

RADOS - Support #23455

osd: large number of inconsistent objects after recover or backfilling

03/24/2018 07:20 AM - Yao Ning

Status: Resolved	% Done: 0%
Priority: Normal	Spent time: 0.00 hour
Assignee:	
Category:	
Target version:	
Tags:	Component(RADOS): OSD
Reviewed:	Pull request ID:
Affected Versions: v10.2.10	

Description

large number of inconsistent objects after recover or backfilling.

reproduce method:

1) create rbd volume and, map to a client. use fio to write data to the volume, as below

```
[job]
ioengine=libaio
direct=1
filename=/dev/rbd0
rw=randwrite
norandommap
bs=8k
iodepth=2
random_distribution=zipf:1.2
time_based
runtime=360000
```

2) stop severel osds
3) create snapshot for this volume
4) and then start osds again
5) finally, do deep-scrub and the inconsistent occurs

like below:

```
host=$(( ${i} % 2 + 1 ))

ssh jewel-ceph-${host} "systemctl stop ceph-osd.target"
echo "stop osd on jewel-ceph-${host}"
sleep 60

rbd snap create nyao@${i}
sleep 60
rbd snap create nyao@s${i}
sleep 60
rbd snap create nyao@ss${i}
sleep 60
rbd snap create nyao@sss${i}
echo "create snap nyao@${i}"

sleep 120
ssh jewel-ceph-${host} "systemctl start ceph-osd.target"
echo "start osd on jewel-ceph-${host}"

for j in {0..5};
do
ceph osd deep-scrub osd.${j}
done
```

```
echo "scrub all osd"  
sleep 3600
```

it seems if I disable fiemap features, then no inconsistent object any more.

History

#1 - 03/24/2018 07:36 AM - Yao Ning

it seems quite similar with issue <http://tracker.ceph.com/issues/21388>

#2 - 03/24/2018 07:46 AM - Yao Ning

it is also affected for v10.2.5. And just affect all snap object, and no head object is affected

#3 - 03/24/2018 07:52 AM - Yao Ning

also affected head object, but very small number of portion.

#4 - 03/24/2018 09:32 AM - Yao Ning

- Assignee deleted (*Sage Weil*)
- Priority changed from *Urgent* to *Normal*

#5 - 04/09/2018 05:17 PM - Patrick Donnelly

- Project changed from *Ceph* to *RADOS*
- Subject changed from *large number of inconsistent objects after recover or backfilling to osd: large number of inconsistent objects after recover or backfilling*
- Category deleted (*OSD*)
- Source set to *Community (user)*
- Release deleted (*jewel*)
- Component(*RADOS*) OSD added

#6 - 04/09/2018 05:18 PM - Greg Farnum

- Tracker changed from *Bug* to *Support*
- Status changed from *New* to *Resolved*

fiemap is disabled by default precisely because there are a number of known bugs in the local filesystems across kernels that we can't do anything about. :(