

RADOS - Bug #39390

filestore pre-split may not split enough directories

04/19/2019 03:50 AM - Jeeqn Chen

Status: Resolved	% Done: 0%
Priority: Normal	Spent time: 0.00 hour
Assignee:	
Category:	
Target version:	
Source:	Affected Versions: v12.2.8
Tags:	ceph-qa-suite:
Backport: luminous, mimic, nautilus	Component(RADOS): FileStore
Regression: No	Pull request ID: 27689
Severity: 3 - minor	Crash signature (v1):
Reviewed:	Crash signature (v2):
Description	
Current HashIndex::pre_split_folder() use the following snippet to figure the number of levels for split.	
<pre>leavies /= subs; while (leavies > 1) { ++level; leavies = leavies >> 4; }</pre>	
The problem is that the number of levels may be far from enough. For example, we expect 512 PGs, about 1 million files in each PG and about 4096 file in each leave directory. So 9 bits are for PG number, which means top 2 level of directories won't be splitted. And "leavies" here is 244 and "subs" here will be 8. Thus we will only get additional 1 level of splitted directories. The number of files in a PG will be $8 * 16 * 4096$ (0.5 million) only.	
The following snippet will be able to assure enough number of leaves:	
<pre>int level = 0; int level_limit = MAX_HASH_LEVEL - dump_num - 1; uint64_t actual_leaves = subs; while (actual_leaves < leavies && level < level_limit) { ++level; actual_leaves <<= 4; }</pre>	
Related issues:	
Copied to RADOS - Backport #39681: luminous: filestore pre-split may not spli...	Rejected
Copied to RADOS - Backport #39682: nautilus: filestore pre-split may not spli...	Resolved
Copied to RADOS - Backport #41291: mimic: filestore pre-split may not split e...	Resolved

History

#1 - 04/19/2019 04:07 AM - Jeeqn Chen

<https://github.com/ceph/ceph/pull/27689>

#2 - 05/09/2019 07:14 AM - Kefu Chai

- Status changed from New to Pending Backport

- Backport set to luminous, nautilus

- Pull request ID set to 27689

#3 - 05/10/2019 10:56 AM - Nathan Cutler

- Copied to Backport #39681: luminous: filestore pre-split may not split enough directories added

#4 - 05/10/2019 10:56 AM - Nathan Cutler

- Copied to Backport #39682: nautilus: filestore pre-split may not split enough directories added

#5 - 08/15/2019 12:11 AM - Kefu Chai

- Backport changed from luminous, nautilus to luminous, mimic, nautilus

#6 - 08/15/2019 09:11 AM - Nathan Cutler

- Copied to Backport #41291: mimic: filestore pre-split may not split enough directories added

#7 - 01/27/2021 07:12 PM - Nathan Cutler

- Status changed from Pending Backport to Resolved

While running with --resolve-parent, the script "backport-create-issue" noticed that all backports of this issue are in status "Resolved" or "Rejected".