



Xinnor xiRAID and OSNexus QuantaStor: A Proven Joint Solution for Multi-Site Ceph Durability



When it comes to building resilient, multi-site Ceph clusters, most architects reach for the same answer: replica=3 spread across three sites. It's a sensible starting point but in practice, it leaves a dangerous gap that many organizations don't discover until they're unexpectedly suffering downtime.

Let's walk through the real math, the hidden risk, and how Xinnor's xiRAID fundamentally changes the durability equation to make metro-clusters with Ceph ultra-reliable.

The Three-Site Replica=3 Assumption

The conventional wisdom goes something like this: deploy three Ceph sites, set your pools to replica=3 with a CRUSH rule that pins one replica to each site, and you have a fault-tolerant metro-cluster. Lose any single site and you still have two copies of every object online. That sounds comfortable and under normal steady-state conditions, it is. The problem is what happens under real world scenarios which we'll get into next.

The Hidden Vulnerability: The Day After a Site Outage

When a site goes offline, your cluster loses all of the data device (OSDs) on that site all at once and all the pools in your cluster become degraded simultaneously. Ceph was design to handle just this kind of thing and because all writes are synchronous you can rest assured that the data is consistent across all sites. Yes, the dashboard will turn red, the cluster will be angry but it'll keep running.. as long as the OSDs on the remaining sites are all healthy.

Once a site is offline your effective replication factor has dropped from 3 to 2. That's still above the default min_size=2 threshold, so your pools stay online, writes and reads continue, no downtime. Most teams breathe a sigh of relief and start planning the recovery. But here's the scenario that keeps storage engineers up at night: **what if, while you're still running on the two remaining sites, a single NVMe OSD or HDD OSD on one of the two remaining sites fails? Or what if there was already a bad OSD on a remaining site when a site outage happens.**

This is not a hypothetical.. drive failures happen at random times and are indifferent to your incident response timeline. The moment you have a bad OSD on one of your remaining two sites you're down to one accessible copy of at least some portion of your data. With the default setting of min_size=2 enforced, Ceph will refuse I/O on any PG that drops below two available copies. Pools stop. Applications stall. What began as a manageable site outage has cascaded into a data availability crisis. One quick fix is to reduce the min

copies required to min_size=1 but that can lead to data consistency issues as there's no longer multiple copies to compare to for correctness.

Why This Is More Common Than You Think

The uncomfortable truth about multi-site cluster deployments (Ceph included) is that **compound failures are not rare events** — they're an expected consequence of scale and you need to design for them. Consider:

- A large cluster might have hundreds (or thousands) of OSDs across three sites
- The annual failure rate for NVMe devices, while low per device, multiplies quickly across a large fleet
- A site outage can last hours or days during which the surviving OSDs continue to age and accumulate wear
- Power events and environmental stresses that cause a site outage can also accelerate hardware degradation on adjacent sites

The gap between "surviving a site outage" and "surviving a site outage plus a single additional disk failure" is the difference between a well-designed system and a truly resilient one.

The xiRAID Solution: Durability at the OSD Layer

This is where Xinnor's xiRAID changes the picture entirely.

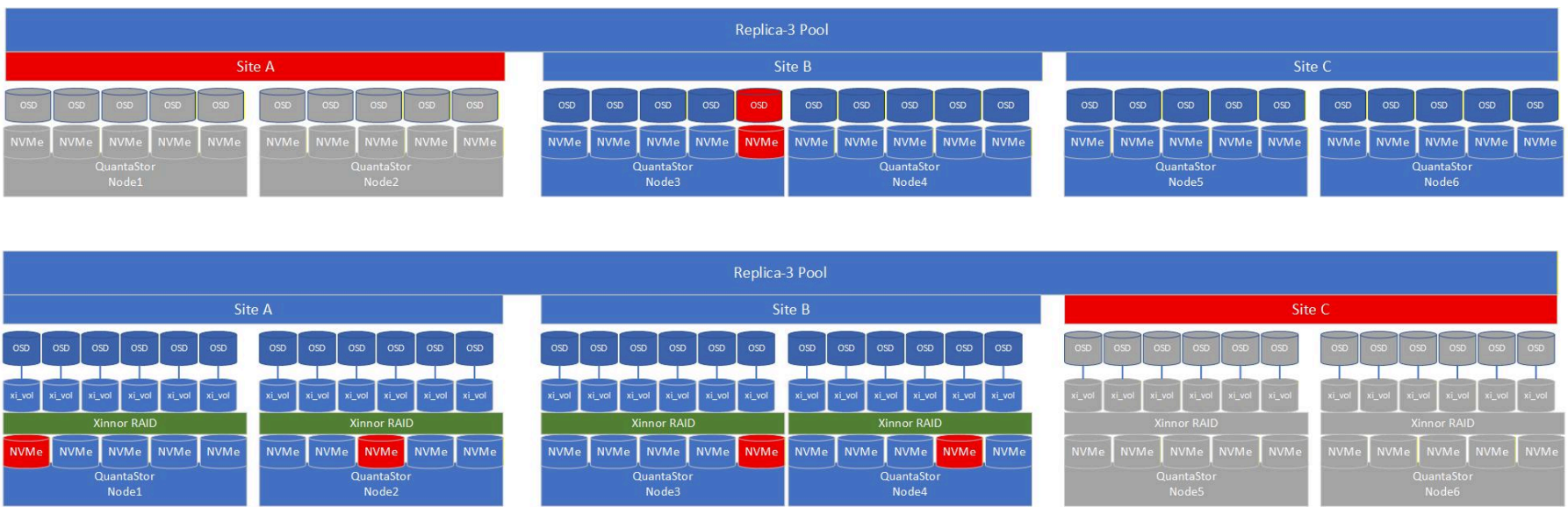
Rather than presenting individual NVMe or HDD devices directly to Ceph as OSDs, xiRAID aggregates the physical drives on each node into high-performance RAID logical devices. Ceph then consumes those logical devices as its OSDs.

The impact on durability is dramatic. A single xiRAID logical device configured with an appropriate RAID level (RAID 6, for example) can sustain multiple simultaneous physical drive failures within a node while the OSD remains fully online and healthy from Ceph's perspective.

Think about what this means in the failure scenario we described:

- **Site goes down** → Ceph drops to effective replica=2 ✓
- **NVMe fails on remaining node** → xiRAID absorbs the failure; OSD stays online ✓
- **Second NVMe fails on same node** → xiRAID absorbs that too ✓
- **Ceph cluster never sees a problem** ✓

The logical OSD that Ceph depends on never goes dark. The cascading failure chain that would normally lead from "site outage" to "pool I/O suspension" is eliminated.



Raising the Floor on Metro-Cluster Reliability

This architectural pattern — xiRAID-backed OSDs in a multi-site Ceph deployment — reframes how we think about durability tiers:

Scenario	Standard OSDs	xiRAID-Backed OSDs
All three sites healthy	Full redundancy	Full redundancy
One site offline	replica=2, one disk failure away from danger	OSD-level RAID absorbs disk failures
Site offline + 1 disk failure	Pool may suspend I/O	Cluster stays fully operational
Site offline + 2 disk failures	Data at risk	Still protected depending on RAID level

The baseline survivability of the cluster shifts upward across the board. Rather than designing for “tolerate a site loss,” you’re designing for “tolerate a site loss and continued hardware degradation on the surviving nodes.”

Performance Without Compromise

A natural concern with RAID at the OSD layer is performance. xiRAID is purpose-built to address this: it is a software-defined, kernel-level RAID solution engineered for modern NVMe hardware, capable of delivering line-rate throughput and low latency even across large RAID stripe sets. The overhead is minimal compared to the durability benefits, and because xiRAID operates below the Ceph OSD layer, it is completely transparent to the rest of the Ceph stack — no tuning changes, no CRUSH adjustments, no protocol modifications required.

Designing Truly Reliable Metro Clusters

A three-site Ceph cluster with replica=3 is a good architecture. A three-site Ceph cluster with replica=3 and xiRAID-backed OSDs is a great architecture and this same durability boost applies as much to Erasure Coded pools as it does for replica based pools.

Ceph + xiRAID accounts for the full probability space of real-world failure modes not just the clean, single-failure scenarios that CRUSH rule documentation tends to describe. It acknowledges that site outages don’t happen in isolation, that drives fail on their own schedule, and that the durability of a storage system should be measured not by its best day but by its worst.

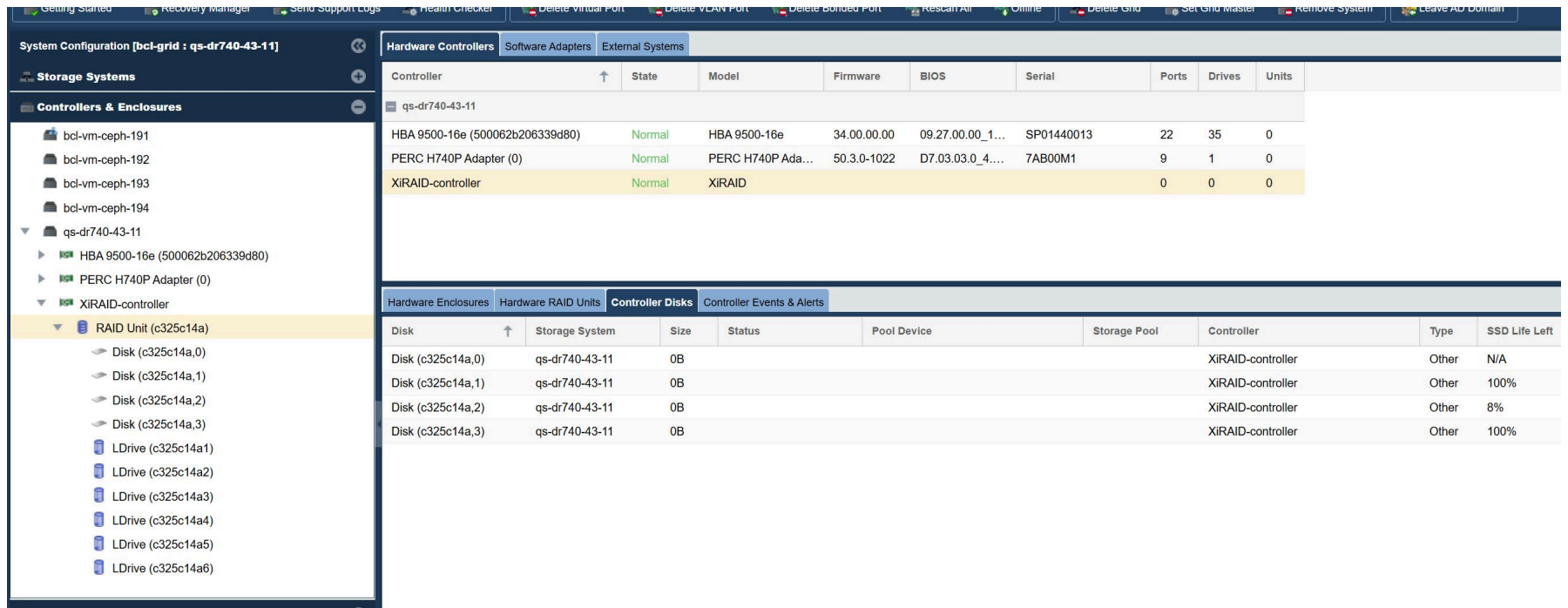
For organizations running mission-critical workloads across metro or regional Ceph deployments, xiRAID provides a straightforward path to a meaningfully higher tier of resilience without adding complexity to the Ceph layer itself.

QuantaStor 6.7: Fully Integrated xiRAID Management

With the QuantaStor 6.7 release (April 8, 2026), deploying and managing xiRAID-backed Ceph clusters has never been simpler. QuantaStor now provides fully integrated monitoring and management of xiRAID directly within the platform — no separate tooling, no context switching.

From the QuantaStor web UI, API, or CLI, you can:

- Create xiRAID RAID groups directly, without ever leaving the QuantaStor management plane
- Monitor RAID group health in real time alongside the rest of your storage infrastructure
- Receive proactive alerts when a RAID group requires maintenance, so drive replacements happen on your schedule rather than in response to a crisis



This integration means that all of the durability benefits described above, namely OSD-level fault absorption, compound failure protection, and metro-cluster resilience are all built into the QuantaStor grid architecture without having to do any additional tooling to support the use of xiRAID. The complexity of standing up and maintaining xiRAID is handled for you, leaving you with a fully integrated storage platform that simply works.

For organizations evaluating multi-site Ceph architectures, QuantaStor 6.7 with native xiRAID integration represents the most operationally straightforward path to enterprise-grade durability available today.