bluestore - Bug #46780

BlueFS Spillover without db being full

07/30/2020 11:12 PM - Seena Fallah

Status:	Closed		% Done:		0%
Priority:	Normal				
Assignee:					
Category:					
Target version:					
Source:			**Affected Versions:**		v14.2.2, v14.2.3, v14.2.4, v14.2.5, v14.2.6, v14.2.7, v14.2.8, v14.2.9
Tags:			**ceph-qa-suite:**		
Backport:			**Pull request ID:**		
Regression:	No		**Crash signature (v1):**		
Severity:	2 - major		**Crash signature (v2):**		
Reviewed:					

Description

Hi. I'm facing a issue that my OSDs are going to spillover but db used that I check from prometheus `ceph_bluefs_db_used_bytes` is just 33GB and my db size is 191GB. The spillover amount is just 65MB and when I do a manual compaction it will resolved.
Sometimes it needs double manual compaction to be cleaned.
I see this occurred when flushing event is triggering in RocksDB logs
I have attached the osd log when spillover occurred.
After a time I do a manual compaction it will again have spillover with this 65MB value.
Why my db has free space but it can't use it?
I'm using nautilus 14.2.9 on ubuntu 18.04.4

History

#1 - 07/30/2020 11:19 PM - Seena Fallah

- File Screenshot from 2020-07-31 03-48-42.png added

- File Screenshot from 2020-07-31 03-48-30.png added

Here is also my db usage for that OSD

#2 - 07/30/2020 11:21 PM - Seena Fallah

I have made a mistack in description the ceph_bluefs_db_used_bytes is 27GB as you see in screen shots and also after while a compacted OSD can be use 40GB of db and doesn't have any spillover but this OSD with 27GB db used has spillover!

#3 - 08/07/2020 08:37 PM - Neha Ojha

- Project changed from RADOS to bluestore

#4 - 08/07/2020 08:39 PM - Josh Durgin

Igor confirmed this was an issue fixed by https://github.com/ceph/ceph/pull/33889 - L4 will go to the main device without this, which is why the spillover occurred.

#5 - 08/07/2020 08:43 PM - Seena Fallah

As I see in PR if I want to use this feature I should change `bluestore_volume_selection_policy` to `use_some_extra`. This option is in `LEVEL_DEV` as seen https://docs.ceph.com/docs/master/dev/config/#levels it's not suggested to use in production environment.
How can I use this PR to solve my problem?

#6 - 01/29/2021 07:41 AM - Konstantin Shalygin

- Status changed from New to Triaged

- Backport deleted (nautilus, octopus)

- Affected Versions v14.2.2, v14.2.3, v14.2.4, v14.2.5, v14.2.6, v14.2.7, v14.2.8 added

Seena, this fixed in 14.2.11, and default in 14.2.12

#7 - 06/08/2021 12:13 PM - Igor Fedotov

- Status changed from Triaged to Closed

Fixed with new bluefs space tracking framework (see #39185) starting v14.2.12

Files

ceph-osd.116.log	46.7 KB	07/30/2020		Seena Fallah
Screenshot from 2020-07-31 03-48-42.png	26.3 KB	07/30/2020		Seena Fallah
Screenshot from 2020-07-31 03-48-30.png	26.4 KB	07/30/2020		Seena Fallah