

Ceph - Bug #5723

OSD seemingly loses objects during crash

07/23/2013 10:01 AM - Mike Lowe

<b>Status:</b>	Resolved	<b>% Done:</b>	0%
<b>Priority:</b>	Urgent	<b>Spent time:</b>	0.00 hour
<b>Assignee:</b>	Samuel Just		
<b>Category:</b>	OSD		
<b>Target version:</b>			
<b>Source:</b>	other	<b>Affected Versions:</b>	
<b>Tags:</b>		<b>ceph-qa-suite:</b>	
<b>Backport:</b>		<b>Pull request ID:</b>	
<b>Regression:</b>	No	<b>Crash signature (v1):</b>	
<b>Severity:</b>	2 - major	<b>Crash signature (v2):</b>	
<b>Reviewed:</b>			
<b>Description</b>			
I had some vm's with the qemu-rbd driver doing a trim operation. One of my osd's crashed and now has an inconsistent pg following restart. OSD 14 was the primary and crashed while OSD 6 is the secondary. The missing objects are like this:			
2.37d osd.14 missing f30b0f7d/rb.0.105b.238e1f29.000000000ff4/head//2			
However it appears to be on disk on both the primary and secondary, just in different places in the directory tree depending on the OSD			
<pre>find /data/osd.14/current/2.37d_head/ -name 'rb.0.105b.238e1f29.000000000ff4*' /data/osd.14/current/2.37d_head/DIR_D/rb.0.105b.238e1f29.000000000ff4__head_F30B0F7D__2</pre>			
<pre>find /data/osd.6/current/2.37d_head/ -name 'rb.0.105b.238e1f29.000000000ff4*' /data/osd.6/current/2.37d_head/DIR_D/DIR_7/rb.0.105b.238e1f29.000000000ff4__head_F30B0F7D__2</pre>			

Associated revisions

Revision 0dc3efdd - 07/25/2013 05:32 PM - Samuel Just

HashIndex: reset attr upon split or merge completion

A replay of an in progress merge or split might make our counts unreliable.

Fixes: #5723  
Signed-off-by: Samuel Just <[sam.just@inktank.com](mailto:sam.just@inktank.com)>  
Reviewed-by: Sage Weil <[sage@inktank.com](mailto:sage@inktank.com)>

Revision b0535fcf - 07/25/2013 05:37 PM - Samuel Just

HashIndex: reset attr upon split or merge completion

A replay of an in progress merge or split might make our counts unreliable.

Fixes: #5723  
Signed-off-by: Samuel Just <[sam.just@inktank.com](mailto:sam.just@inktank.com)>  
Reviewed-by: Sage Weil <[sage@inktank.com](mailto:sage@inktank.com)>

History

---

#1 - 07/23/2013 10:04 AM - Mike Lowe

OSD crashed without writing any logs

#2 - 07/23/2013 03:35 PM - Samuel Just

HashIndex merge needs to verify the collection contents before merging. In the mean time, you can recover by adjusting the cephos.phash.contents for DIR\_D/DIR\_7 from

(02:34:44 PM) jmlowe1: cephos.phash.contents  
(02:34:44 PM) jmlowe1: 0000000: 0109 0000 0000 0000 0000 0002 0000 .....  
(02:34:44 PM) jmlowe1: 0000010: 00

to

(03:34:13 PM) sjust: 0000000: 01dc 0000 0000 0000 0002 0000 0002 0000  
(03:34:13 PM) sjust: 0000010: 00

#3 - 07/23/2013 03:36 PM - Samuel Just

- Assignee set to Samuel Just
- Priority changed from High to Urgent

#4 - 07/23/2013 03:41 PM - Samuel Just

You also need to move all of the objects from DIR\_D to DIR\_D/DIR\_7 again

#5 - 07/23/2013 05:04 PM - Mike Lowe

removing cephos.phash.in\_progress\_op, setting cephos.phash.contents, moving the files and restarting seems to have resolved the missing objects

#6 - 07/25/2013 09:35 AM - Sage Weil

- Status changed from New to Fix Under Review

#7 - 07/25/2013 10:38 AM - Samuel Just

- Status changed from Fix Under Review to Pending Backport

#8 - 07/25/2013 03:34 PM - Sage Weil

- Status changed from Pending Backport to Resolved