
Subject: Openvz7 ploop FS + snapshots = BLOAT
Posted by [nathan.brownrice](#) on Fri, 26 Apr 2024 16:27:39 GMT
[View Forum Message](#) <> [Reply to Message](#)

I'll keep this one short and sweet, and this is a long-shot, but if you know of a solution to this you probably already know what I'm talking about. We've been working through this ourselves with our production machines, and here's an example of the issue:

A VPS container, `df -h` inside the container shows it's using ~15G:

```
vzctl exec 54d7ac3a-3667-44b2-a4c9-67aa394284f0 df -h
Filesystem      Size  Used Avail Use% Mounted on
/dev/ploop16022p1 59G  15G  42G  26% /
none            4.0M  0 4.0M  0% /sys/fs/cgroup
none            3.0G  0 3.0G  0% /dev
tmpfs           3.0G  0 3.0G  0% /dev/shm
tmpfs           1.3G 124K 1.2G  1% /run
```

Same container, in the `/vz/root` (mounted filesystem) path on the host:

```
du -sh /vz/root/54d7ac3a-3667-44b2-a4c9-67aa394284f0/
15G  /vz/root/54d7ac3a-3667-44b2-a4c9-67aa394284f0/
```

Same container, in the `/vz/private` path:

```
du -sh /vz/private/54d7ac3a-3667-44b2-a4c9-67aa394284f0/
86G  /vz/private/54d7ac3a-3667-44b2-a4c9-67aa394284f0/
```

Most of it's stuck in the `/vz/private/54d7ac3a-3667-44b2-a4c9-67aa394284f0/root.hdd` images:

```
du -sh *
4.0K  DiskDescriptor.xml
0     DiskDescriptor.xml.lck
52G   root.hds
29G   root.hds.{4ef2408b-64ea-44d8-b8ca-d0b4b40145c9}
1.8G  root.hds.{c472b739-1cc5-484c-aca9-ced7bed74699}
24K   templates
```

We are running snapshots, and need to run snapshots. However after much experimentation it looks like any "bloat" is locked up in these snapshots and can never be cleared. Running `pcompact` does not free up the locked disk space and gives output saying to remove the

snapshots for better compacting:

```
Inspect 54d7ac3a-3667-44b2-a4c9-67aa394284f0
```

```
Inspect /vz/private/54d7ac3a-3667-44b2-a4c9-67aa394284f0/root.hdd/DiskDescriptor.xml
```

This ploop image contains snapshots. Trying to compact the last delta file.

For best compacting results, remove snapshots.

```
ploop=21118MB image=511MB data=7437MB balloon=19842MB
```

```
Rate: 0.0 (threshold=10)
```

```
Start compacting
```

```
ploop=21118MB image=511MB data=7437MB balloon=19842MB
```

```
End compacting
```

I understand the snapshots take up extra space, but what happens here is bloat over time. Files are added, snapshotted, the base root.hdd expands, but as long as snapshots are in place it will never contract even when files are removed and the running containers disk use is shrunk. Even with the nightly pcompact cron running. We've had quite a few containers that are using an actual ~10GB, with /vz/private dirs that bloat to 100+GB.

We've experimented with ensuring pcompact runs first, before any snapshot creation, but still the bloat gets locked in. The only way to free it is to delete all snapshots, then run pcompact, which shrinks the base root.hdd file down to close to what the container is actually using. Then as months go by things just slowly bloat back out.

Deleting snapshots every few months is not really a solution, so if anyone has run into this, or come up with a solution, I'm all ears!

Subject: Re: Openvz7 ploop FS + snapshots = BLOAT

Posted by [wsap](#) on Wed, 08 May 2024 17:13:38 GMT

[View Forum Message](#) <> [Reply to Message](#)

Interesting! I suspect since this is the same thing pcompact does (but here it's on a per container/ploop basis), it won't do anything different, but could be worth trying:

```
prl_disk_tool compact --hdd /vz/private/$CTID/root.hdd/
```

You might be able to feed it the specific snapshot paths (not certain). It does have a --force option you could try as well.

If these are online snapshots, I wonder if it's actually the system memory snapshot that's eating up a bunch of space (probably includes memory contents AND virtual/vswap). Granted that would be a significant amount of memory still. You could try the --skip-suspend option when creating the snapshot to find out.

Subject: Re: Openvz7 ploop FS + snapshots = BLOAT
Posted by [nathan.brownrice](#) on Fri, 17 May 2024 16:22:29 GMT
[View Forum Message](#) <> [Reply to Message](#)

Hey, thanks for replying!

I've tried that exact command, it's no different from running pcompact as far as I can tell. I've tried running it directly on the root.hdd snapshot files as well, and it does nothing. With the snapshots in place, the disk is "locked" into the base root.hdd file. Only if I delete all of the snapshots first can I free up the wasted disk space in the root.hdd file by running pcompact.

They are live snapshots, and to be honest I haven't tested with offline snapshots. Maybe worth a shot, but we have so many live containers we can't have downtime on them even overnight it would be sloppy.

Subject: Re: Openvz7 ploop FS + snapshots = BLOAT
Posted by [wsap](#) on Thu, 30 May 2024 20:29:42 GMT
[View Forum Message](#) <> [Reply to Message](#)

nathan.brownrice wrote on Fri, 17 May 2024 13:22 They are live snapshots, and to be honest I haven't tested with offline snapshots. Maybe worth a shot, but we have so many live containers we can't have downtime on them even overnight it would be sloppy.

Oh! To clarify, my understanding is that the --skip-suspend option shouldn't result in any downtime. It is supposed to *only* snapshot the disk, and not also the live memory contents. The idea being that if you restore a snapshot that has those memory contents (which they do by default), the container will resume *exactly* as it was (memory state included), whereas with --skip-suspend the disk will still be exactly as it was, but when the container comes online it'll be a fresh boot with clean memory contents.

No clue if that is contributing to your storage issue, but it's the only idea I've got left :)

Subject: Re: Openvz7 ploop FS + snapshots = BLOAT
Posted by [nathan.brownrice](#) on Tue, 02 Jul 2024 22:22:23 GMT
[View Forum Message](#) <> [Reply to Message](#)

Very interesting. We had gone through the ov7 man pages, particularly the snapshotting pages, and never saw the --skip-suspend option. We're using prlctl snapshot uuid to generate our snapshots as per the manual and there are really no options there.

That said, I just tried this and prlctl won't take the --skip-suspend option, but if I run a vzctl snapshot create with that flag it does work, and makes a snapshot super quickly.

I think we'll need to experiment with vzctl snapshot create vs prlctl snapshot create, with this flag, and see if there are any differences in disk usage / bloat.

Thanks for the tip, this is one thing I haven't tried.

Subject: Re: Openvz7 ploop FS + snapshots = BLOAT

Posted by [wsap](#) on Wed, 03 Jul 2024 00:34:18 GMT

[View Forum Message](#) <> [Reply to Message](#)

nathan.brownrice wrote on Tue, 02 July 2024 19:22 Thanks for the tip, this is one thing I haven't tried.

No prob! Hope it helps. Seems unlikely based on the sizes you encountered, but you never know :)
