Ceph - Bug #10059

osd/ECBackend.cc: 876: FAILED assert(0)

11/10/2014 03:20 PM - Samuel Just

Status:	Resolved	**Start date:**	11/10/2014
Priority:	Urgent	**Due date:**	
Assignee:	Samuel Just	**% Done:**	0%
Category:		**Estimated time:**	0.00 hour
Target version:		**Spent time:**	0.00 hour
Source:	other	**Reviewed:**	
Tags:		**Affected Versions:**	
Backport:	giant,firefly	**ceph-qa-suite:**	
Regression:	No	**Pull request ID:**	
Severity:	3 - minor		

Description

```
-1> 2014-11-09 14:13:01.334410 7f8b93c8b700 10 filestore(/var/lib/ceph/osd/ceph-3)
FileStore::read(1.1ds0_head/78348bbd/benchmark_data_plana76_6098_object6568/head//1/18446744073709551615/0) open error:
(2) No such file or directory
0> 2014-11-09 14:13:01.337431 7f8b93c8b700 -1 osd/ECBackend.cc: In function 'void ECBackend::handle_sub_read(pg_shard_t,
ECSubRead&, ECSubReadReply*)' thread 7f8b93c8b700 time 2014-11-09 14:13:01.334418
osd/ECBackend.cc: 876: FAILED assert(0)

ceph version 0.87-573-g6977d02 (6977d02f0d31c453cdf554a8f1796f290c1a3b89)
1: (ceph::__ceph_assert_fail(char const*, char const*, int, char const*)+0x7f) [0xaa870f]
2: (ECBackend::handle_sub_read(pg_shard_t, ECSubRead&, ECSubReadReply*)+0x932) [0x949782]
3: (ECBackend::handle_message(std::tr1::shared_ptr<OpRequest>)+0x4b1) [0x957461]
4: (ReplicatedPG::do_request(std::tr1::shared_ptr<OpRequest>&, ThreadPool::TPHandle&)+0x15a) [0x7d571a]
5: (OSD::dequeue_op(boost::intrusive_ptr<PG>, std::tr1::shared_ptr<OpRequest>, ThreadPool::TPHandle&)+0x17f) [0x656f9f]
6: (OSD::ShardedOpWQ::_process(unsigned int, ceph::heartbeat_handle_d*)+0x65f) [0x6579ff]
7: (ShardedThreadPool::shardedthreadpool_worker(unsigned int)+0x652) [0xa99352]
8: (ShardedThreadPool::WorkThreadSharded::entry()+0x10) [0xa9aa80]
9: (()+0x7e9a) [0x7f8baf2a6e9a]
```

~~5> 2014-11-09 14:13:01.334302 7f8b93c8b700 5 --~~

```
 op tracker -- seq: 65517, time: 2014-11-09 14:13:01.334302, event: reached_pg, op: MOSDECSubOpRea
d(1.1ds0 103 ECSubRead(tid=12070, to_read={78348bbd/benchmark_data_plana76_6098_object6568/head//1
=0,524288}, attrs_to_read=78348bbd/benchmark_data_plana76_6098_object6568/head//1))
    -4> 2014-11-09 14:13:01.334313 7f8b93c8b700 10 osd.3 pg_epoch: 103 pg[1.1ds0( v 92'740 (0'0,92
'740] local-les=92 n=183 ec=9 les/c 92/92 93/98/98) [1,3,0] r=-1 lpr=98 pi=91-97/3 luod=0'0 crt=92
'737 lcod 92'737 active] handle_message: MOSDECSubOpRead(1.1ds0 103 ECSubRead(tid=12070, to_read={
78348bbd/benchmark_data_plana76_6098_object6568/head//1=0,524288}, attrs_to_read=78348bbd/benchmar
k_data_plana76_6098_object6568/he
ad//1)) v1
    -3> 2014-11-09 14:13:01.334333 7f8b93c8b700 15 filestore(/var/lib/ceph/osd/ceph-3) read 1.1ds0
_head/78348bbd/benchmark_data_plana76_6098_object6568/head//1/18446744073709551615/0 0~524288
    -2> 2014-11-09 14:13:01.334397 7f8b93c8b700 10 filestore(/var/lib/ceph/osd/ceph-3) error openi
ng file /var/lib/ceph/osd/ceph-3/current/1.1ds0_head/DIR_D/DIR_B/benchmark\udata\uplana76\u6098\uo
bject6568__head_78348BBD__1_ffffffffffffffff_0 with flags=2: (2) No such file or directory
```

```
2014-11-09 14:12:33.028797 7f8b93c8b700 20 osd.3 pg_epoch: 92 pg[1.1ds0( v 92'740 (0'0,92'740] local-les=92 n=183 ec=9 les/c
92/92 90/91/91) [2147483647,4,5]/[3,4,5] r=0 lpr=91 luod=92'738 crt=92'737 lcod 92'737 mlcod 92'737 active+remapped] check_op
tid 12344: Op(78348bbd/benchmark_data_plana76_6098_object6568/head//1 v=92'739 tt=0'0 tid=12344 reqid=client.4115.0:19498
client_op=osd_op(client.4115.0:19498 benchmark_data_
plana76_6098_object6568 [delete] 1.78348bbd ack+ondisk+write+known_if_redirected e92) pending_commit=5(2)
pending_apply=5(2))
...
2014-11-09 14:12:57.902407 7f8b96c91700 20 merge_log 92'735 (83'362) delete
```

```
4451151d/benchmark_data_plana76_6098_object6501/head//1 by client.4115.0:19431 2014-11-09 14:12:23.055232
2014-11-09 14:12:57.902415 7f8b96c91700 20 merge_log 92'736 (83'363) delete
d91728bd/benchmark_data_plana76_6098_object6505/head//1 by client.4115.0:19435 2014-11-09 14:12:23.447074
2014-11-09 14:12:57.902422 7f8b96c91700 20 merge_log 92'737 (83'364) delete
e9a88cbd/benchmark_data_plana76_6098_object6515/head//1 by client.4115.0:19445 2014-11-09 14:12:24.123917
2014-11-09 14:12:57.902430 7f8b96c91700 20 merge_log 92'738 (83'365) delete
a1d82bbd/benchmark_data_plana76_6098_object6521/head//1 by client.4115.0:19451 2014-11-09 14:12:24.421805
2014-11-09 14:12:57.902751 7f8b96c91700 10 merge_log result log((0'0,92'738], crt=92'738) missing(185) changed=1
2014-11-09 14:12:57.903244 7f8b96c91700  5 osd.3 pg_epoch: 100 pg[1.1ds1( v 92'738 lc 0'0 (0'0,92'738] local-les=0 n=185 ec=9
les/c 92/92 93/98/98) [1,3,0] r=1 lpr=100 pi=91-97/3 crt=92'738 inactive m=185] exit Started/Stray 0.011615 1 0.000080
```

So, the object was deleted, but that entry ended up divergent. A different shard was created on the osd (shard 1) which then the primary did not realize didn't have a copy.

ubuntu@teuthology:/a/sage-2014-11-09_07:49:57-rados-next-testing-basic-multi/592008/remote

Related issues:

Duplicated by Ceph - Bug #10107: Coredump in upgrade:giant-x-next-distro-basi...	**Duplicate**	**11/14/2014**

Associated revisions

Revision c87bde64 - 11/14/2014 11:51 PM - Samuel Just

PG: always clear_primary_state when leaving Primary

Otherwise, entries from the log collection process might leak into the next epoch, where we might end up choosing a different authoritative log. In this case, it resulted in us not rolling back to log entries on one of the replicas prior to trying to recover from an affected object due to the peer_missing not being cleared.

Fixes: #10059
Backport: giant, firefly, dumpling
Signed-off-by: Samuel Just <sjust@redhat.com>

Revision 8b07236c - 03/11/2015 09:39 AM - Samuel Just

PG: always clear_primary_state when leaving Primary

Otherwise, entries from the log collection process might leak into the next epoch, where we might end up choosing a different authoritative log. In this case, it resulted in us not rolling back to log entries on one of the replicas prior to trying to recover from an affected object due to the peer_missing not being cleared.

Fixes: #10059
Backport: giant, firefly, dumpling
Signed-off-by: Samuel Just <sjust@redhat.com>
(cherry picked from commit c87bde64dfccb5d6ee2877cc74c66fc064b1bcd7)

Revision 0c3f7637 - 03/19/2015 09:08 AM - Samuel Just

PG: always clear_primary_state when leaving Primary

Otherwise, entries from the log collection process might leak into the next epoch, where we might end up choosing a different authoritative log. In this case, it resulted in us not rolling back to log entries on one of the replicas prior to trying to recover from an affected object due to the peer_missing not being cleared.

Fixes: #10059
Backport: giant, firefly, dumpling
Signed-off-by: Samuel Just <sjust@redhat.com>
(cherry picked from commit c87bde64dfccb5d6ee2877cc74c66fc064b1bcd7)

History

#1 - 11/10/2014 04:09 PM - Dmitry Smirnov

This bug makes me cry as it is the reason for my cluster to be *completely down* for over 10 days now... Duplicate added...

#2 - 11/13/2014 05:02 PM - Dmitry Smirnov

Any progress?

#3 - 11/14/2014 03:12 PM - Samuel Just

This is almost certainly unrelated to those two bugs. This is a specific edge case in divergent write recovery.

#4 - 11/14/2014 03:45 PM - Samuel Just

- Status changed from New to Testing

#5 - 11/16/2014 02:09 AM - Dmitry Smirnov

Samuel Just wrote:

> This is almost certainly unrelated to those two bugs. This is a specific edge case in divergent write recovery.

Obviously you know better so there is no surprise that c87bde64 did not help #9978...

#6 - 12/12/2014 02:24 PM - Samuel Just

- Status changed from Testing to Pending Backport

- Backport set to giant,firefly

#7 - 12/19/2014 02:08 AM - Dmitry Smirnov

I applied c87bde64 on top of 0.89 on degraded cluster but it caused many PGs to stuck in "inactive" state so I had to revert it...

#8 - 03/11/2015 09:40 AM - Loic Dachary

- firefly backport https://github.com/ceph/ceph/pull/3955

#9 - 03/11/2015 09:42 AM - Loic Dachary

- original pull request https://github.com/ceph/ceph/pull/3154

#10 - 03/19/2015 09:08 AM - Loic Dachary

- giant backport https://github.com/ceph/ceph/pull/4090

#11 - 03/26/2015 01:36 PM - Loic Dachary

0c3f763 PG: always clear_primary_state when leaving Primary (in giant), 8b07236 PG: always clear_primary_state when leaving Primary (in firefly),

#12 - 03/26/2015 01:36 PM - Loic Dachary

- Status changed from Pending Backport to Resolved