

RADOS - Bug #39398

osd: fast_info need update when pglog rewind

04/22/2019 08:43 AM - Zengran Zhang

Status:	Need Review	Start date:	04/22/2019
Priority:	High	Due date:	
Assignee:	Zengran Zhang	% Done:	0%
Category:		Estimated time:	0.00 hour
Target version:		Spent time:	0.00 hour
Source:		Affected Versions:	
Tags:		ceph-qa-suite:	
Backport:		Component(RADOS):	
Regression:	No	Pull request ID:	27698
Severity:	3 - minor	Crash signature:	
Reviewed:			

Description

When the pglog need rewind, the info.last_update will need to change to older value, current impl of PG::_prepare_write_info() will keep fast_info.last_update remain newer value, if osd crash after info be persisted, next reboot fast_info will cover the info with inconsistent last_update!

1. Reproducible Steps

```
MON=1 OSD=2 MDS=0 RGW=0 MGR=1 ../src/vstart.sh --new -x --localhost --bluestore --debug
```

```
bin/ceph osd pool create rep 1
bin/ceph osd pool set rep size 2
bin/ceph osd set norecover
```

```
bin/rados -p rep put test /etc/hosts \# 0'1
bin/ceph daemon osd.1 config set objectstore_blackhole true
bin/rados -p rep put test /etc/hosts \# 0'2 this put will hang because osd.1 set blackhole
```

```
killall ceph-osd
\# !!! cancel the hang put
```

```
bin/init-ceph start osd.1 \# osd.1 active with last_update = 0'1
\#wait pg active
bin/init-ceph start osd.0
\# at here
\# osd.0 pg_info.last_update will rewind 0'2 to 0'1
\# but remain a fast_info with last_update = 0'2 ( *** wrong place )
```

```
killall ceph-osd
```

```
bin/init-ceph start osd.0
```

```
\# check the pg_info.last_update on osd.0 when reload
\# because the fast_info with last_update = 0'2, so it recover the 0'1 in pg_info.last_update
\# 2019-04-20 14:16:34.400 7f46793a5ec0 20 bluestore.OmapIteratorImpl(0x5616d1d09870) valid is at
0x00000000000000418'._biginfo'
\# 2019-04-20 14:16:34.400 7f46793a5ec0 20 bluestore.OmapIteratorImpl(0x5616d1d09870) valid is at
0x00000000000000418'._epoch'
\# 2019-04-20 14:16:34.400 7f46793a5ec0 20 bluestore.OmapIteratorImpl(0x5616d1d09870) valid is at
0x00000000000000418'._fastinfo'
\# 2019-04-20 14:16:34.400 7f46793a5ec0 20 bluestore.OmapIteratorImpl(0x5616d1d09870) valid is at 0x00000000000000418'._info'
\# 2019-04-20 14:16:34.400 7f46793a5ec0 20 bluestore.OmapIteratorImpl(0x5616d1d09870) valid is at
```

```
0x0000000000000418'._infover'  
\# 2019-04-20 14:16:34.400 7f46793a5ec0 20 bluestore.OmaplteratorImpl(0x5616d1d09870) valid is at  
0x0000000000000418'.may_include_deletes_in_missing'  
\# 2019-04-20 14:16:34.400 7f46793a5ec0 20 bluestore.OmaplteratorImpl(0x5616d1d09870) valid is at  
0x0000000000000418'.missing/0000000000000001.5BAA8E04.head.test..  
\# 2019-04-20 14:16:34.404 7f46793a5ec0 10 read_log_and_missing done  
\# 2019-04-20 14:16:34.404 7f46793a5ec0 10 osd.0 pg_epoch: 17 pg[1.0( v 12'2 (0'0,12'2] local-lis/les=16/17 n=1 ec=9/9 lis/c 16/9  
les/c/f 17/10/0 16/16/14) [1,0] r=1 lpr=0 pi=[9,16)/1 crt=12'2 lcod 0'0 unknown m=1 mbc={}] handle_initialize
```

the reproduct log of osd.0 on my environment see the attachment I uploaded.

History

#1 - 04/22/2019 05:08 PM - Neha Ojha

- Status changed from New to Need Review
- Assignee set to Zengran Zhang
- Pull request ID set to 27698

#2 - 05/01/2019 09:03 PM - Neha Ojha

- Priority changed from Immediate to High

Files

osd.0.log	122 KB	04/22/2019	Zengran Zhang
-----------	--------	------------	---------------