PKU--LL202-内存管理

阶段一：存储体系

- Memory Hierarchy: https://en.wikipedia.org/wiki/Memory_hierarchy
  - The term memory hierarchy is used in computer architecture when discussing performance issues in computer architectural design, algorithm predictions, and the lower level programming constructs such as involving locality of reference.
  - A "memory hierarchy" in computer storage distinguishes each level in the "hierarchy" by response time. Since response time, complexity, and capacity are related, the levels may also be distinguished by the controlling technology.
  - There are four major storage levels.
    - Internal – Processor registers and cache.
    - Main – the system RAM and controller cards.
    - On-line mass storage – Secondary storage.
    - Off-line bulk storage – Tertiary and Off-line storage.
  - Locality of reference: https://en.wikipedia.org/wiki/Locality_of_reference
    - In computer science, locality of reference, also known as the principle of locality, is a phenomenon describing the same value, or related storage locations, being frequently accessed.
    - There are two basic types of reference locality – temporal and spatial locality.
      - Temporal locality refers to the reuse of specific data, and/or resources, within a relatively small time duration.
      - Spatial locality refers to the use of data elements within relatively close storage locations.
        - Sequential locality, a special case of spatial locality, occurs when data elements are arranged and accessed linearly, such as, traversing the elements in a one-dimensional array.
- CPU cache: https://en.wikipedia.org/wiki/CPU_cache
  - A CPU cache is a cache used by the central processing unit (CPU) of a computer to reduce the average time to access data from the main memory. The cache is a smaller, faster memory which stores copies of the data from frequently used main memory locations. Most CPUs have different independent caches, including instruction and data caches, where the data cache is usually organized as a hierarchy of more cache levels (L1, L2, etc.)
  - Cache：https://en.wikipedia.org/wiki/Cache_(computing)
    - In computing, a cache (/ kæ / KASH, or / ke / KAYSH in AuE) is a component that stores data so future requests for that data can be served faster; the data stored in a cache might be the results of an earlier computation, or the duplicates of data stored elsewhere.
    - A cache hit occurs when the requested data can be found in a cache, while a cache miss occurs when it cannot. Cache hits are served by reading data from the cache, which is faster than recomputing a result or reading from a slower data store; thus, the more requests can be served from the cache, the faster the system performs.
    - 阅读：The difference between buffer and cache
      - https://en.wikipedia.org/wiki/Cache_(computing)#The_difference_between_buffer_and_cache
  - Cache coherence: https://en.wikipedia.org/wiki/Cache_coherence
    - In computer science, cache coherence is the consistency of shared resource data that ends up stored in multiple local caches.
  - Cache pollution: https://en.wikipedia.org/wiki/Cache_pollution
    - Cache pollution describes situations where an executing computer program loads data into CPU cache unnecessarily, thus causing other needed data to be evicted from the cache into lower levels of the memory hierarchy, potentially all the way down to main memory, degrading performance.
- Computer memory: https://en.wikipedia.org/wiki/Computer_memory
  - In computing, memory refers to the computer hardware devices used to store information for immediate use in a computer; it is synonomous with the term "primary storage".
  - Computer memory operates at a high speed, for example random-access memory (RAM), as a distinction from storage that provides slow-to-access program and data storage but offers higher capacities.
  - Volatile memory: https://en.wikipedia.org/wiki/Volatile_memory
    - Volatile memory, contrary to non-volatile memory, is computer memory that requires power to maintain the stored information; it retains its contents while powered on but when the power is interrupted the stored data is lost very rapidly or immediately.
    - RAM: https://en.wikipedia.org/wiki/Random-access_memory
      - Random-access memory (RAM /ræm/) is a form of computer data storage. A random-access memory device allows data items to be accessed (read or written) in almost the same amount of time irrespective of the physical location of data inside the memory.
  - Non-volatile memory: https://en.wikipedia.org/wiki/Non-volatile_memory
    - Non-volatile memory, nonvolatile memory, NVM or non-volatile storage is computer memory that can retrieve stored information even after having been power cycled (turned off and back on).
    - Examples of non-volatile memory include read-only memory, flash memory, ferroelectric RAM (F-RAM),

most types of magnetic computer storage devices (e.g. hard disks, floppy disks, and magnetic tape), optical discs, and early computer storage methods such as paper tape and punched cards.
    - ROM: https://en.wikipedia.org/wiki/Read-only_memory
        - Read-only memory (ROM) is a class of storage medium used in computers and other electronic devices. Data stored in ROM can only be modified slowly, with difficulty, or not at all, so it is mainly used to distribute firmware (software that is very closely tied to specific hardware, and unlikely to need frequent updates).
    - NVRAM: https://en.wikipedia.org/wiki/Non-volatile_random-access_memory
        - Non-volatile random-access memory (NVRAM) is random-access memory that retains its information when power is turned off (non-volatile). This is in contrast to dynamic random-access memory (DRAM) and static random-access memory (SRAM), which both maintain data only for as long as power is applied.
        - The best-known form of NVRAM memory today is flash memory.
        - Flash memory: https://en.wikipedia.org/wiki/Flash_memory
            - Flash memory is an electronic non-volatile computer storage medium that can be electrically erased and reprogrammed.
            - There are two main types of flash memory, which are named after the NAND and NOR logic gates.
            - The NAND type is primarily used in memory cards, USB flash drives, solid-state drives (those produced in 2009 or later), and similar products, for general storage and transfer of data.

自学：

1. Memory Hierarchy: https://en.wikipedia.org/wiki/Memory_hierarchy
2. Locality of reference: https://en.wikipedia.org/wiki/Locality_of_reference
3. CPU cache: https://en.wikipedia.org/wiki/CPU_cache
4. Cache：https://en.wikipedia.org/wiki/Cache_(computing)
5. Cache coherence: https://en.wikipedia.org/wiki/Cache_coherence
6. Cache pollution: https://en.wikipedia.org/wiki/Cache_pollution
7. Computer memory: https://en.wikipedia.org/wiki/Computer_memory
8. Volatile memory: https://en.wikipedia.org/wiki/Volatile_memory
9. RAM: https://en.wikipedia.org/wiki/Random-access_memory
10. Non-volatile memory: https://en.wikipedia.org/wiki/Non-volatile_memory
11. ROM: https://en.wikipedia.org/wiki/Read-only_memory
12. NVRAM: https://en.wikipedia.org/wiki/Non-volatile_random-access_memory
13. Flash memory: https://en.wikipedia.org/wiki/Flash_memory

阶段二：内存管理基础

- Memory management: https://en.wikipedia.org/wiki/Memory_management_(operating_systems)
    - （Note：wiki还有一个页面是Memory management，主要是讲DMM的）
    - In operating systems, memory management is the function responsible for managing the computer's primary memory.
    - The memory management function keeps track of the status of each memory location, either allocated or free. It determines how memory is allocated among competing processes, deciding who gets memory, when they receive it, and how much they are allowed. When memory is allocated it determines which memory locations will be assigned. It tracks when memory is freed or unallocated and updates the status.
    - Memory management techniques:
        - Single contiguous allocation
        - Partitioned allocation
        - Paged memory management
            - Paging：https://en.wikipedia.org/wiki/Paging
                - In computer operating systems, paging is one of the memory management schemes by which a computer stores and retrieves data from the secondary storage for use in main memory.
                - In the paging memory-management scheme, the operating system retrieves data from secondary storage in same-size blocks called pages.
                - Page：https://en.wikipedia.org/wiki/Page_(computer_memory)
                    - A page, memory page, or virtual page is a fixed-length contiguous block of virtual memory, described by a single entry in the page table. It is the smallest unit of data for memory management in a virtual memory operating system.
                - Demand paging: https://en.wikipedia.org/wiki/Demand_paging
                    - In computer operating systems, demand paging (as opposed to anticipatory paging（预期分页，先行调页）) is a method of virtual memory management.
                    - In a system that uses demand paging, the operating system copies a disk page into physical memory only if an attempt is made to access it and that page is not already in memory (i.e., if a page fault occurs).
                    - It follows that a process begins execution with none of its pages in physical memory, and

many page faults will occur until most of a process's working set of pages is located in physical memory. This is an example of a lazy loading technique.
        - Segmented memory management
            - Memory segmentation: https://en.wikipedia.org/wiki/Memory_segmentation
                - Memory segmentation is the division of a computer's primary memory into segments or sections. In a computer system using segmentation, a reference to a memory location includes a value that identifies a segment and an offset within that segment.
                - Segments or sections are also used in object files of compiled programs when they are linked together into a program image and when the image is loaded into memory.
            - X86 memory segmentation: https://en.wikipedia.org/wiki/X86_memory_segmentation
                - x86 memory segmentation refers to the implementation of memory segmentation in the Intel x86 computer instruction set architecture.
- Virtual memory: https://en.wikipedia.org/wiki/Virtual_memory
    - In computing, virtual memory is a memory management technique that is implemented using both hardware and software. It maps memory addresses used by a program, called virtual addresses, into physical addresses in computer memory.
    - Memory address: https://en.wikipedia.org/wiki/Memory_address
        - In computing, memory address is a data concept used at various levels by software and hardware to access the computer's primary storage memory. Memory addresses are fixed-length sequences of digits conventionally displayed and manipulated as unsigned integers.
        - Byte addressing: https://en.wikipedia.org/wiki/Byte_addressing
            - Byte addressing refers to hardware architectures which support accessing individual bytes of data rather than only larger units called words, which would be word-addressable.
        - Word addressable: https://en.wikipedia.org/wiki/Word-addressable
            - If a computer's memory is word-addressable then each word in memory is assigned its own memory address. That means that the processor is able to address and fetch only complete words from the memory.
    - Physical address: https://en.wikipedia.org/wiki/Physical_address
        - In computing, a physical address (also real address, or binary address), is a memory address that is represented in the form of a binary number on the address bus circuitry in order to enable the data bus to access a particular storage cell of main memory, or a register of memory mapped I/O device.
    - Logical address: https://en.wikipedia.org/wiki/Logical_address
        - In computing, a logical address is the address at which an item (memory cell, storage element, network host) appears to reside from the perspective of an executing application program.
        - Flat memory model: https://en.wikipedia.org/wiki/Flat_memory_model
            - Flat memory model or linear memory model refers to a memory addressing paradigm in which "memory appears to the program as a single contiguous address space." The CPU can directly (and linearly) address all of the available memory locations without having to resort to any sort of memory segmentation or paging schemes.
    - Virtual address space: https://en.wikipedia.org/wiki/Virtual_address_space
        - In computing, a virtual address space (VAS) or address space is the set of ranges of virtual addresses that an operating system makes available to a process.
- Global Descriptor Table: https://en.wikipedia.org/wiki/Global_Descriptor_Table
    - The Global Descriptor Table or GDT is a data structure used by Intel x86-family processors starting with the 80286 in order to define the characteristics of the various memory areas used during program execution, including the base address, the size and access privileges like executability and writability. These memory areas are called segments in Intel terminology.
    - The GDT can hold things other than segment descriptors as well. Every 8-byte entry in the GDT is a descriptor, but these can be Task State Segment (or TSS) descriptors, Local Descriptor Table (LDT) descriptors, or Call Gate descriptors.
- MMU : https://en.wikipedia.org/wiki/Memory_management_unit
    - A memory management unit (MMU), sometimes called paged memory management unit (PMMU), is a computer hardware unit having all memory references passed through itself, primarily performing the translation of virtual memory addresses to physical addresses.
    - An MMU is effectively performing the virtual memory management, handling at the same time memory protection, cache control, bus arbitration and, in simpler computer architectures (especially 8-bit systems), bank switching.
    - Page table: https://en.wikipedia.org/wiki/Page_table
        - A page table is the data structure used by a virtual memory system in a computer operating system to store the mapping between virtual addresses and physical addresses.
        - Virtual addresses are used by the accessing process, while physical addresses are used by the hardware or more specifically to the RAM.
    - TLB: https://en.wikipedia.org/wiki/Translation_lookaside_buffer
        - A translation lookaside buffer (TLB) is a cache that memory management hardware uses to improve virtual address translation speed.
        - The majority of desktop, laptop, and server processors includes one or more TLBs in the memory management hardware, and it is nearly always present in any hardware that utilizes paged or segmented

virtual memory.
    - Memory protection: https://en.wikipedia.org/wiki/Memory_protection
        - Memory protection is a way to control memory access rights on a computer, and is a part of most modern operating systems.
        - The main purpose of memory protection is to prevent a process from accessing memory that has not been allocated to it. This prevents a bug or malware within a process from affecting other processes, or the operating system itself.
        - An attempt to access unowned memory results in a hardware fault, called a segmentation fault or storage violation exception, generally causing abnormal termination of the offending process.
- GPF (IDT 0x0D) and PF (IDT 0x0E):
    - General protection fault: https://en.wikipedia.org/wiki/General_protection_fault
        - A general protection fault (GPF) in the Intel x86 and AMD x86-64 types of computer microprocessor architectures, and other unrelated architectures, is a fault (a type of interrupt) that can encompass several cases in which protection mechanisms within the processor architecture are violated by any of the programs that are running, either the kernel or a user program.
    - Page fault: https://en.wikipedia.org/wiki/Page_fault
        - A page fault (sometimes called #PF or PF) is a type of interrupt, called trap, raised by computer hardware when a running program accesses a memory page that is mapped into the virtual address space, but not actually loaded into main memory.
        - The hardware that detects a page fault is the processor's memory management unit (MMU), while the exception handling software that handles page faults is generally a part of the operating system kernel. When handling a page fault, the operating system generally tries to make the required page accessible at the location in physical memory, or terminates the program in case of an illegal memory access.

自学：

1. Memory management: https://en.wikipedia.org/wiki/Memory_management_(operating_systems)
2. Paging：https://en.wikipedia.org/wiki/Paging
3. Page：https://en.wikipedia.org/wiki/Page_(computer_memory)
4. Demand paging: https://en.wikipedia.org/wiki/Demand_paging
5. Memory segmentation: https://en.wikipedia.org/wiki/Memory_segmentation
6. X86 memory segmentation: https://en.wikipedia.org/wiki/X86_memory_segmentation
7. Virtual memory: https://en.wikipedia.org/wiki/Virtual_memory
8. Memory address: https://en.wikipedia.org/wiki/Memory_address
9. Byte addressing: https://en.wikipedia.org/wiki/Byte_addressing
10. Word addressable: https://en.wikipedia.org/wiki/Word-addressable
11. Physical address: https://en.wikipedia.org/wiki/Physical_address
12. Logical address: https://en.wikipedia.org/wiki/Logical_address
13. Flat memory model: https://en.wikipedia.org/wiki/Flat_memory_model
14. Virtual address space: https://en.wikipedia.org/wiki/Virtual_address_space
15. Global Descriptor Table: https://en.wikipedia.org/wiki/Global_Descriptor_Table
16. MMU：https://en.wikipedia.org/wiki/Memory_management_unit
17. Page table: https://en.wikipedia.org/wiki/Page_table
18. TLB: https://en.wikipedia.org/wiki/Translation_lookaside_buffer
19. Memory protection: https://en.wikipedia.org/wiki/Memory_protection
20. General protection fault: https://en.wikipedia.org/wiki/General_protection_fault
21. Page fault: https://en.wikipedia.org/wiki/Page_fault

阶段三：内存分配

- Memory management: https://en.wikipedia.org/wiki/Memory_management
    - Memory management is the act of managing computer memory at the system level. The essential requirement of memory management is to provide ways to dynamically allocate portions of memory to programs at their request, and free it for reuse when no longer needed.
    - Memory pool：https://en.wikipedia.org/wiki/Memory_pool
        - Memory pools, also called fixed-size blocks allocation, is the use of pools for memory management that allows dynamic memory allocation comparable to malloc or C++'s operator new. As those implementations suffer from fragmentation because of variable block sizes, it is not recommendable to use them in a real time system due to performance. A more efficient solution is preallocating a number of memory blocks with the same size called the memory pool.
        - Slab allocation: https://en.wikipedia.org/wiki/Slab_allocation
            - Slab allocation is a memory management mechanism intended for the efficient memory allocation of kernel objects. It eliminates fragmentation caused by allocations and deallocations. The technique is used to retain allocated memory that contains a data object of a certain type for reuse upon subsequent allocations of objects of the same type. It is analogous to an object pool, but only applies to memory, not other resources.
        - SLUB: https://en.wikipedia.org/wiki/SLUB_(software)
            - SLUB ("the unqueued slab allocator") is a memory management mechanism intended for the efficient

memory allocation of kernel objects which displays the desirable property of eliminating fragmentation caused by allocations and deallocations. The technique is used to retain allocated memory that contains a data object of a certain type for reuse upon subsequent allocations of objects of the same type. It is used in Linux and became the default allocator since 2.6.23.
        - SLOB: https://en.wikipedia.org/wiki/SLOB
            - The SLOB (Simple List Of Blocks) allocator is one of three available memory allocators in the Linux kernel. (The other two are SLAB and SLUB.) The SLOB allocator is designed to require little memory for the implementation and housekeeping, for use in small systems such as embedded systems. Unfortunately, a major limitation of the SLOB allocator is that it suffers greatly from internal fragmentation.
        - Buddy memory allocation: https://en.wikipedia.org/wiki/Buddy_memory_allocation
            - The buddy memory allocation technique is a memory allocation algorithm that divides memory into partitions to try to satisfy a memory request as suitably as possible. This system makes use of splitting memory into halves to try to give a best-fit.
- Fragmentation: https://en.wikipedia.org/wiki/Fragmentation_(computing)
    - In computer storage, fragmentation is a phenomenon in which storage space is used inefficiently, reducing capacity or performance and often both. The exact consequences of fragmentation depend on the specific system of storage allocation in use and the particular form of fragmentation.
    - There are three different but related forms of fragmentation:
        - external fragmentation: External fragmentation arises when free memory is separated into small blocks and is interspersed by allocated memory.
        - internal fragmentation: Due to the rules governing memory allocation, more computer memory is sometimes allocated than is needed.
        - data fragmentation: Data fragmentation occurs when a collection of data in memory is broken up into many pieces that are not close together. It is typically the result of attempting to insert a large object into storage that has already suffered external fragmentation.
- Memory leak: https://en.wikipedia.org/wiki/Memory_leak
    - In computer science, a memory leak is a type of resource leak that occurs when a computer program incorrectly manages memory allocations in such a way that memory which is no longer needed is not released.
    - kmemleak: Documentation/kmemleak.txt
        - Kmemleak provides a way of detecting possible kernel memory leaks in a way similar to a tracing garbage collector, with the difference that the orphan objects are not freed but only reported via /sys/kernel/debug/kmemleak.
        - enable CONFIG_DEBUG_KMEMLEAK
        - See mm/kmemleak.c and mm/kmemleak-test.c
- Working set: https://en.wikipedia.org/wiki/Working_set
    - Working set is a concept in computer science which defines the amount of memory that a process requires in a given time interval.
    - Resident set size: https://en.wikipedia.org/wiki/Resident_set_size
        - In computing, resident set size (RSS) is the portion of memory occupied by a process that is held in main memory (RAM). The rest of the occupied memory exists in the swap space or file system, either because some parts of the occupied memory were paged out, or because some parts of the executable were never loaded.
    - Working set size: https://en.wikipedia.org/wiki/Working_set_size
        - In computing, working set size is the amount of memory needed to compute the answer to a problem. In any computing scenario, but especially high performance computing where mistakes can be costly, this is a significant design-criteria for a given super computer system in order to ensure that the system performs as expected.
    - Thrashing: https://en.wikipedia.org/wiki/Thrashing_(computer_science)
        - In computer science, thrashing occurs when a computer's virtual memory subsystem is in a constant state of paging, rapidly exchanging data in memory for data on disk, to the exclusion of most application-level processing.[1] This causes the performance of the computer to degrade or collapse. The situation may continue indefinitely until the underlying cause is addressed.
    - Out of memory: https://en.wikipedia.org/wiki/Out_of_memory
        - Out of memory (OOM) is an often undesired state of computer operation where no additional memory can be allocated for use by programs or the operating system.
    - Page replacement algorithm: https://en.wikipedia.org/wiki/Page_replacement_algorithm
        - In a computer operating system that uses paging for virtual memory management, page replacement algorithms decide which memory pages to page out (swap out, write to disk) when a page of memory needs to be allocated. Paging happens when a page fault occurs and a free page cannot be used to satisfy the allocation, either because there are none, or because the number of free pages is lower than some threshold.
- 查看内存：
    - free命令
    - cat /proc/meminfo
    - vmstat：https://en.wikipedia.org/wiki/Vmstat
        - vmstat (virtual memory statistics) is a computer system monitoring tool that collects and displays summary

information about operating system memory, processes, interrupts, paging and block I/O. Users of vmstat can specify a sampling interval which permits observing system activity in near-real time.
    - pmap：report memory map of a process
        - `pmap -x `pidof bash``
    - slab信息：
        - `sudo slabtop`
        - `cat /proc/slabinfo`

自学：

1. Memory management: https://en.wikipedia.org/wiki/Memory_management
2. Memory pool：https://en.wikipedia.org/wiki/Memory_pool
3. Slab allocation: https://en.wikipedia.org/wiki/Slab_allocation
4. SLUB: https://en.wikipedia.org/wiki/SLUB_(software)
5. SLOB: https://en.wikipedia.org/wiki/SLOB
6. Buddy memory allocation: https://en.wikipedia.org/wiki/Buddy_memory_allocation
7. Fragmentation: https://en.wikipedia.org/wiki/Fragmentation_(computing)
8. Memory leak: https://en.wikipedia.org/wiki/Memory_leak
9. Working set: https://en.wikipedia.org/wiki/Working_set
10. Resident set size: https://en.wikipedia.org/wiki/Resident_set_size
11. Working set size: https://en.wikipedia.org/wiki/Working_set_size
12. Thrashing: https://en.wikipedia.org/wiki/Thrashing_(computer_science)
13. Out of memory: https://en.wikipedia.org/wiki/Out_of_memory
14. Page replacement algorithm: https://en.wikipedia.org/wiki/Page_replacement_algorithm
15. vmstat：https://en.wikipedia.org/wiki/Vmstat

阶段四：Table walking

- 学习mm/pagewalk.c
    - 理清三层page table和四层page table的一个walking流程

其它建议内容：

- Memory-level parallelism: https://en.wikipedia.org/wiki/Memory-level_parallelism
    - Memory-level parallelism (MLP) is a term in computer architecture referring to the ability to have pending multiple memory operations, in particular cache misses or translation lookaside buffer (TLB) misses, at the same time.
- A20 line: https://en.wikipedia.org/wiki/A20_line
    - The A20 or addressing line 20 is one of the electrical lines that make up the system bus of an x86-based computer system.
    - The A20 line in particular is used to transmit the 21st bit on the address bus.
- PAE: https://en.wikipedia.org/wiki/Physical_Address_Extension
    - In computing, Physical Address Extension (PAE) is a memory management feature for the IA-32 architecture, first introduced in the Pentium Pro. It defines a page table hierarchy of three levels, with table entries of 64 bits each instead of 32, allowing these CPUs to access a physical address space larger than 4 gigabytes (232 bytes).
- Data structure alignment: https://en.wikipedia.org/wiki/Data_structure_alignment
    - Data structure alignment is the way data is arranged and accessed in computer memory.
    - It consists of two separate but related issues: data alignment and data structure padding.
- MTRR: https://en.wikipedia.org/wiki/Memory_type_range_register
    - Memory type range registers (MTRRs) are a set of processor supplementary capabilities control registers that provide system software with control of how accesses to memory ranges by the CPU are cached. It uses a set of programmable model-specific registers (MSRs) which are special registers provided by most modern CPUs.
- Copy-on-write: https://en.wikipedia.org/wiki/Copy-on-write
- Zero page：https://en.wikipedia.org/wiki/Zero_page
- Memory corruption: https://en.wikipedia.org/wiki/Memory_corruption
- User space：https://en.wikipedia.org/wiki/User_space
    - A modern computer operating system usually segregates virtual memory into kernel space and user space.[a] Primarily, this separation serves to provide memory protection that protects data and functionality from faults (by improving fault tolerance) and malicious behaviour (by providing computer security).