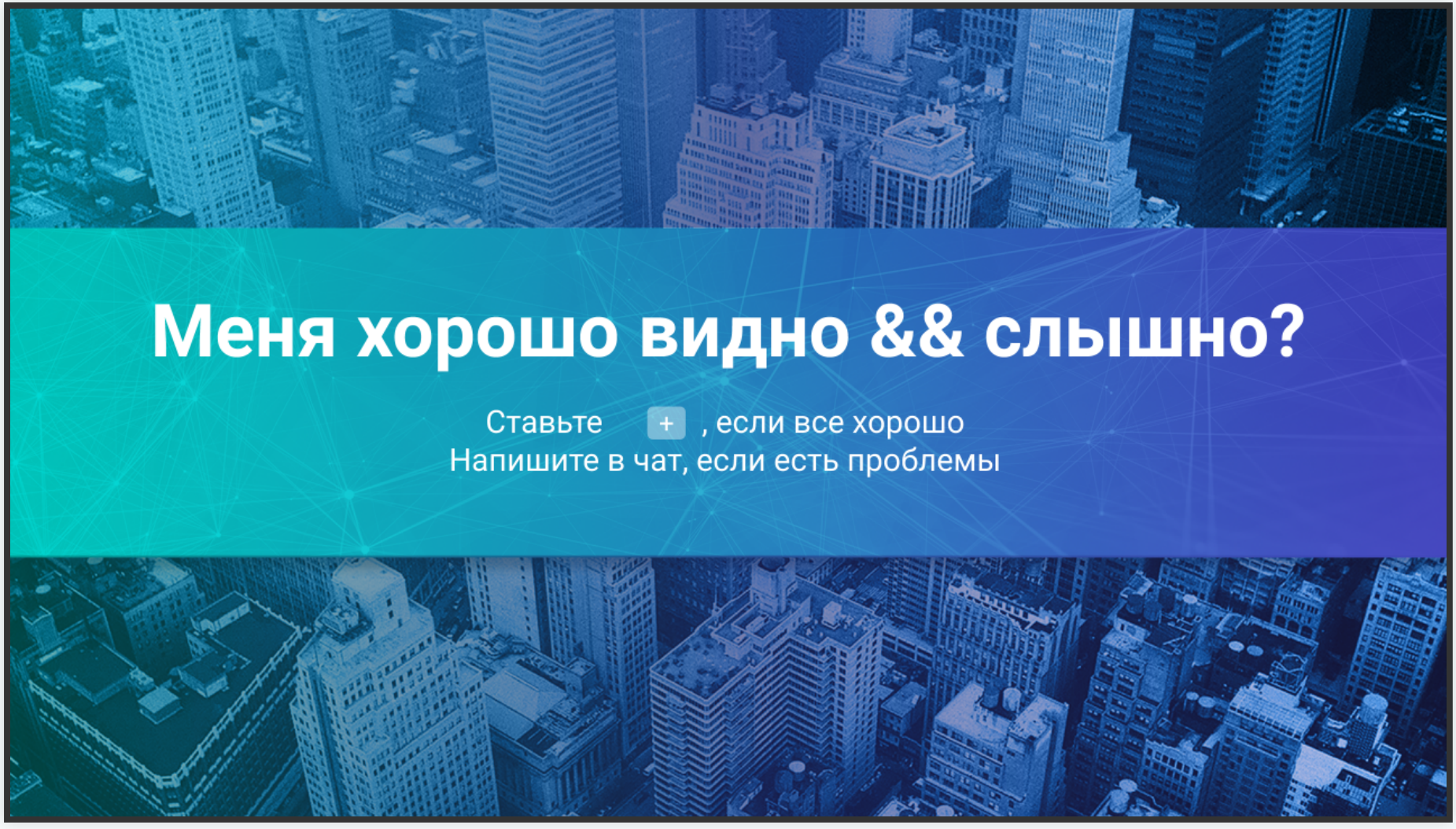
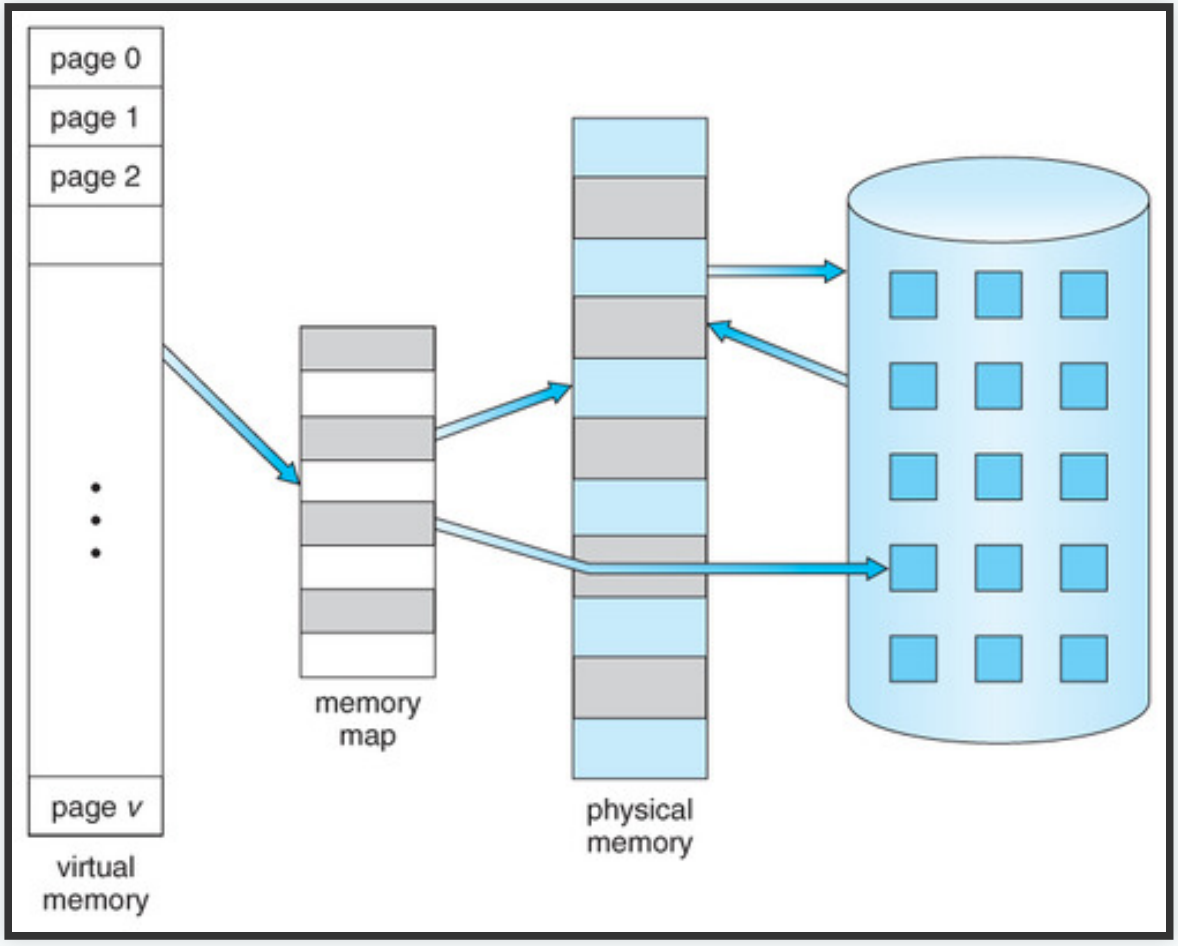

OTUS
ОНЛАЙН-ОБРАЗОВАНИЕ
Онлайн-образование

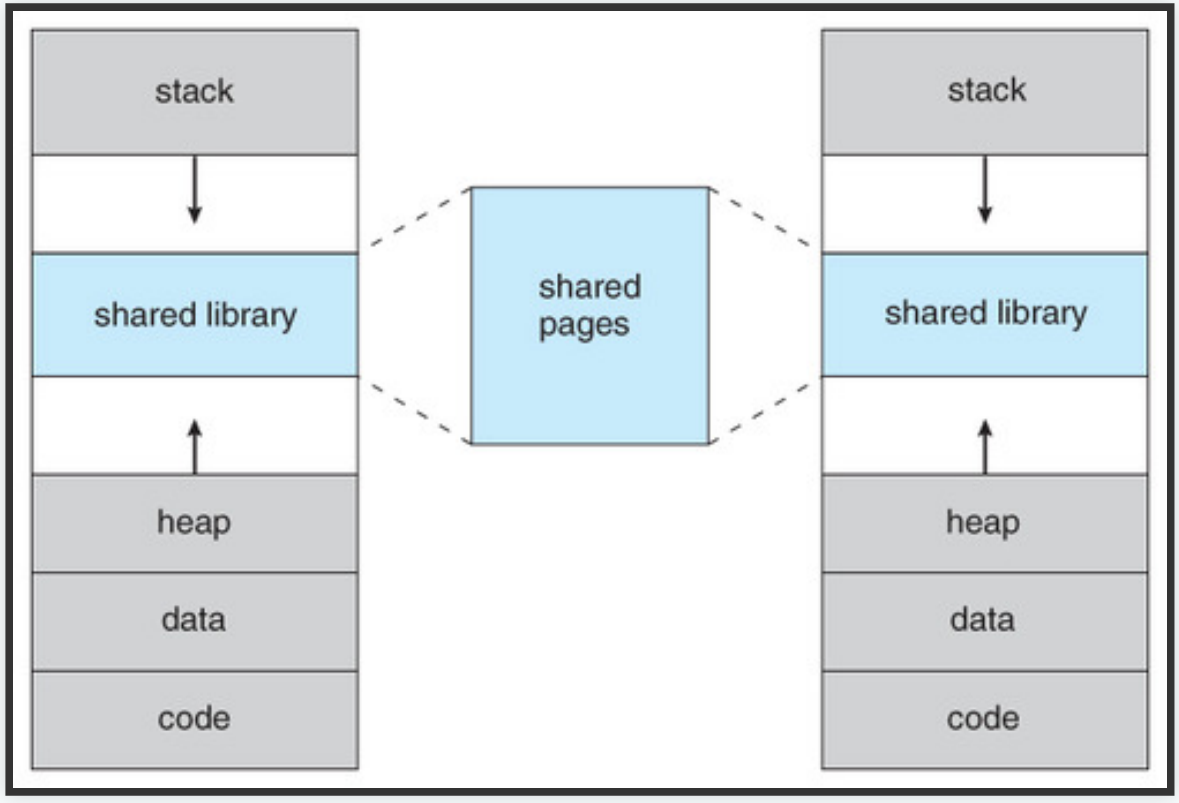

Меня хорошо видно && слышно?
Ставьте + , если все хорошо
Напишите в чат, если есть проблемы

НЕ ЗАБЫТЬ ВКЛЮЧИТЬ ЗАПИСЬ!!!

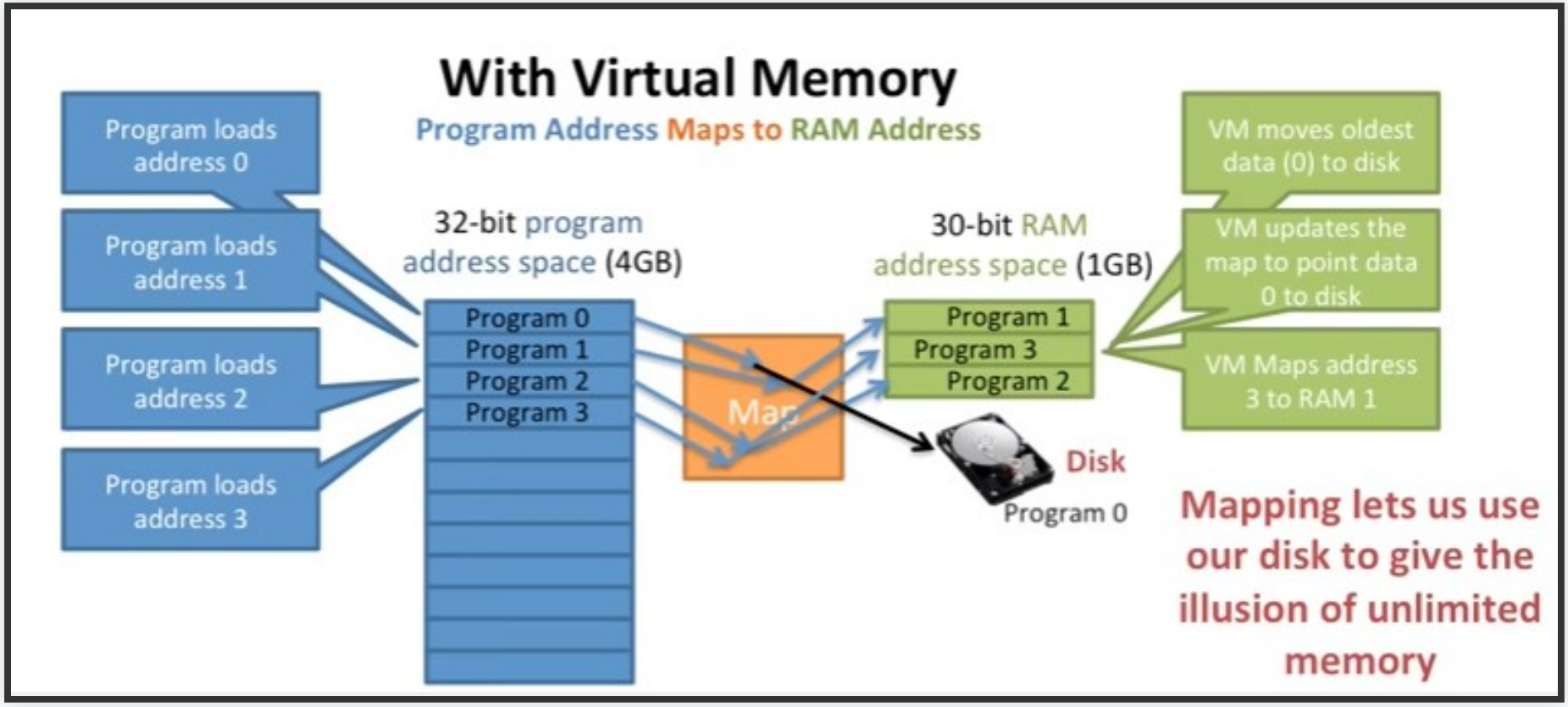
Виртуальная память



Виртуальное адресное пространство процесса



Нехватка памяти



Типы VMA мэппингов

- File based mappings (mmap)
  - Code pages (binaries), libraries
  - Data files
  - Shared memory
  - Devices
- Anonymous mappings:
  - Stack
  - Heap
  - CoW pages
- Различные механизмы для reverse mapping, demand fetching, swapping

PMAP

- pmap
  - Использование памяти процессами. Команда детально расписывает использование оперативной памяти процессами в системе.

```
pmap [options] PID [PID ...]
-x, --extended              show details
-X                          show even more details
-XX                         show everything the kernel provides
-d, --device                show the device format
-p, --show-path             show path in the mapping
```

Информация в /proc/<PID>/maps

```
address perms offset dev inode pathname
```

- карта памяти процесса
  - [heap] - the heap of the program
  - [stack] - the stack of the main process
  - [vdso] - the "virtual dynamic shared object", the kernel system call handler
  - пусто - анонимная область

Информация в /proc/<PID>/smaps

- расширение информации из maps с указанием потребления памяти
  - PSS
    - [own page count] + [shared page count]/[consumer count]
  - RSS
    - занимаемая процессом память
  - Referenced
    - активные page-cache, которые не предполагаются быть выгруженными в ближайшее время
  - VmFlags - флаги свойств области памяти (http://man7.org/linux/man-pages/man5/proc.5.html)

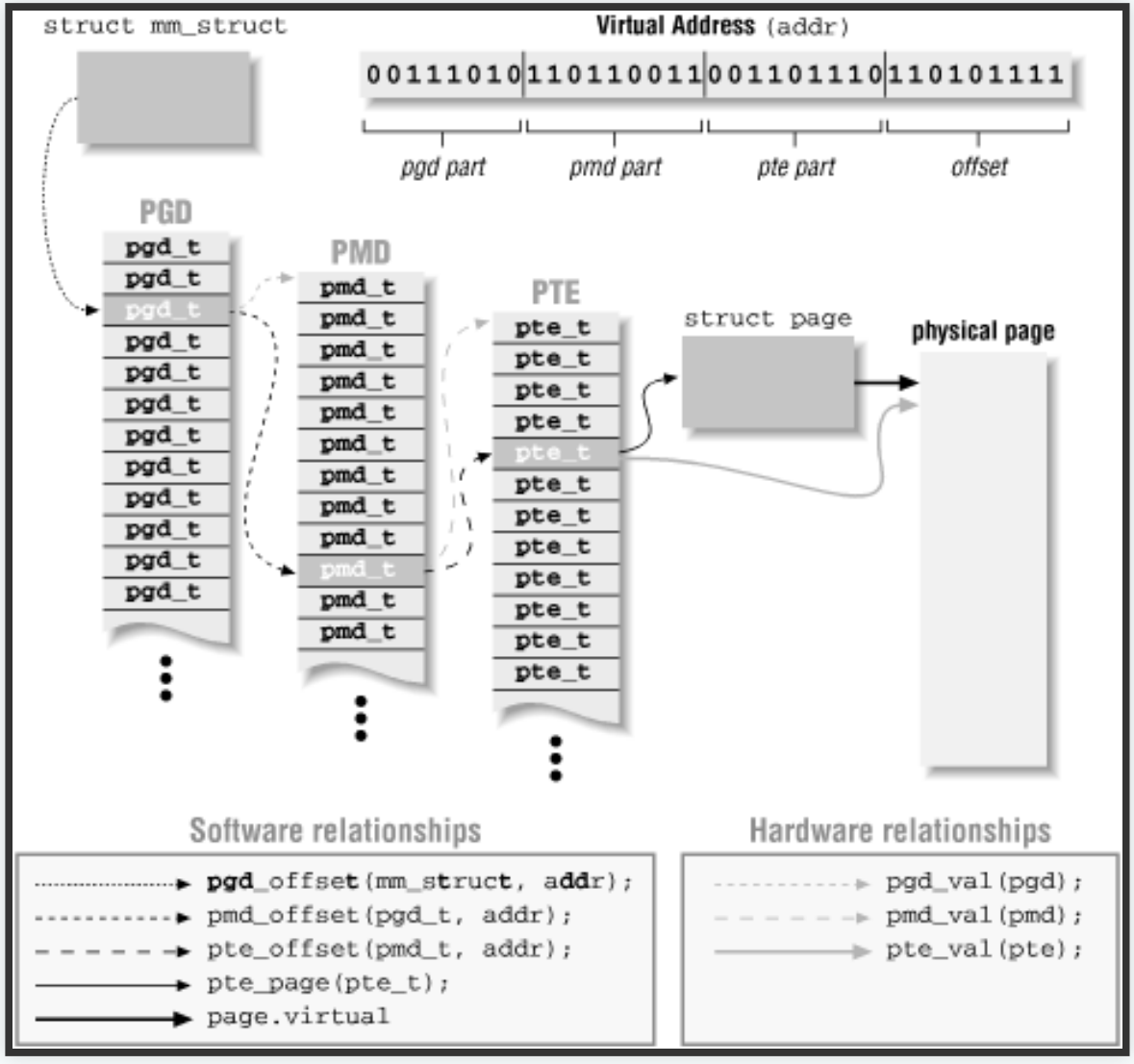
/proc/<PID>/statm

- **size** размер программы (в страницах) (VmSize in status)
- **resident** size of memory portions (pages) (VmRSS in status)
- **shared** кол-во shared страниц (RssFile+RssShmem in status)
- **trs** число страниц 'кода' (not including libs; broken, includes data segment)
- **lrs** число страниц библиотек (always 0 on 2.6)
- **drs** число страниц data/stack (including libs; broken, includes library text)
- **dt** число грязных страниц (always 0 on 2.6)

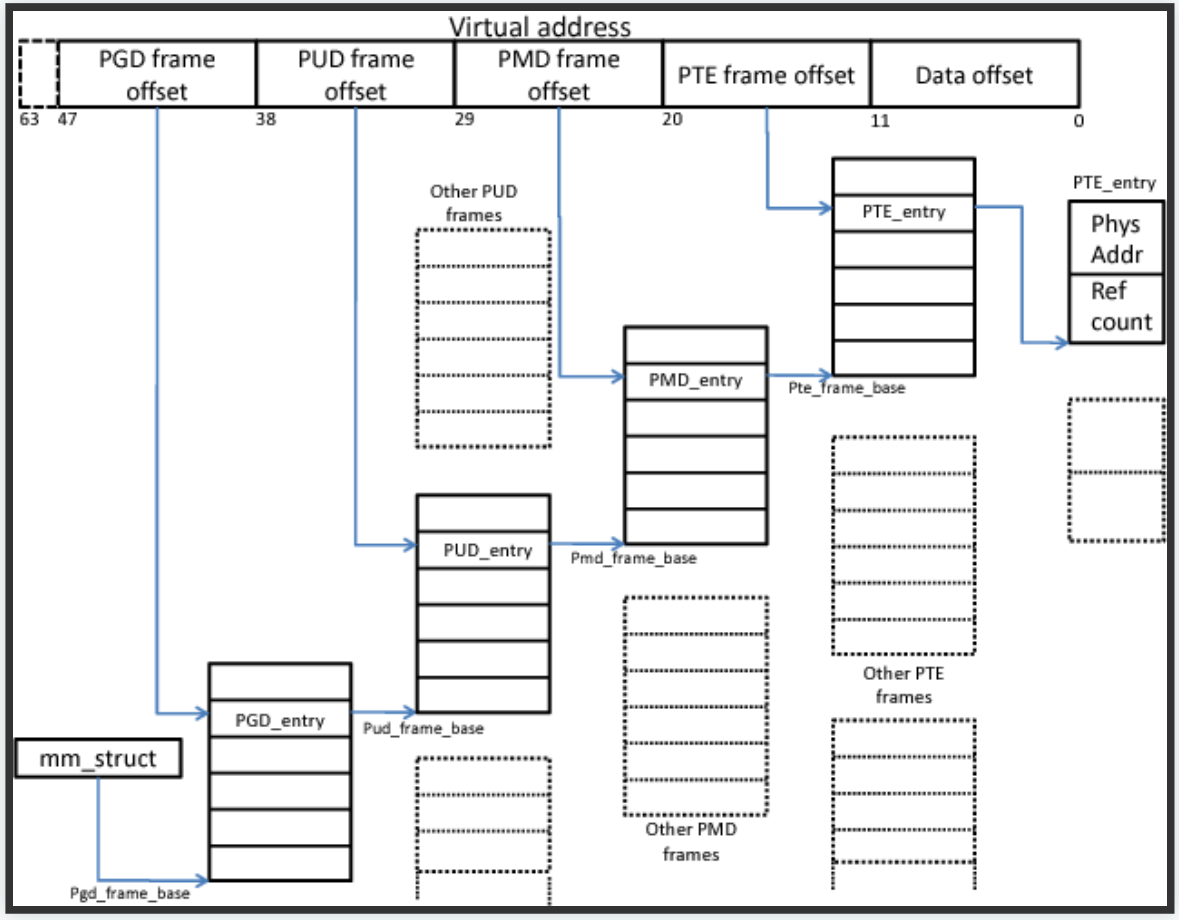
Multi Level Page Table

- PGD: Page Global Directory
- PUD: Page Upper Directory
- PMD: Page Mid-level Directory
- PTE: page table entry
  - все page tables имеют размер 4к
  - Total: grep PageTables /proc/meminfo

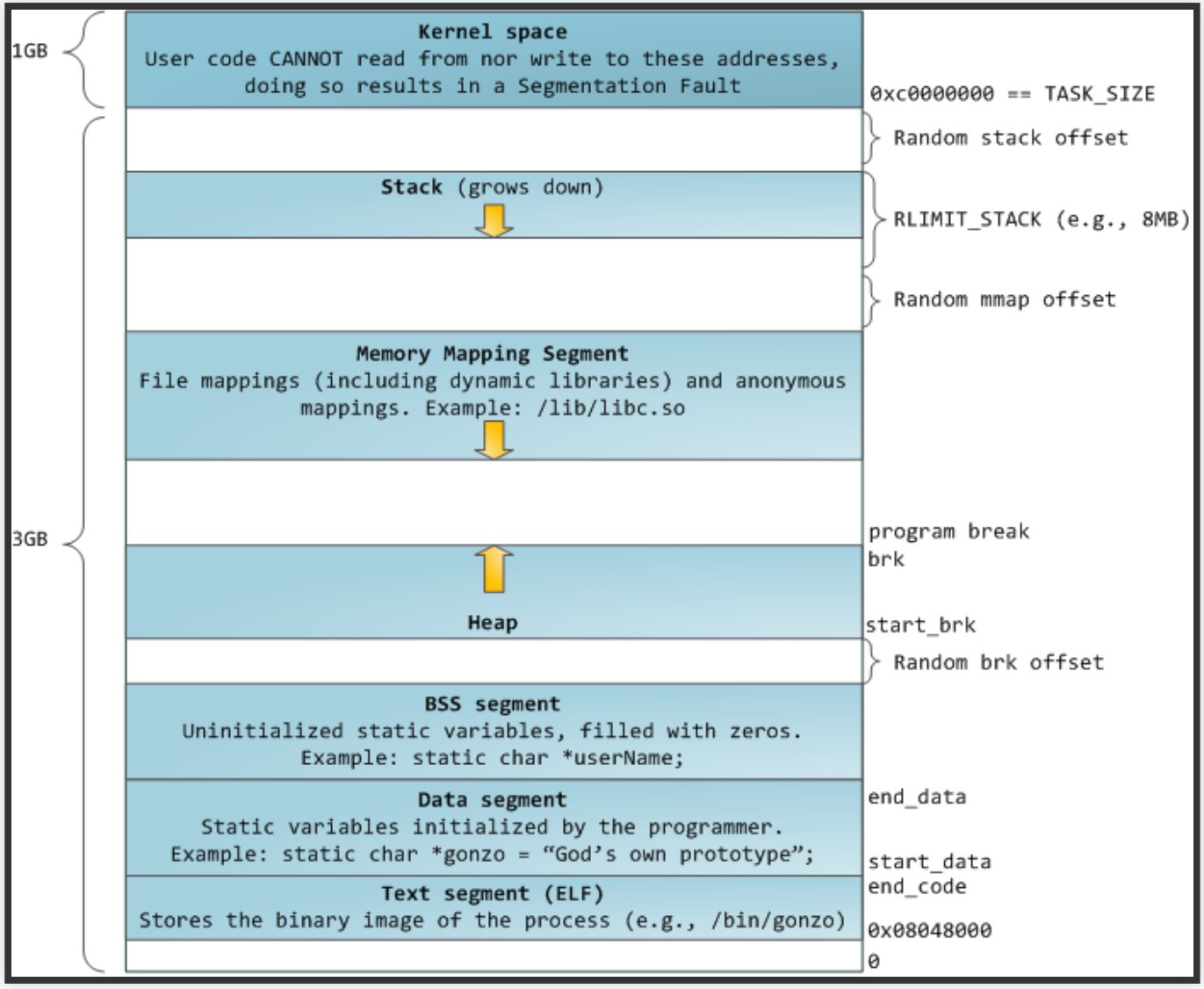
3-level Page Table



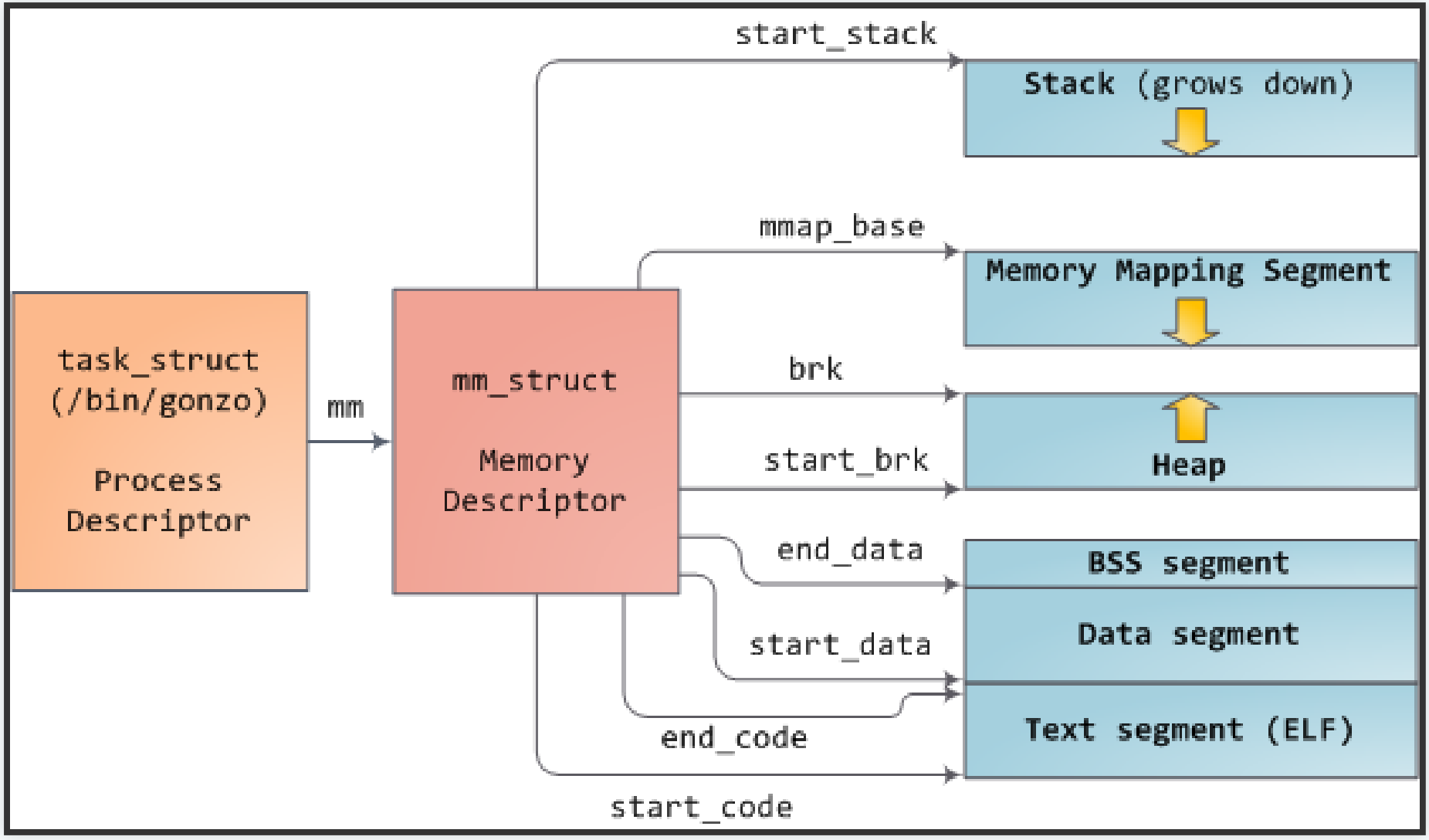
4-level Page Table



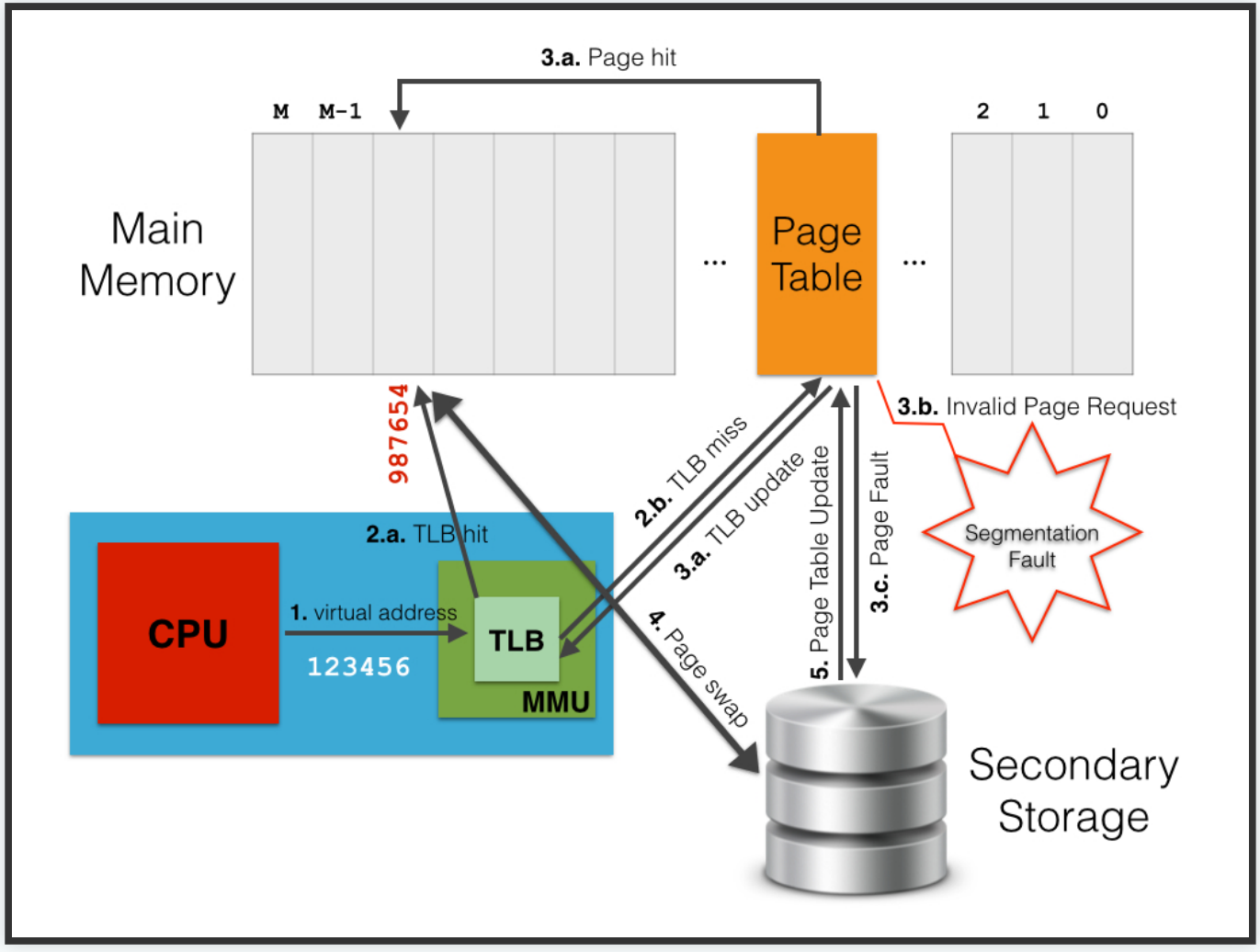
Process memory



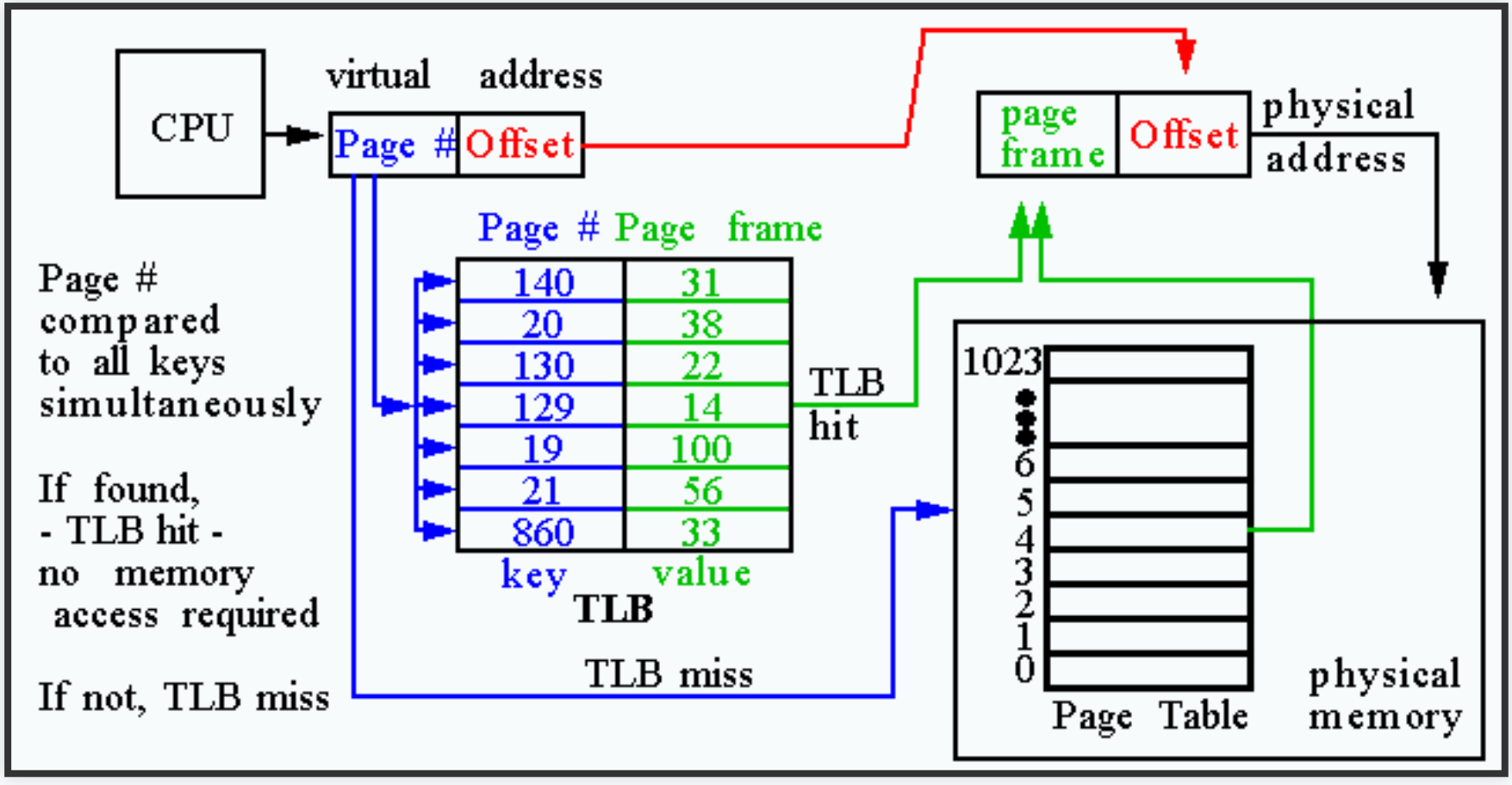
mm_struct


start_stack
Stack (grows down)
mmap_base
Memory Mapping Segment
task_struct
(/bin/gonzo)
Process
Descriptor
mm
mm_struct
Memory
Descriptor
brk
start_brk
Heap
end_data
BSS segment
Data segment
start_data
Text segment (ELF)
end_code
start_code

Memory workflow



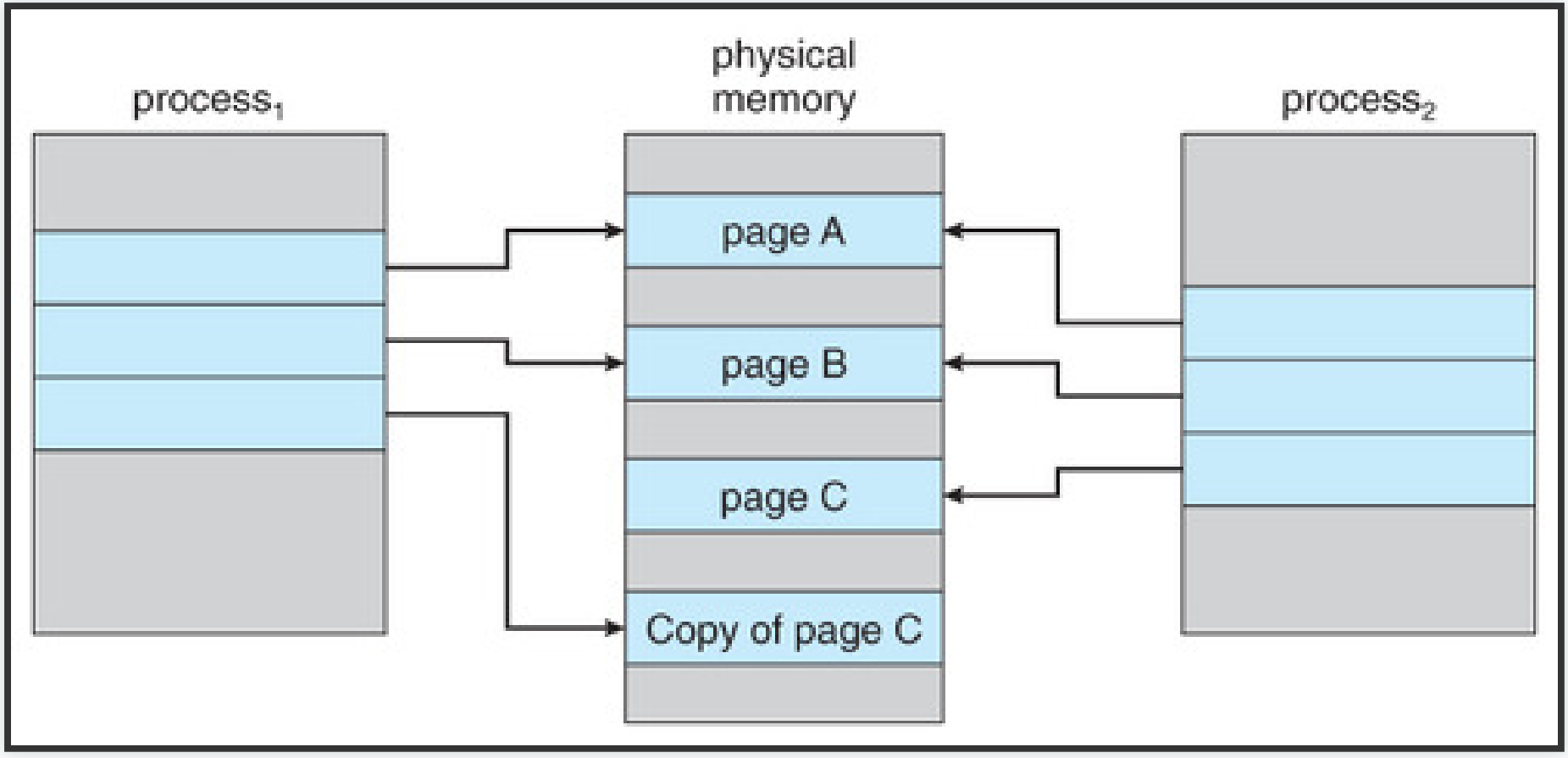
TLB (Translation Lookaside Buffer)



Типы ошибок

- Segmentation fault
  - ошибка выделения виртуальной памяти, выход за пределы разрешенной памяти
- Page fault
  - запрашиваемая страница не загружена в память. Начинается процесс подгрузки страницы с диска

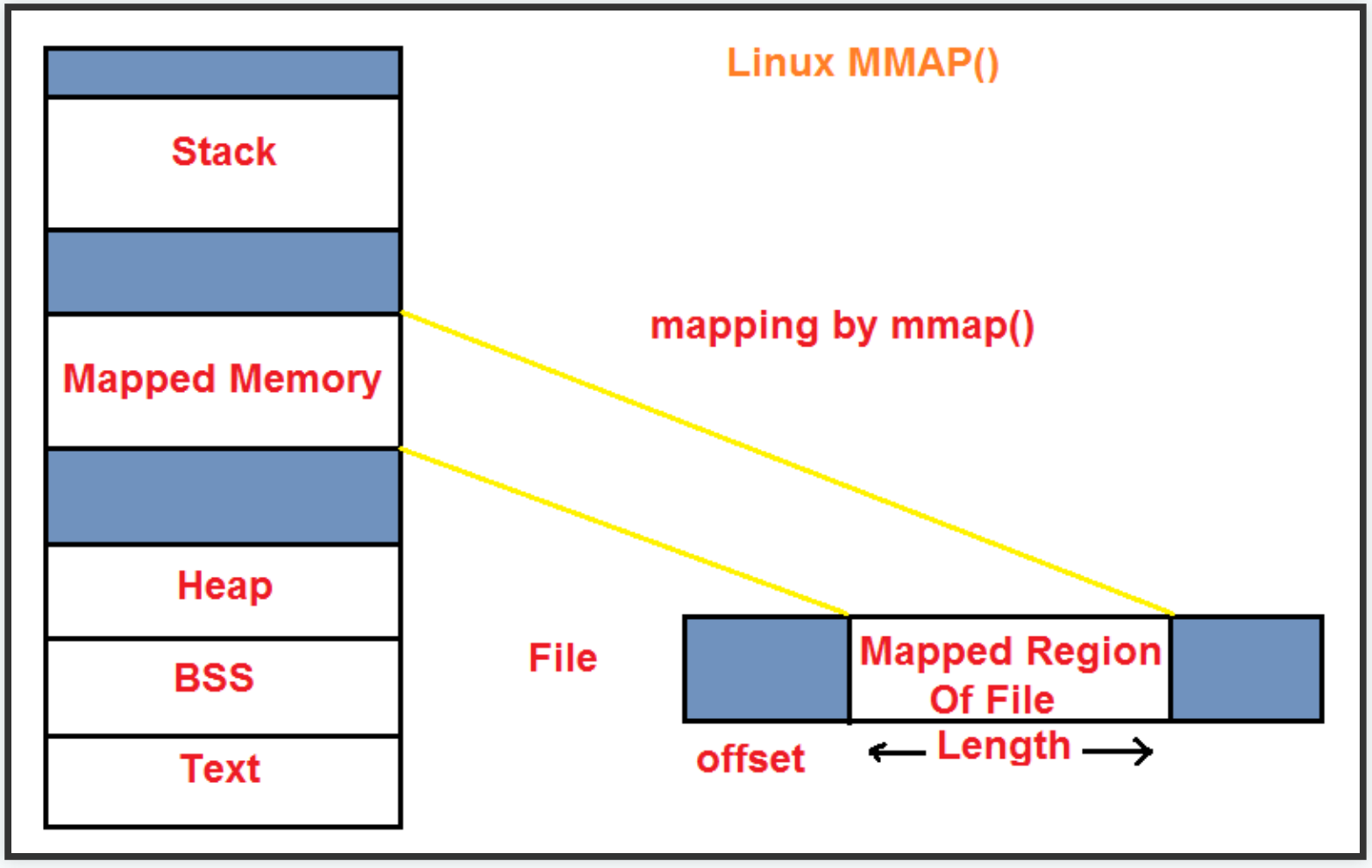
Copy on write



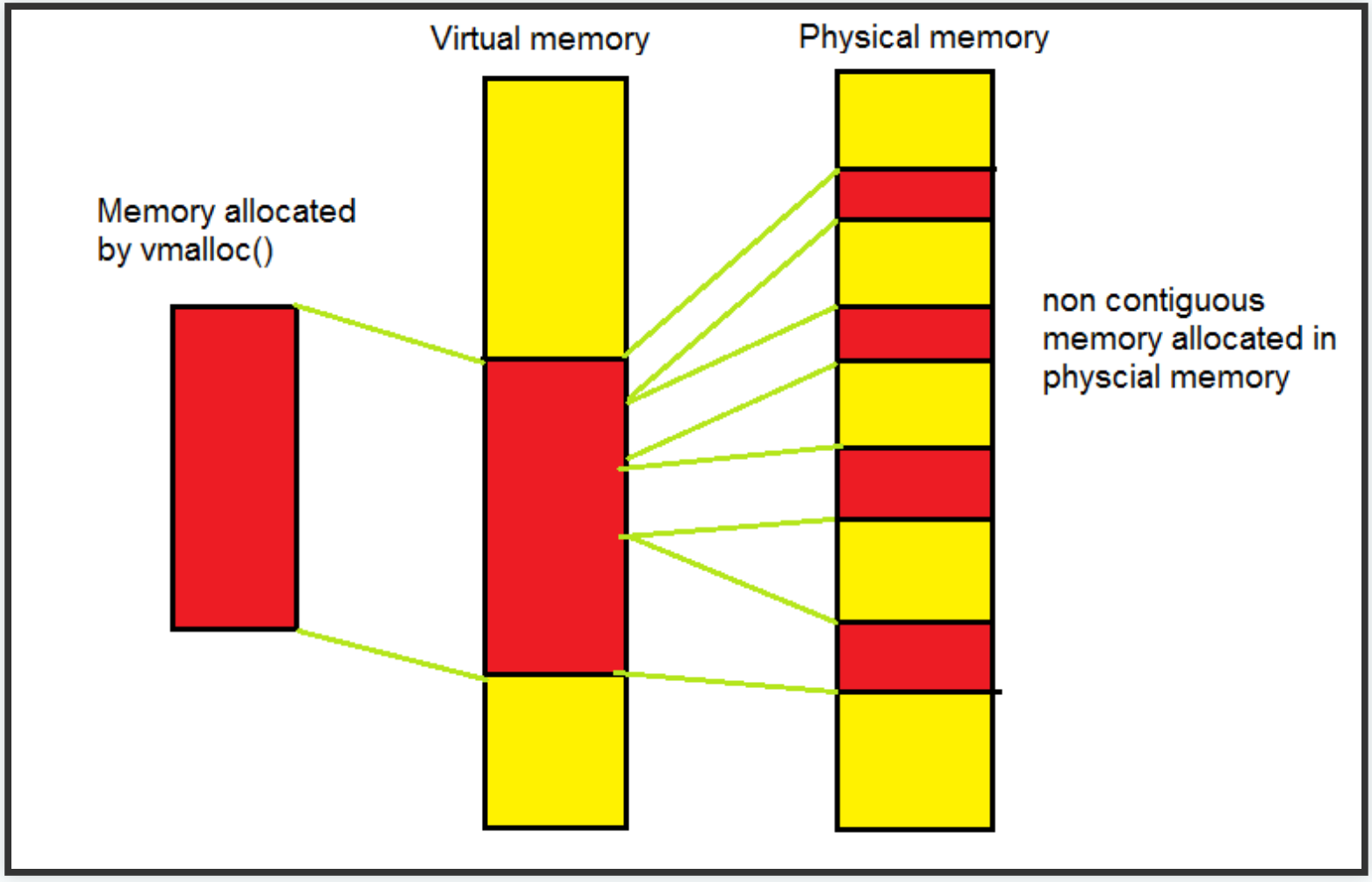
MMAP

mmap — POSIX-совместимый системный вызов Unix, позволяющий выполнить отображение файла или устройства на память. Является методом ввода-вывода через отображение файла на память и естественным образом реализует выделение страниц по запросу, поскольку изначально содержимое файла не читается с диска и не использует физическую память вообще. Реальное считывание с диска производится в «ленивом» режиме, то есть при осуществлении доступа к определённому месту

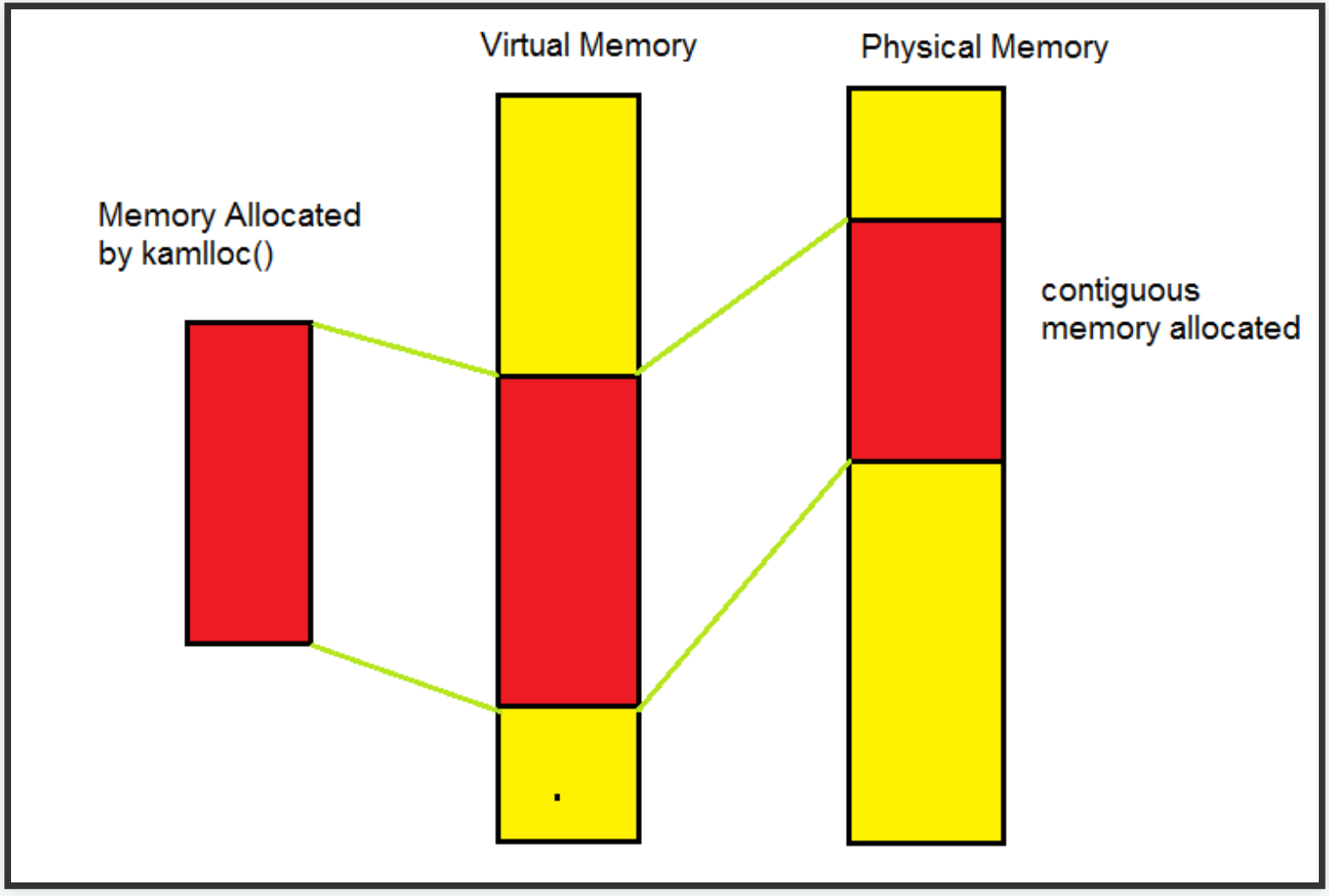
mmap()


Linux MMAP()
Stack
Mapped Memory
Heap
BSS
Text
mapping by mmap()
File
Mapped Region
Of File
offset
Length

vmalloc()


Virtual memory
Physical memory
Memory allocated
by vmalloc()
non contiguous
memory allocated in
physcial memory

kmalloc()


Virtual Memory
Physical Memory
Memory Allocated
by kamlloc()
contiguous
memory allocated

/proc/meminfo

- MemTotal
  - Доступный объем оперативной памяти. Часть физически доступной памяти резервируется во время запуска системы ядром и не входит в указанный здесь объем:
  - `grep Memory: /var/log/dmesg`
- MemFree
  - объем памяти доступеный для немедленного выделения процессам.
- MemAvailable
  - приблизительный объем доступный для старта новых приложений без использования swap
- Shmem
  - общий объем используемой разделяемой памяти

Buffer cache vs Page Cache

```
Buffers:            4172 kB
Cached:           859804 kB
```

- Cached
  - размер занятой памяти под блоки файлов, используемых для ускорения чтения и буферизации записи
- Buffers
  - до ядра 2.4 представляли блоки в памяти для записи на диск. Сейчас это только те блоки которые не принадлежат файлам и не представлены в page cache
- SwapCached
  - память, которая есть и в кэше и на диске

Active and inactive

- Active(anon)(file)
  - недавно используемая
- Inactive(anon)(file)
  - давно неиспользованная, рекомендована для сброса на диск, или освобождения
- Unevictable
  - память недоступная для освобождения или сброса на диск
- Mlocked
  - заблокированная с помощью mlock

/proc/meminfo

- Dirty: память ожидающая сброса на диск
- Writeback: память в настоящий момент записываемая на диск на диск
- AnonPages: страницы памяти не привязанные к файлам (THP)
- Mapped: файлы отображенные в память через mmap
- Slab: кэш структур данных ядра
- PageTables: память выделенная для PageTables

/proc/meminfo

- NFS_Unstable: NFS страницы, посланные на сервер, ожидающие подтверждения
- Bounce: память выделанна для bounce buffers блочных устройств
- CommitLimit: доступная для выделения память операющаяся на vm.overcommit_ratio
- Committed_AS: кол-во памяти выделенной системе в настоящий момент
- VmallocTotal: суммарный размер памяти под vmalloc
- VmallocUsed: используемый объем
- VmallocChunk: наибольший непрерываемый свободный блок vmalloc области

Hugepages

- HugePages_Total: число hugepages выделнных ядром (vm.nr_hugepages)
- HugePages_Free: число свободных hugepages
- HugePages_Rsvd: число выделяемых в настоязий момент hugepages
- Hugepagesize: размер страницы (обычно 2M)
- AnonHugePages - анонимные hugepages
- HugePages_Surp: число страниц в пуле больше vm.nr_hugepages. Контролируется vm.nr_overcommit_hugepages.

transparent hugepages

- встроенный и включенный механизм прозрачно подменяющий обычные страницы на hugepages
- работает с анонимными страницами и tmpfs/shmem
- для систем с СУБД рекомендуется отключать
- подробнее: https://www.kernel.org/doc/Documentation/vm/transhuge.txt

отключение transparent hugetables

```
cat /sys/kernel/mm/redhat_transparent_hugepage/enabled
[always] madvise never

echo never > /sys/kernel/mm/redhat_transparent_hugepage/enabled
echo never > /sys/kernel/mm/redhat_transparent_hugepage/defrag

cat /sys/kernel/mm/redhat_transparent_hugepage/enabled
always madvise [never]
```

Hugepages - ручное выставление

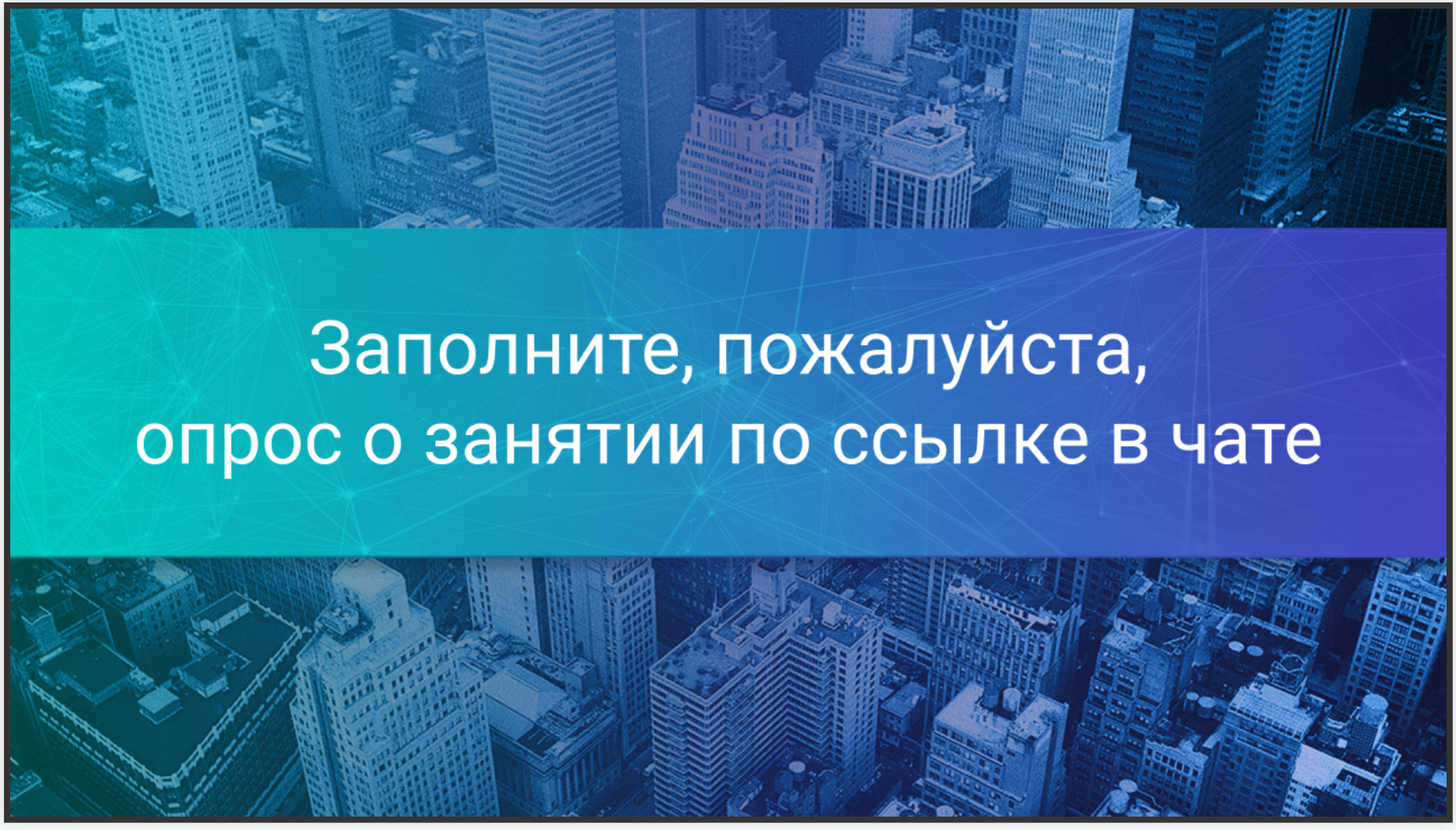
- sysctl
  - vm.nr_hugepages = 20480
- Number Hugepages * Hugepagesize = minimum Memlock
- vi /etc/security/limits.conf

```
oracle                soft    memlock 41943040
oracle                hard    memlock 41943040
```

Рефлексия

Отметьте 3 пункта, которые вам запомнились с вебинара

Что вы будете применять в работе из сегодняшнего вебинара?


Заполните, пожалуйста,
опрос о занятии по ссылке в чате