

openQA Infrastructure - action #57680

o3: optimize fs utilization and resize root disk (was: change /var/log to bind mount to prevent out-of-space)

2019-10-03 15:49 - okurz

Status:	Resolved	Start date:	2019-10-03
Priority:	Normal	Due date:	
Assignee:	nicksinger	% Done:	0%
Category:		Estimated time:	0.00 hour
Target version:	Ready		
Description			
Observation			
<pre>/dev/vda1 9.6G 7.6G 1.6G 83% /</pre>			
<pre># fdisk -l Disk /dev/vda: 10 GiB, 10737418240 bytes, 20971520 sectors Units: sectors of 1 * 512 = 512 bytes Sector size (logical/physical): 512 bytes / 512 bytes I/O size (minimum/optimal): 512 bytes / 512 bytes Disklabel type: dos Disk identifier: 0x025b0d73</pre>			
<pre>Device Boot Start End Sectors Size Id Type /dev/vda1 * 2048 20450744 20448697 9.8G 83 Linux /dev/vda2 20451328 20964824 513497 250.7M 82 Linux swap / Solaris</pre>			
<p>We do not have much headroom and more and more often we run into an out-of-space condition due to mistakes, not frequent enough logrotate or overflowing mail spool dir.</p>			
Suggestion			
<p>We could save more space by moving some directories to other partitions, e.g. /var/log to /space similar to osd where we have /var/log pointing to /srv/log on another partition. Or we simply ask the partition to be increased by engineering infrastructure. For now I tend to do ... both, actually.</p>			

History

#1 - 2019-10-03 18:23 - okurz

- Status changed from New to In Progress
- Assignee set to okurz
- Target version set to Current Sprint

```
systemctl stop systemd-journald apache2 openqa-webui openqa-websockets openqa-livehandler openqa-gru openqa-scheduler && rsync -aHP /var/log/ /space/log/ && mv /var/log/ /var/log.old/ && ln -s /space/log log && systemctl start systemd-journald apache2 openqa-webui openqa-websockets openqa-livehandler openqa-gru openqa-scheduler
```

But then openQA failed to start up because it was denied access to /var/log/openqa by apparmor. Apparently though there is no entry in /var/log/audit/audit.log because auditd did not log any event after 2019-09-20. I restarted auditd and it was fine again.

So for now I rolled back the change and enforced again the apparmor profile:

```
systemctl stop systemd-journald apache2 openqa-webui openqa-websockets openqa-livehandler openqa-gru openqa-scheduler && rsync -aHP /space/log/ /var/log.old/ && rm /var/log && mv /var/log.old/ /var/log/ && aa-enforce /etc/apparmor.d/usr.share.openqa.script.openqa && systemctl start systemd-journald apache2 openqa-webui openqa-websockets openqa-livehandler openqa-gru openqa-scheduler
```

we can now again more carefully look into apparmor denied actions before going to /space/log again.

With aa-logprof -f /var/log/audit/audit.log.1 there is the suggestion to cover

```
+ owner /space/log/openqa wk,
```

in the apparmor profile. I guess it would be better if we use a bind-mount instead of symlink for /space/log aka /var/log. WDYT?

#2 - 2019-10-04 10:48 - okurz

- Status changed from In Progress to Feedback

#3 - 2019-10-04 13:33 - cdywan

okurz wrote:

But then openQA failed to start up because it was denied access to /var/log/openqa by apparmor. Apparently though there is no entry in /var/log/audit/audit.log because auditd did not log any event after 2019-09-20. I restarted auditd and it was fine again.

/var/log/openqa was already in there, wasn't it?

Also, why did auditd break? Might be worth investigating that as well.

With aa-logprof -f /var/log/audit/audit.log.1 there is the suggestion to cover

```
+ owner /space/log/openqa wk,
```

in the apparmor profile. I guess it would be better if we use a bind-mount instead of symlink for /space/log aka /var/log. WDYT?

If /space/log is meant to fully replace /var/log I'd say a bind mount seems like the most natural choice to me. It would avoid apparmor changes or people looking in the wrong place.

What things do we actually want in /var/log btw that shouldn't be covered by systemd journal?

#4 - 2019-10-14 16:21 - okurz

- Subject changed from o3 root volume very limited, nearly out of space soon again to o3: change /var/log to bind mount to prevent out-of-space (was: o3 root volume very limited, nearly out of space soon again)

- Status changed from Feedback to Workable

- Assignee deleted (okurz)

- Priority changed from High to Normal

- Target version changed from Current Sprint to Ready

#5 - 2020-12-08 16:23 - mkittler

- Assignee set to mkittler

#6 - 2020-12-09 09:57 - mkittler

rsync -aHP /var/log /space/var and /space/var/log /var/log bind x-systemd.requires=/space,x-systemd.automount,bind 0 0 in /etc/fstab would technically do it but I'm not sure whether it is a good idea:

1. If the mounting does not work for some reason or happens too late we might end up with a system which doesn't boot anymore.
2. It seems we needed to take some extra effort to ensure no process is writing to /var/log anymore (see <https://serverfault.com/questions/55984/how-can-i-move-var-log-directory> and <https://www.suse.com/support/kb/doc/?id=000018399>).
3. Likely /space is backed up by a slower storage than / so maybe this has a bad impact on the performance (considering our services have very verbose logging).

#7 - 2020-12-09 10:18 - mkittler

- Assignee deleted (mkittler)

Note that the alternative mentioned almost everywhere is to extend the volume. Maybe that's actually an option? We would need to increase the disk vda, e.g. from 10 GiB to 20 GiB. Then we would have to remove vda2 which is currently only used for swap. Since it is only 250 MiB it currently doesn't do much for us anyways. Then we could extend vda1 which is ext4 so that shouldn't be a problem.

Considering the memory utilization of the system is quite low we could also make /tmp a tmpfs. That would also help because /tmp is the 2nd biggest top-level directory.

By the way:

```
/
 2,7 GiB [#####] /var

 2,1 GiB [##### ] /tmp
```

```

1,6 GiB [#####] /lib
1,5 GiB [#####] /usr
167,6 MiB [ ] /home
165,3 MiB [ ] /boot
145,8 MiB [ ] /srv
126,1 MiB [ ] /opt
15,3 MiB [ ] /etc
10,6 MiB [ ] /root
9,1 MiB [ ] /lib64
7,2 MiB [ ] /sbin
1,6 MiB [ ] /bin
1,4 MiB [ ] core
e 16,0 KiB [ ] /lost+found
16,0 KiB [ ] /config
e 4,0 KiB [ ] /selinux
e 4,0 KiB [ ] /mnt
> 0,0 B [ ] /sys
> 0,0 B [ ] /space
> 0,0 B [ ] /run
> 0,0 B [ ] /proc
> 0,0 B [ ] /dev
> 0,0 B [ ] /assets

```

```

/var
2,0 GiB [#####] /log
318,3 MiB [# ] /spool
265,2 MiB [# ] /cache
142,9 MiB [ ] /adm
3,9 MiB [ ] /tmp
3,8 MiB [ ] /lib

```

```

e 4,0 KiB [ ] /yp
e 4,0 KiB [ ] /opt
e 4,0 KiB [ ] /crash
4,0 KiB [ ] .updated
@ 0,0 B [ ] lock
@ 0,0 B [ ] mail
@ 0,0 B [ ] run

```

#8 - 2020-12-09 11:00 - okurz

- Priority changed from Normal to Urgent

I don't mind either way. But considering that / is currently at 95% usage this becomes urgent.

#9 - 2020-12-09 11:19 - nicksinger

- Assignee set to nicksinger

#10 - 2020-12-09 11:25 - nicksinger

First, I made /tmp a tmpfs with the following entry in /etc/fstab:

```
tmpfs /tmp tmpfs nodev,nosuid,size=5G 0 0
```

#11 - 2020-12-09 11:25 - nicksinger

- Status changed from Workable to In Progress

#12 - 2020-12-09 11:41 - mkittler

Thanks, that worked (with minimal disturbance). In my last cleanup I also removed old kernel versions so if we still want more space we could still delete e.g. the last two packages of:

```

martchus@ariel:~> rpm -qa kernel-default
kernel-default-5.3.18-lp152.57.1.x86_64
kernel-default-5.3.18-lp152.54.1.x86_64
kernel-default-5.3.18-lp152.50.1.x86_64
kernel-default-4.12.14-lp151.28.79.1.x86_64

```

#13 - 2020-12-09 12:16 - nicksinger

- Priority changed from Urgent to Normal

I've removed logfiles older than 31 days with `find /var/log/ -type f -mtime +31 -exec rm {} \;`. There was a lot of old stuff laying around there too. We should be fine to reduce priority for now:

```
/dev/vda1          9.6G  6.0G  3.2G  66% /
```

However we really need more space for / - openSUSE is just not designed to run on 10GB :)

#14 - 2020-12-09 12:36 - nicksinger

"#182162: Request for more disk space on ariel/openqa.opensuse.org" created for enginfra. I ask for a increase from 10GB to 20GB and how we can coordinate the enlargement of partitions and such (I've no clue how I could access ariel out-of-band)

#15 - 2020-12-09 12:46 - okurz

- Assignee changed from nicksinger to okurz

[nicksinger](#) thanks for your actions. I can see that we are back to 66% usage, that's good.

mkittler wrote:

Thanks, that worked (with minimal disturbance). In my last cleanup I also removed old kernel versions so if we still want more space we could still delete e.g. the last two packages of:

```
martchus@ariel:~> rpm -qa kernel-default
kernel-default-5.3.18-lp152.57.1.x86_64
kernel-default-5.3.18-lp152.54.1.x86_64
kernel-default-5.3.18-lp152.50.1.x86_64
kernel-default-4.12.14-lp151.28.79.1.x86_64
```

old kernel versions should be deleted automatically. Why it keeps 4.12.4 and also 3 versions for 5.3.18 I don't know. I will check that.

[nicksinger](#) great that you created a ticket. I see no problem with having EngInfra increase the virtual disk, then we reboot, then we call parted to extend the partition and then `resize2fs /`. If for any reason it's not possible to increase the disk size while the VM runs then EngInfra can power down the machine, increase the disk and boot again. We can then do the `partition+f/s` increase afterwards outside.

#16 - 2020-12-09 12:59 - mkittler

- Assignee changed from okurz to nicksinger

- Priority changed from Normal to Urgent

As mentioned in the previous comment we also need to take the swap partition into account when resizing because it is currently on the same disk as the root partition and comes directly after it.

#17 - 2020-12-09 12:59 - okurz

I checked `/etc/zypp/zypp.conf` which is configured for

```
zypp/zypp.conf:## Comma separated list of kernel packages to keep installed in parallel, if the
zypp/zypp.conf:## latest           - Keep kernel with the highest version number
zypp/zypp.conf:## latest-N        - Keep kernel with the Nth highest version number
zypp/zypp.conf:## running         - Keep the running kernel
zypp/zypp.conf:## oldest          - Keep kernel with the lowest version number (the GA kernel)
zypp/zypp.conf:## oldest+N       - Keep kernel with the Nth lowest version number
zypp/zypp.conf:##                 purge-kernels service (via /sbin/purge-kernels).
zypp/zypp.conf:## Default: Do not delete any kernels if multiversion = provides:multiversion(kernel) is set
zypp/zypp.conf:multiversion.kernels = latest,latest-1,latest-2,running
```

so it's correct that it keeps three versions of 5.3.18 . For 4.12.14 I just called `zypper rm -U kernel-default-4.12.14-lp151.28.79.1.x86_64` now.

back to [nicksinger](#)

#18 - 2020-12-09 13:01 - mkittler

- Assignee changed from nicksinger to okurz

- Priority changed from Urgent to Normal

Sorry for changing assignee and prio. Seems like my browser still had the old values in the form.

#19 - 2020-12-09 13:04 - okurz

- Assignee changed from okurz to nicksinger

- Priority changed from Normal to Urgent

that's ok. I wanted to set it back to nicksinger anyway :)

EDIT: I did a bit more cleanup:

```
zypper rm -u cups
rmdir /var/log/{ConsoleKit,cups,news,sa}
```

#20 - 2020-12-09 13:08 - nicksinger

- Assignee changed from nicksinger to okurz

- Priority changed from Urgent to Normal

mkittler wrote:

As mentioned in the previous comment we also need to take the swap partition into account when resizing because it is currently on the same disk as the root partition and comes directly after it.

I don't worry about swap too much but yeah. We have to mention this at least if infra handles it. Any objections of just using a swapfile instead of a dedicated partition?

#21 - 2020-12-09 13:11 - nicksinger

- Assignee changed from okurz to nicksinger

#22 - 2020-12-09 13:12 - okurz

nicksinger wrote:

Any objections of just using a swapfile instead of a dedicated partition?

no objection. I don't see the need though. It's only 256MB which makes sense and it can just as well still live on the same disk as /. There should be no problem to switch it off, enlarge partitions, create new swap and switch it on again

#23 - 2020-12-09 13:17 - nicksinger

- Status changed from In Progress to Blocked

Things left to do:

1. Wait for infra to enlarge disk
2. Remove swap partition
3. Enlarge root partition to full disk size
4. Create /swapfile
5. Enable swapfile
6. [...]
7. Profit :)

#24 - 2020-12-10 07:56 - nicksinger

- Status changed from Blocked to Feedback

- Assignee changed from nicksinger to okurz

gschlotter just increased the root disk. I removed the second partition for swap and enlarged root to 100% with the use of (s)fdisk:

```
ariel:~ # fdisk /dev/vda
```

```
Welcome to fdisk (util-linux 2.33.1).
Changes will remain in memory only, until you decide to write them.
Be careful before using the write command.
```

```
Command (m for help): p
Disk /dev/vda: 20 GiB, 21474836480 bytes, 41943040 sectors
Units: sectors of 1 * 512 = 512 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 512 bytes / 512 bytes
Disklabel type: dos
Disk identifier: 0x025b0d73
```

Device	Boot	Start	End	Sectors	Size	Id	Type
/dev/vda1	*	2048	20450744	20448697	9.8G	83	Linux
/dev/vda2		20451328	20964824	513497	250.7M	82	Linux swap / Solaris

Command (m for help): d
Partition number (1,2, default 2): 2

Partition 2 has been deleted.

Command (m for help): d
Selected partition 1
Partition 1 has been deleted.

Command (m for help): n
Partition type
p primary (0 primary, 0 extended, 4 free)
e extended (container for logical partitions)
Select (default p): p
Partition number (1-4, default 1):
First sector (2048-41943039, default 2048):
Last sector, +/-sectors or +/-size{K,M,G,T,P} (2048-41943039, default 41943039):

Created a new partition 1 of type 'Linux' and of size 20 GiB.
Partition #1 contains a ext4 signature.

Do you want to remove the signature? [Y]es/[N]o: N

Command (m for help): p

Disk /dev/vda: 20 GiB, 21474836480 bytes, 41943040 sectors
Units: sectors of 1 * 512 = 512 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 512 bytes / 512 bytes
Disklabel type: dos
Disk identifier: 0x025b0d73

Device	Boot	Start	End	Sectors	Size	Id	Type
/dev/vda1		2048	41943039	41940992	20G	83	Linux

Command (m for help): a
Selected partition 1
The bootable flag on partition 1 is enabled now.

Command (m for help): p
Disk /dev/vda: 20 GiB, 21474836480 bytes, 41943040 sectors
Units: sectors of 1 * 512 = 512 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 512 bytes / 512 bytes
Disklabel type: dos
Disk identifier: 0x025b0d73

Device	Boot	Start	End	Sectors	Size	Id	Type
/dev/vda1	*	2048	41943039	41940992	20G	83	Linux

Command (m for help): w

The partition table has been altered.
Syncing disks.

Afterwards the filesystem was enlarged and a new swapfile was created:

```
ariel:/ # dd if=/dev/zero of=/swapfile bs=1M count=512 status=progress
512+0 records in
512+0 records out
536870912 bytes (537 MB, 512 MiB) copied, 0.750336 s, 716 MB/s
ariel:/ # chmod 600 /swapfile
ariel:/ # mkswap /swapfile
Setting up swapspace version 1, size = 512 MiB (536866816 bytes)
no label, UUID=b41dd665-08e3-421b-8473-68b25dbfa104
```

I checked with `systemctl --type swap` what unit was previously used to mask it (`systemctl mask dev-vda2.swap`) and see if a unit for `/swapfile` was automatically created which was the case

#25 - 2020-12-10 07:59 - nicksinger

- Subject changed from *o3: change /var/log to bind mount to prevent out-of-space (was: o3 root volume very limited, nearly out of space soon again)* to *o3: optimize fs utilization and resize root disk (was: change /var/log to bind mount to prevent out-of-space)*
- Status changed from *Feedback* to *Resolved*
- Assignee changed from *okurz* to *nicksinger*