

Infra Chapril - Demande #5763

Analyse de IOWait sur l'ensemble de l'infra

29/01/2022 16:21 - pitchum .

Statut:	En cours de traitement	Début:	29/01/2022
Priorité:	Normale	Echéance:	
Assigné à:		% réalisé:	0%
Catégorie:		Temps estimé:	0.00 heure
Version cible:			
Description			
Depuis l'upgrade Bullseye de l'ensemble de l'infra en décembre, on constate de fréquentes alertes de LOAD sur certaines VM (valise notamment) et des alertes IOWAIT DRBD sur les hyperviseurs.			
C'est probablement la cause des erreurs 502 signalées par LordPhoenix sur mastodon			

Historique

#1 - 29/01/2022 16:32 - pitchum .

J'ai essayé de voir si on pouvait "optimiser" nos fichiers qcow2.
Sur coon, j'ai attaché 2 nouveaux disques à la VM ludo :

- un disque qcow2 tout neuf :
qemu-img create -f qcow2 -o preallocation=falloc,cluster_size=262144,lazy_refcounts=on,extended_l2=on /var/lib/libvirt/coon/ludo-bis.qcow2 20G
virsh attach-disk ludo /var/lib/libvirt/coon/ludo-bis.qcow2 vdb --subdriver qcow2 --live --config
- un disque LVM indépendant de DRBD
lvcreate -n vm-ludo-test -L 20G vg_coon
virsh attach-disk ludo /dev/vg_coon/vm-ludo-test vdc --subdriver raw --live --config

Sur ludo, j'ai utilisé chacun de ces nouveaux disques dans un PV dédié et créé un LV dédié, montés (dans des dossier choisis un peu à l'arrache). J'avais donc 3 dossiers dans lesquels j'allais pouvoir faire des tests d'écriture avec la commande dd, en suivant les suggestions fournies sur [cyberciti.biz/howto-linux-unix-test-disk-performance-with-dd-command](#) :

- /var => sur le qcow2 original
- /mnt => sur le qcow2 tout neuf
- /tmp/lvm_disk => sur la partition LVM sur coon

Résultats

disk throughput (write speed)

```
(^)=root@ludo:~# dd if=/dev/zero of=/mnt/test-dd1.img bs=1G count=1 oflag=dsync
1+0 enregistrements lus
1+0 enregistrements écrits
1073741824 octets (1,1 GB, 1,0 GiB) copiés, 49,3794 s, 21,7 MB/s
```

```
(^)=root@ludo:~# dd if=/dev/zero of=/var/test-dd1.img bs=1G count=1 oflag=dsync
1+0 enregistrements lus
1+0 enregistrements écrits
1073741824 octets (1,1 GB, 1,0 GiB) copiés, 50,9006 s, 21,1 MB/s
```

```
(^)=root@ludo:~# dd if=/dev/zero of=/tmp/lvm_disk/test-dd1.img bs=1G count=1 oflag=dsync
1+0 enregistrements lus
1+0 enregistrements écrits
1073741824 octets (1,1 GB, 1,0 GiB) copiés, 45,3781 s, 23,7 MB/s
```

disk latency

```
(^)=root@ludo:~# dd if=/dev/zero of=/var/test-dd2.img bs=512 count=1000 oflag=dsync
1000+0 enregistrements lus
1000+0 enregistrements écrits
512000 octets (512 kB, 500 KiB) copiés, 476,124 s, 1,1 kB/s
```

```
(^)=root@ludo:~# dd if=/dev/zero of=/mnt/test-dd2.img bs=512 count=1000 oflag=dsync
1000+0 enregistrements lus
1000+0 enregistrements écrits
```

512000 octets (512 kB, 500 KiB) copiés, 411,88 s, 1,2 kB/s

```
=(^-^)=root@ludo:~# dd if=/dev/zero of=/tmp/lvm_disk/test-dd2.img bs=512 count=1000 oflag=dsync  
1000+0 enregistrements lus  
1000+0 enregistrements écrits  
512000 octets (512 kB, 500 KiB) copiés, 241,711 s, 2,1 kB/s
```

conclusion

On dirait qu'avec LVM on gagne en latence.
Pas sûr que ça justifie de remettre en cause l'usage de DRBD.

Il y a sûrement d'autres tests à faire.

#2 - 09/02/2022 00:11 - Pierre-Louis Bonicoli

En ce qui concerne le temps de création de la sauvegarde du service Gitea avant sa mise à jour, voici ci-dessous quelques durées. Je n'ai pas l'impression que la mise à jour en Bullseye ait un impact sur la durée de la création de cette sauvegarde. La taille de la sauvegarde est actuellement de 4.9Go.

Actions	Durée de la création de la sauvegarde du service Gitea	Fin (approximative) de la mise à jour
mise à jour 1.14.6	5 minutes	08/08 23h18
mise à jour 1.15.3	8 minutes	25/09 15h42
doublement de la taille de la sauvegarde (en raison d'un nouveau gros dépôt)		
mise à jour 1.15.6	12 minutes	30/10 12h49
migration bullseye		
mise à jour 1.15.7	20 minutes	21/12 17h
mise à jour 1.15.8	13 minutes	24/12 00h24
mise à jour 1.15.9	26 minutes	31/12 01h30 => BACKUP!
mise à jour 1.15.10	12 minutes	20/01 02h38
mise à jour 1.15.11	10 minutes	03/02 00h34

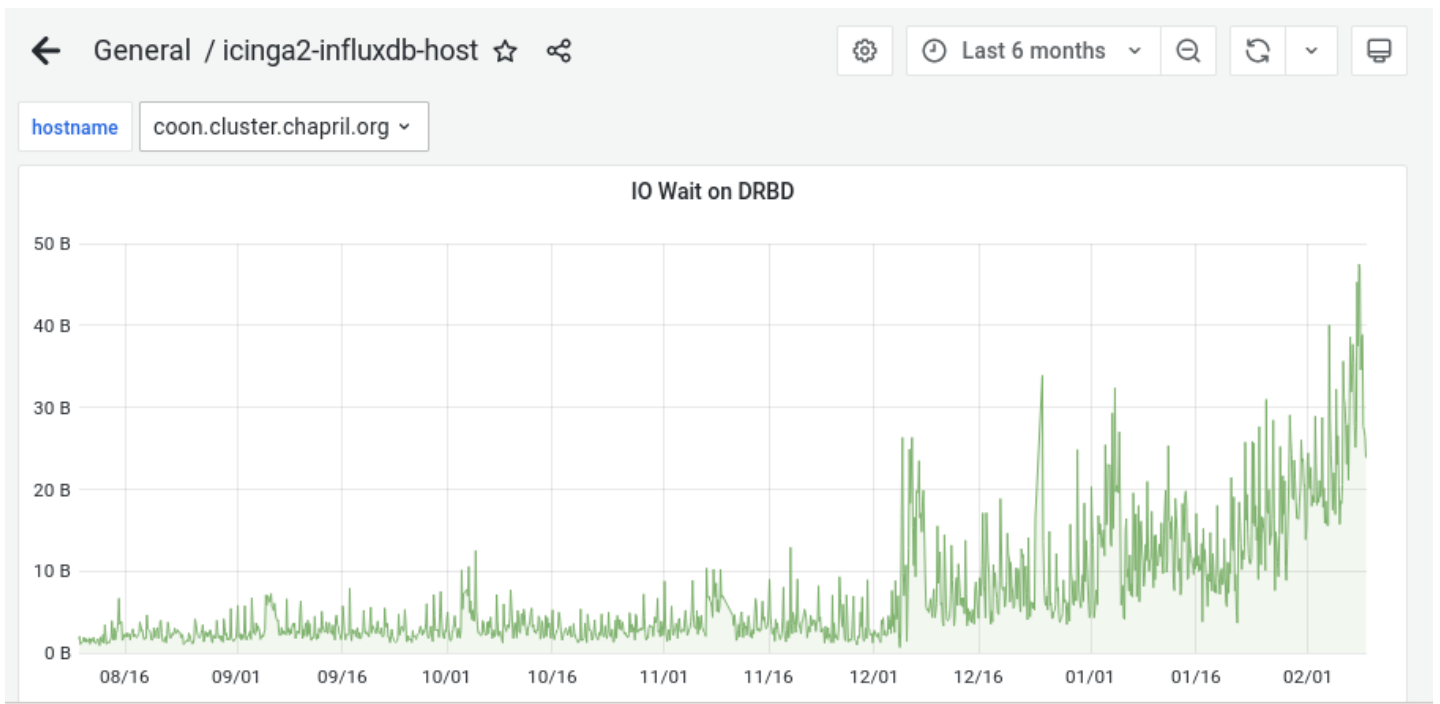
La commande iotop sur les hyperviseurs indique t'elle quelque chose de particulier ? Quelle(s) VM génère(nt) les IO ? Il n'y a pas d'autres IO que celles liées à DRBD ?

#3 - 09/02/2022 10:05 - pitchum .

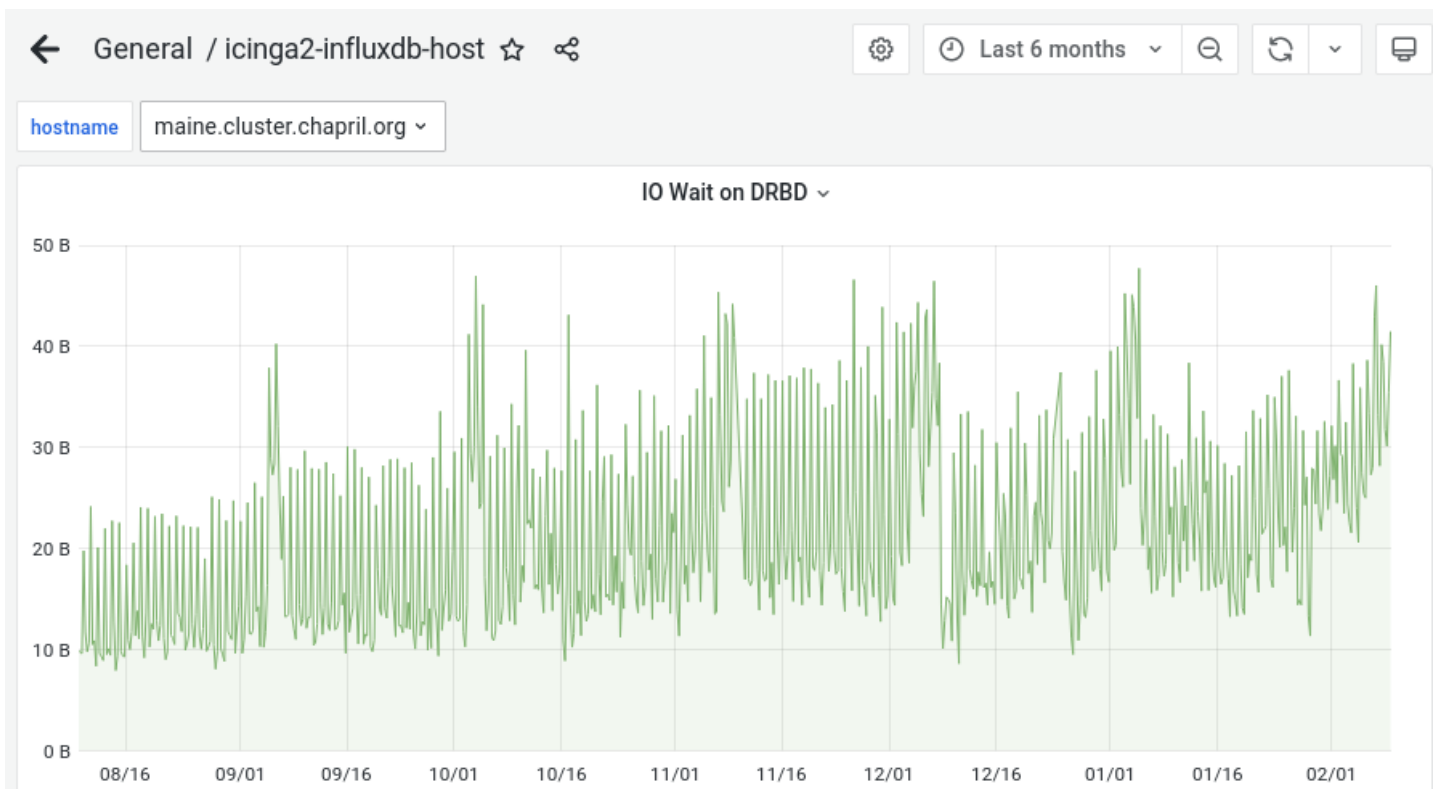
- Fichier `iowait_drbd_coon.png` ajouté
- Fichier `iowait_drbd_maine.png` ajouté
- Fichier `cpu_vms.png` ajouté

Les IO Wait DRBD de maine ont toujours été élevés, ceux de coon ont commencé à grimper en décembre.
(ne pas se fier à l'unité affichée sur les graphes, je n'ai pas trouvé comment la corriger)

[Métérologie pour coon](#)

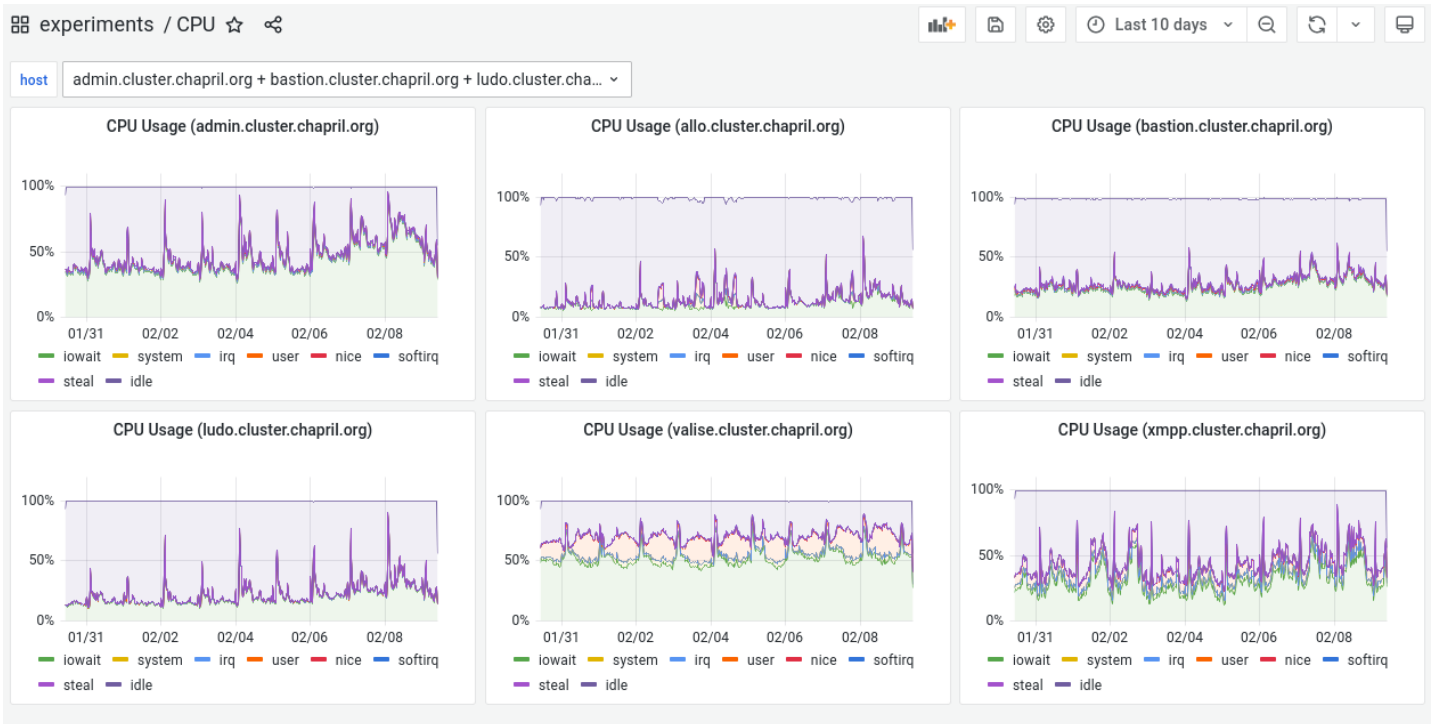


[Météologie pour maine](#)



Et concernant les VMs, nous n'avons de la météo CPU que depuis trop peu de temps pour pouvoir comparer avant/après.

Mais les taux d'iowait me paraissent quand même assez élevés, et ce en permanence, pas seulement pendant les backups. En particulier sur valise qui tourne sur maine.



#4 - 06/04/2022 05:02 - Pierre-Louis Bonicoli

- Version cible changé de Sprint 2022 février à Sprint 2022 avril

#5 - 06/04/2022 16:20 - Chris Mann

conclusion

On dirait qu'avec LVM on gagne en latence.
Pas sûr que ça justifie de remettre en cause l'usage de DRBD.

Il y a sûrement d'autres tests à faire.

Pas tout à fait. Le test est sur un fichier. Je soupçonne le problème étant un nombre important de petits fichiers (dans un nombre de comptes). La nature du volume de data est une piste parmi d'autres à explorer.

#6 - 12/05/2022 10:21 - Pierre-Louis Bonicoli

- Version cible changé de Sprint 2022 avril à Sprint 2022 mai

#7 - 09/06/2022 03:09 - Pierre-Louis Bonicoli

- Version cible changé de Sprint 2022 mai à Sprint 2022 juin

#8 - 07/07/2022 11:48 - Pierre-Louis Bonicoli

- Version cible changé de Sprint 2022 juin à Sprint 2022 juillet-août

#9 - 31/08/2022 02:22 - Pierre-Louis Bonicoli

- Version cible changé de Sprint 2022 juillet-août à Sprint 2022 septembre

#10 - 08/10/2022 15:38 - Pierre-Louis Bonicoli

- Version cible changé de Sprint 2022 septembre à Sprint 2022 novembre

#11 - 22/12/2022 06:05 - Pierre-Louis Bonicoli

- Version cible changé de Sprint 2022 novembre à Sprint 2022 décembre

#12 - 03/01/2023 22:48 - pitchum .

La situation s'est pas mal améliorée dernièrement, mais on à quand même encore régulièrement des WARN sur le IO DRBD.
Comme évoqué en réunion tout à l'heure, j'expérimente à nouveau de stopper collectd pendant quelques jours sur toutes les machines où il est

déployé, voir si ça diminue la fréquence de ces alertes.

```
./do.sh 'systemctl is-enabled collectd && systemctl stop collectd'
```

#13 - 04/01/2023 19:30 - Pierre-Louis Bonicoli

- Version cible changé de Sprint 2022 décembre à Sprint 2023 février

#14 - 08/03/2023 19:13 - Pierre-Louis Bonicoli

- Version cible changé de Sprint 2023 février à Sprint 2023 avril

#15 - 08/03/2023 19:14 - Pierre-Louis Bonicoli

- Version cible changé de Sprint 2023 avril à Sprint 2023 mars

#16 - 04/04/2023 22:23 - Pierre-Louis Bonicoli

- Version cible changé de Sprint 2023 mars à Sprint 2023 avril

#17 - 18/06/2023 23:54 - Pierre-Louis Bonicoli

- Version cible changé de Sprint 2023 avril à Sprint 2023 juin

#18 - 05/07/2023 09:12 - Pierre-Louis Bonicoli

- Version cible changé de Sprint 2023 juin à Sprint 2023 juillet-août

#19 - 06/09/2023 12:52 - Pierre-Louis Bonicoli

- Version cible changé de Sprint 2023 juillet-août à Sprint 2023 septembre

#20 - 04/10/2023 08:12 - Pierre-Louis Bonicoli

- Version cible changé de Sprint 2023 septembre à Sprint 2023 octobre

#21 - 05/12/2023 21:20 - Pierre-Louis Bonicoli

- Version cible changé de Sprint 2023 octobre à Sprint 2023 novembre

#22 - 05/12/2023 22:45 - Pierre-Louis Bonicoli

- Version cible changé de Sprint 2023 novembre à Sprint 2023 décembre

#23 - 20/01/2024 23:53 - Pierre-Louis Bonicoli

- Version cible Sprint 2023 décembre supprimé

Fichiers

iowait_drbd_maine.png	112 ko	09/02/2022	pitchum .
iowait_drbd_conon.png	67,2 ko	09/02/2022	pitchum .
cpu_vms.png	132 ko	09/02/2022	pitchum .