



Performance et diagnostic Méthodologie et outils

- Allo Houston : ça marche pas - c'est lent ...
- Comment faire et ne pas faire

Qui suis-je ?

- Vladislav Pernin
- 15 ans d'expérience
- Responsable technique du pôle Nex'us de la DSI d'Enedis
- Affinités :
 - Performance
 - Monitoring
 - Système

Allo Houston : ça marche pas - c'est lent ...

- Comment procéder ?
- Notions & méthodologie
- Etat des lieux rapide
- Outils et diagnostic plus avancés
- Zoom sur une JVM
- Spécificités avec Docker
- Scénarios
- Surtout en ligne de commande
- Sur Linux
- Aperçu avancé ... de surface
- Pas un guide exhaustif
- Pas un expert système
- Pas de GIF

Notions et méthodologies

Notions de monitoring

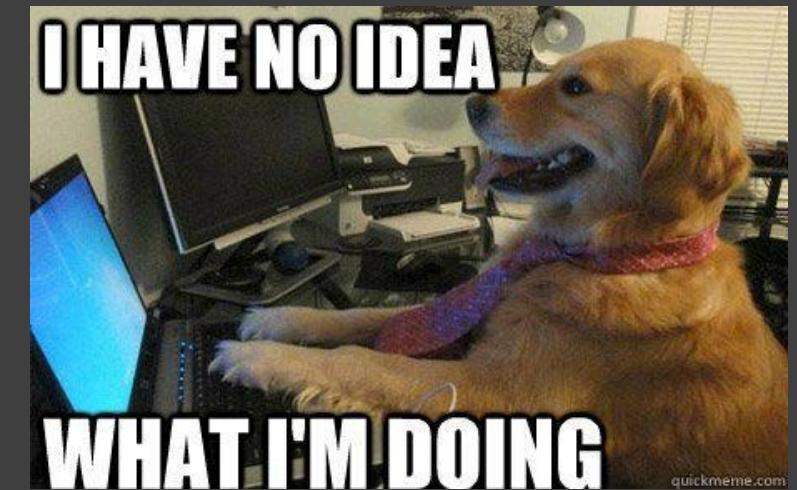
- Boîte noire :
 - Visible depuis l'extérieur, typiquement un user
 - Orienté symptôme
 - Montre des problèmes en cours
 - Base de l'alerting
- Boîte blanche :
 - Métriques exposées par l'intérieur du système
 - Orienté cause
 - Permet aussi de détecter des problèmes à venir, « signaux faibles »
- Prérequis : système observable

Notions de monitoring

- Symptômes ou causes, qu'est ce qui est cassé et pourquoi, ex :
 - Erreurs sur page web utilisateur : job batch sur même machine qui sature le CPU
 - Latence webservice : fibre optique à nettoyer
 - En chaîne, une cause contribuant un souci sur la brique suivante ...
- Sur incident important
 - Rarement une seule root cause
 - Plutôt un ensemble d'éléments contributeurs

Anti méthodologies

- Taper des commandes au hasard ou sauter sur le premier graphe bizarre
- Regarder les logs applicatifs et sauter sur le premier souci
- Googler et appliquer le premier Stack Overflow
- Blame someone else : pas ma faute, c'est le composant ... ou l'équipe ...
- Drunk man anti pattern
- Streetlight anti pattern
- Changer des settings au hasard et regarder



Méthodologie

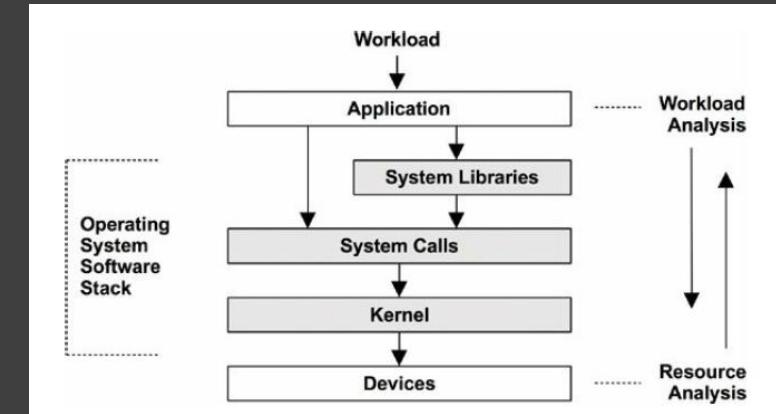
- Enfonçage de porte ouverte mais ...
 - Qu'est ce qui vous fait dire qu'il y a un problème ?
 - Est-ce que le système a déjà bien fonctionné ?
 - Qu'est ce qui a changé récemment ?
 - Qui est concerné, vous ou tout le monde?
 - Quel environnement : logiciels, matériels, versions, configuration, ... ?

Méthodologie

- Avoir une image mentale :
 - De l'architecture de l'application/SI
 - Du chemin des flux, leurs dépendances entre briques
 - Des chemins en lecture et en écriture dans les briques
 - De l'infrastructure sous jacente
 - Indispensable
 - Mais contradictoire à l'heure du cloud
 - Abstraction, « make it someone else problem »

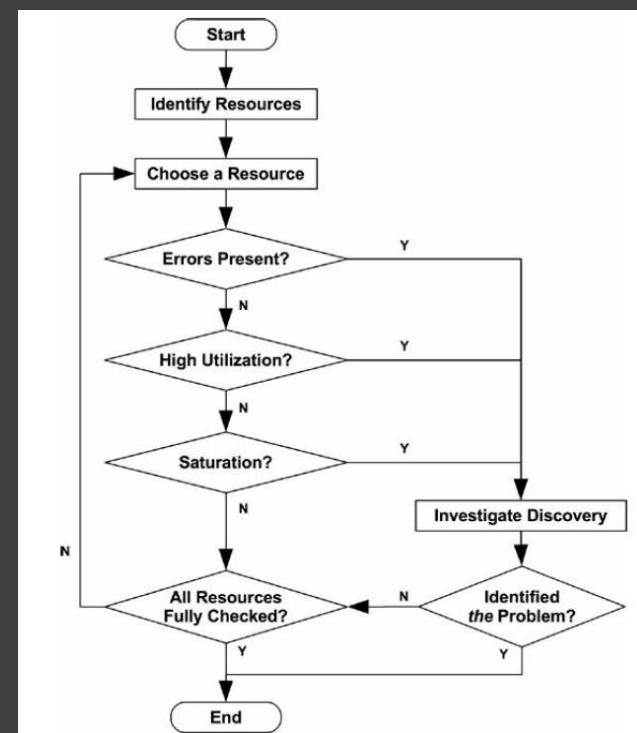
Méthodologie

- Deux approches complémentaires
 - Analyse des ressources :
 - IOPS, throughput, utilization, saturation
 - Pour les ressources système (cpu, mémoire, disque, réseau, ...)
 - Analyse de la charge :
 - Requests
 - Completion/errors
 - Latency
- Bon vernis sur toutes les couches préférable : du système à l'applicatif
- « Measure, don't guess »



USE method : Utilization, Saturation, Errors

- Pour chaque ressource physique : CPUs, mémoire, disque, réseau, ...
- Et aussi logicielles : lock, thread pool, file descriptor, ...
- Vérifier :
 - Utilization : sur un intervalle de temps, le % de temps occupé à travailler
 - Saturation : > 100 %, même occupé, encore possible d'accepter du travail, la saturation définit à quel point cela n'est plus possible
 - Errors : nombre d'erreurs
- Ne fournit que des pistes mais ferment pleins de mauvaises portes



USE method : Utilization, Saturation, Errors

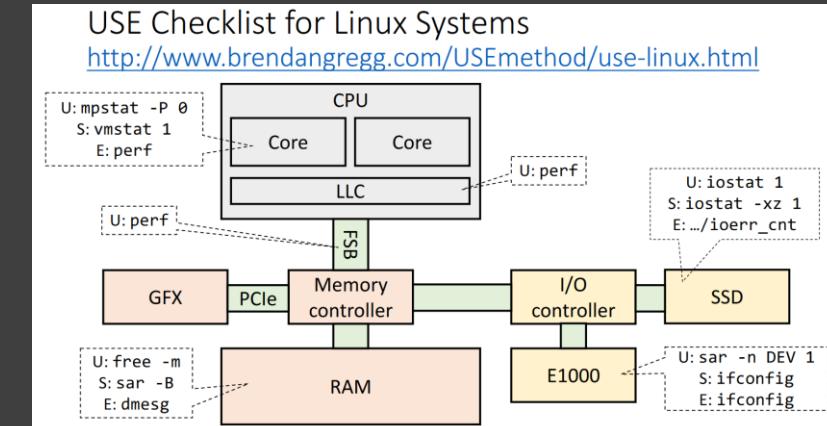
- Certaines métriques sont difficiles à mesurer (erreurs CPU), difficile sur VM, mais plutôt rare.
- Exemple avec le CPU

Resource	Metric	Linux
CPU	errors	perf (LPE) if processor specific error events (CPC) are available; eg, AMD64's "04Ah Single-bit ECC Errors Recorded by Scrubber" [4]
CPU	saturation	system-wide: vmstat 1, "r" > CPU count [2]; sar -q, "runq-sz" > CPU count; dstat -p, "run" > CPU count; per-process: /proc/PID/schedstat 2nd field (sched_info.run_delay); perf sched latency (shows "Average" and "Maximum" delay per-schedule); dynamic tracing, eg, SystemTap schedtimes.stp "queued(us)" [3]
CPU	utilization	system-wide: vmstat 1, "us" + "sy" + "st"; sar -u, sum fields except "%idle" and "%iowait"; dstat -c, sum fields except "idl" and "wai"; per-cpu: mpstat -P ALL 1, sum fields except "%idle" and "%iowait"; sar -P ALL, same as mpstat; per-process: top, "%CPU"; htop, "CPU%"; ps -o pcpu; pidstat 1, "%CPU"; per-kernel-thread: top/htop ("K" to toggle), where VIRT == 0 (heuristic). [1]

USE method : Utilization, Saturation, Errors

- Attention, une faible utilisation peut masquer des pics de charge ou des saturations
- Périmètre :
 - Qu'une sous partie des ressources matérielles ou virtuelles
 - On ne regarde par exemple pas souvent l'interconnect CPU
 - Connaissance : known unkowns, unknow unknowns
- Liste de ressources à vérifier et commandes associées :
<http://www.brendangregg.com/USEmethod/use-rosetta.html>

Resource	Utilization	Saturation	Errors
CPU	<code>mpstat -P ALL 1, sum non-idle fields</code>	<code>vmstat 1, "r"</code>	<code>perf</code>
Memory Capacity	<code>free -m, "used"/"total"</code>	<code>vmstat 1, "si"+"so"; dmesg grep killed</code>	<code>dmesg</code>
Storage I/O	<code>iostat -xz 1, "%util"</code>	<code>iostat -xnz 1, "avgqu-sz" > 1</code>	<code>/sys/.../ioerr_cnt</code> <code>smartctl</code>
Network	<code>nicstat, "%Util"</code>	<code>ifconfig, "overrunns"; netstat -s "retrans..."</code>	<code>ifconfig, "errors"</code>



RED method : Rate, Errors, Duration

- Pour les services
- Pas pour les ressources système
- Mode boîte noire (vue de l'extérieur) :
 - Rate : requêtes/s, séparer celles en succès et celles en erreur
 - Errors : ...
 - Duration (latency) : temps de réponse, temps d'attente/queue inclus



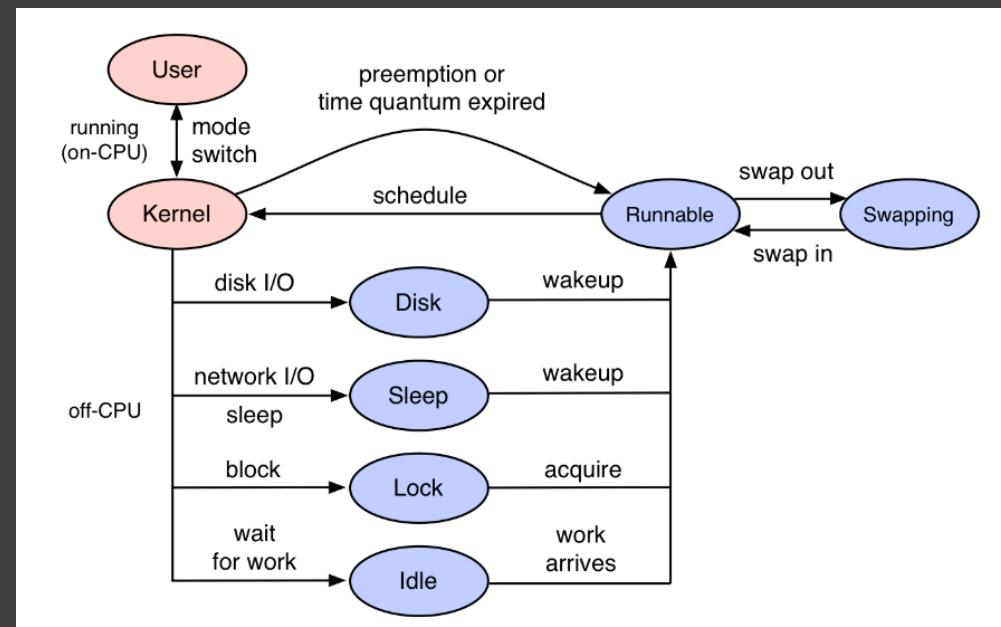
On retient : 5 golden signals

- Google SRE's 4 golden signals : Latency, Traffic, Errors, and Saturation
- « Intersection » de USE & RED
- **USERED** : Utilisation, Saturation, Erreurs, Rate, Duration (Latency)
- Comprendre les deux aspects :
 - Sollicitation utilisateur, service
 - Ressources et leurs comportements sur la charge
- A adapter selon ce que l'on observe

- Nécessite la plupart du temps de l'outillage, ex avec un serveur web :
 - Utilisation : selon le modèle (cnx/process, multi threadé, ...), % de workers occupés
 - Saturation : MaxClients, erreurs 503
 - Nb erreurs/s : parsing access logs, server-status, nb de 4XX, 5XX
 - Nb requêtes/s : idem depuis les logs, à affiner par contexte /uri/site/...
 - Latence : ..., percentile, max

Complément : thread et off CPU analysis

- Majeure partie des profilers et méthodes d'analyse : surtout le temps on CPU
- Pleins de raisons de ne pas être on CPU
- Off CPU : perspective intéressante mais difficile à interpréter
- On peut aussi travailler sur le wakeup
- Pour bien faire, kernel récent > 4.0, voir > 4.8 pas des 2.6 ou 3.10, RHEL 7.6 en beta



Pour faire tout cela correctement

- 4 piliers du monitoring :
 - Logs
 - Métriques
 - Traces
 - Health check

=> alerting
- Plus :
 - Profiling
 - Scripts adhoc
- Mais quand on est sur un serveur sans rien ...
et car on ne peut pas centraliser, historiser tout dans le détail

Etat des lieux rapide

Quelques commandes à connaître

- w
 - uptime
 - Idée de la charge du serveur, sur 1, 5 et 15 minutes
 - Qui est loggué (peut être aussi en train de faire des actions)
- date
 - Vérifier qu'il n'y a pas un écart trop important avec la vraie date ou entre machines
 - Peut être très pénalisant sur des systèmes distribués

```
08:49:02 up 592 days, 19:46, 1 user, load average: 1.62, 2.19, 2.30
USER      TTY      FROM          LOGIN@    IDLE      JCPU      PCPU WHAT
root      pts/1      [REDACTED]  08:48      0.00s   0.15s      0.00s w
```

- uname -srml
 - version du kernel
- cat /etc/redhat-release ou /etc/lsb-release
 - version de l'OS

```
Linux 4.13.0-21-generic x86_64
```

```
Red Hat Enterprise Linux Server release 7.1 (Maipo)
```

- timedatectl
 - Sur systemd

```
[root@... ~]# timedatectl
  Local time: Wed 2019-05-15 14:29:31 CEST
  Universal time: Wed 2019-05-15 12:29:31 UTC
    RTC time: Wed 2019-05-15 13:29:31
    Time zone: Europe/Paris (CEST, +0200)
  NTP enabled: yes
  NTP synchronized: yes
    RTC in local TZ: yes
      DST active: yes
Last DST change: DST began at
                  Sun 2019-03-31 01:59:59 CET
                  Sun 2019-03-31 03:00:00 CEST
Next DST change: DST ends (the clock jumps one hour backwards) at
                  Sun 2019-10-27 02:59:59 CEST
                  Sun 2019-10-27 02:00:00 CET
```

- dmesg

- -T sur RHEL > 7 pour avoir la date
- oom-killer, trace kernel, panne matérielle, ...

```
INFO: task stat:7010 blocked for more than 120 seconds.          x86_64 #1
    Tainted: G          --L-----
"echo 0 > /proc/sys/kernel/hung_task_timeout_secs" disables this message.
stat      D 0000000000000005  0 7010  7007 0x000000080
ffff8804368abbe8 0000000000000082 0000000000000000 ffff8804368abbac
ffff8804368abb48 ffffff81014b19 00bf8afeb86c5ae0 ffffff810b29cf
0000000000000400 0000000d8ae83564 ffff880164d27ad8 ffff8804368abfd8
Call Trace:
[<ffff8804368abbe8>] ? read_tsc+0x9/0x10
[<ffff8804368abb48>] ? ktime_get_ts+0xb/0x100
[<ffff8804368abb48>] ? sync_page+0x0/0x50
[<ffff8804368abb48>] ? io_schedule+0x73/0xc0
[<ffff8804368abb48>] ? sync_page+0x3d/0x50
[<ffff8804368abb48>] ? __wait_on_bit+0x5f/0x90
[<ffff8804368abb48>] ? wait_on_page_bit+0x73/0x80
[<ffff8804368abb48>] ? wake_bit_function+0x0/0x50
[<ffff8804368abb48>] ? pagevec_lookup_tag+0x25/0x40
[<ffff8804368abb48>] ? wait_on_page_writeback_range+0xfb/0x190
[<ffff8804368abb48>] ? filemap_fdatawait+0x2f/0x40
[<ffff8804368abb48>] ? filemap_write_and_wait+0x44/0x60
[<ffff8804368abb48>] ? nfs_getattr+0x185/0x1a0 [nfs]
[<ffff8804368abb48>] ? vfs_getattr+0x51/0x80
[<ffff8804368abb48>] ? vfs_fstatat+0x64/0xa0
[<ffff8804368abb48>] ? vfs_lstat+0x1e/0x20
[<ffff8804368abb48>] ? sys_newlstat+0x24/0x50
[<ffff8804368abb48>] ? audit_syscall_entry+0x1d7/0x200
[<ffff8804368abb48>] ? system_call_fastpath+0x16/0x1b
```

```
Call Trace:
[<ffff8804368abb48>] ? dump_header+0x90/0x1b0
[<ffff8804368abb48>] ? security_real_capable_noaudit+0x3c/0x70
[<ffff8804368abb48>] ? oom_kill_process+0x82/0x2a0
[<ffff8804368abb48>] ? select_bad_process+0xe1/0x120
[<ffff8804368abb48>] ? out_of_memory+0x220/0x3c0
[<ffff8804368abb48>] ? __alloc_pages_nodemask+0x93c/0x950
[<ffff8804368abb48>] ? alloc_pages_current+0xaa/0x110
[<ffff8804368abb48>] ? __page_cache_alloc+0x87/0x90
[<ffff8804368abb48>] ? find_get_page+0x1e/0xa0
[<ffff8804368abb48>] ? filemap_fault+0x1a7/0x500
[<ffff8804368abb48>] ? __do_fault+0x54/0x530
[<ffff8804368abb48>] ? handle_pte_fault+0xf7/0xb20
[<ffff8804368abb48>] ? alloc_pages_current+0xaa/0x110
[<ffff8804368abb48>] ? pte_alloc_one+0x37/0x50
[<ffff8804368abb48>] ? vma_prio_tree_insert+0x41/0x60
[<ffff8804368abb48>] ? handle_mm_fault+0x299/0x3d0
[<ffff8804368abb48>] ? __do_page_fault+0x146/0x500
[<ffff8804368abb48>] ? do_mmap_pgoff+0x335/0x380
[<ffff8804368abb48>] ? do_page_fault+0x3e/0xa0
```

```
.kernel: bnx2x 0000:0d:00.1: eth1: NIC Link is Down
kernel: bnx2x 0000:0d:00.1: eth1: NIC Link is Up, 10000 Mbps full duplex, Flow control: ON - receive & transmit
kernel: [<ffff8804368abb48>] ? bnx2x_rx_int+0x228/0x1530 [bnx2x]
kernel: [<ffff8804368abb48>] ? bnx2x_rx_int+0x228/0x1530 [bnx2x]
kernel: [<ffff8804368abb48>] ? bnx2x_poll+0xab/0xe0 [bnx2x]
kernel: [<ffff8804368abb48>] ? bnx2x_rx_int+0x228/0x1530 [bnx2x]
```

```
Out of memory: Kill process 28845 (influxd) score 777 or sacrifice child
Killed process 28845, UID 492, (influxd) total-vm:178192132kB, anon-rss:26219876kB, file-rss:444kB
```

```
[Wed Sep 12 03:01:15 2018] CPU30: Package temperature above threshold, cpu clock throttled (total events = 3050604)
[Wed Sep 12 03:01:15 2018] mce: [Hardware Error]: Machine check events logged
```

- journalctl -r

• systemctl list-units --type=service

```
-> systemctl list-units --type=service
UNIT          LOAD  ACTIVE SUB   DESCRIPTION
atd.service    loaded active running Job spooling tools
auditt.service loaded active running Security Auditing Service
collectd.service loaded active running Collectd statistics daemon
crond.service  loaded active running Command Scheduler
dbus.service   loaded active running D-Bus System Message Bus
funcd.service  loaded active exited  LSB: Fedora Unified Network Control
getty@tty1.service loaded active running Getty on tty1
goferd.service loaded active running Gofer Agent
gssproxy.service loaded active running GSSAPI Proxy Daemon
● ipmievd.service loaded failed failed  Ipmievdaemon
irqbalance.service loaded active running irqbalance daemon
kmod-static-nodes.service loaded active exited Create list of required static device nodes for
lvm2-lvmetad.service loaded active running LVM2 metadata daemon
lvm2-monitor.service loaded active exited Monitoring of LVM2 mirrors, snapshots etc. usin
lvm2-pvscan@8:2.service loaded active exited LVM2 PV scan on device 8:2
network.service loaded active exited  LSB: Bring up/down networking
node_exporter.service loaded active running Prometheus exporter for machine metrics, written
ntp.service    loaded active running Network Time Service
● ntpdate.service loaded failed failed  Set time via NTP
polkit.service  loaded active running Authorization Manager
postfix.service loaded active running Postfix Mail Transport Agent
rhel-dmesg.service loaded active exited Dump dmesg to /var/log/dmesg
rhel-import-state.service loaded active exited Import network configuration from initramfs
rhel-readonly.service loaded active exited Configure read-only root support
rpcbind.service loaded active running RPC bind service
rsyslog.service loaded active running System Logging Service
sshd.service   loaded active running OpenSSH server daemon
sssd.service   loaded active running System Security Services Daemon
sysstat.service loaded active exited Resets System Activity Logs
systemd-binfmt.service loaded active exited Set Up Additional Binary Formats
systemd-journal-flush.service loaded active exited Flush Journal to Persistent Storage
systemd-journald.service loaded active running Journal Service
systemd-logind.service loaded active exited Load/Save Random Seed
systemd-random-seed.service loaded active exited Remount Root and Kernel File Systems
systemd-remount-fs.service loaded active exited Apply Kernel Variables
systemd-sysctl.service loaded active exited Time & Date Service
systemd-timedated.service loaded active running Create Static Device Nodes in /dev
systemd-tmpfiles-setup-dev.service loaded active exited Create Volatile Files and Directories
systemd-tmpfiles-setup.service loaded active exited udev Coldplug all Devices
systemd-udev-trigger.service loaded active exited udev Kernel Device Manager
systemd-udevd.service  loaded active running Update UTMP about System Boot/Shutdown
systemd-update-utmp.service loaded active exited Permit User Sessions
systemd-user-sessions.service loaded active exited Setup Virtual Console
systemd-vconsole-setup.service loaded active running Dynamic System Tuning Daemon
tuned.service   loaded active running Service for virtual machines hosted on VMware
vmtoolsd.service loaded active running workflow-server
workflow-server.service loaded active running workflow-server
xineted.service loaded active running Xinetd A Powerful Replacement For Inetd

LOAD  = Reflects whether the unit definition was properly loaded.
ACTIVE = The high-level unit activation state, i.e. generalization of SUB.
SUB   = The low-level unit activation state, values depend on unit type.

48 loaded units listed. Pass --all to see loaded but inactive units, too.
To show all installed unit files use 'systemctl list-unit-files'.
```

- `free -m | g`
 - RAM totale
 - Vraiment libre (free), cached ... mais pas forcément mobilisable instantanément
 - Disponible -/+ buffers/cache avant RHEL 7

	<code>total</code>	<code>used</code>	<code>free</code>	<code>shared</code>	<code>buffers</code>	<code>cached</code>
<code>Mem:</code>	193530	188279	5250	3	4189	153336
<code>-/+ buffers/cache:</code>		30752	162777			
<code>Swap:</code>	49151	0	49151			

- Available depuis RHEL 7, ajouter -w pour avoir le détail buffer/cache

	<code>total</code>	<code>used</code>	<code>free</code>	<code>shared</code>	<code>buffers</code>	<code>cache</code>	<code>available</code>
<code>Mem:</code>	3791	1419	1139	189	32	1200	1900
<code>Swap:</code>	2047	1	2046				

- “Estimation of how much memory is available for starting new applications, without swapping. Unlike the data provided by the cache or free fields, this field takes into account page cache and also that not all reclaimable memory slabs will be reclaimed due to items being in use”
- Swap : ca n'est pas parce used != 0 que cela swap, ca peut avoir swappé pour un vrai souci, ou juste le swappiness > 1

- nproc
 - Nombre de core visible
 - Important pour interpréter la charge
- lscpu
 - Plus de détail sur la topologie processeur
 - Nombre de socket, core, hyperthreading
 - Fréquence
 - Nombre de nœuds NUMA
 - dédier « partie de RAM » par socket
 - Attention sur VM : la description est ...

```

Architecture:           x86_64
CPU op-mode(s):        32-bit, 64-bit
Byte Order:             Little Endian
CPU(s):                12
On-line CPU(s) list:   0-11
Thread(s) per core:    1
Core(s) per socket:    1
Socket(s):              12
NUMA node(s):           2
Vendor ID:              GenuineIntel
CPU family:             6
Model:                 47
Stepping:               2
CPU MHz:                1997.834
BogoMIPS:               3995.66
Hypervisor vendor:     VMware
Virtualization type:   full
L1d cache:              32K
L1i cache:              32K
L2 cache:                256K
L3 cache:                18432K
NUMA node0 CPU(s):      0-5
NUMA node1 CPU(s):      6-11

```

```

Architecture:           x86_64
CPU op-mode(s):        32-bit, 64-bit
Byte Order:             Little Endian
CPU(s):                64
On-line CPU(s) list:   0-63
Thread(s) per core:    2
Core(s) per socket:    16
Socket(s):              2
NUMA node(s):           2
Vendor ID:              GenuineIntel
CPU family:             6
Model:                 79
Stepping:               1
CPU MHz:                2600.000
BogoMIPS:               5193.71
Virtualization:         VT-x
L1d cache:              32K
L1i cache:              32K
L2 cache:                256K
L3 cache:                40960K
NUMA node0 CPU(s):      0-15,32-47
NUMA node1 CPU(s):      16-31,48-63

```

- vmstat -w -Sm 2
 - Outil magique pour avoir un vue globale rapide de l'activité d'un serveur : cpu, charge, mémoire, swap, io
 - Ignorer la 1ère ligne

procs	memory				swap		io		system				cpu			
r	b	swpd	free	buff	cache	si	so	bi	bo	in	cs	us	sy	id	wa	st
1	0	945608	438432	564236	513040	0	0	5	33	0	0	25	3	72	0	0
4	0	945608	438120	564236	513072	0	0	0	32	5564	2413	89	7	4	0	0
3	0	945608	438416	564244	513076	0	0	0	96	6328	2397	87	8	6	0	0
6	0	945608	438416	564244	513088	0	0	0	38	5549	2596	87	8	6	0	0

- Souci si r très supérieur au nombre de core
- Souci si ou so (swap in/out) différent de 0
- cs : souci si > 50 000 par core (voir pidstat ...)
- Colonnes cpu :
 - Activité CPU séparée en user, system, idle et iowait (stolen jamais valorisée)
 - Si wa important, attente sur les disques (SAN compris)
 - Si sy > 10-20%, possible souci, dépend aussi de l'activité user
- -a : affichage partie active et inactive

- mpstat -P ALL 2 ou sar -P ALL 2

- Usage détaillé par CPU

02:37:47 PM	CPU	%usr	%nice	%sys	%iowait	%irq	%soft	%steal	%guest	%idle
02:37:49 PM	all	0.25	0.00	0.19	0.00	0.00	0.00	0.00	0.00	99.56
02:37:49 PM	0	0.50	0.00	0.50	0.00	0.00	0.00	0.00	0.00	99.00
02:37:49 PM	1	0.00	0.00	0.51	0.00	0.00	0.00	0.00	0.00	99.49
02:37:49 PM	2	0.50	0.00	0.50	0.00	0.00	0.00	0.00	0.00	99.00
02:37:49 PM	3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00
02:37:49 PM	4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00
02:37:49 PM	5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00
02:37:49 PM	6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00
02:37:49 PM	7	0.50	0.00	0.50	0.00	0.00	0.00	0.00	0.00	99.00

- sar -q 2

- Run queue

02:41:07 PM	runq-sz	plist-sz	ldavg-1	ldavg-5	ldavg-15
02:41:09 PM	0	814	0.11	0.10	0.09
02:41:11 PM	0	814	0.11	0.10	0.09
Average:	0	814	0.11	0.10	0.09

- top -c
 - Vue globale triable cpu P, mémoire M
 - Load average : indice sur la charge du serveur, dépend du nb de core (interprétation subjective)
 - %CPU process : 100% == 1 core
 - Switch H pour afficher les threads
 - Switch 1 pour le détail par core
 - Peut manquer des short-lived process
 - VIRT : mem virtuelle, non allouée physiquement ; souvent écart énorme avec RES (malloc per thread arena)

```
top - 13:50:50 up 652 days, 2:08, 1 user, load average: 2.83, 2.60, 2.54
Tasks: 261 total, 1 running, 260 sleeping, 0 stopped, 0 zombie
Cpu(s): 82.2%us, 4.5%sy, 5.8%ni, 4.5%id, 0.0%wa, 0.5%hi, 2.5%si, 0.0%st
Mem: 3924412k total, 3518540k used, 405872k free, 566552k buffers
Swap: 2097148k total, 945580k used, 1151568k free, 538328k cached

PID USER PR NI VIRT RES SHR S %CPU %MEM TIME+ COMMAND
22916 20 0 4828m 639m 7864 S 175.6 16.7 33999:25 /logiciels/java
21321 39 19 3039m 300m 5632 S 12.3 7.8 15162:23 /logiciels/java/
  12 root 20 0 0 0 S 0.3 0.0 1303:25 [events/1]
  7078 root 20 0 230m 2952 1644 S 0.3 0.1 269:12.41 /usr/sbin/vmtoolsd
12974 20 0 4958m 847m 5356 S 0.3 22.1 475:07.23 /logiciels/
17593 20 0 1302m 9784 2384 S 0.3 0.2 0:06.12 /logiciels/
20745 root 20 0 27620 1496 1024 R 0.3 0.0 0:00.03 top -c
22833 root 20 0 1281m 9.8m 1544 S 0.3 0.3 3720:33 /usr/sbin/collectd -c /etc/collectd.conf -f
  1 root 20 0 23504 1316 1088 S 0.0 0.0 0:15.00 /sbin/init
  2 root 20 0 0 0 S 0.0 0.0 0:03.22 [kthreadd]
  3 root RT 0 0 0 0 S 0.0 0.0 4:44.34 [migration/0]
  4 root 20 0 0 0 S 0.0 0.0 10:35.60 [ksoftirqd/0]

in/java -server -Xms512m -Xmx512m -server -XX:+UseParNewGC -XX:+UseConcMarkSweep
java -XX:+UseParNewGC -XX:+UseConcMarkSweep
in/java -server -Xms3g -Xmx3g -Xmn850m -f /appli/projects/sik/apache_2.4/conf

```

- ps faux

- Vue complète de tout ce qui tourne sur cette machine
- User, pid, commande
- ww pour avoir la ligne de commande complète

```

ntp    2130  0.0  0.0  33616  1652 ?      Ss   2017  1:19 /usr/sbin/ntpd -u ntp:ntp -g
root   12745  0.0  0.0  472328  4516 ?      Ssl  2017  20:31 /usr/sbin/nsrexecd
root   9104  0.0  0.0  6444   556 ?      S<  2017  4:04 /usr/sbin/netatopd
root   25270  0.0  0.0  120420  3056 ?      S1   2017  0:02
root   25349  0.1  0.1  1366604 42748 ?      S1   2017 1111:05
root   23884  0.0  0.0  227764  9504 ?      S   2017  1:48 /bin/python /usr/bin/funcl --daemon
telegraf 20220 0.3  0.0  751908 19280 ?      Ssl  Mar26 1006:00 /usr/bin/telegraf -config /etc/telegraf/telegraf.conf -config-directory /etc/telegraf/telegraf.d
root   6283  0.0  0.0  9364   2176 ?      S   May14  4:36
root   6336  0.0  0.0  5852   1504 ?      S   May14  7:48
root   6338  0.0  0.0  5852   1548 ?      S   May14  1:23
root   30341 0.0  0.1  1000256 42940 ?      Ssl  Jun11  3:05 /usr/sbin/rsyslogd -n
root   31709 0.0  0.0  82996  2164 ?      Ss   Jun26  0:10 /usr/sbin/sshd -D
root   23195 0.1  0.0  161948  5356 ?      Ss   16:39  0:00  \_ sshd:          [priv]
root   23199 0.0  0.0  162080  2524 ?      Ss   16:39  0:00  \_ sshd:          Pts/0
root   23200 0.0  0.0  117136  2004 pts/0   Ss   16:39  0:00  \_ -ksh
root   23235 0.0  0.0  150124 2080 pts/0   R+  16:39  0:00  \_ ps fauxw
root   21902 0.4  0.1  1813400 37360 ?      Ssl  Jul12  652:54 /usr/sbin/collectd
root   11599 0.0  0.0  126332  1496 ?      Ss   Jul19  0:48 /usr/sbin/crond -n
root   31836 0.0  0.0  245000  1964 ?      Ss   Aug06  4:22 /usr/sbin/sssd -D -f
root   31837 0.0  0.0  289696  7220 ?      Ss   Aug06  2:49  \
root   31838 0.0  0.0  255012  3272 ?      Ss   Aug06  1:24  \_ /usr/libexec/sssd/sssd_nss --uid 0 --gid 0 --debug-to-files
root   31839 0.0  0.0  237900  2336 ?      Ss   Aug06  1:07  \_ /usr/libexec/sssd/sssd_pam --uid 0 --gid 0 --debug-to-files
root   31840 0.0  0.0  238012  2772 ?      Ss   Aug06  1:07  \_ /usr/libexec/sssd/sssd_sudo --uid 0 --gid 0 --debug-to-files
zabbix 7683  0.0  0.0  85132   1356 ?      Ss   Aug16  0:00 /usr/sbin/zabbix_agentd -c /etc/zabbix/zabbix_agentd.conf
zabbix 7684  0.1  0.0  85208   2184 ?      Ss   Aug16  182:17 \_ /usr/sbin/zabbix_agentd: collector [idle 1 sec]
zabbix 7685  0.1  0.0  87672   3448 ?      Ss   Aug16  158:37 \_ /usr/sbin/zabbix_agentd: listener #1 [waiting for connection]
zabbix 7686  0.1  0.0  100572  3524 ?      Ss   Aug16  158:42 \_ /usr/sbin/zabbix_agentd: listener #2 [waiting for connection]
zabbix 7687  0.1  0.0  87672   3420 ?      Ss   Aug16  158:03 \_ /usr/sbin/zabbix_agentd: listener #3 [waiting for connection]
14044 189 60.0 1053116184 19696288 ?      Sls  Oct17 14452:02 /bin/java -Xms16g -Xmx16g -Xmn2g -XX:+UseConcMarkSweepGC -XX:CMSInitiatingOccupancyFraction=75 -XX:+UseG1GC
14124 0.0  0.0  72124   1360 ?      S1   Oct17  0:00  \_ /usr/share/elasticsearch/plugins/x-pack/platform/linux-x86_64/bin/controller
32305 1.0  0.0  26308   9400 ?      S<  03:39  7:55 /usr/bin/atom -a -R -w /var/lon/atom/atom_20181022_60

```

- pidstat -u -l -r -d -w -h (-p <pid>) 2
 - Vue par process (filtre possible), par type de cpu, mémoire, IO, context switch (volontaire ou pas)
 - Pratique pour identifier qui est responsable de la consommation de CPU system
 - -t vue par thread
 - Context switch
 - $\frac{cswch/s \times 80\,000}{nb\ core * fréquence\ CPU} > 5\% \Rightarrow$ possible lock contention (plus de 5% des cycles d'horloges passés sur du ctx switch)
 - nvcswch/s : indice de surcharge
 - Ex plusieurs pid : \$(pidof postgres | sed -e 's/ /,/g')
 - %cpu exactement à 100% erroné parfois observé sur VM

#	Time	PID	%usr	%system	%guest	%CPU	CPU	minflt/s	majflt/s	VSZ	RSS	%MEM	kB_rd/s	kB_wr/s	kB_ccwr/s	cswch/s	nvcswch/s	Command
1539950536		9	0.00	0.00	0.00	0.00	1	0.00	0.00	0	0	0.00	0.00	0.00	0.00	2.00	0.00	ksoftirqd/1
1539950536		11	0.00	0.00	0.00	0.00	0	0.00	0.00	0	0	0.00	0.00	0.00	0.00	1.00	0.00	events/0
1539950536		12	0.00	0.00	0.00	0.00	1	0.00	0.00	0	0	0.00	0.00	0.00	0.00	14.00	0.00	events/1
1539950536		846	0.00	0.00	0.00	0.00	1	0.00	0.00	23592	672	0.02	0.00	0.00	0.00	0.50	0.00	rotatelogs
1539950536		933	0.00	0.00	0.00	0.00	0	0.00	0.00	0	0	0.00	0.00	0.00	0.00	1.00	0.00	vmmemctl
1539950536		2223	0.00	0.00	0.00	0.00	1	0.00	0.00	19040	644	0.02	0.00	0.00	0.00	0.50	0.00	rpcbind
1539950536		2778	0.00	0.00	0.00	0.00	1	14.00	0.00	1348368	26288	0.67	0.00	0.00	0.00	1.00	0.00	python
1539950536		3095	0.00	0.00	0.00	0.00	1	17.50	0.00	121460	1148	0.03	0.00	2.00	0.00	1.00	0.00	
1539950536		3320	0.00	0.00	0.00	0.00	1	0.00	0.00	4060	460	0.01	0.00	2.00	0.00	1.00	0.00	atopacctd
1539950536		5110	0.00	0.00	0.00	0.00	1	0.00	0.00	11856	296	0.01	0.00	0.00	0.00	0.50	0.00	epmd
1539950536		7078	0.00	0.00	0.00	0.00	1	0.00	0.00	235752	2952	0.08	0.00	0.00	0.00	10.00	0.00	vmtoolsd
1539950536		10416	0.00	0.00	0.00	0.00	1	0.00	0.00	0	0	0.00	0.00	0.00	0.00	0.50	0.00	flush-253:8
1539950536		12974	0.50	0.00	0.50	0.50	1	1.50	0.00	5077088	867396	22.10	0.00	0.00	0.00	0.00	0.00	java
1539950536		15576	0.00	0.00	0.00	0.00	0	0.00	0.00	71924	1312	0.03	0.00	0.00	0.00	1.00	0.00	httpd_worker
1539950536		18573	0.00	0.00	0.00	0.00	1	0.00	0.00	122740	2092	0.05	0.00	0.00	0.00	2.00	0.00	sshd
1539950536		21321	11.00	1.00	0.00	12.00	1	0.50	0.00	3111992	307736	7.84	0.00	0.00	0.00	0.00	0.00	java
1539950536		21610	0.00	1.00	0.00	1.00	1	390.00	0.00	101144	1048	0.03	0.00	0.00	0.00	0.50	5.00	pidstat
1539950536		22833	0.50	2.00	0.00	2.50	0	299.50	0.00	1312612	10008	0.26	0.00	0.00	0.00	0.50	0.00	collectd
1539950536		22916	157.50	13.00	0.00	170.50	0	306.50	0.00	4944712	655132	16.69	0.00	4.00	0.00	0.00	0.00	java
1539950536		27426	0.00	0.00	0.00	0.00	1	4.50	0.00	198844	16272	0.41	0.00	0.00	0.00	2.50	0.00	
1539950536		27747	0.00	0.00	0.00	0.00	1	0.00	0.00	35960	780	0.02	0.00	0.00	0.00	1.00	0.00	snmpmagt
1539950536		29705	0.00	0.00	0.00	0.00	1	0.00	0.00	22404	556	0.01	0.00	0.00	0.00	3.00	0.00	nukremotexec.xp

- iostat -xmzty 2

- Vue détaillée du profil d'IO, écriture et lecture en Mo/s
- Souci si await > 1000 ms
- Saturation « potentielle » si aq-sz/avgqu-sz > 1
- Souci « potentiel » si %util 100% (100% du temps à écrire)

```

10/19/2018 02:16:50 PM
avg-cpu:  %user   %nice  %system  %iowait  %steal  %idle
      52.74     0.00    5.73     9.31     0.00   32.22

Device:    rrqm/s   wrqm/s     r/s     w/s   rMB/s    wMB/s  avgrrq-sz  avgqu-sz    await  r_await  w_await  svctm  %util
sda        0.00     0.00    1.00    6.00    0.01     0.04    15.29     0.02    2.29    11.00    0.83    1.71    1.20
sdb        0.00     0.00  291.00   85.50   45.33     3.06   263.23    12.32   32.74   41.24    3.82    0.79   29.85
sdc        0.00   863.50   17.50  203.50    7.68     4.24   110.48     0.68    3.19    9.63    2.64    1.61   35.50
sde        0.00     0.00   73.00     0.00   30.32     0.00   850.74     1.74   23.86   23.86    0.00    2.08   15.15
dm-0       0.00     0.00    0.50    5.00     0.00     0.04   16.91     0.00    0.82    0.00    0.90    0.09    0.05
dm-2       0.00     0.00    0.00    0.50     0.00     0.00    5.00     0.00    0.00    0.00    0.00    0.00    0.00
dm-5       0.00     0.00    0.00    0.50     0.00     0.00    7.00     0.00    1.00    0.00    1.00    1.00    0.05
dm-4       0.00     0.00    0.50    1.00     0.00     0.00    8.67     0.00    1.00    1.00    1.00    1.00    0.15
dm-14      0.00     0.00  377.00  1151.50   81.34     7.23   118.67    18.29   11.99   37.08    3.77    0.28   42.35

10/19/2018 02:16:52 PM
avg-cpu:  %user   %nice  %system  %iowait  %steal  %idle
      43.72     0.00    3.39    12.19     0.00   40.70

Device:    rrqm/s   wrqm/s     r/s     w/s   rMB/s    wMB/s  avgrrq-sz  avgqu-sz    await  r_await  w_await  svctm  %util
sdb        0.00     0.50  476.12   136.82   45.70    18.95   216.03     3.43    5.60    5.19    7.03    0.52   32.04
sdc        0.00   945.77  133.83   295.52    2.17    14.18    78.00     2.20    5.13   10.23    2.81    0.80   34.18
sde        0.00     0.00   73.13     0.00   27.49     0.00   769.85     0.85   11.67   11.67    0.00    1.60   11.69
dm-9       0.00     0.00    0.00    0.50     0.00     0.00    3.00     0.00    0.00    0.00    0.00    0.00    0.00
dm-14      0.00     0.00  683.08  1377.11   75.36    33.12   107.84     9.63    4.65    6.89    3.54    0.22   45.17

```

- Sur VM, correspondance fs
 - ls -l /dev/disk/by-id

```

lrwxrwxrwx 1 root root  9 Feb 21  2018 ata-VMware_Virtual_IDE_CDROM_Drive_10000000000000000000000000000001 -> ../../sr0
lrwxrwxrwx 1 root root 10 Feb 21  2018 dm-name-data_vg-                                ../../dm-9
lrwxrwxrwx 1 root root 11 Feb 21  2018 dm-name-data_vg-                                ../../dm-10
lrwxrwxrwx 1 root root 10 May   8 04:02 dm-name-data_vg_lv_appli -> ../../dm-3
lrwxrwxrwx 1 root root 11 Sep   3 12:16 dm-name-data_vg_lv_es_data -> ../../dm-14

```

- sar -n DEV 2
 - Débit réseau entrée et sortie

04:00:04 PM	IFACE	rxpck/s	txpck/s	rxkB/s	txkB/s	rxcmp/s	txcmp/s	rxmcst/s
04:00:06 PM	eth0	82451.50	41220.00	122530.92	2818.03	0.00	0.00	2.00
04:00:06 PM	eth1	2.00	0.00	0.24	0.00	0.00	0.00	2.00
04:00:06 PM	eth2	0.00	0.00	0.00	0.00	0.00	0.00	0.00
04:00:06 PM	lo	0.00	0.00	0.00	0.00	0.00	0.00	0.00

- Aussi nicstat et bwm-ng (taper +)

Average:	IFACE	rxpck/s	txpck/s	rxkB/s	txkB/s	rxcmp/s	txcmp/s	rxmcst/s
Average:	eth0	13789.06	15631.60	5383.62	12313.08	0.00	0.00	0.00
Average:	lo	110.29	110.29	36.91	36.91	0.00	0.00	0.00

- ethtool eth0 **/sbin**
 - Vitesse de la « carte »

```
Settings for eth0:
  Supported ports: [ FIBRE ]
  Supported link modes:  10000baseT/Full
  Supported pause frame use: No
  Supports auto-negotiation: No
  Advertised link modes: 10000baseT/Full
  Advertised pause frame use: No
  Advertised auto-negotiation: No
  Speed: 1000Mb/s
  Duplex: Full
  Port: FIBRE
  PHYAD: 0
  Transceiver: internal
  Auto-negotiation: off
  Current message level: 0x00000000 (0)
```

- sar -n EDEV 2
 - Erreurs réseau en temps réel

```
04:32:52 PM    IFACE  rxerr/s  txerr/s  coll/s  rxdrop/s  txdrop/s  txcarr/s  rxffram/s  rxfifo/s  txfifo/s
04:32:54 PM    eth0    0.00     0.00     0.00     0.00     0.00     0.00     0.00     0.00     0.00
04:32:54 PM    eth1    0.00     0.00     0.00     0.00     0.00     0.00     0.00     0.00     0.00
04:32:54 PM    eth2    0.00     0.00     0.00     0.00     0.00     0.00     0.00     0.00     0.00
04:32:54 PM    lo     0.00     0.00     0.00     0.00     0.00     0.00     0.00     0.00     0.00
```

- netstat -s | egrep -i "drop|error|loss|low|over"
 - Erreurs réseau depuis le boot

```
2 dropped because of missing route
0 packet receive errors
0 receive buffer errors
0 send buffer errors
116 times recovered from packet loss by selective acknowledgements
3 congestion windows recovered without slow start by DSACK
13 congestion windows recovered without slow start after partial ack
6 timeouts after SACK recovery
9 timeouts in loss state
144 retransmits in slow start
TCPLossProbes: 873
TCPLossProbeRecovery: 28
```

- ip -s | /sbin
 - mtu

```
2: eth0: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc mq state UP mode DEFAULT qlen 1000
  Link/ether 00:0c:29:14:dc:1e brd ff:ff:ff:ff:ff:ff
  RX: bytes packets errors dropped overrun mcast
    6298711166 27536318 0      5354   0      9725936
  TX: bytes packets errors dropped carrier collsns
    5206558026 5360080 0      0      0      0
```

- sar -n TCP,ETCP 2

- Nouvelles connexion entrantes et sortantes
- Retransmit TCP

```
04:49:50 PM  active/s  passive/s    iseg/s    oseg/s
04:49:51 PM      0.00     114.12  11941.18  14649.41

04:49:50 PM  atmptf/s   estres/s retrans/s  isegerr/s   orsts/s
04:49:51 PM      0.00      0.00    30.59      0.00    240.00
```

- sar -n DEV,EDEV,TCP,ETCP 2

- Cumulable

```
Linux 3.10.0-327.36.3.el7.x86_64 (el7)        10/22/2018      _x86_64_      (8 CPU)

05:15:53 PM  IFACE  rxpck/s  txpck/s  rxkB/s  txkB/s  rxcmp/s  txcmp/s  rxmcst/s
05:15:55 PM    eth0    3315.50  2302.50  4400.99  5178.99    0.00    0.00    0.00
05:15:55 PM    eth1     30.00     0.00     2.70     0.00    0.00    0.00    0.00
05:15:55 PM    lo      23.50    23.50     8.37     8.37    0.00    0.00    0.00

05:15:53 PM  IFACE  rxerr/s  txerr/s  coll/s  rxdrop/s  txdrop/s  txcarr/s  rxffram/s  rxfifo/s  txfifo/s
05:15:55 PM    eth0     0.00     0.00     0.00     0.00     0.00     0.00     0.00     0.00     0.00
05:15:55 PM    eth1     0.00     0.00     0.00     0.00     0.00     0.00     0.00     0.00     0.00
05:15:55 PM    lo      0.00     0.00     0.00     0.00     0.00     0.00     0.00     0.00     0.00

05:15:53 PM  active/s  passive/s    iseg/s    oseg/s
05:15:55 PM      0.00     3.50    726.00  2300.50

05:15:53 PM  atmptf/s   estres/s retrans/s  isegerr/s   orsts/s
05:15:55 PM      0.00     1.00     0.00      0.00     3.00
```

- `ss -tunal[ropi]` `/usr/sbin`
 - Connexions
 - Etat des connexions : LISTEN, ESTABLISHED ... SYN_SENT, ...
 - Arguments
 - -r : résolution DNS
 - -o : info timer
 - -p : info process (mais lent et coûteux sur RHEL6 par rapport à netstat -antup) `#root`
 - -i : interne TCP
 - Filtre possible : `ss -utan state connected src :5432`

Netid	State	Recv-Q	Send-Q	Local Address:Port	Peer Address:Port
udp	UNCONN	0	0	127.0.0.53%lo:53	0.0.0.0:*
udp	UNCONN	0	0	0.0.0.0:68	0.0.0.0:*
udp	UNCONN	0	0	0.0.0.0:631	0.0.0.0:*
udp	UNCONN	0	0	0.0.0.0:50165	0.0.0.0:*
udp	UNCONN	0	0	0.0.0.0:5353	0.0.0.0:*
udp	UNCONN	0	0	[::]:44989	[::]:*
udp	UNCONN	0	0	[::]:5353	[::]:*
tcp	LISTEN	0	128	127.0.0.1:33897	0.0.0.0:*
tcp	LISTEN	0	5	0.0.0.0:8080	0.0.0.0:*
tcp	LISTEN	0	128	127.0.0.53%lo:53	0.0.0.0:*
tcp	LISTEN	0	5	127.0.0.1:631	0.0.0.0:*
tcp	FIN-WAIT-2	0	0	127.0.0.1:8080	127.0.0.1:53126
tcp	CLOSE-WAIT	0	0	127.0.0.1:53126	127.0.0.1:8080
tcp	LISTEN	0	5	[::1]:631	[::]:*

```

tcp  LISTEN    0      128          localhost.localdomain:81          *:*
users:(("httpd_prefork",1605,5),("httpd_prefork",4059,5),("httpd_prefork",6299,5),("httpd_prefork",9499,5),("httpd_prefork",9529,5),("httpd_prefork",9530,5),("httpd_prefork",16537,5),("httpd_prefork",16562,5),
("httpd_prefork",16563,5),("httpd_prefork",16564,5),("httpd_prefork",16565,5),("httpd_prefork",17140,5))
          rto:1000 cwnd:10
tcp  LISTEN    0      50           localhost.localdomain:10514        *:*
users:(("java",950,16))
          rto:1000 cwnd:10
tcp  LISTEN    0      128          :35858                         *:*
users:(("java",20983,207))
          rto:1000 cwnd:10

```

- **SS -S /usr/sbin**
 - Nombre de connexions

```
Total: 444 (kernel 465)
TCP: 195 (estab 179, closed 0, orphaned 0, synrecv 0, timewait 0/0), ports 25

Transport Total      IP        IPv6
*       465          -         -
RAW      0            0         0
UDP      8            8         0
TCP     195          195        0
INET    203          203        0
FRAG    0            0         0
```

- Tester ouverture d'un port
 - nc -vz <host> <port>
- ```
[root@_ ~]# nc -vz _ 80
Connection to _ port [tcp/http] succeeded!
[root@_ ~]# nc -vz _ 83
nc: connect to _ port 83 (tcp) failed: Connection refused
```
- traceroute -4 -T -O info -p <port> -q 1 <host> #root
    - Alternative si -z impossible
    - <syn,ack> : OK
    - <rst,ack> : KO
  - En bash  
(echo >/dev/tcp/<host>/<port>) &>/dev/null ; if [ \$? -eq 0 ]; then echo OK; else echo KO; fi

- mtr -rw -c 3 <host> #root

- Ping évolué

```
[root@z ~]# mtr -rw -c 3 z
HOST: z
 1. z
 2. z
 Loss% Snt Last Avg Best Wrst StDev
 0.0% 3 2.2 1.4 0.9 2.2 0.7
 0.0% 3 0.3 0.4 0.3 0.8 0.3
```

- df -hTP

```

Filesystem Size Used Avail Use% Mounted on
/dev/mapper/root_vg-lv_root 4.0G 3.5G 580M 86% /
devtmpfs 16G 0 16G 0% /dev
tmpfs 16G 0 16G 0% /dev/shm
tmpfs 16G 1.2G 15G 8% /run
tmpfs 16G 0 16G 0% /sys/fs/cgroup
/dev/mapper/root_vg-lv_tmp 1014M 36M 979M 4% /tmp
/dev/mapper/data_vg-lv_appli 5.0G 47M 5.0G 1% /appli
/dev/mapper/root_vg-lv_var 2.0G 928M 1.1G 46% /var
/dev/mapper/data_vg-lv_opt 2.8G 106M 2.5G 4% /opt
/dev/mapper/data_vg-lv_es_log 4.6G 240M 4.1G 6% /var/log/elasticsearch
/dev/mapper/root_vg-lv_var_log_audit 2.0G 62M 2.0G 4% /var/log/audit
/dev/sda1 509M 130M 380M 26% /boot
/dev/mapper/data_vg-lv_es_data 1.5T 992G 388G 72% /var/lib/elasticsearch/data
/dev/mapper/data_vg-lv_logiciels 10G 266M 9.8G 3% /logiciels
/dev/mapper/data_vg- 014M 33M 982M 4% /var/projects
/dev/mapper/data_vg- 149M 112M 37M 76%
/dev/mapper/data_vg-lv_varopsware 149M 8.6M 141M 6%
/dev/mapper/root_vg-lv_home 1014M 107M 908M 11% /home
/dev/mapper/data_vg- 725M 425M 301M 59%
/dev/mapper/data_vg- 517M 27M 491M 6%
tmpfs 3.2G 0 3.2G 0% /run/user/
tmpfs 3.2G 0 3.2G 0% /run/user/
tmpfs 3.2G 0 3.2G 0% /run/user/
/dev/mapper/data_vg-lv_kibana_log 976M 42M 868M 5% /var/log/kibana
/dev/mapper/data_vg-lv_kibana_data 976M 68M 842M 8% /var/lib/kibana
tmpfs 3.2G 0 3.2G 0% /run/user/

```

- df -iTP pour les inodes

- vgs #root

- Volume group LVM, pour espace disponible

| VG      | #PV | #LV | #SN | Attr   | VSize  | VFree |
|---------|-----|-----|-----|--------|--------|-------|
| data_vg | 3   | 16  | 0   | wz--n- | 1.46t  | 9.97g |
| root_vg | 1   | 6   | 0   | wz--n- | 13.47q | 1.47q |

- du -sh
  - Taille d'un répertoire et de la sous arborescence

```
[root@... log]# du -sh *
18M deploy
3.5M exploit
3.9M history
16K lost+found
4.0G
45M sup
```

- locate <un\_fichier>
  - Localise un fichier/pattern
  - S'appuie sur une DB locale mlocate.db
  - Mise à jour via updatedb souvent croné
  - Attention à l'impact au passage de updatedb
  - Possible de pruner des arborescences dans /etc/updatedb.conf

- lsof -P [-p <pid>] /usr/sbin #root

- Fichiers ouverts [par process] : fichiers, sockets, ...
- Attention au fichier supprimé mais non libéré : file descriptor “deleted”
- lsof sans argument : aussi les threads enfants, pas le cas avec -p <pid>
- [-i :<port>] : qui écoute sur tel port

| COMMAND | PID   | USER | FD   | TYPE               | DEVICE  | SIZE/OFF   | NODE                        | NAME                             |
|---------|-------|------|------|--------------------|---------|------------|-----------------------------|----------------------------------|
| java    | 18615 |      | cwd  | DIR                | 253,1   | 4096       | 2                           | /                                |
| java    | 18615 |      | rtd  | DIR                | 253,1   | 4096       | 2                           | /                                |
| java    | 18615 |      | txt  | REG                | 253,1   | 7734       | 412943                      | /usr/java/                       |
| java    | 18615 |      | mem  | REG                | 253,1   | 157072     | 9315                        | /lib64/ld-2.12.so                |
| java    | 18615 |      | mem  | REG                | 253,1   | 22536      | 9783                        | /lib64/libdl-2.12.so             |
| java    | 18615 |      | mem  | REG                | 253,1   | 1926480    | 9530                        | /lib64/libc-2.12.so              |
| java    | 18615 |      | mem  | REG                | 253,1   | 145896     | 9781                        | /lib64/libpthread-2.12.so        |
| java    | 18615 |      | mem  | REG                | 253,1   | 599392     | 5107                        | /lib64/libm-2.12.so              |
| java    | 18615 |      | mem  | REG                | 253,1   | 47112      | 9804                        | /lib64/librt-2.12.so             |
| java    | 18615 |      | mem  | REG                | 253,1   | 93320      | 27609                       | /lib64/                          |
| java    | 18615 |      | mem  | REG                | 253,1   | 113952     | 14983                       | /lib64/libresolv-2.12.so         |
| java    | 18615 |      | mem  | REG                | 253,1   | 7652       | 413086                      | /usr/java/                       |
| java    | 18615 |      | mem  | REG                | 253,1   | 49341      | 413088                      | /usr/java/                       |
| java    | 18615 |      | mem  | REG                | 253,1   | 15981      | 413057                      | /usr/java/                       |
| java    | 18615 |      | mem  | REG                | 253,1   | 257982     | 413035                      | /usr/java/                       |
| java    | 18615 |      | mem  | REG                | 253,1   | 27424      | 3907                        | /lib64/libnss_dns-2.12.so        |
| java    | 18615 |      | mem  | REG                | 253,1   | 93112      | 413052                      | /usr/java/                       |
| java    | 18615 |      | mem  | REG                | 253,1   | 115814     | 413027                      | /usr/java/                       |
| java    | 18615 |      | mem  | REG                | 253,1   | 50289      | 413063                      | /usr/java/                       |
| java    | 18615 |      | mem  | REG                | 253,1   | 66216625   | 413019                      | /usr/java/                       |
| java    | 18615 |      | mem  | REG                | 253,1   | 124327     | 413050                      | /usr/java/                       |
| java    | 18615 |      | mem  | REG                | 253,1   | 65928      | 3909                        | /lib64/libnss_files-2.12.so      |
| java    | 18615 |      | mem  | REG                | 253,1   | 648816     | 412971                      | /usr/java/                       |
| java    | 18615 |      | mem  | REG                | 253,1   | 225914     | 413077                      | /usr/java/                       |
| java    | 18615 |      | mem  | REG                | 253,1   | 66472      | 413028                      | /usr/java/                       |
| java    | 18615 |      | mem  | REG                | 253,1   | 17013932   | 413081                      | /usr/java/                       |
| java    | 18615 |      | mem  | REG                | 253,1   | 102990     | 412209                      | /usr/java/                       |
| java    | 18615 |      | mem  | REG                | 253,1   | 116446     | 413281                      | /usr/java/                       |
| java    | 18615 |      | mem  | REG                | 253,20  | 32768      | 1330                        | /tmp/hsperrdata_ 18615           |
| java    | 18615 |      | Or   | CHR                | 1,3     | 0t0        | 3905                        | /dev/null                        |
| java    | 18615 |      | Iw   | REG                | 253,5   | 76         | 538                         | /log/-----out                    |
| java    | 18615 | 167r | CHR  | 1,9                | 0t0     | 3910       | /dev/urandom                |                                  |
| java    | 18615 | 167r | CHR  | 1,9                | 0t0     | 3910       | /dev/urandom                |                                  |
| java    | 18615 | 168u | unix | 0xffff880428ce14c0 | 0t0     | 2925582887 | socket                      |                                  |
| java    | 18615 | 169u | IPv4 | 2925582901         | 0t0     | TCP        | :52756->..... (ESTABLISHED) |                                  |
| java    | 18615 | 170r | CHR  | 1,9                | 0t0     | 3910       | /dev/urandom                |                                  |
| java    | 18615 | 171r | REG  | 253,1              | 3505340 | 413251     | /usr/java/                  | /jre/lib/resources.jar           |
| java    | 18615 | 172r | REG  | 253,1              | 42185   | 413104     | /usr/java/                  | /jre/lib/ext/sunec.jar           |
| java    | 18615 | 173r | REG  | 253,1              | 280161  | 413105     | /usr/java/                  | /jre/lib/ext/sunjce_provider.jar |
| java    | 18615 | 174u | IPv4 | 2925583519         | 0t0     | TCP        | :60201->..... (ESTABLISHED) |                                  |
| java    | 18615 | 175u | IPv4 | 2925618721         | 0t0     | TCP        | *:8443 (LISTEN)             |                                  |
| java    | 18615 | 176u | unix | 0xffff8803ff8cec00 | 0t0     | 2925583664 | socket                      |                                  |

- sar -v

- inode, file ...

| 02:44:34 PM | dentunusd | file-nr | inode-nr | pty-nr |
|-------------|-----------|---------|----------|--------|
| 02:44:36 PM | 263453    | 8704    | 138299   | 6      |
| 02:44:38 PM | 263453    | 8704    | 138299   | 6      |

- `atop -r <un_fichier_historise_par_jour>`
  - Vision détaillée par process
  - A faire tourner en daemon et avec historisation (légère empreinte CPU au moment de la collecte)
  - Indispensable pour le post mortem
  - Ne pas oublier le module kernel pour avoir le réseau par process
  - Attention, atop est responsable du flush et rotation d'un fichier alloué par un module kernel.  
Si atop est tué, ce fichier va grossir dans /tmp ...

| 2018/12/05 02:20:02 |     |              |               |               |              |              |              |            |           |         |            |        |         |        |           |         |         |           |         |       |         |
|---------------------|-----|--------------|---------------|---------------|--------------|--------------|--------------|------------|-----------|---------|------------|--------|---------|--------|-----------|---------|---------|-----------|---------|-------|---------|
| PRC                 |     |              | sys           |               |              | user         |              |            | #proc     |         |            | #trun  |         |        | #tslpi    |         |         | #tslpu    |         |       |         |
|                     | CPU | sys          | 46.79%        | user          | 50%          | irq          | 19%          | idle       | 411       | idle    | 8          | idle   | 210%    | wait   | 0%        | idle    | 0       | steal     | 0%      |       |         |
|                     |     | cpu          | 8%            | user          | 67%          | irq          | 19%          | idle       |           | idle    |            | idle   | 7%      | cpu007 | w         | 0%      |         | guest     | 0%      |       |         |
|                     |     | cpu          | 8%            | user          | 65%          | irq          | 0%           | idle       |           | idle    |            | idle   | 27%     | cpu000 | w         | 0%      |         | curf      | 2.30GHz |       |         |
|                     |     | cpu          | 8%            | user          | 65%          | irq          | 0%           | idle       |           | idle    |            | idle   | 27%     | cpu003 | w         | 0%      |         | curf      | 2.30GHz |       |         |
|                     |     | cpu          | 8%            | user          | 65%          | irq          | 0%           | idle       |           | idle    |            | idle   | 27%     | cpu006 | w         | 0%      |         | curf      | 2.30GHz |       |         |
|                     |     | cpu          | 8%            | user          | 65%          | irq          | 0%           | idle       |           | idle    |            | idle   | 27%     | cpu001 | w         | 0%      |         | curf      | 2.30GHz |       |         |
|                     |     | cpu          | 8%            | user          | 64%          | irq          | 0%           | idle       |           | idle    |            | idle   | 28%     | cpu004 | w         | 0%      |         | curf      | 2.30GHz |       |         |
|                     |     | cpu          | 8%            | user          | 62%          | irq          | 0%           | idle       |           | idle    |            | idle   | 30%     | cpu005 | w         | 0%      |         | curf      | 2.30GHz |       |         |
|                     |     | cpu          | 7%            | user          | 56%          | irq          | 0%           | idle       |           | idle    |            | idle   | 37%     | cpu002 | w         | 0%      |         | curf      | 2.30GHz |       |         |
|                     | CPL | avg1         | 5.19          |               | avg5         | 5.90         | avg15        | 3.65       |           |         |            | csw    | 2577031 | intr   | 1470369   |         | numcpu  | 8         | #exit   | 1380  |         |
|                     | MEM | tot          | 11.6G         | free          | 171.8M       | cache        | 1.3G         | dirty      | 0.6M      | buff    | 247.9M     | slab   | 349.6M  | slrec  | 274.6M    | shmem   | 0.3M    | shrss     | 0.0M    | shswp | 0.0M    |
|                     | SWP | tot          | 3.0G          | free          | 3.0G         |              |              |            |           |         |            |        |         | shrss  | 0.0M      | vmbal   | 0.0M    | hptot     | 0.0M    | hpuse | 0.0M    |
|                     | LVM | ta_vg_lv_log | busy          | 0%            | read         | 0            | write        | 321        | Kib/r     | 0       | Kib/w      | 4      | MBr/s   | 0.00   |           |         |         | vmcom     | 10.5G   |       |         |
|                     | LVM | ot_vg_lv_var | busy          | 0%            | read         | 0            | write        | 201        | Kib/r     | 0       | Kib/w      | 4      | MBr/s   | 0.00   |           |         |         | avq       | 6.10    | avio  | 0.57 ms |
|                     | LVM | I_           | busy          | 0%            | read         | 0            | write        | 52         | Kib/r     | 0       | Kib/w      | 2      | MBr/s   | 0.00   |           |         |         | MBr/s     | 0.01    | avq   | 4.56    |
|                     | LVM | ot_vg_lv_tmp | busy          | 0%            | read         | 0            | write        | 243        | Kib/r     | 0       | Kib/w      | 4      | MBr/s   | 0.00   |           |         |         | MBr/s     | 0.00    | avq   | 1.18    |
|                     | LVM | t_vg_lv_root | busy          | 0%            | read         | 0            | write        | 100        | Kib/r     | 0       | Kib/w      | 4      | MBr/s   | 0.00   |           |         |         | MBr/s     | 0.02    | avq   | 5.66    |
|                     | LVM | _logstash_db | busy          | 0%            | read         | 3            | write        | 124        | Kib/r     | 1       | Kib/w      | 1      | MBr/s   | 0.00   |           |         |         | MBr/s     | 0.01    | avq   | 2.89    |
|                     | LVM | #_           | busy          | 0%            | read         | 0            | write        | 73         | Kib/r     | 0       | Kib/w      | 4      | MBr/s   | 0.00   |           |         |         | MBr/s     | 0.00    | avq   | 7.08    |
|                     | LVM | _vg_lv_appli | busy          | 0%            | read         | 0            | write        | 16         | Kib/r     | 0       | Kib/w      | 4      | MBr/s   | 0.00   |           |         |         | MBr/s     | 0.00    | avq   | 3.76    |
|                     | LVM | Tv_logiciels | busy          | 0%            | read         | 0            | write        | 8          | Kib/r     | 0       | Kib/w      | 4      | MBr/s   | 0.00   |           |         |         | MBr/s     | 0.00    | avq   | 1.80    |
|                     | LVM | _varprojects | busy          | 0%            | read         | 0            | write        | 37         | Kib/r     | 0       | Kib/w      | 4      | MBr/s   | 0.00   |           |         |         | MBr/s     | 0.00    | avq   | 1.50    |
|                     | DSK | sda          | busy          | 1%            | read         | 0            | write        | 198        | Kib/r     | 0       | Kib/w      | 10     | MBr/s   | 0.00   |           |         |         | MBr/s     | 0.04    | avq   | 12.22   |
|                     | DSK | sdb          | busy          | 1%            | read         | 3            | write        | 145        | Kib/r     | 1       | Kib/w      | 11     | MBr/s   | 0.00   |           |         |         | MBr/s     | 0.03    | avq   | 1.78    |
|                     | DSK | sdc          | busy          | 0%            | read         | 0            | write        | 80         | Kib/r     | 0       | Kib/w      | 5      | MBr/s   | 0.00   |           |         |         | MBr/s     | 0.01    | avq   | 1.88    |
|                     | NET | transport    | tcpip 1023306 | tcppo 1152459 | udpi 239     | udpo 619     | tcpao 1854   | tcppo 1624 | tcpers 26 | tcpie 0 | tcpor 1095 |        |         |        | tcpip     | 1023306 | udppn 0 | udpie 0   |         |       |         |
|                     | NET | network      | ipi 1024599   | ipfrw 0       | deliv 1025e3 |              |              |            |           |         |            |        |         |        | icmpe 514 |         |         | icmpe 512 |         |       |         |
|                     | NET | eth0         | 6%            | pcki 1061069  | pcko 1181823 | si 40 Mbps   | so 68 Mbps   | coll 0     | mlti 0    | erri 0  | erro 0     | drpi 0 |         |        |           |         |         |           | drpo 0  |       |         |
|                     | NET | eth1         | 0%            | pcki 548      | pcko 284     | si 11 Kbps   | so 8 Kbps    | coll 0     | mlti 0    | erri 0  | erro 0     | drpi 0 |         |        |           |         |         |           | drpo 0  |       |         |
|                     | NET | To           | ---           | pcki 19834    | pcko 19834   | si 2372 Kbps | so 2372 Kbps | coll 0     | mlti 0    | erri 0  | erro 0     | drpi 0 |         |        |           |         |         |           | drpo 0  |       |         |

Consumption per program; use 'a' to toggle between all/active processes

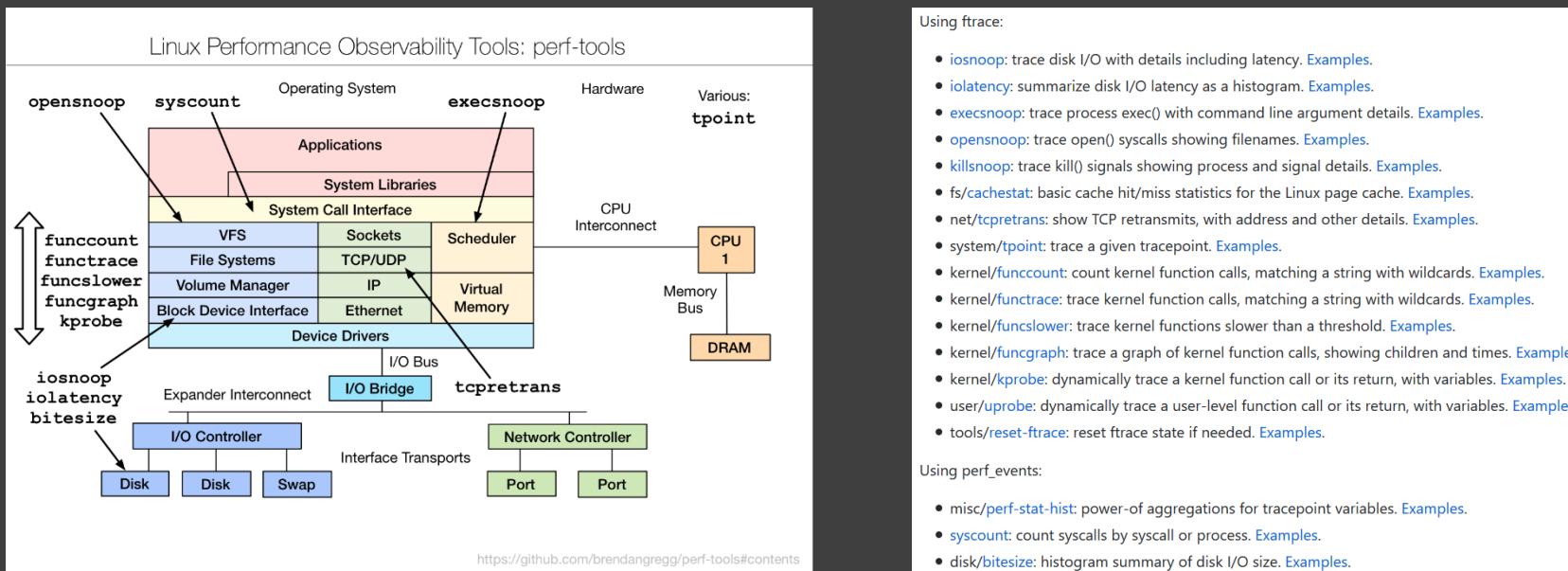
| NPROCS | SYSCPU | USRCPU | VSIZE  | RSIZE  | PSIZE  | SWAPSZ | RDDSK | WRDSK  | RNET  | SNET  | NET | CMD            | 1/9      |  |
|--------|--------|--------|--------|--------|--------|--------|-------|--------|-------|-------|-----|----------------|----------|--|
| 9      | 44.71% | 5m04s  | 32.76  | 9.4G   | 9.3G   | 22824K | 24K   | 5999K  | 100e4 | 113e4 | 95% | java           |          |  |
| 58     | 0.41s  | 0.80s  | 6.2G   | 371.0M | 37044K | OK     | OK    | 20K    | 16444 | 17934 | 5%  |                |          |  |
| 1      | 0.30s  | 0.21s  | 982.1M | 7228K  | 5795K  | 2096K  | OK    | OK     | 143   | 547   | 0%  | collectd       |          |  |
| 1      | 0.08s  | 0.01s  | 48512K | 2752K  | 2466K  | OK     | OK    | OK     | 1962  | 3270  | 0%  |                |          |  |
| 44     | 0.02s  | 1.42s  | OK     | OK     | OK     | OK     | OK    | OK     | 361   | 483   | 0%  | curl           |          |  |
| 46     | 0.00s  | 0.00s  | 39920K | 6768K  | 5226K  | OK     | OK    | OK     | 84    | 84    | 0%  | mtr            |          |  |
| 1      | 0.00s  | 0.01s  | 322.2M | 5804K  | 5024K  | 164K   | OK    | OK     | 32K   | 0     | 132 | 0%             | rsyslogd |  |
| 4      | 0.06s  | 0.22s  | 2.3G   | 68944K | 62743K | 7028K  | OK    | 96K    | 24    | 28    | 0%  | python         |          |  |
| 1      | 0.00s  | 0.00s  | OK     | OK     | OK     | OK     | OK    | OK     | 9     | 17    | 0%  | rpciod/0       |          |  |
| 1      | 0.00s  | 0.01s  | 281.3M | 16960K | 15342K | 464K   | OK    | OK     | 5     | 7     | 0%  | sssd_be        |          |  |
| 1      | 0.00s  | 0.00s  | OK     | OK     | OK     | OK     | OK    | OK     | 3     | 6     | 0%  | rpciod/1       |          |  |
| 1      | 0.00s  | 0.00s  | OK     | OK     | OK     | OK     | OK    | OK     | 3     | 6     | 0%  | rpciod/6       |          |  |
| 1      | 0.00s  | 0.00s  | OK     | OK     | OK     | OK     | OK    | OK     | 2     | 4     | 0%  | rpciod/3       |          |  |
| 1      | 0.03s  | 0.03s  | 197.8M | 18480K | 17292K | 7340K  | OK    | 20K    | 8     | 10    | 0%  |                |          |  |
| 1      | 0.00s  | 0.00s  | OK     | OK     | OK     | OK     | OK    | OK     | 1     | 2     | 0%  | rpciod/2       |          |  |
| 1      | 0.00s  | 0.00s  | OK     | OK     | OK     | OK     | OK    | OK     | 1     | 2     | 0%  | rpciod/4       |          |  |
| 1      | 0.48s  | 0.15s  | 33932K | 16972K | 15535K | OK     | OK    | 80K    | 0     | 0     | 0%  | atop           |          |  |
| 29     | 0.29s  | 0.01s  | OK     | OK     | OK     | OK     | OK    | OK     | 0     | 0     | 0%  | pidof          |          |  |
| 558    | 0.07s  | 0.04s  | 415.5M | 4904K  | 1362K  | OK     | OK    | 11116K | 0     | 0     | 0%  | check_network. |          |  |
| 1      | 0.08s  | 0.03s  | 4060K  | 480K   | 76K    | 28K    | OK    | 100K   | 0     | 0     | 0%  | atopacctd      |          |  |
| 27     | 0.03s  | 0.01s  | 622.1M | 23740K | 5586K  | OK     | OK    | 6432K  | 0     | 0     | 0%  | rotatelogs     |          |  |
| 1      | 0.01s  | 0.03s  | 49888K | 1924K  | 911K   | 240K   | OK    | OK     | 0     | 0     | 0%  | vmtoolsd       |          |  |
| 2      | 0.03s  | 0.00s  | OK     | OK     | OK     | OK     | OK    | OK     | 0     | 0     | 0%  | events/5       |          |  |
| 2      | 0.03s  | 0.00s  | OK     | OK     | OK     | OK     | OK    | OK     | 0     | 0     | 0%  | events/7       |          |  |
| 2      | 0.02s  | 0.00s  | OK     | OK     | OK     | OK     | OK    | OK     | 0     | 0     | 0%  | events/3       |          |  |
| 1      | 0.01s  | 0.00s  | 232.0M | 3012K  | 1652K  | 268K   | OK    | OK     | 0     | 0     | 0%  | sssd_nss       |          |  |
| 32     | 0.01s  | 0.00s  | 119.8M | 2628K  | 705K   | OK     | OK    | 128K   | 0     | 0     | 0%  | keepalive      |          |  |
| 1      | 0.01s  | 0.00s  | 217.4M | 1348K  | 336K   | 348K   | OK    | OK     | 0     | 0     | 0%  | sssd           |          |  |
| 1      | 0.00s  | 0.01s  | 18356K | 608K   | 196K   | 160K   | OK    | OK     | 0     | 0     | 0%  | irqbalance     |          |  |
| 1      | 0.01s  | 0.00s  | 22404K | 532K   | 82K    | 168K   | OK    | OK     | 0     | 0     | 0%  | nukremotexec.x |          |  |
| 2      | 0.01s  | 0.00s  | OK     | OK     | OK     | OK     | OK    | OK     | 0     | 0     | 0%  | events/0       |          |  |
| 1      | 0.01s  | 0.00s  | OK     | OK     | OK     | OK     | OK    | 40K    | 0     | 0     | 0%  | flush-253:25   |          |  |
| 1      | 0.01s  | 0.00s  | OK     | OK     | OK     | OK     | OK    | 44K    | 0     | 0     | 0%  | jbd2/dm-1-8    |          |  |

# D'autres outils plus perfectionnés

La déclinaison de l'état des lieux avec des outils modernes

# perf-tools

- Utilisation des frameworks de tracing de Linux (ftrace et perf) et de scripts
- <https://github.com/brendangregg/perf-tools>
- Designé pour Linux 3.2+, peut fonctionner sur 2.6.x ... ou pas



```

root@...:~/perf-tools/bin# ./iolatency
Tracing block I/O. Output every 1 seconds. Ctrl-C to end.

>=(ms) .. <(ms) : I/O |Distribution
 0 -> 1 : 9 |#####
 1 -> 2 : 63 |#####
 ... : ... |#####

>=(ms) .. <(ms) : I/O |Distribution
 0 -> 1 : 0 |
 1 -> 2 : 218 |#####
 2 -> 4 : 209 |#####
 4 -> 8 : 1 |#
 ... : ... |#####

>=(ms) .. <(ms) : I/O |Distribution
 0 -> 1 : 0 |
 1 -> 2 : 95 |#####
 2 -> 4 : 253 |#####
 4 -> 8 : 7 |##
 ... : ... |#####

>=(ms) .. <(ms) : I/O |Distribution
 0 -> 1 : 0 |
 1 -> 2 : 84 |#####
 2 -> 4 : 268 |#####
 4 -> 8 : 10 |##
 ... : ... |#####

>=(ms) .. <(ms) : I/O |Distribution
 0 -> 1 : 7 |##
 1 -> 2 : 133 |#####
 2 -> 4 : 209 |#####
 4 -> 8 : 7 |##
 ... : ... |#####

```

```

[root@... disk]# ./bitesize
Tracing block I/O size (bytes), until Ctrl-C...
^C
 Kbytes : I/O |Distribution
 -> 0.9 : 2 |#
 1.0 -> 7.9 : 3364 |#####
 8.0 -> 63.9 : 2321 |#####
 64.0 -> 127.9 : 41 |#
 128.0 -> : 83 |#

```

```

./cachestat -t
Counting cache functions... Output every 1 seconds.

TIME HITS MISSES DIRTIES RATIO BUFFERS_MB CACHE_MB
08:28:57 415 0 0 100.0% 1 191
08:28:58 411 0 0 100.0% 1 191
08:28:59 362 97 0 78.9% 0 8
08:29:00 411 0 0 100.0% 0 9
08:29:01 775 20489 0 3.6% 0 89
08:29:02 411 0 0 100.0% 0 89
08:29:03 6069 0 0 100.0% 0 89
08:29:04 15249 0 0 100.0% 0 89
08:29:05 411 0 0 100.0% 0 89
08:29:06 411 0 0 100.0% 0 89
08:29:07 411 0 3 100.0% 0 89

```

```

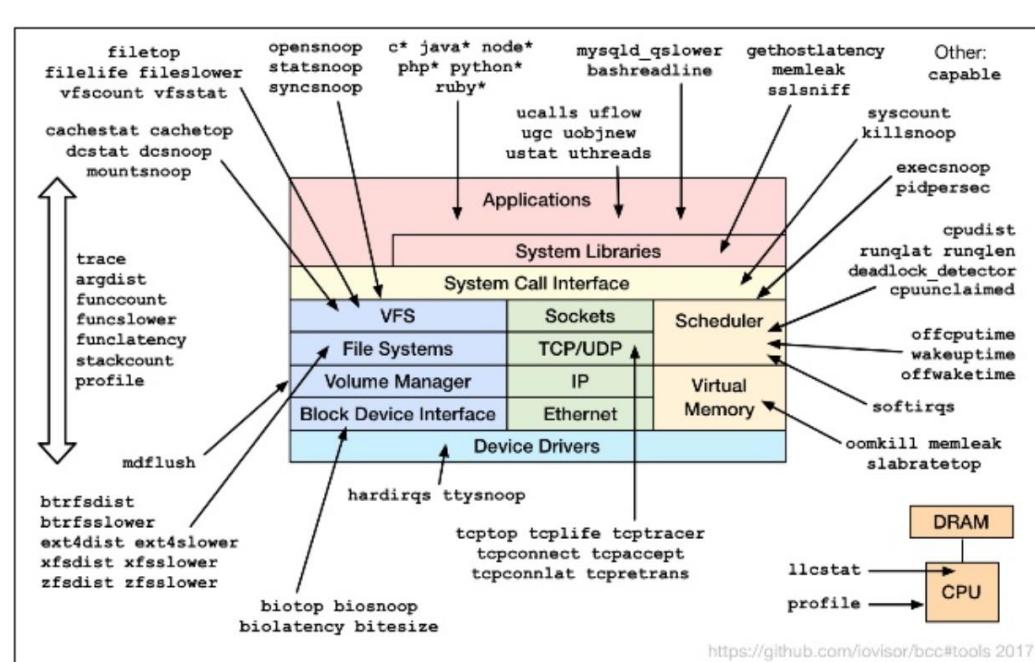
[root@... perf-tools-master]# ./execsnoop
Tracing exec()s. Ctrl-C to end.
Instrumenting sys_execve
 PID PPID ARGS
23807 23805 cat -v trace_pipe
23806 23794 gawk -v o=0 -v opt_name=0 -v name= -v opt_duration=0 [...]
23809 30954 date +%s%
23810 30954 curl --noproxy * -k -u 7...:... -H 'Content-Type: application/x-www-form-urlencoded' -X POST https://...:8086/write?db=jmx [...]
23812 23811 /sbin/modprobe -q -- net-pf-10
23814 30954 date +%s%

```

# Idem mais avec bcc/eBPF

- Sur des kernels récents

1. execsnoop
2. opensnoop
3. ext4slower (...)
4. biolatency
5. biosnoop
6. cachestat
7. tcpconnect
8. tcpaccept
9. tcpretrans
10. gethostlatency
11. runqlat
12. profile



## A Linux Tracing Timeline

- 1990's: Static tracers, prototype dynamic tracers
- 2000: LTT + DProbes (dynamic tracing; not integrated)
- 2004: kprobes (2.6.9)
- 2005: DTrace (not Linux), SystemTap (out-of-tree)
- 2008: ftrace (2.6.27)
- 2009: perf\_events (2.6.31)
- 2009: tracepoints (2.6.32)
- 2010-2016: ftrace & perf\_events enhancements
- 2012: uprobes (3.5)
- 2014-2017: enhanced BPF patches: supporting tracing events
- 2016-2017: ftrace hist triggers

# Etat des lieux rapide : avec bcc/eBPF...

## Discover short-lived process issues using execsnoop

```
execsnoop -t
TIME(s) PCOMM PID PPID RET ARGS
0.031 dirname 23832 23808 0 /usr/bin/dirname /apps/tomcat/bin/catalina.sh
0.888 run 23833 2344 0 ./run
0.889 run 23833 2344 -2 /command/bash
0.889 run 23833 2344 -2 /usr/local/bin/bash
0.889 run 23833 2344 -2 /usr/local/sbin/bash
0.889 bash 23833 2344 0 /bin/bash
0.894 svstat 23835 23834 0 /command/svstat /service/nflx-httd
0.894 perl 23836 23834 0 /usr/bin/perl -e $1=>;$1=-{d+} sec;/print $1||0;
0.899 ps 23838 23837 0 /bin/ps --ppid 1 -o pid,cmd,args
0.900 grep 23839 23837 0 /bin/grep org.apache.catalina
0.900 sed 23840 23837 0 /bin/sed s/^ *//;
0.900 cut 23841 23837 0 /usr/bin/cut -d -f 1
0.901 xargs 23842 23837 0 /usr/bin/xargs
0.912 xargs 23843 23842 -2 /command/echo
0.912 xargs 23843 23842 -2 /usr/local/bin/echo
0.912 xargs 23843 23842 -2 /usr/local/sbin/echo
0.912 echo 23843 23842 0 /bin/echo
[...]
```

## Exonerate or confirm storage latency issues and outliers with ext4slower

```
/usr/share/bcc/tools/ext4slower 1
Tracing ext4 operations slower than 1 ms
TIME COMM PID T BYTES OFF_KB LAT(ms) FILENAME
17:31:42 postdrop 15523 S 0 0 2.32 5630D406E4
17:31:42 cleanup 15524 S 0 0 1.89 57BB7406EC
17:32:09 titus-log-ship 19735 S 0 0 1.94 slurper_checkpoint.db
17:35:37 dhclient 1061 S 0 0 3.32 dhclient.eth0.leases
17:35:39 systemd-journa 504 S 0 0 26.62 system.journal
17:35:39 systemd-journa 504 S 0 0 1.56 system.journal
17:35:39 systemd-journa 504 S 0 0 1.73 system.journal
17:35:45 postdrop 16187 S 0 0 2.41 C0369406B4
17:35:45 cleanup 16188 S 0 0 6.52 C1B90406EC
[...]
```

Tracing at the file system is a more reliable and complete indicator than measuring disk I/O latency  
Also: btrfslower, xfslower, zfslower

## Identify multimodal disk I/O latency and outliers with biolatency

```
biolatency -mT 10
Tracing block device I/O... Hit Ctrl-C to end.

19:19:04
 msecs : count distribution
 0 -> 1 : 238 |*****
 2 -> 3 : 424 |*****
 4 -> 7 : 834 |*****
 8 -> 15 : 506 |*****
 16 -> 31 : 986 |*****
 32 -> 63 : 97 |**
 64 -> 127 : 7 |
 128 -> 255 : 27 |*

19:19:14
 msecs : count distribution
 0 -> 1 : 427 |*****
 2 -> 3 : 424 |*****
```

The "count" column is summarized in-kernel

## # ./fileslower

```
Tracing sync read/writes slower than 10 ms
TIME(s) COMM PID D BYTES LAT(ms) FILENAME
0.000 randread.pl 4762 R 8192 12.70 data1
8.850 randread.pl 4762 R 8192 11.26 data1
12.852 randread.pl 4762 R 8192 10.43 data1
```

This showed a few reads from a "randread.pl" program, each 8 Kbytes in size, and from a "data1" file. These all had over 10 ms latency.

This "latency" is measured from when the read or write was issued at the VFS interface, to when it completed. This spans everything: block device I/O (disk I/O), file system CPU cycles, file system locks, run queue latency, etc. This is a better measure of the latency suffered by applications reading from the file system than measuring this down at the block device interface.

Average latency (iostat/sar) may not be representative with multiple modes or outliers

# Etat des lieux rapide : avec bcc/eBPF...

Efficiently trace TCP sessions with PID, bytes, and duration using `tcplife`

```
/usr/share/bcc/tools/tcplife
PID COMM LADDR LPORT RADDR RPORT TX_KB RX_KB MS
2509 java 100.82.34.63 8078 100.82.130.159 12410 0 0 5.44
2509 java 100.82.34.63 8078 100.82.78.215 55564 0 0 135.32
2509 java 100.82.34.63 60778 100.82.207.252 7001 0 13 15126.87
2509 java 100.82.34.63 38884 100.82.208.178 7001 0 0 15568.25
2509 java 127.0.0.1 4243 127.0.0.1 42166 0 0 0.61
2509 java 127.0.0.1 42166 127.0.0.1 4243 0 0 0.67
12030 upload-mes 127.0.0.1 34020 127.0.0.1 8078 11 0 3.38
2509 java 127.0.0.1 8078 127.0.0.1 34020 0 11 3.41
12030 upload-mes 127.0.0.1 21196 127.0.0.1 7101 0 0 12.61
3964 mesos-slav 127.0.0.1 7101 127.0.0.1 21196 0 0 12.64
12021 upload-sys 127.0.0.1 34022 127.0.0.1 8078 372 0 15.28
2509 java 127.0.0.1 8078 127.0.0.1 34022 0 372 15.31
2235 dockerd 100.82.34.63 13730 100.82.136.233 7002 0 4 18.50
2235 dockerd 100.82.34.63 34314 100.82.64.53 7002 0 8 56.73
[...]
```

Dynamic tracing of TCP set state only; does not trace send/receive  
Also see: `tcpconnect`, `tcpaccept`, `tcpretrans`

Identify DNS latency issues system wide with `gethostlatency`

```
/usr/share/bcc/tools/gethostlatency
TIME PID COMM LATms HOST
18:56:36 5055 mesos-slave 0.01 100.82.166.217
18:56:40 5590 java 3.53 ec2-...-79.compute-1.amazonaws.com
18:56:51 5055 mesos-slave 0.01 100.82.166.217
18:56:53 30166 ncat 0.21 localhost
18:56:56 6661 java 2.19 atlas-alert-...-prod.netflix.net
18:56:59 5589 java 1.50 ec2-...-207.compute-1.amazonaws.com
18:57:03 5370 java 0.04 localhost
18:57:03 30259 sudo 0.07 titusagent-mainvpc-m...3465
18:57:06 5055 mesos-slave 0.01 100.82.166.217
18:57:10 5590 java 3.10 ec2-...-79.compute-1.amazonaws.com
18:57:21 5055 mesos-slave 0.01 100.82.166.217
18:57:29 5589 java 52.36 ec2-...-207.compute-1.amazonaws.com
18:57:36 5055 mesos-slave 0.01 100.82.166.217
18:57:40 5590 java 1.83 ec2-...-79.compute-1.amazonaws.com
18:57:51 5055 mesos-slave 0.01 100.82.166.217
[...]
```

# dbslower mysql -m 0

Tracing database queries for pids 25776 slower than 0 ms...

```
TIME(s) PID MS QUERY
6.003720 25776 2.363 /* mysql-connector-java-5.1.40 (Revision: 402933ef52cad9aa82624e80acbea46e3a701ce6) */SELECT @@session.auto_i...
6.599219 25776 0.068 SET NAMES latin1
6.613944 25776 0.057 SET character_set_results = NULL
6.645228 25776 0.059 SET autocommit=1
6.653798 25776 0.059 SET sql_mode='NO_AUTO_CREATE_USER,NO_ENGINE_SUBSTITUTION,STRICT_TRANS_TABLES'
6.682184 25776 2.526 select * from users where id = 0
6.767888 25776 0.288 select id from products where userid = 0
6.790642 25776 2.255 call getproduct(0)
6.809865 25776 0.218 call getproduct(1)
6.846878 25776 0.248 select * from users where id = 1
6.847623 25776 0.166 select id from products where userid = 1
6.867363 25776 0.244 call getproduct(2)
6.868162 25776 0.107 call getproduct(3)
6.874726 25776 0.208 select * from users where id = 2
6.881722 25776 0.260 select id from products where userid = 2
```

# ./tools/cachestat.py 1

| TOTAL  | MISSES | HITS | DIRTIES | BUFFERS_MB | CACHED_MB |
|--------|--------|------|---------|------------|-----------|
| 1      | 0      | 1    | 0       | 8          | 283       |
| 0      | 0      | 0    | 0       | 8          | 283       |
| 0      | 0      | 0    | 2       | 8          | 283       |
| 0      | 0      | 0    | 0       | 8          | 283       |
| 10009  | 9173   | 836  | 2       | 9          | 369       |
| 152032 | 152032 | 0    | 0       | 9          | 1028      |
| 157408 | 157405 | 3    | 0       | 9          | 1707      |
| 150432 | 150432 | 0    | 0       | 9          | 2331      |
| 0      | 0      | 0    | 0       | 9          | 2331      |

# Etat des lieux rapide : avec bcc/eBPF...

- Injection d'erreurs

```
Let's say we want to fail bio requests when the request is to some specific sector. An example use case would be to fail superblock writes in btrfs. For btrfs, we know that there must be a superblock at 65536 bytes, or sector 128. This allows us to run the following:
```

```
./inject.py bio -v -I 'linux/blkdev.h' '(({struct gendisk *d = bio->bi_disk; struct disk_part_tbl *tbl = d->part_tbl; struct hd_struct **parts = (void *)tbl + sizeof(struct disk_part_tbl); struct hd_struct **partp = parts + bio->bi_partno; struct hd_struct *p = *partp; dev_t disk = p->_dev.devt; disk == MKDEV(254,16);}) && bio->bi_iter.bi_sector == 128)'
```

- Possibilité de tracer des fonctions kernel dans le détail en filtrant sur les arguments, la durée, ...

# Plus avancé

Pour aller plus loin

- numactl -H
- Architecture NUMA du CPU
- Selon la charge, cela a son importance

```
available: 4 nodes (0-3)
node 0 cpus: 0 1 2 3 4 5 6 7 8 9 40 41 42 43 44 45 46 47 48 49
node 0 size: 98192 MB
node 0 free: 94086 MB
node 1 cpus: 10 11 12 13 14 15 16 17 18 19 50 51 52 53 54 55 56 57 58 59
node 1 size: 98304 MB
node 1 free: 94910 MB
node 2 cpus: 20 21 22 23 24 25 26 27 28 29 60 61 62 63 64 65 66 67 68 69
node 2 size: 98304 MB
node 2 free: 95584 MB
node 3 cpus: 30 31 32 33 34 35 36 37 38 39 70 71 72 73 74 75 76 77 78 79
node 3 size: 98304 MB
node 3 free: 95159 MB
node distances:
node 0 1 2 3
 0: 10 21 31 21
 1: 21 10 21 31
 2: 31 21 10 21
 3: 21 31 21 10
```

- numastat
- Ratio hit/miss

```
[root@... ~]# numastat
 node0 node1
numa_hit 8013690576 8217339921
numa_miss 35450493 2913148280
numa_foreign 2913148280 35450493
interleave_hit 568181143 843729705
local_node 7721409091 7897364817
other_node 327731978 3233123384
```

```
[root@... ~]# numastat
 node0 node1
numa_hit 20733718894 24672588487
numa_miss 590500486 512334153
numa_foreign 512334153 590500486
interleave_hit 18103 18078
local_node 20733659578 24672554374
other_node 590559802 512368266
```

- Répartition mémoire par node  
numastat -cm | egrep 'Node|^Huge|Mem'

|                 | Node 0 | Node 1 | Node 2 | Node 3 | Total  |
|-----------------|--------|--------|--------|--------|--------|
| MemTotal        | 98192  | 98304  | 98304  | 98304  | 393104 |
| MemFree         | 53321  | 94979  | 95755  | 95738  | 339792 |
| MemUsed         | 44872  | 3325   | 2549   | 2566   | 53312  |
| HugePages_Total | 32768  | 0      | 0      | 0      | 32768  |
| HugePages_Free  | 1306   | 0      | 0      | 0      | 1306   |
| HugePages_Surp  | 0      | 0      | 0      | 0      | 0      |

- CPU scaling
  - cat /sys/devices/system/cpu/cpu\*/cpu freq/scaling\_governor  
ondemand|performance
- Différence fréquence
  - cat /proc/cpuinfo

model name : Intel(R) Xeon(R) CPU E5-2440 0 @ 2.40GHz  
cpu MHz : 1200.000
- Nombre de threads d'un process
  - cat /proc/<pid>/status | grep Thread
- Scheduler d'IO par disque
  - cat /sys/block/sdc/queue/scheduler  
`noop anticipatory deadline [cfq]`
- Indice de contention mémoire
  - cat /proc/vmstat | grep nr\_vmscan\_write
- Page dirty du cache OS à écrire
  - cat /proc/meminfo | grep Dirty
  - Ou via le plugin vmem de collectd

- tuned
  - Différents profils possible
  - Profil courant : tuned-adm active
  - Contenu des profils dans /etc/tuned-profiles/...
  - Peut surcharger des sysctl, scheduler, ...

```
Available profiles:
- default
- spindown-disk
- laptop-ac-powersave
- latency-performance
- laptop-battery-powersave
- desktop-powersave
- throughput-performance
- virtual-host
- server-powersave
- virtual-guest
- virtual-guest-eee
- enterprise-storage
```

- Bencher rapidement le CPU

- sysbench --cpu-max-prime=2000000 --threads=<n> cpu run
- ou stress --cpu <n> --timeout 10

```
root@...:~# sysbench --cpu-max-prime=2000000 --threads=8 cpu run
sysbench 1.0.8 (using system LuaJIT 2.0.4)

Running the test with following options:
Number of threads: 8
Initializing random number generator from current time

Prime numbers limit: 2000000
Initializing worker threads...

Threads started!

CPU speed:
 events per second: 3.68

General statistics:
 total time: 11.6876s
 total number of events: 43

Latency (ms):
 min: 1685.29
 avg: 2018.19
 max: 2685.91
 95th percentile: 2320.55
 sum: 86781.96

Threads fairness:
 events (avg/stddev): 5.3750/0.48
 execution time (avg/stddev): 10.8477/0.58
```

- Mémoire kernel
  - slabtop -sc

| System-wide slab statistics |          |     |        |         |       |          |                     |      |      |
|-----------------------------|----------|-----|--------|---------|-------|----------|---------------------|------|------|
|                             |          |     |        |         |       |          |                     |      |      |
| OBJS                        | ACTIVE   | USE | OBJ    | SIZE    | SLABS | OBJ/SLAB | CACHE               | SIZE | NAME |
| 37382284                    | 33125921 | 88% | 0.10K  | 1010332 | 37    | 4041328K | buffer_head         |      |      |
| 5951480                     | 5848655  | 98% | 0.19K  | 297574  | 20    | 1190296K | dentry              |      |      |
| 1290240                     | 1290135  | 99% | 0.98K  | 322560  | 4     | 1290240K | ext4_inode_cache    |      |      |
| 955941                      | 889241   | 93% | 0.55K  | 136563  | 7     | 546252K  | radix_tree_node     |      |      |
| 96980                       | 96485    | 99% | 0.19K  | 4849    | 20    | 19396K   | size-192            |      |      |
| 92099                       | 31626    | 34% | 0.06K  | 1561    | 59    | 6244K    | size-64             |      |      |
| 33040                       | 31555    | 95% | 0.03K  | 295     | 112   | 1180K    | size-32             |      |      |
| 32724                       | 32586    | 99% | 0.14K  | 1212    | 27    | 4848K    | sysfs_dir_cache     |      |      |
| 28975                       | 25504    | 88% | 0.20K  | 1525    | 19    | 6100K    | vm_area_struct      |      |      |
| 27804                       | 26325    | 94% | 0.64K  | 4634    | 6     | 18536K   | proc_inode_cache    |      |      |
| 22270                       | 21364    | 95% | 0.11K  | 655     | 34    | 2620K    | Netatop_sockinfo    |      |      |
| 22253                       | 18344    | 82% | 0.05K  | 289     | 77    | 1156K    | anon_vma_chain      |      |      |
| 19905                       | 17750    | 89% | 0.25K  | 1327    | 15    | 5308K    | filp                |      |      |
| 13290                       | 12503    | 94% | 0.12K  | 443     | 30    | 1772K    | size-128            |      |      |
| 11658                       | 10156    | 87% | 0.05K  | 174     | 67    | 696K     | anon_vma            |      |      |
| 9422                        | 9387     | 99% | 16.00K | 9422    | 1     | 150752K  | size-16384          |      |      |
| 8660                        | 8021     | 92% | 1.00K  | 2165    | 4     | 8660K    | size-1024           |      |      |
| 8064                        | 8022     | 99% | 0.58K  | 1344    | 6     | 5376K    | inode_cache         |      |      |
| 7035                        | 6976     | 99% | 0.69K  | 1407    | 5     | 5628K    | sock_inode_cache    |      |      |
| 6460                        | 6444     | 99% | 1.75K  | 3230    | 2     | 12920K   | TCP                 |      |      |
| 4590                        | 3665     | 79% | 0.11K  | 135     | 34    | 540K     | task_delay_info     |      |      |
| 4440                        | 3720     | 83% | 0.12K  | 148     | 30    | 592K     | pid                 |      |      |
| 4440                        | 2578     | 58% | 0.19K  | 222     | 20    | 888K     | cred_jar            |      |      |
| 4110                        | 4096     | 99% | 0.12K  | 137     | 30    | 548K     | ioat2               |      |      |
| 3810                        | 2305     | 60% | 0.12K  | 127     | 30    | 508K     | eventpoll_epi       |      |      |
| 3604                        | 3003     | 83% | 0.07K  | 68      | 53    | 272K     | Acpi-Operand        |      |      |
| 3441                        | 3330     | 96% | 2.61K  | 1147    | 3     | 9176K    | task_struct         |      |      |
| 3213                        | 2911     | 90% | 0.81K  | 357     | 9     | 2856K    | task_xstate         |      |      |
| 3144                        | 3039     | 96% | 0.50K  | 393     | 8     | 1572K    | size-512            |      |      |
| 2884                        | 2697     | 93% | 0.13K  | 103     | 28    | 412K     | cfq_io_context      |      |      |
| 2784                        | 2675     | 96% | 0.23K  | 174     | 16    | 696K     | cfq_queue           |      |      |
| 2650                        | 1468     | 55% | 0.07K  | 50      | 53    | 200K     | eventpoll_pwq       |      |      |
| 2325                        | 1586     | 68% | 0.25K  | 155     | 15    | 620K     | skbuff_head_cache   |      |      |
| 2160                        | 949      | 43% | 0.02K  | 15      | 144   | 60K      | dm_target_io        |      |      |
| 1952                        | 1083     | 55% | 0.12K  | 61      | 32    | 244K     | Netatop_taskinfo    |      |      |
| 1855                        | 1700     | 91% | 0.77K  | 371     | 5     | 1484K    | shmem_inode_cache   |      |      |
| 1474                        | 887      | 60% | 0.05K  | 22      | 67    | 88K      | dm_io               |      |      |
| 1455                        | 1384     | 95% | 0.25K  | 97      | 15    | 388K     | size-256            |      |      |
| 1360                        | 1211     | 89% | 0.19K  | 68      | 20    | 272K     | bio-0               |      |      |
| 1298                        | 479      | 36% | 0.06K  | 22      | 59    | 88K      | fs_cache            |      |      |
| 1295                        | 797      | 61% | 0.10K  | 35      | 37    | 140K     | ext4_prealloc_space |      |      |
| 1224                        | 705      | 57% | 0.11K  | 36      | 34    | 144K     | jbd2_journal_head   |      |      |
| 1163                        | 1142     | 98% | 8.00K  | 1163    | 1     | 9304K    | size-8192           |      |      |

- Quels process swappent ?

- cat /proc/<pid>/status | grep VmSwap

- Ou petit script :

```
for file in /proc/[0-9]*status ; do
 awk '/Tgid|VmSwap|Name/{printf $2 " " $3} END { print ""}' $file;
done | grep kB | sort -k 3 -n -r | head -15
```

```
java 3395 22824 kB
 3846 7340 kB
python 2398 4512 kB
python 2397 2388 kB
nsreexecd 3666 2352 kB
```

- pmap <pid> #root

- Détail des allocations mémoire d'un process et mode (s : shared)

```
[root@... ~]# pmap 4779
4779: java -Xms768m -Xmx768m -XX:+UseG1GC -XX:MetaspaceSize=128m -XX:MaxMetaspaceSize=128m -Djavax.net.ssl.trustStore='.../cacert_health.jks' -Djavax.net.ssl.trustStorePassword='...' -Djava.security.auth.login.config='.../jaas.conf' -Dhttps.protocols=TLS ? -Dsun.rmi.transport.tcp.handshakeTimeout=3000 -XX:+PrintGCCause -XX:+PrintGCDetails -XX:+PrintTenuringDistribution -XX:+PrintGCDateStamps -XX:+PrintGCTimeStamps -XX:+UseGCLogFileRotation -XX:NumberOfGCLogFiles=4 -XX:GCLogFileSize=400000
0000000000400000 4K r-x-- /usr/java/ /bin/java
0000000000600000 4K rw--- /usr/java/ /bin/java
0000000000178e000 940K rw--- [anon]
00000000d0000000 799104K rw--- [anon]
00000000100c60000 1035904K ----- [anon]
0000003a87000000 128K r-x-- /lib64/ld-2.12.so
00007ff876caf000 1016K rw--- [anon]
00007ff876dad000 12K ----- [anon]
00007ff876db0000 3064K rw--- [anon]
00007ff8770ae000 12K ----- [anon]
00007ff8770b1000 1016K rw--- [anon]
00007ff8771af000 12K ----- [anon]
00007ff8771b2000 1016K rw--- [anon]
00007ff8772b0000 200K r-x-- /usr/java/ /jre/lib/amd64/libsunec.so
00007ff8772e2000 2044K ----- /usr/java/ /jre/lib/amd64/libsunec.so
00007ff8774e1000 24K rw--- /usr/java/ /jre/lib/amd64/libsunec.so
00007ff8ad8fd000 252K r--s- /.../elasticsearch/elasticsearch/lib/x-pack-core.jar
00007ff8ad93c000 4K r--s- /.../elasticsearch/elasticsearch/lib/hdfs-sync-1.2.15.jar
```

- Entropy - génération de nombres aléatoires
  - cat /proc/sys/kernel/random/entropy\_avail
  - Souci si < 300 bytes
  - Selon la source /dev/random ou /dev/urandom,  
comportement bloquant pour l'appelant
  - Parfois nécessaire d'installer un générateur tiers, type haveged

- **ulimit -a**

- Limite système sur le nombres de process, la mémoire, ...
- "too many open files", "unable to create new native thread", ...
- Utilisateur courant
- `sudo -u <user> sh -c "ulimit -a"`

```
core file size (blocks, -c) 0
data seg size (kbytes, -d) unlimited
scheduling priority (-e) 0
file size (blocks, -f) unlimited
pending signals (-i) 128323
max locked memory (kbytes, -l) 64
max memory size (kbytes, -m) unlimited
open files (-n) 8192
pipe size (512 bytes, -p) 8
POSIX message queues (bytes, -q) 819200
real-time priority (-r) 0
stack size (kbytes, -s) 10240
cpu time (seconds, -t) unlimited
max user processes (-u) 10000
virtual memory (kbytes, -v) unlimited
file locks (-x) unlimited
```

- Attention, avec systemd, les /etc/security/\*\* ne sont pas pris en compte  
Configuration par service à faire (LimitNOFILE, LimitNPROC, ...)

- PSI : Pressure Stall Information

- Baromètre de « pression » sur mémoire, CPU et IO
- Plus fin et simple que la load average  
Ne dépend pas du nombre de core
- Kernel > 4.20
- cat /proc/pressure/cpu|memory|io
- some : % de temps, un ou plusieurs tasks ayant subi une latence durant les 10, 60, 300 dernières secondes
- full : idem mais pour toutes les tasks
- total : temps total en microsecondes

```
[root@centos ~]$: cat /proc/pressure/cpu
some avg10=0.09 avg60=0.02 avg300=0.00 total=5771445
```

```
[root@centos ~]$: cat /proc/pressure/memory
some avg10=8.10 avg60=2.40 avg300=1.53 total=7224990
full avg10=4.18 avg60=1.37 avg300=1.02 total=5041949
```

# DNS

- Souvent une source de problème ("Everything is a Freaking DNS problem")
- dig +search <hostname> (query time intéressant)

```
[root@... ~]# dig www.google.com

; <>> DiG [tcp] 8.8.8.8.53 +cmd
; global options: +cmd
; Got answer:
;=>>HEADER<<- opcode: QUERY, status: NOERROR, id: 36964
; flags: qr rd ra; QUERY: 1, ANSWER: 1, AUTHORITY: 0, ADDITIONAL: 0

; QUESTION SECTION:
;www.google.com. IN A

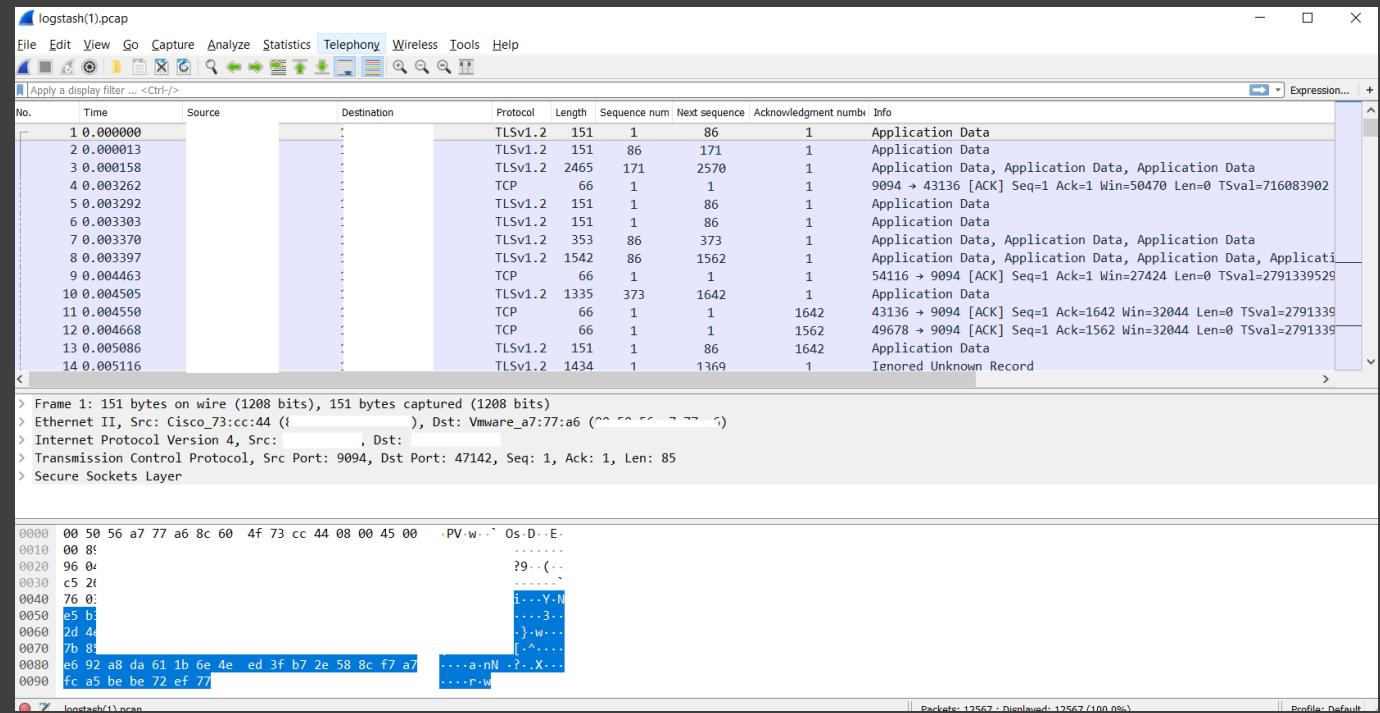
;ANSWER SECTION:
;www.google.com. 3600 IN A 209.85.128.130
; Query time: 3 msec
; SERVER: 128.0.0.1#53(128.0.0.1)
; WHEN: Thu Dec 6 21:21:44 2018
; MSG SIZE rcvd: 54
```

- reverse DNS : dig -x <ip>
- A vérifier :
  - primaire en premier sur le même site
  - sinon latence inter-site et retry
  - Alignement DNS et reverse DNS : cause souvent des problèmes sur les systèmes distribués
- Si dig n'est pas installé : nslookup <host|ip>

- Inspecter le trafic réseau via tcpdump #root

- Attention, énorme overhead
- tcpdump -w <trace.pcap>
  - [-i <une\_interface> port <un\_port>]
  - Si beaucoup de « packets dropped by kernel », les options suivantes peuvent aider :
    - [-nn] : pas de résolution de nom, de protocole
    - [-B 4096] : taille du buffer
    - [-s 68] : ne capture pas le paquet entier

- Analyse ensuite avec Wireshark
- Layout de colonnes (seq, next, ack) intéressant
- Les interprétations données peuvent être erronées.



- ip route    `/sbin`

```
172.17.0.0/16 via 1:2 dev bond1
172.17.0.0/16 dev bond1 proto kernel scope link src 172.17.0.1
172.17.0.0/16 via 1:2 dev bond1
172.17.0.0/16 dev bond0 proto kernel scope link src 172.17.0.1
172.17.0.0/16 via 1:2 dev bond1
default via 172.17.0.1 dev bond0
```

- iptables -S | -L    `#root`

```
[...]\# iptables -L
Chain INPUT (policy ACCEPT)
target prot opt source destination
Chain FORWARD (policy ACCEPT)
target prot opt source destination
Chain OUTPUT (policy ACCEPT)
target prot opt source destination
[...]\# iptables -S
-P INPUT ACCEPT
-P FORWARD ACCEPT
-P OUTPUT ACCEPT
```

- Copie d'écran : configuration vide ou iptables désactivé
- `-n -v -x` : statistiques détaillées avec nombre de paquets et bytes

- Vérifier la bande passante réseau entre deux machines :
  - Sur la cible : iperf -s

```
[root@... ~]# iperf -s

Server listening on TCP port 5001
TCP window size: 512 KByte (default)

[4] local ... port 5001 connected with ... port 54398
[ID] Interval Transfer Bandwidth
[4] 0.0-10.0 sec 11.5 GBytes 9.90 Gbits/sec
```

- Sur le client : iperf -c <hostname\_cible>

```
...> iperf -c ...

client connecting to ..., TCP port 5001
TCP window size: 512 KByte (default)

[3] local ... port 54398 connected with ... port 5001
[ID] Interval Transfer Bandwidth
[3] 0.0-10.0 sec 11.5 GBytes 9.90 Gbits/sec
```

- Tester l'établissement d'une connexion TLS
  - openssl s\_client -connect <host>:<port>

- Tracer tous les appels systèmes d'un programme : strace
  - Overhead très important
  - strace -s 64000000 -T -v -ttt -fp <pid> > strace.log 2>&1

```
[pid 18580] 1544168204.563811 futex(0x7fc8c5a97f28, FUTEX_WAKE_PRIVATE, 1) = 0 <0.000035>
[pid 18580] 1544168204.563898 futex(0x7fc8c5a97f54, FUTEX_WAIT_BITSET_PRIVATE, 1, {49138120, 820297286}, ffffffff) = -1 ETIMEDOUT (Connection timed out) <0.050098>
[pid 18580] 1544168204.614078 futex(0x7fc8c5a97f28, FUTEX_WAKE_PRIVATE, 1) = 0 <0.000035>
[pid 18580] 1544168204.614164 futex(0x7fc8c5a97f54, FUTEX_WAIT_BITSET_PRIVATE, 1, {49138120, 870562813}, ffffffff <unfinished ...>
[pid 18716] 1544168204.616625 <... restart_syscall resumed> = -1 ETIMEDOUT (Connection timed out) <2.498966>
[pid 18716] 1544168204.616684 futex(0x7fc8c6af9228, FUTEX_WAKE_PRIVATE, 1) = 0 <0.000045>
[pid 18734] 1544168204.616978 <... futex resumed> = -1 ETIMEDOUT (Connection timed out) <1.000052>
[pid 18734] 1544168204.617013 futex(0x2419828, FUTEX_WAKE_PRIVATE, 1) = 0 <0.000017>
[pid 18734] 1544168204.617065 futex(0x2419854, FUTEX_WAIT_BITSET_PRIVATE, 1, {49138121, 823471133}, ffffffff <unfinished ...>
[pid 18716] 1544168204.618332 write(565, "\0\0\7\301\32\10\335\212\32\32\22Re_...rt \1\244\177\ n \n\24: \20\224]\30\260\312\366\330\363,\22\377~\10\0\20\277\365\335\35\30\2477 \254]" \306\1\ n\205\1\10\1\22\200\1`-.-`- -filtrage-cache.data,-1948518176_20170119-22:24:21.925_407943_13260572,1527001268268.b9e4f112e510b24e1665cc3861ec28f2.\20\1\30\2 \304\21(\324\0010\0008\08\327\275\16H\36P,X,"2h\376\24p\240\6x\377\223\245\1\0\0\200?\210\1\237\365\343\367\366,\222\1\10\ n\1E\20\377\223\245\1^n\314\1\ n\203\1\10\1\22\177\`-.-`- -correlation-cache.data,29541036_20171117-01:40:17.380_382407_2449909,1527000371597.F939d357f62fc36a8db75d9f7f369812.\20\1\30\2 \212\31(\272\0030\0008\08\216\233\317\1H\270\202\206\3P\205\325\vx\205\325\vx\`3h\214\34p\200\6x\274\350\365(\205\1\0\0\200?\210\1\242\232\301\253\370,\222\1\10\ n\1E\20\274\350\365(`240\1\ nZ\10\1\22\Evenement,<62824R0006-01863\0\0\1W\233\234\7&y,1527034248844.34571f534eb2cb977316bfd1a3a16c0d.\20\1\30\1 \275\246\1(\356\0310\0008\08\227\1H\331hP\215\215\230\10X\215\215\230\10\`20h\274\245\1p\300\34x\260\355Q\205\1\0\0\200?\210\1\30\1\375\301\266\370,\222\1\17\ n\1e\20\260\355Q\`314\1\ n\203\1\10\1\22\177`-.-`-filtrage-cache.data,-2031036142_20170426-10:18:42.265_111681_1433386,1527001268268.9ff9e010642d255b550314901f9112f3.\20\1\30\2 \200\260\1\266\0250\0008\08\241\3510H\332\251\1P\315\226\4X\315\226\4`th\361\320\1p\240Ax\247\371\333\r\205\1\0\0\2007\210\1\240\305\211\304\367,\222\1\10\ n\1E\20\247\371\333\r\`257\1\ n\10\1\22e`-.-`-cache.data,dictionaryPd
```

- Appels aux librairies : ltrace

- RedHat Transparent Huge Pages (THP)

- Fréquente source de « hiccup »
- A désactiver

```
[root@... ~]# cat /sys/kernel/mm/redhat_transparent_hugepage/enabled
always madvise [never]
```

- Swap

- Un système peut swapper même si il y a de l'espace libre
- Mauvais pour une JVM
- Paramètre swappiness à 1 `/sbin`

```
[root@... ~]# sysctl vm.swappiness
vm.swappiness = 1
```

## • perf top [-p <pid>] #root

Samples: 169K of event 'cpu-clock', Event count (approx.): 26351879957  
 Overhead Shared Object

| Overhead | Shared Object      | Symbol                                                  |
|----------|--------------------|---------------------------------------------------------|
| 7.10%    | libjvm.so          | [.] CMTTask::deal_with_reference                        |
| 3.53%    | libjvm.so          | [.] oopDesc::size                                       |
| 2.34%    | libjvm.so          | [.] InstanceKlass::oop_oop_iterate_nv                   |
| 2.30%    | libjvm.so          | [.] Dictionary::find                                    |
| 2.17%    | [kernel]           | [k] _spin_unlock_irqrestore                             |
| 1.94%    | libjvm.so          | [.] CMTTask::drain_local_queue                          |
| 1.44%    | libjvm.so          | [.] ConcurrentMark::count_region                        |
| 1.22%    | [kernel]           | [k] finish_task_switch                                  |
| 0.87%    | libjvm.so          | [.] ObjPtrQueue::should_enqueue_buffer                  |
| 0.83%    | [kernel]           | [k] e1000_xmit_frame                                    |
| 0.75%    | libjvm.so          | [.] ClassLoaderDataGraphKlassIteratorAtomic::next_klass |
| 0.75%    | libjvm.so          | [.] CMTTask::do_marking_step                            |
| 0.73%    | libjvm.so          | [.] SymbolTable::lookup                                 |
| 0.71%    | libjvm.so          | [.] SymbolTable::possibly_parallel_unlink               |
| 0.64%    | libjvm.so          | [.] CodeHeap::find_start                                |
| 0.60%    | libjvm.so          | [.] java_lang_Throwable::fill_in_stack_trace            |
| 0.59%    | libjvm.so          | [.] ObjArrayKlass::oop_oop_iterate_nv                   |
| 0.53%    | libjvm.so          | [.] BacktraceBuilder::push                              |
| 0.51%    | libjvm.so          | [.] ObjectSynchronizer::deflate_idle_monitors           |
| 0.42%    | libjvm.so          | [.] ClassLoaderData::oops_do                            |
| 0.41%    | libjvm.so          | [.] CMBitMapClosure::do_bit                             |
| 0.34%    | libjvm.so          | [.] OtherRegionsTable::add_reference                    |
| 0.33%    | libjvm.so          | [.] Dictionary::roots_oops_do                           |
| 0.32%    | libjvm.so          | [.] nmethod::do_unloading_parallel                      |
| 0.32%    | libjvm.so          | [.] G1ParScanThreadState::trim_queue                    |
| 0.31%    | libjvm.so          | [.] CMTTask::drain_global_stack                         |
| 0.28%    | libjvm.so          | [.] InstanceKlass::vtable_length                        |
| 0.27%    | libjvm.so          | [.] G1ParScanThreadState::copy_to_surv                  |
| 0.25%    | [kernel]           | [k] __do_softirq                                        |
| 0.25%    | libpthread-2.12.so | [.] pthread_getspecific                                 |
| 0.24%    | perf-28554.map     | [.] 0x00007f170627dc61                                  |
| 0.24%    | [kernel]           | [k] e1000_clean                                         |
| 0.22%    | libjvm.so          | [.] G1KlassScanClosure::do_klass                        |
| 0.22%    | libjvm.so          | [.] StringTable::possibly_parallel_oop                  |
| 0.22%    | perf-28554.map     | [.] 0x00007f17032ab76e                                  |
| 0.21%    | libjvm.so          | [.] InstanceKlass::clean_weak_instance                  |
| 0.21%    | [kernel]           | [k] e1000_intr                                          |
| 0.20%    | libjvm.so          | [.] Method::set_on_stack                                |
| 0.20%    | libjvm.so          | [.] MetadataAwareOopClosure::do_class_loader_data       |
| 0.19%    | libjvm.so          |                                                         |
| 0.19%    | libjvm.so          |                                                         |
| 0.19%    | libjvm.so          |                                                         |
| 3.85%    | libjvm.so          |                                                         |
| 2.52%    | libjvm.so          |                                                         |
| 2.16%    | perf-7936.map      |                                                         |
| 1.92%    | perf-7936.map      |                                                         |
| 1.61%    | libzip.so          |                                                         |
| 1.49%    | libjvm.so          |                                                         |
| 1.41%    | libjvm.so          |                                                         |
| 1.41%    | libjvm.so          |                                                         |
| 1.40%    | [kernel]           |                                                         |
| 1.28%    | libjvm.so          |                                                         |
| 1.18%    | libjvm.so          |                                                         |

| Overhead | Shared Object | Symbol                                                               |
|----------|---------------|----------------------------------------------------------------------|
| 6.01%    | [kernel]      | [k] __do_softirq                                                     |
| 3.55%    | libjvm.so     | [.] java_lang_Throwable::fill_in_stack_trace                         |
| 2.55%    | libjvm.so     | [.] BacktraceBuilder::push                                           |
| 1.90%    | [kernel]      | [k] _raw_spin_unlock_irqrestore                                      |
| 1.61%    | libjvm.so     | [.] CodeHeap::find_start                                             |
| 1.46%    | libzip.so     | [.] longest_match                                                    |
| 1.12%    | [kernel]      | [k] finish_task_switch                                               |
| 1.10%    | libzip.so     | [.] deflate_slow                                                     |
| 0.79%    | libjvm.so     | [.] InstanceKlass::oop_oop_iterate_nv                                |
| 0.71%    | libjvm.so     | [.] ParNewGeneration::copy_to_survivor_space_avoiding_promotion_undo |

| Overhead | Shared Object | Symbol                                       |
|----------|---------------|----------------------------------------------|
| 3.85%    | libjvm.so     | [.] Dictionary::find                         |
| 2.52%    | libjvm.so     | [.] CMTTask::deal_with_reference             |
| 2.16%    | perf-7936.map | [.] 0x00007f1c7c4b522                        |
| 1.92%    | perf-7936.map | [.] 0x00007f1c33fd5f6                        |
| 1.61%    | libzip.so     | [.] ZIP_GetEntry2                            |
| 1.49%    | libjvm.so     | [.] java_lang_Throwable::fill_in_stack_trace |
| 1.41%    | libjvm.so     | [.] BacktraceBuilder::push                   |
| 1.41%    | libjvm.so     | [.] oopDesc::size                            |
| 1.40%    | [kernel]      | [k] _spin_unlock_irqrestore                  |
| 1.28%    | libjvm.so     | [.] StringTable::intern                      |
| 1.18%    | libjvm.so     | [.] CodeHeap::find_start                     |

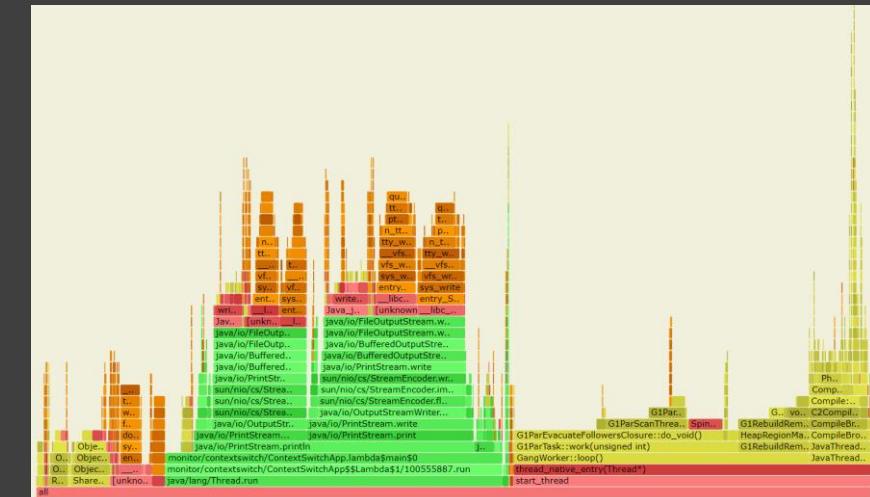
- `perf stat -ad -- sleep 10 #root`
  - [-A] detail par cpu
  - Analyse IPC : sur du physique ou virtuel si exposé (vPMC sur ESX)
  - If your IPC is < 1.0, you are likely memory stalled, and software tuning strategies include reducing memory I/O, and improving CPU caching and memory locality, especially on NUMA systems. Hardware tuning includes using processors with larger CPU caches, and faster memory, busses, and interconnects.
  - If your IPC is > 1.0, you are likely instruction bound. Look for ways to reduce code execution: eliminate unnecessary work, cache operations, etc. CPU flame graphs are a great tool for this investigation. For hardware tuning, try a faster clock rate, and more cores/hyperthreads.
  - **Plus de core ou plus rapide pas forcément utile**

| Performance counter stats for 'system wide': |                         |                                |           |
|----------------------------------------------|-------------------------|--------------------------------|-----------|
| 240344.655818                                | task-clock (msec)       | # 24.033 CPUs utilized         | (100.00%) |
| 144,079                                      | context-switches        | # 0.599 K/sec                  | (100.00%) |
| 2,439                                        | cpu-migrations          | # 0.010 K/sec                  | (100.00%) |
| 58,853                                       | page-faults             | # 0.245 K/sec                  |           |
| 18,041,072,111                               | cycles                  | # 0.075 GHz                    | (30.02%)  |
| 31,631,154,710                               | stalled-cycles-frontend | # 175.33% frontend cycles idle | (40.00%)  |
| 29,407,722,150                               | stalled-cycles-backend  | # 163.00% backend cycles idle  | (40.01%)  |
| 7,618,970,547                                | instructions            | # 0.42 insns per cycle         |           |
|                                              |                         | # 4.15 stalled cycles per insn | (50.00%)  |
| 1,349,590,739                                | branches                | # 5.615 M/sec                  | (50.00%)  |
| 23,085,732                                   | branch-misses           | # 1.71% of all branches        | (50.00%)  |
| 1,956,327,322                                | L1-dcache-loads         | # 8.140 M/sec                  | (49.95%)  |
| 111,858,077                                  | L1-dcache-load-misses   | # 5.72% of all L1-dcache hits  | (20.01%)  |
| 47,939,203                                   | LLC-loads               | # 0.199 M/sec                  | (29.99%)  |
| 13,146,794                                   | LLC-load-misses         | # 27.42% of all LL-cache hits  | (39.99%)  |
| 10.000798555 seconds time elapsed            |                         |                                |           |

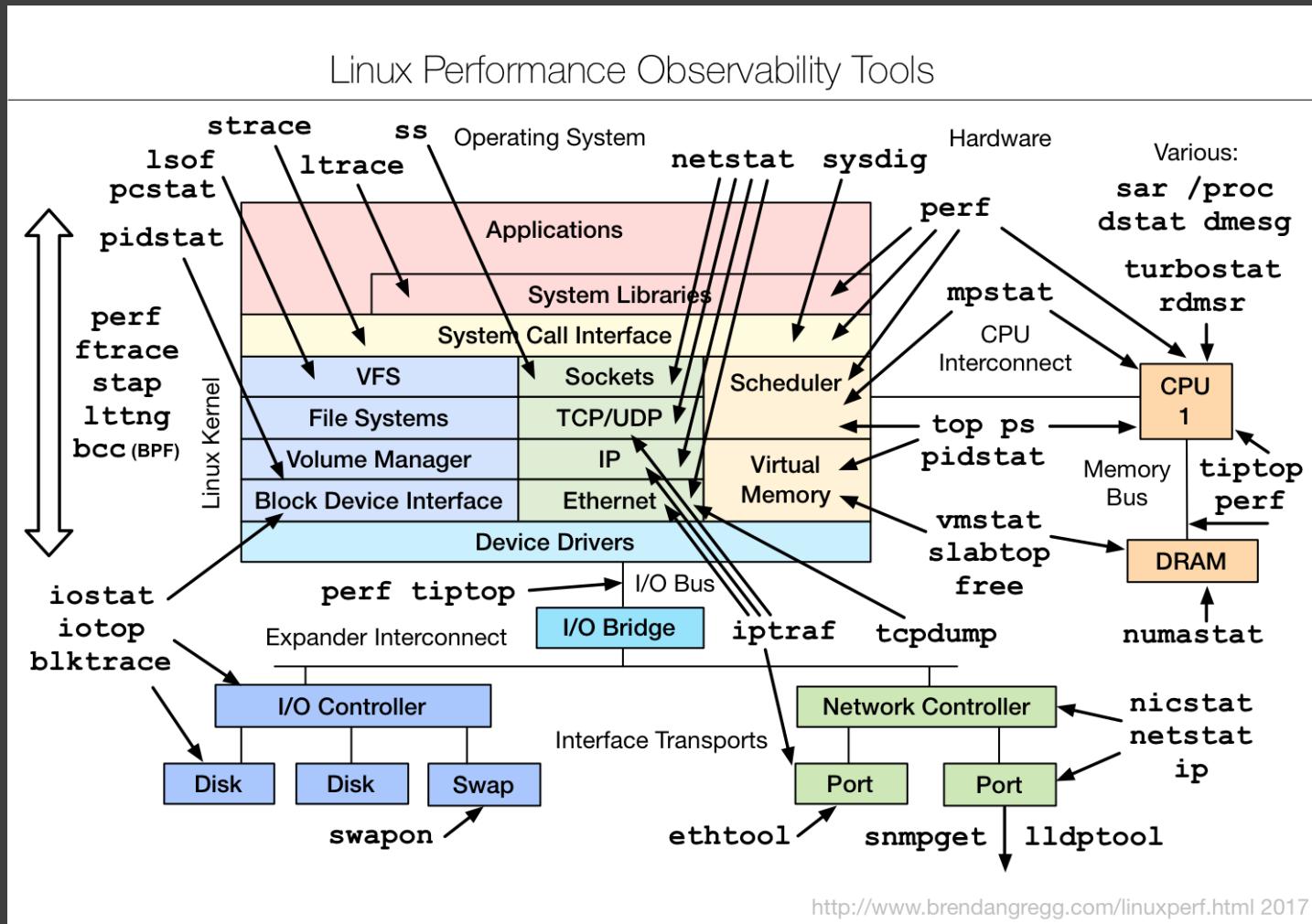
# FlameGraph via perf

#root

- Sur une charge fortement dominée par du CPU
  - Visualiser rapidement le comportement d'une application (on-cpu)
  - Axe Y : stack d'appel, feuille en haut et on-cpu, thread en bas
  - Axe X : tri alphabétique, la largeur dépend du temps passé (nb de samples)
  - Couleur random (ou pas)
- Linux 2.6.x
  - `perf record -F 99 -a --call-graph dwarf -- sleep 10`
  - `git --depth 1 clone https://github.com/brendangregg/FlameGraph`
  - `perf script [-i perf.data]| ./stackcollapse-perf.pl | ./FlameGraph/flamegraph.pl > perf.svg`
- Linux 4.9+
  - `git clone --depth 1 https://github.com/brendangregg/FlameGraph`
  - `git clone --depth 1 https://github.com/iovisor/bcc`
  - `./bcc/tools/profile.py -df -F 99 10 | ./FlameGraph/flamegraph.pl > perf.svg`



# Beaucoup d'autres



# Remarques

- Comment est ce mesuré, échantillonné, avec quel overhead ?
- Principalement du parsing de /proc
- Stats : attention, une moyenne cache les outliers, moyenne de moyenne, de percentiles, ...
- Bon ou mauvais CPU : si beaucoup d'IO wait, plus de CPU ne sert à rien
- % CPU IO wait : comparaison délicate, élevé peut être car disques très lents ; charge consommant énormément de CPU user, masquant les IO wait
- Compréhension délicate, parfois pas de doc outre les sources du kernel
- Métriques disques : % d'utilisation, IOPS de quelle taille, IO sync/async, ...

- Une perspective
- Performance is a field where "the more you know, the more you don't know".  
It's the same principle: the more you learn about systems, the more unknown unknowns you become aware of, which are then known-unknowns that you can check on.

# Ecosystème Java

Plus dans le détail sur une JVM

- 1) USERED
- 2) Dominant consumer
  - Comportement CPU user ou system
  - Oriente les investigations
    - CPU system dominant > 10-15% : réseau, IO, ...
    - CPU user dominant
      - Utilisation mémoire / consommation GC : tuning GC ..., correction applicative
      - Sinon profiling application
    - ...
- Utilisation d'outils spécifiques à l'écosystème Java riche

# Activité des threads

- Photo de l'état des thread : thread dump
  - freeze la JVM un court instant
  - kill -3
  - sortie dans /proc/<pid>/fd/1
  - jstack fonctionne aussi mais d'expérience, pas sur une JVM mal en point
  - ou jcmt <pid> Thread.print -l -e > <un\_fichier>
  - corrélation consommation cpu
- top -b -H -n 1 -p<pid> | head -15
  - printf "%x\n" <tid>
  - Recherche dans le thread dump de "nid=0x..."

| PID   | USER | PR | NI | VIRT  | RES  | SHR | S | %CPU | %MEM | TIME+    | COMMAND |
|-------|------|----|----|-------|------|-----|---|------|------|----------|---------|
| 21327 | v    | 20 | 0  | 8855m | 4.1g | 15m | S | 13.7 | 26.2 | 20:00.22 | java    |
| 21340 | v    | 20 | 0  | 8855m | 4.1g | 15m | S | 9.8  | 26.2 | 19:35.09 | java    |
| 21931 | v    | 20 | 0  | 8855m | 4.1g | 15m | S | 3.9  | 26.2 | 0:34.41  | java    |
| 21146 | v    | 20 | 0  | 8855m | 4.1g | 15m | S | 2.0  | 26.2 | 25:27.17 | java    |
| 21331 | v    | 20 | 0  | 8855m | 4.1g | 15m | S | 2.0  | 26.2 | 13:21.39 | java    |
| 21333 | v    | 20 | 0  | 8855m | 4.1g | 15m | S | 2.0  | 26.2 | 19:52.15 | java    |
| 21343 | v    | 20 | 0  | 8855m | 4.1g | 15m | S | 2.0  | 26.2 | 20:55.80 | java    |
| 21729 | v    | 20 | 0  | 8855m | 4.1g | 15m | S | 2.0  | 26.2 | 18:38.02 | java    |

```
[root@534f ~]# printf "%x\n" 21327
```

```
"[ACTIVE] ExecuteThread: '4' for queue: 'weblogic.kernel.Default (self-tuning)'" daemon prio=10 tid=0x00007f436f54e800 nid=0x534f runnable [0x00007f42d7efa000]
 java.lang.Thread.State: RUNNABLE
 at java.net.SocketInputStream.socketRead0(Native Method)
 at java.net.SocketInputStream.read(SocketInputStream.java:152)
 at java.net.SocketInputStream.read(SocketInputStream.java:122)
 at oracle.net.ns.Packet.receive(Packet.java:300)
 at oracle.net.ns.DataPacket.receive(DataPacket.java:106)
 at oracle.net.ns.NetInputStream.getNextPacket(NetInputStream.java:315)
 at oracle.net.ns.NetInputStream.read(NetInputStream.java:260)
 at oracle.net.ns.NetInputStream.read(NetInputStream.java:185)
 at oracle.net.ns.NetInputStream.read(NetInputStream.java:102)
 at oracle.jdbc.driver.T4CSocketInputStreamWrapper.readNextPacket(T4CSocketInputStreamWrapper.java:124)
 at oracle.jdbc.driver.T4CSocketInputStreamWrapper.read(T4CSocketInputStreamWrapper.java:80)
 at oracle.jdbc.driver.T4CMAREngine.unmarshalUB1(T4CMAREngine.java:1137)
 at oracle.jdbc.driver.T4CTTIfun.receive(T4CTTIfun.java:290)
 at oracle.jdbc.driver.T4CTTIfun.doRPC(T4CTTIfun.java:192)
 at oracle.jdbc.driver.T4C80all.doALL(T4C80all.java:531)
 at oracle.jdbc.driver.T4CPreparedStatement.doAll18(T4CPreparedStatement.java:207)
 at oracle.jdbc.driver.T4CPreparedStatement.executeForRows(T4CPreparedStatement.java:1044)
 at oracle.jdbc.driver.OracleStatement.doExecuteWithTimeout(OracleStatement.java:1329)
 at oracle.jdbc.driver.OraclePreparedStatement.executeInternal(OraclePreparedStatement.java:3593)
 at oracle.jdbc.driver.OraclePreparedStatement.execute(OraclePreparedStatement.java:3694)
 - locked <0x00000007d318c4d0> (a oracle.jdbc.driver.T4CConnection)
 at oracle.jdbc.driver.OraclePreparedStatementWrapper.execute(OraclePreparedStatementWrapper.java:1378)
 at weblogic.jdbc.wrapper.PreparedStatement.execute(PreparedStatement.java:99)
 at org.apache.ibatis.executor.statement.PreparedStatementHandler.update(PreparedStatementHandler.java:46)
 at org.apache.ibatis.executor.statement.RoutingStatementHandler.update(RoutingStatementHandler.java:74)
 at org.apache.ibatis.executor.SimpleExecutor.doUpdate(SimpleExecutor.java:50)
 at org.apache.ibatis.executor.BaseExecutor.update(BaseExecutor.java:117)
 at sun.reflect.GeneratedMethodAccessor615.invoke(Unknown Source)
 at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
 at java.lang.reflect.Method.invoke(Method.java:606)
 at org.apache.ibatis.plugin.Invocation.proceed(Invocation.java:49)
 at fr.edf.distribution.linky.technique.mybatisplus.TimeoutConfigurerInterceptor.intercept(TimeoutConfigurerInterceptor.java:69)
 at org.apache.ibatis.plugin.Plugin.invoke(Plugin.java:61)
 at com.sun.proxy.$Proxy2238.update(Unknown Source)
 at org.apache.ibatis.session.defaults.DefaultSqlSession.update(DefaultSqlSession.java:198)
 at fr.edf.distribution.linky.technique.mybatisplus.DispatchSqlSession.update(DispatchSqlSession.java:172)
 at sun.reflect.GeneratedMethodAccessor652.invoke(Unknown Source)
 at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
 at java.lang.reflect.Method.invoke(Method.java:606)
 at org.mybatis.spring.SqlSessionTemplate$SqlSessionInterceptor.invoke(SqlSessionTemplate.java:433)
 at com.sun.proxy.$Proxy817.update(Unknown Source)
 at org.mybatis.spring.SqlSessionTemplate.update(SqlSessionTemplate.java:294)
 at org.apache.ibatis.binding.MapperMethod.execute(MapperMethod.java:62)
 at org.apache.ibatis.binding.MapperProxy.invoke(MapperProxy.java:59)
 at com.sun.proxy.$Proxy1011.updatePlanifIntervention(Unknown Source)
 at sun.reflect.GeneratedMethodAccessor1936.invoke(Unknown Source)
 at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
```

# Activité des threads

- Outils d'aide à l'analyse :
  - <http://spotify.github.io/threaddump-analyzer/>
  - <https://mkbrv.github.io/tda/>
- jcmd <pid> PerfCounter.print | grep java.threads

```
java.threads.daemon=114
java.threads.live=116
java.threads.livePeak=116
java.threads.started=121
```

- Depuis Java 9, le nom des threads Java est transmis à l'OS (16 premiers caractères)

```
root@...:/opt# pstree -pt $(pidof java)
java(3397)---{C1 CompilerThre}(3409)
 |---{C2 CompilerThre}(3408)
 |---{Common-Cleaner}(3413)
 |---{ContainerBackgr}(3425)
 |---{Finalizer}(3406)
 |---{G1 Conc#0}(3401)
 |---{G1 Conc#1}(3420)
 |---{G1 Main Marker}(3400)
 |---{G1 Refine#0}(3402)
 |---{G1 Refine#1}(3431)
 |---{G1 Young RemSet}(3403)
 |---{GC Thread#0}(3399)
 |---{GC Thread#1}(3415)
 |---{GC Thread#2}(3416)
 |---{GC Thread#3}(3419)
 |---{GC Thread#4}(3430)
 |---{NioBlockingSele}(3432)
 |---{Reference Handl}(3405)
 |---{Service Thread}(3411)
 |---{Signal Dispatch}(3407)
 |---{SimplePauseDete}(3424)
 |---{Sweeper thread}(3410)
 |---{Thread-3}(3423)
 |---{VM Periodic Tas}(3412)
 |---{VM Thread}(3404)
 |---{container-0}(3426)
 |---{http-nio-8080-A}(3445)
 |---{http-nio-8080-A}(3446)
 |---{http-nio-8080-C}(3443)
 |---{http-nio-8080-C}(3444)
 |---{http-nio-8080-e}(3433)
 |---{http-nio-8080-e}(3434)
 |---{http-nio-8080-e}(3435)
 |---{http-nio-8080-e}(3436)
 |---{http-nio-8080-e}(3437)
 |---{http-nio-8080-e}(3438)
 |---{http-nio-8080-e}(3439)
 |---{http-nio-8080-e}(3440)
 |---{http-nio-8080-e}(3441)
 |---{http-nio-8080-e}(3442)
 |---{java}(3398)
 |---{spring.cloud.in}(3429)

top - 10:29:58 up 5:17, 1 user, load average: 1,28, 0,54, 0,28
Threads: 564 total, 8 en cours, 556 en veille, 0 arrêté, 0 zombie
%Cpu(s): 39,6 ut, 2,5 sy, 0,0 ni, 57,4 id, 0,0 wa, 0,0 hi, 0,5 si, 0,0 st
KiB Mem : 10162568 total, 7995728 libr, 1265892 util, 900948 tmap/cache
KiB Éch: 483800 total, 483800 libr, 0 util. 8576908 dispo Mem

 PID UTIL. PR NI VIRT RES SHR S %CPU %MEM TEMPS+ COM.
 3312 20 0 7183972 464892 22316 R 91,1 4,6 0:17.07 C2 CompilerThre
 3352 20 0 7183972 464892 22316 R 90,1 4,6 0:06.71 C2 CompilerThre
 3313 20 0 7183972 464892 22316 S 19,5 4,6 0:05.56 C1 CompilerThre
 3344 20 0 7183972 464892 22316 S 8,6 4,6 0:00.55 http-nio-8080-e
 3339 20 0 7183972 464892 22316 S 8,3 4,6 0:00.58 http-nio-8080-e
 3346 20 0 7183972 464892 22316 S 8,3 4,6 0:00.56 http-nio-8080-e
 3338 20 0 7183972 464892 22316 S 7,9 4,6 0:00.60 http-nio-8080-e
 3341 20 0 7183972 464892 22316 S 7,9 4,6 0:00.57 http-nio-8080-e
 3342 20 0 7183972 464892 22316 S 7,9 4,6 0:00.54 http-nio-8080-e
 3343 20 0 7183972 464892 22316 S 7,9 4,6 0:00.53 http-nio-8080-e
 3345 20 0 7183972 464892 22316 S 7,9 4,6 0:00.54 http-nio-8080-e
 3337 20 0 7183972 464892 22316 S 7,6 4,6 0:00.88 http-nio-8080-e
 3340 20 0 7183972 464892 22316 S 7,6 4,6 0:00.56 http-nio-8080-e
```

- Depuis Java 11, temps CPU et elapsed

```
"C1 CompilerThread0" #7 daemon prio=9 os_prio=0 cpu=58,26ms elapsed=49,36s tid=0x00007f3ef419e800 nid=0x3524 waiting on condition [0x0000000000000000]
 java.lang.Thread.State: RUNNABLE
 No compile task

"Sweeper thread" #8 daemon prio=9 os_prio=0 cpu=1,86ms elapsed=49,36s tid=0x00007f3ef41a0800 nid=0x3525 runnable [0x0000000000000000]
 java.lang.Thread.State: RUNNABLE

"Service Thread" #9 daemon prio=9 os_prio=0 cpu=0,06ms elapsed=49,32s tid=0x00007f3ef421f000 nid=0x3526 runnable [0x0000000000000000]
 java.lang.Thread.State: RUNNABLE

"Common-Cleaner" #10 daemon prio=8 os_prio=0 cpu=0,17ms elapsed=49,31s tid=0x00007f3ef4233000 nid=0x3528 in Object.wait() [0x00007f3ecd114000]
 java.lang.Thread.State: TIMED_WAITING (on object monitor)
 at java.lang.Object.wait(java.base@11.0.1/Native Method)
 - waiting on <0x000000076e822568> (a java.lang.ref.ReferenceQueue$Lock)
 at java.lang.ref.ReferenceQueue.remove(java.base@11.0.1/ReferenceQueue.java:155)
 - waiting to re-lock in wait() <0x000000076e822568> (a java.lang.ref.ReferenceQueue$Lock)
 at jdk.internal.ref.CleanerImpl.run(java.base@11.0.1/CleanerImpl.java:148)
 at java.lang.Thread.run(java.base@11.0.1/Thread.java:834)
 at jdk.internal.misc.InnocuousThread.run(java.base@11.0.1/InnocuousThread.java:134)

"Thread-3" #14 prio=5 os_prio=0 cpu=32676,75ms elapsed=49,28s tid=0x00007f3ef4289000 nid=0x352c runnable [0x00007f3ecc8e7000]
"Thread-4" #15 prio=5 os_prio=0 cpu=32060,82ms elapsed=49,28s tid=0x00007f3ef428b000 nid=0x352d runnable [0x00007f3ecc7e6000]
"Thread-5" #16 prio=5 os_prio=0 cpu=31776,09ms elapsed=49,28s tid=0x00007f3ef428c800 nid=0x352e runnable [0x00007f3ecc6e5000]
```

- jcmt -e : allocated et defined\_classes

```
"http-nio-8080-exec-8" #29 daemon prio=5 os_prio=0 cpu=2807.51ms elapsed=462.36s allocated=163M defined_classes=33 tid=0x00007f4055b26000 nid=0x36 runnable [0x00007f3fc3cf9000]
 java.lang.Thread.State: RUNNABLE
 at java.lang.StringLatin1.hashCode(java.base@11.0.3/Unknown Source)
 at java.lang.String.hashCode(java.base@11.0.3/Unknown Source)
 at java.util.HashMap.hash(java.base@11.0.3/Unknown Source)
 at java.util.HashMap.get(java.base@11.0.3/Unknown Source)
 at org.apache.catalina.connector.Request.getAttribute(Request.java:866)
 at org.apache.catalina.connector.RequestFacade.getAttribute(RequestFacade.java:282)
```

# Activités threads, gc ... tout en un

- Vision rapide : jvm-tools
  - <https://github.com/aragozin/jvm-tools>
  - Top thread, allocation, cpu user et system
  - GC
- A installer sur toutes machines hébergeant une JVM
- [sudo -u <user>] java -jar sjk-plus-<version>.jar ttop -o CPU -n 10 -p <pid>

```
2018-12-06T17:41:23.043+0100 Process summary
process cpu=73.70%
application cpu=70.59% (user=60.71% sys=9.88%)
other: cpu=3.11%
thread count: 596
GC time=0.45% (young=0.45%, old=0.00%)
heap allocation rate 111mb/s
safe point rate: 0.7 (events/s) avg. safe point pause: 9.84ms
safe point sync time: 0.02% processing time: 0.68% (wallclock time)
[006072] user=11.28% sys= 0.49% alloc= 52mb/s - jmsContainerOptimTournee-175
[006539] user= 8.46% sys= 1.77% alloc= 4552kb/s - RMI TCP Connection(386)-
[000010] user= 7.95% sys= 0.08% alloc= 1947kb/s - Agent Execution
[000494] user= 3.52% sys= 0.27% alloc= 8735kb/s - [AC]
[000022] user= 3.42% sys= 0.20% alloc= 7187kb/s - [AC]
[000162] user= 2.01% sys= 0.35% alloc= 4427kb/s - [AC]
[000019] user= 1.91% sys= 0.24% alloc= 3463kb/s - [AC]
[000165] user= 1.51% sys= 0.31% alloc= 2895kb/s - [AC]
[000171] user= 1.51% sys= 0.19% alloc= 3207kb/s - [AC]
[000185] user= 1.41% sys= 0.28% alloc= 2306kb/s - [AC]
```

# Mémoire

- Contenu de la mémoire de la JVM : heap dump
  - Freeze potentiellement très long
  - Plusieurs moyen de le déclencher :
    - Sur OOM avec -XX:+HeapDumpOnOutOfMemoryError -XX:HeapDumpPath=...
    - En JMX
    - Avec jmap -dump:format=b,file=<dump.hprof> <pid>
    - Avec jcmt <pid> GC.heap\_dump <dump.hprof>
    - Avec core dump, très rapide à générer mais nécessite gdb et la JVM qui faisait tourner le process
      - gdb --pid=<pid>
      - gcore <coredump.core>
      - detach
      - quit
      - Java < 9 # jmap -dump:format=b,file=<dump.hprof> java <coredump.core>
      - Java > 9 # jhsdb jmap --exe <java> --core <coredump.core> --binaryheap --dumpfile <dump.hprof>
    - Sur JVM ko, coredump et kill la JVM : kill -6
  - De préférence avec la même version du JDK que la cible
  - Attention à ne pas faire swapper ou « oomer » la machine en écrivant dans /dev/shm
  - Parfois impossible à lire  
  - À analyser avec Eclipse MAT

# Mémoire

- Histogramme du contenu des objets de la heap
  - jmap -histo:live <pid> ><un\_fichier.log>
  - jcnd <pid> GC.class\_histogram ><un\_fichier.log>  
necessite -XX:+UnlockDiagnosticVMOptions sauf sur JDK récent (testé avec 11)
- Outil de diff : <https://github.com/phraktle/histodiff>

| #   | Instances | Bytes     | Type                                              |
|-----|-----------|-----------|---------------------------------------------------|
| 1:  | 5384166   | 862913328 | [C                                                |
| 2:  | 213638    | 209679912 | [B                                                |
| 3:  | 917264    | 131467096 | <constMethodKlass>                                |
| 4:  | 5371053   | 128905272 | java.lang.String                                  |
| 5:  | 917264    | 117426160 | <methodKlass>                                     |
| 6:  | 105681    | 101191072 | <constantPoolKlass>                               |
| 7:  | 2716859   | 86939488  | java.util.HashMap\$Entry                          |
| 8:  | 105646    | 74774456  | <instanceKlassKlass>                              |
| 9:  | 88187     | 52630400  | <constantPoolCacheKlass>                          |
| 10: | 558755    | 52174400  | [Ljava.util.HashMap\$Entry;                       |
| 11: | 618804    | 49504320  | java.lang.reflect.Method                          |
| 12: | 1014688   | 48705024  | java.util.HashMap                                 |
| 13: | 688386    | 40057136  | [Ljava.lang.Object;                               |
| 14: | 814689    | 32587560  | java.util.LinkedHashMap\$Entry                    |
| 15: | 1342965   | 32231160  | javax.management.ObjectName\$Property             |
| 16: | 760145    | 24324640  | java.util.concurrent.ConcurrentHashMap\$HashEntry |
| 17: | 40782     | 20786640  | <methodDataKlass>                                 |
| 18: | 475409    | 18993648  | [Ljavax.management.ObjectName\$Property;          |
| 19: | 77346     | 16087968  | weblogic.jms.frontend.FESession                   |
| 20: | 272912    | 15283072  | java.util.LinkedHashMap                           |
| 21: | 635330    | 15247920  | java.util.ArrayList                               |
| 22: | 298947    | 14349456  | com.sun.org.apache.xerces.internal.dom.AttrNSImpl |

| root@: | #: ./jdk- | /bin/jcnd 2403 GC.class_histogram                                                          |
|--------|-----------|--------------------------------------------------------------------------------------------|
| 2403:  | num       | #instances #bytes class name (module)                                                      |
|        | 1:        | 2623 128048 [B (java.base@11.0.1)                                                          |
|        | 2:        | 739 91136 java.lang.Class (java.base@11.0.1)                                               |
|        | 3:        | 1014 70264 [Ljava.lang.Object; (java.base@11.0.1)                                          |
|        | 4:        | 2494 59856 java.lang.String (java.base@11.0.1)                                             |
|        | 5:        | 1695 54240 java.util.HashMap\$Node (java.base@11.0.1)                                      |
|        | 6:        | 10 33648 [L (java.base@11.0.1)                                                             |
|        | 7:        | 1027 32864 java.util.concurrent.ConcurrentHashMap\$Node (java.base@11.0.1)                 |
|        | 8:        | 298 31792 [Ljava.util.HashMap\$Node; (java.base@11.0.1)                                    |
|        | 9:        | 312 14976 java.util.HashMap (java.base@11.0.1)                                             |
|        | 10:       | 23 14832 [Ljava.util.concurrent.ConcurrentHashMap\$Node; (java.base@11.0.1)                |
|        | 11:       | 354 8496 java.lang.module.ModuleDescriptor\$Exports (java.base@11.0.1)                     |
|        | 12:       | 172 6984 [I (java.base@11.0.1)                                                             |
|        | 13:       | 278 6672 java.util.ImmutableCollections\$Set12 (java.base@11.0.1)                          |
|        | 14:       | 17 6392 java.lang.Thread (java.base@11.0.1)                                                |
|        | 15:       | 69 4880 java.net.URI (java.base@11.0.1)                                                    |
|        | 16:       | 132 4224 java.lang.module.ModuleDescriptor\$Requires (java.base@11.0.1)                    |
|        | 17:       | 282 4192 java.lang.Integer (java.base@11.0.1)                                              |
|        | 18:       | 164 3936 java.util.ImmutableCollections\$SetN (java.base@11.0.1)                           |
|        | 19:       | 61 3994 java.net.URL (java.base@11.0.1)                                                    |
|        | 20:       | 60 3840 java.lang.module.ModuleDescriptor (java.base@11.0.1)                               |
|        | 21:       | 66 3696 java.lang.module (java.base@11.0.1)                                                |
|        | 22:       | 77 3696 java.lang.invoke.MemberName (java.base@11.0.1)                                     |
|        | 23:       | 121 3600 [Ljava.lang.Class; (java.base@11.0.1)                                             |
|        | 24:       | 60 3368 jdk.internal.module.ModuleReferenceImpl (java.base@11.0.1)                         |
|        | 25:       | 188 3088 java.util.HashSet (java.base@11.0.1)                                              |
|        | 26:       | 44 2992 [Ljava.lang.ref.SoftReference; (java.base@11.0.1)                                  |
|        | 27:       | 33 2964 java.lang.reflect.Method (java.base@11.0.1)                                        |
|        | 28:       | 11 2488 [Ljava.lang.invoke.MethodHandle; (java.base@11.0.1)                                |
|        | 29:       | 73 2336 java.lang.invoke.LambdaFormName (java.base@11.0.1)                                 |
|        | 30:       | 66 2112 java.lang.invoke.LambdaFormKind (java.base@11.0.1)                                 |
|        | 31:       | 33 2112 java.util.concurrent.ConcurrentHashMap (java.base@11.0.1)                          |
|        | 32:       | 49 1968 java.lang.invoke.MethodType (java.base@11.0.1)                                     |
|        | 33:       | 121 1936 java.util.Collections\$UnmodifiableSet (java.base@11.0.1)                         |
|        | 34:       | 57 1824 java.lang.invoke.MethodType\$ConcurrentWeakInternSet\$weakEntry (java.base@11.0.1) |

# Mémoire

- Mémoire totale d'une JVM
  - Heap (Xmx), metaspace
  - Code cache
  - Thread
  - GC internals
  - Symbol et internals
  - Direct buffer et mapped file
- Diagnostic mémoire plus fin via le NativeMemoryTracking
  - -XX:NativeMemoryTracking=summary ou -XX:NativeMemoryTracking=detail
  - jcmd <pid> VM.native\_memory summary
  - Détection leak : jcmd <pid> VM.native\_memory baseline et summary.diff
  - Overhead de perf de 5 à 10%
  - Possible de rapprocher ces résultats (pour le détail) avec le contenu du fichier /proc/[pid]/smaps qui liste les pages mémoires allouées par le process

#### Native Memory Tracking:

```
Total: reserved=2470622KB, committed=232554KB
 Java Heap (reserved=1048576KB, committed=159744KB)
 (mmap: reserved=1048576KB, committed=159744KB)

 Class (reserved=1056899KB, committed=5763KB)
 (classes #730)
 (instance classes #637, array classes #93)
 (malloc=131KB #817)
 (mmap: reserved=1056768KB, committed=5632KB)
 (Metadata:)
 (reserved=8192KB, committed=5120KB)
 (used=4996KB)
 (free=124KB)
 (waste=0KB = 0,00%)
 (Class space:)
 (reserved=1048576KB, committed=512KB)
 (used=453KB)
 (free=59KB)
 (waste=0KB = 0,00%)

 Thread (reserved=26822KB, committed=1846KB)
 (thread #26)
 (stack: reserved=26700KB, committed=1724KB)
 (malloc=91KB #129)
 (arena=30KB #50)

 Code (reserved=247732KB, committed=7593KB)
 (malloc=45KB #528)
 (mmap: reserved=247686KB, committed=7548KB)

 GC (reserved=87028KB, committed=54044KB)
 (malloc=15080KB #1423)
 (mmap: reserved=71948KB, committed=38964KB)

 Compiler (reserved=262KB, committed=262KB)
 (malloc=131KB #92)
 (arena=131KB #5)

 Internal (reserved=583KB, committed=583KB)
 (malloc=543KB #1233)
 (mmap: reserved=40KB, committed=40KB)

 Symbol (reserved=2122KB, committed=2122KB)
 (malloc=1283KB #3166)
 (arena=839KB #1)

 Native Memory Tracking (reserved=151KB, committed=151KB)
 (malloc=6KB #80)
 (tracking overhead=144KB)

 Arena Chunk (reserved=367KB, committed=367KB)
 (malloc=367KB)

 Logging (reserved=4KB, committed=4KB)
 (malloc=4KB #179)

 Arguments (reserved=17KB, committed=17KB)
 (malloc=17KB #467)

 Module (reserved=58KB, committed=58KB)
 (malloc=58KB #1025)
```

#### Virtual memory map:

```
[0x00000000c0000000 - 0x0000000100000000] reserved 1048576KB for Java Heap from
 [0x000007f9a31a070fb] ReservedHeapSpace::try_reserve_heap(unsigned long, unsigned long, bool, char*)+0xcb
 [0x000007f9a31a07fd5] ReservedHeapSpace::initialise_compressed_heap(unsigned long, unsigned long, bool)+0x915
 [0x000007f9a31a08142] ReservedHeapSpace::ReservedHeapSpace(unsigned long, unsigned long, bool, char const*)+0xa2
 [0x000007f9a319c8efd] Universe::reserve_heap(unsigned long, unsigned long)+0xd4

[0x00000000c0000000 - 0x00000000c9c00000] committed 159744KB from
 [0x000007f9a3139771f] G1PageBasedVirtualSpace::commit_preferred_pages(unsigned long, unsigned long)+0x5f
 [0x000007f9a3139798e] G1PageBasedVirtualSpace::commit(unsigned long, unsigned long)+0x5e
 [0x000007f9a313a3760] G1RegionsLargerThanCommitSizeMapper::commit_regions(unsigned int, unsigned long, WorkGang*)+0x40
 [0x000007f9a3141ea19] HeapRegionManager::commit_regions(unsigned int, unsigned long, WorkGang*)+0x89

[0x0000000100000000 - 0x0000000140000000] reserved 1048576KB for Class from
 [0x000007f9a31a06dc2] ReservedSpace::ReservedSpace(unsigned long, unsigned long, bool, char*)+0x82
 [0x000007f9a31757689] Metaspace::allocate_metaspace_compressed_klass_ptrs(char*, unsigned char*)+0x39
 [0x000007f9a31757cd5] Metaspace::global_initialise()+0x75
 [0x000007f9a319c8b72] universe_init()+0x72

[0x0000000100000000 - 0x0000000100060000] committed 384KB from
 [0x000007f9a31a05c47] VirtualSpace::expand_by(unsigned long, bool)+0x197
 [0x000007f9a31a04b1f] metaspaces::VirtualSpaceNode::expand_by(unsigned long, unsigned long)+0x5f
 [0x000007f9a31a03cba] metaspaces::VirtualSpaceList::get_new_chunk(unsigned long, unsigned long)+0x12a
 [0x000007f9a31756450] ClassLoaderMetaspace::ClassLoaderMetaspace(Mutex*, Metaspace::MetaspaceType)+0x1c0

[0x0000000100060000 - 0x0000000100080000] committed 128KB from
 [0x000007f9a31a05c47] VirtualSpace::expand_by(unsigned long, bool)+0x197
 [0x000007f9a31a04b1f] metaspaces::VirtualSpaceNode::expand_by(unsigned long, unsigned long)+0x5f
 [0x000007f9a31a03cba] metaspaces::VirtualSpaceList::get_new_chunk(unsigned long, unsigned long)+0x12a
 [0x000007f9a318f0ef9] metaspaces::SpaceManager::grow_and_allocate(unsigned long)+0x2e9
```

#### Details:

```
[0x000007f9a313581dc] G1CollectedHeap::G1CollectedHeap(G1CollectorPolicy*)+0x81c
[0x000007f9a31348dc9] G1Arguments::create_heap()+0x69
[0x000007f9a319c8882] Universe::initialise_heap()+0x22
[0x000007f9a319c8b46] universe_init()
 (malloc=7168KB type=GC #7)

[0x000007f9a3136c471] G1ConcurrentMark::G1ConcurrentMark(G1CollectedHeap*, G1RegionToSpaceMapper*, G1RegionToSpaceMapper*)+0x7e1
[0x000007f9a3135c99a] G1CollectedHeap::initialise()+0x65a
[0x000007f9a319c8892] Universe::initialise_heap()+0x32
[0x000007f9a319c8b46] universe_init()
 (malloc=7168KB type=GC #7)

[0x000007f9a319069e3] StringTable::StringTable()+0x1b3
[0x000007f9a319c8e0e] universe_init()+0x30e
[0x000007f9a31437cf7] init_globals()+0x57
[0x000007f9a319a7da7] Threads::create_vm(JavaVMInitArgs*, bool*)+0x327
 (malloc=512KB type=Symbol #1)
```

# Metaspace en détail

- capacity (used + free + waste + overhead)
- Souci si beaucoup de deallocated
- Virtual space : bien du virtuel et pas du resident
- 1GB de reserved si pas de MaxMetapaceSize

```
./opt/jdk-11.0.2+9/bin# ./jcmd $(pidof java) VM.metaspase
3134: -
```

Total Usage ( 1119 loaders):

|            |                        |                            |                                         |                       |
|------------|------------------------|----------------------------|-----------------------------------------|-----------------------|
| Non-Class: | 2900 chunks,           | 55,19 MB capacity,         | 54,37 MB ( 99%) used,                   | 658,30 KB ( 1%) free, |
|            | 3,24 KB ( <1%) waste,  | 181,25 KB ( <1%) overhead, | deallocated: 1017 blocks with 233,62 KB |                       |
| Class:     | 1346 chunks,           | 8,18 MB capacity,          | 7,66 MB ( 94%) used,                    | 447,95 KB ( 5%) free, |
|            | 88 bytes ( <1%) waste, | 84,12 KB ( 1%) overhead,   | deallocated: 209 blocks with 53,95 KB   |                       |
| Both:      | 4246 chunks,           | 63,37 MB capacity,         | 62,03 MB ( 98%) used,                   | 1,08 MB ( 2%) free,   |
|            | 3,33 KB ( <1%) waste,  | 265,38 KB ( <1%) overhead, | deallocated: 1226 blocks with 287,57 KB |                       |

Virtual space:

|                  |                     |                           |
|------------------|---------------------|---------------------------|
| Non-class space: | 56,00 MB reserved,  | 55,62 MB (>99%) committed |
| Class space:     | 56,00 MB reserved,  | 8,38 MB ( 15%) committed  |
| Both:            | 112,00 MB reserved, | 64,00 MB ( 57%) committed |

Chunk freelist:

Non-Class:

|                     |                         |
|---------------------|-------------------------|
| specialized chunks: | 55, capacity 55,00 KB   |
| small chunks:       | 85, capacity 340,00 KB  |
| medium chunks:      | (none)                  |
| humongous chunks:   | (none)                  |
| Total:              | 140, capacity=395,00 KB |

Class:

|                     |                         |
|---------------------|-------------------------|
| specialized chunks: | 25, capacity 25,00 KB   |
| small chunks:       | 85, capacity 170,00 KB  |
| medium chunks:      | (none)                  |
| humongous chunks:   | (none)                  |
| Total:              | 110, capacity=195,00 KB |

Waste (percentages refer to total committed size 64,00 MB):

|                                 |                                |
|---------------------------------|--------------------------------|
| Committed unused:               | 52,00 KB ( <1%)                |
| Waste in chunks in use:         | 3,33 KB ( <1%)                 |
| Free in chunks in use:          | 1,08 MB ( 2%)                  |
| Overhead in chunks in use:      | 265,38 KB ( <1%)               |
| In free chunks:                 | 590,00 KB ( <1%)               |
| Deallocated from chunks in use: | 287,57 KB ( <1%) (1226 blocks) |
| -total-:                        | 2,25 MB ( 4%)                  |

MaxMetapaceSize: 64,00 MB  
InitialBootClassLoaderMetaspaseSize: 4,00 MB  
UseCompressedClassPointers: true  
CompressedClassSpaceSize: 56,00 MB

# Profiling mémoire native : jemalloc

- Utile pour avoir une vue sur les chemins d'allocation mémoire

Installer graphviz

```
export LD_PRELOAD=/usr/local/lib/libjemalloc.so
```

```
export MALLOC_CONF=prof:true,lg_prof_interval:30,lg_prof_sample:17
```

```
java -jar ...
```

```
jeprof --show_bytes --svg /path/to/java jeprof*.heap > memory-profiling.svg
```

- % de mémoire référencée par la JVM : osmalloc@...

os:malloc doit être à 98-99%

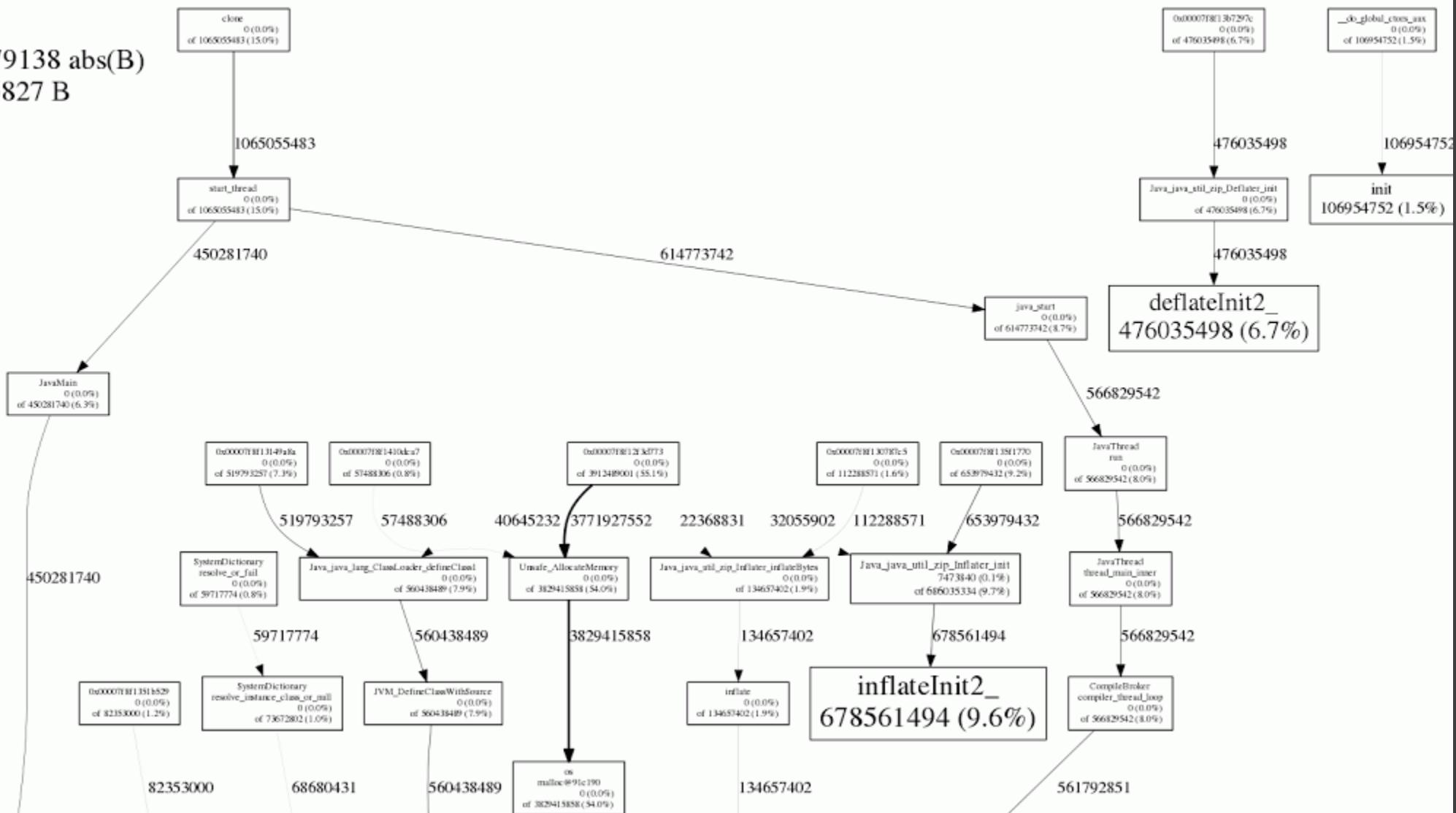
/usr/lib/jvm/java-8-oracle/jre/bin/java

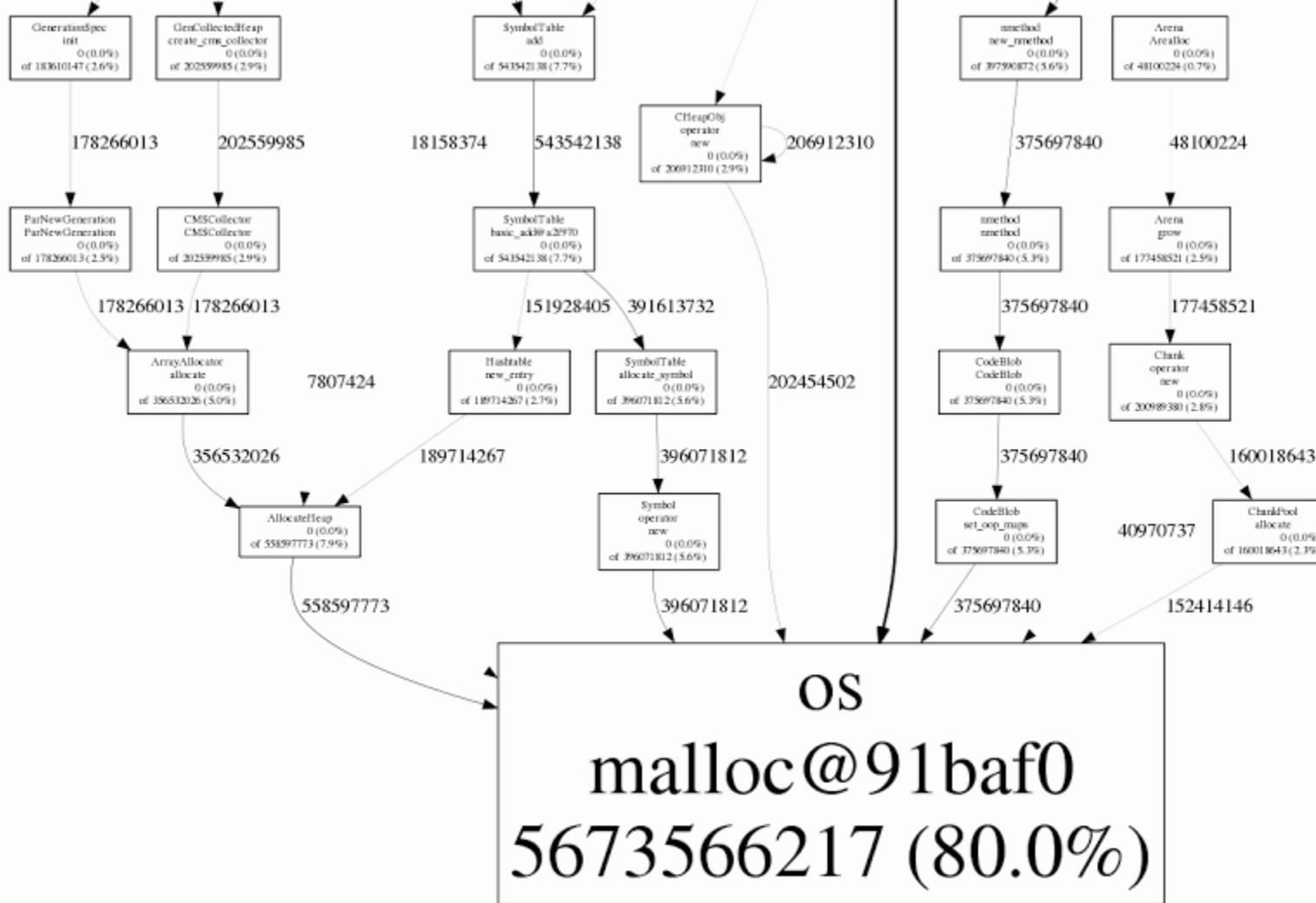
Total B: 7095827787

Focusing on: 7095827787

Dropped nodes with <= 35479138 abs(B)

Dropped edges with <= 7095827 B





# Garbage collector

- Logs GC à activer partout, y compris en production
  - JDK8: -XX:+PrintGCCause -XX:+PrintGCDetails -XX:+PrintTenuringDistribution -XX:+PrintGCDateStamps -XX:+PrintGCTimeStamps -XX:+UseGCLogFileRotation -XX:NumberOfGCLogFiles=10 -XX:GCLogFileSize=10M -XX:+PrintGCAapplicationStoppedTime -XX:+PrintGCAapplicationConcurrentTime -Xloggc:<un\_fichier>
  - JDK9+ : -Xlog:gc\*,age\*=trace,safepoint:/app/log/gc-%p-%t.log:t,level,tags:filecount=10,filesize=10M
- Quelques évidences visuelles ou greppable comme le « failure » de « (concurrent mode failure) » ou le « failed » de « (promotion failed) » du CMS
- Sinon via un outil comme gceasy.io, GC Log Viewer ou Censum (commercial ... Microsoft)
- Via jstat -gc <pid> 2s

```
[.. ~]$.. /jstat -gc 21106 2s
 SOC S1C SOU S1U EC EU OC OU PC PU YGC YGCT FGC FGCT GCT
104832.0 104832.0 42296.0 0.0 838912.0 203913.3 2097152.0 1580143.3 524288.0 496037.9 6460 220.414 121 43.031 263.444
104832.0 104832.0 42296.0 0.0 838912.0 361353.2 2097152.0 1580143.3 524288.0 496037.9 6460 220.414 121 43.031 263.444
104832.0 104832.0 42296.0 0.0 838912.0 419344.1 2097152.0 1580143.3 524288.0 496037.9 6460 220.414 121 43.031 263.444
104832.0 104832.0 42296.0 0.0 838912.0 504507.3 2097152.0 1580143.3 524288.0 496037.9 6460 220.414 121 43.031 263.444
104832.0 104832.0 42296.0 0.0 838912.0 545860.2 2097152.0 1580143.3 524288.0 496037.9 6460 220.414 121 43.031 263.444
```

# Garbage collector

- Activable et modifiable au runtime Java > 9
- jcfd <pid> VM.log output="file=gc-%p-%t.log"  
what="gc=debug,gc\*,age\*=trace,safepoint" decorators="time,level"  
output\_options="filecount=10,filesize=10M"
- Via jcfd <pid> PerfCounter.print | egrep 'safepointTime|applicationTime'
- % de temps dans des safepoint, pour la plupart du GC :  
 $100 * \text{safepointTime} / (\text{safepointTime} + \text{applicationTime})$

# Garbage collector

- Sujet complexe
- Phases du GC
  - Durée et fréquence
  - Bloquant ou pas
  - Concurrent ou pas
  - Multithreadé
- Points d'attention
  - Taille du Live data set, des espace (Young : eden + survivor), Old, Perm | Metaspace
  - Allocation rate et taille récupérée
  - Premature promotion
  - Cause des GC
  - ...
- Souci si throughput applicatif < 90 % (GC prend plus de 10%)

- jcmande
- Déjà vu pour le GC, NMT ...
- Beaucoup d'autres outils :
  - Détail sur le JIT
  - Classloader et hiérarchie
  - Activation de flags
  - VM.info : beaucoup d'informations en une commande
  - Flight Recorder
  - Démarrage JMX : jcmande <pid> ManagementAgent.start config.file=<path.to.file>
  - ...

# JMX

- Le plus simple, un seul fichier de configuration

```
cat jmxremote.properties
com.sun.management.jmxremote=true
com.sun.management.jmxremote.port=<port>
com.sun.management.jmxremote.ssl=true
com.sun.management.jmxremote.rmi.port=<port>
com.sun.management.jmxremote.password.file=jmxremote.password
com.sun.management.jmxremote.access.file=jmxremote.access
com.sun.management.jmxremote.registry.ssl=true
com.sun.management.jmxremote.ssl.enabled.protocols=TLSv1.2
com.sun.management.jmxremote.ssl.enabled.cipher.suites=TLS_ECDHE_ECDSA_WITH_AES_256_GCM_SHA384
com.sun.management.jmxremote.ssl.config.file=jmxremote.properties

javax.net.ssl.keyStore=jmxremote.jks
javax.net.ssl.keyStorePassword=tls pass
javax.net.ssl.trustStore=truststore.jks
javax.net.ssl.trustStorePassword=tls pass
```

# Quelles options sur ma JVM ?

- Listes des options : défauts, surcharge et ergonomics
  - java -XX:+PrintFlagsFinal -XX:+UnlockDiagnosticVMOptions -version
  - | grep ':=' valeurs modifiés manuellement ou par les ergonomics
  - jcmd <pid> VM.flags –all
    - Mais n'affiche pas tout ...
  - Depuis le JDK11, c'est encore plus clair, on sait si c'est la « command line », les « ergonomics », ...

```
intx CICompilerCount := 12 {product}
uintx InitialHeapSize := 2147483648 {product}
uintx MaxHeapSize := 2147483648 {product}
uintx MaxMetaspaceSize := 268435456 {product}
uintx MaxNewSize := 1073741824 {product}
uintx MinHeapDeltaBytes := 524288 {product}
uintx NewSize := 1073741824 {product}
uintx OldSize := 1073741824 {product}
bool PrintFlagsFinal := true {diagnostic}
bool UnlockDiagnosticVMOptions := true {product}
bool UseCompressedClassPointers := true {[lp64_product]}
bool UseCompressedOoops := true {[lp64_product]}
bool UseParallelOldGC := true {product}
' "1.8.0_111"
```

```
[Global flags]
 intx CICompilerCount = 3 {product} {ergonomic}
 intx CompressedClassSpaceSize = 260046848 {product} {ergonomic}
 size_t InitialHeapSize = 2147483648 {product} {command line}
 size_t MaxHeapSize = 2147483648 {product} {command line}
 size_t MaxMetaspaceSize = 268435456 {product} {command line}
 size_t MaxNewSize = 1073741824 {product} {command line}
 size_t MinHeapDeltaBytes = 524288 {product} {ergonomic}
 size_t NewSize = 1073741824 {product} {command line}
 uintx NonNMethodCodeHeapSize = 5830092 {pd product} {ergonomic}
 uintx NonProfiledCodeHeapSize = 122914074 {pd product} {ergonomic}
 size_t OldSize = 1073741824 {product} {ergonomic}
 bool PrintFlagsFinal = true {product} {command line}
 uintx ProfiledCodeHeapSize = 122914074 {pd product} {ergonomic}
 uintx ReservedCodeCacheSize = 251658240 {pd product} {ergonomic}
 bool SegmentedCodeCache = true {product} {ergonomic}
 bool UnlockDiagnosticVMOptions = true {diagnostic} {command line}
 bool UseCompressedClassPointers = true {[lp64_product]} {ergonomic}
 bool UseCompressedOoops = true {[lp64_product]} {command line}
 bool UseParallelOldGC = true {product} {command line}
```

- Plusieurs sites avec la liste complète par version ...  
[https://chriswhocodes.com/hotspot\\_options\\_jdk11.html](https://chriswhocodes.com/hotspot_options_jdk11.html)

- Beaucoup d'outils s'appuie sur le fichier hsperfdata
- Attention au flag systemd PrivateTmp avec pour conséquence une localisation du hsperfdata exotique :  
`/tmp/systemd-private-xxx-elasticsearch.service-xxxx/tmp/hsperfdata_elasticsearch/`

# Diagnostic avancé

- Arthas
- Opensourcé par Alibaba, outil de diagnostic, attachement live (nécessite telnet)
- <http://repository.sonatype.org/service/local/artifact/maven/redirect?r=central-proxy&g=com.taobao.arthas&a=arthas-packaging&e=zip&c=bin&v=LATEST./as.sh>
- Certains nombres de commandes accélérant le diagnostic

**Arthas**

## • Dashboard

| ID                     | NAME                                             | GROUP                           | PRIORITY | STATE         | %CPU                             | TIME  | INTERRUPTED | DAEMON |
|------------------------|--------------------------------------------------|---------------------------------|----------|---------------|----------------------------------|-------|-------------|--------|
| 4812                   | Timer-for-arthas-dashboard-d6d432c2-002f-43fc-ac | system                          | 10       | RUNNABLE      | 68                               | 0:0   | false       | true   |
| 496                    | HiccupRecorder                                   | Pooled Threads                  | 5        | TIMED_WAITING | 5                                | 1:44  | false       | true   |
| 12                     | Intelligent Instrumentation Thread               | main                            | 10       | TIMED_WAITING | 4                                | 1:53  | false       | true   |
| 388                    | kafka-poller-consumer-cmd-executor-0-processing- | Pooled Threads                  | 5        | RUNNABLE      | 3                                | 0:19  | false       | false  |
| 397                    | kafka-poller-consumer-cmd-executor-1-processing- | Pooled Threads                  | 5        | RUNNABLE      | 3                                | 0:21  | false       | false  |
| 417                    | kafka-poller-consumer-cmd-executor-3-processing- | Pooled Threads                  | 5        | RUNNABLE      | 2                                | 0:25  | false       | false  |
| 460                    | kafka-poller-consumer-indexation-es-pdk-0-proces | Pooled Threads                  | 5        | RUNNABLE      | 2                                | 0:11  | false       | false  |
| 423                    | kafka-coordinator-heartbeat-thread               | Pooled Threads                  | 5        | TIMED_WAITING | 1                                | 0:11  | false       | true   |
| 442                    | kafka-poller-consumer-action-executor-2-processi | Pooled Threads                  | 5        | RUNNABLE      | 1                                | 0:11  | false       | false  |
| 95                     | Abandoned connection cleanup thread              | Pooled Threads                  | 5        | WAITING       | 0                                | 0:0   | false       | true   |
| 14                     | Agent Execution                                  | Agent                           | 5        | WAITING       | 0                                | 2:28  | false       | true   |
| 8                      | Agent Heartbeat                                  | Agent                           | 5        | TIMED_WAITING | 0                                | 0:8   | false       | true   |
| 137                    | Agent ServerConnection                           | Agent                           | 5        | TIMED_WAITING | 0                                | 0:0   | false       | true   |
| 4789                   | AsyncAppender-Worker-arthas-cache.result.AsyncAp | system                          | 9        | WAITING       | 0                                | 0:0   | false       | true   |
| 4589                   | Attach Listener                                  | system                          | 9        | RUNNABLE      | 0                                | 0:0   | false       | true   |
| 482                    | CmdExecutorQuartzScheduler                       | QuartzScheduler:cmdExecutorQuar | 5        | TIMED_WAITING | 0                                | 0:1   | false       | false  |
| 136                    | Command Queue Heartbeat                          | Agent                           | 5        | TIMED_WAITING | 0                                | 0:1   | false       | true   |
| 10                     | Configuration Watch Heartbeat                    | Heartbeat                       | 5        | TIMED_WAITING | 0                                | 0:0   | false       | true   |
| 85                     | DoManager                                        | Pooled Threads                  | 6        | TIMED_WAITING | 0                                | 0:0   | false       | true   |
| 65                     | F                                                | M Thread Group for Queue        | 5        | RUNNABLE      | 0                                | 0:13  | false       | true   |
| 66                     | F                                                | M Thread Group for Queue        | 5        | RUNNABLE      | 0                                | 0:10  | false       | true   |
| 67                     | F                                                | M Thread Group for Queue        | 5        | RUNNABLE      | 0                                | 0:17  | false       | true   |
| 68                     | L                                                | M Thread Group for Queue        | 5        | RUNNABLE      | 0                                | 0:9   | false       | true   |
| 3                      | Finalizer                                        | system                          | 8        | WAITING       | 0                                | 0:1   | false       | true   |
| 257                    | GC Daemon                                        | system                          | 2        | TIMED_WAITING | 0                                | 0:0   | false       | true   |
| <hr/>                  |                                                  |                                 |          |               |                                  |       |             |        |
| Memory                 |                                                  |                                 |          |               |                                  |       |             |        |
|                        | used                                             | total                           | max      | usage         | GC                               |       |             |        |
| heap                   | 4369M                                            | 5120M                           | 5120M    | 85,35%        | gc.gc1_young_generation.count    | 921   |             |        |
| g1_eden_space          | 2058M                                            | 2232M                           | -1       | 92,20%        | gc.gc1_young_generation.time(ms) | 94803 |             |        |
| g1_survivor_space      | 112M                                             | 112M                            | -1       | 100,00%       | gc.gc1_old_generation.count      | 0     |             |        |
| g1_old_gen             | 2199M                                            | 2776M                           | 5120M    | 42,96%        | gc.gc1_old_generation.time(ms)   | 0     |             |        |
| nonheap                | 1069M                                            | 1135M                           | 2288M    | 46,76%        |                                  |       |             |        |
| code_cache             | 185M                                             | 223M                            | 240M     | 77,15%        |                                  |       |             |        |
| metaspace              | 780M                                             | 803M                            | 1024M    | 76,25%        |                                  |       |             |        |
| compressed_class_space | 103M                                             | 109M                            | 1024M    | 10,15%        |                                  |       |             |        |
| direct                 | 466M                                             | 466M                            | -        | 100,00%       |                                  |       |             |        |
| mapped                 | 8M                                               | 8M                              | -        | 100,00%       |                                  |       |             |        |
| <hr/>                  |                                                  |                                 |          |               |                                  |       |             |        |
| Runtime                |                                                  |                                 |          |               |                                  |       |             |        |
| os.name                | Linux                                            |                                 |          |               |                                  |       |             |        |
| os.version             | 5.10.0-1024.1.2.el8_6.x86_64                     |                                 |          |               |                                  |       |             |        |
| java.version           | 1                                                |                                 |          |               |                                  |       |             |        |
| java.home              | /logiciels/java/j                                |                                 |          |               |                                  |       |             |        |
| systemload.average     | 0.06                                             |                                 |          |               |                                  |       |             |        |
| processors             | 8                                                |                                 |          |               |                                  |       |             |        |
| uptime                 | 10381s                                           |                                 |          |               |                                  |       |             |        |

- Classloader statistics

| \$ classloader | name                                                       | numberOfInstances | loadedCountTotal |
|----------------|------------------------------------------------------------|-------------------|------------------|
|                | weblogic.utils.classloaders.ChangeAwareClassLoader         | 7                 | 118764           |
|                | com.oracle.classloader.weblogic.LaunchClassLoader          | 1                 | 20419            |
|                | BootstrapClassLoader                                       | 1                 | 8167             |
|                | sun.reflect.DelegatingClassLoader                          | 6274              | 6274             |
|                | com.taobao.artdas.agent.ArthasClassLoader                  | 2                 | 2069             |
|                |                                                            | 29                | 736              |
|                |                                                            | 7                 | 714              |
|                | org.apache.cxf.common.utilASMHelper\$TypeHelperClassLoader | 50                | 399              |
|                | java.net.URLClassLoader                                    | 4                 | 153              |
|                | sun.misc.Launcher\$ExtClassLoader                          | 1                 | 71               |
|                | sun.misc.Launcher\$AppClassLoader                          | 1                 | 68               |
|                | weblogic.utils.classloaders.GenericClassLoader             | 17                | 26               |
|                |                                                            | 5                 | 10               |
|                | org.eclipse.persistence.internal.jaxb.JAXBClassLoader      | 6                 | 6                |
|                | org.jvnet.hk2.internal.DelegatingClassLoader               | 1                 | 3                |
|                | javax.management.remote.rmi.NoCallStackClassLoader         | 2                 | 2                |
|                | weblogic.servlet.jsp.JspClassLoader                        | 1                 | 1                |
|                | weblogic.servlet.jsp.TagFileClassLoader                    | 1                 | 0                |

- Décompilation à la volée : jad

```

import java.util.Collection;
import java.util.Collections;
import org.springframework.beans.factory.annotation.Autowired;
import org.springframework.beans.factory.annotation.Qualifier;
import org.springframework.beans.factory.annotation.Value;
import org.springframework.security.core.GrantedAuthority;
import org.springframework.stereotype.Component;

@Component
public class CmdExecutor
extends AbstractKafkaMessageListener {
 private static final LoggerAudit LOGGER_AUDIT = new ChainedLogAudit();
 private static final Log4ERDFLogger LOGGER_TECH;
 private static final String COMMANDS_TIMER_NAME = "commands-timer";
 private static final String COMMANDS_COMMITKAFKA_TIMER_NAME = "commands.commit-kafka-timer";
 private static final String COMMAND_IDEMPOTENCE_TIMER_NAME = "command.idempotence-timer";
 @Autowired
 @Qualifier(value="cmdExecutorKafkaListenerConfiguration")
 private KafkaConsumerConfig kafkaConsumerConfig;
 @Value(value="${sup.kafka.cmd-executor.thread.count}")
 private int numberOfProcessingThreads;
 @Value(value="${sup.kafka.cmd-executor.consumers.count}")
 private int numberOfWorkers;
 @Autowired
 private ICommandExecutorService commandExecutorService;
 @Autowired
 private IIidempotenceService idempotenceService;
 public Collection<? extends GrantedAuthority> getAuthorities() {
 return Collections.emptyList();
 }
}

```

- Analyse des threads facilitée : thread

| ID   | NAME                                           | GROUP          | PRIORITY | STATE         | %CPU | TIME | INTERRUPTED | DAEMON |
|------|------------------------------------------------|----------------|----------|---------------|------|------|-------------|--------|
| 4708 | as-command-execute-daemon                      | system         | 10       | RUNNABLE      | 21   | 0:0  | false       | true   |
| 322  | Timer-9                                        | Pooled Threads | 5        | TIMED_WAITING | 10   | 0:0  | false       | true   |
| 389  | kafka-poller-consumer-cmd-executor-0-processin | Pooled Threads | 5        | RUNNABLE      | 4    | 0:20 | false       | false  |
| 402  | kafka-poller-consumer-cmd-executor-1-processin | Pooled Threads | 5        | RUNNABLE      | 4    | 0:24 | false       | false  |
| 405  | kafka-poller-consumer-cmd-executor-2-processin | Pooled Threads | 5        | RUNNABLE      | 4    | 0:18 | false       | false  |
| 417  | kafka-poller-consumer-cmd-executor-3-processin | Pooled Threads | 5        | RUNNABLE      | 4    | 0:23 | false       | false  |
| 354  | elasticsearch[_client_][transport_client_boss] | Pooled Threads | 5        | RUNNABLE      | 2    | 0:4  | false       | true   |
| 556  | elasticsearch[_client_][transport_client_boss] | Pooled Threads | 5        | RUNNABLE      | 2    | 0:5  | false       | true   |
| 356  | elasticsearch[_client_][transport_client_boss] | Pooled Threads | 5        | RUNNABLE      | 2    | 0:5  | false       | true   |
| 372  | elasticsearch[_client_][transport_client_boss] | Pooled Threads | 5        | RUNNABLE      | 2    | 0:5  | false       | true   |
| 557  | elasticsearch[_client_][transport_client_boss] | Pooled Threads | 5        | RUNNABLE      | 2    | 0:4  | false       | true   |
| 357  | elasticsearch[_client_][transport_client_boss] | Pooled Threads | 5        | RUNNABLE      | 2    | 0:3  | false       | true   |
| 373  | elasticsearch[_client_][transport_client_boss] | Pooled Threads | 5        | RUNNABLE      | 2    | 0:4  | false       | true   |
| 558  | elasticsearch[_client_][transport_client_boss] | Pooled Threads | 5        | RUNNABLE      | 2    | 0:4  | false       | true   |
| 427  | kafka-poller-consumer-action-executor-0-proces | Pooled Threads | 5        | RUNNABLE      | 2    | 0:10 | false       | false  |
| 435  | kafka-poller-consumer-action-executor-1-proces | Pooled Threads | 5        | RUNNABLE      | 2    | 0:11 | false       | false  |
| 444  | kafka-poller-consumer-action-executor-2-proces | Pooled Threads | 5        | RUNNABLE      | 2    | 0:11 | false       | false  |
| 450  | kafka-poller-consumer-action-executor-3-proces | Pooled Threads | 5        | RUNNABLE      | 2    | 0:10 | false       | false  |
| 496  | HiccupRecorder                                 | Pooled Threads | 5        | TIMED_WAITING | 1    | 1:28 | false       | true   |
| 12   | Intelligent Instrumentation Thread             | main           | 10       | TIMED_WAITING | 1    | 1:37 | false       | true   |
| 358  | elasticsearch[_client_][transport_client_boss] | Pooled Threads | 5        | RUNNABLE      | 1    | 0:3  | false       | true   |
| 487  | kafka-poller-consumer-histo-event-executor-0-p | Pooled Threads | 5        | RUNNABLE      | 1    | 0:12 | false       | false  |
| 489  | kafka-poller-consumer-histo-event-executor-1-p | Pooled Threads | 5        | RUNNABLE      | 1    | 0:12 | false       | false  |
| 465  | kafka-poller-consumer-indexation-es-pdc-0-proc | Pooled Threads | 5        | RUNNABLE      | 1    | 0:10 | false       | false  |
| 95   | Abandoned connection cleanup thread            | Pooled Threads | 5        | WAITING       | 0    | 0:0  | false       | true   |

- Monitoring dynamique de méthode d'une classe : monitor

```
$ monitor -c 2 org.HdrHistogram.SingleWriterRecorder *
Press Ctrl+C to abort.
Affect(class-cnt:1 , method-cnt:11) cost in 32 ms.
 timestamp class method total success fail avg-rt(ms) fail-rate
2018-12-14 17:31:02 org.HdrHistogram.SingleWriterRecorder recordValueWithExpectedInterval 2 2 0 1.22 0.00%
 timestamp class method total success fail avg-rt(ms) fail-rate
2018-12-14 17:31:04 org.HdrHistogram.SingleWriterRecorder performIntervalSample 1 1 0 0.07 0.00%
2018-12-14 17:31:04 org.HdrHistogram.SingleWriterRecorder getIntervalHistogram 1 1 0 0.09 0.00%
2018-12-14 17:31:04 org.HdrHistogram.SingleWriterRecorder recordValueWithExpectedInterval 1738 1738 0 0.02 0.00%
2018-12-14 17:31:04 org.HdrHistogram.SingleWriterRecorder validateFitAsReplacementHistogram 1 1 0 0.04 0.00%
 timestamp class method total success fail avg-rt(ms) fail-rate
2018-12-14 17:31:06 org.HdrHistogram.SingleWriterRecorder performIntervalSample 0 0 0 0.00 0.00%
2018-12-14 17:31:06 org.HdrHistogram.SingleWriterRecorder getIntervalHistogram 0 0 0 0.00 0.00%
2018-12-14 17:31:06 org.HdrHistogram.SingleWriterRecorder recordValueWithExpectedInterval 1796 1796 0 0.01 0.00%
2018-12-14 17:31:06 org.HdrHistogram.SingleWriterRecorder validateFitAsReplacementHistogram 0 0 0 0.00 0.00%
$ reset
Affect(class-cnt:1 , method-cnt:0) cost in 16 ms.
```

- Paramètres d'une méthode : watch

```
$ watch org.HdrHistogram.SingleWriterRecorder recordValueWithExpectedInterval params[0]
Press Ctrl+C to abort.
Affect(class-cnt:1 , method-cnt:1) cost in 24 ms.
ts=2018-12-14 17:33:36; [cost=2.155953ms] result=@Long[67003]
ts=2018-12-14 17:33:36; [cost=0.047445ms] result=@Long[73366]
```

- Valeur de retour

```
$ watch, ... messaging.consumer.kafka.KafkaConsumerImpl getClientId returnObj
Press Ctrl+C to abort.
Affect(class-cnt:2 , method-cnt:2) cost in 1027 ms.
ts=2018-12-14 17:38:12; [cost=3.92359ms] result=@String[cmd-executor-3]ts=2018-12-14 17:38:12; [cost=11.772356ms] result=@String[indexation-es-da-1]ts=2018-12-14 17:38:12; [cost=3.312547ms] result=@String[action-executor-0]ts=2018-12-14 17:38:12; [cost=6.979195ms] result=@String[indexation-es-pdk-0]ts=2018-12-14 17:38:12; [cost=5.859155ms] result=@String[indexation-es-pdk-1]
```

- Les deux

```
$ watch -f -x 2 -n 3 cor... `````` agent.jmx.Filter matches "{params,returnObj}"
Press Ctrl+C to abort.
Affect(class-cnt:1 , method-cnt:2) cost in 581 ms.
ts=2018-12-17 15:47:07; [cost=3.015918ms] result=@ArrayList[
 @Object[] [
 @String[JMX] `````` |JMSConnectionRuntime=connection31010104|JMSRuntime="com.sun.jmx.mbeanserver.JMXRuntime:jms|Name=session31010106|ServerRuntime=l...|User=STM|Type=JMSSessionRuntime:Prod
ucersHighCount],
 @Integer[0],
 @String[WebAppComponentRuntime*:OpenSessionsCurrentCount],
 @Integer[0],
],
 @Boolean[false],
]
]
```

- Filtrage possible sur arguments ('params[0].contains("jms")'), duration ('#cost>1'), ...

```
$ trace com... `````` .agent.jmx.Filter matches 'params[0].contains("jms")'
Press Ctrl+C to abort.
Affect(class-cnt:1 , method-cnt:2) cost in 642 ms.
`---ts=2018-12-17 16:45:16;thread_name=Agent Execution;id=a;is_daemon=true;priority=5;TCCL=sun.misc.Launcher$AppClassLoader@4d29dcc0
 `---[5.988165ms] com... `````` .agent.jmx.Filter:matches()
 +---[0.079993ms] java.lang.String:length()
 `---[2.675884ms] com... `````` .agent.jmx.Filter:matches()
 `---[2.593797ms] com.v... `````` .agent.jmx.Filter:matches()
 +---[min=7.31E-4ms,max=0.00227ms,total=0.155638ms,count=152] java.lang.String:length()
 `---[min=7.23E-4ms,max=0.011516ms,total=0.321399ms,count=308] java.lang.String:charAt()
```

- Qui m'appelle ?

```
$ stack -n 5 org.apache.storm.shade.org.apache.zookeeper.ClientCnxn$SendThread getZkState
Press Ctrl+C to abort.
ts=2018-12-17 17:23:51;thread_name=Thread-10-$mastercoord-bg0-executor[1 1]-SendThread(0.0.0.0:2181->127.0.0.1:2181);id=4f;is_daemon=true;priority=5;TCCL=sun.misc.Launcher$AppClassLoader@764c12b6
 @org.apache.storm.shade.org.apache.zookeeper.ClientCnxnSocketNIO.doTransport()
 at org.apache.storm.shade.org.apache.zookeeper.ClientCnxn$SendThread.run(null:-1)
```

- Enregister et avoir le détail d'un appel

| \$ tt -t -n 5 org.apache.storm.shade.org.apache.zookeeper.ClientCnxn\$SendThread getZkState |                     |          |        |        |            |                        |            |
|---------------------------------------------------------------------------------------------|---------------------|----------|--------|--------|------------|------------------------|------------|
| Press Ctrl+C to abort.                                                                      |                     |          |        |        |            |                        |            |
| Affect(class-cnt:1 , method-cnt:1) cost in 112 ms.                                          |                     |          |        |        |            |                        |            |
| INDEX                                                                                       | TIMESTAMP           | COST(ms) | IS-RET | IS-EXP | OBJECT     | CLASS                  | METHOD     |
| 1001                                                                                        | 2018-12-17 17:26:34 | 0.044672 | true   | false  | 0x1c7e4a4  | ClientCnxn\$SendThread | getZkState |
| 1000                                                                                        | 2018-12-17 17:26:34 | 0.678031 | true   | false  | 0x77dd8f02 | ClientCnxn\$SendThread | getZkState |
| 1003                                                                                        | 2018-12-17 17:26:34 | 0.023868 | true   | false  | 0x1c7e4a4  | ClientCnxn\$SendThread | getZkState |
| 1002                                                                                        | 2018-12-17 17:26:34 | 0.023776 | true   | false  | 0x77dd8f02 | ClientCnxn\$SendThread | getZkState |

```
$ tt -i 1001
INDEX 1001
GMT-CREATE 2018-12-17 17:26:34
COST(ms) 0.044672
OBJECT 0x1c7e4a4
CLASS org.apache.storm.shade.org.apache.zookeeper.ClientCnxn$SendThread
METHOD getZkState
IS-RETURN true
IS-EXCEPTION false
RETURN-OBJ @States[
 CONNECTING=@States [CONNECTING] ,
 ASSOCIATING=@States [ASSOCIATING] ,
 CONNECTED=@States [CONNECTED] ,
 CONNECTEDREADONLY=@States [CONNECTEDREADONLY] ,
 CLOSED=@States [CLOSED] ,
 AUTH_FAILED=@States [AUTH_FAILED] ,
 NOT_CONNECTED=@States [NOT_CONNECTED] ,
 $VALUES=@States [] [isEmpty=false;size=7] ,
]

```

# Crash

- En cas de crash, la JVM génère un fichier `hs_err_pid<pid>.log` (chemin customisable via `-XX:ErrorFile`) qui contient pleins d'informations intéressantes pour le diagnostic
- Doc Oracle :
  - The operating exception or signal that provoked the fatal error
  - Version and configuration information
  - Details on the thread that provoked the fatal error and thread's stack trace
  - The list of running threads and their state
  - Summary information about the heap
  - The list of native libraries loaded
  - Command line arguments
  - Environment variables
  - Details about the operating system and CPU
- Un fichier core peut aussi être généré

```

A fatal error has been detected by the Java Runtime Environment:
#
SIGBUS (0x7) at pc=0x000000382b0896e1, pid=741, tid=0x00007f33e3626700
#
JRE version: Java(TM) SE Runtime Environment () (build
Java VM: Java HotSpot(TM) 64-Bit Server VM (mixed mode linux-amd64 compressed oops)
Problematic frame:
C [libc.so.6 memcpy+0x11
#
Core dump written. Default location: /usr/hdp/ /hadoop-yarn/core or core.741
#
If you would like to submit a bug report, please visit:
http://bugreport.java.com/bugreport/crash.jsp
The crash happened outside the Java Virtual Machine in native code.
See problematic frame for where to report the bug.
#
----- T H R E A D -----
Current thread (0x00007f3413661800): JavaThread "IPC Server handler 8 on 45454" daemon [_thread_in_native, id=1039, stack(0x00007f33e3526000,0x00007f33e3627000)]
siginfo: si_signo: 7 (SIGBUS), si_code: 2 (BUS_ADRERR), si_addr: 0x00007f33e4c56000

Registers:
RAX=0x00007f33e4c56000, RBX=0x00007f341334a4f0, RCX=0x000000000000
RSP=0x00007f33e3623e38, RBP=0x0000000000000007, RSI=0x00007f33e38c
R8 =0x00000000000000181, R9 =0x0000000000000000, R10=0x000000000000
R12=0x0000000000000007, R13=0x00007f33e3623eb0, R14=0x00007f33e3d
RIP=0x000000382b0896e1, EFLAGS=0x000000000010202, CSFS=0x0000
TRAPNO=0x000000000000000e

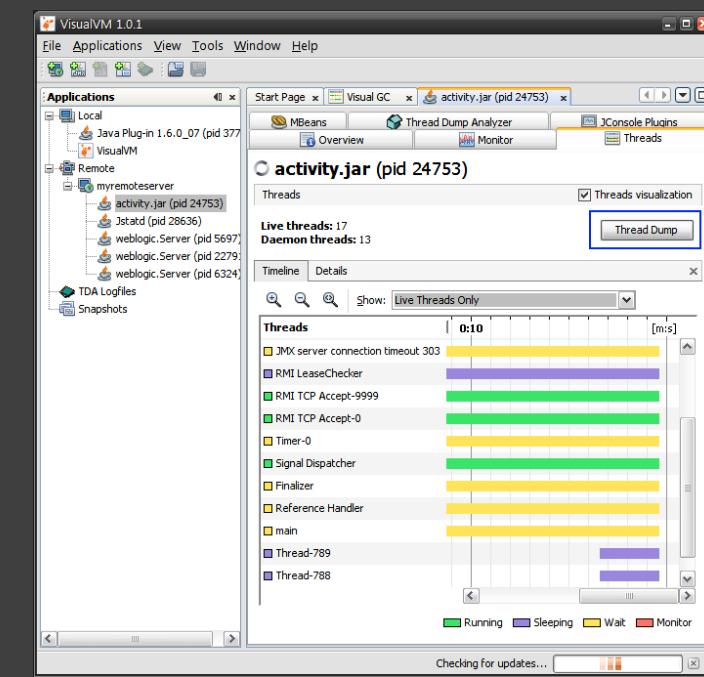
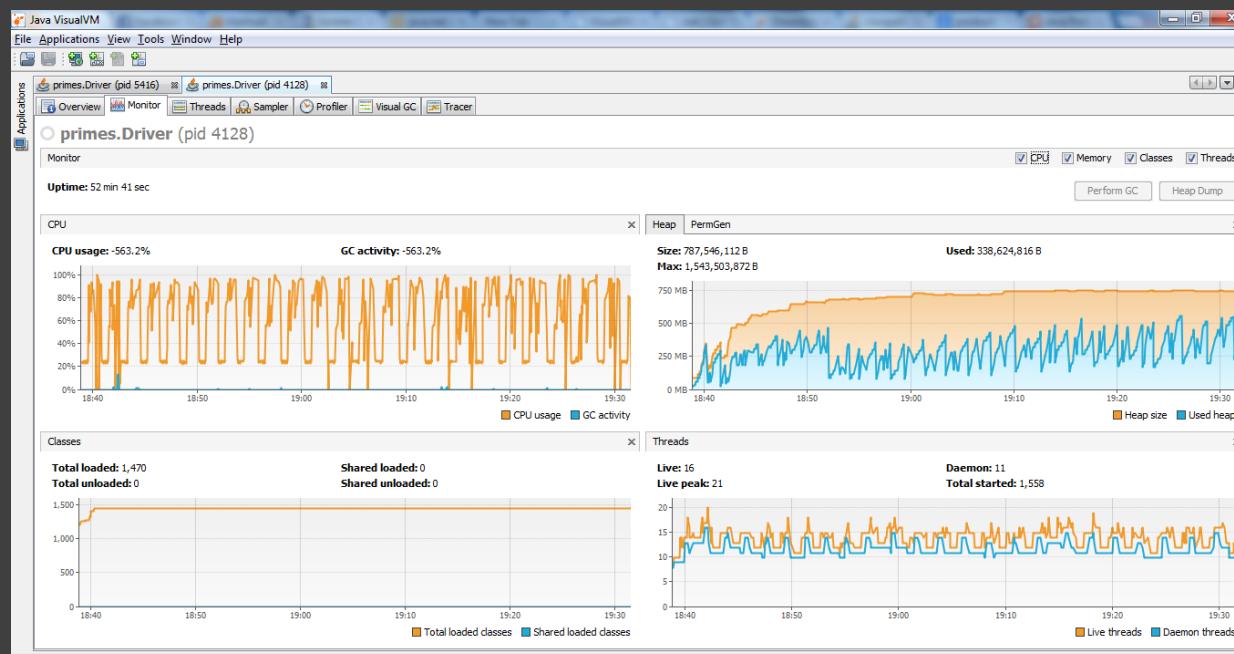
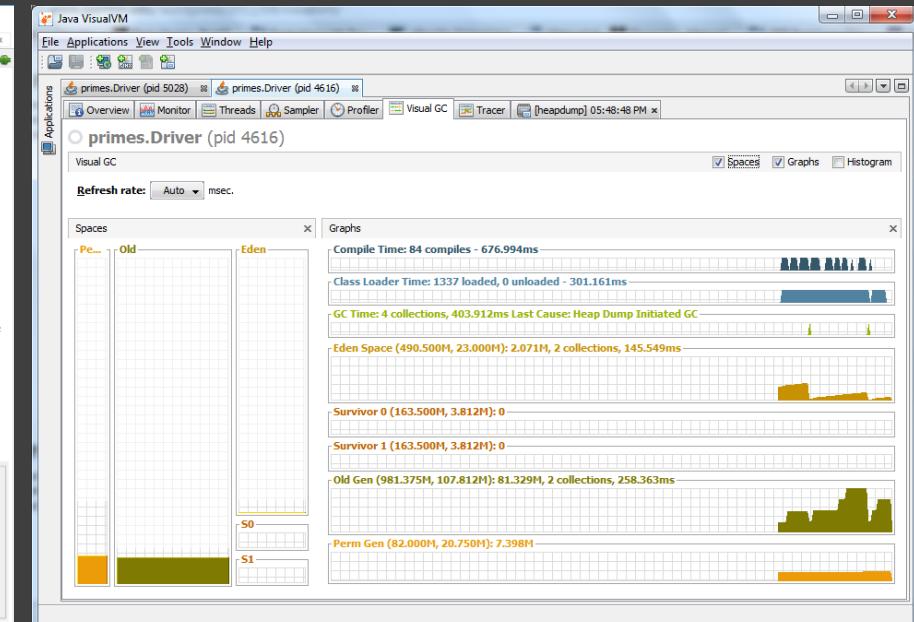
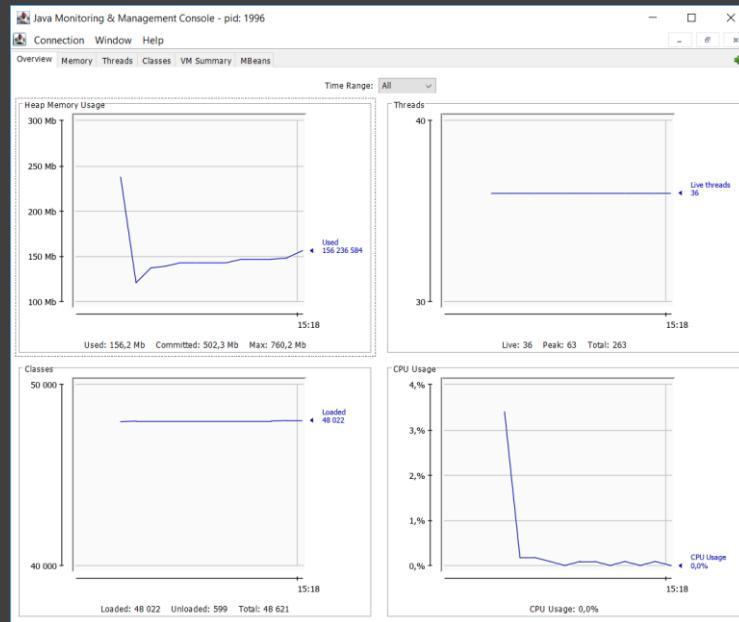
Top of Stack: (sp=0x00007f33e3623e38)
0x00007f33e3623e38: 00007f33ea60e80e 00007f33e3623eb0
0x00007f33e3623e48: 00007f33e3623f10 0000000000000049
0x00007f33e3623e58: 00007f34119f39a0 00007f33f4348d18
0x00007f33e3623e68: 00007f33ea5f7ab5 00007f33e3623eb0
0x00007f33e3623e78: 0000000000000007 00000000fd7627a0
0x00007f33e3623e88: 00000000c107a2c8 00007f33e3623f18
0x00007f33e3623e98: 00007f34014e01a4 00000000c107a2c8
0x00007f33e3623ea8: 00000000fd75fb28 0001004929e1671
0x00007f33e3623eb8: 00000000fd762808 0000000000000001
0x00007f33e3623ec8: 000000000000049 0000000000000049
0x00007f33e3623ed8: 00007f34119f39a0 00007f33f4348d18
0x00007f33e3623ee8: 00007f33ea5f7c1e 0000000000000001
0x00007f33e3623ef8: 000007f340268131d 00007f33e3624008
0x00007f33e3623f08: 00007f34119f3b40 0000000000020049
0x00007f33e3623f18: 00007f34119f3ce8 00007f34119f4958
0x00007f33e3623f28: 00007f33e3624090 00007f34119f3ce8
0x00007f33e3623f38: 0000000000000000 00007f33e3623fa0

Stack: [0x00007f33e3526000,0x00007f33e3623e38], free space=1015k
Native frames: (J=compiled Java code, j=interpreted, Vv=VM code, C=native code)
C [libc.so.6+0x896e1] memcpy+0x11
j org.fusesource.LevelDBJNI.Delete(JLorg/fusesource/leveldbjni/internal/NativeWriteOptions;Lorg/fusesource/leveldbjni/internal/NativeSlice;)J+0
j org.fusesource.LevelDBJNI.Delete(Lorg/fusesource/leveldbjni/internal/NativeWriteOptions;Lorg/fusesource/leveldbjni/internal/NativeSlice;)V+10
j org.fusesource.LevelDBJNI.Delete(Lorg/fusesource/leveldbjni/internal/NativeWriteOptions;Lorg/fusesource/leveldbjni/internal/NativeBuffer;)V+10
j org.fusesource.LevelDBJNI.Delete([B)Lorg/fusesource/leveldb/writeOptions;)Lorg/iq80/leveldb/Snapshot;+27
j org.fusesource.LevelDBJNI.JniDB.delete([B)V+9
j org.apache.hadoop.yarn.server.nodemanager.recovery.NMContainerTokenService.removeContainerToken(Lorg/apache/hadoop/yarn/api/records/ContainerId;)V+28
j org.apache.hadoop.yarn.server.nodemanager.security.NMContainerSecretManager.removeAnyContainerTokenIfExpired()V+103
j org.apache.hadoop.yarn.server.nodemanager.security.NMContainerSecretManager.isValidStartContainerRequest(Lorg/apache/hadoop/yarn/security/ContainerTokenIdentifier;)Z+1
j org.apache.hadoop.yarn.server.nodemanager.ContainerManagerImpl.authorizeStartRequest(Lorg/apache/hadoop/yarn/security/NMTokenIdentifier;Lorg/apache/hadoop/yarn/security/ContainerTokenIdentifier;)V+97
j org.apache.hadoop.yarn.server.nodemanager.ContainerManagerImpl.startContainerInternal(Lorg/apache/hadoop/yarn/security/NMTokenIdentifier;Lorg/apache/hadoop/yarn/security/ContainerTokenIdentifier;Lorg/apache/hadoop/yarn/protocolrecords/StartContainerRequest;)V+3
j org.apache.hadoop.yarn.server.nodemanager.ContainerManagerImpl.startContainers(Lorg/apache/hadoop/yarn/api/protocolrecords/StartContainersRequest;)Lorg/apache/hadoop/yarn/api/protocolrecords/StartContainersResponse;+127
j org.apache.hadoop.yarn.api.impl.pb.service.ContainerManagementProtocolPBServiceImpl.startContainers(Lcom/google/protobuf/RpcController;Lorg/apache/hadoop/yarn/proto/YarnServiceProtos$StartContainersRequestProto;)Lorg/apache/hadoop/yarn/proto/YarnServiceProtos$StartContainersResponseProto;+14
j org.apache.hadoop.yarn.proto.ContainerManagementProtocol$ContainerManagementProtocolService$2.callBlockingMethod(Lcom/google/protobuf/Descriptors$MethodDescriptor;Lcom/google/protobuf/RpcController;Lcom/google/protobuf/Message;)Lcom/google/protobuf/Message;+61
j org.apache.hadoop.ipc.ProtobufRpcEngine$Server$ProtoBufRpcInvoker.call(Lorg/apache/hadoop/ipc/RPC$Server;Ljava/lang/String;Lorg/apache/hadoop/io/Writable;)Lorg/apache/hadoop/io/Writable;+246
j org.apache.hadoop.ipc.RPC$Server.call(Lorg/apache/hadoop/ipc/RPC$RpcKind;Ljava/lang/String;Lorg/apache/hadoop/io/Writable;)Lorg/apache/hadoop/io/Writable;+9
j org.apache.hadoop.ipc.Server$Handler$1.run()Lorg/apache/hadoop/io/Writable;+38
j org.apache.hadoop.ipc.Server$Handler$1.run()Ljava/lang/Object;+1
v ~StubRoutines::call_stub
J 321 java.security.AccessController.doPrivileged(Ljava/security/PrivilegedExceptionAction;Ljava/security/AccessControlContext;)Ljava/lang/Object; (0 bytes) @ 0x00007f340134fea3 [0x00007f340134fe40+0x63]
J 9306 C1 javax.security.auth.Subject.doAs(Ljavax/security/auth/Subject;Ljava/security/PrivilegedExceptionAction;)Ljava/lang/Object; (46 bytes) @ 0x00007f3402ab2ed4 [0x00007f3402ab2840+0x694]
J 10517 C1 org.apache.hadoop.security.UserGroupInformation.doAs(Ljava/security/PrivilegedExceptionAction;)Ljava/lang/Object; (178 bytes) @ 0x00007f3402dee0a4 [0x00007f3402dedfe0+0xc4]
j org.apache.hadoop.ipc.Server$Handler.run()V+315
v ~StubRoutines::call_stub

----- P R O C E S S -----
Java Threads: (=> current thread)
0x00007f33f483d800 JavaThread "Keep-Alive-Timer" daemon [_thread_blocked, id=49429, stack(0x00007f33e111b000,0x00007f33e121c000)]
0x00007f33f5b1d800 JavaThread "RMI TCP Connection(idle)" daemon [_thread_blocked, id=45218, stack(0x00007f33d9ae1000,0x00007f33d9be2000)]
0x00007f33dcfea800 JavaThread "399847377@qtp-1903051549-5" daemon [_thread_blocked, id=40473, stack(0x00007f33d9ee3000,0x00007f33d9fe4000)]
0x00007f33fc257000 JavaThread "LogAggregationService #18" [_thread_blocked, id=37275, stack(0x00007f33d9de2000,0x00007f33d9ee3000)]
0x00007f33fc252000 JavaThread "LogAggregationService #17" [_thread_blocked, id=19628, stack(0x00007f33daef000,0x00007f33dabf0000)]
0x00007f33fc38a000 JavaThread "LogAggregationService #16" [_thread_blocked, id=59087, stack(0x00007f33dbf03000,0x00007f33dbf14000)]
0x00000000001b66800 JavaThread "Thread-349" [_thread_in_native, id=51301, stack(0x00007f33da0e5000,0x00007f33dale6000)]
```

# GUI

- JConsole : le VisualVM historique
- VisualVM : vue CPU (+GC), différent espace mémoire, thread, mbeans JMX avec plugin, ...
- JMC : le futur, mbean sans plugin, ..., GUI jcmd  
il faut s'habituer ...
- JMC/FlightRecorder : utilisable en ligne de commande ... fantastique ...  
MBeans JMX combinable ... allocation/lock/exception ... opensourcé par Oracle récemment depuis JDK 11 (commercial sur JDK 8) mais les early build ont disparus ... et l'accounting du profiling est erroné et profiling arrayCopy() impossible



**Mission Control - Eclipse SDK**

**JOverflow Analysis**

**Object Selection** Referrer Memory KB Overhead KB

- All Objects Unknown GC root 3 543 (12%) 0 (0%)
- Duplicate Strings String[] 2 472 (8%) 0 (0%)
- Sparse Large Collections [HashMap] 2 425 (8%) 0 (0%)
- Small Collections [HashMap] 1 501 (5%) 0 (0%)
- Sparse Small Collections java.lang.reflect.Field[] 820 (2%) 0 (0%)
- Arrays with Underlined Elements Object[] 767 (2%) 0 (0%)
- Arrays with Underlined Elements sun.reflect.UnsafeStaticObjectFieldAccessorImplField 723 (2%) 0 (0%)
- Arrays with One Element org.eclipse.core.internal.registry.KeyedElement[] 673 (2%) 0 (0%)
- Empty Arrays [WeakHashMap] 522 (1%) 0 (0%)
- Empty Collections org.eclipse.core.internal.registry.ReferenceMap\$Entry[] 457 (1%) 0 (0%)

**Ancestor referrer** Ancestor referrer Memory KB Overhead KB

- [HashMap] N/A 3 927 (33%) 0 (0%)
- String[] 3 546 (12%) 0 (0%)
- java.lang.reflect.Field[] 2 472 (8%) 0 (0%)
- Object[] 820 (2%) 0 (0%)
- sun.reflect.UnsafeStaticObjectFieldAccessorImplField[] 767 (2%) 0 (0%)
- org.eclipse.core.internal.registry.K... 736 (2%) 0 (0%)
- [WeakHashMap] 673 (2%) 0 (0%)
- org.eclipse.equinox.internal.p2.c... 528 (1%) 0 (0%)
- org.eclipse.core.internal.registry.R... 458 (1%) 0 (0%)
- [Hashtable] 457 (1%) 0 (0%)
- [ArrayList] 366 (1%) 0 (0%)
- org.eclipse.core.internal.registry.C... 316 (1%) 0 (0%)
- [LinkedHashMap] 288 (0%) 0 (0%)
- [HashMap] 268 (0%) 0 (0%)
- java.lang.reflect.FieldAccessor 259 (0%) 0 (0%)
- Collections\$SynchronizedMap.m 238 (0%) 0 (0%)
- org.eclipse.osgi.framework.interna... 205 (0%) 0 (0%)
- java.util.jar.Attributes.map 203 (0%) 0 (0%)

**Class**

**SQL Statements**

**Database** Events Operative Set Interval: 3 min 11 s (all)

2012-02-01 15:29:28

**JDBC Operations** Shows JDBC data reported by the WebLogic Diagnostics Framework (WLDF).

**Log** Events Operative Set Interval: 1 min (all) Synchronize Selection

2013-06-17 13:38:56

**Event Log**

| Start Time              | End Time                | Duration | Thread                   | Event Type              |
|-------------------------|-------------------------|----------|--------------------------|-------------------------|
| 2013-06-17 13:38:56.755 | 2013-06-17 13:38:56.755 | 0 s      | main                     | Allocation in new TLAB  |
| 2013-06-17 13:38:56.762 | 2013-06-17 13:38:56.762 | 0 s      | Service Thread           | Allocation outside TLAB |
| 2013-06-17 13:38:56.762 | 2013-06-17 13:38:56.762 | 0 s      | main                     | Allocation in new TLAB  |
| 2013-06-17 13:38:56.763 | 2013-06-17 13:38:56.763 | 0 s      | RMI TCP Connection(idle) | Allocation outside TLAB |
| 2013-06-17 13:38:56.764 | 2013-06-17 13:38:56.764 | 0 s      | RMI TCP Connection(idle) | Allocation in new TLAB  |
| 2013-06-17 13:38:56.764 | 2013-06-17 13:38:56.764 | 0 s      | RMI TCP Connection(idle) | Allocation outside TLAB |
| 2013-06-17 13:38:56.765 | 2013-06-17 13:38:56.765 | 0 s      | RMI TCP Connection(idle) | Allocation in new TLAB  |
| 2013-06-17 13:38:56.765 | 2013-06-17 13:38:56.765 | 0 s      | RMI TCP Connection(idle) | Allocation outside TLAB |
| 2013-06-17 13:38:56.765 | 2013-06-17 13:38:56.765 | 0 s      | RMI TCP Connection(idle) | Allocation in new TLAB  |
| 2013-06-17 13:38:56.801 | 2013-06-17 13:38:56.801 | 0 s      | main                     | Allocation in new TLAB  |
| 2013-06-17 13:38:56.803 | 2013-06-17 13:38:56.803 | 0 s      | Service Thread           | Allocation in new TLAB  |

**Event Attributes**

Name Value

- StartTime 2013-06-17 13:38:56.764
- EndTime 2013-06-17 13:38:56.764
- Duration 0 s
- Class 856 bytes
- AllocationSize 856 bytes
- EventThread

**Oracle Java Mission Control**

**JVM Brow** Event Typ localhost flight\_recording\_2013-09-11\_11\_29\_46.jfr

**Overview**

**Dashboard**

Used Java Heap Memory: Now: 13,00 MB Max: 19,44 MB

JVM CPU Usage: Now: 0,00% Mac: 27,78%

Live Set + Fragmentation: Now: 41,27% Max: 41,27%

**MBeans**

The JVM Running Mission Control

- MBean Server
- Flight Recorder
- No Recordings
- URockit 10776 (10776)
- URockit 14928 (14928)
- localhost
- MBean Server
- Flight Recorder

**Runtime**

**Processor**

Machine CPU Usage: Now: 0,00% Mac: 27,78%

JVM CPU Usage: Now: 0,00% Mac: 27,78%

**Memory**

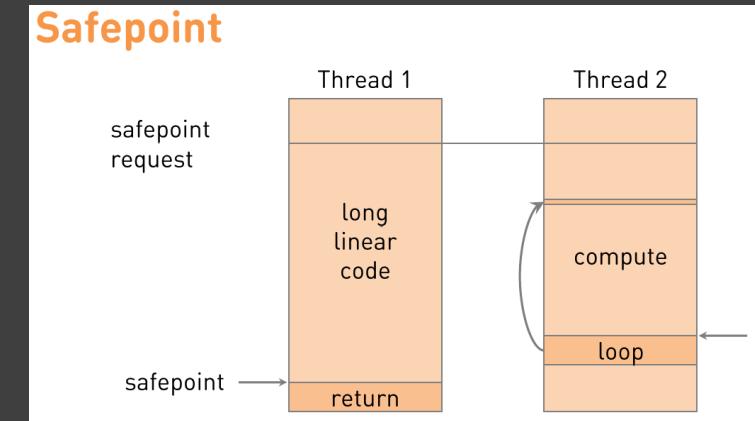
Used Physical Memory (%): Now: 50% Mac: 50%

Used Java Heap (%): Now: 41,27% Mac: 41,27%

**Server Information**

# Profiling

- Profilers (VisualVM, JProfiler, JMC, ...)
  - Souffrent tous des problématiques liées aux safepoints
  - « Stop the world » tous les threads pour sampler
  - Ne voit que le « Java », pas le reste : OS, kernel, runtime JVM (GC, compiler), librairie native, code interprété, inlining, ...
  - Modifie le code, donc pas compilé de la même manière
- S'appuyer sur perf et AsyncGetCallTrace
  - A installer tout le temps (il peut y avoir qq settings sécurité)
  - AsyncGetCallTrace (API interne JDK)



# Profiling

- **async-profiler pour « tout avoir »**
  - <https://github.com/jvm-profiling-tools/async-profiler>
  - Kernel, stack natives, compteurs matériels (page fault, ctx switch, ...) (si exposé), stack Java complete, allocations, locks
  - Ok sur Java 8+

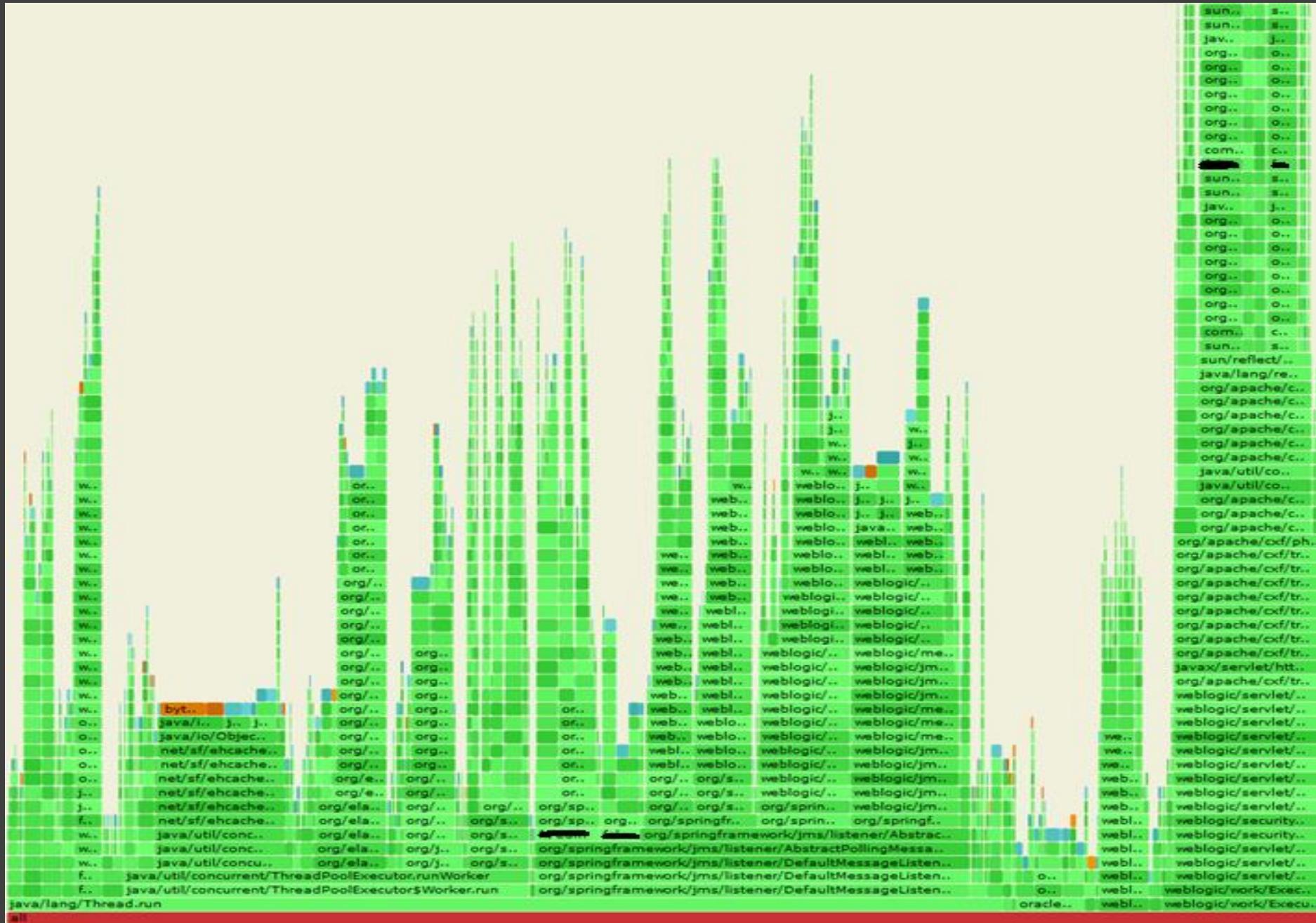
|                       | JFR     | perf           | async-profiler |
|-----------------------|---------|----------------|----------------|
| Java stack            | Yes     | No interpreted | Yes            |
| Native stack          | No      | Yes            | Yes            |
| Kernel stack          | No      | Yes            | Yes            |
| JDK support           | 8*, 11+ | 8u60+          | 6+             |
| OS support            | All     | Linux only     | Linux, macOS*  |
| Permanent overhead    | 0       | 2-5%           | 0              |
| System-wide profiling | No      | Possible       | No             |

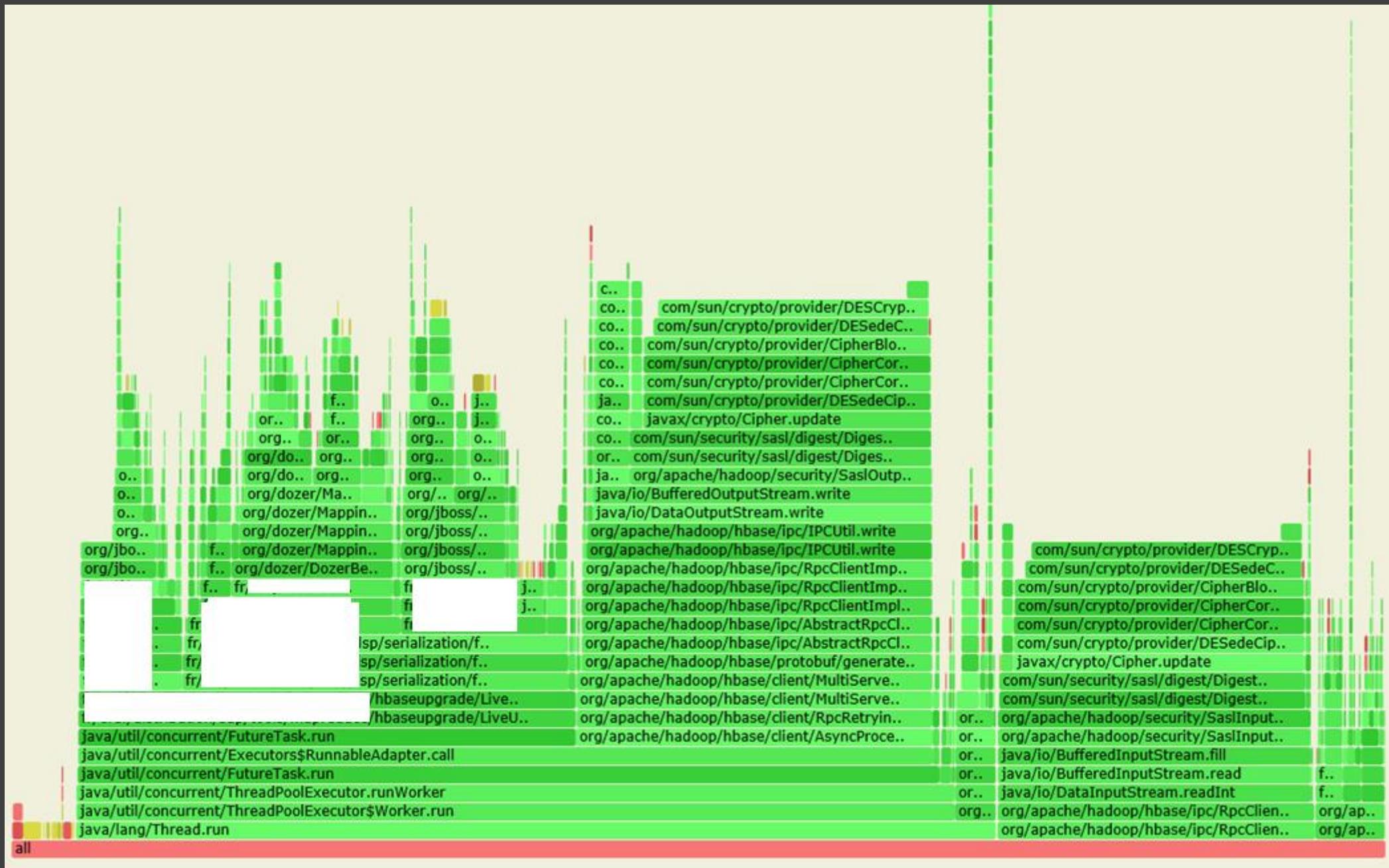
- Bug JVM : <https://bugs.openjdk.java.net/browse/JDK-8178287>  
AsyncGetCallTrace fails to traverse valid Java stacks

# FlameGraph via async-profiler

- Quelques settings à modifier
  - sysctl -w kernel.perf\_event\_max\_stack=4096
  - /proc/sys/kernel/perf\_event\_mlock\_kb
  - echo 1 > /proc/sys/kernel/perf\_event\_paranoid
  - echo 0 > /proc/sys/kernel/kptr\_restrict
  - Si pas loadé en tant qu'agent, pour être plus fiable :  
-XX:+UnlockDiagnosticVMOptions -XX:+DebugNonSafepoints
- ./profiler.sh -d 10 -f perf.svg -i 500us -b 3000000 <pid>
- Palette de couleur
  - green : Java compilé ou classique // aqua : Java inliné
  - brown : kernel // yellow : C++ // red : code natif
- Aussi
  - Profiling des allocations : -e alloc
  - Profiling wall clock time : -e wall -t
  - Origine d'une exception sans log : -e Java\_java\_lang\_Throwable\_fillInStackTrace  
Chercher dans le code : JNICALL object JNICALL

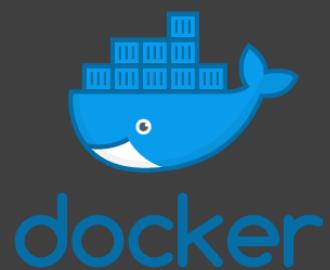






WIP

# Spécificités avec Docker



# Problématique

- Dans le container
  - Pas ou peu d'outillage système (maintenir petite taille image et faible surface d'attaque)
  - JDK ou JRE ?
  - Mémoire limitée
  - PIDs différents
  - syscalls bloqués par seccomp
  - Throttling tooling du fait des quotas
- Sur le host Linux
  - Nécessite des privilèges élevés, pas simple sur plateforme mutualisée
  - Doit pouvoir rentrer dans les namespaces
  - Avec un JDK (>10 pour avoir Attach API dans Docker), montage /tmp éventuel
  - Avec tout l'outillage classique (sysstat, perf, ...)
  - Volume d'échange assez pratique (log, heap dump, flame graph, ...)

- **systemd-cgtop**
  - top pour cgroup

| Control Group                                                            | Tasks | %CPU | Memory | Input/s | Output/s |
|--------------------------------------------------------------------------|-------|------|--------|---------|----------|
| /                                                                        | -     | 20.1 | 1.6G   | 0B      | 23.6K    |
| /user.slice                                                              | 292   | 19.1 | 805.4M | -       | -        |
| /system.slice                                                            | 132   | 0.5  | 275.8M | -       | -        |
| /system.slice/docker.service                                             | 51    | 0.5  | 90.2M  | -       | -        |
| /docker                                                                  | 46    | 0.3  | 686.6M | -       | -        |
| /docker/f9853868effe9755003c9a3141762669fbe6e751eab2525a22c242ec6d544a23 | 46    | 0.3  | 606.0M | -       | -        |
| /system.slice/fwupd.service                                              | 5     | 0.0  | 2.5M   | -       | -        |
| /system.slice/NetworkManager.service                                     | 4     | 0.0  | 12.8M  | -       | -        |
| /system.slice/vboxadd-service.service                                    | 8     | 0.0  | 1.8M   | -       | -        |
| /system.slice/packagekit.service                                         | 3     | 0.0  | 98.1M  | -       | -        |
| /init.scope                                                              | 1     | -    | 5.3M   | -       | -        |
| /system.slice/ModemManager.service                                       | 3     | -    | 2.6M   | -       | -        |
| /system.slice/accounts-daemon.service                                    | 3     | -    | 2.1M   | -       | -        |
| /system.slice/acpid.service                                              | 1     | -    | 336.0K | -       | -        |
| /system.slice/avahi-daemon.service                                       | 2     | -    | 1.1M   | -       | -        |
| /system.slice/colord.service                                             | 3     | -    | 2.8M   | -       | -        |
| /system.slice/cron.service                                               | 1     | -    | 3.3M   | -       | -        |
| /system.slice/cups-browsed.service                                       | 3     | -    | 1.1M   | -       | -        |
| /system.slice/cups.service                                               | 1     | -    | 1.6M   | -       | -        |
| /system.slice/dbus.service                                               | 1     | -    | 3.2M   | -       | -        |
| /system.slice/dev-hugepages.mount                                        | -     | -    | 80.0K  | -       | -        |
| /system.slice/dev-mqueue.mount                                           | -     | -    | 140.0K | -       | -        |
| /system.slice/gdm.service                                                | 3     | -    | 2.4M   | -       | -        |
| /system.slice/irqbalance.service                                         | 1     | -    | 636.0K | -       | -        |
| /system.slice/kerneloops.service                                         | 1     | -    | 780.0K | -       | -        |
| /system.slice/polkit.service                                             | 3     | -    | 4.3M   | -       | -        |
| /system.slice/rsyslog.service                                            | 4     | -    | 1.6M   | -       | -        |
| /system.slice/rtkit-daemon.service                                       | 3     | -    | 520.0K | -       | -        |
| /system.slice/snapd.service                                              | 12    | -    | 7.1M   | -       | -        |
| /system.slice/swappiness.swap                                            | -     | -    | 116.0K | -       | -        |
| /system.slice/sys-fs-fuse-connections.mount                              | -     | -    | 104.0K | -       | -        |
| /system.slice/sys-kernel-config.mount                                    | -     | -    | 60.0K  | -       | -        |
| /system.slice/sys-kernel-debug.mount                                     | -     | -    | 64.0K  | -       | -        |

- **docker stats**
  - top pour container

| CONTAINER    | CPU %   | MEM USAGE / LIMIT     | MEM %  | NET I/O           | BLOCK I/O    | PIDS |
|--------------|---------|-----------------------|--------|-------------------|--------------|------|
| f9853868effe | 210.88% | 611.8 MiB / 1.949 GiB | 30.65% | 34.3 MB / 34.1 MB | 0 B / 4.1 kB | 48   |

- **docker system df**
  - Utilisation du disque

| root@         | ~# docker system df |        |          |                |
|---------------|---------------------|--------|----------|----------------|
| TYPE          | TOTAL               | ACTIVE | SIZE     | RECLAIMABLE    |
| Images        | 4                   | 2      | 2.24 GB  | 360.3 MB (16%) |
| Containers    | 24                  | 2      | 25.68 MB | 0 B (0%)       |
| Local Volumes | 22                  | 22     | 209.8 MB | 0 B (0%)       |

- dive : exploration des layers
- <https://github.com/wagoodman/dive>
- dive <image\_id>

**[● Layers]**

| Cmp    | Size                                                                                            | Command              |
|--------|-------------------------------------------------------------------------------------------------|----------------------|
| 102 MB | FROM sha256:604cbde1                                                                            |                      |
| 745 B  | set -xe && echo '#!/bin/sh' > /usr/sbin/policy-rc.d && echo 'exit 101' >> /usr/sbin/policy-rc.d |                      |
| 0 B    | rm -rf /var/lib/apt/lists/*                                                                     |                      |
| 7 B    | mkdir -p /run/systemd && echo 'docker' > /run/systemd/container                                 |                      |
| 17 MB  | rm -rf /var/lib/apt/lists/* && apt-get clean && apt-get update -y & apt-get upgrade -y          |                      |
| 126 MB | set -eux; ARCH=\$(dpkg --print-architecture); case "\$ARCH" in                                  | ppc64el pp           |
| 41 MB  | jib-maven-plugin:1.1.2                                                                          | drwxr-xr-x 0:0 14 kB |
| 2.2 kB | jib-maven-plugin:1.1.2                                                                          | drwxr-xr-x 0:0 14 kB |
| 14 kB  | jib-maven-plugin:1.1.2                                                                          | drwxr-xr-x 0:0 14 kB |
| 186 B  | jib-maven-plugin:1.1.2                                                                          | drwxr-xr-x 0:0 14 kB |

**[Layer Details]**

```
Digest: sha256:63e67e987db15533d6ed0f78bb6a7b4a1b92d5c219e976206f33976339602e90
Command:
jib-maven-plugin:1.1.2
```

**[Image Details]**

```
Total Image size: 285 MB
Potential wasted space: 42 MB
Image efficiency score: 87 %

Count Total Space Path
ges 2 6.2 MB /var/lib/apt/lists/archive.ubuntu.com_ubuntu_dists_bionic_main_binary-amd64_Packa -rw-r--r-- 0:0 4.2 MB
ges 2 5.8 MB /var/cache/apt/pkgcache.bin -rw-r--r-- 0:0 144 kB
ges 2 5.8 MB /var/cache/apt/srcpkgcache.bin -rw-r--r-- 0:0 4.1 kB
ges 2 4.0 MB /var/lib/apt/lists/archive.ubuntu.com_ubuntu_dists_bionic-updates_main_binary-amd -rw-r--r-- 0:0 245 kB
64_Packages 2 3.2 MB /usr/lib/x86_64-linux-gnu/libstdc++.so.6.0.25 -rw-r--r-- 0:0 9.3 kB
n 2 3.2 MB /var/lib/apt/lists/archive.ubuntu.com_ubuntu_dists_bionic_main_i18n_Translation-e -rw-r--r-- 0:0 56 kB
lation-en 2 2.4 MB /var/lib/apt/lists/archive.ubuntu.com_ubuntu_dists_bionic-updates_main_i18n_Trans -rw-r--r-- 0:0 20 kB
lation-en 2 2.2 MB /var/lib/apt/lists/security.ubuntu.com_ubuntu_dists_bionic-security_main_binary-a -rw-r--r-- 0:0 14 kB
md64_Packages 2 1.6 MB /var/lib/apt/lists/security.ubuntu.com_ubuntu_dists_bionic-security_main_i18n_Tra -rw-r--r-- 0:0 9.3 kB
nslation-en 2 966 kB /var/cache/debconf/templates.dat -rw-r--r-- 0:0 9.7 kB
nslation-en 2 742 kB /var/lib/dpkg/info/libstdc++6:amd64.symbols -rw-r--r-- 0:0 51 kB
2 376 kB /var/log/dpkg.log -rw-r--r-- 0:0 66 kB
2 252 kB /usr/sbin/useradd -rw-r--r-- 0:0 324 kB
2 248 kB /usr/sbin/username -rw-r--r-- 0:0 1.4 MB
2 242 kB /var/lib/apt/lists/archive.ubuntu.com_ubuntu_dists_bionic_InRelease -rw-r--r-- 0:0 33 kB
```

**[Current Layer Contents]**

| Permission | UID:GID | Size   | Filetree                                       |
|------------|---------|--------|------------------------------------------------|
| drwxr-xr-x | 0:0     | 41 MB  | app                                            |
| -rwxr--r-- | 0:0     | 14 kB  | classes                                        |
| -rwxr--r-- | 0:0     | 14 kB  | fr                                             |
| -rwxr--r-- | 0:0     | 14 kB  | java                                           |
| -rwxr--r-- | 0:0     | 14 kB  | javadiag                                       |
| -rwxr--r-- | 0:0     | 1.1 kB | JavadiagApplication.class                      |
| -rwxr--r-- | 0:0     | 720 B  | PrimeNumbers.class                             |
| -rwxr--r-- | 0:0     | 1.4 kB | ServiceHeavyObject.class                       |
| -rwxr--r-- | 0:0     | 2.7 kB | Service.class                                  |
| -rwxr--r-- | 0:0     | 7.4 kB | WebController.class                            |
| -rwxr--r-- | 0:0     | 3.0 kB | tls                                            |
| -rwxr--r-- | 0:0     | 3.8 kB | TLSv13Test\$EchoServer.class                   |
| -rwxr--r-- | 0:0     | 8 B    | TLSv13Test.class                               |
| drwxr-xr-x | 0:0     | 41 MB  | libs                                           |
| -rwxr--r-- | 0:0     | 297 kB | FastInfoSet-1.2.13.jar                         |
| -rwxr--r-- | 0:0     | 114 kB | HdrHistogram-2.1.9.jar                         |
| -rwxr--r-- | 0:0     | 30 kB  | LatencyUtils-2.0.3.jar                         |
| -rwxr--r-- | 0:0     | 69 kB  | activation-1.1.1.jar                           |
| -rwxr--r-- | 0:0     | 122 kB | aspectjrt-1.9.2.jar                            |
| -rwxr--r-- | 0:0     | 2.1 kB | aspectjweaver-1.9.2.jar                        |
| -rwxr--r-- | 0:0     | 245 kB | attoparser-2.0.5.RELEASE.jar                   |
| -rwxr--r-- | 0:0     | 796 kB | bcpkix-jdk15on-1.60.jar                        |
| -rwxr--r-- | 0:0     | 144 kB | bcprov-jdk15on-1.60.jar                        |
| -rwxr--r-- | 0:0     | 4.2 kB | brave-5.5.2.jar                                |
| -rwxr--r-- | 0:0     | 144 kB | brave-context-log4j2-5.5.2.jar                 |
| -rwxr--r-- | 0:0     | 4.1 kB | brave-instrumentation-http-5.5.2.jar           |
| -rwxr--r-- | 0:0     | 24 kB  | brave-instrumentation-httpsyncclient-5.5.2.jar |
| -rwxr--r-- | 0:0     | 14 kB  | brave-instrumentation-httpClient-5.5.2.jar     |
| -rwxr--r-- | 0:0     | 9.3 kB | brave-instrumentation-jms-5.5.2.jar            |
| -rwxr--r-- | 0:0     | 56 kB  | brave-instrumentation-kafka-clients-5.5.2.jar  |
| -rwxr--r-- | 0:0     | 14 kB  | brave-instrumentation-servlet-5.5.2.jar        |
| -rwxr--r-- | 0:0     | 14 kB  | brave-instrumentation-spring-rabbit-5.5.2.jar  |
| -rwxr--r-- | 0:0     | 9.3 kB | brave-instrumentation-spring-web-5.5.2.jar     |
| -rwxr--r-- | 0:0     | 9.7 kB | brave-instrumentation-spring-webmvc-5.5.2.jar  |
| -rwxr--r-- | 0:0     | 51 kB  | cache-api-1.1.0.jar                            |
| -rwxr--r-- | 0:0     | 66 kB  | classmate-1.4.0.jar                            |
| -rwxr--r-- | 0:0     | 1.6 kB | ehcache-3.6.3.jar                              |
| -rwxr--r-- | 0:0     | 1.2 kB | hibernate-validator-6.0.13.Final.jar           |
| -rwxr--r-- | 0:0     | 25 kB  | istack-commons-runtime-3.6.5.jar               |
| -rwxr--r-- | 0:0     | 66 kB  | Jackson-annotations-2.9.0.jar                  |
| -rwxr--r-- | 0:0     | 324 kB | Jackson-core-2.9.7.jar                         |
| -rwxr--r-- | 0:0     | 1.4 kB | Jackson-databind-2.9.7.jar                     |
| -rwxr--r-- | 0:0     | 33 kB  | Jackson-datatype-jdk8-2.9.7.jar                |

**AC quit Tab Switch view AF Filter L Show layer changes A Show aggregated changes**

- cgroup cpu

- docker run --cpus --cpu-shares ...
- docker ps --no-trunc : id complet
- cat /sys/fs/cgroup/cpuacct/docker/<container-id>/cpuacct.stat

```
user 27942
system 16176
```

- cat /sys/fs/cgroup/cpu,cpuacct/docker/<container-id>/cpu.stat  
Ex : throttling moyen de 148 ms

```
nr_periods 1936
nr_throttled 1588
throttled_time 235421460795
echo '235421460795 / 1588 / 1000000' | bc
```

148

- CPU share et cap causent du context switch involontaire  
grep nonvoluntary\_ctxt\_switches /proc/<pid>/task/\*/status

- cgroup mémoire

- docker run --memory --kernel-memory ...
- cat /sys/fs/cgroup/memory/docker/<container-id>/memory.stat
- attention aux JVMs : « oomkilled » avant le heapdump

- cgroup io

- docker run --device-read-bps --device-read-iops
- cat /sys/fs/cgroup/memory/docker/<container-id>/
  - blkio.throttle.read\_bps\_device ou blkio.throttle.write\_bps\_device
  - blkio.throttle.read\_iops\_device ou blkio.throttle.write\_iops\_device

```
cache 69705728
rss 384110592
rss_huge 0
shmem 40960
mapped_file 21327872
dirty 16384
writeback 0
pgpgin 123757
pgpgout 12962
pgfault 107843
pgmajfault 178
inactive_anon 20480
active_anon 384131072
inactive_file 50548736
active_file 19116032
unevictable 0
hierarchical_memory_limit 524288000
total_cache 69705728
total_rss 384110592
total_rss_huge 0
total_shmem 40960
total_mapped_file 21327872
total_dirty 16384
total_writeback 0
total_pgpgin 123757
total_pgpgout 12962
total_pgfault 107843
total_pgmajfault 178
total_inactive_anon 20480
total_active_anon 384131072
total_inactive_file 50548736
total_active_file 19116032
total_unevictable 0
```

- nsenter
  - Exécution de commandes « dans un container » mais sans entrer dans les cgroups
  - Aussi docker exec -it <container-id> <une\_commande>
  - -m: mount, -u: uts, -i: ipc, -n: net, -p: pid, -U: user
  - nsenter -t <pid> -u hostname
  - nsenter -t <pid> -n netstat -antp
  - nsenter -t <pid> -n ss -s
  - nsenter -t <pid> -m -p top
- sidecar container
  - docker run -it --privileged  
  --pid=container:<name|id>  
  --volumes-from=<name|id>  
  --network=container:<name|id>  
  --ipc=container:<name|id>  
  <tooling\_image> /bin/bash

- Confusion entre métriques du host ou du container
  - docker run ... -m800m

```
root@be0f29e122a1:/# free -m
 total used free shared buff/cache available
Mem: 1996 1039 478 19 478 778
Swap: 472 421 51
```

vue JDK 11 : jcmd <pid> VM.info | grep memory\_limit

```
memory_limit_in_bytes: 838860800
```

- docker run ... --cpus=1

```
root@be0f29e122a1:/# nproc
4
```

vue JDK 11 : jcmd <pid> VM.info | grep cpu

```
cpu quota: 100000
```

- idem loadavg, iostat

- Container Metrics system

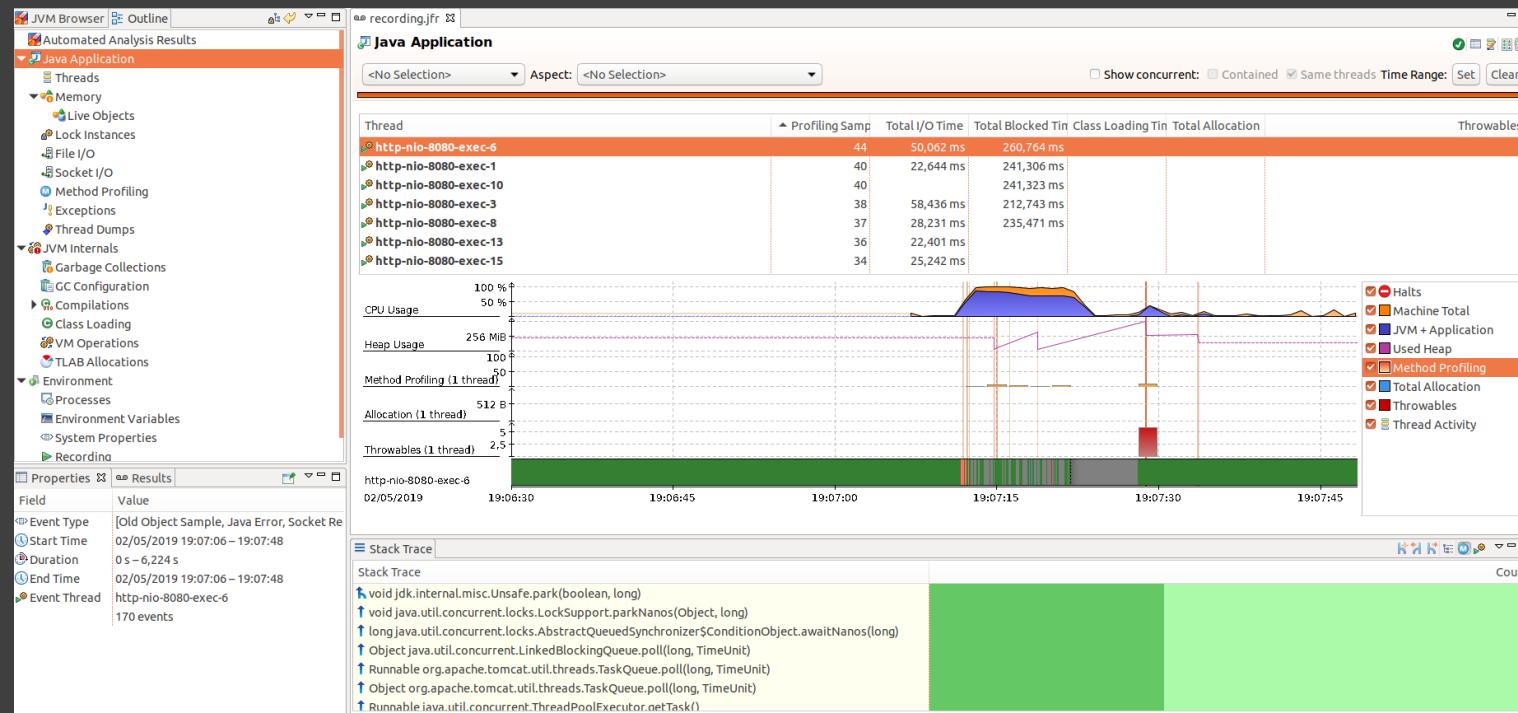
- API disponible en Java 11
- Pas encore exposée en JMX

```
webadm@e05ef69fdb4:~$ java -XshowSettings:system -version
Operating System Metrics:
 Provider: cgroupv1
 Effective CPU Count: 1
 CPU Period: 1000000us
 CPU Quota: 100000us
 CPU Shares: -1
 List of Processors, 4 total:
 0 1 2 3
 List of Effective Processors, 4 total:
 0 1 2 3
 List of Memory Nodes, 1 total:
 0
 List of Available Memory Nodes, 1 total:
 0
 CPUSet Memory Pressure Enabled: false
 Memory Limit: 800.00M
 Memory Soft Limit: Unlimited
 Memory & Swap Limit: 0.00K
 Kernel Memory Limit: Unlimited
 TCP Memory Limit: Unlimited
 Out Of Memory Killer Enabled: true

openjdk version "11.0.3" 2019-04-16
OpenJDK Runtime Environment AdoptOpenJDK (build 11.0.3+7)
OpenJDK 64-Bit Server VM AdoptOpenJDK (build 11.0.3+7, mixed mode)
```

- **async-profiler**
  - Selon version de kernel, de Docker, ...
  - `echo 1 > /proc/sys/kernel/perf_event_paranoid`
  - `docker run --cap-add SYS_ADMIN`  
ou  
`docker run --security-opt seccomp:unconfined`  
ou profile seccomp avec whitelisting des syscall nécessaires
  - Si on ne peut pas changer `perf_event_paranoid` : `./profile ... --all-user`
  - Mettre `async-profiler` à la racine d'un volume partagé
  - `./profile ... -f <volume>/perf.svg`

- Exemple avec FlightRecorder depuis le host
  - jcmsg <pid> JFR.start duration=30s filename=/path/in/container/perf-tbl.jfr
  - jcmsg <pid> JFR.dump
  - docker cp <container-id>:/path/in/container/perf-tbl.jfr /path/in/host ou plus rapide sur un volume
  - Analyse avec Mission Control

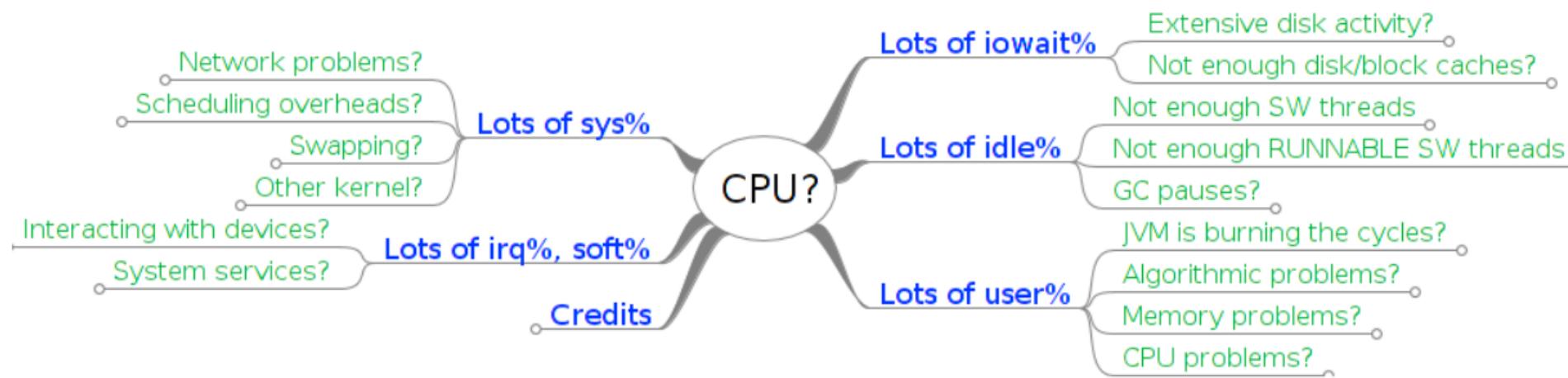


# Scénarios

# Exemple sur une JVM

- Processus d'analyse complet (issue d'une conférence de Aleksey Shipilev et Kirk Pepperdine)
- Niveau système :
  - USE method sur les ressources CPU, RAM, IO ...
- Niveau JVM :
  - GC, thread, JIT ...
- Niveau applicatif :
  - algo en premier, thread/pooling, lock
- Bas niveau
  - cache CPU, alignement, miss, ...

- Commencer par le CPU et explorer les branches



# • Niveau système



## Not particularly the application code fault

- Most obvious contender is network I/O
- Then, scheduling overheads
- Then, swapping
- Then, in minor cases, other kernel



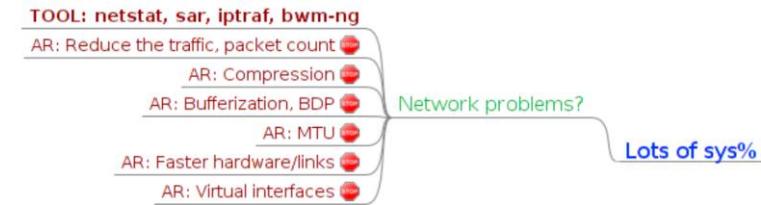
## The symptom of the unbalanced threading

- Lots of voluntary context switches (thread thrashing)
- Lots of involuntary context switches (over-saturation)



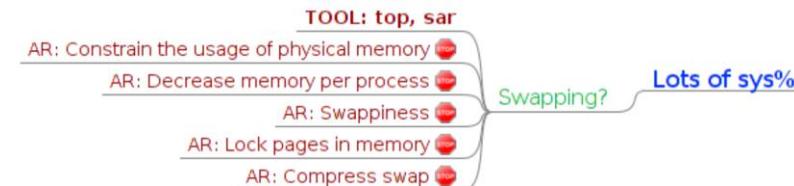
## Sometimes kernel is your enemy

- Unusual API choices from the JVM and/or application
- (Un)known bugs



## One of the major contributors to sys%

- In many cases, hardware/OS configuration is enough
- In other cases, application changes might be necessary



## Swapping is the killer for Java performance

- The target is to avoid swapping at all costs
- Swapping out other processes to save the memory is good



## Usual thing when interacting with the devices

- Sometimes IRQ balancing is required
- Sometimes IRQ balancing is expensive



### Expected contributor with disk I/O

- Watch for disk activity
- Watch for disk throughput
- Watch for disk IOPS



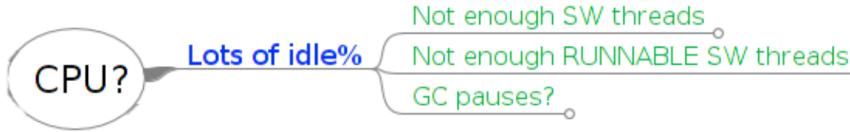
### Is that amount of I/O really required?

- Caching, bufferization are your friends
- More (faster) disks can solve throughput/IOPS problems



### More caching helps?

- Reduce other physical memory usages, free up for caches
- Trade in performance over consistency



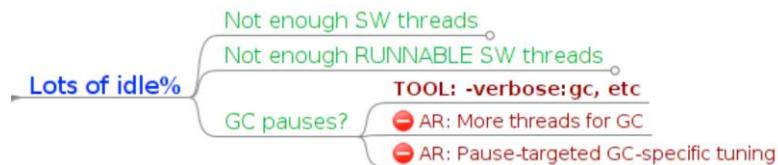
### **There are resources, but nobody uses them?**

- This is admittedly easy to diagnose
- ...and very easy to miss



### **Running low-threaded applications on manycore hosts**

- The signal for you to start parallelizing
- Or, reduce the number of available HW strands



### **Very rare, and surprising case**

- Application is highly threaded
- GC is frequently running with low thread count
- The average CPU utilization is low



### **There are not enough threads ready to run**

- Locking?
- Waiting for something else?

# • Niveau JVM

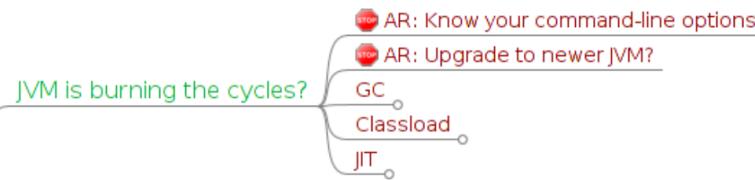


Lots of user%

- JVM is burning the cycles?
- Algorithmic problems?
- Memory problems?
- CPU problems?

## Application/JVM is finally busy

- This is where most people start
- This is where profilers start to be actually useful



## JVM is the new abstraction level

- Interacts with the application, mangles into application
- JVM performance affects application performance

JVM is burning the cycles?

GC

- TOOL: `-verbose:gc, -XX:+PrintGCDetails, VisualGC`
- AR: Tune Java heap, generations, and regions
  - AR: Thread stack size
  - AR: (Un)usual tuning

## GC

- Most usual contender in JVM layer
- Lots of things to try fixing (not covered here, see elsewhere)

JVM is burning the cycles?

Classload

TOOL: `verbose:class, MXBeans`

- AR: Turn off bytecode verification: `--no-verify`
- AR: Turn on CDS: `-Xshare:on`
- AR: Recompile your Java code with updated javac
- AR: Increase the size of system dictionary
- AR: Repackage classes into small amount of larger JARs

## Classload

- Important for startup metrics; not really relevant for others
- Removing the loading obstacles is the road to awe

JVM is burning the cycles?

JIT

- TOOL: `PrintCompilation, MXBeans`
- AR: Choose the compiler
    - server
    - client
    - `-XX:+TieredCompilation`
  - AR: Low-level tuning
  - AR: Go to OpenJDK ML and ask

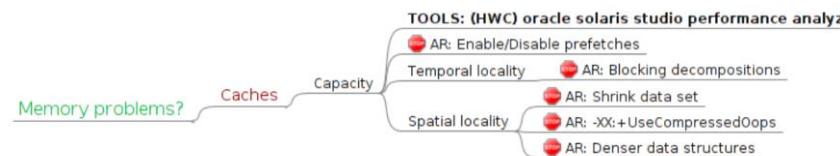
## JIT

- Very cool to have your code compiled
- Sometimes it's even cooler to get the code compiled better



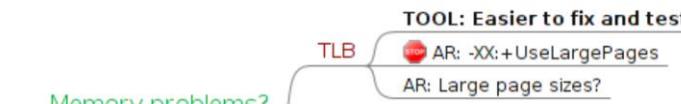
## Memory

- The gem and the curse of von-Neumann architectures
- Dominates most of the applications (in different forms)



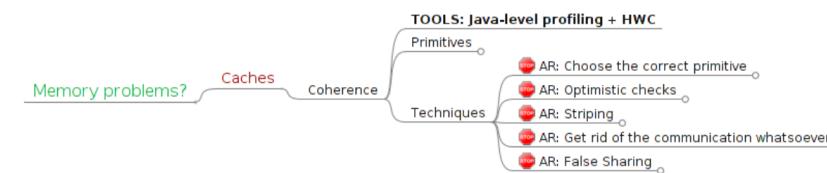
## CPU caches: capacity

- Important to hide memory latency (and bandwidth) issues
- Virtually all applications today are memory/cache-bounded



## TLB

- Very important for memory-bound workloads
- “Invisible” artifact of virtual memory system



## CPU Caches: coherence

- Inter-CPU communication is managed via cache coherence
- Understanding this is the road to master the communication



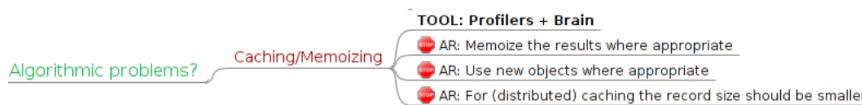
## Memory Bandwidth

- Once caches run out, you face the memory
- Dominates the cache miss performance
- Faster memory, multiple channels help



### Application level

- In many, many cases, silly oversights in algorithms use
- Cargo cult of approaches, patterns, code reuse

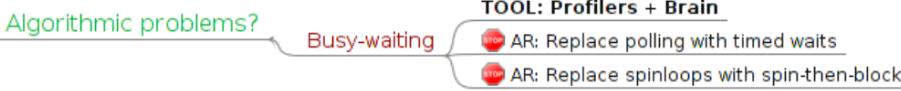


### Application Caching

- Seems to be the answer to most performance problems?
- In fact, blows up the footprint, heap occupancy, etc

### Algorithmic Complexity

- Figuring out the straight-forward code has huge complexity
- Sometimes, the low-O code is slower than high-O code



### Application Busy-Waits

- The natural instinct: blocked waits (with helping)
- For latency-oriented: busy-waits are profitable

# Scénarios

- Démos
- Différents type de charge et vmstat :  
[https://www.thomas-krenn.com/en/wiki/Linux Performance Measurements using vmstat](https://www.thomas-krenn.com/en/wiki/Linux_Performance_Measurements_using_vmstat)

# Conclusion

- Un peu de théorie
- Méthodologie et « à ne pas faire »
- Etat des lieux rapides, quelques commandes à connaître
- Outils plus perfectionnés
- Outillage spécifique JVM
- Interprétation toujours délicate : pas de conclusions hâtives



# Sources & références

- Expériences personnelles
- Brendan Gregg, Sasha Goldshtein, Alexey Ragozin, Kirk Pepperdine, Charlie Hunt, Andrei Pangin, Nitsan Wakart, Richard Waburton, Thomas Stüfe, Evan Jones, Jonatan Kazmierczak, ...
- Google SRE
- ...

# Annexes

<http://j.mp/PerfMindMap>

