

Ceph - Bug #847

OSD: Recovery does not complete, PG's stay active

02/28/2011 11:51 AM - Wido den Hollander

Status:	Resolved	Start date:	02/28/2011
Priority:	Normal	Due date:	
Assignee:		% Done:	0%
Category:	OSD	Estimated time:	0.00 hour
Target version:	v0.25	Spent time:	15.50 hours
Source:		Reviewed:	
Tags:		Affected Versions:	
Backport:		ceph-qa-suite:	
Regression:	No	Pull request ID:	
Severity:	3 - minor	Crash signature:	

Description

After I've expanded my cluster from 4 to 8 OSD's the recovery would not complete.

The state right now:

```
pg v112366: 8240 pgs: 48 active, 8192 active+clean; 260 GB data, 523 GB used, 14364 GB / 14904 GB avail; 3126/134690 degraded (2.321%)
```

First it were 64 PG's which stayed active, after some OSD restarts it got back to 48.

```
3.7c9 46 46 46 92 188416 192937984 4704 4704 active 316'242 13
81'458 [4,6] [4,6] 316'242 2011-02-22 07:29:12.919625
```

That PG for example, that one stays "active".

Browsing through the logs I found:

```
Feb 25 22:27:23 atom osd.6[18705]: 7f81df605700 osd6 1258 pg[3.7c9( v 316'242 lc 0'0 (316'238,316'
242)+backlog n=46 ec=2 les=1186 1258/1258/1248) [4,6] r=1 crashed+stray m=46] clear_recovery_state
Feb 25 22:27:23 atom osd.6[18705]: 7f81df605700 osd6 1258 pg[3.7c9( v 316'242 lc 0'0 (316'238,316'
242)+backlog n=46 ec=2 les=1186 1258/1258/1248) [4,6] r=1 crashed+stray m=46] up [4] -> [4,6], act
ing [4] -> [4,6], role -1 -> 1
Feb 25 22:27:23 atom osd.6[18705]: 7f81df605700 osd6 1258 pg[3.7c9( v 316'242 lc 0'0 (316'238,316'
242)+backlog n=46 ec=2 les=1186 1258/1258/1248) [4,6] r=1 crashed+stray m=46] on_role_change
Feb 25 22:27:23 atom osd.6[18705]: 7f81df605700 osd6 1258 pg[3.7c9( v 316'242 lc 0'0 (316'238,316'
242)+backlog n=46 ec=2 les=1186 1258/1258/1248) [4,6] r=1 crashed+stray m=46] cancel_generate_back
log
Feb 25 22:27:30 atom osd.4[18838]: 7f3a0d7b3700 osd4 1258 pg[3.7c9( v 316'242 lc 0'0 (316'238,316'
242)+backlog n=46 ec=2 les=1185 1258/1258/1248) [4,6] r=0 mlcod 0'0 crashed+replay+degraded m=46 u
=46] cancel_recovery
```

This process keeps repeating for this PG, the recovery gets cancelled?

In about 20 min a lot of these lines are printed, which then ends with:

```
Feb 25 22:47:44 atom osd.4[21301]: 7faa5870f700 osd4 1379 pg[3.7c9( v 316'242 lc 0'0 (316'238,316'
242)+backlog n=46 ec=2 les=
1186 1377/1377/1377) [4,6] r=0 mlcod 0'0 crashed+replay m=46 u=46] update_stats 1377'437
Feb 25 22:48:29 noisy osd.1[27292]: 7f31fb7a9700 osd1 1380 pg[3.7c9( v 316'242 (316'238,316'242)]+b
```

```
acklog n=46 ec=2 les=1185 1
380/1380/1380) [6] r=-1 lcod 0'0 stray] up [4,6] -> [6], acting [4,6] -> [6], role -1 -> -1
Feb 25 22:48:29 noisy osd.1[27292]: 7f31fb7a9700 osd1 1380 pg[3.7c9( v 316'242 (316'238,316'242)]+b
acklog n=46 ec=2 les=1185 1
380/1380/1380) [6] r=-1 lcod 0'0 stray] on_change
Feb 25 22:48:29 noisy osd.1[27292]: 7f31fb7a9700 osd1 1380 pg[3.7c9( v 316'242 (316'238,316'242)]+b
acklog n=46 ec=2 les=1185 1
380/1380/1380) [6] r=-1 lcod 0'0 stray] [4,6] -> [6], acting primary 4 -> 6
```

After that, it starts printing lines like, after which the PG seems "idle":

```
Feb 26 00:24:11 noisy osd.1[27292]: 7f3201fb6700 osd1 1389 on 2011-02-22 07:29:12.919625 3.7c9
Feb 26 00:24:12 noisy osd.1[27292]: 7f3201fb6700 osd1 1389 on 2011-02-22 07:29:12.919625 3.7c9
Feb 26 00:24:14 noisy osd.1[27292]: 7f3201fb6700 osd1 1389 on 2011-02-22 07:29:12.919625 3.7c9
```

What is osd1 doing with this PG? (This keeps repeating for thousands of loglines)

I do see this PG being scrubbed multiple times, but other than that, I see the "osd1" lines repeating.

I just restarted osd4 again and saw:

```
Feb 28 20:56:39 atom osd.4[24320]: 7f434613b720 filestore(/var/lib/ceph/osd.4) getattr /var/lib/ce
ph/osd.4/current/3.7c9_head/rb.0.1.000000002dd0_head '_'
Feb 28 20:56:39 atom osd.4[24320]: 7f434613b720 osd4 1415 pg[3.7c9( v 316'242 lc 0'0 (316'238,316'
242)]+backlog n=46 ec=2 les=1389 1381/1381/1381) [] r=0 mlcod 0'0 inactive m=40 u=40] read_log mis
sing 316'78 (316'77) b rb.0.1.000000002dd0/head by client4304.0:64776 2011-02-21 12:31:58.473405
Feb 28 20:56:39 atom osd.4[24320]: 7f434613b720 filestore(/var/lib/ceph/osd.4) getattr /var/lib/ce
ph/osd.4/current/3.7c9_head/rb.0.1.00000000117f_head '_' = -2
Feb 28 20:56:39 atom osd.4[24320]: 7f434613b720 filestore(/var/lib/ceph/osd.4) getattr /var/lib/ce
ph/osd.4/current/3.7c9_head/rb.0.1.0000000010be_head '_' = -2
Feb 28 20:56:39 atom osd.4[24320]: 7f434613b720 osd4 1415 pg[3.7c9( v 316'242 lc 0'0 (316'238,316'
242)]+backlog n=46 ec=2 les=1389 1381/1381/1381) [] r=0 mlcod 0'0 inactive m=44 u=44] read_log mis
sing 316'59 (316'58) b rb.0.1.0000000010be/head by client4304.0:32191 2011-02-21 11:18:13.436974
Feb 28 20:56:39 atom osd.4[24320]: 7f434613b720 filestore(/var/lib/ceph/osd.4) getattr /var/lib/ce
ph/osd.4/current/3.7c9_head/rb.0.1.00000000a638_head '_'
Feb 28 20:56:39 atom osd.4[24320]: 7f434613b720 osd4 1415 pg[3.7c9( v 316'242 lc 0'0 (316'238,316'
242)]+backlog n=46 ec=2 les=1389 1381/1381/1381) [] r=0 mlcod 0'0 inactive m=45 u=45] read_log mis
sing 316'55 (316'54) b rb.0.1.00000000a638/head by client4304.0:21256 2011-02-21 11:00:31.259864
Feb 28 20:56:39 atom osd.4[24320]: 7f434613b720 osd4 1415 pg[3.7c9( v 316'242 lc 0'0 (316'238,316'
242)]+backlog n=46 ec=2 les=1389 1381/1381/1381) [] r=0 mlcod 0'0 inactive m=46 u=46] read_log done
Feb 28 20:56:42 atom osd.4[24320]: 7f433b388700 osd4 1416 pg[3.7c9( v 316'242 lc 0'0 (316'238,316'
242)]+backlog n=46 ec=2 les=1389 1416/1416/1416) [6] r=-1 inactive m=46] cancel_recovery
```

The recovery still seems to get cancelled? Due to the missing rb.* file?

The recovery stuff is still a bit new to me, so I'm not sure what to look at.

The logs are available on logger.ceph.widodh.nl/srv/ceph/remote-syslog when you will find "ceph-osd.log", 2.gz and 3.gz are the ones where I got this data from.

On this cluster I'm running: c78b29a47d7211a4b8b1585112ac22b8435a82c7

Associated revisions

Revision 19bcd747 - 03/01/2011 12:05 AM - Sage Weil

osd: trigger discover_all_missing after replay delay

We were calling discover_all_missing only when we went immediately active, not after we were in the replay state (which triggers from a timer event that calls OSD::activate_pg()). Move the call into PG::activate() so that we catch both callers.

This requires passing in a query_map from the caller. While we're at it, clean up some other instances where we are defining a new query_map deep within the call tree.

Fixes: #847 (I hope)

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

Revision 7cd8ac0b - 03/02/2011 12:49 PM - Sage Weil

osd: trigger discover_all_missing after replay delay

We were calling discover_all_missing only when we went immediately active, not after we were in the replay state (which triggers from a timer event that calls OSD::activate_pg()). Move the call into PG::activate() so that we catch both callers.

This requires passing in a query_map from the caller. While we're at it, clean up some other instances where we are defining a new query_map deep within the call tree.

Fixes: #847 (I hope)

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

History

#1 - 02/28/2011 11:58 AM - Wido den Hollander

osd4 started printing some more lines after it's restart:

```
Feb 28 20:59:41 atom osd.4[24320]: 7f4341495700 osd4 1436 pg[3.7c9( v 316'242 lc 0'0 (316'238,316'242])+backlog
n=46 ec=2 les=
1436 1434/1434/1434) [4,6] r=0 mlcod 0'0 active m=46 u=46] activate - snap_trimq []
Feb 28 20:59:41 atom osd.4[24320]: 7f4341495700 osd4 1436 pg[3.7c9( v 316'242 lc 0'0 (316'238,316'242])+backlog
n=46 ec=2 les=
1436 1434/1434/1434) [4,6] r=0 mlcod 0'0 active m=46 u=46] write_info info 309
Feb 28 20:59:41 atom osd.4[24320]: 7f4341495700 osd4 1436 pg[3.7c9( v 316'242 lc 0'0 (316'238,316'242])+backlog
n=46 ec=2 les=
1436 1434/1434/1434) [4,6] r=0 mlcod 0'0 active m=46 u=46] activate - starting recovery
Feb 28 20:59:41 atom osd.4[24320]: 7f4341495700 osd4 1436 pg[3.7c9( v 316'242 lc 0'0 (316'238,316'242])+backlog
n=46 ec=2 les=
1436 1434/1434/1434) [4,6] r=0 mlcod 0'0 active m=46 u=46] activate peer osd6 sending log(0'0,0'0) missing(0)
Feb 28 20:59:41 atom osd.4[24320]: 7f4341495700 osd4 1436 pg[3.7c9( v 316'242 lc 0'0 (316'238,316'242])+backlog
n=46 ec=2 les=
1436 1434/1434/1434) [4,6] r=0 mlcod 0'0 active m=46 u=46] is_all_uptodate: primary has 46
Feb 28 20:59:41 atom osd.4[24320]: 7f4341495700 osd4 1436 pg[3.7c9( v 316'242 lc 0'0 (316'238,316'242])+backlog
n=46 ec=2 les=
1436 1434/1434/1434) [4,6] r=0 mlcod 0'0 active m=46 u=46] is_all_uptodate: osd6 has 46 missing
...
...
...
1436 1434/1434/1434) [4,6] r=0 mlcod 0'0 active m=46 u=46] recover_replicas: rb.0.1.00000000d1b0/head still un
found
Feb 28 20:59:56 atom osd.4[24320]: 7f4338382700 osd4 1436 pg[3.7c9( v 316'242 lc 0'0 (316'238,316'242])+backlog
n=46 ec=2 les=1436 1434/1434/1434) [4,6] r=0 mlcod 0'0 active m=46 u=46] recover_replicas: rb.0.1.00000000d4a
a/head still unfound
Feb 28 20:59:56 atom osd.4[24320]: 7f4338382700 osd4 1436 pg[3.7c9( v 316'242 lc 0'0 (316'238,316'242])+backlog
n=46 ec=2 les=1436 1434/1434/1434) [4,6] r=0 mlcod 0'0 active m=46 u=46] recover_replicas: rb.0.1.00000000f6d
2/head still unfound
```

```
Feb 28 20:59:56 atom osd.4[24320]: 7f4338382700 osd4 1436 pg[3.7c9( v 316'242 lc 0'0 (316'238,316'242]+backlog
n=46 ec=2 les=1436 1434/1434/1434) [4,6] r=0 mlcod 0'0 active m=46 u=46] recover_replicas: rb.0.1.00000001029
3/head still unfound
Feb 28 20:59:56 atom osd.4[24320]: 7f4338382700 osd4 1436 pg[3.7c9( v 316'242 lc 0'0 (316'238,316'242]+backlog
n=46 ec=2 les=1436 1434/1434/1434) [4,6] r=0 mlcod 0'0 active m=46 u=46] is_all_uptodate: osd6 has 46 missing
Feb 28 20:59:56 atom osd.4[24320]: 7f4338382700 osd4 1436 pg[3.7c9( v 316'242 lc 0'0 (316'238,316'242]+backlog
n=46 ec=2 les=1436 1434/1434/1434) [4,6] r=0 mlcod 0'0 active m=46 u=46] start_recovery_ops: some OSDs are no
t up-to-date yet. Recovery isn't finished yet.
```

#2 - 02/28/2011 04:06 PM - Sage Weil

I believe this will fix it: [19bcd747b452b79c80744620e635f036bfc1b738](https://bugzilla.redhat.com/show_bug.cgi?id=19bcd747b452b79c80744620e635f036bfc1b738)

#3 - 02/28/2011 04:35 PM - Sage Weil

- Status changed from New to Testing

#4 - 03/01/2011 05:06 AM - Wido den Hollander

I've tried your patch, but unfortunately it did not work out.

Did a little digging and out:

```
Mar 1 13:53:33 atom osd.4[29656]: 7f9e891f4700 osd4 1528 pg[3.7c9( v 316'242 lc 0'0 (316'238,316'242]+backlog
n=46 ec=2 les=1513 1504/1504/1504) [4,6] r=0 mlcod 0'0 active m=46] search_for_missing missing {rb.0.1.000000
000446/head=316'242,rb.0.1.0000000010be/head=316'59,rb.0.1.00000000117f/head=316'63,rb.0.1.000000001397/head=3
16'67,rb.0.1.0000000020c9/head=316'75,rb.0.1.000000002dd0/head=316'78,rb.0.1.000000002f5a/head=316'82,rb.0.1.0
00000003122/head=316'86,rb.0.1.00000000395a/head=316'90,rb.0.1.000000003fb2/head=316'95,rb.0.1.000000004188/he
ad=316'99,rb.0.1.00000000482a/head=316'103,rb.0.1.000000005cd5/head=316'107,rb.0.1.0000000061a9/head=316'112,r
b.0.1.000000006408/head=316'116,rb.0.1.00000000648c/head=316'120,rb.0.1.000000006ff5/head=316'125,rb.0.1.00000
0007386/head=316'129,rb.0.1.0000000074f8/head=316'133,rb.0.1.00000000787d/head=316'137,rb.0.1.000000007bda/hea
d=316'141,rb.0.1.000000008411/head=316'145,rb.0.1.0000000089b7/head=316'149,rb.0.1.00000000a27f/head=316'153,r
b.0.1.00000000a638/head=316'155,rb.0.1.00000000a96c/head=316'155,rb.0.1.00000000adcc/head=316'159,rb.0.1.000000
00aee6/head=316'165,rb.0.1.00000000b12e/head=316'169,rb.0.1.00000000b36b/head=316'173,rb.0.1.00000000b6e8/head
=316'177,rb.0.1.00000000bc0d/head=316'181,rb.0.1.00000000c1c9/head=316'186,rb.0.1.00000000c3c2/head=316'190,rb
.0.1.00000000c4c5/head=316'194,rb.0.1.00000000c929/head=316'199,rb.0.1.00000000d1b0/head=316'203,rb.0.1.000000
00d4aa/head=316'207,rb.0.1.00000000e17d/head=316'211,rb.0.1.00000000e2ce/head=316'215,rb.0.1.00000000e353/head
=316'219,rb.0.1.00000000f6c0/head=316'223,rb.0.1.00000000f6d2/head=316'227,rb.0.1.000000010293/head=316'231,rb
.0.1.00000001046a/head=316'235,rb.0.1.000000010669/head=316'239}
Mar 1 13:53:33 atom osd.4[29656]: 7f9e891f4700 osd4 1528 pg[3.7c9( v 316'242 lc 0'0 (316'238,316'242]+backlog
n=46 ec=2 les=1513 1504/1504/1504) [4,6] r=0 mlcod 0'0 active m=46] ignoring osd1 log, pg is already active
Mar 1 13:53:33 atom osd.4[29656]: 7f9e899f5700 osd4 1529 pg[3.7c9( v 316'242 lc 0'0 (316'238,316'242]+backlog
n=46 ec=2 les=1513 1504/1504/1504) [4,6] r=0 mlcod 0'0 active m=46] unaffected with [4,6]/[4,6] up/acting
Mar 1 13:53:37 atom osd.4[29656]: 7f9e891f4700 osd4 1529 pg[3.7c9( v 316'242 lc 0'0 (316'238,316'242]+backlog
n=46 ec=2 les=1513 1504/1504/1504) [4,6] r=0 mlcod 0'0 active m=46] got dup osd2 info 3.7c9( v 316'242 (316'2
38,316'242) n=46 ec=2 les=514 1504/1504/1504), identical to ours
...
...
...
Mar 1 13:53:42 atom osd.4[29656]: 7f9e891f4700 osd4 1530 pg[3.7c9( v 316'242 lc 0'0 (316'238,316'242]+backlog
n=46 ec=2 les=1513 1504/1504/1504) [4,6] r=0 mlcod 0'0 active m=46] update_stats 1504'499
Mar 1 13:53:43 atom osd.4[29656]: 7f9e891f4700 osd4 1530 pg[3.7c9( v 316'242 lc 0'0 (316'238,316'242]+backlog
n=46 ec=2 les=1513 1504/1504/1504) [4,6] r=0 mlcod 0'0 active m=46] got dup osd2 info 3.7c9( v 316'242 (316'2
38,316'242) n=46
ec=2 les=514 1504/1504/1504), identical to ours
Mar 1 13:55:25 atom osd.4[29656]: 7f9e8fb02700 osd4 1530 on 2011-02-22 07:29:12.919625 3.7c9
```

osd4 keeps saying that osd6 is missing objects, that's what I see repeating, over and over.

#5 - 03/01/2011 01:41 PM - Sage Weil

Oh, I see the problem:

2011-03-01 22:39:59.563426 log 2011-03-01 22:27:18.519821 osd1 [2a00:f10:113:1:230:48ff:fe8d:a21e]:6804/4281 34 : [ERR] osd4 tried to pull rb.0.1.000000001277/head in 3.c9 but got error 2: No such file or directory

and osd recovery max active = 1. FBOFW the recovery slot isn't freed up when these errors are detected.

#6 - 03/01/2011 04:11 PM - Sage Weil

Pushed an osd_recovery branch that should continue recovering, skipping over the missing objects. If there are other available replicas, it will try to use those to recover.

#7 - 03/01/2011 04:37 PM - Sage Weil

- Target version set to v0.25

#8 - 03/03/2011 03:14 AM - Wido den Hollander

I tried the new branch, but it isn't working as expected yet.

When following the progress I saw:

```

2011-03-03 11:47:27.448864 pg v133244: 8240 pgs: 48 active, 8192 active+clean; 260 GB data, 523 GB used, 14
364 GB / 14904 GB avail; 3106/134690 degraded (2.306%)
2011-03-03 11:47:27.956132 log 2011-03-03 11:47:04.404287 osd1 [2a00:f10:113:1:230:48ff:fe8d:a21e]:6804/2928
6 204 : [ERR] 3.7c9 osd4 tried to pull rb.0.1.0000000020c9/head but got error 2: No such file or directory
2011-03-03 11:47:27.956132 log 2011-03-03 11:47:04.407369 osd1 [2a00:f10:113:1:230:48ff:fe8d:a21e]:6804/2928
6 205 : [ERR] 3.7c9 osd4 tried to pull rb.0.1.000000002dd0/head but got error 2: No such file or directory
2011-03-03 11:47:27.956132 log 2011-03-03 11:47:04.414558 osd1 [2a00:f10:113:1:230:48ff:fe8d:a21e]:6804/2928
6 206 : [ERR] 3.7c9 osd4 tried to pull rb.0.1.000000002f5a/head but got error 2: No such file or directory
2011-03-03 11:47:27.956132 log 2011-03-03 11:47:04.420450 osd1 [2a00:f10:113:1:230:48ff:fe8d:a21e]:6804/2928
6 207 : [ERR] 3.7c9 osd4 tried to pull rb.0.1.000000003122/head but got error 2: No such file or directory
2011-03-03 11:47:27.956132 log 2011-03-03 11:47:04.429385 osd1 [2a00:f10:113:1:230:48ff:fe8d:a21e]:6804/2928
6 208 : [ERR] 3.7c9 osd4 tried to pull rb.0.1.00000000395a/head but got error 2: No such file or directory
2011-03-03 11:47:27.956132 log 2011-03-03 11:47:04.436475 osd1 [2a00:f10:113:1:230:48ff:fe8d:a21e]:6804/2928
6 209 : [ERR] 3.7c9 osd4 tried to pull rb.0.1.000000003fb2/head but got error 2: No such file or directory
2011-03-03 11:47:27.956132 log 2011-03-03 11:47:04.448904 osd1 [2a00:f10:113:1:230:48ff:fe8d:a21e]:6804/2928
6 210 : [ERR] 3.7c9 osd4 tried to pull rb.0.1.000000004188/head but got error 2: No such file or directory
2011-03-03 11:47:27.956132 log 2011-03-03 11:47:04.455656 osd1 [2a00:f10:113:1:230:48ff:fe8d:a21e]:6804/2928
6 211 : [ERR] 3.7c9 osd4 tried to pull rb.0.1.00000000482a/head but got error 2: No such file or directory
2011-03-03 11:47:27.956132 log 2011-03-03 11:47:04.468122 osd1 [2a00:f10:113:1:230:48ff:fe8d:a21e]:6804/2928
6 212 : [ERR] 3.7c9 osd4 tried to pull rb.0.1.000000005cd5/head but got error 2: No such file or directory
2011-03-03 11:47:27.956132 log 2011-03-03 11:47:04.472866 osd1 [2a00:f10:113:1:230:48ff:fe8d:a21e]:6804/2928
6 213 : [ERR] 3.7c9 osd4 tried to pull rb.0.1.0000000061a9/head but got error 2: No such file or directory
2011-03-03 11:47:27.956132 log 2011-03-03 11:47:04.484156 osd1 [2a00:f10:113:1:230:48ff:fe8d:a21e]:6804/2928
6 214 : [ERR] 3.7c9 osd4 tried to pull rb.0.1.000000006408/head but got error 2: No such file or directory
2011-03-03 11:47:27.956132 log 2011-03-03 11:47:04.497592 osd1 [2a00:f10:113:1:230:48ff:fe8d:a21e]:6804/2928
6 215 : [ERR] 3.7c9 osd4 tried to pull rb.0.1.00000000648c/head but got error 2: No such file or directory
2011-03-03 11:47:27.956132 log 2011-03-03 11:47:04.506712 osd1 [2a00:f10:113:1:230:48ff:fe8d:a21e]:6804/2928
6 216 : [ERR] 3.7c9 osd4 tried to pull rb.0.1.000000006ff5/head but got error 2: No such file or directory
2011-03-03 11:47:27.956132 log 2011-03-03 11:47:04.526531 osd1 [2a00:f10:113:1:230:48ff:fe8d:a21e]:6804/2928
6 217 : [ERR] 3.7c9 osd4 tried to pull rb.0.1.000000007386/head but got error 2: No such file or directory
2011-03-03 11:47:27.956132 log 2011-03-03 11:47:04.535226 osd1 [2a00:f10:113:1:230:48ff:fe8d:a21e]:6804/2928
6 218 : [ERR] 3.7c9 osd4 tried to pull rb.0.1.0000000074f8/head but got error 2: No such file or directory
2011-03-03 11:47:27.956132 log 2011-03-03 11:47:04.560339 osd1 [2a00:f10:113:1:230:48ff:fe8d:a21e]:6804/2928
6 219 : [ERR] 3.7c9 osd4 tried to pull rb.0.1.00000000787d/head but got error 2: No such file or directory
2011-03-03 11:47:27.956132 log 2011-03-03 11:47:04.572708 osd1 [2a00:f10:113:1:230:48ff:fe8d:a21e]:6804/2928
6 220 : [ERR] 3.7c9 osd4 tried to pull rb.0.1.000000007bda/head but got error 2: No such file or directory
2011-03-03 11:47:27.956132 log 2011-03-03 11:47:04.574649 osd1 [2a00:f10:113:1:230:48ff:fe8d:a21e]:6804/2928
6 221 : [ERR] 3.7c9 osd4 tried to pull rb.0.1.000000008411/head but got error 2: No such file or directory
2011-03-03 11:47:27.956132 log 2011-03-03 11:47:04.581441 osd1 [2a00:f10:113:1:230:48ff:fe8d:a21e]:6804/2928
6 222 : [ERR] 3.7c9 osd4 tried to pull rb.0.1.0000000089b7/head but got error 2: No such file or directory
2011-03-03 11:47:27.956132 log 2011-03-03 11:47:04.601985 osd1 [2a00:f10:113:1:230:48ff:fe8d:a21e]:6804/2928
6 223 : [ERR] 3.7c9 osd4 tried to pull rb.0.1.00000000a27f/head but got error 2: No such file or directory
2011-03-03 11:47:27.956132 log 2011-03-03 11:47:04.604099 osd1 [2a00:f10:113:1:230:48ff:fe8d:a21e]:6804/2928
6 224 : [ERR] 3.7c9 osd4 tried to pull rb.0.1.00000000a96c/head but got error 2: No such file or directory

```

2011-03-03 11:47:27.956132 log 2011-03-03 11:47:04.629478 osd1 [2a00:f10:113:1:230:48ff:fe8d:a21e]:6804/2928
6 225 : [ERR] 3.7c9 osd4 tried to pull rb.0.1.00000000adcc/head but got error 2: No such file or directory
2011-03-03 11:47:27.956132 log 2011-03-03 11:47:04.643160 osd1 [2a00:f10:113:1:230:48ff:fe8d:a21e]:6804/2928
6 226 : [ERR] 3.7c9 osd4 tried to pull rb.0.1.00000000aee6/head but got error 2: No such file or directory
2011-03-03 11:47:27.956132 log 2011-03-03 11:47:04.662869 osd1 [2a00:f10:113:1:230:48ff:fe8d:a21e]:6804/2928
6 227 : [ERR] 3.7c9 osd4 tried to pull rb.0.1.00000000b12e/head but got error 2: No such file or directory
2011-03-03 11:47:27.956132 log 2011-03-03 11:47:04.676948 osd1 [2a00:f10:113:1:230:48ff:fe8d:a21e]:6804/2928
6 228 : [ERR] 3.7c9 osd4 tried to pull rb.0.1.00000000b36b/head but got error 2: No such file or directory
2011-03-03 11:47:27.956132 log 2011-03-03 11:47:04.697511 osd1 [2a00:f10:113:1:230:48ff:fe8d:a21e]:6804/2928
6 229 : [ERR] 3.7c9 osd4 tried to pull rb.0.1.00000000b6e8/head but got error 2: No such file or directory
2011-03-03 11:47:27.956132 log 2011-03-03 11:47:04.712442 osd1 [2a00:f10:113:1:230:48ff:fe8d:a21e]:6804/2928
6 230 : [ERR] 3.7c9 osd4 tried to pull rb.0.1.00000000bc0d/head but got error 2: No such file or directory
2011-03-03 11:47:27.956132 log 2011-03-03 11:47:04.730804 osd1 [2a00:f10:113:1:230:48ff:fe8d:a21e]:6804/2928
6 231 : [ERR] 3.7c9 osd4 tried to pull rb.0.1.00000000c1c9/head but got error 2: No such file or directory
2011-03-03 11:47:27.956132 log 2011-03-03 11:47:04.742646 osd1 [2a00:f10:113:1:230:48ff:fe8d:a21e]:6804/2928
6 232 : [ERR] 3.7c9 osd4 tried to pull rb.0.1.00000000c3c2/head but got error 2: No such file or directory
2011-03-03 11:47:27.956132 log 2011-03-03 11:47:04.747028 osd1 [2a00:f10:113:1:230:48ff:fe8d:a21e]:6804/2928
6 233 : [ERR] 3.7c9 osd4 tried to pull rb.0.1.00000000c4c5/head but got error 2: No such file or directory
2011-03-03 11:47:27.956132 log 2011-03-03 11:47:04.752920 osd1 [2a00:f10:113:1:230:48ff:fe8d:a21e]:6804/2928
6 234 : [ERR] 3.7c9 osd4 tried to pull rb.0.1.00000000c929/head but got error 2: No such file or directory
2011-03-03 11:47:27.956132 log 2011-03-03 11:47:04.757412 osd1 [2a00:f10:113:1:230:48ff:fe8d:a21e]:6804/2928
6 235 : [ERR] 3.7c9 osd4 tried to pull rb.0.1.00000000d1b0/head but got error 2: No such file or directory
2011-03-03 11:47:27.956132 log 2011-03-03 11:47:04.763600 osd1 [2a00:f10:113:1:230:48ff:fe8d:a21e]:6804/2928
6 236 : [ERR] 3.7c9 osd4 tried to pull rb.0.1.00000000d4aa/head but got error 2: No such file or directory
2011-03-03 11:47:27.956132 log 2011-03-03 11:47:04.768695 osd1 [2a00:f10:113:1:230:48ff:fe8d:a21e]:6804/2928
6 237 : [ERR] 3.7c9 osd4 tried to pull rb.0.1.00000000e17d/head but got error 2: No such file or directory
2011-03-03 11:47:27.956132 log 2011-03-03 11:47:04.774378 osd1 [2a00:f10:113:1:230:48ff:fe8d:a21e]:6804/2928
6 238 : [ERR] 3.7c9 osd4 tried to pull rb.0.1.00000000e2ce/head but got error 2: No such file or directory
2011-03-03 11:47:27.956132 log 2011-03-03 11:47:04.779276 osd1 [2a00:f10:113:1:230:48ff:fe8d:a21e]:6804/2928
6 239 : [ERR] 3.7c9 osd4 tried to pull rb.0.1.00000000e353/head but got error 2: No such file or directory
2011-03-03 11:47:27.956132 log 2011-03-03 11:47:04.785520 osd1 [2a00:f10:113:1:230:48ff:fe8d:a21e]:6804/2928
6 240 : [ERR] 3.7c9 osd4 tried to pull rb.0.1.00000000f6c0/head but got error 2: No such file or directory
2011-03-03 11:47:27.956132 log 2011-03-03 11:47:04.786110 osd1 [2a00:f10:113:1:230:48ff:fe8d:a21e]:6804/2928
6 241 : [ERR] 3.7c9 osd4 tried to pull rb.0.1.00000000f6d2/head but got error 2: No such file or directory
2011-03-03 11:47:27.956132 log 2011-03-03 11:47:04.793151 osd1 [2a00:f10:113:1:230:48ff:fe8d:a21e]:6804/2928
6 242 : [ERR] 3.7c9 osd4 tried to pull rb.0.1.000000010293/head but got error 2: No such file or directory
2011-03-03 11:47:27.956132 log 2011-03-03 11:47:04.793747 osd1 [2a00:f10:113:1:230:48ff:fe8d:a21e]:6804/2928
6 243 : [ERR] 3.7c9 osd4 tried to pull rb.0.1.00000001046a/head but got error 2: No such file or directory
2011-03-03 11:47:32.452360 pg v133245: 8240 pgs: 48 active, 8192 active+clean; 260 GB data, 523 GB used, 14
364 GB / 14904 GB avail; 3102/134690 degraded (2.303%)
2011-03-03 11:48:46.140553 pg v133246: 8240 pgs: 48 active, 8192 active+clean; 260 GB data, 523 GB used, 14
364 GB / 14904 GB avail; 3102/134690 degraded (2.303%)
2011-03-03 11:48:46.636801 pg v133247: 8240 pgs: 48 active, 8192 active+clean; 260 GB data, 523 GB used, 14
364 GB / 14904 GB avail; 3102/134690 degraded (2.303%)
2011-03-03 11:48:46.925668 pg v133248: 8240 pgs: 48 active, 8192 active+clean; 260 GB data, 523 GB used, 14
364 GB / 14904 GB avail; 3102/134690 degraded (2.303%)
2011-03-03 11:48:49.292291 pg v133249: 8240 pgs: 48 active, 8192 active+clean; 260 GB data, 523 GB used, 14
364 GB / 14904 GB avail; 3102/134690 degraded (2.303%)
2011-03-03 11:48:51.341869 pg v133250: 8240 pgs: 48 active, 8192 active+clean; 260 GB data, 523 GB used, 14
364 GB / 14904 GB avail; 3102/134690 degraded (2.303%)
2011-03-03 11:49:32.560627 pg v133251: 8240 pgs: 48 active, 8192 active+clean; 260 GB data, 523 GB used, 14
364 GB / 14904 GB avail; 3102/134690 degraded (2.303%)

It did went a bit further, but than stopped again.

In the logs I saw:

Mar 3 11:55:02 atom osd.4[7377]: 7f19de10a700 osd4 1597 do_recovery starting 2 (0/2 rops) on pg[3.7c9 (v 316'
242 lc 0'0 (316'238,316'242)+backlog n=46 ec=2 les=1597 1594/1594/1594) [4,6] r=0 mlcod 0'0 active m=44 u=44]
Mar 3 11:55:02 atom osd.4[7377]: 7f19de10a700 osd4 1597 pg[3.7c9 (v 316'242 lc 0'0 (316'238,316'242)+backlog
n=46 ec=2 les=1597 1594/1594/1594) [4,6] r=0 mlcod 0'0 active m=44 u=44] peer osd6 missing {rb.0.1.0000000010
be/head=316'59,rb.0.1.00000000117f/head=316'63,rb.0.1.000000001397/head=316'67,rb.0.1.0000000020c9/head=316'75
,rb.0.1.000000002dd0/head=316'78,rb.0.1.000000002f5a/head=316'82,rb.0.1.000000003122/head=316'86,rb.0.1.000000
00395a/head=316'90,rb.0.1.000000003fb2/head=316'95,rb.0.1.000000004188/head=316'99,rb.0.1.00000000482a/head=31
6'103,rb.0.1.000000005cd5/head=316'107,rb.0.1.0000000061a9/head=316'112,rb.0.1.000000006408/head=316'116,rb.0.
1.00000000648c/head=316'120,rb.0.1.000000006ff5/head=316'125,rb.0.1.000000007386/head=316'129,rb.0.1.000000007
4f8/head=316'133,rb.0.1.00000000787d/head=316'137,rb.0.1.000000007bda/head=316'141,rb.0.1.000000008411/head=31
6'145,rb.0.1.0000000089b7/head=316'149,rb.0.1.00000000a27f/head=316'153,rb.0.1.00000000a638/head=316'55,rb.0.1.
.00000000a96c/head=316'155,rb.0.1.00000000adcc/head=316'159,rb.0.1.00000000aee6/head=316'165,rb.0.1.00000000b1
2e/head=316'169,rb.0.1.00000000b36b/head=316'173,rb.0.1.00000000b6e8/head=316'177,rb.0.1.00000000bc0d/head=316
'181,rb.0.1.00000000c1c9/head=316'186,rb.0.1.00000000c3c2/head=316'190,rb.0.1.00000000c4c5/head=316'194,rb.0.1.
.00000000c929/head=316'199,rb.0.1.00000000d1b0/head=316'203,rb.0.1.00000000d4aa/head=316'207,rb.0.1.00000000e1
7d/head=316'211,rb.0.1.00000000e2ce/head=316'215,rb.0.1.00000000e353/head=316'219,rb.0.1.00000000f6c0/head=316

```
'223,rb.0.1.00000000f6d2/head=316'227,rb.0.1.000000010293/head=316'231,rb.0.1.00000001046a/head=316'235}
Mar 3 11:55:02 atom osd.4[7377]: 7f19de10a700 osd4 1597 pg[3.7c9( v 316'242 lc 0'0 (316'238,316'242)+backlog
n=46 ec=2 les=1597 1594/1594/1594) [4,6] r=0 mlcod 0'0 active m=44 u=44] recover_replicas: rb.0.1.000000001397
/head still unfound
Mar 3 11:55:02 atom osd.4[7377]: 7f19de10a700 osd4 1597 pg[3.7c9( v 316'242 lc 0'0 (316'238,316'242)+backlog
n=46 ec=2 les=1597 1594/1594/1594) [4,6] r=0 mlcod 0'0 active m=44 u=44] recover_replicas: rb.0.1.0000000020c9
/head still unfound
Mar 3 11:55:02 atom osd.4[7377]: 7f19de10a700 osd4 1597 pg[3.7c9( v 316'242 lc 0'0 (316'238,316'242)+backlog
n=46 ec=2 les=1597 1594/1594/1594) [4,6] r=0 mlcod 0'0 active m=44 u=44] recover_replicas: rb.0.1.000000004188
/head still unfound
Mar 3 11:55:02 atom osd.4[7377]: 7f19de10a700 osd4 1597 pg[3.7c9( v 316'242 lc 0'0 (316'238,316'242)+backlog
n=46 ec=2 les=1597 1594/1594/1594) [4,6] r=0 mlcod 0'0 active m=44 u=44] recover_replicas: rb.0.1.000000006ff5
/head still unfound
Mar 3 11:55:02 atom osd.4[7377]: 7f19de10a700 osd4 1597 pg[3.7c9( v 316'242 lc 0'0 (316'238,316'242)+backlog
n=46 ec=2 les=1597 1594/1594/1594) [4,6] r=0 mlcod 0'0 active m=44 u=44] recover_replicas: rb.0.1.000000007386
/head still unfound
Mar 3 11:55:02 atom osd.4[7377]: 7f19de10a700 osd4 1597 pg[3.7c9( v 316'242 lc 0'0 (316'238,316'242)+backlog
n=46 ec=2 les=1597 1594/1594/1594) [4,6] r=0 mlcod 0'0 active m=44 u=44] recover_replicas: rb.0.1.0000000089b7
/head still unfound
Mar 3 11:55:02 atom osd.4[7377]: 7f19de10a700 osd4 1597 pg[3.7c9( v 316'242 lc 0'0 (316'238,316'242)+backlog
n=46 ec=2 les=1597 1594/1594/1594) [4,6] r=0 mlcod 0'0 active m=44 u=44] recover_replicas: rb.0.1.00000000a27f
/head still unfound
Mar 3 11:55:02 atom osd.4[7377]: 7f19de10a700 osd4 1597 pg[3.7c9( v 316'242 lc 0'0 (316'238,316'242)+backlog
n=46 ec=2 les=1597 1594/1594/1594) [4,6] r=0 mlcod 0'0 active m=44 u=44] recover_replicas: rb.0.1.00000000b36b
/head still unfound
Mar 3 11:55:02 atom osd.4[7377]: 7f19de10a700 osd4 1597 pg[3.7c9( v 316'242 lc 0'0 (316'238,316'242)+backlog
n=46 ec=2 les=1597 1594/1594/1594) [4,6] r=0 mlcod 0'0 active m=44 u=44] recover_replicas: rb.0.1.00000000b6e8
/head still unfound
Mar 3 11:55:02 atom osd.4[7377]: 7f19de10a700 osd4 1597 pg[3.7c9( v 316'242 lc 0'0 (316'238,316'242)+backlog
n=46 ec=2 les=1597 1594/1594/1594) [4,6] r=0 mlcod 0'0 active m=44 u=44] recover_replicas: rb.0.1.00000000c4c5
/head still unfound
Mar 3 11:55:02 atom osd.4[7377]: 7f19de10a700 osd4 1597 pg[3.7c9( v 316'242 lc 0'0 (316'238,316'242)+backlog
n=46 ec=2 les=1597 1594/1594/1594) [4,6] r=0 mlcod 0'0 active m=44 u=44] recover_replicas: rb.0.1.00000000c929
/head still unfound
Mar 3 11:55:02 atom osd.4[7377]: 7f19de10a700 osd4 1597 pg[3.7c9( v 316'242 lc 0'0 (316'238,316'242)+backlog
n=46 ec=2 les=1597 1594/1594/1594) [4,6] r=0 mlcod 0'0 active m=44 u=44] recover_replicas: rb.0.1.00000000e353
/head still unfound
Mar 3 11:55:02 atom osd.4[7377]: 7f19de10a700 osd4 1597 pg[3.7c9( v 316'242 lc 0'0 (316'238,316'242)+backlog
n=46 ec=2 les=1597 1594/1594/1594) [4,6] r=0 mlcod 0'0 active m=44 u=44] recover_replicas: rb.0.1.00000000f6c0
/head still unfound
Mar 3 11:55:02 atom osd.4[7377]: 7f19de10a700 osd4 1597 pg[3.7c9( v 316'242 lc 0'0 (316'238,316'242)+backlog
n=46 ec=2 les=1597 1594/1594/1594) [4,6] r=0 mlcod 0'0 active m=44 u=44] start_recovery_ops: some OSDs are not
up-to-date yet. Recovery isn't finished yet.
Mar 3 11:55:02 atom osd.4[7377]: 7f19de10a700 osd4 1597 do_recovery started 0 (0/2 rops) on pg[3.7c9( v 316'2
42 lc 0'0 (316'238,316'242)+backlog n=46 ec=2 les=1597 1594/1594/1594) [4,6] r=0 mlcod 0'0 active m=44 u=44]
```

I've set 'osd recovery max active = 2', but that wouldn't help either.

#9 - 03/03/2011 08:35 AM - Sage Weil

Can you look manually on the OSDs for the objects that it's showing as 'still unfound'? rb.0.1.00000000f6c0/head becomes rb.0.1.00000000f6c0_head. I think some objects got lost somewhere. Also, can you change your syslog so it retains logs for longer? When I was looking at this the other day they didn't go far back enough to find out when/how some of the objects got lost.

Also, is the picture similar for the other non-clean PGs? (ceph pg dump -o - | grep -v clean, then grep for that pgid in the logs and see if there are similar no such file or directory errors).

#10 - 03/03/2011 10:51 AM - Wido den Hollander

I searched for the object:

```
root@noisy:~# find /var/lib/ceph -name 'rb.0.1.00000000f6c0_head'
/var/lib/ceph/osd.2/snap_623655/3.7c9_head/rb.0.1.00000000f6c0_head
/var/lib/ceph/osd.2/current/3.7c9_head/rb.0.1.00000000f6c0_head
/var/lib/ceph/osd.2/snap_630809/3.7c9_head/rb.0.1.00000000f6c0_head
/var/lib/ceph/osd.2/snap_630833/3.7c9_head/rb.0.1.00000000f6c0_head
root@noisy:~#
```

In all the logs I saw osd1, osd4 and osd6 involved in this object. It seems to be on osd2, but it isn't searching there?

I checked other PG's, like 3.6a1, but those are also missing objects:

```
Mar 3 11:54:46 atom osd.4[7377]: 7f19e0a10700 osd4 1597 pg[3.6a1( v 316'153 lc 0'0 (316'151,316'153)+backlog
n=37 ec=2 les=1597 1594/1594/1594) [4,5] r=0 mlcod 0'0 active m=37 u=37] search_for_missing missing {rb.0.1.00
0000000499/head=172'5,rb.0.1.0000000008ee/head=242'9,rb.0.1.000000001060/head=316'15,rb.0.1.0000000010f2/head=
316'19,rb.0.1.000000001bd8/head=316'23,rb.0.1.000000002b4f/head=316'27,rb.0.1.000000002e76/head=316'31,rb.0.1.
00000000354e/head=316'35,rb.0.1.000000004285/head=316'40,rb.0.1.00000000493d/head=316'44,rb.0.1.000000004b1f/h
ead=316'48,rb.0.1.000000004c72/head=316'52,rb.0.1.000000004e67/head=316'56,rb.0.1.000000005391/head=316'60,rb.
0.1.000000005b97/head=316'65,rb.0.1.000000005c3b/head=316'13,rb.0.1.000000005d72/head=316'69,rb.0.1.0000000060
7a/head=316'73,rb.0.1.00000000635c/head=316'78,rb.0.1.000000008138/head=316'82,rb.0.1.0000000088f7/head=316'86
,rb.0.1.0000000098bf/head=316'91,rb.0.1.000000009c79/head=316'95,rb.0.1.000000009ff4/head=316'99,rb.0.1.000000
00b1c1/head=316'104,rb.0.1.00000000b3c0/head=316'109,rb.0.1.00000000bcb4/head=316'113,rb.0.1.00000000c6d0/head
=316'117,rb.0.1.00000000d3ea/head=316'121,rb.0.1.00000000df4b/head=316'125,rb.0.1.00000000ea84/head=316'129,rb
.0.1.00000000ebca/head=316'133,rb.0.1.00000000f20b/head=316'137,rb.0.1.00000000f575/head=316'141,rb.0.1.000000
010020/head=316'145,rb.0.1.0000000105da/head=316'149,rb.0.1.000000010b30/head=316'153}
```

Searching for that objects shows:

```
root@noisy:~# find /var/lib/ceph -name 'rb.0.1.000000000499_head'
/var/lib/ceph/osd.1/snap_754352/3.6a1_head/rb.0.1.000000000499_head
/var/lib/ceph/osd.1/current/3.6a1_head/rb.0.1.000000000499_head
/var/lib/ceph/osd.1/snap_762081/3.6a1_head/rb.0.1.000000000499_head
/var/lib/ceph/osd.1/snap_762090/3.6a1_head/rb.0.1.000000000499_head
/var/lib/ceph/osd.2/snap_623655/3.6a1_head/rb.0.1.000000000499_head
/var/lib/ceph/osd.2/current/3.6a1_head/rb.0.1.000000000499_head
/var/lib/ceph/osd.2/snap_630809/3.6a1_head/rb.0.1.000000000499_head
/var/lib/ceph/osd.2/snap_630833/3.6a1_head/rb.0.1.000000000499_head
root@noisy:~#
```

3.6a1	37	37	37	74	151552	155189248	3724	3724	active	316'153	1599'528
[4,5]	[4,5]	316'153	2011-02-22	07:09:20.128641							

For another PG I found:

```
Mar 3 11:54:46 atom osd.4[7377]: 7f19e0a10700 osd4 1597 pg[3.5c9( v 316'123 lc 0'0 (316'117,316'123)+backlog
n=30 ec=2 les=1597 1594/1594/1594) [4,6] r=0 mlcod 0'0 active m=28 u=7] search_for_missing rb.0.1.00000000cf1d
/head 316'94 is on osd1
```


But searching for the object shows:

```
root@noisy:~# find /var/lib/ceph -name 'rb.0.1.00000000cf1d_head'
/var/lib/ceph/osd.2/snap_623655/3.5c9_head/rb.0.1.00000000cf1d_head
/var/lib/ceph/osd.2/current/3.5c9_head/rb.0.1.00000000cf1d_head
/var/lib/ceph/osd.2/snap_630809/3.5c9_head/rb.0.1.00000000cf1d_head
/var/lib/ceph/osd.2/snap_630833/3.5c9_head/rb.0.1.00000000cf1d_head
root@noisy:~#
```

It isn't on osd1, but on osd2.

I'll try to fix the syslog, my logrotate configuration says 14 days, no idea why it rotates that fast.

#11 - 03/03/2011 11:35 AM - Wido den Hollander

Just checked another PG (just to make sure), 3.22b in this case:

```
Mar  3 11:54:45 atom osd.4[7377]: 7f19e0a10700 osd4 1597 pg[3.22b( v 316'138 lc 0'0 (316'135,316'138)+backlog
n=33 ec=2 les=1592 1594/1594/1594) [4,1] r=0 mlcod 0'0 peering m=33 u=33] search_for_missing rb.0.1.00000000db
84/head 316'117 also missing on osd1
```

Searching for this object shows:

```
root@noisy:~# find /var/lib/ceph -name 'rb.0.1.00000000db84_head'
/var/lib/ceph/osd.2/snap_623655/3.22b_head/rb.0.1.00000000db84_head
/var/lib/ceph/osd.2/current/3.22b_head/rb.0.1.00000000db84_head
/var/lib/ceph/osd.2/snap_630897/3.22b_head/rb.0.1.00000000db84_head
/var/lib/ceph/osd.2/snap_630898/3.22b_head/rb.0.1.00000000db84_head
/var/lib/ceph/osd.0/snap_784865/3.22b_head/rb.0.1.00000000db84_head
/var/lib/ceph/osd.0/current/3.22b_head/rb.0.1.00000000db84_head
/var/lib/ceph/osd.0/snap_796258/3.22b_head/rb.0.1.00000000db84_head
/var/lib/ceph/osd.0/snap_796259/3.22b_head/rb.0.1.00000000db84_head
root@noisy:~#
```

#12 - 03/03/2011 01:58 PM - Sage Weil

Starting looking through your logs to debug this but it really looks like they are incomplete.. either syslog is dropping something or the new logging code is. Can we go back to logging to a file (over NFS?) for these nodes? At least temporarily?

Thanks!

#13 - 03/04/2011 02:17 AM - Wido den Hollander

I think it was syslog. The logger machine is not on the same location, could be that there is a bandwidth issue which caused the UDP packages to be dropped.

I setup a NFS server at the same location where the cluster is running, right now the machines are logging over NFS, all OSD's are at 'debug osd = 20'.

#14 - 03/04/2011 04:47 AM - Wido den Hollander

I've searched the logs again for rb.0.1.00000000a638/head and I found:

```
2011-03-04 10:34:31.649578 7fee49878700 osd4 1712 pg[3.7c9( v 316'242 lc 0'0 (316'238,316'242)+backlog n=46 ec
=2 les=1712 1702/1702/1699) [4,6] r=0 mlcod 0'0 active m=44 u=44] recover_replicas: rb.0.1.00000000a638/head s
till unfound
2011-03-04 10:39:55.904877 16'55 (316'54) b rb.0.1.00000000a638/head by client4304.0:21256 2011-02-21 11:00:31
.259864 indexed
2011-03-04 10:39:55.907236 7fee4c17e700 osd4 1728 pg[3.7c9( v 316'242 lc 0'0 (316'238,316'242)+backlog n=46 ec
=2 les=1712 1702/1702/1699) [4,6] r=0 mlcod 0'0 active m=44 u=21] search_for_missing rb.0.1.00000000a638/head
316'55 is on osd1
```

It thinks this object is on osd1, but a find tells me:

```
root@noisy:~# find /var/lib/ceph -type f -name 'rb.0.1.00000000a638_head'
/var/lib/ceph/osd.2/snap_634414/3.7c9_head/rb.0.1.00000000a638_head
/var/lib/ceph/osd.2/current/3.7c9_head/rb.0.1.00000000a638_head
/var/lib/ceph/osd.2/snap_640162/3.7c9_head/rb.0.1.00000000a638_head
/var/lib/ceph/osd.2/snap_640165/3.7c9_head/rb.0.1.00000000a638_head
root@noisy:~#
```

Looking on osd1 in 3.7c9 I found:

```
root@noisy:~# ls /var/lib/ceph/osd.1/current/3.7c9_head/
rb.0.1.000000000446_head rb.0.1.000000010669_head
root@noisy:~#
```

Looking on osd2:

```
root@noisy:~# ls /var/lib/ceph/osd.2/current/3.7c9_head/
rb.0.1.000000000446_head rb.0.1.000000005cd5_head rb.0.1.00000000a638_head rb.0.1.00000000d1b0_head
rb.0.1.0000000010be_head rb.0.1.0000000061a9_head rb.0.1.00000000a96c_head rb.0.1.00000000d4aa_head
rb.0.1.00000000117f_head rb.0.1.000000006408_head rb.0.1.00000000adcc_head rb.0.1.00000000e17d_head
rb.0.1.000000001397_head rb.0.1.00000000648c_head rb.0.1.00000000aee6_head rb.0.1.00000000e2ce_head
rb.0.1.0000000020c9_head rb.0.1.000000006ff5_head rb.0.1.00000000b12e_head rb.0.1.00000000e353_head
rb.0.1.000000002dd0_head rb.0.1.000000007386_head rb.0.1.00000000b36b_head rb.0.1.00000000f6c0_head
rb.0.1.000000002f5a_head rb.0.1.0000000074f8_head rb.0.1.00000000b6e8_head rb.0.1.00000000f6d2_head
rb.0.1.000000003122_head rb.0.1.00000000787d_head rb.0.1.00000000bc0d_head rb.0.1.000000010293_head
rb.0.1.00000000395a_head rb.0.1.000000007bda_head rb.0.1.00000000c1c9_head rb.0.1.00000001046a_head
rb.0.1.000000003fb2_head rb.0.1.000000008411_head rb.0.1.00000000c3c2_head rb.0.1.000000010669_head
rb.0.1.000000004188_head rb.0.1.0000000089b7_head rb.0.1.00000000c4c5_head
rb.0.1.00000000482a_head rb.0.1.00000000a27f_head rb.0.1.00000000c929_head
root@noisy:~#
```

If I understand it correctly, the PG (3.7c9) is on osd1, osd4 and osd6, but it isn't, since most of the data of this PG is on osd2 where osd4 and osd6 are still missing it:

```
root@atom:/var/log/ceph# ls /var/lib/ceph/osd.4/current/3.7c9_head/
rb.0.1.000000000446_head rb.0.1.000000010669_head
root@atom:/var/log/ceph# ls /var/lib/ceph/osd.6/current/3.7c9_head/
rb.0.1.000000000446_head rb.0.1.000000010669_head
```

```
root@atom:/var/log/ceph#
```

I'm just trying to understand what is going on here. The recovery code isn't looking for the objects on osd2, since it thinks they are not there?

But in the logs I see:

```
2011-03-04 11:01:43.784979 7fee4c17e700 osd4 1731 pg[3.7c9( v 316'242 lc 0'0 (316'238,316'242]+backlog n=46 ec
=2 les=1712 1702/1702/1699) [4,6] r=0 mlcod 0'0 active m=44] got dup osd2 info 3.7c9( v 316'242 (316'238,316'2
42] n=46 ec=2 les=514 1702/1702/1699), identical to ours
```

#15 - 03/04/2011 02:23 PM - Sage Weil

- Status changed from Testing to Resolved

phew! all pgs are now active+clean. hopefully scrub won't turn anything up :)