RADOS - Bug #38322

luminous: mons do not trim maps until restarted

02/14/2019 01:31 PM - Sage Weil

Status:	Closed	**% Done:**	0%
Priority:	High	**Spent time:**	0.00 hour
Assignee:	Joao Eduardo Luis		
Category:	Correctness/Safety		
Target version:			
Source:		**Affected Versions:**	
Tags:		**ceph-qa-suite:**	
Backport:		**Component(RADOS):**	Monitor
Regression:	No	**Pull request ID:**	19076
Severity:	3 - minor	**Crash signature:**	
Reviewed:			

Description	
Reported by several users, most recently at https://marc.info/?l=ceph-devel&m=154955388914036&w=2	

Related issues:	
Related to Ceph - Backport #45403: luminous: mon/OSDMonitor: maps not trimmed...	**Rejected**
Related to Ceph - Bug #45400: mon/OSDMonitor: maps not trimmed if osds are down	**Pending Backport**

History

#1 - 02/15/2019 10:53 AM - Dan van der Ster

Here's an example on v12.2.8. The mon db is normally trim at around 700MB, but after some backfilling its currently at 6.3GB on all mons.

```
cephbeesly-mon-2a00f134e5.cern.ch:
6.3G    /var/lib/ceph/mon/

p01001532021656.cern.ch:
6.3G    /var/lib/ceph/mon/

p05517715d82373.cern.ch:
6.3G    /var/lib/ceph/mon/

p05517715y01595.cern.ch:
6.3G    /var/lib/ceph/mon/

p05517715y58557.cern.ch:
6.3G    /var/lib/ceph/mon/
```

We follow Sage's procedure to debug:

```
- enable debug_mon = 20 on all mons (*before* restarting)
    ceph tell mon.* injectargs '--debug-mon 20'
- wait for 10 minutes or so to generate some logs
- add 'debug mon = 20' to ceph.conf (on mons only)
- restart the monitors
- wait for them to start trimming
- remove 'debug mon = 20' from ceph.conf (on mons only)
```

Following the mon restarts the db's shrunk to ~500MB:

```
cephbeesly-mon-2a00f134e5.cern.ch:
532M    /var/lib/ceph/mon

p01001532021656.cern.ch:
532M    /var/lib/ceph/mon

p05517715d82373.cern.ch:
532M    /var/lib/ceph/mon

p05517715y01595.cern.ch:
532M    /var/lib/ceph/mon

p05517715y58557.cern.ch:
532M    /var/lib/ceph/mon
```

The logs are here: ceph-post-file: 877cc29d-e697-4f76-9d52-70f08511cfca

#2 - 02/18/2019 02:30 PM - Joao Eduardo Luis

I have a feeling this is actually due to what led me to open this PR: https://github.com/ceph/ceph/pull/19076

The problem was that I was unable to reproduce it in mimic, so that was annoying enough to just let it linger. I'll get a branch with this for luminous, and hopefully someone will be able to test it.

#3 - 02/18/2019 02:31 PM - Joao Eduardo Luis

- Category set to Correctness/Safety

- Assignee set to Joao Eduardo Luis

- Component(RADOS) Monitor added

#4 - 03/04/2019 10:51 AM - Swami Reddy

seen this issue with 10.2.4

#5 - 08/21/2019 07:45 PM - Neha Ojha

- Status changed from Need More Info to Fix Under Review

- Pull request ID set to 19076

#6 - 07/31/2020 11:34 AM - Joao Eduardo Luis

- Status changed from Fix Under Review to Resolved

#7 - 07/31/2020 11:34 AM - Joao Eduardo Luis

- Status changed from Resolved to Closed

#8 - 07/31/2020 11:35 AM - Joao Eduardo Luis

- Related to Backport #45403: luminous: mon/OSDMonitor: maps not trimmed if osds are down added

#9 - 07/31/2020 11:35 AM - Joao Eduardo Luis

- Related to Bug #45400: mon/OSDMonitor: maps not trimmed if osds are down added