rbd - Feature #40072

[journal] tweak config defaults to improve small-IO performance

05/29/2019 06:30 PM - Jason Dillaman

<table>
<thead>
<tr>
<th>Status:</th>
<th>Resolved</th>
<th>% Done:</th>
<th>0%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Priority:</td>
<td>Normal</td>
<td>Spent time:</td>
<td>0.00 hour</td>
</tr>
<tr>
<td>Assignee:</td>
<td>Jason Dillaman</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Category:</td>
<td></td>
<td>Reviewed:</td>
<td></td>
</tr>
<tr>
<td>Target version:</td>
<td></td>
<td>Affected Versions:</td>
<td></td>
</tr>
<tr>
<td>Source:</td>
<td></td>
<td>Pull request ID:</td>
<td>28539</td>
</tr>
<tr>
<td>Tags:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Backport:</td>
<td>nautilus,mimic,luminous</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Description

Small IO throughput w/ journaling enabled is substantially worse than the expected worse-case of 2x the latency / 0.5x the IOPS. Tweak the config defaults to try to reduce the performance hit for small IO.

I suggest we create a new "rbd_journal_object_writethrough_until_flush"-style object (i.e. similar to "rbd_cache_writethrough_until_flush") which is defaulted to true. Then we can set "rbd_journal_object_flush_bytes" to 1MiB (i.e. similar to the current cache defaults when combined w/ the next point) w/o the need to worry about inconsistent IO in the case of failure since we know the app is sending flushes. Finally, increase the default "rbd_journal_splay_width" to 16 since (1) it's the default for "rbd bench", (2) it should match the max expected queue depth of workload.

The above are just ideas and should be tested to verify the best performance for small IO workloads.

Related issues:

Copied to rbd - Backport #40509: luminous: [journal] tweak config defaults to... Rejected
Copied to rbd - Backport #40510: mimic: [journal] tweak config defaults to im... Rejected
Copied to rbd - Backport #40511: nautilus: [journal] tweak config defaults to... Resolved

History

#1 - 06/11/2019 05:02 PM - Jason Dillaman
- Status changed from New to In Progress
- Assignee set to Jason Dillaman

#2 - 06/13/2019 12:19 AM - Jason Dillaman
- Backport set to nautilus

#3 - 06/13/2019 10:16 PM - Jason Dillaman
- Status changed from In Progress to Fix Under Review
- Pull request ID set to 28539

#4 - 06/13/2019 10:17 PM - Jason Dillaman
- Backport changed from nautilus to nautilus,mimic,luminous

Skip last commit (enabling writethrough mode) during backport

#5 - 06/20/2019 04:04 PM - Mykola Golub
- Status changed from Fix Under Review to Pending Backport

#6 - 06/24/2019 10:05 AM - Nathan Cutler
- Copied to Backport #40509: luminous: [journal] tweak config defaults to improve small-IO performance added
#7 - 06/24/2019 10:05 AM - Nathan Cutler
- Copied to Backport #40510: mimic: [journal] tweak config defaults to improve small-IO performance added

#8 - 06/24/2019 10:06 AM - Nathan Cutler
- Copied to Backport #40511: nautilus: [journal] tweak config defaults to improve small-IO performance added

#9 - 11/05/2019 01:11 AM - Jason Dillaman
See additional unit test fixes: #42598

#10 - 01/27/2021 08:06 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".