Subject: [SOLVED] live migration with extra mounts in VE
Posted by frodo on Sun, 02 Dec 2007 15:38:22 GMT

View Forum Message <> Reply to Message

Hi,

I'm trying to do the following:

live migration of a VE which has an extra partition mounted.

Live migration without the extra partition works like a charm.
I then added access to /dev/sda and /dev/sda1 to the VE with
the "vzctl set <VEID> --devices 8:0:rw --devices 8:1:rw --save
In the VE I can access the block device, create a filesystem on
it, and mount it.

Since /dev/sda is an iscsi lun on shared storage, it is also available on the node I want to migrate
to, with the same major and minor device numbers.

I am not using a bind mount, since lvie migration has been documented not to be possible in that
way.

I have also tried using lvm and using a logical volume instead of a primary partition on a disk.
Results are the same.

Any ideas ?

The reason for this is the ability to use SAN space for the data and local storage for the VE OS.
This would cut down on the syncing time needed, and make fast SAN disks available to the VE's.

Kind regards.

Subject: Re: live migration with extra mounts in VE
Posted by Vasily Tarasov on Mon, 03 Dec 2007 05:25:33 GMT

View Forum Message <> Reply to Message

Hello,

what exact errors do you see on the screen? What is in your dmesg?

Thank you,
Vasily.

Subject: Re: live migration with extra mounts in VE
Posted by frodo on Mon, 03 Dec 2007 13:31:44 GMT

Hey Vasily,

The output of the vzmigrate command:

(dumping seems to work fine, undumping on the other node produces an error, in bold)

[root@sv-ard-drp8 dev]# vzmigrate -r no -v --online 192.168.254.30 120 OPT:-r
OPT:-v
OPT:--online
OPT:192.168.254.30
Starting online migration of VE 120 on 192.168.254.30
OpenVZ is running...
   Loading /etc/vz/vz.conf and /etc/vz/conf/120.conf files
   Check IPs on destination node:
Preparing remote node
   Copying config file
120.conf   100% 1933   1.9KB/s 00:00
Saved parameters for VE 120
   Creating remote VE root dir
   Creating remote VE private dir
Initializing remote quota
   Quota init
   Turning remote quota on
Syncing private
Live migrating VE
   Suspending VE
Setting up checkpoint...
     suspend...
     get context...
Checkpointing completed succesfully
   Dumping VE
Setting up checkpoint...
     join context..
     dump...
Checkpointing completed succesfully
   Copying dumpfile
dump.120   100% 1581KB 1.5MB/s 00:00
   Syncing private (2nd pass)
Syncing 2nd level quota
   Dumping 2nd level quota
   Copying 2nd level quota
quotadump.120   100% 21   0.0KB/s 00:00
   Load 2nd level quota
   Undumping VE
Restoring VE ...
Starting VE ...
vzquota : (warning) Quota is running for id 120 already

VE is mounted
        undump...
Setting CPU units: 1000
Setting devices
Configure meminfo: 49152
Configure veth devices: veth120.0
Initializing interface veth120.0 for VE120.
Adding interface veth120.0 to the bridge br1.
Error: undump failed: No such file or directory
Restoring failed:
-2 mounting /data/ftp ext3 40000000
rst_namespace: -2
VE start failed
Stopping VE ...
VE was stopped
VE is unmounted
Error: Failed to undump VE
Resuming...
vzquota : (error) Quota is not running for id 120
[root@sv-ard-drp8 dev]#

The dmesg output on the remote node gives me these error messages:

CPT ERR: df8a3800,120 :-2 mounting /data/ftp ext3 40000000
CPT ERR: df8a3800,120 :rst_namespace: -2

I am working with a logical volume created with lvm2.
This logical volume is visible on both nodes (I can mount it
simultaneously on both nodes, which is very dangerous, i know

I have the following in my <VEID>.conf for this volume:

DEVNODES="sda:rw mapper/vg_120-lv_ftpdata:rw"

sda is the iscsi LUN that is shared between the nodes, and which contains the logical volume.

Regards,

frodo

Subject: Re: live migration with extra mounts in VE
Posted by Andrey Mirkin on Tue, 04 Dec 2007 06:47:23 GMT
View Forum Message <> Reply to Message

Hello!

Can you please scpecify what kernel version you are using?

CPT ERR: df8a3800,120 :-2 mounting /data/ftp ext3 40000000

Flag 0x40000000 is stands for external mount, so it seems that during checkpointing this partition was recognized as external partition (mounted from VE0).

We had a bug in this place, but this bug was fixed long time ago in 028stab040.1 kernel.

Subject: Re: live migration with extra mounts in VE
Posted by frodo on Tue, 04 Dec 2007 08:17:22 GMT
View Forum Message <> Reply to Message

Andrey,

I'm using the 2.6.18-8.1.15.el5.028stab049.1 RHEL5 kernel.

I mounted the LV from inside the VE, this is the /proc/mounts
for VE 120. I added it to /etc/fstab on the VE, to have it
mounted at VE start.

```
vzmig120:/# cat /proc/mounts
simfs / simfs rw 0 0
tmpfs /lib/init/rw tmpfs rw,nosuid 0 0
proc /proc proc rw,nosuid,nodev,noexec 0 0
sysfs /sys sysfs rw,nosuid,nodev,noexec 0 0
tmpfs /dev/shm tmpfs rw,nosuid,nodev 0 0
devpts /dev/pts devpts rw,nosuid,noexec 0 0
/dev/mapper/vg_120-lv_ftpdata /data/ftp ext3 rw,data=ordered 0 0
```

But since there is only one kernel running I presume it should
be visible on the HN too.

Is it even possible to have live migration in this way ?
It would greatly help in getting OpenVZ adopted here

The /proc/mounts from the HN tells me indeed the mount is visible in VE0:

```
[root@sv-ard-drp8 dev]# cat /proc/mounts
/dev/root / ext3 rw,data=ordered 0 0
/dev /dev tmpfs rw 0 0
/proc /proc proc rw 0 0
/sys /sys sysfs rw 0 0
/proc/bus/usb /proc/bus/usb usbfs rw 0 0
devpts /dev/pts devpts rw 0 0
/dev/cciss/c0d0p3 /data ext3 rw,data=ordered 0 0
```

tmpfs /dev/shm tmpfs rw 0 0
none /proc/sys/fs/binfmt_misc binfmt_misc rw 0 0
sunrpc /var/lib/nfs/rpc_pipefs rpc_pipefs rw 0 0
/etc/auto.misc /misc autofs rw,fd=6,pgrp=6555,timeout=300,minproto=5,maxproto=5, indirect 0 0
-hosts /net autofs rw,fd=11,pgrp=6555,timeout=300,minproto=5,maxproto=5,indirec t 0 0
/vz/private/110 /data/vz/root/110 simfs rw 0 0
tmpfs /data/vz/root/110/lib/init/rw tmpfs rw,nosuid 0 0
proc /data/vz/root/110/proc proc rw,nosuid,nodev,noexec 0 0
sysfs /data/vz/root/110/sys sysfs rw,nosuid,nodev,noexec 0 0
tmpfs /data/vz/root/110/dev/shm tmpfs rw,nosuid,nodev 0 0
devpts /data/vz/root/110/dev/pts devpts rw,nosuid,noexec 0 0
/vz/private/500 /data/vz/root/500 simfs rw 0 0
tmpfs /data/vz/root/500/lib/init/rw tmpfs rw,nosuid 0 0
proc /data/vz/root/500/proc proc rw,nosuid,nodev,noexec 0 0
sysfs /data/vz/root/500/sys sysfs rw,nosuid,nodev,noexec 0 0
tmpfs /data/vz/root/500/dev/shm tmpfs rw,nosuid,nodev 0 0
devpts /data/vz/root/500/dev/pts devpts rw,nosuid,noexec 0 0
/vz/private/120 /data/vz/root/120 simfs rw 0 0
tmpfs /data/vz/root/120/lib/init/rw tmpfs rw,nosuid 0 0
proc /data/vz/root/120/proc proc rw,nosuid,nodev,noexec 0 0
sysfs /data/vz/root/120/sys sysfs rw,nosuid,nodev,noexec 0 0
tmpfs /data/vz/root/120/dev/shm tmpfs rw,nosuid,nodev 0 0
devpts /data/vz/root/120/dev/pts devpts rw,nosuid,noexec 0 0
/dev/mapper/vg_120-lv_ftpdata /data/vz/root/120/data/ftp ext3 rw,data=ordered 0 0
[root@sv-ard-drp8 dev]#

Regards,

frodo

Subject: Re: live migration with extra mounts in VE
Posted by Andrey Mirkin on Wed, 05 Dec 2007 10:58:36 GMT
View Forum Message <> Reply to Message

Hello!

I have prepared debug patch for kernel. Can you please try to checkpoint VE with it and post here logs from /var/log/messages.

Thanks for your help in advance!

File Attachments
1) diff-cpt-external-mnt-debug-2, downloaded 365 times

Subject: Re: live migration with extra mounts in VE

Posted by [frodo](#) on Thu, 06 Dec 2007 14:02:43 GMT

Hey Andrey,

I compiled and installed the kernel with the debug patch.
Upgraded to 2.6.18-53.el5.028stab051.1 as well.

The error message I get (in dmesg) is:

CPT ERR: f5b3a000,120 :external mount: mnt rootfs / rootfs, mnt_ub 0, exec_ub 120
CPT ERR: f5b3a000,120 :external mount: mnt /vz/private/120 / simfs, mnt_ub 0, exec_ub 120

Hopefully this helps you on the way.

Regards,

frodo

Subject: Re: live migration with extra mounts in VE
Posted by [Andrey Mirkin](#) on Thu, 06 Dec 2007 15:10:22 GMT

Hello,

Hmm... It is strange...
I have expected to see some logs about /data/ftp, but you don't have them.
Can you please give me an access to your node to investigate this issue. If you can, then please
send me credentials via PM.

BTW, can you please just try to do following on one node:
1. vzctl chkpnt VEID
2. vzctl restore VEID

Is this bug reproducable with this scenario?

Subject: Re: live migration with extra mounts in VE
Posted by [frodo](#) on Thu, 06 Dec 2007 15:44:50 GMT

Andrey,

I'm looking into the access to the machine. I hope the security
engineers allow it.

The bug is indeed reproducible with vzctl chkpnt <VEID> and vzctl restore <VEID>.
I get the following in dmesg:

CPT ERR: f5b90800,120 :external mount: mnt rootfs / rootfs, mnt_ub 0, exec_ub 120
CPT ERR: f5b90800,120 :external mount: mnt /vz/private/120 / simfs, mnt_ub 0, exec_ub 120

The mount does appear automatically when I start the VE, as can
be seen below. It just doesn't survive a chkpnt-restore cycle.

[root@sv-ard-drp8 ~]# vzctl start 120
Starting VE ...
VE is mounted
Setting CPU units: 1000
Setting devices
Configure meminfo: 49152
Set hostname: vzmig120.test.ardatis.com
Configure veth devices: veth120.0
Initializing interface veth120.0 for VE120.
Adding interface veth120.0 to the bridge br1.
VE start in progress...
[root@sv-ard-drp8 ~]# vzctl enter 120
entered into VE 120
vzmig120:/# cat /proc/mounts
simfs / simfs rw 0 0
tmpfs /lib/init/rw tmpfs rw,nosuid 0 0
proc /proc proc rw,nosuid,nodev,noexec 0 0
sysfs /sys sysfs rw,nosuid,nodev,noexec 0 0
tmpfs /dev/shm tmpfs rw,nosuid,nodev 0 0
devpts /dev/pts devpts rw,nosuid,noexec 0 0
/dev/mapper/vg_120-lv_ftpdata /data/ftp ext3 rw,data=ordered 0 0
vzmig120:/# logout
exited from VE 120
[root@sv-ard-drp8 ~]# vzctl chkpnt 120
Setting up checkpoint...
	suspend...
	dump...
	kill...
VE is unmounted
Checkpointing completed succesfully
[root@sv-ard-drp8 ~]# vzctl restore 120
Restoring VE ...
Starting VE ...
VE is mounted
	undump...
Setting CPU units: 1000
Setting devices
Configure meminfo: 49152
Configure veth devices: veth120.0

Initializing interface veth120.0 for VE120.
Adding interface veth120.0 to the bridge br1.
Error: undump failed: No such file or directory
Restoring failed:
Error: -2 mounting /data/ftp ext3 40000000
Error: rst_namespace: -2
VE start failed
Stopping VE ...
VE was stopped
VE is unmounted
[root@sv-ard-drp8 ~]#

Subject: Re: live migration with extra mounts in VE
Posted by Andrey Mirkin on Fri, 07 Dec 2007 06:46:56 GMT
View Forum Message <> Reply to Message

Hello,

Can you please upload somewhere dump file of this VE and provide a link. Or you can just send it
to me. I'll send you my e-mail via PM.
Usualy it is stored in /vz/dump/Dump.VEID.

But to be sure you can prepare it with following command:
vzctl chkpnt VEID --dumpfile=<path>

Subject: [SOLVED] Re: live migration with extra mounts in VE
Posted by frodo on Tue, 18 Dec 2007 11:19:57 GMT
View Forum Message <> Reply to Message

Aaarggh, my bad.

Putting the mount in /etc/fstab of the VE has been the culprit.

When I create a <VEID>.mount file in /etc/sysconfig/vz-scripts
all is well, even the live migration with a shared iSCSI disk
with LVM on top.
So don't put mounts in /etc/fstab, use the mount scripts on the
HN if you want live migration.

Regards,

frodo