sr. No.	Logical memory address	Physical Memory address
1	generated by CPU	location in a memory unit (Computed by MMU)
2	Logical Address Space is set of all logical addresses generated by CPU in reference to a program.	Physical Address is set of all physical addresses mapped to the corresponding logical addresses.
3	The process can use the logical address to access the physical address.	The process can indirectly access physical address but not directly.

- Process is getting divided into equal size of pages (virtual memory pages) and mapped with main memory frames (physical memory pages).
- The mapping from virtual to physical address is done by the memory management unit (MMU) which is a hardware device and this mapping is known as paging technique.
- The Physical Address Space is conceptually divided into a number of fixed-size blocks, called frames.
- The Logical address Space is also splitted into fixed-size blocks, called pages.
- **Page Size = Frame Size i.e. size of page in process (generated by CPU) = size of frame in main memory**

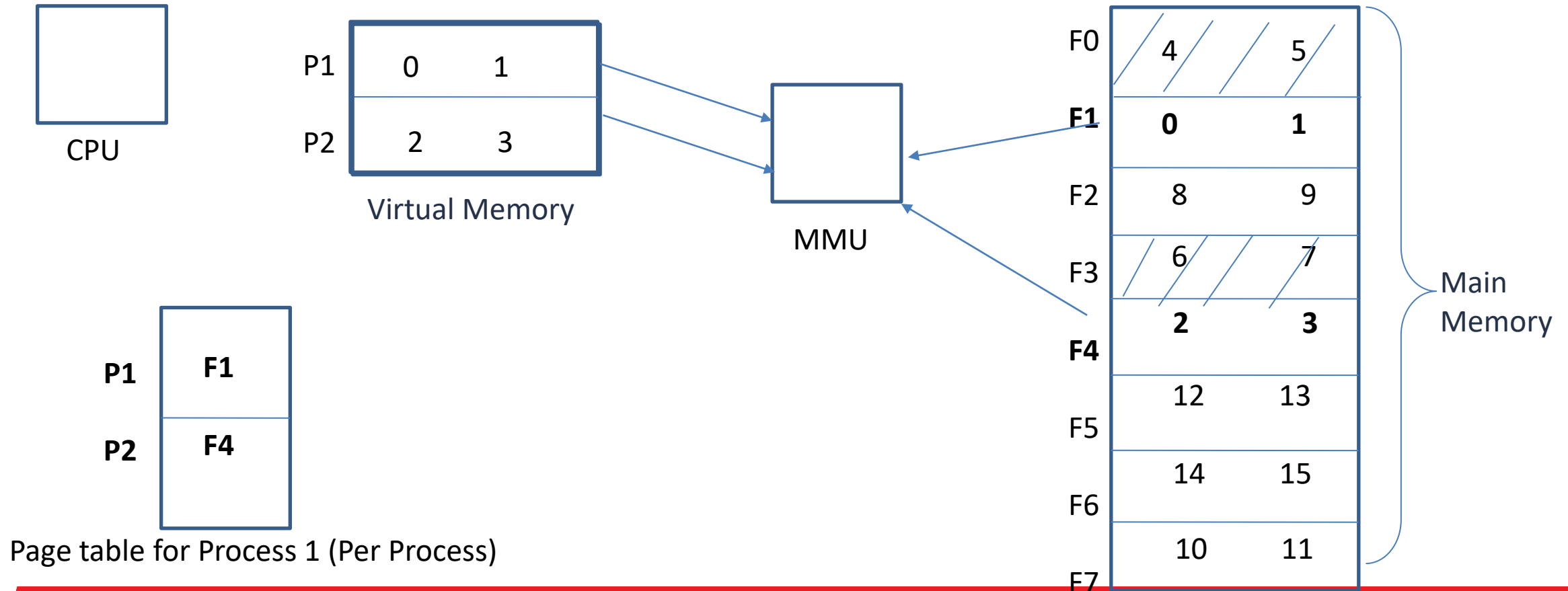
Example

Let's say we have process P1 having size 4B and main memory having size 16B.

We are aware that Page size=frame size=2B. so size required in main memory to store this process will be 2 frames each having size of 2B.

M/M size = 16B and frame size=2B so No. of frames=16B/2B=8 frames.

Process size=4B and page size = 2B so No. of pages for this process will be 4B/2B=2 pages.



- **Virtual Memory = RAM + SWAP**
- Virtual Memory is a storage allocation scheme in which secondary memory can be addressed as though it were part of main memory.
- **Advantages:**
 - Useful when process size is more than the size of main memory
 - Useful in multi-processing i.e. multiple number of processes can be in main memory using page replacement strategy.
- **Page Fault:** Page demanded by CPU for processing is absent in main memory.
- **Page fault service steps:** after page fault Kernel generates interrupt and mark that process in waiting state.
- Kernel looks it into SWAP area and if found place it into main memory, updates the process table and for execution.

- **Swapping:**

- Swapping a process out means removing all of its pages from memory, or marking them so that they will be removed by the normal page replacement process.
- Suspending a process ensures that it is not runnable while it is swapped out.
- At some later time, the system swaps back the process from the secondary storage to main memory.

- **Paging:**

- Paging is the procedure of memory allocation where different non-contiguous blocks of memory are assigned a fixed size.
- The size is generally of 4KB. The paging is always performed between active pages.

Difference between Paging and Swapping :

Swapping	Paging
It is procedure of copying out the entire process.	It is a technique of memory allocation.
Swapping occurs when whole process is transferred to disk.	Paging occurs when some part of process is transferred to disk.
Swapping is done for inactive processes.	Only active process can perform paging.

Major Fault and Minor Fault



- A minor fault means the page is in memory but not allocated to the requesting process or not marked as present in the memory management unit.
- A major fault means the page is no longer in memory.

Commands to check memory



- `free -m`
- `swapon -s`
- `vmstat 5 10` (5 updates for every 10 secs.) Need to observe columns free, si (swap in) and so (swap out))