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!