Ceph - Bug #7328

osd: reweight-by-utilization ended up with stuck remapped pgs

02/03/2014 06:07 PM - Tyler Brekke

Status:	Resolved	**% Done:**	0%
Priority:	Urgent	**Spent time:**	0.00 hour
Assignee:	Sage Weil		
Category:			
Target version:			
Source:	Support	**Affected Versions:**	
Tags:		**ceph-qa-suite:**	
Backport:		**Pull request ID:**	
Regression:	No	**Crash signature (v1):**	
Severity:	3 - minor	**Crash signature (v2):**	
Reviewed:			

Description

Running ceph osd reweight-by-utilization resulted in stuck pgs.

```
health HEALTH_WARN 204 pgs stuck unclean; recovery 4136/163619494 objects degraded (0.003%)
monmap e1: 1 mons at {cs-compute03=192.168.181.13:6789/0}, election epoch 1, quorum 0 cs-compute03

osdmap e2996: 120 osds: 120 up, 120 in
pgmap v665045: 6128 pgs, 4 pools, 164 TB data, 79892 kobjects
329 TB used, 106 TB / 435 TB avail
4136/163619494 objects degraded (0.003%)
5917 active+clean
204 active+remapped
```

I believe the cause is related to the crush configuration. 2 copies split across 2 rooms. Weighting the OSDs back up to 1 resolves the remapped pgs.

```
root root2 {
    id -7        # do not change unnecessarily
    # weight 435.600
    alg straw
    hash 0    # rjenkins1
    item room1 weight 217.800
    item room2 weight 217.800
}

rule testrule {
    ruleset 3
    type replicated
    min_size 2
    max_size 4
    step take root2
    step chooseleaf firstn 0 type room
    step emit
}
```

If this is just caused from the crush configuration reweight-by-utilization should be smarter about weighting down OSDs.

More cluster information is available in ZD Ticket #928

History

#1 - 02/04/2014 09:41 AM - Ian Colle

- Assignee set to Sage Weil

- Priority changed from Normal to Urgent

#2 - 02/20/2014 10:52 AM - Sage Weil

- Status changed from New to Resolved

this came down to a crush flaw. there is a new tunable to address it in firefly, although it will remain off for the time being. (users who have all-firefly clients can make use of it with 'ceph osd crush tunables optimal')

e88f843c99c0a45390f71c2a5e53a52305c041fd and related commits.