Unit-05 : Memory Management

- Introduction to Memory management,
- Contiguous allocation,
  - Fixed Partition,
  - Dynamic partition,
- Non-contiguous allocation –
  - paging,
  - segmentation.
- Introduction to Virtual-memory management –
- Demand paging,
- Cop-on-write,
- page replacement,
- Allocation of frames,
- Thrashing.

Questions to be discussed :

1. Describe memory management. How memory management is required ?
2. Compare between internal and external fragmentation.
3. What are the difference between static & dynamic memory allocation ?
4. Define logical address and physical address.
5. Differentiate between paging and segmentation ?
6. What are the difference between paging and segmentation ?
7. What is virtual memory ? Discuss the benefits of virtual memory technique in details.
8. What is demand paging ? Describe the process of how can demand paging can be implemented with virtual memory.
9. What is page faults ? Explain in detail the step involve in handling a page fault.
10. What is thrashing ? What is the cause of thrashing ? Also discuss how does the system detect thrashing.
11. Explain different types of page replacement policy.
    a) First In First Out (FIFO)
    b) Least Recently Used (LRU)
    c) Optimal Page Replacement
12. Concider the reference string :

    7, 0, 1, 2, 0, 3, 0,4, 2, 3, 0, 3, 2, 1, 2, 0, 1, 7, 0, 1 for a memory with three frames. Trace FIFO, Optimal and LRU page replacement algorithms.
13. Write short notes on following :
    a) Cop-on write
    b) Demand paging

What is memory management ?

- The task of dividing the memory among different processes is called memory management.
- It is the function responsible for allocating and managing main memory of computer.
- It keeps track of each and every memory location, either it is allocated to some process or free.
- It checks how much memory is to be allocated to processes.
- It decides which process will get memory at what time.
- It tracks whenever some memory gets freed or unallocated then it updates the status.
- The main aim of memory management is to achieve efficient utilization of memory.

Why Memory Management is required ?

- To minimize fragmentation issues.
- To proper utilization of main memory.
- To keep track of used memory space by processes.
- To maintain data integrity while executing of process.
- Allocate and de-allocate memory before and after process execution.

Function of memory management :

- Allocation
- Protection
- Free space management
- Deallocation

Goals :

- Effective utilization of memory space.
- Run larger program in smaller memory.

There are two types of memory management techniques :

1. Contiguous(process stored in continuous manner)
2. Non-contiguous(process stored at different-different location)

Memory management techniques

Contiguous

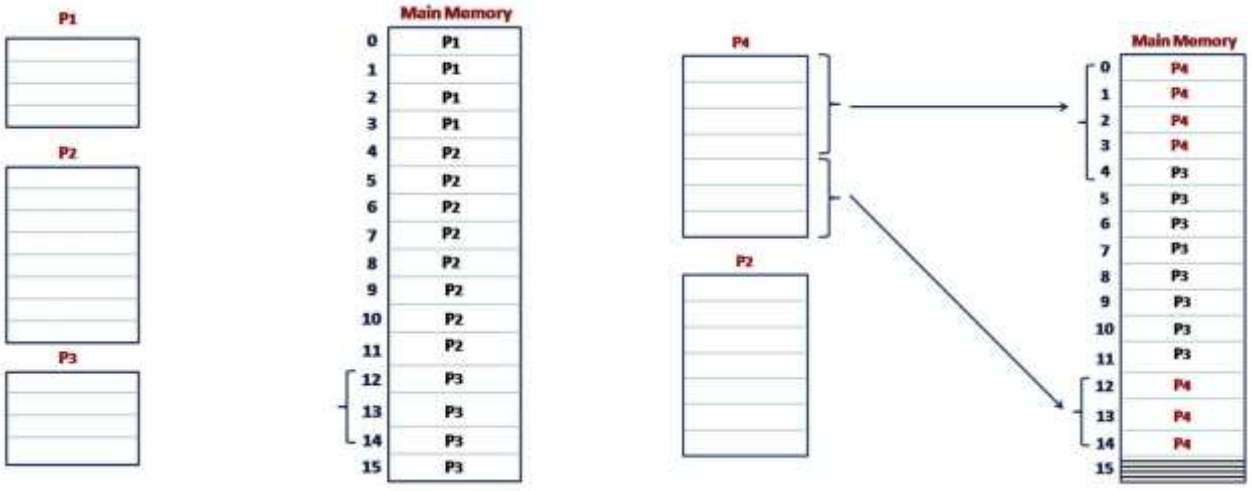
- Partitioning
- Overlays

Non-contiguous

- Paging
- Segmentation
- Vertual memory

Difference between Contiguous & Noncontiguous Memory Allocation :

Contiguous Memory Allocation	Non-Contiguous Memory Allocation
It allocates consecutive blocks of memory.	It allocates non-consecutive blocks of memory.
Memory space present at same location.	Memory space present at different location.
Faster in Execution.	Slower in Execution.
It is easier for the OS to control.	It is difficult for the OS to control.
Both Internal fragmentation and external fragmentation occurs.	Only External fragmentation occurs.

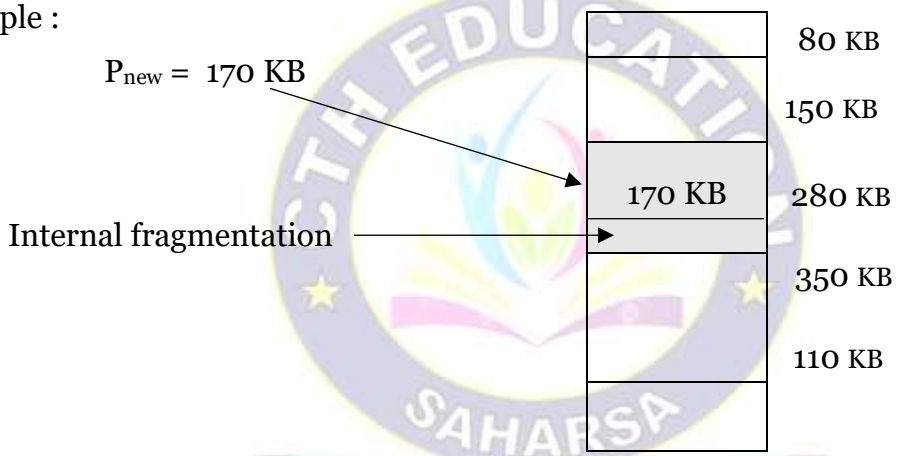


What is fragmentation ?

- Wastage of memory space within the computer is called fragmentation.
- There are two types of fragmentation :
  1. Internal fragmentation
  2. External fragmentation

Internal fragmentation :

- It happens when the process assigned is smaller than the memory.
- It is the difference between allocated memory to a process and required memory.
- If the memory block assigned to process is bigger then some portion of memory is left unused and it cannot be used by another process.
- The wastage of memory in this situation is known as internal fragmentation.
- Example :



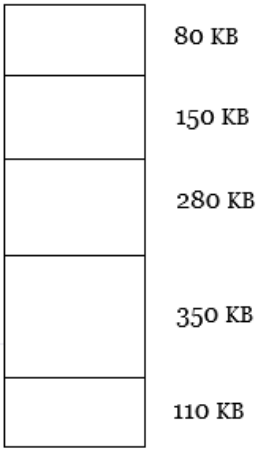
External fragmentation :

- If total memory space is enough to satisfy a request of a process , but we can't accommodate it into the memory because it is not contiguous, so it cannot be used.
- The wastage of memory in this situation is known as external fragmentation.

Example :

P_{new} = 360 KB

Here, external fragmentation occurs because, A new process of size 360 KB can not able to Accommodate while total 870 KB memory size is available.



Compare between internal and external fragmentation.

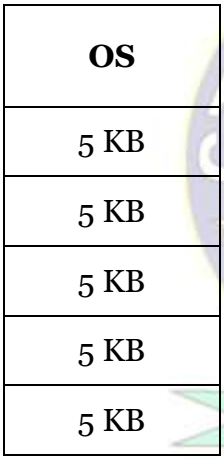
Internal fragmentation(IF)	External fragmentation(EF)
The difference between allocated memory and required memory is called IF.	The unused spaces are too small to accommodate a new process, is called EF.
It happens when the process is smaller than the memory.	It happens when the process is not satisfy any big enough free memory to accommodate it.
It occurs with fixed partitioning.	It occurs with dynamic partitioning.
It occurs in worst fit memory allocation method.	It occurs in best fit and first fit memory allocation method.
The solution of internal fragmentation is best-fit allocation algorithm.	Solution of external fragmentation is compaction.

There are two types of contiguous memory allocation technique :

1. Fixed (static) partitioning
2. Variable (dynamic) partitioning

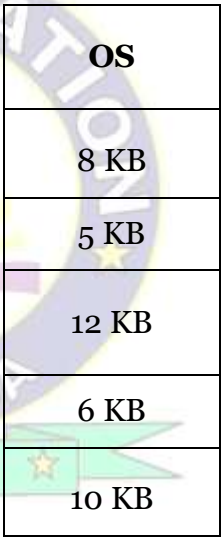
Fixed partitioning :

- It is a contiguous memory management technique.
- In this method the number of partitions are fixed.
- Size of partition may or may not be same.
- It is also known as static partitioning.
- It is the oldest and simplest technique.
- It is a contiguous allocation so, spanning is not allowed.
- Here partitions are made before execution or during system configure.



Equal size partitions

No. of partitions = fixed



Unequal size partitions
No. of partitions = fixed

Advantages of Fixed Partitioning :

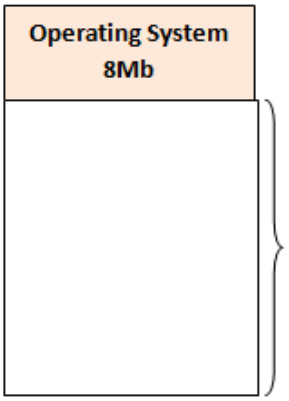
- Easy to implement

Disadvantages of Fixed Partitioning :

- Internal Fragmentation
- External Fragmentation
- Limit process size
- Limitation on Degree of Multiprogramming:

Variable Partitioning :

- It is a contiguous memory management technique in which the main memory is not divided into partitions before coming process in memory.
- In this method the memory is allocated to the process according to their need.
- It is also known as dynamic partitioning.
- Here, internal fragmentation is not occurs but it suffer from external fragmentation.
- It also provides the concept of compaction to remove external fragmentation.
- In compaction the spaces that are free and the spaces which not allocated to the process are combined and single large memory space is made.



Advantages of variable partitioning :

- No Internal Fragmentation
- No restriction on Degree of Multiprogramming
- No Limitation on the size of the process

Disadvantages of variable vartitioning :

- Difficult Implementation
- External Fragmentation

Difference between fixed & variable partitioning :

Fixed partitioning	Variable partitioning
Memory is divided into fixed sized partitions.	Memory is not divided into fixed sized partitions.
It is also known as static partitioning.	It is also known as dynamic partitioning.
Only one process can be placed in partition.	Memory is allocated according to need of process.
It does not utilize main memory effectively.	It utilizes the main memory effectively.
Internal and external fragmentation occurs.	Only external fragmentation occurs.
Limitation on degree of multi-programming.	No limitation on degree of multiprogramming.
It is more easier to implement.	It is difficult to implement.
There is limitation on size of process.	There is no limitation on size of process.

Partition allocation policies :

- There are different placement/aloocation Algorithm :
  1. First Fit
  2. Best Fit
  3. Worst Fit
  4. Next Fit

First Fit :

- Allocate the process in the first free big-enough of the partition.
- It scans memory from the beginning and choose the first available block that is big enough.

Example :

P_{new} =100 KB

80 KB
150 KB
280 KB
350 KB
110 KB

Best Fit :

- Allocate the process to the partition which is the first smallest sufficient partition among the free available partition.
- It searches the entire list of holes to find the smallest hole whose size is greater than or equal to the size of the process.

Example :

P_{new} =100 k

80 k
150 k
280 k
350 k
110 k

Worst Fit :

- Allocate the process to the partition which is the largest sufficient among the freely available partitions available in the main memory.
- It searches the entire list of holes to find the largest hole and allocate it to process.
- It is opposite to the best-fit algorithm.

Example :

P_{new} =100 k

80 k
150 k
280 k
350 k
110 k

Next Fit :

- Next fit is similar to the first fit but it will search for the first sufficient partition from the last allocation point.

Example :

P_{new} =100 k

80 k
150 k
280 k(Used)
350 k
110 k

Q. Consider the requests from processes in given order 300K, 25K, 125K, and 50K. Let there be two blocks of memory available of size 150K followed by a block size 350K.

Which of the following partition allocation schemes can satisfy the above requests?

A) Best fit but not first fit.

B) First fit but not best fit.

C) Both First fit & Best fit.

D) neither first fit nor best fit.

What is Overlays ?

- The main problem in fixed partitioning is the size of a process has to be limited.
- To solve this problem, earlier people have used some solution which is called as Overlays.
- So overlay is a technique to run a larger program in smaller size of memory by keeping only those instructions and data that are needed at any given time.
- Divide the program into modules in such a way that not all modules need to be in the memory at the same time.
- The concept of overlays is that whenever a process is running it will not use the complete program at the same time, it will use only some part of it.
- Then overlays concept says that whatever part you required, you load it and once the part is done, then you just unload it.

What is address space ?

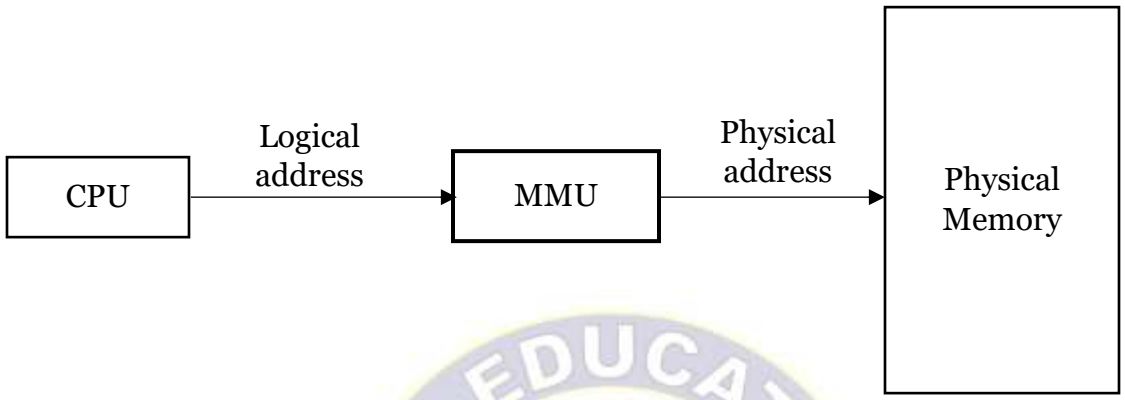
- Each & every instruction or data are stored at a particular address in memory.
- The set of all addresses is known as the address space.
- The address space is the set of addresses generated by programs.
- There are two types of address space :
    1. Logical address space
    2. Physical address space

Difference between logical & physical address :

LOGICAL ADDRESS	PHYSICAL ADDRESS
It is generated by CPU.	Actual location in main memory unit.
User can access the logical address.	User can never access physical address.
The user can use the logical address to access the physical address.	The user can indirectly access physical address but not directly.
Logical address can be change.	Physical address will not change.
Generated by the CPU	Computed by MMU
It is also known as virtual address.	It is also known as real address.
Logical Address Space is set of all logical addresses generated by CPU.	Physical Address is set of all physical addresses mapped to the corresponding logical addresses.

Memory Management Unit(MMU) :

- MMU is a hardware device.
- It convert the logical address into the physical address during the program execution.
- It is located within the Central Processing Unit.
- A memory management unit is also known as a paged memory management unit.



What is Paging ?

- Paging is the concept of non-contiguous memory management technique.
- The main purpose of paging is to divide each process in the form of equal size pages & load them into the main memory.
- In paging process address space is broken into blocks of the same size pages.
- The main memory will also be divided in the form of frames.
- In this mechanism the page size and frame size must be same.

Advantages and Disadvantages of Paging :

- Paging reduces external fragmentation, but still suffer from internal fragmentation.
- Paging is simple to implement and assumed as an efficient memory management technique.
- Due to equal size of the pages and frames, swapping becomes very easy.
- Page table requires extra memory space, so may not be good for a system having small RAM.

What is Segmentation ?

- Segmentation is the concept of non-contiguous memory management technique.
- In this, the process is divided into the variable size segments.
- The details about each segment are stored in a table called as segment table.
- Segment table contains base address & length of the segments.

Advantages of Segmentation

- No internal fragmentation
- Average Segment Size is larger than the actual page size.
- It is easier to relocate segments than entire address space.
- The segment table is of lesser size as compare to the page table in paging.

Disadvantages

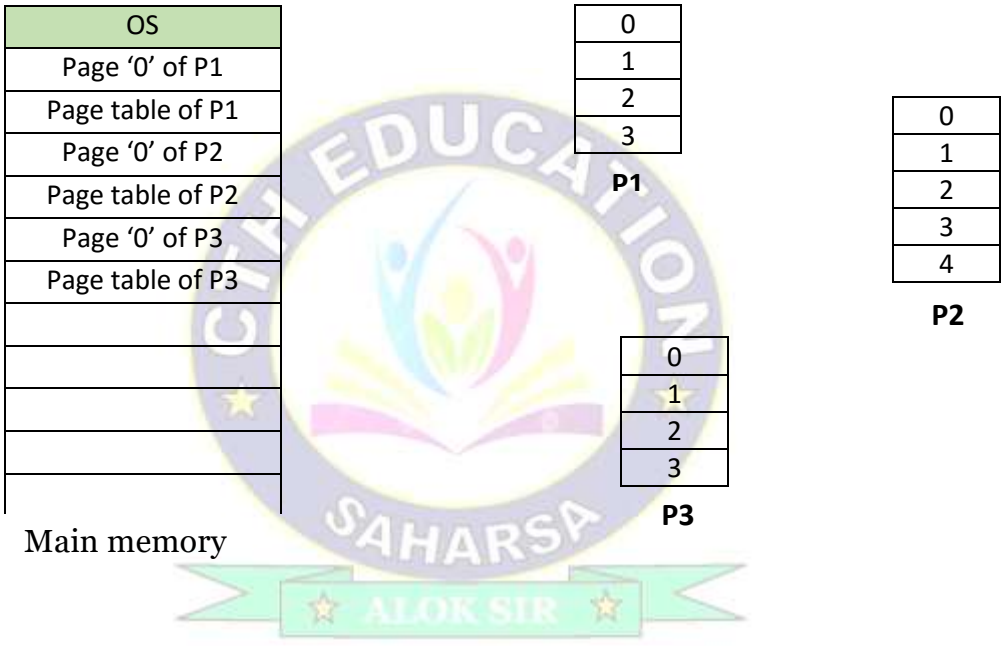
- It can have external fragmentation.
- It is difficult to allocate contiguous memory to variable sized partition.
- Costly memory management algorithms.

Difference between paging and segmentation :

Paging	Segmentation
Program is divided into fixed size called pages.	Program is divided into variable size called frames.
For paging operating system is accountable.	For segmentation compiler is accountable.
Paging is faster.	Segmentation is slow.
Internal fragmentation occurs in paging.	Here, External fragmentation is occurs.
In paging, logical address is split into page number and page offset.	Here, logical address is split into segment number and segment offset.
Paging is invisible to the user.	Segmentation is visible to the user.
In paging, processor needs page number, offset to calculate absolute address.	In segmentation, processor uses segment number, offset to calculate full address.

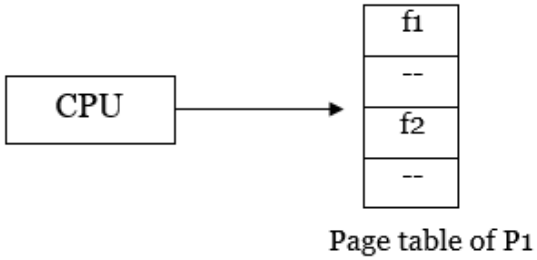
Virtual memory :

- It provides the illusion to the programmer that a process whose size is bigger than the size of main memory cal also be executed.
- It is a storage scheme that provides user an illusion of having a very big main memory.
- Virtual memory increase the degree of multiprogramming.
- In this concept, load only required pages of process in the main memory.
- It is implemented by demand paging.
- This concept is handled by operating system.
- Now a days, virtual memory concept uses almost every computer system.



Advantage :

- Illusion of large main memory.
- Increase degree of multiprogramming.

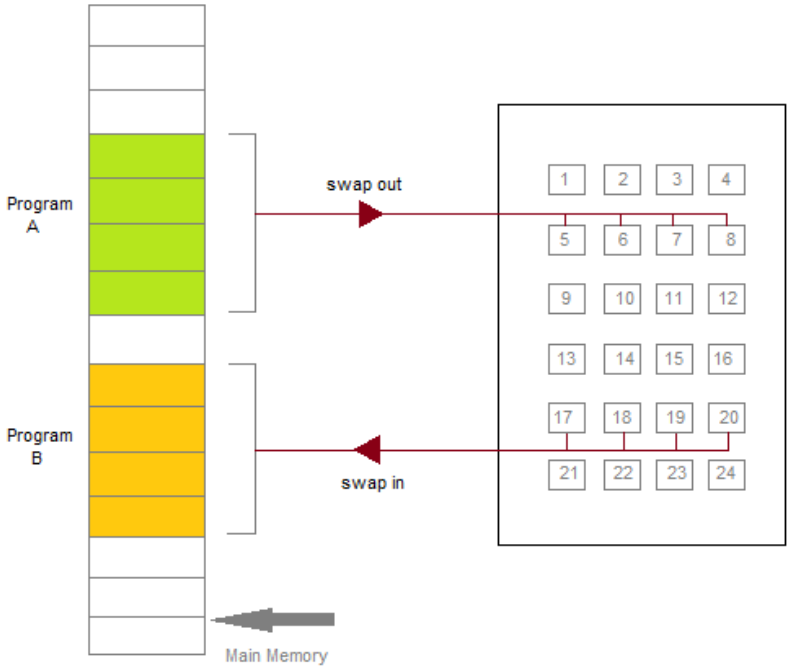


Disadvantage

- It may increase page fault
- The system becomes slower since swapping takes time.
- It takes more time in switching between applications.

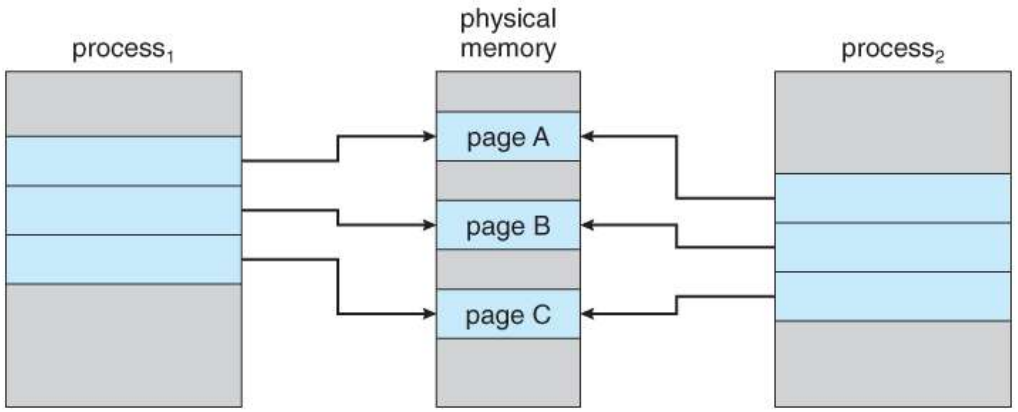
Demand paging :

- This concept tells that not load any page until it is required or request by CPU.
- The process of loading the page into memory on demand (whenever page fault occurs) is known as demand paging.
- If the CPU tries to access page that is currently not available in the main memory, it generates a request.
- The OS will search required page in the logical address space.
- The required page will be brought from logical address space to physical address space.
- The page replacement algorithms are used for the decision-making of replacing the page in physical address space.



Copy on Write(COW) :

- Copy on Write is a resource management technique.
- In OS, fork() system call creates a duplicate process of the parent process which is called as the child process.
- When a parent process creates a child process then both of these processes initially will share the same pages in memory.
- These shared pages will be marked as copy-on-write which means that if any of these processes will try to modify the shared pages then only a copy of these pages will be created and the modifications will be done.



Page replacement

- It is a process of swapping out an existing page from the frame of a main memory and replacing it with the required page.
- Page replacement algorithms help to decide which page must be swapped out from the main memory.
- Various page replacement algorithms are :
  1. FIFO Page Replacement Algorithm
  2. LRU Page Replacement Algorithm
  3. Optimal Page Replacement Algorithm

Note : A good page replacement algorithm is one that minimizes the number of page faults.

FIFO Page Replacement Algorithm :

- This algorithm works on the principle of "**First in First out**".
- It replaces the oldest page that has been present in the main memory for the longest time.

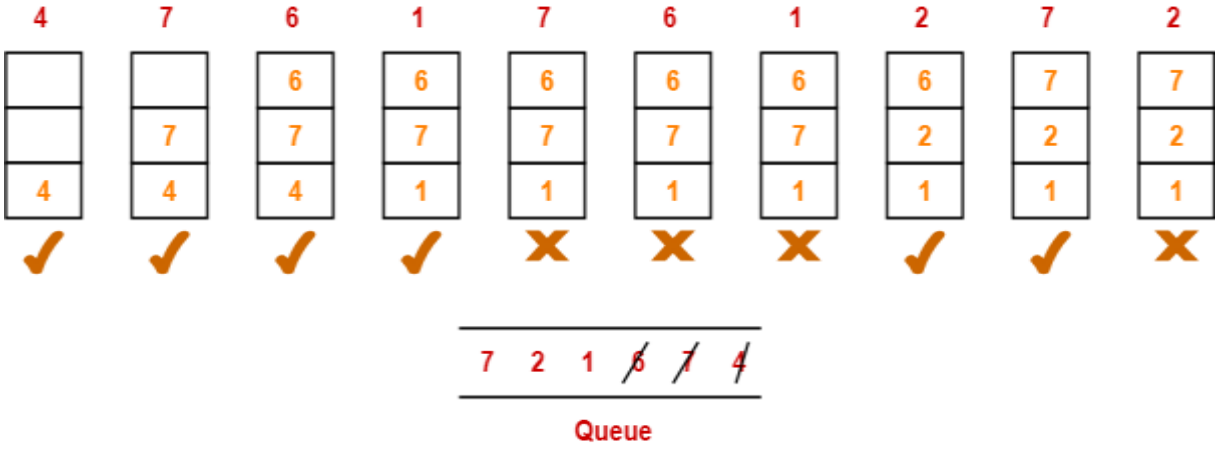
Problem-01: A system uses 3 page frames for storing process pages in main memory. It uses the FIFO page replacement policy. Assume that all the page frames are initially empty. What is the total number of page faults that will occur while processing the page reference string given below-

4 , 7, 6, 1, 7, 6, 1, 2, 7, 2

Also calculate the hit ratio and miss ratio.

Solution :

Total number of references = 10



Total number of page faults occurred = 6

Calculating Hit ratio :

Total number of page hits = Total number of references – Total number of page faults

= 10 – 6 = 4

Hit ratio = Total number of page hits / Total number of references

= 4 / 10

= 0.4 or 40%

Calculating Miss ratio :

Miss ratio = 1 – Hit ratio

= 1 – 0.4

= 0.6 or 60%

Optimal Page Replacement Algorithm :

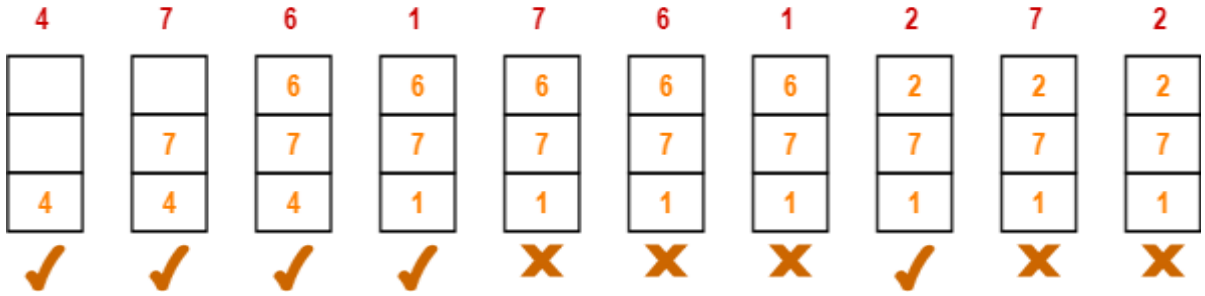
- This algorithm replaces the page that will not be referred by the CPU in future for the longest time.
- It is the best known algorithm and gives the least number of page faults.
- It is practically impossible to implement this algorithm.

Problem-02 : A system uses 3 page frames for storing process pages in main memory. It uses the Optimal page replacement policy. Assume that all the page frames are initially empty. What is the total number of page faults that will occur while processing the page reference string given below-

4 , 7, 6, 1, 7, 6, 1, 2, 7, 2

Also calculate the hit ratio and miss ratio.

Solution : Total number of references = 10



Total number of page faults occurred = 5

- Hit ratio = 0.5 or 50%
- Miss ratio = 0.5 or 50%

LRU Page Replacement Algorithm :

- This algorithm works on the principle of "**Least Recently Used**".
- It replaces the page that has not been referred by the CPU for the longest time.

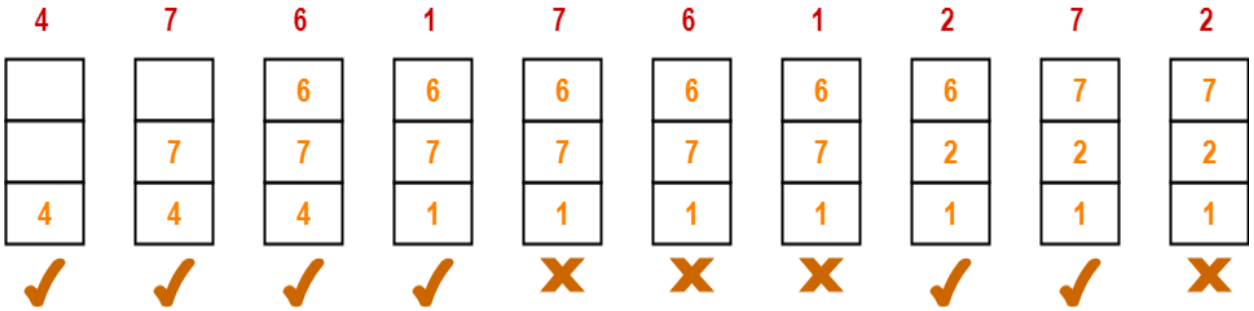
Problem-03 : A system uses 3 page frames for storing process pages in main memory. It uses the LRU page replacement policy. Assume that all the page frames are initially empty. What is the total number of page faults that will occur while processing the page reference string given below-

4 , 7, 6, 1, 7, 6, 1, 2, 7, 2

Also calculate the hit ratio and miss ratio.

Solution :

Total number of references = 10



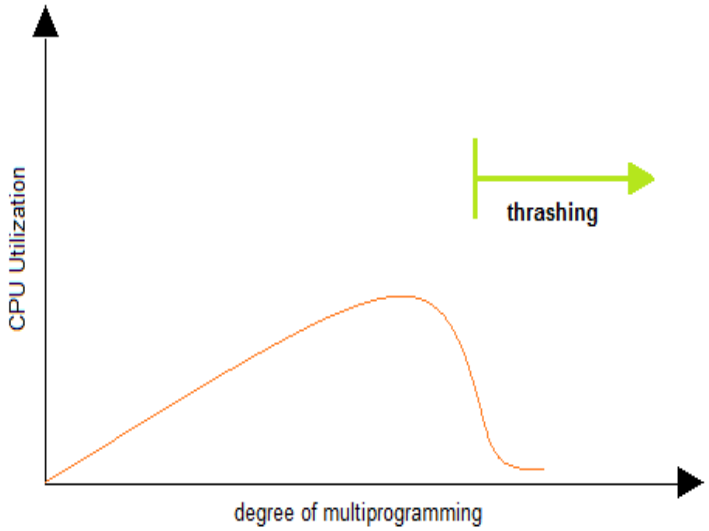
Total number of page faults occurred = 6

In the similar manner as above-

- Hit ratio = 0.4 or 40%
- Miss ratio = 0.6 or 60%

What is thrashing ?

- A high page fault rate is called thrashing.
- It is also an undesirable state of the system like deadlock.
- Lack of memory(frames) cause thrashing.
- Thrashing occurs due to increasing of degree of multiprogramming.
- To prevent thrashing we must provide processes with as many frames as they really need "right now".



Causes of Thrashing :

- High degree of multiprogramming.
- Lacks of Frames.

Recovery of Thrashing :

- Do not allow the system to go into thrashing by instructing the long-term scheduler not to bring the processes into memory after the threshold.
- If the system is already thrashing then instruct the mid-term scheduler to suspend some of the processes so that we can recover the system from thrashing.