

## Ceph - Bug #1694

### monitor crash: FAILED assert(get\_max\_osd() >= crush.get\_max\_devices())

11/08/2011 07:01 AM - Wido den Hollander

<b>Status:</b>	Resolved	<b>% Done:</b>	0%
<b>Priority:</b>	Normal	<b>Spent time:</b>	4.50 hours
<b>Assignee:</b>	Sage Weil		
<b>Category:</b>	Monitor		
<b>Target version:</b>	v0.39		
<b>Source:</b>		<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>	3 - minor	<b>Crash signature (v2):</b>	
<b>Reviewed:</b>			

#### Description

I just did a fresh install of my cluster and after starting I saw my monitors go down with:

```
Nov 8 14:40:35 monitor-sec mon.sec[1611]: ./osd/OSDMap.h: In function 'int OSDMap::_pg_to_osds(const pg_pool_t&, pg_t, std::vector<int>&)', in thread '7f9c04742700'#012./osd/OSDMap.h: 454: FAILED assert(get_max_osd() >= crush.get_max_devices())
```

```
Nov 8 14:40:35 monitor-sec mon.sec[1611]: ceph version 0.37-314-g40843eb (commit:40843eb36c3c029925d62f35aa8a4dee2876381c)#012 1: /usr/bin/ceph-mon() [0x45c27f]#012 2: (PGMonitor::send_pg_create_s()+0x15b4) [0x4c6974]#012 3: (PGMonitor::update_from_paxos()+0x4ef) [0x4c959f]#012 4: (PaxosService::_active()+0x39) [0x4861e9]#012 5: (Context::complete(int)+0xa) [0x472aca]#012 6: (finish_contexts(CephContext*, std::list<Context*, std::allocator<Context*> >&, int)+0xca) [0x482e2a]#012 7: (Paxos::handle_lease(MMonPaxos*)+0x36b) [0x47d35b]#012 8: (Paxos::dispatch(PaxosServiceMessage*)+0x21b) [0x481d3b]#012 9: (Monitor::_ms_dispatch(Message*)+0x84e) [0x47062e]#012 10: (Monitor::ms_dispatch(Message*)+0x35) [0x479485]#012 11: (SimpleMessenger::dispatch_entry()+0x84b) [0x56d05b]#012 12: (SimpleMessenger::DispatchThread::entry()+0x1c) [0x45fd5c]#012 13: (()+0x7efc) [0x7f9c08007efc]#012 14: (clone()+0x6d) [0x7f9c06a4189d]
```

```
Nov 8 14:40:35 monitor-sec mon.sec[1611]: ceph version 0.37-314-g40843eb (commit:40843eb36c3c029925d62f35aa8a4dee2876381c)#012 1: /usr/bin/ceph-mon() [0x45c27f]#012 2: (PGMonitor::send_pg_create_s()+0x15b4) [0x4c6974]#012 3: (PGMonitor::update_from_paxos()+0x4ef) [0x4c959f]#012 4: (PaxosService::_active()+0x39) [0x4861e9]#012 5: (Context::complete(int)+0xa) [0x472aca]#012 6: (finish_contexts(CephContext*, std::list<Context*, std::allocator<Context*> >&, int)+0xca) [0x482e2a]#012 7: (Paxos::handle_lease(MMonPaxos*)+0x36b) [0x47d35b]#012 8: (Paxos::dispatch(PaxosServiceMessage*)+0x21b) [0x481d3b]#012 9: (Monitor::_ms_dispatch(Message*)+0x84e) [0x47062e]#012 10: (Monitor::ms_dispatch(Message*)+0x35) [0x479485]#012 11: (SimpleMessenger::dispatch_entry()+0x84b) [0x56d05b]#012 12: (SimpleMessenger::DispatchThread::entry()+0x1c) [0x45fd5c]#012 13: (()+0x7efc) [0x7f9c08007efc]#012 14: (clone()+0x6d) [0x7f9c06a4189d]
```

This seems to be due to [885d71481bf06915569fad938a0245097f2a9e0](https://bugzilla.ceph.com/show_bug.cgi?id=885d71481bf06915569fad938a0245097f2a9e0)

On that specific monitor I checked the OSD maps and this showed:

```
root@monitor-sec:/var/lib/ceph/mon.sec/osdmap_full# osdmactool --print 3
osdmactool: osdmap file '3'
epoch 3
fsid 4bd06b88-1d07-53de-ea22-73f1fb4fe0c4
created 2011-11-08 14:09:40.450317
modified 2011-11-08 14:39:01.498778
flags
```

```
pool 0 'data' rep size 2 crush_ruleset 0 object_hash rjenkins pg_num 2496 pgp_num 2496 lpg_num 2 1
pgp_num 2 last_change 1 owner 0 crash_replay_interval 45
```

```
pool 1 'metadata' rep size 2 crush_ruleset 1 object_hash rjenkins pg_num 2496 pgp_num 2496 lpg_num
 2 lpgp_num 2 last_change 1 owner 0
pool 2 'rbd' rep size 2 crush_ruleset 2 object_hash rjenkins pg_num 2496 pgp_num 2496 lpg_num 2 lp
gp_num 2 last_change 1 owner 0

max_osd 39
osd.17 up in weight 1 up_from 2 up_thru 0 down_at 0 last_clean_interval [0,0] [2a00:f10:11b:cef
0:225:90ff:fe33:49cc]:6803/5177 [2a00:f10:11b:cef0:225:90ff:fe33:49cc]:6804/5177 [2a00:f10:11b:cef
0:225:90ff:fe33:49cc]:6805/5177
osd.20 up in weight 1 up_from 3 up_thru 0 down_at 0 last_clean_interval [0,0] [2a00:f10:11b:cef
0:225:90ff:fe33:497c]:6800/8492 [2a00:f10:11b:cef0:225:90ff:fe33:497c]:6801/8492 [2a00:f10:11b:cef
0:225:90ff:fe33:497c]:6802/8492
osd.22 up in weight 1 up_from 3 up_thru 0 down_at 0 last_clean_interval [0,0] [2a00:f10:11b:cef
0:225:90ff:fe33:497c]:6806/8693 [2a00:f10:11b:cef0:225:90ff:fe33:497c]:6807/8693 [2a00:f10:11b:cef
0:225:90ff:fe33:497c]:6808/8693
osd.29 up in weight 1 up_from 3 up_thru 0 down_at 0 last_clean_interval [0,0] [2a00:f10:11b:cef
0:225:90ff:fe33:499c]:6803/5478 [2a00:f10:11b:cef0:225:90ff:fe33:499c]:6804/5478 [2a00:f10:11b:cef
0:225:90ff:fe33:499c]:6805/5478
osd.30 up in weight 1 up_from 3 up_thru 0 down_at 0 last_clean_interval [0,0] [2a00:f10:11b:cef
0:225:90ff:fe33:499c]:6806/5600 [2a00:f10:11b:cef0:225:90ff:fe33:499c]:6807/5600 [2a00:f10:11b:cef
0:225:90ff:fe33:499c]:6808/5600
osd.31 up in weight 1 up_from 3 up_thru 0 down_at 0 last_clean_interval [0,0] [2a00:f10:11b:cef
0:225:90ff:fe33:499c]:6809/5711 [2a00:f10:11b:cef0:225:90ff:fe33:499c]:6810/5711 [2a00:f10:11b:cef
0:225:90ff:fe33:499c]:6811/5711
osd.36 up in weight 1 up_from 3 up_thru 0 down_at 0 last_clean_interval [0,0] [2a00:f10:11b:cef
0:225:90ff:fe33:49a4]:6800/5663 [2a00:f10:11b:cef0:225:90ff:fe33:49a4]:6801/5663 [2a00:f10:11b:cef
0:225:90ff:fe33:49a4]:6802/5663
osd.37 up in weight 1 up_from 3 up_thru 0 down_at 0 last_clean_interval [0,0] [2a00:f10:11b:cef
0:225:90ff:fe33:49a4]:6803/5753 [2a00:f10:11b:cef0:225:90ff:fe33:49a4]:6804/5753 [2a00:f10:11b:cef
0:225:90ff:fe33:49a4]:6805/5753

root@monitor-sec:/var/lib/ceph/mon.sec/osdmap_full#
```

I extracted the crushmap (attached) out the osdmap and that shows:

```
root@monitor-sec:~# cat /root/crushmap.txt |grep item|grep osd|wc -l
40
root@monitor-sec:~#
```

I don't see why this assert should come up? 39 (max\_osd) is less then 40 (max devices).

Could it be a problem that the "devices" were renamed to "items" in the crushmap? I haven't dumped max\_devices out of the crushmap to test it though.

## History

### #1 - 11/08/2011 07:18 AM - Wido den Hollander

I just made a small adjustment to crushtool so it would print max\_devices:

```
root@monitor-sec:~# ./crushtool -d crushmap -o crushmap.txt
max_devices 40
root@monitor-sec:~#
```

That seems OK?

**#2 - 11/08/2011 09:17 AM - Sage Weil**

- Target version set to v0.39

max\_osd in the osdmap needs to be >= the max\_devices in the crush map. how did you set up the cluster? did mkcephfs generate the crush map or did you feed one in manually?

**#3 - 11/08/2011 10:06 AM - Wido den Hollander**

Aha! Read that wrong, tnx.

I used mkcephfs to generate the crushmap, I did not write my own.

**#4 - 11/08/2011 09:28 PM - Sage Weil**

Can you try this and see if there is a mismatch?

```
osdmapprootool --create-from-conf -c your.ceph.conf osdmap
osdmapprootool -p osdmap | grep max
osdmapprootool --export-crush crushmap osdmap
crushtool -d crushmap | grep device | tail -1
```

**#5 - 11/08/2011 10:20 PM - Sage Weil**

- Status changed from New to Need More Info

**#6 - 11/09/2011 02:25 AM - Wido den Hollander**

Ok, I've ran those commands and it gives me:

```
root@monitor:~# osdmapprootool -p osdmap | grep max
max_osd 39
root@monitor:~#
```

So, that is one short of the 40 I have.

```
root@monitor:~# crushtool -d crushmap | grep device | tail -1
device 39 osd.39
root@monitor:~#
```

That is correct, also, if I check the crushmap:

```
root@monitor:~# crushtool -d crushmap | grep device | grep osd | wc -l
40
root@monitor:~#
```

So, max\_osd is set one short in the osdmap.

**#7 - 11/09/2011 07:55 AM - Wido den Hollander**

The monitor that was generating the osdmap was running [5bd029ef01fcb59bea9170af563c3499cce1e8c4](#) and that failed.

I just ran it with the latest master and that gave me a map with max\_osd = 40, but I don't see a change in the last 24 hours, or did I miss that?

I'll update asap with a test with the latest master.

**#8 - 11/09/2011 11:06 AM - Sage Weil**

- Assignee set to Sage Weil

Great. Can you attach (or email) the ceph.conf you're using?

Thanks!

**#9 - 11/09/2011 11:09 AM - Sage Weil**

oh nevermind, didn't see that second comment. the fix is [0bcdd4f3b2a2dba405639122b84f7aad978f347b](#), which comes after [5bd029ef01fcb59bea9170af563c3499cce1e8c4](#).

**#10 - 11/09/2011 11:09 AM - Sage Weil**

- Status changed from Need More Info to Resolved

**Files**

---

crushmap.txt	3.61 KB	11/08/2011	Wido den Hollander
--------------	---------	------------	--------------------