

Who's Using Ceph RBD Mirroring for Kubernetes Storage in Production?

Julien Laurenceau | Published 2025-12-15 | Updated 2026-10-01

<https://pepitedata.com/ceph-rbd-mirroring-kubernetes-production/>

PepiteData | LinkedIn: Julien Laurenceau | X: @Julien_Lau

RBD mirroring in production is one of the patterns I audit and design: [Ceph storage consulting](#).

Alpha Feature in Production... Good Idea?

The most promising approach, **journal-based mirroring**, offers near real-time replication and faster failover. However, it's currently an **alpha feature** in the Ceph CSI driver and relies on **rbd-nbd**, which introduces significant risks:

- If the CSI plugin pods restart, it behaves like a node reboot, and **all RBD mount points vanish**, disrupting workloads.
- It's more resource-intensive than the recommended **krbd**, consuming additional CPU and memory.

For documentation on Ceph RBD mirroring see [IBM documentation](#)

For details, see the [rbd-nbd design proposal](#).

Why Snapshot-Based Mirroring Is the Safer Option

Until journal-based mirroring matures, **snapshot-based mirroring** is the only viable option for production. It works by taking periodic snapshots and replicating them to a standby Ceph cluster. However, it has limitations:

- **Data loss risk:** Changes between snapshots are not captured.
- **Manual failover:** Requires reinstalling the Ceph driver, recreating PVs, and promoting standby volumes to master.

While stable, it's far from seamless. Learn more in the [RBD mirroring design proposal](#).

Do you really need RBD mirroring ?

With RBD mirroring the storage requirement is 6x replication for a two-zone design: 3 replicas in each zone with `min_size=2`. Before choosing mirroring, price that multiplier, check whether the application already replicates, and compare the 2.5-zone layout, which survives one zone loss at replication 4 with `min_size=2` inside a single cluster.

Another potent alternative when dealing with the constraint of 2 zones is to go for a 2.5 zones layout. A single ceph cluster spanning 3 zones for the control plane with limited resources footprint and no sensitive user data on this control plane combined with 2 zones for the storage plane. The advantage of this design is that it is then possible to sustain the loss of one zone by using `replication=4` and `min_size=2`.

What's Next?

The Ceph community is working on solutions like the [RBD Volume Healer](#), which could simplify failover and recovery.

What's Your Experience?

Are you using Ceph RBD mirroring in production? How are you handling its current limitations? Let's discuss.

<https://www.linkedin.com/>

Update 2026-09-29. This post is now part of [Ceph in Production: RBD, RGW and CephFS](#), the field guide to Ceph in production.

Related posts

- [Ceph in Production: RBD, RGW and CephFS](#)
- [Migrating unmanaged OSDs to managed OSDs in Ceph Squid](#)
- [Upgrading a Ceph multi-site S3 platform from Squid to Tentacle](#)
- [Choosing the right storage backend for Kubernetes PVCs](#)

PDF snapshot of the public article. The online version is authoritative; links remain clickable.