Memory Virtualization

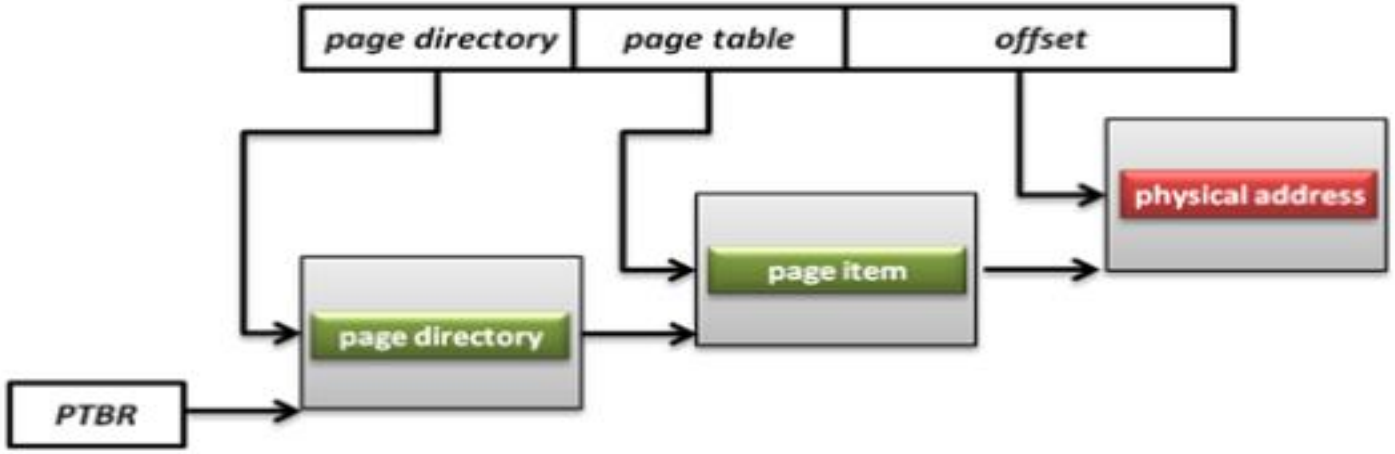
I/O Virtualization

Memory Virtualization

- Memory management in OS
  - Traditionally, OS fully controls all physical memory space and provide a continuous addressing space to each process.
  - In server virtualization, VMM should make all virtual machines share the physical memory space without knowing the fact.
- Goals of memory virtualization :
  - Address Translation
    - Control table-walking hardware that accesses translation tables in main memory.
  - Memory Protection
    - Define access permission which uses the Access Control Hardware.

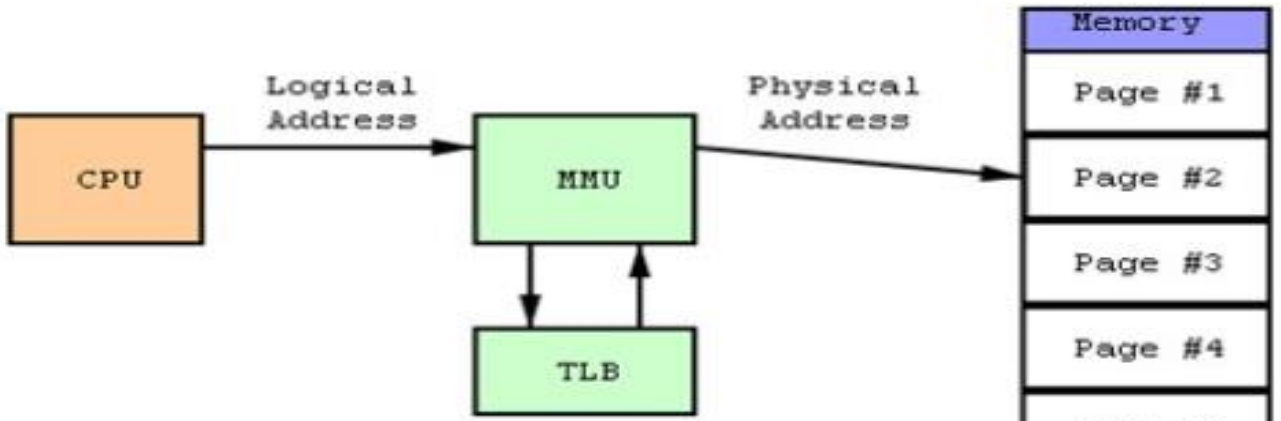
Memory Architecture

- Memory Management Unit (MMU)
  - What is MMU ?
    - A computer hardware component responsible for handling accesses to memory requested by the CPU.
    - Its functions include translation of virtual addresses to physical addresses, memory protection, cache control, bus arbitration and etc.
  - What is PTBR ?
    - Page Table Base Register (PTBR) is a register point to the base of page table for MMU.



Memory Architecture

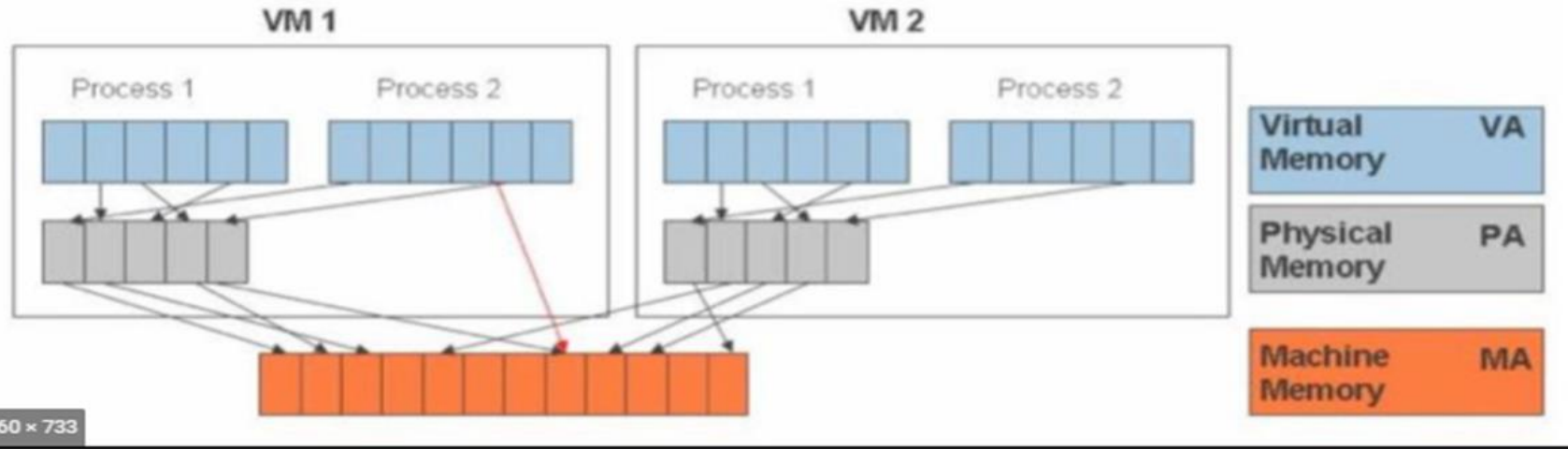
- Translation Lookaside Buffer (TLB)
  - What is TLB ?
    - A CPU cache that memory management hardware uses to improve virtual address translation speed.



TLB: Translation Look-aside Buffer
MMU: Memory Management Unit
CPU: Central Processing Unit

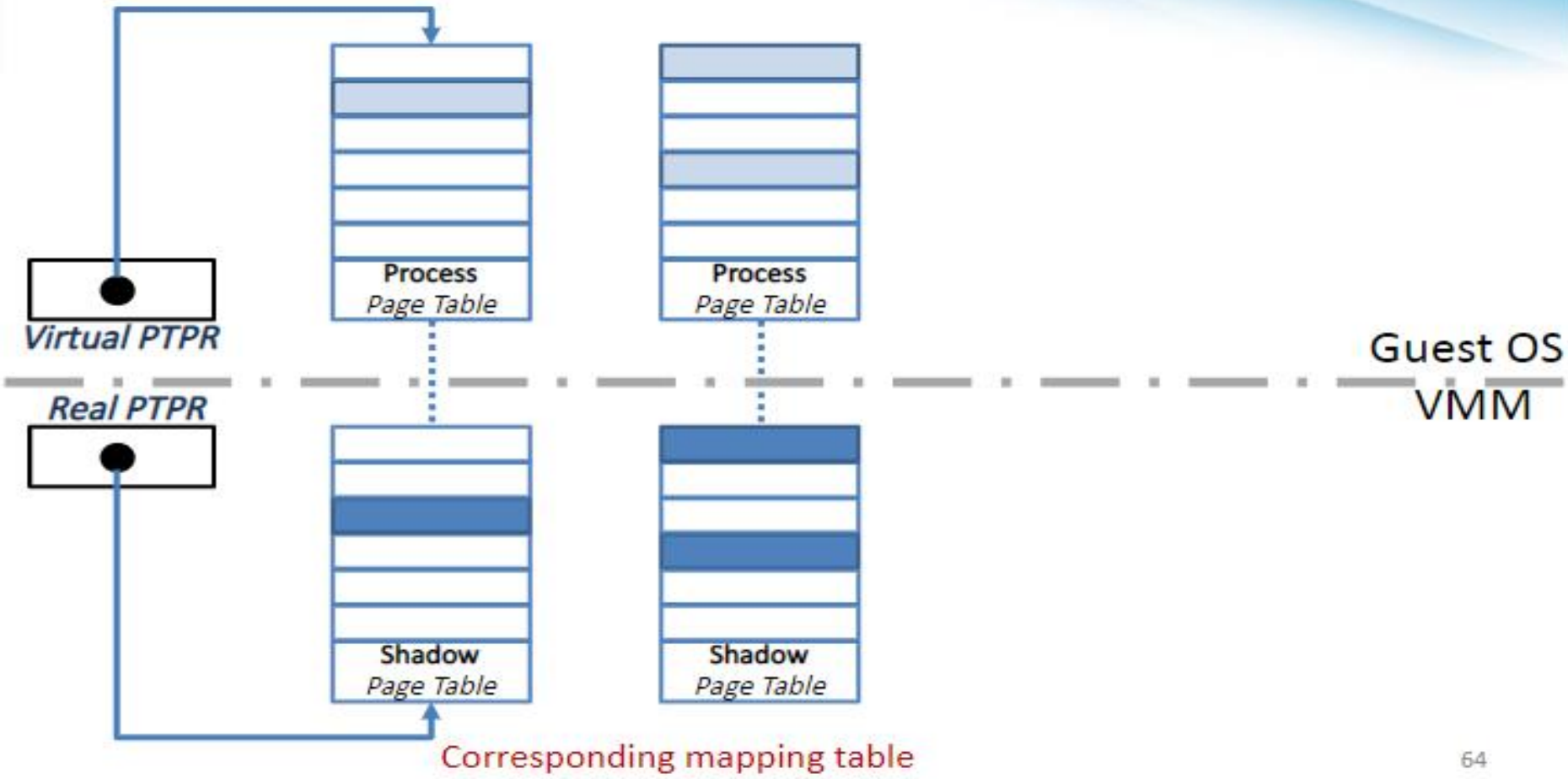
Memory Virtualization

- The guest OS continues to control the mapping of virtual addresses to the guest memory physical addresses, but the guest OS cannot have direct access to the actual machine memory.
- The VMM is responsible for mapping guest physical memory to the actual machine memory, and it uses shadow page tables to accelerate the mappings.
- The VMM uses TLB hardware to map the virtual memory directly to the machine memory to avoid the two levels of translation on every access.



Shadow Page Table

- Shadow page table operations :



How does shadow Page Table works?

- VMM should make MMU virtualized
- VMM manages the real PTBR and a virtual PTBR for each VM
- When the guest OS is activated, the real PTBR points to a shadow page table
- When guest OS modifies the virtual PTBR, it is trapped by VMM
- VMM will walk the page table of the guest and modify the related shadow page table to make MMU get host physical address

Hardware Solution

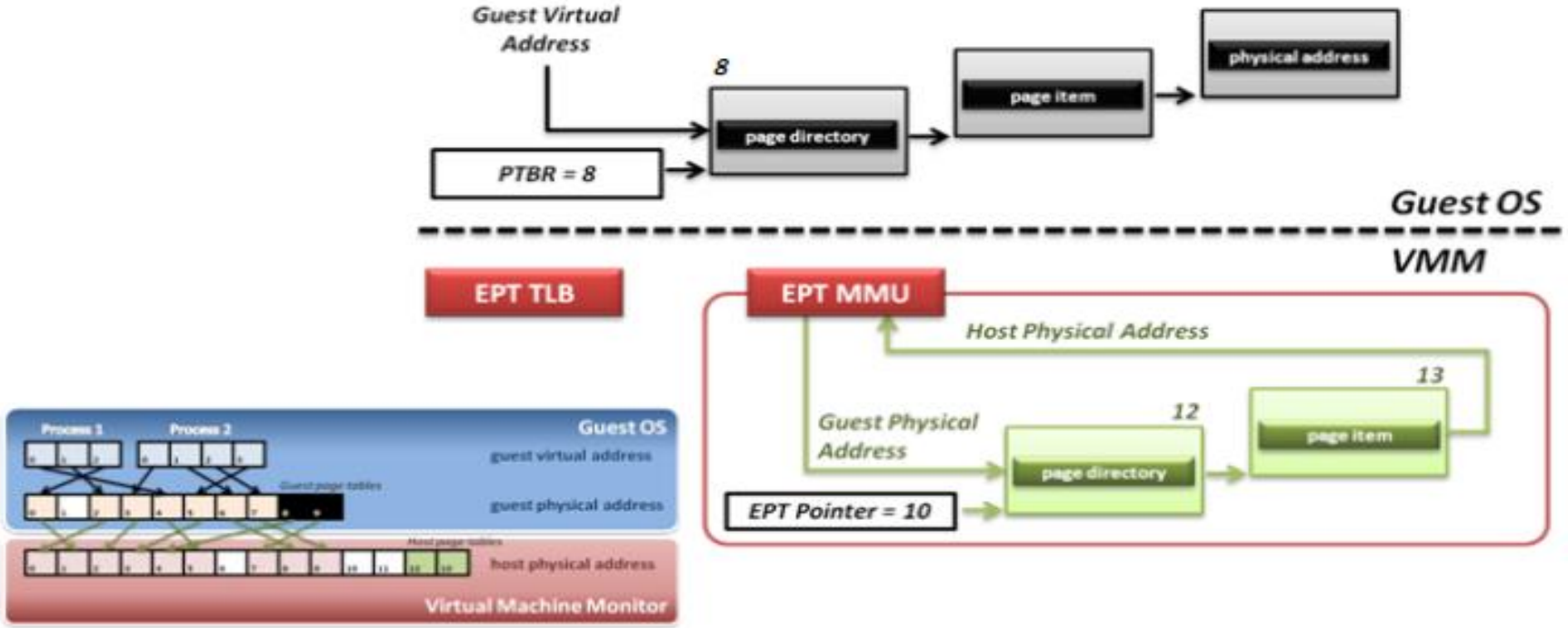
- Difficulties of shadow page table technique :
  - Shadow page table implementation is extremely complex.
  - Page fault mechanism and synchronization issues are critical.
  - Host memory space overhead is considerable.
- But why we need this technique to virtualize MMU ?
  - MMU do not first implemented for virtualization.
  - MMU is knowing nothing about two level page address translation.
- Now, let us consider hardware solution.

Extended Page Table

- Concept of Extended Page Table (EPT) :
  - Instead of walking along with only one page table hierarchy, EPT technique implement one more page table hierarchy.
    - One page table is maintained by guest OS, which is used to generate guest physical address.
    - The other page table is maintained by VMM, which is used to map guest physical address to host physical address.
  - For each memory access operation, EPT MMU will directly get guest physical address from guest page table, and then get host physical address by the VMM mapping table automatically.

Extended Page Table

- Memory operation :



Memory Virtualization Summary

- Software implementation
  - Memory architecture
    - MMU (memory management unit)
    - TLB (translation lookaside buffer)
  - Shadow page table
    - MMU virtualization by virtual PTBR
    - Shadow page table construction
    - Page fault and page table protection
- Hardware assistance
  - Extended page table
    - Hardware walk guest and host page table simultaneously

Memory Virtualization

Summary

- SW-based memory virtualization has been the most complex part in VMM
  - Before HW support, Xen continued optimizing its shadow page tables up to ver3
  - Virtual memory itself is already complicated, but virtualizing virtual memory is horrible
- HW-based memory virtualization significantly reduces VMM complexity
  - The most complex and heavy part is now offloaded to HW

Memory Virtualization
I/O Virtualization

IO Virtualization

- Goal :
  - Share or create IO devices for virtual machines.

- Traditional IO techniques :
  - Direct memory Access (DMA)

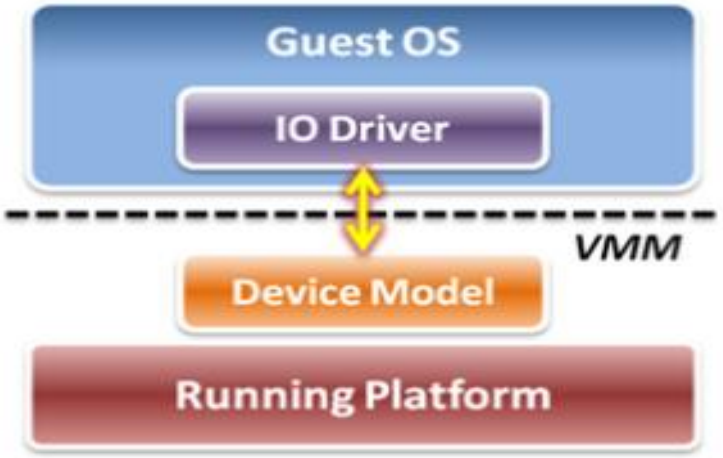
- What is DMA ?
  - Allow certain hardware subsystems within the computer to access system memory for reading and/or writing independently of the central processing unit.

IO Virtualization

- Implementation Layers :
  - System call
    - The interface between applications and guest OS.
  - Driver call
    - The interface between guest OS and IO device drivers.
  - IO operation
    - The interface between IO device driver of guest OS and virtualized hardware (in VMM).

Device Model

- Focus on IO operation level implementation.
  - This is an approach of full virtualization.
- Logic relation between guest OS and VMM :
  - VMM intercepts IO operations from guest OS.
  - Pass these operations to device model on a running platform.
  - Device model needs to emulate the IO operation interfaces.
    - Port mapped IO
    - Memory mapped IO
    - DMA
    - ... etc.



Device Model

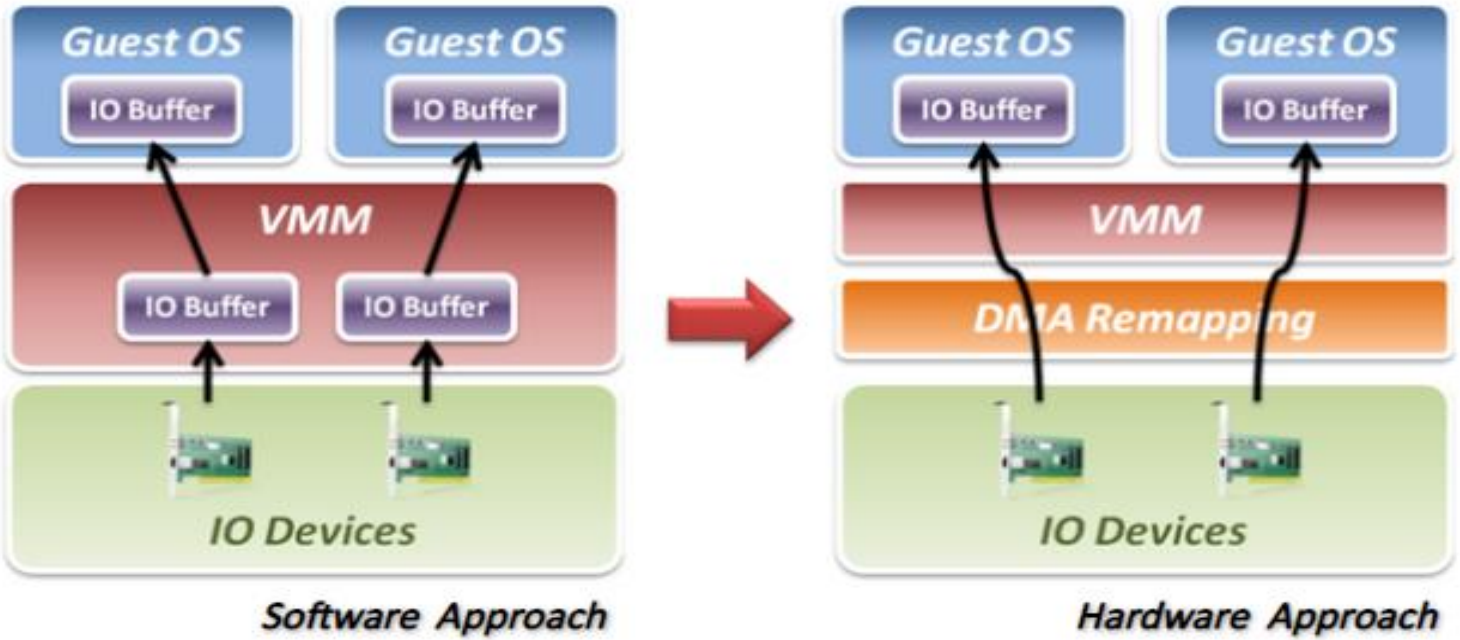
- IO virtualization flow
  - Initialization – device discovery
    - VMM will make guest OS discover the virtualized IO devices.
    - Then guest OS will load the corresponding device driver.
  - Operation – access interception
    - When guest OS executes IO operations, VMM will intercept those accesses.
    - After virtual device operations, VMM returns the control to guest OS.
  - Virtualization – device virtualization
    - Device model should emulate the real electronic logic to satisfy all device interface definition and its effects.
    - VMM may share physical devices to all virtual machines.

Hardware Solution

- Difficulty :
  - Software cannot make data access directly from devices.
- Tow hardware solutions :
  - Implement DMA remapping in hardware
    - Remap DMA operations automatically by hardware.
    - For example, ***Intel VT-d*** .

Intel VT-d

- Add DMA remapping hardware component.



References

- Cloud computing Black Book by Kailash Jayaswal
- Virtualization and Cloud Computing Lecture 6: Memory Virtualization Techniques https://www.youtube.com/watch?v=SiVuXTqwYWk
- Server Virtualization: https://slideplayer.com/slide/5103233/
- Virtualization and Cloud Computing Lecture 7: I/O Virtualization Techniques https://www.youtube.com/watch?v=gwMrdCONERo