ㅇ 전체 메모리 구성

MemTotal = MemFree + User space memory + kernel space memory

User Memory = File-backed memory + Anonymous memory

1. File-backed memory : file에서 읽어 왔거나 앞으로 다시 file로 저장될 것들을 file-backed memory 영역으로 해서 메모리 관리자가 관리하며

Buffers, Cached, SwapCached가 있음

1. Anonymous memory : 메모리에 있으나 file에서 오지 않았으며, 앞으로도 File에 쓰여지지 않을 것들을 Anonymous memory로 해서 관리합니다
2. Acitve : 실행 중인 프로그램이나 데이터 , 실행할 예정인 프로그램이나 데이터

(Active 에는 Anonymous 와 File-backed memory)

1. Inactive : 실행 하지 않고 대기 중인(비활성) 프로그램이나 데이터

(Inacitve에도 Anonymous 와 File-backed memory)

ㅇ 메모리 사용주체

1. 응용프르그램 : application 용으로
2. 시스템 커널 : page cache , dentry cache, inode cache 등

ㅇ 시스템이 메모리가 부족할때 하는 2가지 메모리 회수 방법

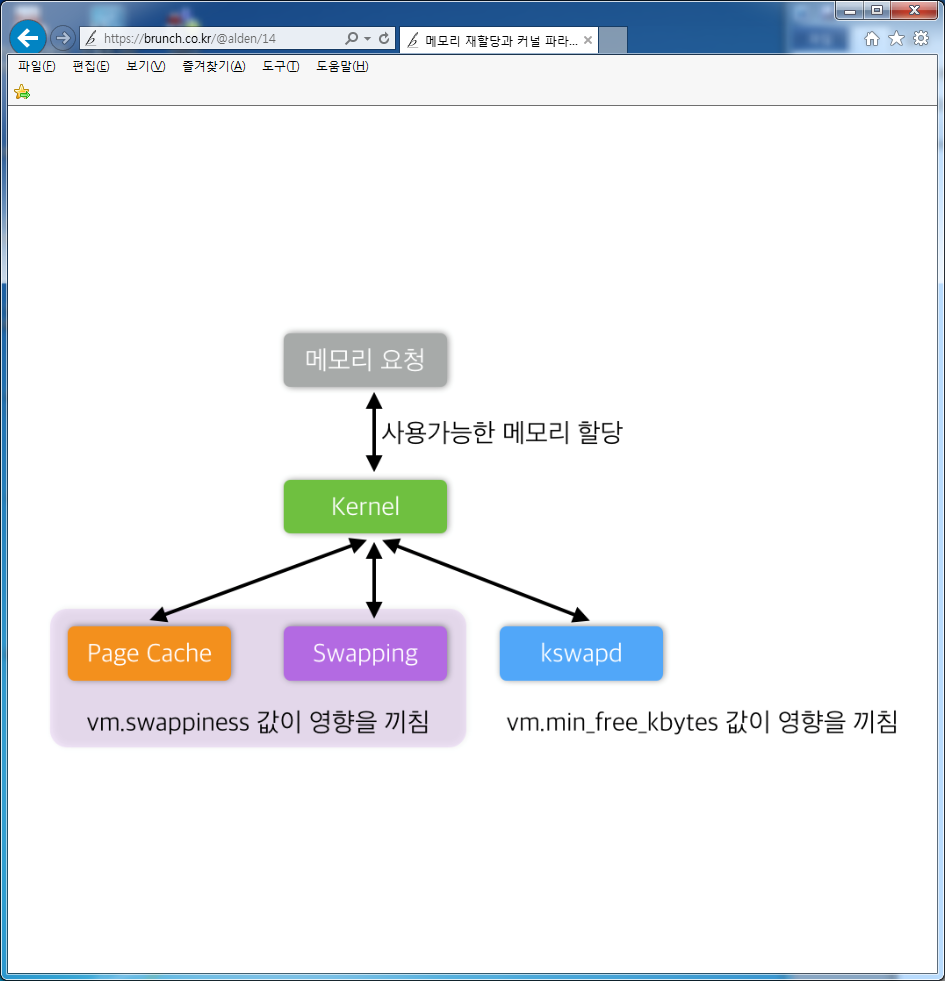
1. page cache 등 사용하지 않는 메모리를 다른 프로그램에 재할당 (cache 삭제)

* cache 삭제방식으로 시스템에 영향이 주지 않음

1. 가장 덜 사용되는 응용프로그램 메모리(inactive)를 회수하여 swap 파티션으로 이동후 다른 프로그램에 할당

* swaping 방식으로 메모리를 disk로 이동시키기 때문에 성능에 영향을 줌

ㅇ 메모리 회수 logic 3가지



ㅇ스왑(Swap) 이란?

스왑은 한정된 물리 메모리를 가지고 가상 메모리를 지원하는데 있어서 필수적인 기능이다. 태스크가 페이지 요청 시 물리적인 메모리가 부족할 경우, 페이지 폴트 발생하게 된다. 이 경우 스와핑(swapping)이 발생하며, 물리적인 메모리의 페이지 일부를 다른 장치(일반적으로 하드 디스크의 스왑영역을 사용)에 저장하고, 필요한 페이지를 할당하여 할당하게 된다.

ㅇ kswapd

커널 스왑 관리 데몬은 다른 커널 쓰레드들과 같이 깨어날 때까지 대기 상태로 전환하며, 스왑 요청에 의해 깨어나면 작업을 수행 하고 대기 상태로 다시 전환되는 작업을 반복한다. 커널 스왑 관리 데몬은 kswapd() 함수로부터 시작된다. 스와핑은 앞에서도 간략히 언급했듯이 페이지 폴트 익셉션이 발생하는 경우에 발생하게 된다. 페이지 폴트는 "1. 사용자 페이지"가 "2. 읽기/쓰기 접근 시"에 "3. 페이지가 존재하지 않는 경우" 세가지 조건이 만족될 때 발생하게 된다. 페이지 폴트가 발생하면, do\_page\_fault() 함수가 호출되며, 이 함수를 쫓아가다보면, wakeup\_kswapd() 함수를 호출하여 인터럽트를 발생시키고, 커널 스왑 관리 데몬을 깨우게 된다

ㅇ swap tendency 계산식

swap tendency= mapped ratio/2 + distress + swappiness => 100 인 경우 swap 발생

1. mapped\_ratiod: 현재 사용중인 메모리 비율( 조절 불가) (0<=x<=50)

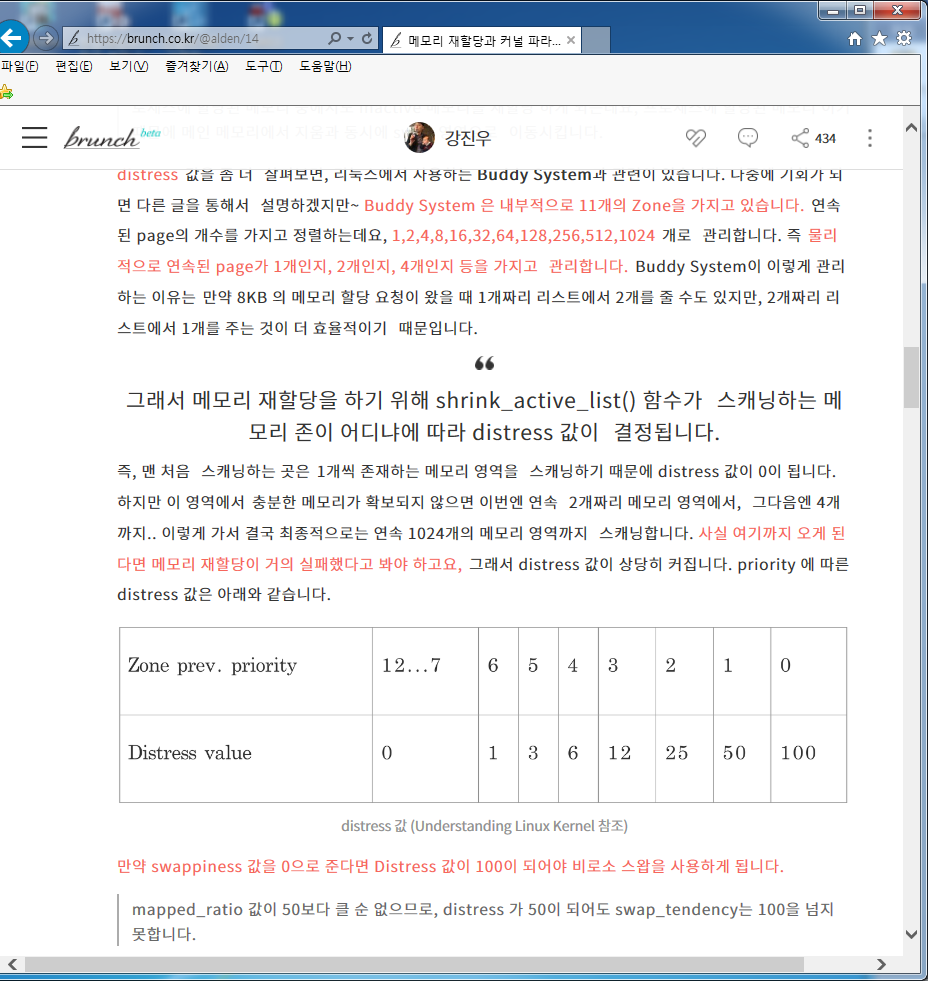
만약 전체 16G중에 8G 사용되고 있으면 mapped\_ration는 50%

1. distress: 메모리 확보가 얼마나 어려운 상태인지를 나타내는 값 (조절 불가)(0<=x<=100)

리눅스 커널이 활성 리스트에서 적은 개수의 페이지 검색을 시도해보고 페이지 회수에 실패하는 경우에 점점 더 많은 개수의 페이지를 scan하게됨 이때 priority값이 12->0으로 떨어짐에 따라

distress 값도 아래 테이블과 같이 변함 (distress =100/(2^priority)

scan시 연속된 메모리 영역이 커짐에 따라 priority값은 작아지고 distress value는 커짐



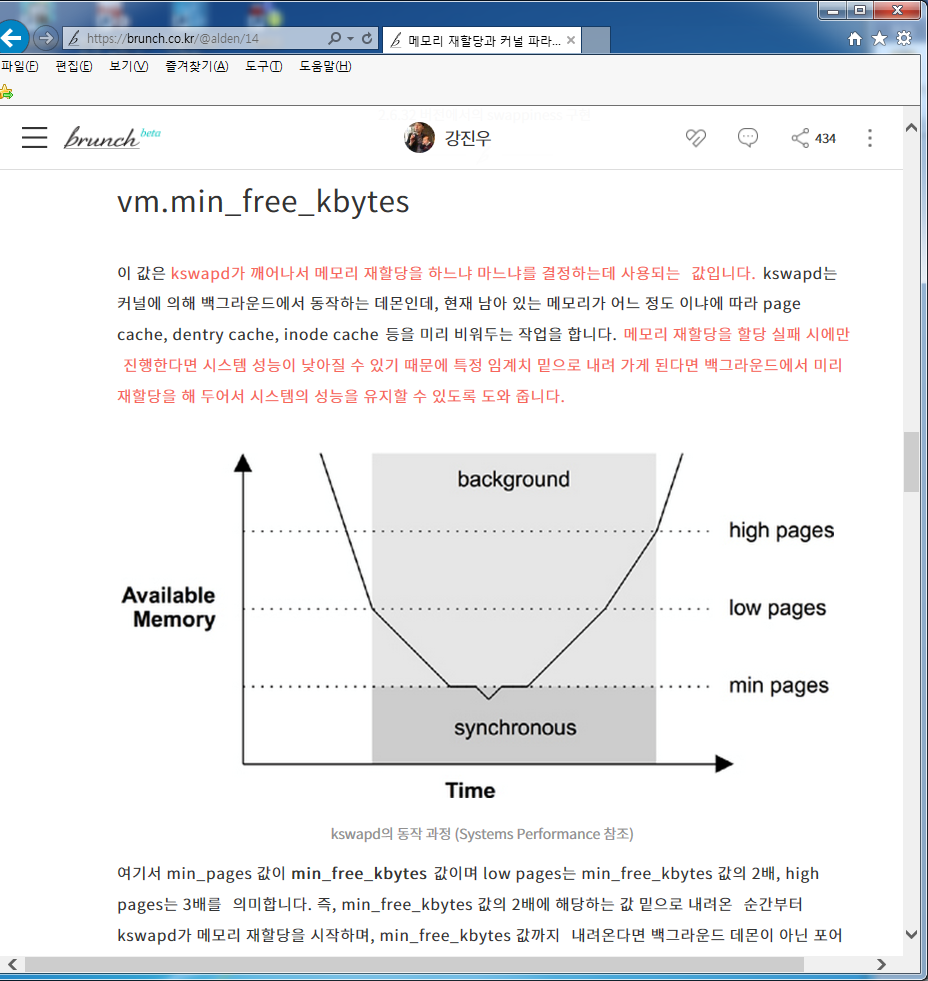
1. swappiness: vm.swappiness 커널 파라미터 (control 가능)

**vm.swapiness** : swap 활용수준 (조절 가능)(0<=x<=100)

|  |  |
| --- | --- |
| **값** | **설명** |
| vm.swappiness = 0 | 스왑 사용안함[1] |
| vm.swappiness = 1 | 스왑 사용 최소화 |
| vm.swappiness = 60 | 기본값 |
| vm.swappiness = 100 | 적극적으로 스왑 사용 |

**vfs\_cache\_pressure**: 이 값을 크게 하면 응용프로그램이 사용하지 않는 메모리는 바로 free되지 않고 buffer cache(dentry cache, inode cache) 에 있는데 이를 free로 만들어 주는 역할

**min\_free\_kbytes**: kswapd 데몬 프로세스가 현재 사용 가능한 메모리가 어느 정도 이냐에 따라 page cache, dentry cache, inode cache 등을 미리 비워두는 작업



min\_free\_kbytes\*3값

min\_free\_kbytes값

min\_free\_kbytes\*2값

low pages 밑으로 내려가면 kswapd가 백그라운드로 메모리 재할당을 시작하고, min pages 밑으로 내려오면 포그라운드로 메모리 재할당하며 high pages까지 재할당이 되면 동작을 멈춤

ㅇ swap 메모리 사용량순으로 프로세스 정렬

[root@controller Desktop(keystone\_admin)]# ps -eo user,pid,ppid,rss,size,vsize,pmem,pcpu,cmd --sort -size | head -n 11

USER PID PPID RSS SIZE VSZ %MEM %CPU CMD

mysql 1876 1 69956 1082928 1178304 1.9 0.7 /usr/libexec/mysqld --basedir=/usr

root 1838 1 9312 952560 1301740 0.2 0.0 /usr/sbin/libvirtd --listen

nova 2349 1 121988 813240 1206040 3.4 1.2 /usr/bin/python2 /usr/bin/nova-compute

root 3372 3150 178108 766736 1481004 5.0 0.9 /usr/bin/gnome-shell

[root@controller Desktop(keystone\_admin)]# free -m

total used free shared buff/cache available

Mem: 3416 3040 48 7 326 182

Swap: 1435 201 1234

[root@controller Desktop(keystone\_admin)]# cat /proc/meminfo

MemTotal: 3498276 kB : 약간의 예약된 byte들과 Kernel binary code를 제외한 물리적인 메모리의 사이즈

MemFree: 59816 kB : 시스템에 의해 사용되지 않고 남아있는 물리적인 메모리의 양

MemAvailable: 191720 kB :

Buffers: 4 kB : file buffer들의 위해 사용되는 물리적인 메모리의 양

Cached: 146928 kB : The amount of physical RAM used as cache memory

SwapCached: 44032 kB : cache 메모리로 사용되는 swap의 양. 한번 swap out 된 메모리는 다시 swap in 되지만, 여전히 swapfile안에는 남아있다.

Active: 1456908 kB : buffer 또는 cache 메모리의 총 양이며, 이것이 active이다. 이 메모리는 최근에 사용된 메모리이며 절대적으로 필요하지 않는다면, 보통 반환되지 않는다

Inactive: 1602372 kB : free하고 이용가능한 buffer 또는 cache 메모리의 총 양. 이 메모리는 최근에 사용되지 않은 메모리 이며, paging algorithm에 의해 다른 목적을 위해 재 반환 할 수 있다

Active(anon): 1403132 kB

Inactive(anon): 1527844 kB

Active(file): 53776 kB

Inactive(file): 74528 kB

Unevictable: 39212 kB

Mlocked: 39212 kB

SwapTotal: 1470460 kB : 물리적인 swap memory의 총 양

SwapFree: 1260152 kB : swap 메모리의 free 양

Dirty: 52 kB : 디스크로 write되기를 기다리는 메모리의 총 양.

Writeback: 0 kB : 디스크로 active하게 write 중인 메모리의 총 양

AnonPages: 2922192 kB

Mapped: 58364 kB : mmap명령을 사용하여 device, 파일 또는 librayr들을 mapping하는데 사용되는 메모리의 총 양.

Shmem: 8024 kB : 프로세스가 공유하는 메모리 영역

Slab: 183348 kB : 커널 자신이 사용하는 data structure를 cache하기 위해 커널에 의해 사용 되는 메모리의 총 양

SReclaimable: 64992 kB

SUnreclaim: 118356 kB

KernelStack: 15664 kB : 커널에서 프로세스가 전환이 일어날때 stack에 현재 상태 저장하는데 사용하는 메모리 영역

PageTables: 50316 kB : page 정보를 담고 있는 메타데이인 page table이 차지하는 메모리

NFS\_Unstable: 0 kB

Bounce: 0 kB

WritebackTmp: 0 kB

CommitLimit: 3219596 kB

Committed\_AS: 7063568 kB

VmallocTotal: 34359738367 kB : Virtual address 공간으로 할당된 총 메모리의 양.

VmallocUsed: 192448 kB : Virtual address 공간으로 사용되는 메모리의 총 양

VmallocChunk: 34359533896 kB : Virtual address 공간으로 이용가능한 가장 큰 연속적인 메모리의 block.

HardwareCorrupted: 0 kB

AnonHugePages: 245760 kB

HugePages\_Total: 0 : 시스템에 대한 총 hugepage의 수

HugePages\_Free: 0 : 시스템에 대한 이용가능한 총  hugepage의 수

HugePages\_Rsvd: 0

HugePages\_Surp: 0

Hugepagesize: 2048 kB : Hugepage  기본 크기(단위 KB),

DirectMap4k: 163648 kB

DirectMap2M: 3489792 kB

DirectMap1G: 0 kB

[root@controller Desktop(keystone\_admin)]# cat /proc/swaps

Filename Type Size Used Priority

/dev/dm-1 partition 1470460 210284 -1

coredump\_filter io mountstats personality stat

[root@localhost haproxy]# cat /proc/9332/status

Name: haproxy

State: S (sleeping)

Tgid: 9332

Ngid: 0

Pid: 9332

PPid: 1

TracerPid: 0

Uid: 188 188 188 188

Gid: 188 188 188 188

FDSize: 64 : 현재 할당되어 있는 file descriptor 슬롯의 갯수

Groups:

VmPeak: 49988 kB : 최고로 할당된 가상메모리 크기

VmSize: 49988 kB : 가상 메모리 크기

VmLck: 0 kB : lock된 메모리 크기

VmPin: 0 kB

VmHWM: 1692 kB : swap되지 않고 설정된 가장 큰 크기

VmRSS: 1664 kB : swap되지 않고 설정된 크기

VmData: 1136 kB : Data (정적변수) 세그먼트 크기

VmStk: 136 kB : 스택 (동적변수) 세그먼트 크기

VmExe: 824 kB : 텍스트(프로그램 코드와 상수) 세그먼트 크기

VmLib: 6504 kB : 라이브러리 코드의 크기

VmPTE: 108 kB : 페이지 테이블 엔트리 크기

VmSwap: 0 kB

Threads: 1

SigQ: 0/3790

SigPnd: 0000000000000000

ShdPnd: 0000000000000000

SigBlk: 0000000000000000

SigIgn: 0000000000001000

SigCgt: 0000000180300205

CapInh: 0000000000000000

CapPrm: 0000000000000000

CapEff: 0000000000000000

CapBnd: 0000001fffffffff

Seccomp: 0

Cpus\_allowed: ffffffff,ffffffff,ffffffff,ffffffff

Cpus\_allowed\_list: 0-127

Mems\_allowed: 00000000,00000000,00000000,00000000,00000000,00000000,00000000,00000000,00000000,00000000,00000000,00000000,00000000,00000000,00000000,00000000,00000000,00000000,00000000,00000000,00000000,00000000,00000000,00000000,00000000,00000000,00000000,00000000,00000000,00000000,00000000,00000001

Mems\_allowed\_list: 0

voluntary\_ctxt\_switches: 17176

nonvoluntary\_ctxt\_switches: 23

[root@localhost haproxy]# vmstat 5 5

procs -----------memory---------- ---swap-- -----io---- -system-- ------cpu-----

r b swpd free buff cache si so bi bo in cs us sy id wa st

4 0 36668 101128 0 328480 0 2 98 48 88 157 2 1 97 0 0

1 0 36668 101112 0 328480 0 0 0 0 281 440 6 1 93 0 0

0 0 36668 101128 0 328480 0 0 0 0 148 286 5 1 94 0 0

0 0 36668 101208 0 328448 0 0 0 0 104 203 1 0 98 0 0

0 0 36668 102080 0 328448 0 0 0 0 519 677 11 2 88 0 0



[root@localhost haproxy]# iostat

Linux 3.10.0-327.el7.x86\_64 (localhost) 11/29/2016 \_x86\_64\_ (1 CPU)

avg-cpu: %user %nice %system %iowait %steal %idle

1.58 0.48 0.79 0.03 0.00 97.11

Device: tps kB\_read/s kB\_wrtn/s kB\_read kB\_wrtn

sda 2.89 95.00 46.27 2176094 1059812

dm-0 2.76 94.54 44.54 2165559 1020252

dm-1 0.45 0.21 1.64 4724 37500

