

RADOS - Bug #38786

autoscale down can lead to max_pg_per_osd limit

03/16/2019 06:33 PM - Sage Weil

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

Description

we adjust pgp_num all the way down to the target, which can make osds hit the max_pgs_per_osd if it's going too far.

saw this on the lab cluster,

```
pool 4 'libvirt-pool' replicated size 3 min_size 2 crush_rule 0 object_hash rjenkins pg_num 3541 p
gp_num 12 pg_num_target 4 pgp_num_target 4 autoscale_mode on last_change 1096029 lfor 0/1096029/10
96025 flags hashspool min_write_recency_for_promote 1 stripe_width 0 application libvirt
```

```
root@reesi006:~# ceph pg ls activating
```

PG	OBJECTS	DEGRADED	MISPLACED	UNFOUND	BYTES	OMAP_BYTES*	OMAP_KEYS*	LOG	STATE
SINCE VERSION	REPORTED	UP	ACTING	SCRUB_STAMP	DEEP_S	CRUB_STAMP			
4.1a4	0	0	0	0	0	0	0	0	activating
6m	0'0	1097345:1340	[24,69,46]p24	[24,69,46]p24	2019-03-14 09:14:01.948061	2019-03-14 09:14:01.948061			
4.1b4	1	3	0	0	4194304	0	0	2	activating+degraded
6m	405308'2	1097345:1336	[24,69,46]p24	[24,69,46]p24	2019-03-15 19:01:51.539002	2019-03-15 19:01:51.539002			
4.1d4	1	3	0	0	4194304	0	0	2	activating+degraded
6m	405540'2	1097345:1331	[24,69,46]p24	[24,69,46]p24	2019-03-14 13:00:44.455889	2019-03-14 13:00:44.455889			
4.1e4	1	3	0	0	4194304	0	0	2	activating+degraded
6m	405300'2	1097345:1329	[24,69,46]p24	[24,69,46]p24	2019-03-15 16:37:55.135432	2019-03-15 16:37:55.135432			
4.1f4	0	0	0	0	0	0	0	0	activating
6m	0'0	1097345:1327	[24,69,46]p24	[24,69,46]p24	2019-03-15 20:01:49.959489	2019-03-15 20:01:49.959489			
4.224	0	0	0	0	0	0	0	0	activating
6m	0'0	1097345:1329	[24,69,46]p24	[24,69,46]p24	2019-03-14 18:03:28.050687	2019-03-14 18:03:28.050687			
4.234	0	0	0	0	0	0	0	0	activating
6m	0'0	1097345:1326	[24,69,46]p24	[24,69,46]p24	2019-03-15 20:32:26.246396	2019-03-15 20:32:26.246396			
4.244	0	0	0	0	0	0	0	0	activating
6m	0'0	1097345:1353	[24,69,46]p24	[24,69,46]p24	2019-03-14 12:07:57.114855	2019-03-14 12:07:57.114855			
4.274	0	0	0	0	0	0	0	0	activating
6m	0'0	1097345:1328	[24,69,46]p24	[24,69,46]p24	2019-03-15 20:27:12.199661	2019-03-15 20:27:12.199661			
4.284	0	0	0	0	0	0	0	0	activating
6m	0'0	1097345:1368	[24,69,46]p24	[24,69,46]p24	2019-03-14 17:03:26.116879	2019-03-14 17:03:26.116879			

```

3-10 02:07:07.974526
4.294      0      0      0      0      0      0      0      0      0      activating
6m      0'0      1097345:1327      [24,69,46]p24      [24,69,46]p24      2019-03-14      04:59:38.598132      2019-0
3-09 08:07:24.415034
4.2c4      0      0      0      0      0      0      0      0      0      activating
6m      0'0      1097345:1353      [24,69,46]p24      [24,69,46]p24      2019-03-14      14:44:03.602640      2019-0
3-09 11:39:35.133023
4.2e4      0      0      0      0      0      0      0      0      0      activating
6m      0'0      1097345:1327      [24,69,46]p24      [24,69,46]p24      2019-03-14      09:30:22.009430      2019-0
3-13 07:00:56.840736
4.314      0      0      0      0      0      0      0      0      0      activating
6m      0'0      1097345:1329      [24,69,46]p24      [24,69,46]p24      2019-03-13      23:23:02.233771      2019-0
3-10 02:29:07.039327
4.324      0      0      0      0      0      0      0      0      0      activating
6m      0'0      1097345:1327      [24,69,46]p24      [24,69,46]p24      2019-03-14      08:54:31.642665      2019-0
3-12 22:16:31.374791
4.334      1      3      0      0      4194304      0      0      2      activating+degraded
6m      405540'2      1097345:1329      [24,69,46]p24      [24,69,46]p24      2019-03-14      21:06:03.101677      2019-0
3-14 21:06:03.101677
4.34c      0      0      0      0      0      0      0      0      0      activating
6m      0'0      1097345:1328      [24,69,46]p24      [24,69,46]p24      2019-03-14      11:19:44.965680      2019-0
3-13 00:53:39.903505
4.364      0      0      0      0      0      0      0      0      0      activating
6m      0'0      1097345:1328      [24,69,46]p24      [24,69,46]p24      2019-03-14      10:35:56.589251      2019-0
3-13 09:02:07.524679
...

```

and on osd.69,

```

2019-03-16 18:27:06.104 7f0d5de25700 10 osd.69 1097345 handle_pg_create_info hit max pg, dropping
2019-03-16 18:27:06.104 7f0d5de25700 10 osd.69 1097345 handle_pg_create_info hit max pg, dropping
2019-03-16 18:27:06.112 7f0d5e626700 10 osd.69 1097345 handle_pg_create_info hit max pg, dropping
2019-03-16 18:27:06.112 7f0d5de25700 10 osd.69 1097345 handle_pg_create_info hit max pg, dropping

```

Related issues:

Copied to RADOS - Backport #39271: nautilus: autoscale down can lead to max_p...

Resolved

History

#1 - 04/09/2019 10:17 PM - Sage Weil

- Status changed from 12 to Fix Under Review
- Backport set to nautilus

<https://github.com/ceph/ceph/pull/27473>

#2 - 04/12/2019 11:57 AM - Sage Weil

- Status changed from Fix Under Review to Pending Backport

#3 - 04/12/2019 12:04 PM - Nathan Cutler

- Copied to Backport #39271: nautilus: autoscale down can lead to max_pg_per_osd limit added

#4 - 04/16/2019 10:53 AM - Nathan Cutler

- Status changed from Pending Backport to Resolved