Subject: /proc/swaps needed for Oracle 11g RAC
Posted by smaliz on Fri, 24 Apr 2009 17:48:59 GMT
View Forum Message <> Reply to Message

Hi Folks:
Please pardon my ignorance as i am DBA and not that intimate with the workings of openvz kernel.

I have published my efforts before with the following howto

http://wiki.openvz.org/Oracle_10g_VE

I am hoping to do the same for an Oracle 11g RAC.

I have been trying to implement an Oracle 11g cluster (RAC) on openvz. I have created two containers and all the requirements for
Oracle RAC are being met (as reported by Oracle Cluster verification Utility).
There is the /proc/swaps problem though which i have been unable to find a work around.
As have been reported in several posts like this one

(http://forum.openvz.org/index.php?t=tree&goto=26398&
&srch=%2Fproc%2Fswaps#msg_26398)

Oracle 11g installer simply refuses to start without /proc/swaps.
Doing an strace on the installer reveals its doing a cat on /proc/swaps.
I tried fooling the installer by replacing /bin/cat with a script that returned the contents of /proc/swaps from the HN, but that did not work either.

I was hoping someone here can help me create a temporary /proc/swaps solution so that the installer would finish.

Any help is greatly appreciated.
thx

Subject: Re: /proc/swaps needed for Oracle 11g RAC
Posted by kir on Sat, 25 Apr 2009 14:53:22 GMT
View Forum Message <> Reply to Message

Your workaround didn't work probably because you have copy-pasted the /proc/swaps content. This replaces tabs with spaces -- so in order to fix it properly do not use copy/paste.

Proper solution is to use RHEL5-based (i.e. stable) kernel >= 028stab060.2 and vzctl >= 3.0.24. Since vzctl-3.0.24 is not yet released you have to use the one from git repo.

It adds a new parameter called --swappages, which is described in vzctl(man page. Here's the link to man page diff:

http://git.openvz.org/?p=vzctl;a=blobdiff;f=man/vzctl.8;h=01c9d59139f5f778ce5eb23a2b0934610bda4603;hp=92d49e3d3351f4d64f22f1e043e886e83436f247;hb=6b819fb60634a109200b7a749aa40d595fcd4895;hpb=4059ec910609421087bb9143d1de8424a81ad4d0

Subject: Re: /proc/swaps needed for Oracle 11g RAC
Posted by bpolarsk on Tue, 28 Apr 2009 13:35:16 GMT
View Forum Message <> Reply to Message

I am just curious : How did you did for the shared disk?

Subject: Re: /proc/swaps needed for Oracle 11g RAC
Posted by smaliz on Tue, 28 Apr 2009 18:10:31 GMT
View Forum Message <> Reply to Message

Hi Kir:
Thank you much for the help.

I tried the workaround with tab instead of spaces.
Doing an strace, it seems the installer is moving past /proc/swaps check. However i am still getting a segmentation fault from the Oracle installer.

I am onto trying the solution with vzctl-3.0.24.
I have updated the kernel to 2.6.18-92.1.18.el5.028stab060.8ent and trying to grab vzctl from git.

However when i run this command

git clone http://git.openvz.org/pub/vzctl

It eventually fails with the following

---
got 9cdc0376d15031c0734b9640fb5b77c2fb02e402
error: File 019b021a6f92238ef4b22d67b21d0e679378d88c has bad hash
error: Unable to find 019b021a6f92238ef4b22d67b21d0e679378d88c under
http://git.openvz.org/pub/vzctl
Cannot obtain needed tree 019b021a6f92238ef4b22d67b21d0e679378d88c
while processing commit 413b39c870543afe64c85613b42f66e165cff9b8.
fatal: Fetch failed.

Any ideas what may be the issue.

thankyou again for your help.

Subject: Re: /proc/swaps needed for Oracle 11g RAC
Posted by smaliz on Tue, 28 Apr 2009 18:20:23 GMT
View Forum Message <> Reply to Message

Hi bpolarsk:
I have not yet been able to install 11g, so i dont know if its actually going to work. But here is what i have done

I am trying to use RAW volumes created from LVM for the OCR and Voting disk and couple of physical disks (block devices) for the ASM data disks.

Data Disk on HN
==========
/dev/ida/c0d1

Flash Disk on HN
==========
/dev/ida/c0d2

OCR and Voting Disk Logical Volumes on HN
==========
/dev/oravg/ocr
/dev/oravg/vote

Check their major and minor numbers

ls -l /dev/oravg/ocr /dev/oravg/vote /dev/ida/c0d1 /dev/ida/c0d2

ls -l /dev/oravg/ocr /dev/oravg/vote /dev/ida/c0d1 /dev/ida/c0d2
brw-r----- 1 root disk 72, 16 Nov 5 11:56 /dev/ida/c0d1
brw-r----- 1 root disk 72, 32 Nov 5 11:56 /dev/ida/c0d2
lrwxrwxrwx 1 root root   21 Nov 5 16:45 /dev/oravg/ocr -> /dev/mapper/oravg-ocr
lrwxrwxrwx 1 root root   22 Nov 5 16:46 /dev/oravg/vote -> /dev/mapper/oravg-vote

ls -l /dev/mapper/oravg-ocr
brw-rw---- 1 root disk 253, 9 Nov 5 16:45 /dev/mapper/oravg-ocr

ls -l /dev/mapper/oravg-vote
brw-rw---- 1 root disk 253, 10 Nov 5 16:46 /dev/mapper/oravg-vote

Add these devices to the two containers

vzctl set 3001 --devices b:72:32:rw --devices b:72:16:rw --devices b:253:9:rw --devices b:253:10:rw --save

vzctl exec 3001 mknod /dev/ocr b 253 9
vzctl exec 3001 mknod /dev/vote b 253 10

vzctl exec 3001 mknod /dev/data_dg1_d1 b 72 16
vzctl exec 3001 mknod /dev/flash_dg1_d1 b 72 32

vzctl set 3002 --devices b:72:32:rw --devices b:72:16:rw --devices b:253:9:rw --devices b:253:10:rw --save

vzctl exec 3002 mknod /dev/ocr b 253 9
vzctl exec 3002 mknod /dev/vote b 253 10

vzctl exec 3002 mknod /dev/data_dg1_d1 b 72 16
vzctl exec 3002 mknod /dev/flash_dg1_d1 b 72 32

Subject: Re: /proc/swaps needed for Oracle 11g RAC
Posted by smaliz on Tue, 28 Apr 2009 19:06:24 GMT
View Forum Message <> Reply to Message

Hi Kir:

I was able to download the vzctl tree by using the git protocol instead of http:
However the downloaded tree seems in-complete.
I dont see the configure script in the vzctl directory.
Here is what i see

acinclude.m4 bin configure.ac etc INSTALL man pathsubst.am src
autogen.sh ChangeLog COPYING include Makefile.am paths.am scripts vzctl.spec

Did i do something incorrect while fetching via git?

Also i noticed vzctl-3.0.24 is not available even via git.
The vzctl.spec file shows 3.0.23.

thx

Subject: Re: /proc/swaps needed for Oracle 11g RAC
Posted by kir on Wed, 29 Apr 2009 05:24:30 GMT
View Forum Message <> Reply to Message

git is not very reliable over http:// (although I have just checked it, i.e. did a clean git clone http://git.openvz.org/pub/vzctl and it worked for me).

git:// protocol is more reliable, try using git-clone git://git.openvz.org/pub/vzctl

---

Subject: Re: /proc/swaps needed for Oracle 11g RAC
Posted by kir on Wed, 29 Apr 2009 05:28:50 GMT
View Forum Message <> Reply to Message

Quote:I dont see the configure script in the vzctl directory.

this is because configure is auto-generated from configure.ac. You have to run autogen.sh (and have auto* tools installed).

Quote:Also i noticed vzctl-3.0.24 is not available even via git.
The vzctl.spec file shows 3.0.23.

It's because the current git is somewhere in between 3.0.23 and 3.0.24... and the last thing we do -- just before the release -- is changing version.

You can actually take vzctl-3.0.23-1.src.rpm, then unpack it, pick just one patch (http://git.openvz.org/?p=vzctl;a=commitdiff_plain;h=6b819fb6 0634a109200b7a749aa40d595fcd4895) and add it into vzctl.spec and rebuild it, if you know how to do that (nothing really fancy).

Subject: Here is my procedure to create virutal disk for CT
Posted by bpolarsk on Wed, 29 Apr 2009 09:41:46 GMT
View Forum Message <> Reply to Message

I can creates a virtual shared disk, but did not tested if it works for RAC :

On the Vz Server (Ip host vz 192.168.89.7):

yum whatprovides tgtd
--> scsi-target-utils.x86_64

yum install scsi-target-utils.x86_64 # to be found on the Centos repository, not on the RedHat one.
yum install iscsi-initiator-utils.x86_64 # also needed

#Start isci daemon:
/etc/init.d/iscsi start
--> Setting up iSCSI targets: iscsiadm: No records found! [OK]

create an empty file to be the virtual disk:
dd if=/dev/zero of=/vz/shared_disk/shared_disk01.vz bs=1M count=512
create a iscsi group
tgtadm --lld iscsi --op new --mode target --tid 1 -T shared_disk01.vz # this is group name
a a disk to this iscsi group
tgtadm --lld iscsi --op new --mode logicalunit --tid 1 --lun 1 -b /vz/shared_disk/shared_disk01.vz

```
tgtadm --lld iscsi --op bind --mode target --tid 1 -I 192.168.89.7
# show what we have done:
tgtadm --lld iscsi --op show --mode target
iscsiadm --mode discovery --type sendtargets --portal 192.168.89.7
--> # 192.168.89.7:3260,1 shared_disk01.vz

iscsiadm --mode node --targetname shared_disk01.vz --portal 192.168.89.7:3260 --login
--> Logging in to [iface: default, target: shared_disk01.vz, portal: 192.168.89.7,3260]
--> Login to [iface: default, target: shared_disk01.vz, portal: 192.168.89.7,3260]: successful

[bdglo0001vz]:/vz/shared_disk> fdisk -l
--> Disk /dev/cciss/c0d0: 73.3 GB, 73372631040 bytes
--> 255 heads, 63 sectors/track, 8920 cylinders
--> Units = cylinders of 16065 * 512 = 8225280 bytes
-->
-->           Device Boot      Start         End      Blocks   Id  System
--> /dev/cciss/c0d0p1   *           1          19      152586   83  Linux
--> /dev/cciss/c0d0p2              20        1294    10241437+  82  Linux swap / Solaris
--> /dev/cciss/c0d0p3            1295        8920    61255845   8e  Linux LVM
-->
--> Disk /dev/sda: 20 MB, 20971520 bytes
--> 1 heads, 40 sectors/track, 1024 cylinders
--> Units = cylinders of 40 * 512 = 20480 bytes
--> Disk /dev/sda doesn't contain a valid partition table

fdsik /dev/sda   # create here a partition primary on the whole disk, then quit
partprobe /dev/sda
mke2fs /dev/sda1


# Now I mount this Virtual disk on a directory on the node, then I export this directory into 2
Container
# Fist mount the Virt. Disk on the VZ host:
mount /dev/sda1 /vz/shared_disk/mnt1
# Next shared this mount with both target CT 140001 and CT 140003:

mount --bind /vz/shared_disk/mnt1 /vz/root/140001/app
mount --bind /vz/shared_disk/mnt1 /vz/root/140003/app

#On container 140001:
 vzctl enter 140001
--> entered into CT 140001
--> [root@ap01 /]# df
--> Filesystem           1K-blocks      Used Available Use% Mounted on
--> simfs                20314748   9902384   9363788  52% /
--> /dev/sda1               19807       173     18611   1% /app
--> [root@ap01 /]# ls -l app
--> total 13
```

--> -rw-r--r-- 1 root root   12 Apr 29 11:25 hello
--> drwx------ 2 root root 12288 Apr 29 11:00 lost+found

#On container 140003:
vzctl enter 140003
--> entered into CT 140003
--> [root@ap03 app]# df
--> Filesystem   1K-blocks   Used Available Use% Mounted on
--> simfs   20314748   9902384   9363788 52% /
--> /dev/sda1   19807   173   18611   1% /app
--> [root@ap03 app]# ls -l
--> total 13
--> -rw-r--r-- 1 root root   12 Apr 29 11:25 hello
--> drwx------ 2 root root 12288 Apr 29 11:00 lost+found

Subject: Re: /proc/swaps needed for Oracle 11g RAC
Posted by smaliz on Thu, 30 Apr 2009 22:01:04 GMT
View Forum Message <> Reply to Message

Hi Kir:
Ok vzctl is now compiled and installed. When i tried to use the vzctl set command or even any other vzctl command i am getting this message

lockdir is not set
Unable to lock container

I verified that the /etc/vz/vz.conf file has the LOCDIR specified and it is a valid directory..

Is vzctl looking for the vz.conf file somewhere else, if so how can i tell it the location.

thx

Subject: Re: /proc/swaps needed for Oracle 11g RAC
Posted by kir on Fri, 01 May 2009 12:10:51 GMT
View Forum Message <> Reply to Message

When running ./configure you have to supply a few params, otherwise you'll end up with vzctl installed at (and using) /usr/local. See vzctl.spec for params to ./configure.

Yet better, build an RPM (just run 'make rpms'); do not forget to change Release: field in spec file.

Subject: Re: /proc/swaps needed for Oracle 11g RAC
Posted by smaliz on Fri, 01 May 2009 21:39:46 GMT
View Forum Message <> Reply to Message

Thx a bunch Kir.
Recompiled vzctl with the prefix /usr and it seems to be working fine.

Ran the following two commands to setup swappages

vzctl set 3001 --meminfo pages:4147520 --save
vzctl set 3001 --swappages 1147520 --save

Oracle Installer now seems much happier. It is not throwing a segmentation fault but complaining that swap space is zero.

When i do a cat on /proc/swaps i only see this.

[root@simorcl03a /]# cat /proc/swaps
Filename Type Size Used Priority
---

free command also returns zero swap space.

Is that expected?

thx

Subject: Re: /proc/swaps needed for Oracle 11g RAC
Posted by rongcc on Tue, 26 May 2009 07:33:44 GMT
View Forum Message <> Reply to Message

There is BUG.

src/lib/ub.c

```
void add_ub_limit(struct ub_struct *ub, int res_id, unsigned long *limit)
{
#define ADD_UB_PARAM(name, id)                          \
if (res_id == id) {                                     \
        ub->name = limit;                               \
        return;                                         \
}

        ADD_UB_PARAM(kmemsize, PARAM_KMEMSIZE)
```

```
	ADD_UB_PARAM(lockedpages, PARAM_LOCKEDPAGES)
	ADD_UB_PARAM(privvmpages, PARAM_PRIVVMPAGES)
	ADD_UB_PARAM(shmpages, PARAM_SHMPAGES)
	ADD_UB_PARAM(numproc, PARAM_NUMPROC)
	ADD_UB_PARAM(physpages, PARAM_PHYSPAGES)
	ADD_UB_PARAM(vmguarpages, PARAM_VMGUARPAGES)
	ADD_UB_PARAM(oomguarpages, PARAM_OOMGUARPAGES)
	ADD_UB_PARAM(numtcpsock, PARAM_NUMTCPSOCK)
	ADD_UB_PARAM(numflock, PARAM_NUMFLOCK)
	ADD_UB_PARAM(numpty, PARAM_NUMPTY)
	ADD_UB_PARAM(numsiginfo, PARAM_NUMSIGINFO)
	ADD_UB_PARAM(tcpsndbuf, PARAM_TCPSNDBUF)
	ADD_UB_PARAM(tcprcvbuf, PARAM_TCPRCVBUF)
	ADD_UB_PARAM(othersockbuf, PARAM_OTHERSOCKBUF)
	ADD_UB_PARAM(dgramrcvbuf, PARAM_DGRAMRCVBUF)
	ADD_UB_PARAM(numothersock, PARAM_NUMOTHERSOCK)
	ADD_UB_PARAM(numfile, PARAM_NUMFILE)
	ADD_UB_PARAM(dcachesize, PARAM_DCACHESIZE)
	ADD_UB_PARAM(numiptent, PARAM_NUMIPTENT)
	ADD_UB_PARAM(avnumproc, PARAM_AVNUMPROC)
#undef ADD_UB_PARAM
}
```

The swappages don't set, add ADD_UB_PARAM(swappages, PARAM_SWAPPAGES).

Subject: Re: /proc/swaps needed for Oracle 11g RAC
Posted by kir on Tue, 26 May 2009 10:31:02 GMT
View Forum Message <> Reply to Message

Thanks for the report. Fixed in git:

http://git.openvz.org/?p=vzctl;a=commit;h=593a9fbe97f26c1c87 2f7cfea9f252bc6e9db0e6
