Subject: *KNOWN ISSUE* VZquota and Shared VEs with NFS
Posted by seanfulton on Tue, 29 May 2007 17:29:37 GMT
View Forum Message <> Reply to Message

I am having a problem using an NFS shared drive with OpenVZ. What I want to do is have a single NFS share that is mounted on three machines, weba, webb and webc.

weba will run vz 1-50, while webb will run vz51-100. webc will take-over for either a or b should a or b fail.

I followed the instructions in   http://wiki.openvz.org/HA_cluster_with_DRBD_and_Heartbeat#Installing_the_two_nodes

The problem I am having is #1, the quota info on weba is messed up--it shows minus disk space. #2, when I try to run a vz off of web2, it tells me:

Starting VE ...
vzquota : (error) can't lock quota file, some quota operations are performing for id 7003
vzquota on failed [7]

Even though vz7003 is not running on any of the machines.

It looks to me like an NFS lock problem, but none of the other files that are shared are having the same problem.

Any ideas or insight would be appreciated.

sean

Subject: Re: Shared VZs with NFS
Posted by seanfulton on Fri, 01 Jun 2007 18:58:38 GMT
View Forum Message <> Reply to Message

Any help on this? I've done some further testing and determined:
a) Everything works fine with Quota's turned off in the config, and
b) Each time I get the vzquota error, the following shows up in the logs:

Jun 1 09:45:09 web2 kernel: lockd: cannot monitor 209.2.69.12
Jun 1 09:45:09 web2 kernel: lockd: failed to monitor 209.2.69.12

This is the *only* time the error appears. I can read/write to the partition with absolutel no problems. But vzquota complains that the quota file is locked and something is still running on the vz. It seems pretty clear to me that this is an NFS/VZQUOTA issue.

Here is the mount command from my fstab if that helps:

fs-cluster1:/share/disk1 /mnt/disk1 nfs acl,udp,soft,bg,rsize=8192,wsize=8192 0 0

Subject: Re: Shared VZs with NFS
Posted by Vasily Tarasov on Mon, 04 Jun 2007 06:43:49 GMT
View Forum Message <> Reply to Message

Hello,

OpenVZ disk quota doesn't work on NFS filesystem:
http://bugzilla.openvz.org/show_bug.cgi?id=219

So, please, switch OVZ disk quota off in /etc/vz/vz.conf file.

HTH,
Vasily

Subject: Re: *KNOWN ISSUE* VZquota and Shared VEs with NFS
Posted by seanfulton on Mon, 04 Jun 2007 11:18:41 GMT
View Forum Message <> Reply to Message

I'm not so sure any of this is stable on NFS.

Over the weekend I switched to NFS4. I used both the 33 kernel and the older 31 kernel. Just installing packages into the VZ I repeatedly corrupted the RPM databases, so badly that I have had to completely erase the VE and start over. Three times in fact. I thought it might be bad RPMS, but that doesn't appear to be the case. I've done benchark tests writing to the NFS share and it seems to be stable.

I removed the NFS drive, re-installed everything VZ using local storage. Everything worked fine. Installed RPMS that had given me trouble on NFS with no problem.

Then I re-mounted the NFS drive onto /vz, copied all of the directories in, restarted, everything worked fine. Then this morning I installed the mysql-server RPM and it whacked the RPM database.

BTW: My NFS "Drive" is in /share/disk1. I tried both: ln -s /share/disk1/vz /vz and mount --bind /share/disk1/vz /vz and both seem to have the same results.

Is there any documentation or suggestions from anyone who has VZ running on NFS? This is turning out to be way harder than I thought it would be. Turning of VZQuota is not a big deal, but if it's going to corrupt data on the VE, I need to find another network filesystem or use another approach.

Any ideas?

Subject: Re: *KNOWN ISSUE* VZquota and Shared VEs with NFS
Posted by dev on Mon, 04 Jun 2007 12:20:26 GMT
View Forum Message <> Reply to Message

1. My personal IMHO - NFS is not suitable for HA solutions.
  a) Every release we add fixes for NFS over and over again (you can check git).
  b) RPM uses mmap'ed files, so probably the recent fixes in
    028stab034 can help - there was a corruption in mainstream/RHEL :/
    Was RPM database (when it was corrupted) placed on NFS in your cases?
   The list of problems with NFS is really huge and we do our best to work them out...
    BTW, do you use RHEL5 or mainstream kernel?
  b) NFS has one bottleneck - NFS server, so for HA you should HA your NFS server DRBD on
the other hand does mirroring, i.e. slave node and master are interchangeable.

2. host node quota (including vzquota) doesn't work with NFS, since quota is calculated/checked
on NFS server and client can't track quotas.

3. The article you reference is about DRBD. what has it to do with NFS?

Subject: Re: *KNOWN ISSUE* VZquota and Shared VEs with NFS
Posted by seanfulton on Mon, 04 Jun 2007 12:32:45 GMT
View Forum Message <> Reply to Message

The article about DRBD just showed me how to set the mounts and links to use VZ on a shared
drive.

I am using DRBD on two servers (A and B), but I want them both to actively share the storage
along with a third server that will act as a hot stand-by. So A and B share the DRBD drive via
NFS, A runs half the VEs, B runs the other half. The drive is also shared out to Machine C which
monitors *all* the VEs and will take over for any that fail.

The DRBD article only uses one active server and one stand-by. The DRBD-NFS-HA solution
theoretically lets me run two servers with one stand-by.

I was using the stock RHEL kernel. The entire /vz partition was on NFS, and it was the RPM
database inside the VE that was corrupted. So yes, it was on the NFS share. Is 34 out yet?

I looked into other ways of sharing network mirrored storage out to the network, specifically GFS,
but I couldn't find any good documentation or reports of wide-spread use. The DRBD-NFS-HA
solution seems to be in wide-spread use, so I chose that. Using VZ over DRBD provided a means
to "share" the VZ data, so I used that as well.

If you know of or can recommend a better setup, I would be glad to try it. The real object here is to
have multiple primary servers with one stand-by ready to take-over at any time.

sean

Subject: Re: *KNOWN ISSUE* VZquota and Shared VEs with NFS
Posted by seanfulton on Fri, 08 Jun 2007 03:18:51 GMT
View Forum Message <> Reply to Message

Any word on this NFS problem? NFS is unusable even for non VE activities. Also, you mention a 34 kernel for RHEL5, but I don't see it on the FTP site or anywhere else. Can you tell me where to get it??

sean

Subject: Re: *KNOWN ISSUE* VZquota and Shared VEs with NFS
Posted by dev on Fri, 08 Jun 2007 07:27:47 GMT
View Forum Message <> Reply to Message

1. it was replied that vzquota doesn't work with NFS due to NFS quota are calculated on server.

2. the problem with rpm on NFS will be checked.

3. Denis is currently working on messages like this:
Jun 1 09:45:09 web2 kernel: lockd: cannot monitor 209.2.69.12
Jun 1 09:45:09 web2 kernel: lockd: failed to monitor 209.2.69.12

So my estimation is 1-2 weeks for these problems.

034 test kernel will be available today-tomorrow.
released - later.

Subject: Re: *KNOWN ISSUE* VZquota and Shared VEs with NFS
Posted by den on Fri, 08 Jun 2007 08:04:22 GMT
View Forum Message <> Reply to Message

Unfortunately, NFSv4 is not supported to use in VE.

As for NFS "lockd" stalls - we are working with them now and they are a task with highest priority for us. There are many problems with them and a lot of issues.

Subject: Re: *KNOWN ISSUE* VZquota and Shared VEs with NFS

OK, well, I'm asking specifically about the NFS problem now. I've discovered strange behavior over an NFS mount having nothing to do with VZ. Writing with CPIO or copy returns errors of that there is no more disk space, when in fact there is *plenty* of space.

This works with the 33 kernel on both machines, and with the drive NFS3 mounted across the network so I suspect it's related to the other issue.

Also, have you any reports of people using GFS or OCFS2 for shared drives with VEs on them?

sean