

## Admins - Demande #4450

### Améliorer la gestion du swap

02/05/2020 11:22 - Christian P. Momon

<b>Statut:</b>	Fermé	<b>Début:</b>	02/05/2020
<b>Priorité:</b>	Normale	<b>Echéance:</b>	
<b>Assigné à:</b>	Christian P. Momon	<b>% réalisé:</b>	0%
<b>Catégorie:</b>		<b>Temps estimé:</b>	0.00 heure
<b>Version cible:</b>	Juin 2020	<b>Temps passé:</b>	0.00 heure
<b>Difficulté:</b>	2 Facile		

#### Description

Nous constatons des problèmes d'I/O sur le SI April. Ça peut venir de débordement en swap.

Le swap est connu pour être très défavorable aux performances I/O.

Actuellement, la présence de swap sur les VM s'appuie sur deux considérations :

- en fonctionnement nominal, le swap est inutile sur les VM ;
- en cas de manque de RAM, le swap peut permettre :
  - d'amortir les problèmes,
  - de détecter qu'il faut faire quelque chose :
    - paramétrer un processus,
    - rajouter de la RAM.

État des swap le 02/05/2020 à 10h27 (uptime global : 2 jours) :

	total (Mo)	used (Mo)	free (Mo)	swappiness	vf
s_cache_pressure					
==== bastion =====					
100	Swap:	1903	93	1810	1
==== admin =====					
100	Swap:	1903	204	1699	1
==== dns =====					
100	Swap:	951	3	948	1
==== mail =====					
100	Swap:	951	40	911	1
==== sympa =====					
100	Swap:	951	380	571	1
==== adl =====					
100	Swap:	951	127	824	1
==== lamp =====					
100	Swap:	951	112	839	1
==== agir =====					
100	Swap:	951	146	805	1
==== bots =====					
100	Swap:	951	42	909	1
==== dtc =====					
100	Swap:	951	133	818	1
==== drupal6 =====					
100	Swap:	951	31	920	1
==== republique-numerique =====					
100	Swap:	951	43	908	1
==== mumble =====					
100	Swap:	951	22	929	1
==== candidatsfr =====					
100	Swap:	951	311	640	1
==== pad =====					
100	Swap:	951	104	847	1
==== scm =====					
100	Swap:	951	9	942	1
==== pouet =====					
100	Swap:	951	157	794	1
==== webchat =====					
	Swap:	951	12	939	1

100	==== photos =====	Swap:	951	88	863	1
100	==== cms-dev =====	Swap:	951	36	915	1
100	==== virola.april.org =====	Swap:	6143	176	5967	1
100	==== calamus.april.org =====	Swap:	6143	114	6029	1
100	==== galanga.april.org =====	Swap:	951	95	856	1
100						

## /proc/sys/vm/swappiness

<https://www.kernel.org/doc/Documentation/sysctl/vm.txt>

### swappiness

This control is used to define how aggressive the kernel will swap memory pages. Higher values will increase aggressiveness, lower values decrease the amount of swap. A value of 0 instructs the kernel not to initiate swap until the amount of free and file-backed pages is less than the high water mark in a zone.

The default value is 60.

## /proc/sys/vm/vfs\_cache\_pressure

### vfs\_cache\_pressure

This percentage value controls the tendency of the kernel to reclaim the memory which is used for caching of directory and inode objects.

At the default value of `vfs_cache_pressure=100` the kernel will attempt to reclaim dentries and inodes at a "fair" rate with respect to pagecache and swapcache reclaim. Decreasing `vfs_cache_pressure` causes the kernel to prefer to retain dentry and inode caches. When `vfs_cache_pressure=0`, the kernel will never reclaim dentries and inodes due to memory pressure and this can easily lead to out-of-memory conditions. Increasing `vfs_cache_pressure` beyond 100 causes the kernel to prefer to reclaim dentries and inodes.

Increasing `vfs_cache_pressure` significantly beyond 100 may have negative performance impact. Reclaim code needs to take various locks to find freeable directory and inode objects. With `vfs_cache_pressure=1000`, it will look for ten times more freeable objects than there are.

La valeur par défaut est 60.

## Articles

<https://haydenjames.io/linux-performance-almost-always-add-swap-space/>

## Questions

- 1) alors que le swappiness est à 1, pourquoi ça swap sur les vm ?
- 2) faut-il modifier le paramétrage ? swappiness à 0 ? `vfs_cache_pressure` à 1 ? `vfs_cache_pressure` à 0 ?
- 3) comment différencier swap intelligent (page extrêmement rarement utilisée) du swap par manque de ram (celui qui explose les I/O) ?
- 4) faut-il désactiver le swap sur les VM ?
  - A priori, non. Le swap est censé protéger des OOMKILL malheureux.

## Demandes liées:

## Historique

### #1 - 04/05/2020 00:43 - Christian P. Momon

- Statut changé de Nouveau à En cours de traitement
- Assigné à mis à Christian P. Momon

[https://fr.wikipedia.org/wiki/Espace\\_d%27%C3%A9change](https://fr.wikipedia.org/wiki/Espace_d%27%C3%A9change) :

Valeur	Stratégie
vm.swappiness = 0	(Depuis Linux 3.5) Fonctionnement particulier : le noyau ne va utiliser le swap que pour éviter les erreurs de manque de mémoire.
vm.swappiness = 1	Utilisation minimale du swap avec le fonctionnement général. Correspond à swappiness = 0 avant Linux 3.5.

Le nouveau cluster du SI April a été fait en Jessie dont le noyau était antérieur à Linux 3.5. D'après l'extrait Wikipédia ci-dessus, à l'époque les valeurs 0 correspondait à « utilisation minimale du swap ». Depuis Linux 3.5, il faut mettre la valeur à 1 pour activer ce mode.

### #2 - 04/05/2020 01:23 - Christian P. Momon

La valeur `vm.swappiness` est définie dans le fichier `/etc/sysctl.d/ee.conf` :

```
(April) root@calamus:/etc/sysctl.d[master]# cat ee.conf
kernel.printk = 4 4 1 7
# A server should try to not swap ...
vm.swappiness = 1
```

Ce fichier dépend du paquet `eeinstall` :

```
(April) root@calamus:/etc/sysctl.d[master]# dpkg -S ee.conf
eeinstall: /etc/sysctl.d/ee.conf

(April) root@calamus:/etc/sysctl.d[master]# apt policy eeinstall
eeinstall:
  Installed: 9.00~ee100+5
  Candidate: 9.00~ee100+5
  Version table:
 *** 9.00~ee100+5 100
    100 http://apt.easter-eggs.com/debian buster-ee/main amd64 Packages
    100 /var/lib/dpkg/status
```

Historique Git de `/etc/sysctl.d/ee.conf` :

```
2016-04-28 09:10 +0000 root I Initial commit
+vm.swappiness = 0

2018-11-09 16:28 +0000 root o committing changes in /etc after apt run
-vm.swappiness = 0
+vm.swappiness = 1
```

### #3 - 04/05/2020 01:38 - Christian P. Momon

Interprétation historique :

- en 2016, à l'installation du cluster en Jessie, la valeur `vm.swappiness=0` signifiait « Utilisation minimale du swap ».
- à partir du noyau 3.5 :
  - la valeur `vm.swappiness=0` change de sens et signifie « n'utiliser le swap que pour éviter un OOMKILL »,
  - l'équivalent de l'ancien `vm.swappiness=0` devient `vm.swappiness=1`,
  - logiquement, pour garantir un iso-fonctionnement, EE met à jour le paquet `eeinstall` pour passer la valeur à `vm.swappiness=1`.

Questions à poser à EE :

- l'interprétation historique est-elle juste ?
- est-il intéressant de mettre `vm.swappiness=0` (au lieu de 1) sur le SI April ?
- EE va-t-il mettre `vm.swappiness=0` (au lieu de 1) dans le paquet `eeinstall` ?

#### #4 - 04/05/2020 09:53 - François Poulain

Le swap est connu pour être très défavorable aux performances I/O.

Notons qu'on pourrait gagner pas mal en perfs à ce niveau en déclarant un swap file qui court-circuite la virtualisation et drbd.

#### #5 - 04/05/2020 10:31 - François Poulain

en fonctionnement nominal, le swap est inutile sur les VM ;

Ça contredit ce qui est écrit ici, lié depuis la page de wikipédia (note-1) : <https://chrisdown.name/2018/01/02/in-defence-of-swap.html>

#### #6 - 04/05/2020 10:57 - Christian P. Momon

François Poulain a écrit :

en fonctionnement nominal, le swap est inutile sur les VM ;

Ça contredit ce qui est écrit ici, lié depuis la page de wikipédia (note-1) : <https://chrisdown.name/2018/01/02/in-defence-of-swap.html>

L'article cite les cas suivants où le swap est utile :

- éviter les OOMKILL (ça peut casser des bases de données...);
- faire confiance au noyau pour swapper les pages ultra rarement utilisées et donc se donner plus de place pour les buffer/cache ; c'est tentant mais dans notre cas quelle quantité de pages cela concerne-t-il vraiment ? On peut espérer que c'est un nombre très faible sinon ça veut dire qu'on pourrait rajouter de la mémoire. Donc du coup on peut s'en passer.

Donc, « nominalement », le swap est inutile, mais dans la pratique c'est bien pour au moins se protéger des OOMKILL. D'où la proposition de mettre `vm.swappiness=0`.

#### #7 - 21/05/2020 18:42 - Christian P. Momon

La réponse de Emmanuel Lacour d'Ester-egg :

```
Subject: [Easter-eggs #73359] Re: Question de swappiness pour le SI
April (pas urgent)
From: Emmanuel Lacour via RT <support@easter-eggs.com>
To: brenard@easter-eggs.com
CC: cmomon@april.org
Date: Wed, 13 May 2020 16:55:51 +0200
Bonjour,
```

alors, c'est vrai que la doc officielle est moyennement précise, mais la page suivante détaille bien le fonctionnement:  
<https://www.howtogeek.com/449691/what-is-swappiness-on-linux-and-how-to-change-it/>

Le swappiness ne permet donc pas de diminuer l'utilisation de la swap, mais d'indiquer au kernel quel type de mémoire il doit mettre en swap pour en libérer (de la mémoire). Globalement, sauf incompréhension de ma part bien sûr, il est quand même préférable d'aller taper dans les pages de cache fichier plutôt que dans les pages anonymes si on veut minimiser les IO disques (en terme d'IOPS, pas forcément de bande passante). Mais mettre une valeur de 1 permet quand même de ne pas complètement désactiver la récupération des pages anonymes ... pour le cas où.

Donc AMHA, il vaut mieux rester sur une valeur basse de swappiness, mais pas 0, tout en comprenant bien (il suffit de tester ;) ) que ça ne va pas désactiver la swap (qui par ailleurs est nécessaire, même vide, pour la gestion de l'overcommit de mémoire).

--

Emmanuel Lacour :: support-eggs.com :: 01xxxxxxxxx

#### #8 - 21/05/2020 18:43 - Christian P. Momon

Oki, donc on a garder `vm.swappiness=1`.

**#9 - 21/05/2020 18:44 - Christian P. Momon**

Mais quand même, je tente une demande complément au support Easter-eggs :

```
Le 13/05/2020 à 16:55, Emmanuel Lacour via RT a écrit :
> Donc AMHA, il vaut mieux rester sur une valeur basse de swappiness,
> mais pas 0
  Bonjour Emmanuel \o/
```

Un grand merci d'avoir pris la peine de creuser ce sujet vraiment pas évident :D

D'accord pour swappiness=1.

Mais alors je reformule la question. Nos vm n'ont absolument aucune raison de swapper, sauf pour les deux cas suivants pour lesquels nous gardons activé le swap afin d'éviter de malheureux OOMKILL :

1) incident de processus : ponctuellement, un processus part en vrille ;

2) mémoire mal dimensionnée : du coup la consommation de swap est un indice qu'il faut augmenter la mémoire.

Or, nous avons des cas illogique de swap :

A) exemple sur une vm

```
(April) root@candidatsfr:~# free -m
      total        used         free   shared  buff/cache   available
Mem:    986         429         145        38        412         340
Swap:   951         371         580
```

```
(April) root@candidatsfr:~# psswap |head -1
nom          pid      swap
mysqld       616     307908 kB
systemd-journal 239    14420 kB
apache2     25377   5960 kB
apache2     17516   3860 kB
apache2     6007    3828 kB
apache2     25383   3828 kB
```

Quel paramétrage faire pour éviter cette situation ?  
Comment interdire à certains processus de swapper ?  
Devons-nous envisager un cron quotidien pour vider le swap ?

B) exemple sur une pm

Cas ridicule, nos machines physiques swappent aussi :

```
(April) root@calamus:~# free -m
      total        used         free   shared  buff/cache   available
Mem:  32003    12856         264        273    18882    18416
Swap:  6143         87        6056
```

```
(April) root@calamus:~# psswap |head
systemd-journal 334    17540 kB
icinga2         834    5612 kB
libvirtd        830    4152 kB
fail2ban-server 833    1428 kB
```

Bien que ce soit faible, peut-on faire quelque chose ?  
Sachant que les machines physiques ne font tourner que des vm, leur mettre un swap est-il utile ?

Voilà, nous sommes preneur de vos expériences et expertises. Merci par avance :D

Librement,

Christian (Cpm).

**#10 - 21/05/2020 18:45 - Christian P. Momon**

- Statut changé de En cours de traitement à Attente d'information

**#11 - 27/05/2020 21:58 - Quentin Gibeaux**

- *Version cible changé de Mai 2020 à Juin 2020*

**#12 - 01/07/2020 19:59 - Christian P. Momon**

- *Statut changé de Attente d'information à Résolu*

Le ticket a été fermé côté Easter Eggs :

Le 10/06/2020 à 11:27, Emmanuel Lacour via RT a écrit :

> [english version below]

>

> Votre requête a été marquée comme "résolue". N'hésitez pas à répondre à ce message si vous avez des questions ou commentaires sur le sujet de votre demande initiale.

>

> Pour toute nouvelle demande, merci de créer un ticket via un nouvel email.

En l'absence d'information supplémentaire, je ferme le ticket.

**#13 - 01/07/2020 21:45 - Quentin Gibeaux**

- *Statut changé de Résolu à Fermé*

**#14 - 17/08/2020 17:28 - Christian P. Momon**

- *Lié à Demande #4664: Améliorer la gestion du swap (suite) ajouté*