Module 4

Memory Management

- How can main/primary memory be directly accessed?

  by the central processing unit (CPU)

- Main memory is volatile, meaning

  data is stored temporarily and can be lost in the event of a power failure

- What is secondary memory?

  a physical storage device to permanently store data, regardless of the computer being turned on or off

  includes hard drives and external media like disks, tapes, flash drives

  this memory is slower to access but there is much more storage capacity

- Every program and file accessed from must be loaded from where?

  the secondary memory into the primary memory

- What requirements must memory management address?
  - Relocation
    - Active processes are swapped into and out of the main memory, to make room for other processes, and it is not always possible to return a specific process into the same main memory location it was in before, since that location may be occupied by another process.
    - The process needs to be relocated, keeping track of the new location, or address.
  - Protection

    - Every process must be protected against interference from other processes
- Sharing
    - There needs to be controlled access to shared areas of memory while maintaining protection
- Logical organization
    - An OS and computer hardware must deal with a different organizational structure for main memory
- Physical organization
    - The essence memory management is the task of moving information between the primary and secondary memory

▼ What are common techniques for loading multiple processes into main memory?

- fixed partitioning
- dynamic partitioning

▼ What is fixed partitioning?

- partitioning with a specific (fixed) number of partitions of either all the same size or varying sizes
- the problem of “internal fragmentation” arises when the partition is larger than what is loaded into it, resulting in wasted, unused space

▼ What is dynamic partitioning?

- it solves internal fragmentation by not creating partitions until the process is run, so the partition is sized according to the need of the process
- efficient use of memory
- can accommodate large processes
- however, this makes for more complex processing during a program's run time to have to allocate the memory for the partitions, and also creates “holes” in the memory, as processes are swapped into partitions that were created for a previous, larger process

- over time, this leads to many small gaps that cannot accommodate enough processes, so the active processes need to be shifted to free up more space for another partition, but that takes up additional time and system resources

▼ How can internal and external fragmentation can be fixed?

by dividing both memory and processes into small partitions and using reference tables to keep track of these partitions and the location of the processs in memory

▼ What is paging (fixed size)?

- frames: the fixed-sized portions or blocks of physical memory
- pages: the fixed-sized blocks of logical memory
- To run a process, the process pages from logical memory space are loaded into the frames of physical memory address space, with the page number used as part of the address for accessing the frame. This information is stored in the page table.

▼ What is segmentation (variable size)?

- Each process is divided into variable size segments and loaded to the logical memory address space.
- To run a process, the segments from logical memory space are loaded to the physical memory space, storing the mapping of the logical address to the physical address in the segment table.