Ceph - Backport #14231

hammer: ceph-disk fails to work with udev generated symlinks

01/05/2016 09:21 AM - Alexey Sheplyakov

Status:	Rejected	**Spent time:**	0.00 hour
Priority:	Normal		
Assignee:	Loïc Dachary		
Target version:			
Release:	hammer	**Crash signature (v2):**	
Crash signature (v1):			

Description

description

~# ceph-deploy osd prepare node-9:/dev/sdc3:/dev/disk/by-id/ata-INTEL_SSDSC2BW240A4_PHDA410301812403GN-part3

fails here:

[node-9][WARNIN] DEBUG:ceph-disk:Journal /dev/disk/by-id/ata-INTEL_SSDSC2BW240A4_PHDA410301812403GN-part3 was previously prepared with ceph-disk. Reusing it.
[node-9][WARNIN] INFO:ceph-disk:Running command: /sbin/sgdisk -i 2 /dev/disk/by-id/ata-INTEL_SSDSC
[node-9][WARNIN] Problem opening /dev/disk/by-id/ata-INTEL_SSDSC for reading! Error is 2.

workaround

~# ceph-deploy osd prepare node-9:/dev/sdc3:$(ssh node-9 realpath /dev/disk/by-id/ata-INTEL_SSDSC2BW240A4_PHDA410301812403GN-part3)

History

#1 - 01/06/2016 10:11 PM - Loïc Dachary

- Assignee set to Loïc Dachary

https://github.com/ceph/ceph/pull/7123

#2 - 02/25/2016 10:55 AM - Loïc Dachary

- Status changed from New to 12

We don't have good enough test coverage in hammer for me to feel comfortable with the proposed change. It may introduce a subtle regression, even with a careful review. We could however add a regexp to the existing function that is specific to your use case. Something like:

```
    if 'loop' in dev or 'cciss' in dev or 'nvme' in dev:
        return re.match('(.*\d+)p(\d+)', dev).group(1, 2)
    elif 'ata-INTEL' in dev:
        return re.match('(.*)-part(\d+)', dev).group(1, 2)
    else:
        return re.match('(\D+)(\d+)', dev).group(1, 2)
```


What do you think ?

#3 - 03/01/2016 10:14 AM - Alexey Sheplyakov

We don't have good enough test coverage in hammer

We (Mirantis) have been using a similar patch for 2 months without any problems (that doesn't prove the patch is bug free, though).

elif 'ata-INTEL' in dev:

The problem is not limited to Intel devices, most hard drives have digits in their model names.
For instance on my workstation:

```
$ ls -1 /dev/disk/by-id/ata*
/dev/disk/by-id/ata-INTEL_SSDSC2BP240G4_BTJR529502MY240AGN
/dev/disk/by-id/ata-INTEL_SSDSC2BP240G4_BTJR529502MY240AGN-part1
/dev/disk/by-id/ata-INTEL_SSDSC2BP240G4_BTJR529502MY240AGN-part2
/dev/disk/by-id/ata-ST1000DM003-1CH162_S1DGTCYM
/dev/disk/by-id/ata-ST1000DM003-1CH162_S1DGTCYM-part1
/dev/disk/by-id/ata-ST1000DM003-1CH162_S1DGTCYM-part2
/dev/disk/by-id/ata-ST1000DM003-1CH162_S1DGTCYM-part5
/dev/disk/by-id/ata-WDC_WD20EURS-63SPKY0_WD-WMC300967502
/dev/disk/by-id/ata-WDC_WD20EURS-63SPKY0_WD-WMC300967502-part1
/dev/disk/by-id/ata-WDC_WD20EURS-63SPKY0_WD-WMC300967502-part2
/dev/disk/by-id/ata-WDC_WD20EURS-63SPKY0_WD-WMC300967502-part3
```

Also there are scsi, sas, and less common types of drives (hw raid and so on).
I think handling those with regexps is very error prone (and even more intrusive than the suggested patch).

#4 - 03/01/2016 10:16 AM - Alexey Sheplyakov

I think handling those with regexps is very error prone (and even more intrusive than the suggested patch).

On the other hand we could try matching with '(.*)-part(\d+)' first, and falling back to '(\D+)(\d+)'

#5 - 03/01/2016 10:27 AM - Loïc Dachary

@Alexey Sheplyakov would you be so kind as to push said similar patch in a pull request ? That would be most helpful :-)

#6 - 03/01/2016 12:03 PM - Loïc Dachary

@Alexey Sheplyakov or maybe the similar patch you're refering to is https://github.com/ceph/ceph/pull/7123/files ?

#7 - 03/02/2016 10:27 AM - Alexey Sheplyakov

@Loic Dachary

> would you be so kind as to push said similar patch in a pull request?

Actually that patch is a cherry-pick of

https://github.com/ceph/ceph/commit/0e34742b968e72aa6ce4a0c95a885dced435b3bc
https://github.com/ceph/ceph/commit/3bc95dfc1b88c01e16c3df04e96acced777b344a
https://github.com/ceph/ceph/commit/77ff7c3dc6dd6861b094e5a53d329de0802f3032
https://github.com/ceph/ceph/commit/b3c7cb098195111b9c642e5a9b726b63717f2e0d

(the last one has been already merged to 0.94.6)

The pool request I've submitted (https://github.com/ceph/ceph/pull/7123) is a reduced version of that patch (I've skipped multipath related changes).

#8 - 03/02/2016 11:27 AM - Loïc Dachary

Interesting :-) Are you using the patch that you proposed at https://github.com/ceph/ceph/pull/7123 ? Or another version of it ?

#9 - 03/03/2016 05:51 AM - Loïc Dachary

- Tracker changed from Bug to Backport

#10 - 03/03/2016 05:53 AM - Loïc Dachary

- Description updated

Original description

~# ceph-deploy osd prepare node-9:/dev/sdc3:/dev/disk/by-id/ata-INTEL_SSDSC2BW240A4_PHDA410301812403GN-part3

fails here:

[node-9][WARNIN] DEBUG:ceph-disk:Journal /dev/disk/by-id/ata-INTEL_SSDSC2BW240A4_PHDA410301812403GN-part3 was previously prepared with ceph-disk. Reusing it.
[node-9][WARNIN] INFO:ceph-disk:Running command: /sbin/sgdisk -i 2 /dev/disk/by-id/ata-INTEL_SSDSC
[node-9][WARNIN] Problem opening /dev/disk/by-id/ata-INTEL_SSDSC for reading! Error is 2.

split_dev_base_partnum [1] fails to guess the base device, so ceph-disk tries to open
a wrong (non-existent) device.

The problem has been fixed in the master branch as a side effect of the following commits:

https://github.com/ceph/ceph/commit/0e34742b968e72aa6ce4a0c95a885dced435b3bc
https://github.com/ceph/ceph/commit/3bc95dfc1b88c01e16c3df04e96acced777b344a
https://github.com/ceph/ceph/commit/77ff7c3dc6dd6861b094e5a53d329de0802f3032

[1] https://github.com/ceph/ceph/blob/v0.94.5/src/ceph-disk#L2361-L2365

#11 - 03/03/2016 05:53 AM - Loïc Dachary

- Subject changed from ceph-disk fails to work with udev generated symlinks to hammer: ceph-disk fails to work with udev generated symlinks

#12 - 03/03/2016 06:19 AM - Loïc Dachary

Since there is no satisfactory solution that would be both well tested and generic enough, I'm tempted to rule this as "Won't fix" since it's fixed in a
generic way in infernalis.

The simple workaround being to ceph-disk prepare $(realpath /dev/disk/by-id/ata-INTEL_SSDSC2BP240G4_BTJR529502MY240AGN) if using these
paths is somehow required.

What do you think ?

#13 - 03/03/2016 06:23 AM - Loïc Dachary

- Description updated

#14 - 04/11/2016 04:33 PM - Alexey Sheplyakov

Are you using the patch that you proposed at https://github.com/ceph/ceph/pull/7123?

Upd: we've been using the very same patch for 3 weeks, no problems so far.
Note that usually we run ~ 10 deployment tests per a day (most of those are VM based, though).
We are going to ship that patch with Mirantis OpenStack 9.0.

The simple workaround being to ceph-disk prepare $(realpath /dev/disk/by-id/ata-INTEL_SSDSC2BP240G4_BTJR529502MY240AGN)

It's a bit counter-intuitive and breaks the existing deployment tools.

#15 - 04/11/2016 04:54 PM - Loïc Dachary

thanks for the update, much appreciated :-)

#16 - 09/03/2017 06:59 PM - Nathan Cutler

- Status changed from 12 to Rejected

Hammer is EOL.