Ceph - Bug #521

objecter: crash in osdmap assert

10/27/2010 11:45 AM - Sage Weil

Status:	Resolved	**Start date:**	10/27/2010
Priority:	High	**Due date:**	
Assignee:		**% Done:**	0%
Category:		**Estimated time:**	0.00 hour
Target version:	v0.23	**Spent time:**	1.00 hour
Source:		**Reviewed:**	
Tags:		**Affected Versions:**	
Backport:		**ceph-qa-suite:**	
Regression:	No	**Pull request ID:**	
Severity:	3 - minor	**Crash signature:**	

Description

```
When accessing multiple RBD-Volumes from one VM in parallel, we are
receiving an assertion:

./osd/OSDMap.h: In function 'entity_inst_t OSDMap::get_inst(int)':
./osd/OSDMap.h:460: FAILED assert(exists(osd) && is_up(osd))
ceph version 0.22.1 (commit:c6f403a6f441184956e00659ce713eaee7014279)
1: (Objecter::op_submit(Objecter::Op*)+0x6c2) [0x38658854c2]
2: /usr/lib64/librados.so.1() [0x3865855dc9]
3: (RadosClient::aio_write(RadosClient::PoolCtx&, object_t, long,
ceph::buffer::list const&, unsigned long,
RadosClient::AioCompletion*)+0x24b) [0x386585724b]
4: (rados_aio_write()+0x9a) [0x386585741a]
5: /usr/bin/qemu-kvm() [0x45a305]
6: /usr/bin/qemu-kvm() [0x45a430]
7: /usr/bin/qemu-kvm() [0x43bb73]
NOTE: a copy of the executable, or `objdump -rdS <executable>` is
needed to interpret this.
./osd/OSDMap.h: In function 'entity_inst_t OSDMap::get_inst(int)':
./osd/OSDMap.h:460: FAILED assert(exists(osd) && is_up(osd))
ceph version 0.22.1 (commit:c6f403a6f441184956e00659ce713eaee7014279)
1: (Objecter::op_submit(Objecter::Op*)+0x6c2) [0x38658854c2]
2: /usr/lib64/librados.so.1() [0x3865855dc9]
3: (RadosClient::aio_write(RadosClient::PoolCtx&, object_t, long,
ceph::buffer::list const&, unsigned long,
RadosClient::AioCompletion*)+0x24b) [0x386585724b]
4: (rados_aio_write()+0x9a) [0x386585741a]
5: /usr/bin/qemu-kvm() [0x45a305]
6: /usr/bin/qemu-kvm() [0x45a430]
7: /usr/bin/qemu-kvm() [0x43bb73]
NOTE: a copy of the executable, or `objdump -rdS <executable>` is
needed to interpret this.
terminate called after throwing an instance of 'ceph::FailedAssertion'
*** Caught signal (ABRT) ***
ceph version 0.22.1 (commit:c6f403a6f441184956e00659ce713eaee7014279)
1: (sigabrt_handler(int)+0x91) [0x3865922b91]
2: /lib64/libc.so.6() [0x3c0c032a30]
3: (gsignal()+0x35) [0x3c0c0329b5]
4: (abort()+0x175) [0x3c0c034195]
5: (__gnu_cxx::__verbose_terminate_handler()+0x12d) [0x3c110beaad]

This is reproducible by doing the following inside a VM:
```

```
# mkfs.btrfs /dev/vdb /dev/vdc /dev/vdd /dev/vde
# mount /dev/vdb /mnt
# cd /mnt
# bonnie++ -u root -d /mnt -f

Any hints are welcome...

Thanks,
```

History

#1 - 10/30/2010 04:02 PM - Sage Weil

- Target version changed from v0.22.2 to v0.22.3

#2 - 11/05/2010 01:36 PM - Sage Weil

- Status changed from New to Feedback

#3 - 11/09/2010 09:45 AM - Sage Weil

- Source set to 0

latest from ML:

```
Subject:    Re: AW: ./osd/OSDMap.h:460: FAILED assert(exists(osd) && is_up(osd))
From:       Christian Brunner <chb () muc ! de>
Date:       2010-11-05 12:40:39
Message-ID: AANLkTi=2=oaobOtwacs8EZRGnBRzctShTFPs_2E-_gcq () mail ! gmail ! com
[Download message RAW]

Hi Sage,

I'm sorry, I was busy with some other things, but I was able to look
at this now:

Now I can confirm that the problem is related to a missing osd, as I
had to stop one of the osds to reproduce it. When I split up the two
asserts, the error occurs in:

./osd/OSDMap.h:460: FAILED assert(exists(osd))

and here is the gdb backtrace:

#0  0x0000003c0c0329b5 in raise () from /lib64/libc.so.6
#1  0x0000003c0c034195 in abort () from /lib64/libc.so.6
#2  0x0000003c110beaad in __gnu_cxx::__verbose_terminate_handler() ()
   from /usr/lib64/libstdc++.so.6
#3  0x0000003c110bcc36 in ?? () from /usr/lib64/libstdc++.so.6
#4  0x0000003c110bcc63 in std::terminate() () from /usr/lib64/libstdc++.so.6
#5  0x0000003c110bcd5e in __cxa_throw () from /usr/lib64/libstdc++.so.6
#6  0x00007ffc36df6136 in ceph::__ceph_assert_fail (
    assertion=0x7ffc36e22c59 "exists(osd)", file=<value optimized out>,
    line=460, func=<value optimized out>) at common/assert.cc:30
#7  0x00007ffc36d8014f in OSDMap::get_inst (this=<value optimized out>,
    osd=<value optimized out>) at osd/OSDMap.h:460
#8  0x00007ffc36d7b4c2 in Objecter::op_submit (this=0x159a4a0,
    op=0x7ffb29e264c0) at osdc/Objecter.cc:461
#9  0x00007ffc36d4bdc9 in Objecter::write (this=0x159a4a0, oid=..., ol=...,
    off=<value optimized out>, len=1638400, snapc=..., bl=..., mtime=...,
    onack=0x7ffc00811d20, oncommit=0x7ffc000461a0, flags=0)
    at osdc/Objecter.h:606
#10 0x00007ffc36d4d24b in RadosClient::aio_write (this=0x15916e0, pool=...,
    oid=..., off=917504, bl=..., len=1638400, c=0x7ffc03d4c010)
    at librados.cc:949
#11 0x00007ffc36d4d41a in rados_aio_write (pool=0x159a000,
    o=<value optimized out>, off=917504, buf=<value optimized out>,
    len=1638400, completion=0x7ffc03d4c010) at librados.cc:2119
#12 0x000000000045a305 in rbd_aio_rw_vector (bs=<value optimized out>,
    sector_num=<value optimized out>, qiov=<value optimized out>,
```

```
    nb_sectors=917504, cb=<value optimized out>, opaque=<value optimized out>,
    write=1) at block/rbd.c:769
#13 0x0000000000045a430 in rbd_aio_writev (bs=<value optimized out>,
    sector_num=<value optimized out>, qiov=<value optimized out>,
    nb_sectors=<value optimized out>, cb=<value optimized out>,
    opaque=<value optimized out>) at block/rbd.c:802
#14 0x0000000000043bb73 in bdrv_aio_writev (bs=0x159dd20, sector_num=2279168,
    qiov=0x7ffb29e26480, nb_sectors=3200, cb=<value optimized out>,
    opaque=<value optimized out>) at block.c:2019
#15 0x0000000000043bb73 in bdrv_aio_writev (bs=0x159d3f0, sector_num=2279168,
    qiov=0x7ffb29e26480, nb_sectors=3200, cb=<value optimized out>,
    opaque=<value optimized out>) at block.c:2019
#16 0x0000000000043ca2c in bdrv_aio_multiwrite (bs=0x159d3f0,
    reqs=0x7ffc0f5fd690, num_reqs=<value optimized out>) at block.c:2228
#17 0x0000000000041cca5 in virtio_submit_multiwrite (bs=<value optimized out>,
    mrb=0x7ffc0f5fd690) at /usr/src/debug/qemu-kvm-0.13.0/hw/virtio-blk.c:241
#18 0x0000000000041d30c in virtio_blk_handle_output (vdev=0x1e56c30,
    vq=<value optimized out>)
    at /usr/src/debug/qemu-kvm-0.13.0/hw/virtio-blk.c:359
#19 0x0000000000042dc5d in kvm_handle_io (env=0x15c4b20)
    at /usr/src/debug/qemu-kvm-0.13.0/kvm-all.c:763
#20 kvm_run (env=0x15c4b20) at /usr/src/debug/qemu-kvm-0.13.0/qemu-kvm.c:645
#21 0x0000000000042dd89 in kvm_cpu_exec (env=<value optimized out>)
    at /usr/src/debug/qemu-kvm-0.13.0/qemu-kvm.c:1238
#22 0x0000000000042f181 in kvm_main_loop_cpu (_env=0x15c4b20)
    at /usr/src/debug/qemu-kvm-0.13.0/qemu-kvm.c:1495
#23 ap_main_loop (_env=0x15c4b20)
    at /usr/src/debug/qemu-kvm-0.13.0/qemu-kvm.c:1541
#24 0x0000003c0c4077e1 in start_thread () from /lib64/libpthread.so.0
#25 0x0000003c0c0e151d in clone () from /lib64/libc.so.6

I hope this helps.

Regards,
Christian

2010/11/4 Sage Weil <sage@newdream.net>:
> Hi Christian,
>
> On Tue, 26 Oct 2010, Christian Brunner wrote:
>> I can't promise this for tomorrow, but I think I can do this on Thursday.
>
> Have you had a chance to look into this one at all?
>
> Thanks-
> sage
```

#4 - 11/09/2010 09:56 AM - Sage Weil

- Target version changed from v0.22.3 to v0.23

#5 - 11/09/2010 04:26 PM - Sage Weil

Can you try with something like

```
diff --git a/src/osdc/Objecter.cc b/src/osdc/Objecter.cc
index 0fe4d65..d2f4c54 100644
--- a/src/osdc/Objecter.cc
+++ b/src/osdc/Objecter.cc
@@ -494,6 +494,11 @@ tid_t Objecter::op_submit(Op *op)
     if (op->priority)
       m->set_priority(op->priority);

+    if (!osdmap->exists(pg.primary())) {
+      dout(0) << "pgid " << op->pgid << " acting " << pg.acting
+              << " primary dne, osdmap epoch " << osdmap->get_epoch() << dendl;
+    }
+
     messenger->send_message(m, osdmap->get_inst(pg.primary()));
   } else
     maybe_request_map();
```

(That is against the latest 'rc' branch.)

Once you get the osdmap epoch, can you dump that (ceph osd dump <epoch> -o -). And map the pg explicitly, 'ceph pg map 1.23', or 'ceph osd getmap <epoch> -o /tmp/foo ; osdmaptool /tmp/foo --test-map-pg 1.23').

#6 - 11/10/2010 09:43 AM - Sage Weil

- Status changed from Feedback to Resolved

- Source changed from 0 to 2

586c9e7a80b425802ca77d8c09bb00da5c25d616