

CEPH performance

- [BlueStore Config Reference: Sizing](#)
- https://yourcmc.ru/wiki/Ceph_performance
- [Ceph Performance Tuning Checklist](#)
- [New to Ceph, HDD pool is extremely slow](#)
- [Ceph Storage Performance](#)

Performance tips

Ceph is build for scale and works great in large clusters. In small cluster every node will be heavily loaded.

- adapt PG to number of OSDs to spread traffic evenly
- use krbd
- enable writeback on VMs (possible data loss on consumer SSDs)

performance on small cluster

- [Configuring Small Ceph Clusters for Optimal Performance - Josh Salomon, Red Hat](#)
- number of PG should be power of 2 (or middle between powers of 2)
- same utilization (% full) per device
- same number of PG per OSD := same number of request per device
- same number of primary PG per OSD = read operations spread evenly
 - primary PG - original/first PG - others are replicas. Primary PG is used for read.
- use relatively more PG than for big cluster - better balance, but handling PGs consumes resources (RAM)

balancer

```
ceph mgr module enable balancer
ceph balancer on
ceph balancer mode upmap
```

CRUSH reweight

If possible use balancer

Override default CRUSH assignment.

PG autoscaler

Better to use in warn mode, to do not put unexpected load when PG number will change.

```
ceph mgr module enable pg_autoscaler
#ceph osd pool set <pool> pg_autoscale_mode <mode>
ceph osd pool set rbd pg_autoscale_mode warn
```

It is possible to set desired/target size of pool. This prevents autoscaler to move data every time new data are stored.

check cluster balance

ceph -s ceph osd df - shows standard deviation

no tools to show primary PG balancing. Tool on

https://github.com/JoshSalomon/Cephalocon-2019/blob/master/pool_pgs_osd.sh

performance on slow HDDs

From:

<https://niziak.spox.org/wiki/> - **niziak.spox.org**

Permanent link:

<https://niziak.spox.org/wiki/vm:proxmox:ceph:performance>

Last update: **2024/05/17 19:33**

