

## bluestore - Bug #23540

### FAILED assert(0 == "can't mark unloaded shard dirty") with compression enabled

04/02/2018 08:22 PM - Francisco Freire

<b>Status:</b> Resolved	<b>% Done:</b> 0%
<b>Priority:</b> High	
<b>Assignee:</b> Igor Fedotov	
<b>Category:</b>	
<b>Target version:</b>	
<b>Source:</b>	<b>Reviewed:</b>
<b>Tags:</b>	<b>Affected Versions:</b>
<b>Backport:</b> mimic,luminous	<b>ceph-qa-suite:</b>
<b>Regression:</b> No	<b>Pull request ID:</b>
<b>Severity:</b> 1 - critical	<b>Crash signature:</b>
<b>Description</b>	
<p>We are using the latest ceph luminous version (12.2.4), and we have a SATA pool tiered by an SSD pool. All using bluestore, and this bug only occurs on the SSD pool. I changed some OSD's do filestore and everything works fine, i got this error like 2 or 3 times a day on EACH osd, causing then to go down and restart. I have to keep the noout flag on cluster to get everything running.</p> <p>This ceph cluster is used on Openstack on vm disks (nova) and volumes (cinder)</p> <p>Thanks!</p> <p>/build/ceph-12.2.4/src/os/bluestore/BlueStore.cc: 2714: FAILED assert(0 == "can't mark unloaded shard dirty")</p> <pre>ceph version 12.2.4 (52085d5249a80c5f5121a76d6288429f35e4e77b) luminous (stable)  1: (ceph::__ceph_assert_fail(char const*, char const*, int, char const*)+0x102) [0x561c785ce872]  2: (BlueStore::ExtentMap::dirty_range(unsigned int, unsigned int)+0x54a) [0x561c7841927a]  3: (BlueStore::_do_write(BlueStore::TransContext*, boost::intrusive_ptr&lt;BlueStore::Collection&gt;&amp;, boost::intrusive_ptr&lt;BlueStore::Onode&gt;&amp;, unsigned long, unsigned long, ceph::buffer::list&amp;, unsigned int)+0x4d9) [0x561c7847e4b9]  4: (BlueStore::_write(BlueStore::TransContext*, boost::intrusive_ptr&lt;BlueStore::Collection&gt;&amp;, boost::intrusive_ptr&lt;BlueStore::Onode&gt;&amp;, unsigned long, unsigned long, ceph::buffer::list&amp;, unsigned int)+0xfc) [0x561c7847ef9c]  5: (BlueStore::_txc_add_transaction(BlueStore::TransContext*, ObjectStore::Transaction*)+0x1b34) [0x561c78485ea4]  6: (BlueStore::queue_transactions(ObjectStore::Sequencer*, std::vector&lt;ObjectStore::Transaction, std::allocator&lt;ObjectStore::Transaction&gt;&gt;&amp;, boost::intrusive_ptr&lt;TrackedOp&gt;, ThreadPool::TPHandle*)+0x52e) [0x561c7848702e]  7: (PrimaryLogPG::queue_transactions(std::vector&lt;ObjectStore::Transaction, std::allocator&lt;ObjectStore::Transaction&gt;&gt;&amp;, boost::intrusive_ptr&lt;OpRequest&gt;&amp;)+0x66) [0x561c781ae256]  8: (ReplicatedBackend::do_repop(boost::intrusive_ptr&lt;OpRequest&gt;&amp;)+0xc34) [0x561c782d37a4]  9: (ReplicatedBackend::_handle_message(boost::intrusive_ptr&lt;OpRequest&gt;&amp;)+0x294) [0x561c782dc834] 10: (PGBackend::handle_message(boost::intrusive_ptr&lt;OpRequest&gt;&amp;)+0x50) [0x561c781ebca0] 11: (PrimaryLogPG::do_request(boost::intrusive_ptr&lt;OpRequest&gt;&amp;, ThreadPool::TPHandle&amp;)+0x543) [0x561c781509d3] 12: (OSD::dequeue_op(boost::intrusive_ptr&lt;PG&gt;, boost::intrusive_ptr&lt;OpRequest&gt;, ThreadPool::TPHandle&amp;)+0x3a9) [0x561c77fca3b9] 13: (PGQueueable::RunVis::operator() (boost::intrusive_ptr&lt;OpRequest&gt;&amp; const&amp;)+0x57) [0x561c7826d047] 14: (OSD::ShardedOpWQ::_process(unsigned int, ceph::heartbeat_handle_d*)+0x130e) [0x561c77ff29ae] 15: (ShardedThreadPool::shardedthreadpool_worker(unsigned int)+0x884) [0x561c785d3664] 16: (ShardedThreadPool::WorkThreadSharded::entry()+0x10) [0x561c785d66a0] 17: (()+0x76ba) [0x7f040b0e96ba] 18: (clone()+0x6d) [0x7f040a16041d] NOTE: a copy of the executable, or `objdump -rdS &amp;lt;executable&gt;` is needed to interpret this.</pre>	
<b>Related issues:</b>	

Copied to bluestore - Backport #24798: luminous: FAILED assert(0 == "can't ma...

Resolved

Copied to bluestore - Backport #24799: mimic: FAILED assert(0 == "can't mark ...

Resolved

## History

---

### #1 - 04/02/2018 09:56 PM - Igor Fedotov

Hi Francisco,  
wondering if you have compression enabled for any of your pools or the whole bluestore?

### #2 - 04/02/2018 10:40 PM - Igor Fedotov

- Project changed from Ceph to bluestore  
- Category deleted (OSD)

### #3 - 04/02/2018 11:50 PM - Francisco Freire

Yeah. The whole cluster has compression enabled

### #4 - 04/04/2018 02:53 PM - Francisco Freire

I disabled the compression for a while and no OSD's get the error, after enabled it again they back to get the problem, for sure this is caused by compression.

### #5 - 04/04/2018 03:12 PM - Igor Fedotov

Francisco,  
thanks for the update, very appreciated.

Curious if you can collect a log for the crushing OSD, with debug bluestore set to 20. It would be very helpful for troubleshooting. This might impact cluster performance though...

### #6 - 04/04/2018 03:12 PM - Igor Fedotov

- Assignee set to Igor Fedotov

### #7 - 04/09/2018 01:19 PM - Francisco Freire

Hello,  
We are running for a few days without problem (with compression disabled), to get a debug i need to enable compression again, il do it on a timeframe and post here.

Thanks

### #8 - 04/12/2018 02:47 PM - Sage Weil

- Priority changed from Normal to High

### #9 - 04/23/2018 10:13 PM - Sage Weil

- Subject changed from FAILED assert(0 == "can't mark unloaded shard dirty") to FAILED assert(0 == "can't mark unloaded shard dirty") with compression enabled  
- Status changed from New to Need More Info

Francisco, any update?

**#10 - 05/11/2018 02:22 PM - Yohay Azulay**

I had the same issue with 3 new clusters. compression set to FORCE, once I changed to none and restarted the whole cluster, the osds stopped flapping.

Any logs I can send that can help?

**#11 - 05/14/2018 09:34 AM - Igor Fedotov**

Hi Yohay,  
could you please collect a log for the crash with debug bluestore set to 20?

**#12 - 05/14/2018 01:50 PM - Yohay Azulay**

- File *ceph-osd.15.zip* added

That can be a problem because I disabled the Compression and cluster is running in production. if I enable compression it will crush again.. :(

log file I have is 110MB and 8MB compressed, it repeats it self and I attached a sample of the log.

Igor Fedotov wrote:

Hi Yohay,  
could you please collect a log for the crash with debug bluestore set to 20?

**#13 - 06/15/2018 06:37 PM - Sage Weil**

Hi everyone,

Is someone willing to enable compression on a bluestore osd with debugging enabled (debug bluestore = 20) so that we can capture a complete log leading up to the crash that would be extremely helpful. Igor has been unable to reproduce this in his own environment so we currently don't have much to go on. Perhaps enabling debugging and compression on only a handful of OSDs would be sufficient with a small impact on the production environment (performance, /var/log/ceph disk usage, etc.).

**#14 - 06/30/2018 07:28 AM - Yohay Azulay**

Got it.. here it is: <http://77.247.180.45/ceph-osd.11.log.debug.gz>

Sage Weil wrote:

Hi everyone,

Is someone willing to enable compression on a bluestore osd with debugging enabled (debug bluestore = 20) so that we can capture a complete log leading up to the crash that would be extremely helpful. Igor has been unable to reproduce this in his own environment so we currently don't have much to go on. Perhaps enabling debugging and compression on only a handful of OSDs would be sufficient with a small impact on the production environment (performance, /var/log/ceph disk usage, etc.).

**#15 - 07/02/2018 01:42 PM - Igor Fedotov**

Yohay,  
I can't access the file at the link you provided, "Not found" returned..

**#16 - 07/02/2018 01:45 PM - Yohay Azulay**

arghh. my mistake, <http://77.247.180.45/download/ceph-osd.11.log.debug.gz>

Igor Fedotov wrote:

Yohay,  
I can't access the file at the link you provided, "Not found" returned..

**#17 - 07/02/2018 01:47 PM - Igor Fedotov**

Much better now :)  
Thanks a lot!!!

**#18 - 07/03/2018 04:06 PM - Peter Gervai**

ceph-post-file: 1b1d42bb-6cae-430a-8fe7-974ce077b8dc  
May (or may not) help, it's around loglevel5 I guess.

**#19 - 07/05/2018 11:46 AM - Igor Fedotov**

- Status changed from Need More Info to Fix Under Review

<https://github.com/ceph/ceph/pull/22873>

**#20 - 07/05/2018 02:34 PM - Sage Weil**

- Backport set to *mimic,luminous*

**#21 - 07/06/2018 04:12 PM - Igor Fedotov**

- Related to Backport #24798: *luminous: FAILED assert(0 == "can't mark unloaded shard dirty") with compression enabled added*

**#22 - 07/06/2018 04:25 PM - Igor Fedotov**

- Related to Backport #24799: *mimic: FAILED assert(0 == "can't mark unloaded shard dirty") with compression enabled added*

**#23 - 07/06/2018 06:43 PM - Sage Weil**

- Status changed from Fix Under Review to Pending Backport

**#26 - 07/06/2018 09:14 PM - Nathan Cutler**

- Related to deleted (Backport #24798: *luminous: FAILED assert(0 == "can't mark unloaded shard dirty") with compression enabled*)

**#27 - 07/06/2018 09:14 PM - Nathan Cutler**

- Related to deleted (Backport #24799: *mimic: FAILED assert(0 == "can't mark unloaded shard dirty") with compression enabled*)

**#28 - 07/06/2018 09:15 PM - Nathan Cutler**

- Copied to Backport #24798: luminous: FAILED assert(0 == "can't mark unloaded shard dirty") with compression enabled added

**#29 - 07/06/2018 09:15 PM - Nathan Cutler**

- Copied to Backport #24799: mimic: FAILED assert(0 == "can't mark unloaded shard dirty") with compression enabled added

**#30 - 07/06/2018 09:21 PM - Nathan Cutler**

@Igor - please use "Copied To" instead of "Related To" for the links to the backport issues. All good otherwise, thanks!

Note: you can automate creation of backport issues for issues in Pending Backport status using src/script/backport-create-issue (but this script currently does not know how to limit itself to a single issue - it loops through them all).

**#32 - 08/06/2018 04:37 PM - Igor Fedotov**

- Status changed from Pending Backport to Resolved

**Files**

---

ceph-osd.15.zip	197 KB	05/14/2018	Yohay Azulay
-----------------	--------	------------	--------------