

Ceph - Bug #1659

Upgrade from 0.27 -> 0.37 going wrong, OSDs miss map updates

10/28/2011 01:01 AM - Wido den Hollander

Status: Can't reproduce	Start date: 10/28/2011
Priority: Normal	Due date:
Assignee: Samuel Just	% Done: 0%
Category: OSD	Estimated time: 0.00 hour
Target version:	Spent time: 8.00 hours
Source:	Reviewed:
Tags:	Affected Versions:
Backport:	ceph-qa-suite:
Regression: No	Pull request ID:
Severity: 3 - minor	

Description

Hi,

Like I mentioned on IRC, I had some problems with upgrading my cluster from 0.27 to 0.37.

It was a big step, but I tried anyway.

The OSDs booted and starting converting the on-disk format, that all went well, but then I saw multiple crashes, all seemed to relate back to the on-disk change.

First, the logs:

- <http://logger.ceph.widodh.nl/ceph/ceph-osd-22-10-2011.log.gz>
- <http://logger.ceph.widodh.nl/ceph/ceph-osd-23-10-2011.log.gz>
- <http://logger.ceph.widodh.nl/ceph/ceph-mon-oct-2011.log.gz>

I started seeing that OSDs came up and crashed after communicating with the monitor about their OSDmap.

```
Oct 22 22:10:47 node10 osd.5[10558]: 7f2d34cbc700 osd.5 794 _got_boot_version mon has osdmaps 795..805
Oct 22 22:10:47 node10 osd.5[10558]: 7f2d34cbc700 osd.5 794 send_boot
Oct 22 22:10:47 node10 osd.5[10558]: 7f2d34cbc700 osd.5 794 assuming cluster_addr ip matches client_addr
Oct 22 22:10:47 node10 osd.5[10558]: 7f2d34cbc700 osd.5 794 assuming hb_addr ip matches cluster_addr
Oct 22 22:10:47 node10 osd.5[10558]: 7f2d34cbc700 osd.5 794 client_addr [2a00:f10:10a:8:230:48ff:fe59:5186]:6800/10557, cluster_addr [2a00:f10:10a:8:230:48ff:fe59:5186]:6801/10557, hb addr [2a00:f10:10a:8:230:48ff:fe59:5186]:6802/10557
Oct 22 22:10:47 node10 osd.5[10558]: 7f2d31bb5700 osd.5 794 update_osd_stat osd_stat(6703 MB used, 86971 MB avail, 95366 MB total, peers []/[])
Oct 22 22:10:47 node10 osd.5[10558]: 7f2d31bb5700 osd.5 794 heartbeat: osd_stat(6703 MB used, 86971 MB avail, 95366 MB total, peers []/[])
Oct 22 22:10:47 node10 osd.5[10558]: 7f2d374c1700 osd.5 794 _dispatch 0x2b9a5a0 osd_map(805..806 src has 795..806) v2
Oct 22 22:10:47 node10 osd.5[10558]: 7f2d374c1700 osd.5 794 handle_osd_map epochs [805,806], i have 794, src has [795,806]
Oct 22 22:10:47 node10 osd.5[10558]: 7f2d374c1700 osd.5 794 handle_osd_map message skips epochs 795..804
Oct 22 22:10:47 node10 osd.5[10558]: 7f2d374c1700 osd.5 794 locking handle_osd_map permissions
Oct 22 22:10:47 node10 osd.5[10558]: 7f2d374c1700 osd.5 794 handle_osd_map got inc map for epoch 805
Oct 22 22:10:47 node10 osd.5[10558]: 7f2d374c1700 osd.5 794 add_map_inc_bl 805 121 bytes
Oct 22 22:10:47 node10 osd.5[10558]: 7f2d374c1700 osd.5 794 get_map 804 - loading and decoding 0x2b62900
Oct 22 22:10:47 node10 osd.5[10558]: 7f2d374c1700 filestore(/var/lib/ceph/osd.5) read meta/osdmap.
```

```

804/0 0~0
Oct 22 22:10:47 node10 osd.5[10558]: 7f2d374c1700 filestore(/var/lib/ceph/osd.5) FileStore::read(m
eta/osdmap.804/0): open error error 2: No such file or directory
Oct 22 22:10:47 node10 osd.5[10558]: *** Caught signal (Aborted) **#012 in thread 7f2d374c1700
Oct 22 22:10:47 node10 osd.5[10558]: 7f2d313b4700 osd.5 794 No AuthAuthorizeHandler found for prot
ocol 1
Oct 22 22:10:47 node10 osd.5[10558]: 7f2d313b4700 -- [2a00:f10:10a:8:230:48ff:fe59:5186]:6800/1055
7 >> [2a00:f10:10a:8:230:48ff:fe56:ed63]:6800/13206 pipe(0x2b55c80 sd=16 pgs=0 cs=0 l=1).accept ba
d authorizer
Oct 22 22:10:47 node10 osd.5[10558]: 7f2d313b4700 osd.5 794 No AuthAuthorizeHandler found for prot
ocol 1
Oct 22 22:10:47 node10 osd.5[10558]: 7f2d313b4700 -- [2a00:f10:10a:8:230:48ff:fe59:5186]:6800/1055
7 >> [2a00:f10:10a:8:230:48ff:fe56:ed63]:6800/13206 pipe(0x2b55c80 sd=16 pgs=0 cs=0 l=1).accept ba
d authorizer
Oct 22 22:10:47 node10 osd.5[10558]: 7f2d313b4700 osd.5 794 No AuthAuthorizeHandler found for prot
ocol 1
Oct 22 22:10:47 node10 osd.5[10558]: 7f2d313b4700 -- [2a00:f10:10a:8:230:48ff:fe59:5186]:6800/1055
7 >> [2a00:f10:10a:8:230:48ff:fe56:ed63]:6800/13206 pipe(0x2b55a00 sd=16 pgs=0 cs=0 l=1).accept ba
d authorizer
Oct 22 22:10:47 node10 osd.5[10558]: ceph version 0.37-96-g1b846f4 (commit:1b846f43f181ac31a1a191
1c74b7cf3ae063f35d)#012 1: /usr/bin/ceph-osd() [0x67fd94]#012 2: (()+0xf8f0) [0x7f2d42f818f0]#012
3: (gsignal()+0x35) [0x7f2d4170ba75]#012 4: (abort()+0x180) [0x7f2d4170f5c0]#012 5: (__gnu_cxx::__
verbose_terminate_handler()+0x115) [0x7f2d41fc18e5]#012 6: (()+0xcad16) [0x7f2d41fbfd16]#012 7: ((
)+0xcad43) [0x7f2d41fbfd43]#012 8: (()+0xcae3e) [0x7f2d41fbfe3e]#012 9: (ceph::buffer::list::itera
tor::copy(unsigned int, char*)+0x13e) [0x5c3dae]#012 10: (OSDMap::decode(ceph::buffer::list&)+0x84
) [0x5774d4]#012 11: (OSD::get_map(unsigned int)+0x221) [0x539b11]#012 12: (OSD::handle_osd_map(MO
SDMap*)+0x225b) [0x55616b]#012 13: (OSD::_dispatch(Message*)+0x258) [0x557f58]#012 14: (OSD::ms_di
spatch(Message*)+0x103) [0x558e23]#012 15: (SimpleMessenger::dispatch_entry()+0x7f9) [0x58a489]#01
2 16: (SimpleMessenger::DispatchThread::entry()+0x1c) [0x4a6b2c]#012 17: (()+0x69ca) [0x7f2d42f789
ca]#012 18: (clone()+0x6d) [0x7f2d417be70d]

```

And indeed, osdmap 804 was missing. I manually scp'd it from the monitor to the OSD and a minute later it would complain that 807 was not present (it's own boot did trigger a osdmap change which never propagated to the OSD).

You can take a look through the logs, but I'm not sure if you'll find anything useful in them. Debugging was enabled at random to debug why OSDs went down, but not consistently.

Besides the osdmap problems (which occurred on almost every OSD) there are various crashed to be seen, like:

```

Oct 22 21:55:47 node03 osd.0[15276]: 7f7337a7a720 filestore(/var/lib/ceph/osd.0) _destroy_collecti
on /var/lib/ceph/osd.0/current/_temp
Oct 22 21:55:47 node03 osd.0[15276]: 7f7337a7a720 filestore(/var/lib/ceph/osd.0) _destroy_collecti
on /var/lib/ceph/osd.0/current/_temp = -39
Oct 22 21:55:47 node03 osd.0[15276]: os/FileStore.cc: In function 'unsigned int FileStore::_do_tra
nsaction(ObjectStore::Transaction&)', in thread '7f7337a7a720'#012os/FileStore.cc: 2416: FAILED as
sert(0 == "ENOTEMPTY suggests garbage data in osd data dir")
Oct 22 21:55:47 node03 osd.0[15276]: ceph version 0.37-96-g1b846f4 (commit:1b846f43f181ac31a1a191
1c74b7cf3ae063f35d)#012 1: (FileStore::_do_transaction(ObjectStore::Transaction&)+0x410b) [0x6c181
b]#012 2: (FileStore::do_transactions(std::list<ObjectStore::Transaction*, std::allocator<ObjectSt
ore::Transaction*> >&, unsigned long)+0x75) [0x6c1e75]#012 3: (JournalingObjectStore::journal_repl
ay(unsigned long)+0x793) [0x6ccfc3]#012 4: (FileStore::mount()+0x2803) [0x6abb13]#012 5: (OSD::con
vertfs(std::string const&, std::string const&)+0xc8) [0x524f18]#012 6: (main()+0x2c8d) [0x4a55dd]#
012 7: (__libc_start_main()+0xfd) [0x7f7335e09c4d]#012 8: /usr/bin/ceph-osd() [0x4a2529]
Oct 22 21:55:47 node03 osd.0[15276]: ceph version 0.37-96-g1b846f4 (commit:1b846f43f181ac31a1a191
1c74b7cf3ae063f35d)#012 1: (FileStore::_do_transaction(ObjectStore::Transaction&)+0x410b) [0x6c181
b]#012 2: (FileStore::do_transactions(std::list<ObjectStore::Transaction*, std::allocator<ObjectSt
ore::Transaction*> >&, unsigned long)+0x75) [0x6c1e75]#012 3: (JournalingObjectStore::journal_repl
ay(unsigned long)+0x793) [0x6ccfc3]#012 4: (FileStore::mount()+0x2803) [0x6abb13]#012 5: (OSD::con
vertfs(std::string const&, std::string const&)+0xc8) [0x524f18]#012 6: (main()+0x2c8d) [0x4a55dd]#
012 7: (__libc_start_main()+0xfd) [0x7f7335e09c4d]#012 8: /usr/bin/ceph-osd() [0x4a2529]
Oct 22 21:55:47 node03 osd.0[15276]: *** Caught signal (Aborted) **#012 in thread 7f7337a7a720
Oct 22 21:55:47 node03 osd.0[15276]: ceph version 0.37-96-g1b846f4 (commit:1b846f43f181ac31a1a191
1c74b7cf3ae063f35d)#012 1: /usr/bin/ceph-osd() [0x67fd94]#012 2: (()+0xf8f0) [0x7f733765a8f0]#012
3: (gsignal()+0x35) [0x7f7335e1ea75]#012 4: (abort()+0x180) [0x7f7335e225c0]#012 5: (__gnu_cxx::__
verbose_terminate_handler()+0x115) [0x7f73366d48e5]#012 6: (()+0xcad16) [0x7f73366d2d16]#012 7: ((

```

```
)+0xcad43) [0x7f73366d2d43]#012 8: ((+0xcae3e) [0x7f73366d2e3e]#012 9: (ceph::__ceph_assert_fail(
char const*, char const*, int, char const*)+0x39f) [0x5b561f]#012 10: (FileStore::_do_transaction(
ObjectStore::Transaction&)+0x410b) [0x6c181b]#012 11: (FileStore::do_transactions(std::list<Object
Store::Transaction*, std::allocator<ObjectStore::Transaction*> >&, unsigned long)+0x75) [0x6c1e75]
#012 12: (JournalingObjectStore::journal_replay(unsigned long)+0x793) [0x6ccfc3]#012 13: (FileStor
e::mount()+0x2803) [0x6abb13]#012 14: (OSD::convertfs(std::string const&, std::string const&)+0xc8
) [0x524f18]#012 15: (main()+0x2c8d) [0x4a55dd]#012 16: (__libc_start_main()+0xfd) [0x7f7335e09c4d
]#012 17: /usr/bin/ceph-osd() [0x4a2529]
```

I guess this one is related to the on-disk change which not went smoothly.

```
Oct 22 23:18:06 node07 osd.4[11761]: 7fd5b69b3720 filestore(/var/lib/ceph/osd.4) error error 17:
File exists not handled
Oct 22 23:18:06 node07 osd.4[11761]: os/FileStore.cc: In function 'unsigned int FileStore::_do_tra
nSACTION(ObjectStore::Transaction&)', in thread '7fd5b69b3720'#012os/FileStore.cc: 2423: FAILED as
sert(0 == "unexpected error")
Oct 22 23:18:06 node07 osd.4[11761]: ceph version 0.37-96-g1b846f4 (commit:1b846f43f181ac31a1a191
1c74b7cf3ae063f35d)#012 1: (FileStore::_do_transaction(ObjectStore::Transaction&)+0x40c) [0x6bdb1c
]#012 2: (FileStore::do_transactions(std::list<ObjectStore::Transaction*, std::allocator<ObjectSto
re::Transaction*> >&, unsigned long)+0x75) [0x6c1e75]#012 3: (JournalingObjectStore::journal_repla
y(unsigned long)+0x793) [0x6ccfc3]#012 4: (FileStore::mount()+0x2803) [0x6abb13]#012 5: (OSD::conv
ertfs(std::string const&, std::string const&)+0xc8) [0x524f18]#012 6: (main()+0x2c8d) [0x4a55dd]#0
12 7: (__libc_start_main()+0xfd) [0x7fd5b4d43c4d]#012 8: /usr/bin/ceph-osd() [0x4a2529]
Oct 22 23:18:06 node07 osd.4[11761]: ceph version 0.37-96-g1b846f4 (commit:1b846f43f181ac31a1a191
1c74b7cf3ae063f35d)#012 1: (FileStore::_do_transaction(ObjectStore::Transaction&)+0x40c) [0x6bdb1c
]#012 2: (FileStore::do_transactions(std::list<ObjectStore::Transaction*, std::allocator<ObjectSto
re::Transaction*> >&, unsigned long)+0x75) [0x6c1e75]#012 3: (JournalingObjectStore::journal_repla
y(unsigned long)+0x793) [0x6ccfc3]#012 4: (FileStore::mount()+0x2803) [0x6abb13]#012 5: (OSD::conv
ertfs(std::string const&, std::string const&)+0xc8) [0x524f18]#012 6: (main()+0x2c8d) [0x4a55dd]#0
12 7: (__libc_start_main()+0xfd) [0x7fd5b4d43c4d]#012 8: /usr/bin/ceph-osd() [0x4a2529]
Oct 22 23:18:06 node07 osd.4[11761]: *** Caught signal (Aborted) **#012 in thread 7fd5b69b3720
Oct 22 23:18:06 node07 osd.4[11761]: ceph version 0.37-96-g1b846f4 (commit:1b846f43f181ac31a1a191
1c74b7cf3ae063f35d)#012 1: /usr/bin/ceph-osd() [0x67fd94]#012 2: ((+0xf8f0) [0x7fd5b65948f0]#012
3: (gsignal()+0x35) [0x7fd5b4d58a75]#012 4: (abort()+0x180) [0x7fd5b4d5c5c0]#012 5: (__gnu_cxx::__
verbose_terminate_handler()+0x115) [0x7fd5b560e8e5]#012 6: ((+0xcad16) [0x7fd5b560cd16]#012 7: ((
)+0xcad43) [0x7fd5b560cd43]#012 8: ((+0xcae3e) [0x7fd5b560ce3e]#012 9: (ceph::__ceph_assert_fail(
char const*, char const*, int, char const*)+0x39f) [0x5b561f]#012 10: (FileStore::_do_transaction(
ObjectStore::Transaction&)+0x40c) [0x6bdb1c]#012 11: (FileStore::do_transactions(std::list<ObjectS
tore::Transaction*, std::allocator<ObjectStore::Transaction*> >&, unsigned long)+0x75) [0x6c1e75]#
012 12: (JournalingObjectStore::journal_replay(unsigned long)+0x793) [0x6ccfc3]#012 13: (FileStore
::mount()+0x2803) [0x6abb13]#012 14: (OSD::convertfs(std::string const&, std::string const&)+0xc8)
[0x524f18]#012 15: (main()+0x2c8d) [0x4a55dd]#012 16: (__libc_start_main()+0xfd) [0x7fd5b4d43c4d]
#012 17: /usr/bin/ceph-osd() [0x4a2529]
```

Hope this helps!

History

#1 - 10/31/2011 03:34 PM - Samuel Just

- Assignee set to Samuel Just

#2 - 11/14/2011 04:12 PM - Anonymous

I saw a very similar stack trace in the 11/11 Nightly
[1862/remotel@ubuntu@sepi9.ceph.dreamhost.com/log/osd.5.log.gz](https://bugzilla.redhat.com/show_bug.cgi?id=1862)

#3 - 02/07/2012 10:37 PM - Sage Weil

- Status changed from New to Can't reproduce