Subject: Stale mtab entries following stop
Posted by laotse on Mon, 17 Aug 2009 08:35:21 GMT
View Forum Message <> Reply to Message

I have a couple of containers, which use the mount/unmount action scripts to mount physical partitions into the root file system, e.g.

mkdir -p /var/lib/vz/root/1005/srv/mail
mount -t xfs /dev/r5/mail /var/lib/vz/root/1005/srv/mail
umount /var/lib/vz/root/1005/srv/mail

However, following a vzctl stop the device is unmounted, but it is still listed in /etc/mtab. And each time I cycle the container, another instance pops up.

If I unmount the device manually with the VE stopped, umount moans that the device is not mounted, but all entries are removed at once.

Any idea, what might cause this?

Subject: Re: Stale mtab entries following stop
Posted by maratrus on Wed, 19 Aug 2009 13:39:52 GMT
View Forum Message <> Reply to Message

I can only guess what was the real reason but I think I know what it could be.

Most of VEs contain /etc/mtab as a symlink to /proc/mounts. You can manually do it for this particular VE.

A VE's mount script which is supposed to mount a partition put an appropriate information to /etc/mtab. When a VE is going to be stopped, one of the init-scripts have to umount all mounted filesystems gracefully. Perhaps it umounted all filesystems indeed but somehow missed to clear /etc/mtab.
IMHO you must not create an umount script because all filesystems have to be umounted by init-scripts (like it happen on a standalonw computer).
So, I'd recommend you to delete umount script and check if init-scripts umount all filesystem gracefully or you can mount a partition with a "-n" option.

Subject: Re: Stale mtab entries following stop
Posted by laotse on Thu, 27 Aug 2009 12:29:38 GMT
View Forum Message <> Reply to Message

maratrus wrote on Wed, 19 August 2009 15:39Most of VEs contain /etc/mtab as a symlink to /proc/mounts. You can manually do it for this particular VE.

The mtab of the VE is sane and indeed is the link you suggested. The issue happens with the mtab of the node!

Quote:A VE's mount script which is supposed to mount a partition put an appropriate information to /etc/mtab.

The mount script does not fiddle with the mtab of the VE for mtab is a link to /proc/mounts. The mount shown in the original post is listed as:
/dev/mapper/r5-mail /srv/mail xfs rw,attr2,nobarrier,noquota 0 0

Quote:When a VE is going to be stopped, one of the init-scripts have to umount all mounted filesystems gracefully. Perhaps it umounted all filesystems indeed but somehow missed to clear /etc/mtab.

Since the init script runs inside the VE, it cannot access the mtab of the node. Neither could an init script during startup access the block device for mounting. Mounting and unmounting has to be performed by the node.

But, if there is something like the VE trying to unmount the mount performed by the node, then we probably have some start for investigating what is actually happening.

Quote:or you can mount a partition with a "-n" option.

This is probably a work-around for this issue to avoid flooding mtab. But I'd really like to know how mount/unmount all running in the node's context can perform such a misbehaviour in the first place.

Thanks for the help,
 - lars.

Subject: Re: Stale mtab entries following stop
Posted by maratrus on Sun, 30 Aug 2009 10:37:03 GMT
View Forum Message <> Reply to Message

Hi,

does information about /dev/mapper/r5-mail is listed in /proc/mounts when VE is running?
If yes, then init scripts have to umount this mount point when VE is being stopped.
In this case a little problem seems to exist in OpenVZ. The record that is written to /etc/mtab on the HN is never deleted.

Subject: Re: Stale mtab entries following stop
Posted by laotse on Sun, 30 Aug 2009 13:25:42 GMT
View Forum Message <> Reply to Message

maratrus wrote on Sun, 30 August 2009 12:37does information about /dev/mapper/r5-mail is listed in /proc/mounts when VE is running?

Yes, it's listed in both the VE and the node.

To summarize your hypothesis this is what presumably happens:

The mount script running in the node's context mounts the drive and writes the record to the mtab of the node.
When the VE is brought down, the init scripts unmount the drive in VE context, since it appears in /proc/mounts = /etc/mtab of the VE.
When the unmount script in node context is run, the drive is already unmounted and umount fails, so the record stays in the mtab of the node.

What I still do not understand, is why all stale entries in the node's mtab disappear, when I manually unmount the drive while the VE is down. Can anyone imagine, what's the difference in the unmount script context and a root shell with the VE stopped?

Regards,
 - lars.

Subject: Re: Stale mtab entries following stop
Posted by maratrus on Sun, 30 Aug 2009 13:49:51 GMT
View Forum Message <> Reply to Message

Quote:
What I still do not understand, is why all stale entries in the node's mtab disappear, when I manually unmount the drive while the VE is down. Can anyone imagine, what's the difference in the unmount script context and a root shell with the VE stopped?

Do you want to say that in both cases the mount point actually is not mounted but "umount" command on the HN clears all entries in /etc/mtab in spite of being failed?

Subject: Re: Stale mtab entries following stop
Posted by laotse on Sun, 30 Aug 2009 16:15:43 GMT
View Forum Message <> Reply to Message

[quote title=maratrus wrote on Sun, 30 August 2009 15:49]Quote:Do you want to say that in both cases the mount point actually is not mounted but "umount" command on the HN clears all entries in /etc/mtab in spite of being failed?

Exactly! See this:node:~# cat /proc/mounts | grep 1002
/var/lib/vz/private/1002 /var/lib/vz/root/1002 simfs rw 0 0
/dev/mapper/base-proxy /var/lib/vz/root/1002/srv/proxy xfs rw,attr2,nobarrier,noquota 0 0
tmpfs /var/lib/vz/root/1002/lib/init/rw tmpfs rw,nosuid,mode=755 0 0

```
proc /var/lib/vz/root/1002/proc proc rw,nosuid,nodev,noexec 0 0
sysfs /var/lib/vz/root/1002/sys sysfs rw,nosuid,nodev,noexec 0 0
tmpfs /var/lib/vz/root/1002/dev/shm tmpfs rw,nosuid,nodev 0 0
devpts /var/lib/vz/root/1002/dev/pts devpts rw,nosuid,noexec,gid=5,mode=620 0 0
node:~# vzctl stop 1002
Stopping VE ...
VE was stopped
VE is unmounted
node:~# grep 1002 /etc/mtab
/dev/mapper/base-proxy /var/lib/vz/root/1002/srv/proxy xfs rw 0 0
/dev/mapper/base-proxy /var/lib/vz/root/1002/srv/proxy xfs rw 0 0
node:~# grep 1002 /proc/mounts
node:~# umount /var/lib/vz/root/1002/srv/proxy
umount: /dev/mapper/base-proxy: not mounted
umount: /var/lib/vz/root/1002/srv/proxy: not found
umount: /dev/mapper/base-proxy: not mounted
umount: /var/lib/vz/root/1002/srv/proxy: not found
umount: /dev/mapper/base-proxy: not mounted
umount: /var/lib/vz/root/1002/srv/proxy: not found
node:~# grep 1002 /proc/mounts
node:~# grep 1002 /etc/mtab
node:~# vzctl start 1002
Starting VE ...
VE is mounted
Adding IP address(es): XXX.XXX.XXX.XXX
Setting CPU units: 1000
Configure meminfo: 262144
Set hostname: vename.local
File resolv.conf was modified
VE start in progress...
node:~#
```