

Ceph - Bug #12814

flipping the overlay from forward to seems to reorder writes

08/27/2015 11:54 PM - Samuel Just

Status: Can't reproduce	Start date: 08/27/2015
Priority: Urgent	Due date:
Assignee: Sage Weil	% Done: 0%
Category:	Estimated time: 0.00 hour
Target version:	Spent time: 0.00 hour
Source: Q/A	Reviewed:
Tags:	Affected Versions:
Backport:	ceph-qa-suite: rados
Regression: No	Pull request ID:
Severity: 3 - minor	
Description ubuntu@teuthology:/a/samuelj-2015-08-26_12:18:34-rados-wip-sam-testing-distro-basic-multi/1033948	
Related issues:	
Related to Ceph - Backport #15171: hammer: osd: corruption when min_read_rece...	Resolved
Duplicated by Ceph - Bug #13013: Cache tier thrashing causes out of order write	Duplicate 09/09/2015
Duplicated by Ceph - Bug #14306: "RadosModel.h: 854: FAILED assert(0)" in rad...	Duplicate 01/08/2016

Associated revisions

Revision d7da6884 - 03/28/2016 08:57 PM - Sage Weil

osd: add 'proxy' cache mode

The 'forward' mode is flawed for the same reason that redirects are flawed: they allow the reordering of requests. Add a proxy mode that can be used instead of forward. (Later we can consider dropping forward entirely?)

Fixes: #12814

Signed-off-by: Sage Weil <sage@redhat.com>

History

#1 - 09/06/2015 05:49 PM - Sage Weil

maybe /a/sage-2015-09-06_05:36:07-rados-wip-sage-testing---basic-multi/1044769 ?

```
2015-09-06T07:42:40.338 INFO:tasks.rados.rados.0.burnupi24.stderr:Error: finished tid 2 when last_acked_tid was 6
```

#2 - 11/14/2015 06:12 PM - Sage Weil

/a/sage-2015-11-14_04:50:22-rados-wip-sage-testing---basic-multi/1149284

#3 - 11/17/2015 03:57 PM - Sage Weil

- Duplicated by Bug #13013: Cache tier thrashing causes out of order write added

#4 - 11/23/2015 05:41 PM - Samuel Just

/a/teuthology-2015-11-21_21:00:01-rados-hammer-distro-basic-openstack/17756

#5 - 11/26/2015 01:36 PM - Sage Weil

/a/sage-2015-11-25_07:26:30-rados-wip-sage-testing---basic-multi/1160057

#6 - 11/26/2015 02:07 PM - Sage Weil

```
teuthology:1160057 05:57 AM $ egrep 'client.4119.0:3052[456789]' remote/*/log/*osd* | grep === | sort -k 2 |
grep e23
remote/burnupi26/log/ceph-osd.5.log:2015-11-25 14:38:24.300618 7fb4a7bab700 1 -- 10.214.135.38:6808/10423 <==
client.4119 10.214.135.38:0/3782606168 87 ==== osd_op(client.4119.0:30524 1.65b401d9 (undecoded) ack+ondisk+r
etry+write+redirected+known_if_redirected e23) v7 ==== 243+0+734444 (3093478145 0 1938170131) 0x7fb4d860e680 c
on 0x7fb4d7267200
remote/mira057/log/ceph-osd.2.log:2015-11-25 14:38:24.320914 7f1b33bbe700 1 -- 10.214.134.136:6800/3338 <== c
lient.4119 10.214.135.38:0/3782606168 36 ==== osd_op(client.4119.0:30525 2.65b401d9 (undecoded) ack+ondisk+wri
te+known_if_redirected e23) v7 ==== 205+0+509614 (4268985705 0 4255468290) 0x7f1b64847c80 con 0x7f1b64272300
remote/burnupi26/log/ceph-osd.5.log:2015-11-25 14:38:24.325428 7fb4a7bab700 1 -- 10.214.135.38:6808/10423 <==
client.4119 10.214.135.38:0/3782606168 88 ==== osd_op(client.4119.0:30529 1.65b401d9 (undecoded) ack+ondisk+r
etry+write+redirected+known_if_redirected e23) v7 ==== 205+0+509614 (3250234563 0 4255468290) 0x7fb4d8f9b440 c
on 0x7fb4d7267200
remote/mira057/log/ceph-osd.2.log:2015-11-25 14:38:24.331382 7f1b33bbe700 1 -- 10.214.134.136:6800/3338 <== c
lient.4119 10.214.135.38:0/3782606168 37 ==== osd_op(client.4119.0:30526 2.65b401d9 (undecoded) ack+ondisk+wri
te+known_if_redirected e23) v7 ==== 205+0+281605 (2638626519 0 3821793884) 0x7f1b63505e40 con 0x7f1b64272300
remote/mira057/log/ceph-osd.2.log:2015-11-25 14:38:24.331509 7f1b33bbe700 1 -- 10.214.134.136:6800/3338 <== c
lient.4119 10.214.135.38:0/3782606168 38 ==== osd_op(client.4119.0:30527 2.65b401d9 (undecoded) ack+ondisk+wri
te+known_if_redirected e23) v7 ==== 243+0+65 (3223704887 0 2138031401) 0x7f1b63695440 con 0x7f1b64272300
remote/mira057/log/ceph-osd.2.log:2015-11-25 14:38:24.331607 7f1b33bbe700 1 -- 10.214.134.136:6800/3338 <== c
lient.4119 10.214.135.38:0/3782606168 39 ==== osd_op(client.4119.0:30528 2.65b401d9 (undecoded) ack+read+rword
ered+known_if_redirected e23) v7 ==== 205+0+0 (1396123015 0 0) 0x7f1b63422ec0 con 0x7f1b64272300
```

the problematic op is the first one, client.4119.0:30524. why does it get sent with flag 'redirected' to osd.5?

#7 - 11/27/2015 12:45 PM - Loic Dachary

testing infernalis `rados singleton/{all/thrash_cache_writeback_forward_none.yaml fs/btrfs.yaml msgr/simple.yaml msgr-failures/few.yaml}`, ran into it, just in case it may be helpful to know

```
2015-11-26T09:06:05.170 INFO:tasks.rados.rados.0.target167160.stdout:2400: finishing write tid 1 to target167160.teuthology9942-2400
2015-11-26T09:06:05.171 INFO:tasks.rados.rados.0.target167160.stderr:Error: finished tid 1 when last_acked_tid was 5
2015-11-26T09:06:05.178 INFO:tasks.rados.rados.0.target167160.stderr:./test/osd/RadosModel.h: In function 'virtual void WriteOp::_finish(TestOp::CallbackInfo*)' thread 7f649bfff700 time 2015-11-26 09:06:05.080725
2015-11-26T09:06:05.178 INFO:tasks.rados.rados.0.target167160.stderr:./test/osd/RadosModel.h: 862: FAILED assert(0)
2015-11-26T09:06:05.180 INFO:tasks.rados.rados.0.target167160.stderr: ceph version 9.2.0-46-g97293b5 (97293b554026cff96da59760aeb577660867a2ac)
2015-11-26T09:06:05.180 INFO:tasks.rados.rados.0.target167160.stderr: 1: (ceph::__ceph_assert_fail(char const*, char const*, int, char const*)+0x8b) [0x7f64b341ffdb]
2015-11-26T09:06:05.180 INFO:tasks.rados.rados.0.target167160.stderr: 2: (WriteOp::_finish(TestOp::CallbackInfo*)+0x4b8) [0x7f64b340b8a8]
2015-11-26T09:06:05.180 INFO:tasks.rados.rados.0.target167160.stderr: 3: (write_callback(void*, void*)+0x19) [0x7f64b3417379]
2015-11-26T09:06:05.180 INFO:tasks.rados.rados.0.target167160.stderr: 4: (()+0x9ccad) [0x7f64b0c33cad]
2015-11-26T09:06:05.181 INFO:tasks.rados.rados.0.target167160.stderr: 5: (()+0x77fc9) [0x7f64b0c0efc9]
2015-11-26T09:06:05.181 INFO:tasks.rados.rados.0.target167160.stderr: 6: (()+0x13bef8) [0x7f64b0cd2ef8]
2015-11-26T09:06:05.181 INFO:tasks.rados.rados.0.target167160.stderr: 7: (()+0x8182) [0x7f64b076b182]
2015-11-26T09:06:05.181 INFO:tasks.rados.rados.0.target167160.stderr: 8: (clone()+0x6d) [0x7f64af2f547d]
2015-11-26T09:06:05.181 INFO:tasks.rados.rados.0.target167160.stderr: NOTE: a copy of the executable, or `objdump -rdS <executable>` is needed to interpret this.
```

#8 - 11/27/2015 04:01 PM - Sage Weil

- Status changed from New to In Progress
- Assignee set to Sage Weil
- Source changed from other to Q/A

#9 - 12/02/2015 01:30 AM - bo cai

Hi,sage,how is this going?

#10 - 01/08/2016 06:25 PM - Samuel Just

- Duplicated by Bug #14306: "RadosModel.h: 854: FAILED assert(0)" in rados-hammer-distro-basic-openstack added

#11 - 02/25/2016 06:27 AM - David Zafman

Seen again in my hammer backport testing:

description: rados:singleton:all/thrash_cache_writeback_forward_none.yaml

teuthology:/a/dzafman-2016-02-24_20:32:40-rados:singleton:all-wip-14077-hammer---basic-smithi/26379

#12 - 03/08/2016 10:13 PM - Yuri Weinstein

Seen in run:

http://pulpito.ovh.sepia.ceph.com:8081/teuthology-2016-03-08_02:00:01-rados-infernalis-distro-basic-openstack/

Job: 11576

Logs: http://teuthology.ovh.sepia.ceph.com/teuthology/teuthology-2016-03-08_02:00:01-rados-infernalis-distro-basic-openstack/11576/teuthology.log

```
2016-03-08T03:39:03.627 INFO:tasks.rados.rados.0.target078044.stdout:9241: writing target078044.ovh.sepia.ceph.com13368-9241 from 1639782 to 2066647 tid 2
2016-03-08T03:39:03.627 INFO:tasks.rados.rados.0.target078044.stdout: waiting on 16
2016-03-08T03:39:03.691 INFO:tasks.rados.rados.0.target078044.stdout:9213: finishing write tid 2 to target078044.ovh.sepia.ceph.com13368-9213
2016-03-08T03:39:03.691 INFO:tasks.rados.rados.0.target078044.stderr:Error: finished tid 2 when last_acked_tid was 6
2016-03-08T03:39:03.701 INFO:tasks.rados.rados.0.target078044.stderr:./test/osd/RadosModel.h: In function 'virtual void WriteOp::_finish(TestOp::CallbackInfo*)' thread 7f8e22ffd700 time 2016-03-08 03:39:03.617390
2016-03-08T03:39:03.701 INFO:tasks.rados.rados.0.target078044.stderr:./test/osd/RadosModel.h: 862: FAILED assert(0)
2016-03-08T03:39:03.703 INFO:tasks.rados.rados.0.target078044.stderr: ceph version 9.2.1-5-g5eb2d6c (5eb2d6cf1ecb0c2d9b3caee10eff99c24d82b484)
2016-03-08T03:39:03.703 INFO:tasks.rados.rados.0.target078044.stderr: 1: (ceph::__ceph_assert_fail(char const*, char const*, int, char const*)+0x8b) [0x7f8e31f85f9b]
2016-03-08T03:39:03.704 INFO:tasks.rados.rados.0.target078044.stderr: 2: (WriteOp::_finish(TestOp::CallbackInfo*)+0x4b8) [0x7f8e31f71868]
2016-03-08T03:39:03.704 INFO:tasks.rados.rados.0.target078044.stderr: 3: (write_callback(void*, void*)+0x19) [0x7f8e31f7d339]
2016-03-08T03:39:03.704 INFO:tasks.rados.rados.0.target078044.stderr: 4: (()+0x9cc7d) [0x7f8e2f799c7d]
2016-03-08T03:39:03.704 INFO:tasks.rados.rados.0.target078044.stderr: 5: (()+0x77f89) [0x7f8e2f774f89]
2016-03-08T03:39:03.704 INFO:tasks.rados.rados.0.target078044.stderr: 6: (()+0x13be88) [0x7f8e2f838e88]
2016-03-08T03:39:03.705 INFO:tasks.rados.rados.0.target078044.stderr: 7: (()+0x8182) [0x7f8e2f2d1182]
2016-03-08T03:39:03.705 INFO:tasks.rados.rados.0.target078044.stderr: 8: (clone()+0x6d) [0x7f8e2de5b47d]
2016-03-08T03:39:03.705 INFO:tasks.rados.rados.0.target078044.stderr: NOTE: a copy of the executable, or `objdump -rds <executable>` is needed to interpret this.
```

#13 - 03/08/2016 10:14 PM - Yuri Weinstein

- Release set to *infernalis*
- *ceph-qa-suite rados* added

#14 - 03/16/2016 07:36 PM - Sage Weil

- Status changed from *In Progress* to *Verified*
- Assignee deleted (*Sage Weil*)

#15 - 03/16/2016 07:42 PM - Greg Farnum

Looks like a user on the mailing list ran into this: <https://www.mail-archive.com/ceph-users@lists.ceph.com/msg27680.html>
Apparently switching between writeback and forward is useful in some cases for getting a read cache even if your cache pool can't handle the write throughput. 0_o

#16 - 03/17/2016 01:15 PM - Sage Weil

- Related to Backport #15171: *hammer: osd: corruption when min_read_recency_for_promote > 1* added

#17 - 03/17/2016 04:55 PM - Ricardo Dias

The same problem occurred in the following job:
http://pulpito.ceph.com/joao-2016-03-17_07:09:30-rados:thrash-wip-joao-testing---basic-multi/68628

#18 - 03/18/2016 03:16 PM - Sage Weil

- Assignee set to *Sage Weil*

#19 - 03/18/2016 03:55 PM - Sage Weil

- Status changed from *Verified* to *In Progress*

creating a 'proxy' mode that replaces 'forward'.

#20 - 03/28/2016 03:21 PM - Sage Weil

- Priority changed from *Urgent* to *Immediate*

#21 - 03/30/2016 12:24 PM - Sage Weil

- Status changed from *In Progress* to *Resolved*

#22 - 03/30/2016 10:43 PM - Sage Weil

- Status changed from *Resolved* to *In Progress*

http://pulpito.ceph.com/sage-2016-03-30_13:23:09-rados-wip-sage-testing-distro-basic-smithi/98188

```
75276277278279280281282283284285286287288289290291292293294295296297298299 version is 6730 and expected 8776
2016-03-30T15:25:29.310 INFO:tasks.rados.rados.0.smithi027.stderr:./test/osd/RadosModel.h: In function 'virtua
l void ReadOp::_finish(TestOp::CallbackInfo*)' thread 7fab15ffb700 time 2016-03-30 22:25:29.293395
2016-03-30T15:25:29.311 INFO:tasks.rados.rados.0.smithi027.stderr:./test/osd/RadosModel.h: 1169: FAILED assert
(version == old_value.version)
```

#23 - 04/04/2016 10:32 AM - Loic Dachary

on the hammer branch

```
description: rados/singleton/{all/thrash_cache_writeback_forward_none.yaml fs/btrfs.yaml
  msgr-failures/many.yaml}
...
2016-04-03T19:42:03.114 INFO:tasks.rados.rados.0.target167114240148.stdout:5143: finishing write tid 2 to target167114240148.teuthology9910-5143
2016-04-03T19:42:03.114 INFO:tasks.rados.rados.0.target167114240148.stderr:Error: finished tid 2 when last_ack_ed_tid was 7
2016-04-03T19:42:03.115 INFO:tasks.rados.rados.0.target167114240148.stderr:./test/osd/RadosModel.h: In function 'virtual void WriteOp::_finish(TestOp::CallbackInfo*)' thread 7f7754ff9700 time 2016-04-03 19:42:03.103186
2016-04-03T19:42:03.115 INFO:tasks.rados.rados.0.target167114240148.stderr:./test/osd/RadosModel.h: 854: FAILED assert(0)
2016-04-03T19:42:03.116 INFO:tasks.rados.rados.0.target167114240148.stderr: ceph version 0.94.6-201-g7cfc739 (7cfc739a513da84d76b959efc51dd21c1cc552d6)
2016-04-03T19:42:03.116 INFO:tasks.rados.rados.0.target167114240148.stderr: 1: (ceph::__ceph_assert_fail(char const*, char const*, int, char const*)+0x8b) [0x4e101b]
2016-04-03T19:42:03.116 INFO:tasks.rados.rados.0.target167114240148.stderr: 2: (WriteOp::_finish(TestOp::CallbackInfo*)+0x46e) [0x4c8dce]
2016-04-03T19:42:03.116 INFO:tasks.rados.rados.0.target167114240148.stderr: 3: (write_callback(void*, void*)+0x19) [0x4d8619]
2016-04-03T19:42:03.116 INFO:tasks.rados.rados.0.target167114240148.stderr: 4: (()+0x94fcd) [0x7f77617d8fcd]
2016-04-03T19:42:03.117 INFO:tasks.rados.rados.0.target167114240148.stderr: 5: (()+0x70319) [0x7f77617b4319]
2016-04-03T19:42:03.117 INFO:tasks.rados.rados.0.target167114240148.stderr: 6: (()+0x134348) [0x7f7761878348]
2016-04-03T19:42:03.117 INFO:tasks.rados.rados.0.target167114240148.stderr: 7: (()+0x8182) [0x7f7761318182]
2016-04-03T19:42:03.117 INFO:tasks.rados.rados.0.target167114240148.stderr: 8: (clone()+0x6d) [0x7f775fc9d47d]
2016-04-03T19:42:03.117 INFO:tasks.rados.rados.0.target167114240148.stderr: NOTE: a copy of the executable, or `objdump -rdS <executable>` is needed to interpret this.
2016-04-03T19:42:03.117 INFO:tasks.rados.rados.0.target167114240148.stderr:terminate called after throwing an instance of 'ceph::FailedAssertion'
```

#24 - 04/04/2016 07:34 PM - Sage Weil

removed the test from the hammer suite <https://github.com/ceph/ceph-ga-suite/pull/931>

#25 - 04/06/2016 02:37 PM - Sage Weil

```
2016-03-30 22:25:14.308757 7fb988321700 10 osd.5 pg_epoch: 20 pg[2.2( v 20'6027 (19'2932,20'6027) local-les=8
```

n=14 ec=6 les/c/f 8/8/0 6/6/6) [5,1] r=0 lpr=6 luod=20'6026 crt=20'6025 lcod 20'6025 mlcod 20'6025 active+clea
n] do_op osd_op(client.4118.0:46216 2.1aeca46e smithi0272108-7783 01234567891011121314151617181920212223242526
27282930313233343536373839404142434445464748495051525354555657585960616263646566676869707172737475767778798081
82838485868788899091929394959697989910010110210310410510610710810911011111211311411511611711811912012112212312
41251261271281291301311321331341351361371381391401411421431441451461471481491501511521531541551561571581591601
61162163164165166167168169170171172173174175176177178179180181182183184185186187188189190191192193194195196197
19819920020120220320420520620720820921021121221321421521621721821922022122222322422522622722822923234235236237
42352362372382392402412422432442452462472482492502512522532542552562572582592602612622632642652662672682692702
71272273274275276277278279280281282283284285286287288289290291292293294295296297298299 [write 712067~706736] s
nappc 0=[] ack+ondisk+write+known_if_redirected e20) v7 may_write -> write-ordered flags ack+ondisk+write+known
_if_redirected
2016-03-30 22:25:14.308785 7fb988321700 10 osd.5 pg_epoch: 20 pg[2.2(v 20'6027 (19'2932,20'6027) local-les=8
n=14 ec=6 les/c/f 8/8/0 6/6/6) [5,1] r=0 lpr=6 luod=20'6026 crt=20'6025 lcod 20'6025 mlcod 20'6025 active+clea
n] get_object_context: obc NOT found in cache: 2:76253758::smithi0272108-7783 0123456789101112131415161718192
02122232425262728293031323334353637383940414243444546474849505152535455565758596061626364656667686970717273747
57677787980818283848586878889909192939495969798991001011021031041051061071081091101111121131141151161171181191
20121122123124125126127128129130131132133134135136137138139140141142143144145146147148149150151152153154155156
15715815916016116216316416516616716816917017117217317417517617717817918018118218318418518618718818919019119219
3194195196197198199200201202203204205206207208209210211212213214215216217218219220221222223224225226227228229
30231232233234235236237238239240241242243244245246247248249250251252253254255256257258259260261262263264265266
267268269270271272273274275276277278279280281282283284285286287288289290291292293294295296297298299:head
2016-03-30 22:25:14.308822 7fb988321700 15 filestore(/var/lib/ceph/osd/ceph-5) getattr 2.2_head/#2:76253758::
smithi0272108-7783 0123456789101112131415161718192021222324252627282930313233343536373839404142434445464748495
05152535455565758596061626364656667686970717273747576777879808182838485868788899091929394959697989910010110210
31041051061071081091101111121131141151161171181191201211221231241251261271281291301311321331341351361371381391
40141142143144145146147148149150151152153154155156157158159160161162163164165166167168169170171172173174175176
17717817918018118218318418518618718818919019119219319419519619719819920020120220320420520620720820921021121221
32142152162172182192202212222232242252262272282292302312322332342352362372382392402412422432442452462472482492
50251252253254255256257258259260261262263264265266267268269270271272273274275276277278279280281282283284285286
287288289290291292293294295296297298299:head# '_'
2016-03-30 22:25:14.308944 7fb988321700 10 filestore(/var/lib/ceph/osd/ceph-5) error opening file /var/lib/cep
h/osd/ceph-5/current/2.2_head/smithi0272108-7783 0123456789101112131415161718192021222324252627282930313233343
53637383940414243444546474849505152535455565758596061626364656667686970717273747576777879808182838485868788899
0919293949596979899100101102103104105_44cfbf811b9e7fd5862e_0_long with flags=2: (2) No such file or directory
2016-03-30 22:25:14.308953 7fb988321700 10 filestore(/var/lib/ceph/osd/ceph-5) getattr 2.2_head/#2:76253758::
smithi0272108-7783 0123456789101112131415161718192021222324252627282930313233343536373839404142434445464748495
05152535455565758596061626364656667686970717273747576777879808182838485868788899091929394959697989910010110210
31041051061071081091101111121131141151161171181191201211221231241251261271281291301311321331341351361371381391
40141142143144145146147148149150151152153154155156157158159160161162163164165166167168169170171172173174175176
17717817918018118218318418518618718818919019119219319419519619719819920020120220320420520620720820921021121221
32142152162172182192202212222232242252262272282292302312322332342352362372382392402412422432442452462472482492
50251252253254255256257258259260261262263264265266267268269270271272273274275276277278279280281282283284285286
287288289290291292293294295296297298299:head# '_' = -2
2016-03-30 22:25:14.308962 7fb988321700 10 osd.5 pg_epoch: 20 pg[2.2(v 20'6027 (19'2932,20'6027) local-les=8
n=14 ec=6 les/c/f 8/8/0 6/6/6) [5,1] r=0 lpr=6 luod=20'6026 crt=20'6025 lcod 20'6025 mlcod 20'6025 active+clea
n] get_object_context: no obc for soid 2:76253758::smithi0272108-7783 012345678910111213141516171819202122232
4252627282930313233343536373839404142434445464748495051525354555657585960616263646566676869707172737475767778
798081828384858687888990919293949596979899100101102103104105106107108109110111112113114115116117118119120121122
12312412512612712812913013113213313413513613713813914014114214314414514614714814915015115215315415515615715815
91601611621631641651661671681691701711721731741751761771781791801811821831841851861871881891901911921931941951
96197198199200201202203204205206207208209210211212213214215216217218219220221222223224225226227228229230231232
23323423523623723823924024124224324424524624724824925025125225325425525625725825926026126226326426526626726826
9270271272273274275276277278279280281282283284285286287288289290291292293294295296297298299:head but can_creat
e
2016-03-30 22:25:14.308979 7fb988321700 15 filestore(/var/lib/ceph/osd/ceph-5) getattr 2.2_head/#2:76253758::
smithi0272108-7783 0123456789101112131415161718192021222324252627282930313233343536373839404142434445464748495
05152535455565758596061626364656667686970717273747576777879808182838485868788899091929394959697989910010110210
31041051061071081091101111121131141151161171181191201211221231241251261271281291301311321331341351361371381391
40141142143144145146147148149150151152153154155156157158159160161162163164165166167168169170171172173174175176
17717817918018118218318418518618718818919019119219319419519619719819920020120220320420520620720820921021121221
32142152162172182192202212222232242252262272282292302312322332342352362372382392402412422432442452462472482492
50251252253254255256257258259260261262263264265266267268269270271272273274275276277278279280281282283284285286
287288289290291292293294295296297298299:snapdir# 'snapset'
2016-03-30 22:25:14.309051 7fb988321700 10 filestore(/var/lib/ceph/osd/ceph-5) error opening file /var/lib/cep
h/osd/ceph-5/current/2.2_head/smithi0272108-7783 0123456789101112131415161718192021222324252627282930313233343
53637383940414243444546474849505152535455565758596061626364656667686970717273747576777879808182838485868788899
0919293949596979899100101102103104105_a23dc54bec6480cad5a6_0_long with flags=2: (2) No such file or directory
2016-03-30 22:25:14.309057 7fb988321700 10 filestore(/var/lib/ceph/osd/ceph-5) getattr 2.2_head/#2:76253758::
smithi0272108-7783 0123456789101112131415161718192021222324252627282930313233343536373839404142434445464748495
05152535455565758596061626364656667686970717273747576777879808182838485868788899091929394959697989910010110210
31041051061071081091101111121131141151161171181191201211221231241251261271281291301311321331341351361371381391
40141142143144145146147148149150151152153154155156157158159160161162163164165166167168169170171172173174175176
17717817918018118218318418518618718818919019119219319419519619719819920020120220320420520620720820921021121221
32142152162172182192202212222232242252262272282292302312322332342352362372382392402412422432442452462472482492
50251252253254255256257258259260261262263264265266267268269270271272273274275276277278279280281282283284285286

287288289290291292293294295296297298299:snaptid# 'snapset' = -2
2016-03-30 22:25:14.309084 7fb988321700 10 osd.5 pg_epoch: 20 pg[2.2(v 20'6027 (19'2932,20'6027] local-les=8
n=14 ec=6 les/c/f 8/8/0 6/6/6) [5,1] r=0 lpr=6 luod=20'6026 crt=20'6025 lcod 20'6025 mlcod 20'6025 active+clea
n] create_object_context 0x7fb9ae22f180 2:76253758::smithi0272108-7783 01234567891011121314151617181920212223
24252627282930313233343536373839404142434445464748495051525354555657585960616263646566676869707172737475767778
798081828384858687888990919293949596979899100101102103104105106107108109110111112113114115116117118119120121122
123124125126127128129130131132133134135136137138139140141142143144145146147148149150151152153154155156157158159160161162163164165166167168169170171172173174175176177178179180181182183184185186187188189190191192193194195
19619719819920020120220320420520620720820921021121221321421521621721821922022122223224225226227228229230231232
233234235236237238239240241242243244245246247248249250251252253254255256257258259260261262263264265266267268269
270271272273274275276277278279280281282283284285286287288289290291292293294295296297298299:head
2016-03-30 22:25:14.309099 7fb988321700 10 osd.5 pg_epoch: 20 pg[2.2(v 20'6027 (19'2932,20'6027] local-les=8
n=14 ec=6 les/c/f 8/8/0 6/6/6) [5,1] r=0 lpr=6 luod=20'6026 crt=20'6025 lcod 20'6025 mlcod 20'6025 active+clea
n] populate_obc_watchers 2:76253758::smithi0272108-7783 01234567891011121314151617181920212223242526272829303
13233343536373839404142434445464748495051525354555657585960616263646566676869707172737475767778798081828384858
68788899091929394959697989910010110210310410510610710810911011111211311411511611711811912012112212312412512612
71281291301311321331341351361371381391401411421431441451461471481491501511521531541551561571581591601611621631
64165166167168169170171172173174175176177178179180181182183184185186187188189190191192193194195196197198199200
2012022032042052062072082092102112122132142152162172182192202212222322422522622722822923023123223323423523623
72382392402412422432442452462472482492502512522532542552562572582592602612622632642652662672682692702712722732
74275276277278279280281282283284285286287288289290291292293294295296297298299:head
2016-03-30 22:25:14.309111 7fb988321700 20 osd.5 pg_epoch: 20 pg[2.2(v 20'6027 (19'2932,20'6027] local-les=8
n=14 ec=6 les/c/f 8/8/0 6/6/6) [5,1] r=0 lpr=6 luod=20'6026 crt=20'6025 lcod 20'6025 mlcod 20'6025 active+clea
n] ReplicatedPG::check_blacklisted_obc_watchers for obc 2:76253758::smithi0272108-7783 0123456789101112131415
16171819202122232425262728293031323334353637383940414243444546474849505152535455565758596061626364656667686970
71727374757677787980818283848586878889909192939495969798991001011021031041051061071081091101111121131141151161
17118119120121122123124125126127128129130131132133134135136137138139140141142143144145146147148149150151152153
15415515615715815916016116216316416516616716816917017117217317417517617717817918018118218318418518618718818919
0191192193194195196197198199200201202203204205206207208209210211212213214215216217218219220221222232242252262
27228229230231232233234235236237238239240241242243244245246247248249250251252253254255256257258259260261262263
264265266267268269270271272273274275276277278279280281282283284285286287288289290291292293294295296297298299:h
ead
2016-03-30 22:25:14.309123 7fb988321700 10 osd.5 pg_epoch: 20 pg[2.2(v 20'6027 (19'2932,20'6027] local-les=8
n=14 ec=6 les/c/f 8/8/0 6/6/6) [5,1] r=0 lpr=6 luod=20'6026 crt=20'6025 lcod 20'6025 mlcod 20'6025 active+clea
n] get_object_context: 0x7fb9ae22f180 2:76253758::smithi0272108-7783 0123456789101112131415161718192021222324
25262728293031323334353637383940414243444546474849505152535455565758596061626364656667686970717273747576777879
80818283848586878889909192939495969798991001011021031041051061071081091101111121131141151161171181191201211221
23124125126127128129130131132133134135136137138139140141142143144145146147148149150151152153154155156157158159
16016116216316416516616716816917017117217317417517617717817918018118218318418518618718818919019119219319419519
6197198199200201202203204205206207208209210211212213214215216217218219220221222232242252262272282292302312322
33234235236237238239240241242243244245246247248249250251252253254255256257258259260261262263264265266267268269
270271272273274275276277278279280281282283284285286287288289290291292293294295296297298299:head rwstate(none n
=0 w=0) oi: 2:76253758::smithi0272108-7783 012345678910111213141516171819202122232425262728293031323334353637
38394041424344454647484950515253545556575859606162636465666768697071727374757677787980818283848586878889909192
93949596979899100101102103104105106107108109110111112113114115116117118119120121122123124125126127128129130131
13213313413513613713813914014114214314414514614714814915015115215315415515615715815916016116216316416516616716
8169170171172173174175176177178179180181182183184185186187188189190191192193194195196197198199200201202203204
0520620720820921021121221321421521621721821922022122223224225226227228229230231232233234235236237238239240241
24224324424524624724824925025125225325425525625725825926026126226326426526626726826927027127227327427527627727
8279280281282283284285286287288289290291292293294295296297298299:head(0'0 unknown.0.0:0 s 0 uv 0) ssc: 0x7fb9a
e80b2c0 snapset: 0=[]:[]
2016-03-30 22:25:14.309142 7fb988321700 10 osd.5 pg_epoch: 20 pg[2.2(v 20'6027 (19'2932,20'6027] local-les=8
n=14 ec=6 les/c/f 8/8/0 6/6/6) [5,1] r=0 lpr=6 luod=20'6026 crt=20'6025 lcod 20'6025 mlcod 20'6025 active+clea
n] find_object_context 2:76253758::smithi0272108-7783 0123456789101112131415161718192021222324252627282930313
23334353637383940414243444546474849505152535455565758596061626364656667686970717273747576777879808182838485868
78889909192939495969798991001011021031041051061071081091101111121131141151161171181191201211221231241251261271
28129130131132133134135136137138139140141142143144145146147148149150151152153154155156157158159160161162163164
16516616716816917017117217317417517617717817918018118218318418518618718818919019119219319419519619719819920020
1202203204205206207208209210211212213214215216217218219220221222232242252262272282292302312322332342352362372
38239240241242243244245246247248249250251252253254255256257258259260261262263264265266267268269270271272273274
275276277278279280281282283284285286287288289290291292293294295296297298299:head @head oi=2:76253758::smithi0
272108-7783 01234567891011121314151617181920212223242526272829303132333435363738394041424344454647484950515253
54555657585960616263646566676869707172737475767778798081828384858687888990919293949596979899100101102103104105
10610710810911011111211311411511611711811912012112212312412512612712812913013113213313413513613713813914014114
21431441451461471481491501511521531541551561571581591601611621631641651661671681691701711721731741751761771781
7918018118218318418518618718818919019119219319419519619719819920020120220320420520620720820921021121213214215
2162172182192202212222322422522622722822923023123223323423523623723823924024124224324424524624724824925025125
22532542552562572582592602612622632642652662672682692702712722732742752762772782792802812822832842852862872882
89290291292293294295296297298299:head(0'0 unknown.0.0:0 s 0 uv 0)
2016-03-30 22:25:14.309162 7fb988321700 20 osd.5 pg_epoch: 20 pg[2.2(v 20'6027 (19'2932,20'6027] local-les=8
n=14 ec=6 les/c/f 8/8/0 6/6/6) [5,1] r=0 lpr=6 luod=20'6026 crt=20'6025 lcod 20'6025 mlcod 20'6025 active+clea
n] agent_choose_mode0x7fb9ac6fc800 delaying, ignored
2016-03-30 22:25:14.309179 7fb988321700 10 osd.5 pg_epoch: 20 pg[2.2(v 20'6027 (19'2932,20'6027] local-les=8
n=14 ec=6 les/c/f 8/8/0 6/6/6) [5,1] r=0 lpr=6 luod=20'6026 crt=20'6025 lcod 20'6025 mlcod 20'6025 active+clea
n] start_copy 2:76253758::smithi0272108-7783 0123456789101112131415161718192021222324252627282930313233343536

37383940414243444546474849505152535455565758596061626364656667686970717273747576777879808182838485868788899091
92939495969798991001011021031041051061071081091101111121131141151161171181191201211221231241251261271281291301
31132133134135136137138139140141142143144145146147148149150151152153154155156157158159160161162163164165166167
16816917017117217317417517617717817918018118218318418518618718818919019119219319419519619719819920020120220320
42052062072082092102112122132142152162172182192202212222232242252262272282292302312322332342352362372382392402
41242243244245246247248249250251252253254255256257258259260261262263264265266267268269270271272273274275276277
278279280281282283284285286287288289290291292293294295296297298299:head from 2:76253758::smithi0272108-7783 0
12345678910111213141516171819202122232425262728293031323334353637383940414243444546474849505152535455565758596
06162636465666768697071727374757677787980818283848586878889909192939495969798991001011021031041051061071081091
10111112113114115116117118119120121122123124125126127128129130131132133134135136137138139140141142143144145146
14714814915015115215315415515615715815916016116216316416516616716816917017117217317417517617717817918018118218
31841851861871881891901911921931941951961971981992002012022032042052062072082092102112122132142152162172182192
2022122223224225226227228229230231232233234235236237238239240241242243244245246247248249250251252253254255256
25725825926026126226326426526626726826927027127227327427527627727827928028128228328428528628728828929029129229
3294295296297298299:head @1 v0 flags 30 mirror_snapset

in epoch 20, the map is in proxy mode. i can't figure out why it didn't do a proxy write from the log. unfortunately the mon data tarball isn't complete, so i can't fully verify the osdmap content.

i think we need to wait for this to trigger again. enabling a higher osd log level.

#26 - 04/06/2016 02:37 PM - Sage Weil

- Status changed from *In Progress* to *Need More Info*

#27 - 04/08/2016 01:47 PM - Sage Weil

- Priority changed from *Immediate* to *Urgent*

#28 - 08/25/2016 04:56 PM - Samuel Just

- Status changed from *Need More Info* to *Can't reproduce*