

# Ceph Quarterly

Issue # 2 *An overview of the past three months of Ceph upstream development.* Oct. 2023

Pull request (PR) numbers are provided for many of the items in the list below. To see the PR associated with a list item, append the PR number to the string <https://github.com/ceph/ceph/pull/>. For example, to see the PR for the first item in the left column below, append the string 48038 to the string <https://github.com/ceph/ceph/pull/> to make this string: <https://github.com/ceph/ceph/pull/48038>.

## CephFS

1. A non-blocking I/O API for libcephfs has been added to Ceph: **48038**
2. A cause of potential deadlock in Python libcephfs has been fixed, which also affected the mgr modules using it: **52290**
3. MDS: acquisition throttles have been adjusted to more sensible defaults: **43546**
4. MDS: Buggy clients are now evicted in order to keep MDS available: **52944**

## Cephadm

1. Support for init containers has been added. Init containers allow custom actions to run before the daemon container starts: **52178**
2. We announce the deployment of the NVMe-oF gateway: **50423, 52691**
3. LV devices are now reported by ceph-volume in the inventory list, and can be prepared as OSDs: **52877**
4. cephadm is now split into multiple files in order to make it easier for humans to read and understand. A new procedure has been added to the documentation that describes how to acquire this new version of cephadm: **53052**

## Crimson

1. Support for multicore has been added to Crimson: **51147, 51770, 51916, 52306**
2. Infrastructure to support erasure coding has been added to Crimson: **52211**
3. We announce the introduction of a performance test suite for Crimson: **50458**

## Dashboard

1. RGW multisite configuration can now be imported from a secondary cluster or exported to a secondary cluster: **50706**
2. We announce several upgrades to the Cluster User Interface and the Cluster API: **52351, 52395, 52903, 52919, 52222, 53022**
3. More detail has been added to the RGW overview. This includes more granular information about daemons,

zoning, buckets, sers, used capacity (the capacity used by all the pools in the cluster). Cards detailing these assets have been added to the rgw overview dashboard: **52317, 52405, 52915**

4. It is now possible to manage CephFS subvolumes from the dashboard. This includes creating subvolumes, editing subvolumes, removing subvolumes, creating subvolume groups, editing subvolume groups, removing subvolume groups, removing subvolume groups with snapshots, and displaying subvolume groups in the CephFS subvolume tab: **52786, 52861, 52869, 52886, 52898, 53018, 53182, 53246**

## RADOS

1. Improved robustness against disk corruption of key data structures—OSD superblock data is now replicated in an onode's OMAP, which makes recovery possible even when onode data is corrupted: **50326**
2. We introduce a disk-fragmentation histogram for bluestore, accessible through an admin socket: **51820**
3. The amount of metadata necessary to create a snapshot has been significantly reduced, making snapshot creation and deletion more efficient: **53178**
4. ceph-mgr is now more resilient to blocklisting—cluster connections between MGRs and RADOS clients are now reopened when this occurs: **50291**
5. LZ4 compression can now be enabled for Bluestore RocksDB: **53343**

## RBD

1. Windows driver now runs RBD functional tests in CI: **50141**
2. The handling of blocklisting in rbd-mirror and krbd has been improved: **52990**, Linux kernel 6.5-rc5
3. Various rbd-mirror bug fixes: **52057, 52086, 53251**

## RGW

1. Support for versioned objects has been added to emergency bucket repair tools, and rgw-restore-bucket-index can now recognize versioned buckets and RGW objects that have names beginning with underscores: **51071**
2. Multisite sync load is now distributed fairly across gateways: **51493**
3. Several issues have been fixed, and Trino/TPCDS benchmark now runs with s3select: **52651**, <https://github.com/ceph/s3select/pull/132>
4. An initial version of a posix-fs backend for RGW has been developed: **52933**

## Telemetry

1. Over the past two months, Telemetry reports a storage capacity of 27PB from  $\approx 170$  Reef clusters. Approximately 2,800 clusters report  $\approx 1.2\text{EB}$ , including the 27PB from the Reef clusters. See [telemetry-public.ceph.com](https://telemetry-public.ceph.com).

## User+Dev Meeting Relunched

1. Neha Ojha and Laura Flores organized the relaunch of the User+Dev meeting, which was held on 22 Sep 2023. Two members of the Ceph upstream community gave presentations at this meeting. These were: Cory Snyder of 11:11 Systems's "What To Do When Ceph isn't Cephing" ([PDF link](#)) and Jonas Sterr of Thomas Krenn AG's "Ceph Usability Improvements" ([PDF link](#)). The minutes of this meeting are available here: <https://pad.ceph.com/p/ceph-user-dev-monthly-minutes>.
2. The User+Dev Meeting is held at 14:00 UTC on the third Thursday of every month.
3. <https://meet.jit.si/ceph-user-dev-monthly> is where the meeting is held.

CQ is a production of the Ceph Foundation. To support or join the Ceph Foundation, contact [membership@linuxfoundation.org](mailto:membership@linuxfoundation.org).

Send all inquiries and comments to Zac Dover at [zac.dover@proton.me](mailto:zac.dover@proton.me)