Subject: DRBD inside VE (again)
Posted by bogomolov on Mon, 22 Oct 2007 00:16:08 GMT
View Forum Message <> Reply to Message

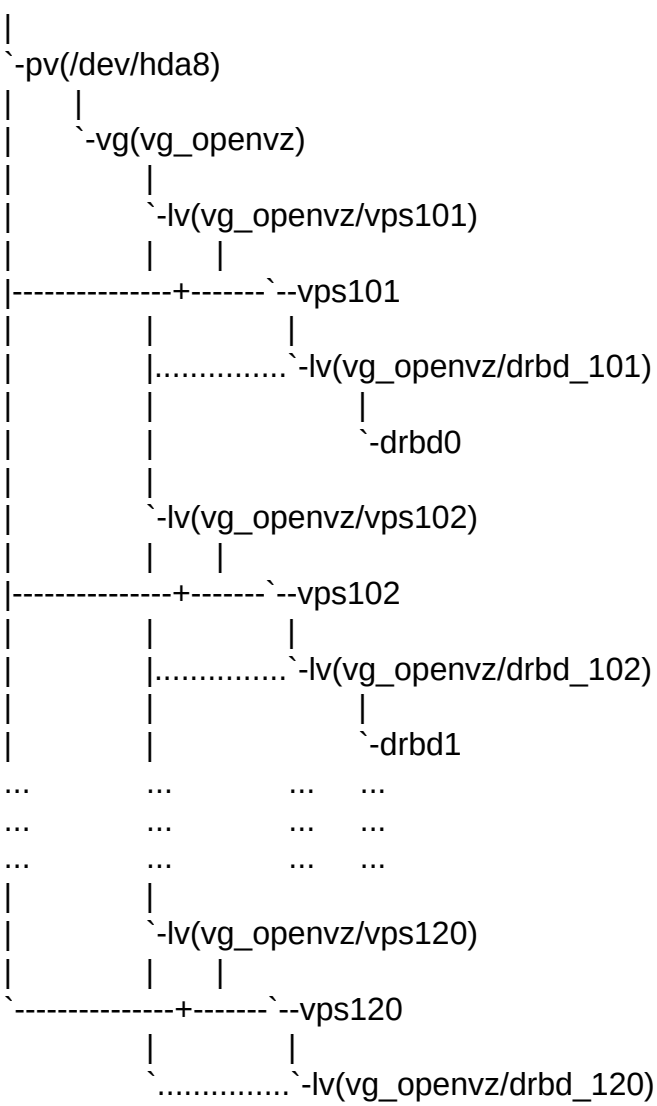
Hi all

Im in trouble now - my layout of HA appear dont work with OpenVZ

What i want: Each VPS running at top of a LV volume, and a DRBD layer running inside each VPS at top of another LV volume (to sync data with another VPS in another HN). Confuse? me too XD - maybe this help (or confuse more...)

Update[1] - Changed the ASCII map to most close real layout

Clue: PV (Physical Volume);VG (Volume Group);LV (Logical Volume).



```
                    |
                    `-drbd19
```

On the another HN, same layout. Someone probably will ask "WTF this dude need this stupid/complex thing, instead of this - http://wiki.openvz.org/HA_cluster_with_DRDD_and_Heartbeat" . The answer is: i want both HNs working and still have some flexibility to make backups (LVM snapshots), modify VPS volumes and leave the memory for programs, not for infrastructure/overhead (like Xen/VMware).

What (supposed) i need to run DRBD inside VPSs?

1 - HN kernel with DRBD support (OK)

Debian GNU/Linux 4.0 \n \l

Linux hn01 2.6.18-028stab031.1-openvz-smp #1 SMP Fri May 4 15:27:21 CEST 2007 i686 GNU/Linux
ii drbd0.7-module-2.6.18-028stab031.1-openvz-smp
ii drbd0.7-module-source
ii drbd0.7-utils

2 - Module present and loaded (OK)

root@hn01:~# locate drbd.ko;cat /etc/modules;lsmod|grep drbd;cat /proc/drbd
/lib/modules/2.6.18-028stab031.1-openvz-smp/kernel/drivers/block/drbd.ko
/etc/modules: kernel modules to load at boot time.
#
This file contains the names of kernel modules that should be loaded
at boot time, one per line. Lines beginning with "#" are ignored.

loop
drbd minor_count=2 #just 2 instances for now

drbd   143220 0
version: 0.7.21 (api:79/proto:74)
SVN Revision: 2326 build by root@hn01, 2007-10-14 12:31:14
 0: cs:Unconfigured
 1: cs:Unconfigured

3 - Devices/Devnodes included in config of VPS (OK)

root@hn01:~# tail -n4 /etc/vz/conf/*99*.conf

```
==> /etc/vz/conf/199.conf <==
IP_ADDRESS="192.168.0.199"
NAMESERVER="200.165.132.155 200.149.55.140"
NAME="vps199"
DEVNODES="drbd1:rw vg_openvz/drbd_199:rw mapper/vg_openvz-drbd_199:rw"

==> /etc/vz/conf/99.conf <==
IP_ADDRESS="192.168.0.99"
NAMESERVER="200.165.132.155 200.149.55.140"
NAME="vps99"
DEVNODES="drbd0:rw vg_openvz/drbd_99:rw mapper/vg_openvz-drbd_99:rw"
```

4 - LV volumes up (OK)

```
root@hn01:~# lvdisplay |grep drbd -B1 -A11
  --- Logical volume ---
  LV Name                /dev/vg_openvz/drbd_99
  VG Name                vg_openvz
  LV UUID                tBPivk-69zn-HvW1-mHh5-Wo3h-Rjo1-SnJnBa
  LV Write Access        read/write
  LV Status              available
  # open                 0
  LV Size                500,00 MB
  Current LE             125
  Segments               1
  Allocation             inherit
  Read ahead sectors     0
  Block device           253:1
--
  --- Logical volume ---
  LV Name                /dev/vg_openvz/drbd_199
  VG Name                vg_openvz
  LV UUID                quiRp4-WB91-WKII-iekB-rDYt-UqaW-hF9kSq
  LV Write Access        read/write
  LV Status              available
  # open                 0
  LV Size                500,00 MB
  Current LE             125
  Segments               1
  Allocation             inherit
  Read ahead sectors     0
  Block device           253:2
```

5 - DRBD tools working in VPS (OK)

```
root@vps99:/# dpkg -l|grep drbd|awk '{print($1,$2)}'
```

```
ii drbd0.7-utils
root@vps199:/# dpkg -l|grep drbd|awk '{print($1,$2)}'
ii drbd0.7-utils
```

So, with all this i configure DRBD in VPSs: (im trying make a drbd link between 2 VPSs because im only with one HN now)

```
root@vps99:/# cat /etc/drbd.conf
resource r0 {
  protocol C;
  incon-degr-cmd "echo '!DRBD! pri on incon-degr' | wall ; sleep 60 ; halt -f";

  startup {
    degr-wfc-timeout 120;
  }

  net {
    on-disconnect reconnect;
  }

  disk {
    on-io-error   pass_on;
  }

  syncer {
    rate 30M;
    group 1;
    al-extents 257;
  }

  on vps99 {
    device     /dev/drbd0;
    disk       /dev/vg_openvz/drbd_99;
    address    192.168.0.99:7788;
    meta-disk  internal;
  }

  on vps199 {
    device     /dev/drbd1;
    disk       /dev/vg_openvz/drbd_199;
    address    192.168.0.199:7788;
    meta-disk  internal;
  }
root@vps99:/# cat /etc/hosts
127.0.0.1  vps99  localhost localhost.localdomain
192.168.0.99    vps99
192.168.0.199   vps199
```

.:The problem

Firing-up DRDB:

```
root@vps99:/# /etc/init.d/drbd start
Starting DRBD resources:    [ d0 ioctl(,SET_DISK_CONFIG,) failed: Operation not permitted

cmd /sbin/drbdsetup /dev/drbd0 disk /dev/vg_openvz/drbd_99 internal -1 --on-io-error=pass_on
failed!
```

WTH! All devices are ok?

```
root@vps99:/# ll /dev/drbd0 /dev/vg_openvz /dev/mapper/
brw-rw---- 1 root root 147, 0 Oct 21 19:51 /dev/drbd0

/dev/mapper/:
total 0
brw-r----- 1 root root 253, 1 Oct 21 21:32 vg_openvz-drbd_99

/dev/vg_openvz:
total 0
brw-r----- 1 root root 253, 1 Oct 21 22:38 drbd_99
```

Module working inside VPS

```
root@vps99:/# cat /proc/drbd
version: 0.7.21 (api:79/proto:74)
SVN Revision: 2326 build by root@hn01, 2007-10-14 12:31:14
 0: cs:Unconfigured
 1: cs:Unconfigured
```

I have access to LV volume? - yup

```
root@vps99:/# mkfs.ext3 /dev/vg_openvz/drbd_99
mke2fs 1.40-WIP (14-Nov-2006)
Filesystem label=
OS type: Linux
Block size=1024 (log=0)
Fragment size=1024 (log=0)
128016 inodes, 512000 blocks
25600 blocks (5.00%) reserved for the super user
First data block=1
Maximum filesystem blocks=67633152
```

63 block groups
8192 blocks per group, 8192 fragments per group
2032 inodes per group
Superblock backups stored on blocks:
        8193, 24577, 40961, 57345, 73729, 204801, 221185, 401409

Writing inode tables: done
Creating journal (8192 blocks): done
Writing superblocks and filesystem accounting information: done

This filesystem will be automatically checked every 23 mounts or
180 days, whichever comes first. Use tune2fs -c or -i to override.

root@vps99:/# mkdir /tmp/test;mount /dev/vg_openvz/drbd_99 /tmp/tt/;mount
tmpfs on /lib/init/rw type tmpfs (rw,nosuid,mode=0755)
proc on /proc type proc (rw,noexec,nosuid,nodev)
sysfs on /sys type sysfs (rw,noexec,nosuid,nodev)
tmpfs on /dev/shm type tmpfs (rw,nosuid,nodev)
devpts on /dev/pts type devpts (rw,noexec,nosuid,gid=5,mode=620)
simfs on / type simfs (rw)
/dev/vg_openvz/drbd_99 on /tmp/tt type ext3 (rw)

Ha! so - the problem are in DRBD, right? no
DRBD starts fine on HN...

root@hn01:~# drbdsetup /dev/drbd1 disk /dev/vg_openvz/drbd_199 internal -1
--on-io-error=detach
root@hn01:~# drbdsetup /dev/drbd0 disk /dev/vg_openvz/drbd_99 internal -1 --on-io-error=detach

root@hn01:~# cat /proc/drbd
version: 0.7.21 (api:79/proto:74)
SVN Revision: 2326 build by root@hn01, 2007-10-14 12:31:14
 0: cs:StandAlone st:Secondary/Unknown ld:Inconsistent
    ns:0 nr:0 dw:0 dr:0 al:0 bm:0 lo:0 pe:0 ua:0 ap:0
 1: cs:StandAlone st:Secondary/Unknown ld:Inconsistent
    ns:0 nr:0 dw:0 dr:0 al:0 bm:0 lo:0 pe:0 ua:0 ap:0

So... anyone have some light on this?

Sorry for long post

Subject: Re: DRBD inside VE (again)
Posted by piavlo on Mon, 22 Oct 2007 07:54:51 GMT
View Forum Message <> Reply to Message

bogomolov wrote on Mon, 22 October 2007 02:16

root@vps99:/# /etc/init.d/drbd start
Starting DRBD resources: [d0 ioctl(,SET_DISK_CONFIG,) failed: Operation not permitted

cmd /sbin/drbdsetup /dev/drbd0 disk /dev/vg_openvz/drbd_99 internal -1 --on-io-error=pass_on
failed!

I suspect you might need to enable an aprorpiate cabablility inside the VE with
vzctl set VEID --capability capability:on --save
But i am not sure which capability is needed by DRBD for the operation, try the
following:net_admin, net_raw, sys_rawio, mknod.

Subject: Re: DRBD inside VE (again)
Posted by bogomolov on Wed, 24 Oct 2007 11:02:18 GMT
View Forum Message <> Reply to Message

Hi piavlo

I enabled various capability's for my vps's, but still with same error, see:

CAPABILITY's witch i enabled:

root@hn01:~# tail -n4 /etc/vz/conf/*99*.conf
==> /etc/vz/conf/199.conf <==
NAMESERVER="200.165.132.155 200.149.55.140"
NAME="vps199"
DEVNODES="drbd1:rw vg_openvz/drbd_199:rw mapper/vg_openvz-drbd_199:rw"
CAPABILITY="CHOWN:on DAC_OVERRIDE:on DAC_READ_SEARCH:on SETUID:on
NET_ADMIN:on NET_RAW:on SYS_MODULE:on SYS_RAWIO:on SYS_PACCT:on MKNOD:on
VE_ADMIN:on "
==> /etc/vz/conf/99.conf <==
NAMESERVER="200.165.132.155 200.149.55.140"
NAME="vps99"
DEVNODES="drbd0:rw vg_openvz/drbd_99:rw mapper/vg_openvz-drbd_99:rw "
CAPABILITY="CHOWN:on DAC_OVERRIDE:on DAC_READ_SEARCH:on SETUID:on
NET_ADMIN:on NET_RAW:on SYS_MODULE:on SYS_RAWIO:on SYS_PACCT:on MKNOD:on
VE_ADMIN:on "

the problem persists...

root@hn01:~# vzctl exec2 99 /etc/init.d/drbd start
Starting DRBD resources: [d0 ioctl(,SET_DISK_CONFIG,) failed: Operation not permitted

cmd /sbin/drbdsetup /dev/drbd0 disk /dev/vg_openvz/drbd_99 internal -1 --on-io-error=pass_on failed!

checking if CAPABILITY's are really on

```
root@hn01:~# vzctl exec2 99 lcap
Current capabilities: 0xFFFFFEFF
   0) *CAP_CHOWN                1) *CAP_DAC_OVERRIDE
   2) *CAP_DAC_READ_SEARCH      3) *CAP_FOWNER
   4) *CAP_FSETID               5) *CAP_KILL
   6) *CAP_SETGID               7) *CAP_SETUID
   8)  CAP_SETPCAP              9) *CAP_LINUX_IMMUTABLE
  10) *CAP_NET_BIND_SERVICE    11) *CAP_NET_BROADCAST
  12) *CAP_NET_ADMIN           13) *CAP_NET_RAW
  14) *CAP_IPC_LOCK            15) *CAP_IPC_OWNER
  16) *CAP_SYS_MODULE          17) *CAP_SYS_RAWIO
  18) *CAP_SYS_CHROOT          19) *CAP_SYS_PTRACE
  20) *CAP_SYS_PACCT           21) *CAP_SYS_ADMIN
  22) *CAP_SYS_BOOT            23) *CAP_SYS_NICE
  24) *CAP_SYS_RESOURCE        25) *CAP_SYS_TIME
  26) *CAP_SYS_TTY_CONFIG      27) *CAP_MKNOD
  28) *CAP_LEASE               29) *CAP_AUDIT_WRITE
  30) *CAP_AUDIT_CONTROL
    * = Capabilities currently allowed
```

have another tip for me?

Subject: Re: DRBD inside VE (again)
Posted by seanfulton on Sat, 23 Jun 2012 18:49:50 GMT
View Forum Message <> Reply to Message

Did you ever get this working?

I am interested in doing the same thing.

sean

Subject: Re: DRBD inside VE (again)
Posted by mustardman on Tue, 26 Jun 2012 02:42:34 GMT
View Forum Message <> Reply to Message

You can only do it between nodes. Not between VPS's. Maybe with ploop it will be possible since

it is a block device. Not sure about that but I'm sure you can't do it between OpenVZ VPS's running in their own container file space.