Subject: Hung Tasks on NFS (maybe not a OpenVZ Problem) - How to forcefully kill a container ?
Posted by svensirk on Fri, 30 Mar 2012 11:03:13 GMT
View Forum Message <> Reply to Message

Hi everyone,

I am running a lot of CTs with their roots located on an nfs share.
Once in a while it happens that a process gets stuck which I fear has
something to do with the nfs mount.
See the dmesg out below.
The problem now is that I can't kill this process anymore.
This results into beeing unable to stop the CT running this process.
vzctl stop <CTID> runs into a timeout.
It is totally impossible to kill the process - The only solution is a
reboot of the Host-System.

Is there a way to forcefully kill the CT ?
In this case I don't care if the process remains running.
I just want the rest of the CT to be stopped so I can start the CT again.

Here is the dmes output:

[194043.649945] INFO: task which:810615 blocked for more than 120 seconds.
[194043.650077] "echo 0 > /proc/sys/kernel/hung_task_timeout_secs"
disables this message.
[194043.650274] which   D ffff882f74146d50   0 810615 682640
125 0x00000084
[194043.650281] ffff8825ba732f98 0000000000000086 0000000000000017
ffff8825ba732fc0
[194043.650293] 0000000000000000 0000000000000000 ffff8824d0186d70
0000000000000007
[194043.650301] ffff8825ba732f88 ffff882f74147308 ffff8825ba733fd8
ffff8825ba733fd8
[194043.650308] Call Trace:
[194043.650318] [<ffffffff81122480>] ? sync_page+0x0/0x50
[194043.650325] [<ffffffff814e6b73>] io_schedule+0x73/0xc0
[194043.650330] [<ffffffff811224bd>] sync_page+0x3d/0x50
[194043.650334] [<ffffffff814e73da>] __wait_on_bit_lock+0x5a/0xc0
[194043.650339] [<ffffffff81122457>] __lock_page+0x67/0x70
[194043.650345] [<ffffffff81095550>] ? wake_bit_function+0x0/0x50
[194043.650351] [<ffffffff8113ac72>] ? pagevec_lookup+0x22/0x30
[194043.650357] [<ffffffff8113cc8e>] truncate_inode_pages_range+0x43e/0x450
[194043.650397] [<ffffffffa0314b80>] ? nfs_dq_delete_inode+0x0/0xd0 [nfs]
[194043.650406] [<ffffffffa03a7c7f>] ? vzquota_data_unlock+0x2f/0x40 [vzdquota]
[194043.650421] [<ffffffffa0314b80>] ? nfs_dq_delete_inode+0x0/0xd0 [nfs]
[194043.650426] [<ffffffff8113ccb5>] truncate_inode_pages+0x15/0x20
[194043.650441] [<ffffffffa0314b9f>] nfs_dq_delete_inode+0x1f/0xd0 [nfs]

```
[194043.650447] [<ffffffff811ac736>] generic_delete_inode+0xd6/0x1c0
[194043.650451] [<ffffffff811ac885>] generic_drop_inode+0x65/0x80
[194043.650456] [<ffffffff811ab3e2>] iput+0x62/0x70
[194043.650467] [<ffffffffa02fa48e>] nfs_dentry_iput+0x3e/0x60 [nfs]
[194043.650472] [<ffffffff811a7c1b>] dentry_iput+0x8b/0x110
[194043.650476] [<ffffffff811a7d9c>] d_kill+0x3c/0x70
[194043.650480] [<ffffffff811a9533>] dput+0xa3/0x1d0
[194043.650485] [<ffffffff8119e30a>] path_put+0x1a/0x40
[194043.650497] [<ffffffffa0301982>] __put_nfs_open_context+0xc2/0xf0 [nfs]
[194043.650510] [<ffffffffa0301a90>] put_nfs_open_context+0x10/0x20 [nfs]
[194043.650524] [<ffffffffa0311029>] nfs_commitdata_release+0x29/0x40 [nfs]
[194043.650537] [<ffffffffa03116c1>] nfs_commit_release+0x31/0x40 [nfs]
[194043.650564] [<ffffffffa029dde7>] rpc_release_calldata+0x17/0x20 [sunrpc]
[194043.650576] [<ffffffffa029e090>] rpc_free_task+0x50/0x80 [sunrpc]
[194043.650588] [<ffffffffa029e115>] rpc_final_put_task+0x55/0x60 [sunrpc]
[194043.650600] [<ffffffffa029e150>] rpc_do_put_task+0x30/0x40 [sunrpc]
[194043.650612] [<ffffffffa029e190>] rpc_put_task+0x10/0x20 [sunrpc]
[194043.650626] [<ffffffffa03105c1>] nfs_initiate_commit+0x131/0x190 [nfs]
[194043.650640] [<ffffffffa0311a89>] nfs_commit_inode+0x199/0x250 [nfs]
[194043.650646] [<ffffffff8100bb0e>] ? common_interrupt+0xe/0x13
[194043.650658] [<ffffffffa02fe426>] nfs_release_page+0x86/0xa0 [nfs]
[194043.650662] [<ffffffff81121800>] try_to_release_page+0x30/0x60
[194043.650668] [<ffffffff8113fc77>] shrink_page_list+0x817/0x9f0
[194043.650673] [<ffffffff81140227>] shrink_inactive_list+0x3d7/0xa40
[194043.650678] [<ffffffff81141308>] shrink_zone+0x5d8/0x9d0
[194043.650684] [<ffffffff81063c4b>] ? dequeue_task_fair+0x12b/0x130
[194043.650689] [<ffffffff8114240d>] __zone_reclaim+0x22d/0x2f0
[194043.650694] [<ffffffff8113eb30>] ? isolate_pages_global+0x0/0x520
[194043.650698] [<ffffffff811425e7>] zone_reclaim+0x117/0x150
[194043.650703] [<ffffffff8113261c>] get_page_from_freelist+0x6ac/0x840
[194043.650709] [<ffffffff814e8eab>] ? _spin_unlock_bh+0x1b/0x20
[194043.650714] [<ffffffff81125177>] ? mempool_free_slab+0x17/0x20
[194043.650720] [<ffffffff81134266>] __alloc_pages_nodemask+0x116/0xb40
[194043.650734] [<ffffffffa03148a9>] ? nfs_dq_update_shrink+0x29/0x120 [nfs]
[194043.650739] [<ffffffff8112228e>] ? find_get_page+0x1e/0xa0
[194043.650743] [<ffffffff81123bbc>] ? filemap_fault+0xfc/0x5d0
[194043.650750] [<ffffffff81174e6a>] alloc_pages_vma+0x9a/0x150
[194043.650755] [<ffffffff81155a67>] handle_pte_fault+0xa87/0xf60
[194043.650759] [<ffffffff81156124>] handle_mm_fault+0x1e4/0x2b0
[194043.650765] [<ffffffff811904ea>] ? do_sync_read+0xfa/0x140
[194043.650770] [<ffffffff81042aa9>] __do_page_fault+0x139/0x480
[194043.650776] [<ffffffff814ebe2e>] do_page_fault+0x3e/0xa0
[194043.650780] [<ffffffff814e91d5>] page_fault+0x25/0x30
```

many thanks and best regards,

Sirk

--

--

Subject: Re: Hung Tasks on NFS (maybe not a OpenVZ Problem) - How to forcefully kill a container ?
Posted by Todd Lyons on Fri, 30 Mar 2012 12:40:14 GMT
View Forum Message <> Reply to Message

On Fri, Mar 30, 2012 at 4:03 AM, Sirk Johannsen
<s.johannsen@satzmedia.de> wrote:
> Hi everyone,
>
> I am running a lot of CTs with their roots located on an nfs share.
> Once in a while it happens that a process gets stuck which I fear has
> something to do with the nfs mount.
> See the dmesg out below.
> The problem now is that I can't kill this process anymore.
> This results into beeing unable to stop the CT running this process.
> vzctl stop <CTID> runs into a timeout.
> It is totally impossible to kill the process - The only solution is a
> reboot of the Host-System.

Yep, that's correct. Your only real option is to migrate all of the
other CT's to another host node, then reboot this host node, then
migrate the other CT's back.

> Is there a way to forcefully kill the CT ?

Nope, the kernel is hung waiting for IO which will wait until the cows
come home. Are you using TCP or UDP nfs mounts? Try switching from
one to the other and see if that affects your nfs timeout issue.

> In this case I don't care if the process remains running.
> I just want the rest of the CT to be stopped so I can start the CT again.

I don't think it can be done.

> Here is the dmes output:
> [194043.649945] INFO: task which:810615 blocked for more than 120 seconds.
> [194043.650077] "echo 0 > /proc/sys/kernel/hung_task_timeout_secs"
> disables this message.
> [194043.650274] which   D ffff882f74146d50   0 810615 682640
> 125 0x00000084

So the "which" command was scanning the directories in the path and

---

that's what caused the nfs fault. Maybe you can tune your tcp
settings to compensate (assuming your using tcp nfs mounts) but I've
never tried to do anything like that, so I don't know if settings
could actually fix anything. Network congestion is likely your
biggest issue.

> [194043.650308] Call Trace:

Yeah, once you get a call trace, you're hosed.

Does slabtop show that your nfs slabs are using up extremely large
chunks of memory?

...Todd
--
Always code as if the guy who ends up maintaining your code will be a
violent psychopath who knows where you live. -- Martin Golding

---

Subject: Re: Hung Tasks on NFS (maybe not a OpenVZ Problem) - How to forcefully kill a container ?
Posted by Aleksandar Ivanisevic on Mon, 02 Apr 2012 10:08:18 GMT
View Forum Message <> Reply to Message

Sirk Johannsen <s.johannsen@satzmedia.de>
writes:

>
> Is there a way to forcefully kill the CT ?
> In this case I don't care if the process remains running.
> I just want the rest of the CT to be stopped so I can start the CT again.

try this:

vzctl chkpnt VEID --kill

don't know where I got it, but it worked for me in a few cases; in
some others it did not though.

Subject: Re: Re: Hung Tasks on NFS (maybe not a OpenVZ Problem) - How to forcefully kill a container ?
Posted by Kirill Korotaev on Mon, 02 Apr 2012 10:31:15 GMT
View Forum Message <> Reply to Message

Vzctl stop --fast

However it wont't help in case of tasks in D state. You need to mount nfs with softintr option for that.

Sent from my iPhonespam SPAMSPAM

On 02.04.2012, at 14:22, "Aleksandar Ivanisevic" <aleksandar@ivanisevic.de> wrote:

> Sirk Johannsen <s.johannsen@satzmedia.de>
> writes:
>
>>
>> Is there a way to forcefully kill the CT ?
>> In this case I don't care if the process remains running.
>> I just want the rest of the CT to be stopped so I can start the CT again.
>
> try this:
>
> vzctl chkpnt VEID --kill
>
> don't know where I got it, but it worked for me in a few cases; in
> some others it did not though.
>

Subject: Re: Re: Hung Tasks on NFS (maybe not a OpenVZ Problem) - How to forcefully kill a container ?
Posted by svensirk on Mon, 02 Apr 2012 13:24:22 GMT
View Forum Message <> Reply to Message

Thanks for all the responses.
I already have the NFS Mount mounted with intr but the processes still
stay in D state and are not killable.
Anyway, I'll try to convert all CTs to ploop this night and hope not
to see stuck processes on NFS afterwards :-)

best regards,

Sirk

2012/4/2 Kirill Korotaev <dev@parallels.com>:
> Vzctl stop --fast
> However it wont't help in case of tasks in D state. You need to mount nfs with softintr option for that.
>
> Sent from my iPhonespam SPAMSPAM
>
> On 02.04.2012, at 14:22, "Aleksandar Ivanisevic" <aleksandar@ivanisevic.de> wrote:
>

>> Sirk Johannsen <s.johannsen@satzmedia.de>
>> writes:
>>
>>>
>>> Is there a way to forcefully kill the CT ?
>>> In this case I don't care if the process remains running.
>>> I just want the rest of the CT to be stopped so I can start the CT again.
>>
>> try this:
>>
>> vzctl chkpnt VEID --kill
>>
>> don't know where I got it, but it worked for me in a few cases; in
>> some others it did not though.
>>
--
Satzmedia GmbH

Altonaer Poststraße 9
22767 Hamburg
Tel: +49 (0) 40 - 1 888 969 - 140
Fax: +49 (0) 40 - 1 888 969 - 200
E-Mail: s.johannsen@satzmedia.de
E-Business-Lösungen: http://www.satzmedia.de
Amtsgericht Hamburg, HRB 71729
Ust-IDNr. DE201979921
Geschäftsführer:
Dipl.-Kfm. Christian Satz
Dipl.-Inform. Markus Meyer-Westphal

--

--

Subject: Re: Re: Hung Tasks on NFS (maybe not a OpenVZ Problem) - How to
forcefully kill a container ?
Posted by kir on Mon, 02 Apr 2012 15:56:51 GMT
View Forum Message <> Reply to Message

On 04/02/2012 05:24 PM, Sirk Johannsen wrote:
> Thanks for all the responses.
> I already have the NFS Mount mounted with intr but the processes still
> stay in D state and are not killable.

Kirill probably meant the "soft" option. Plus, intr/nointr are
deprecated and ignored since 2.6.25 so they make no difference.

More details: man 5 nfs

> Anyway, I'll try to convert all CTs to ploop this night and hope not
> to see stuck processes on NFS afterwards :-)
>
> best regards,
>
> Sirk
>
> 2012/4/2 Kirill Korotaev<dev@parallels.com>:
>> Vzctl stop --fast
>> However it wont't help in case of tasks in D state. You need to mount nfs with softintr option for
that.
>>
>> Sent from my iPhonespam SPAMSPAM
>>
>> On 02.04.2012, at 14:22, "Aleksandar Ivanisevic"<aleksandar@ivanisevic.de> wrote:
>>
>>> Sirk Johannsen<s.johannsen@satzmedia.de>
>>> writes:
>>>
>>>> Is there a way to forcefully kill the CT ?
>>>> In this case I don't care if the process remains running.
>>>> I just want the rest of the CT to be stopped so I can start the CT again.
>>> try this:
>>>
>>> vzctl chkpnt VEID --kill
>>>
>>> don't know where I got it, but it worked for me in a few cases; in
>>> some others it did not though.
>>>