

Ceph - Bug #996

osd: ReplicatedPG::do_osd_ops

04/10/2011 01:32 AM - Wido den Hollander

Status:	Can't reproduce	% Done:	0%
Priority:	Normal	Spent time:	1.00 hour
Assignee:			
Category:	OSD		
Target version:	v0.27.1		
Source:		Affected Versions:	
Tags:		ceph-qa-suite:	
Backport:		Pull request ID:	
Regression:	No	Crash signature (v1):	
Severity:	3 - minor	Crash signature (v2):	
Reviewed:			

Description

I had an issue that my cluster would not recover, spoke with joshd about this on IRC (<http://irclogs.ceph.widodh.nl/index.php?date=2011-04-08>)

After I recovered my cluster back to 39/40 OSD's everything came back to normal again, but my kclient would not see the new osdmap, so I rebooted that client (force reboot with sysrq).

Doing so seemed to crash 13 of the online OSD's. I rebooted the client somewhere around 08:15, at the same moment all these OSD's went down with:

```
(gdb) bt
#0 0x00007f0f2eaa7bb in raise () from /lib/libpthread.so.0
#1 0x000000000620ce3 in reraise_fatal (signum=18759) at common/signal.cc:63
#2 0x000000000621a0b in handle_fatal_signal (signum=6) at common/signal.cc:110
#3 <signal handler called>
#4 0x00007f0f2d5eaa75 in raise () from /lib/libc.so.6
#5 0x00007f0f2d5ee5c0 in abort () from /lib/libc.so.6
#6 0x00007f0f2dea08e5 in __gnu_cxx::__verbose_terminate_handler() () from /usr/lib/libstdc++.so.6
#7 0x00007f0f2de9ed16 in ?? () from /usr/lib/libstdc++.so.6
#8 0x00007f0f2de9ed43 in std::terminate() () from /usr/lib/libstdc++.so.6
#9 0x00007f0f2de9ee3e in __cxa_throw () from /usr/lib/libstdc++.so.6
#10 0x000000000607caa in ceph::__ceph_assert_fail (assertion=<value optimized out>, file=<value optimized out>, line=<value optimized out>,
    func=0x63f320 "int ReplicatedPG::do_osd_ops(ReplicatedPG::OpContext*, std::vector<OSDOp, std::allocator<OSDOp> >&, ceph::bufferlist&)" at common/assert.cc:86
#11 0x0000000004b8f62 in ReplicatedPG::do_osd_ops (this=0x3002000, ctx=0x6eb5480, ops=<value optimized out>, odata=<value optimized out>) at osd/ReplicatedPG.cc:1211
#12 0x0000000004b9bcf in ReplicatedPG::prepare_transaction (this=0x3002000, ctx=0x4c12) at osd/ReplicatedPG.cc:2229
#13 0x0000000004bb2e9 in ReplicatedPG::do_op (this=0x3002000, op=0x102f6240) at osd/ReplicatedPG.cc:501
#14 0x0000000004f54ad in OSD::dequeue_op (this=0x2a86000, pg=0x3002000) at osd/OSD.cc:5419
#15 0x00000000060ab39 in ThreadPool::worker (this=0x2a863f0) at common/WorkQueue.cc:44
#16 0x000000000524efd in ThreadPool::WorkThread::entry() ()
#17 0x00007f0f2ea119ca in start_thread () from /lib/libpthread.so.0
#18 0x00007f0f2d69d70d in clone () from /lib/libc.so.6
#19 0x0000000000000000 in ?? ()
```

The following OSD's went down:

- osd2

- osd11
- osd13
- osd14
- osd18
- osd19
- osd20
- osd28
- osd29
- osd33
- osd34
- osd38
- osd39

The last log lines I found for example osd11:

```
Apr  9 08:16:53 atom2 osd.11[29340]: 7f320fa46700 -- [2a00:f10:113:1:225:90ff:fe33:49f2]:6813/2933
9 >> [2a00:f10:113:1:225:90ff:fe33:49fe]:6805/1552 pipe(0x2fa9c80 sd=13 pgs=49 cs=2 l=0).connect c
laims to be [2a00:f10:113:1:225:90ff:fe33:49fe]:6805/1448 not [2a00:f10:113:1:225:90ff:fe33:49fe]:
6805/1552 - wrong node!
Apr  9 08:17:07 atom2 osd.11[29340]: 7f32188b1700 osd11 1797 pg[0.7e7( v 1499'634 (1205'631,1499'6
34) n=591 ec=2 les=1513 1509/1509/1509) [11,34,20] r=0 mlcod 0'0 active+clean] oi.user_version=0'0
is_modify=0
Apr  9 08:17:07 atom2 osd.11[29340]: 7f32188b1700 log [ERR] : 0.7e7 10000006e1d.00000fff/head oi.s
ize 0 but stat got -2 size 0
Apr  9 08:17:07 atom2 osd.11[29340]: osd/ReplicatedPG.cc: In function 'int ReplicatedPG::do_osd_op
s(ReplicatedPG::OpContext*, std::vector<OSDOp, std::allocator<OSDOp> >&, ceph::bufferlist&)', in t
hread '0x7f32188b1700'#012osd/ReplicatedPG.cc: 1211: FAILED assert(0 == "oi disagrees with stat")
Apr  9 08:17:07 atom2 osd.11[29340]: ceph version 0.25-588-g24caedc (commit:24caedc8f549eeeba48b2
d4a44927ee16e65c42a)#012 1: (ReplicatedPG::do_osd_ops(ReplicatedPG::OpContext*, std::vector<OSDOp,
std::allocator<OSDOp> >&, ceph::buffer::list&)+0xa2e2) [0x4b8f62]#012 2: (ReplicatedPG::prepare_t
ransaction(ReplicatedPG::OpContext*)+0x6f) [0x4b9bcf]#012 3: (ReplicatedPG::do_op(MOSDOp*)+0xa09)
[0x4bb2e9]#012 4: (OSD::dequeue_op(PG*)+0x37d) [0x4f54ad]#012 5: (ThreadPool::worker()+0x519) [0x6
0ab39]#012 6: (ThreadPool::WorkThread::entry()+0xd) [0x524efd]#012 7: (()+0x69ca) [0x7f322583c9ca]
#012 8: (clone()+0x6d) [0x7f32244c870d]
Apr  9 08:17:07 atom2 osd.11[29340]: ceph version 0.25-588-g24caedc (commit:24caedc8f549eeeba48b2
d4a44927ee16e65c42a)#012 1: (ReplicatedPG::do_osd_ops(ReplicatedPG::OpContext*, std::vector<OSDOp,
std::allocator<OSDOp> >&, ceph::buffer::list&)+0xa2e2) [0x4b8f62]#012 2: (ReplicatedPG::prepare_t
ransaction(ReplicatedPG::OpContext*)+0x6f) [0x4b9bcf]#012 3: (ReplicatedPG::do_op(MOSDOp*)+0xa09)
[0x4bb2e9]#012 4: (OSD::dequeue_op(PG*)+0x37d) [0x4f54ad]#012 5: (ThreadPool::worker()+0x519) [0x6
0ab39]#012 6: (ThreadPool::WorkThread::entry()+0xd) [0x524efd]#012 7: (()+0x69ca) [0x7f322583c9ca]
#012 8: (clone()+0x6d) [0x7f32244c870d]
Apr  9 08:17:07 atom2 osd.11[29340]: *** Caught signal (Aborted) **#012 in thread 0x7f32188b1700
Apr  9 08:17:07 atom2 osd.11[29340]: ceph version 0.25-588-g24caedc (commit:24caedc8f549eeeba48b2
d4a44927ee16e65c42a)#012 1: /usr/bin/cosd() [0x6217ee]#012 2: (()+0xf8f0) [0x7f32258458f0]#012 3:
(gsignal()+0x35) [0x7f3224415a75]#012 4: (abort()+0x180) [0x7f32244195c0]#012 5: (__gnu_cxx::__ver
bose_terminate_handler()+0x115) [0x7f3224ccb8e5]#012 6: (()+0xcad16) [0x7f3224cc9d16]#012 7: (()+0
xcad43) [0x7f3224cc9d43]#012 8: (()+0xcae3e) [0x7f3224cc9e3e]#012 9: (ceph::__ceph_assert_fail(char
const*, char const*, int, char const*)+0x36a) [0x607caa]#012 10: (ReplicatedPG::do_osd_ops(Repli
catedPG::OpContext*, std::vector<OSDOp, std::allocator<OSDOp> >&, ceph::buffer::list&)+0xa2e2) [0x
4b8f62]#012 11: (ReplicatedPG::prepare_transaction(ReplicatedPG::OpContext*)+0x6f) [0x4b9bcf]#012
12: (ReplicatedPG::do_op(MOSDOp*)+0xa09) [0x4bb2e9]#012 13: (OSD::dequeue_op(PG*)+0x37d) [0x4f54ad
]#012 14: (ThreadPool::worker()+0x519) [0x60ab39]#012 15: (ThreadPool::WorkThread::entry()+0xd) [0
x524efd]#012 16: (()+0x69ca) [0x7f322583c9ca]#012 17: (clone()+0x6d) [0x7f32244c870d]
```

And osd38:

```
Apr  9 08:17:04 atom9 osd.39[16064]: 7f2c6795b700 osd39 1797 pg[0.10f( v 1492'289 (1205'287,1492'2
89) n=257 ec=2 les=1155 1152/1152/1152) [39,14,19] r=0 mlcod 1294'288 active+clean] oi.user_versio
n=0'0 is_modify=0
Apr  9 08:17:04 atom9 osd.39[16064]: 7f2c6795b700 log [ERR] : 0.10f 10000006e1f.00000fff/head oi.s
ize 0 but stat got -2 size 0
Apr  9 08:17:04 atom9 osd.39[16064]: osd/ReplicatedPG.cc: In function 'int ReplicatedPG::do_osd_op
s(ReplicatedPG::OpContext*, std::vector<OSDOp, std::allocator<OSDOp> >&, ceph::bufferlist&)', in t
hread '0x7f2c6795b700'#012osd/ReplicatedPG.cc: 1211: FAILED assert(0 == "oi disagrees with stat")
Apr  9 08:17:04 atom9 osd.39[16064]: ceph version 0.25-588-g24caedc (commit:24caedc8f549eeeba48b2
```

```
d4a44927ee16e65c42a)#012 1: (ReplicatedPG::do_osd_ops(ReplicatedPG::OpContext*, std::vector<OSDOp,
std::allocator<OSDOp> >&, ceph::buffer::list&)+0xa2e2) [0x4b8f62]#012 2: (ReplicatedPG::prepare_t
ransaction(ReplicatedPG::OpContext*)+0x6f) [0x4b9bcf]#012 3: (ReplicatedPG::do_op(MOSDOp*)+0xa09)
[0x4bb2e9]#012 4: (OSD::dequeue_op(PG*)+0x37d) [0x4f54ad]#012 5: (ThreadPool::worker()+0x519) [0x6
0ab39]#012 6: (ThreadPool::WorkThread::entry()+0xd) [0x524efd]#012 7: (()+0x69ca) [0x7f2c748e69ca]
#012 8: (clone()+0x6d) [0x7f2c7357270d]
Apr 9 08:17:04 atom9 osd.39[16064]: ceph version 0.25-588-g24caedc (commit:24caedc8f549eeeba48b2
d4a44927ee16e65c42a)#012 1: (ReplicatedPG::do_osd_ops(ReplicatedPG::OpContext*, std::vector<OSDOp,
std::allocator<OSDOp> >&, ceph::buffer::list&)+0xa2e2) [0x4b8f62]#012 2: (ReplicatedPG::prepare_t
ransaction(ReplicatedPG::OpContext*)+0x6f) [0x4b9bcf]#012 3: (ReplicatedPG::do_op(MOSDOp*)+0xa09)
[0x4bb2e9]#012 4: (OSD::dequeue_op(PG*)+0x37d) [0x4f54ad]#012 5: (ThreadPool::worker()+0x519) [0x6
0ab39]#012 6: (ThreadPool::WorkThread::entry()+0xd) [0x524efd]#012 7: (()+0x69ca) [0x7f2c748e69ca]
#012 8: (clone()+0x6d) [0x7f2c7357270d]
Apr 9 08:17:04 atom9 osd.39[16064]: *** Caught signal (Aborted) **#012 in thread 0x7f2c6795b700
Apr 9 08:17:04 atom9 osd.39[16064]: ceph version 0.25-588-g24caedc (commit:24caedc8f549eeeba48b2
d4a44927ee16e65c42a)#012 1: /usr/bin/cosd() [0x6217ee]#012 2: (()+0xf8f0) [0x7f2c748ef8f0]#012 3:
(gsignal()+0x35) [0x7f2c734bfa75]#012 4: (abort()+0x180) [0x7f2c734c35c0]#012 5: (__gnu_cxx::__ver
bose_terminate_handler()+0x115) [0x7f2c73d758e5]#012 6: (()+0xcad16) [0x7f2c73d73d16]#012 7: (()+
xcad43) [0x7f2c73d73d43]#012 8: (()+0xcae3e) [0x7f2c73d73e3e]#012 9: (ceph::__ceph_assert_fail(char
const*, char const*, int, char const*)+0x36a) [0x607caa]#012 10: (ReplicatedPG::do_osd_ops(Repli
catedPG::OpContext*, std::vector<OSDOp, std::allocator<OSDOp> >&, ceph::buffer::list&)+0xa2e2) [0x
4b8f62]#012 11: (ReplicatedPG::prepare_transaction(ReplicatedPG::OpContext*)+0x6f) [0x4b9bcf]#012
12: (ReplicatedPG::do_op(MOSDOp*)+0xa09) [0x4bb2e9]#012 13: (OSD::dequeue_op(PG*)+0x37d) [0x4f54ad
]#012 14: (ThreadPool::worker()+0x519) [0x60ab39]#012 15: (ThreadPool::WorkThread::entry()+0xd) [0
x524efd]#012 16: (()+0x69ca) [0x7f2c748e69ca]#012 17: (clone()+0x6d) [0x7f2c7357270d]
```

I'm not sure if it is related to the reboot of my client, but it did happen at the same time.

History

#1 - 04/10/2011 09:28 AM - Greg Farnum

The OSD code was briefly broken last week -- it looks like you managed to pull down a bad one. Get the latest master and this won't be a problem. :)

#2 - 04/11/2011 09:22 PM - Sage Weil

- Target version set to v0.28

#3 - 04/11/2011 09:22 PM - Sage Weil

- Target version changed from v0.28 to v0.26.1

#4 - 04/23/2011 01:37 PM - Sage Weil

- Target version changed from v0.26.1 to v0.27.1

#5 - 05/02/2011 06:15 PM - Sage Weil

- Status changed from New to Can't reproduce