

Ceph Disaggregated Tiering Architecture with OpenFlex® Data24 4000 and Ultrastar® Data60 3000



Abstract

This architecture redefines infrastructure by decoupling compute and storage, replacing rigid HCI models with a flexible, tiered Ceph object storage platform that scales independently, reduces costs, and aligns resources with real workload demands.

July 2026

Table of Contents

Introduction	3
Problem Statement	3
Solution Highlights.....	4
Technology Overview.....	4
OpenFlex Data24 4000 Series Storage Platform	4
Ultrastar Data60 3000 Storage Platforms.....	5
Ceph Tentacle Object Storage	6
Ceph Tentacle Tiered Storage with OpenFlex Data24 4000 and Ultrastar Data60 3000.....	8
Ceph Installation Details	10
Ceph OSD Details	11
Ceph Pool Details.....	11
Ceph Storage class Tiring Details.....	12
Ceph Bucket Life Cycle Rule for Storage Tiring and Life Cycle Progress Status.....	12
How Lifecycle Works?	12
Lifecycle policies act as the intelligence layer of the tiering solution by:	12
Lifecycle Policy Constructs in Ceph	13
Storage Tier Mapping in this Architecture.....	13
Lifecycle Workflow Details	13
Conclusion	17

Introduction

Modern data growth, especially driven by analytics, AI/ML workloads, and unstructured data, is pushing traditional infrastructure beyond its limits. Hyperconverged Infrastructure (HCI), while simple and easy to deploy, couples compute and storage in a tightly integrated model that becomes inefficient at scale. Organizations often find themselves forced to scale compute resources simply to gain additional storage capacity, leading to suboptimal resource utilization, rising capital expenditures (CAPEX), and operational complexity.

This reference architecture proposes a disaggregated object storage solution using Ceph, combined with OpenFlex Data24 4000 (NVMe™-based hot tier) and Ultrastar Data60 (high-density cold tier). This architecture decouples compute from storage using lightweight servers and dedicated storage enclosures. It enables independent scaling, optimized performance, and tier-aware data placement.

Customers transitioning from HCI to this architecture gain:

- **Efficiency:** Scale storage and compute independently.
- **Performance:** NVMe-based OpenFlex tier for high IOPS and low latency workloads.
- **Cost Optimization:** High-density Data60 for economical bulk storage.
- **Resilience:** Distributed storage with no single point of failure.
- **Flexibility:** Policy-based tiering to automatically place data where it belongs.

The business value lies in transforming storage from a rigid, expensive resource into a flexible, scalable, and cost-efficient service aligned with workload requirements.

Organizations can modernize their infrastructure to effectively support both latency-sensitive and data-intensive workloads by adopting a tiered, disaggregated storage approach. By combining Ceph object storage with the high-performance capabilities of OpenFlex Data24's composable NVMe architecture, this solution enables efficient utilization of NVMe performance while maintaining operational simplicity, flexibility, and broad ecosystem compatibility. The OpenFlex Data24 4000 series, delivering up to ~1474TB¹ of low-latency dual-port NVMe SSD capacity in a dense 2U, 24-bay form factor, extends NVMe performance to shared storage environments using NVMe-over-Fabrics (NVMe-oF™). Powered by Western Digital's RapidFlex™ A2000 Fabric Bridge, it provides up to 12x100GbE connectivity with support for both RDMA and TCP transports, ensuring high throughput and low latency access from multiple hosts. Complementing the hot tier, the Ultrastar Data60 platform serves as a highly efficient cold storage layer, offering up to 60 drives in a 4U enclosure designed for high-density, cost-optimized capacity. This makes it ideal for storing large volumes of infrequently accessed or archival data. Together, OpenFlex Data24 4000 and Ultrastar Data60 3000 enable a balanced architecture where performance-critical data resides on NVMe flash while colder data is seamlessly tiered to high-capacity HDD storage, delivering an optimal mix of performance, scalability, and total cost efficiency.

By leveraging Ceph object storage with OpenFlex Data24 4000 and Ultrastar Data60 3000, enterprises can adopt a future ready storage architecture that balances performance, cost efficiency, and ease of deployment—positioning their infrastructure to meet evolving application and business demands.

Problem Statement

Enterprises today face significant challenges managing rapidly growing volumes of unstructured data. Traditional HCI environments introduce inefficiencies such as overprovisioning compute resources, limited scalability, and inability to differentiate storage tiers effectively. Additionally, performance-sensitive workloads compete with archival data for the same infrastructure, resulting in unpredictable performance and rising costs.

A disaggregated Ceph-based tiered storage architecture with OpenFlex Data24 4000 and Ultrastar Data60 3000 enables efficient resource utilization, predictable performance, and cost optimization by aligning data placement with workload requirements.

¹ One terabyte (TB) is equal to one trillion bytes and one petabyte (PB) is equal to 1,000 TB. Actual user capacity may be less due to operating environment.

Solution Highlights

The Modern storage environments require intelligent data placement to balance performance and cost. This solution leverages Ceph Tentacle's tiering capabilities to distribute data across high-performance and capacity-optimized storage platforms.

The architecture introduces a tiered storage model:

Hot Tier: Manages latency-sensitive workloads such as analytics, AI/ML, and transactional operations.

Cold Tier: Stores infrequently accessed data, archives, and backups at lower cost.

The design leverages:

- Ceph's distributed, software defined architecture.
- OpenFlex's NVMe based, disaggregated storage fabric.
- Ultrastar Data60 3000 is a high-density, hybrid HDD-based storage platform.

Key Components:

- **OpenFlex Data24 4000 (Hot Tier)**

Provides ultra-low latency NVMe storage

Used for active datasets requiring high throughput

- **Ultrastar Data60 3000 (Cold Tier)**

Provides dense, cost-efficient HDD storage

Ideal for archival and infrequently accessed data

- **Benefits of Tiering**

Improved performance for critical workloads

Reduced storage cost through intelligent data placement

Automated movement of data between tiers

Optimized resource utilization across the infrastructure

Enhanced scalability without linear cost increase

Technology Overview

OpenFlex Data24 4000 Series Storage Platform

Western Digital's OpenFlex Data24 4200 series NVMe-oF storage platform extends the high performance of NVMe flash to shared storage. It provides low latency sharing of NVMe SSDs over a high-performance Ethernet fabric to deliver similar performance to locally attached NVMe SSDs. Industry-leading connectivity, using Western Digital RapidFlex network adapters, allows up to 12 hosts to be attached without a switch.

OpenFlex Data24 4200 series uses Western Digital's RapidFlex A2000 Fabric Bridge devices to provide 12-ports of 100GbE which can connect to RDMA and/or TCP configured host ports. While RoCE (RDMA over Converged Ethernet) connections have historically been preferred in data centers, TCP offers greater ease-of-use and is sometimes preferred. OpenFlex Data24 4200 series offers the flexibility of connecting to either RoCE or TCP host ports for optimum usage. OpenFlex Data24 4200 series enables PCIe® Gen4 performance throughout the chassis, bringing the full performance capability of each SSD to the Ethernet fabric. PCIe Gen4 SSDs from multiple brands are supported.

NVMe-over-Fabrics, or NVMe-oF, is a networked storage protocol that allows storage to be disaggregated from compute to make that storage widely available to multiple applications and servers. By enabling applications to share a common pool of storage capacity, data can be easily shared between applications or needed capacity can be allocated to an application to respond to application needs. OpenFlex Data24 4200 series NVMe-oF storage platform can also be used as a disaggregated storage resource in an open composable infrastructure environment using the Open Composable API. OpenFlex Data24 4200 series is built to deliver high availability and enterprise-class reliability. The entire platform, including SSDs, is backed with a 5-year limited warranty.



Specification	Value
Form Factor	2U
Front Drive bays	Up to 24 x U.2 NVMe SSDs
Power Supply	2x 800W Titanium 100-240VAC, CRPS, Hot Plug
Max raw data storage capacity	1474 TB (24 x 15.36TB NVMe SSDs)
Fabric Adapter Slots	12x 100 GbE ports
Fabric Adapter(s)	Western Digital RapidFlex A2000 NVMe-oF Fabric Bridge ASICs
Cabling Passive	Cabling Passive (1 - 5m) and Active Optical (5m - 20m)
HA Redundancy	Dual IOMs, PSUs, and N+2 fans

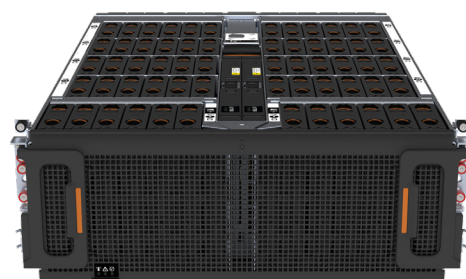
Features

- NVMe-over-Fabrics (NVMe-oF) architecture
- Bandwidth match between PCIe Gen4 SSDs (Storage) and I/O (Network) – No oversubscription
- RESTful API support for simplified management
- N+2 fan redundancy eliminates the need for hot-swappable fan modules
- Connect directly to an Ethernet NIC in your host or through an Ethernet switch
- Device Sharing optional configuration utilizes NVMe Non-Transparent Bridging to allow all Ethernet ports to access any SSD
- High IOPS and throughput performance
- Software-defined composability
- Supports Ethernet-based disaggregated storage
- High availability with redundant components
- Ideal for AI/ML, real-time analytics, and transactional workload

For more details regarding OpenFlex Data24 4000 please connect with enterprisesupport@wdc.com.

Ultrastar Data60 3000 Storage Platforms

The Ultrastar Data60 3000 Hybrid Storage Platform redefines what is possible in modern storage infrastructure. As a foundational building block for disaggregated and software-defined storage (SDS) architectures, it delivers massive capacity without sacrificing flexibility — empowering organizations to strike the right balance between performance and cost across any workloads. Packed with up to 1.92PB of raw storage in a compact, efficient form factor, the Data60 3000 is purpose-built for the demands of scalable data-intensive environments. Western Digital HelioSeal® drives ensure cool running, quiet operation, and high reliability. This platform addresses the demanding storage needs of large enterprise customers, storage OEMs, cloud service providers and resellers/integrators that require dense, shared HDD storage. The Ultrastar Data60 3000 Hybrid Storage Platform provides flexibility to specify the HDD combinations to balance capacity, performance and cost.



Developing storage devices and platforms side-by-side, we address these challenges through Silicon to Systems

design, a set of technologies developed based on a holistic view of devices, platforms, and their interactions. The first two of these innovative technologies are IsoVibe™ and ArcticFlow™.

IsoVibe reduces vibration-induced performance degradation, while ArcticFlow mitigates cooling issues by introducing cool air into the middle of the platform. Both these technologies contribute to long-term reliability, enabling our five-year limited warranty on the entire platform.

Specification	Value
Max Drives	60 x 3.5in drive bays
Drive Interface	12Gb/s SAS 6Gb/s SATA
Available Drive Capacities	HDD up to 26TB CMR
Host Interface	Dual redundant HEMs, 6 Mini-SAS HD ports per HEM
Power	Titanium: Dual 200-240V AC, 1600W
LED Indicators	Front/Rear: Power, ID, Fault Drive: Activity, Fault
Cooling	4 main enclosure fans front-to-rear cooling with zero-loss backflow prevention 1 IO module fan Dual PSUs with built-in fans
Serviceability	Cable-free hot-swappable IOM, HEM, power supply, fans, and drives

Features

- Up to 60 Ultrastar HDDs (SAS or SATA)
- Up to 1.92PB1 of raw storage² in 4U
- Up to 4 units may be daisy-chained for a total raw capacity of 7.68PB
- Patented IsoVibe technology ensures maximum performance even in heavy workloads
- Improved cooling from innovative ArcticFlow technology
- Together, IsoVibe and ArcticFlow contribute to a 62% reduction in drive returns²
- Choose dual-port SAS for high availability or single-port SATA for lower cost
- Up to 12 x 24Gb/s SAS-4 host connections
- Enterprise-grade redundant and hot-swappable drives, PSUs, IO Modules, HEMs, and fans
- Rack-mounted top cover for quick and easy service

For more details regarding OpenFlex Data24 4000 please connect with enterprisesupport@wdc.com.

Ceph Tentacle Object Storage

Ceph Object Storage provides a highly scalable, software-defined storage platform that delivers S3-compatible object storage capabilities. Ceph offers object storage tiering capabilities to optimize cost and performance by seamlessly moving data between storage classes. These tiers can be configured locally within an on-premises infrastructure or extended to include cloud-based storage classes, providing a flexible and scalable solution for diverse workloads. With policy-based automation, administrators can define lifecycle policies to migrate data between high-performance storage and cost-effective archival tiers, ensuring the right balance of speed, durability, and cost-efficiency. Ceph Tentacle introduces performance and efficiency improvements, particularly for erasure coded workloads.

² Based on observed and projected drive return data compared to prior generation product without features. Does not change product specifications or alter product warranty

Key Capabilities

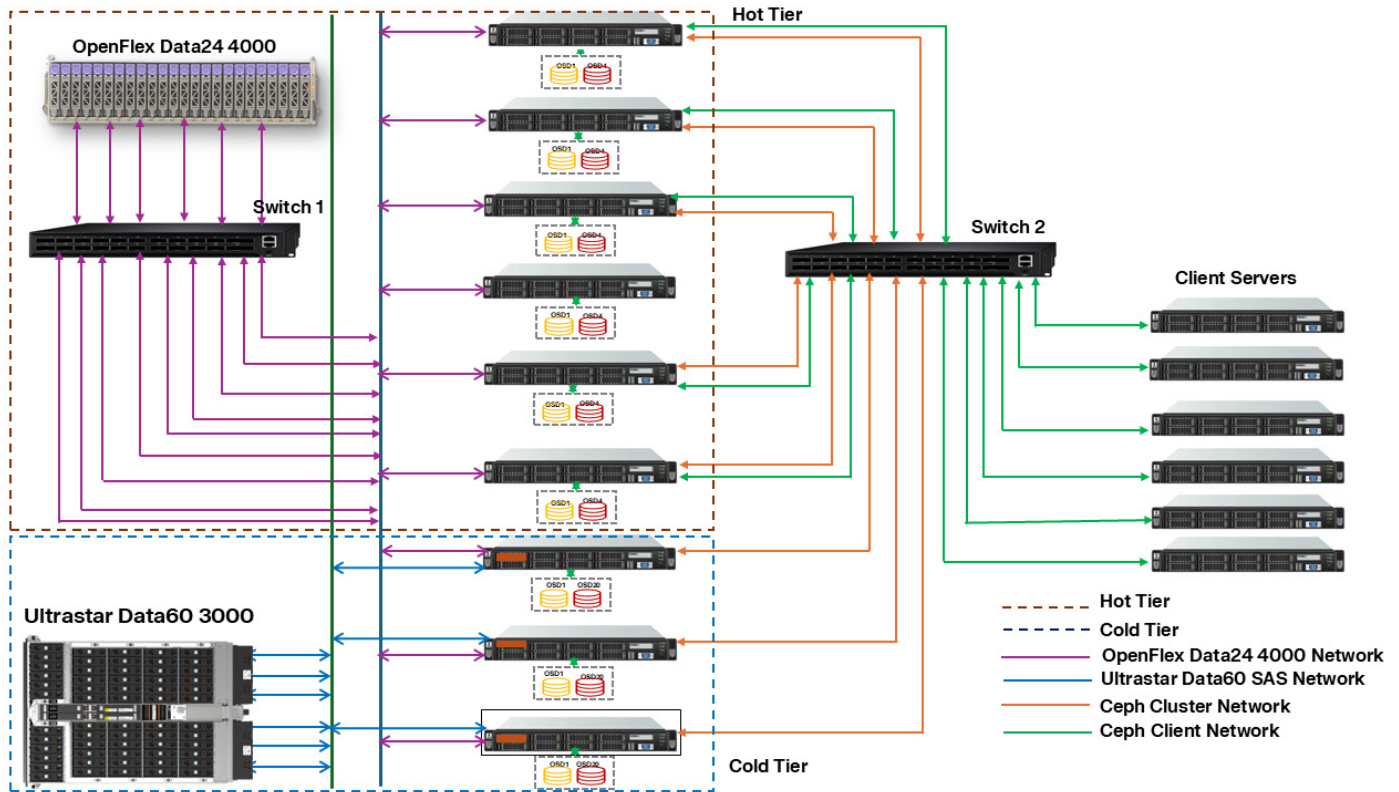
- **S3/Swift-compatible APIs** via RADOS Gateway (RGW)
- **CRUSH-based data placement** – Eliminates centralized metadata and enables deterministic, scalable, and highly resilient data distribution.
- **Massive horizontal scalability** – Scales from a few nodes to exabytes of object data without architectural changes.
- **Built-in data protection** – Offers flexible data protection schemes to balance durability, performance, and cost with replication and erasure coding.
- **Horizontal scalability** – Offers horizontal scalability with no single point of failure
- **Self-healing and self-managing** – Automatically detect failures and rebalances data with minimal administrative effort.
- **Native S3 and Swift APIs** – Provides industry standard object interfaces for seamless application integration.
- **Strong consistency model** – Ensures immediate data consistency, unlike many eventually consistent object stores.
- **Hardware and vendor agnostic** – Runs on commodity hardware and supports diverse storage media (HDD, SSD, NVMe).
- **Multi-tenancy with quotas and isolation** – Enables secure sharing of object storage across teams or workloads.
- **Unified storage platform** – Object storage coexists with block and file storage under the same Ceph cluster.
- **Open-source and cloud-native** – Backed by a strong community, avoiding vendor lock in and supporting hybrid or private clouds. Ceph object storage provides S3 compatible access with strong consistency, self healing, and horizontal scalability.
- **Tentacle Innovations** – Offers Faster BlueStore WAL and improved compression, Optimized RGW bucket listing via faster OMAP iteration, Enhanced dashboard and Mult-cluster management, Performance improvements for EC (FastEC).
- **Multi-site replication and disaster recovery support** – Enables automatic data replication across geographically distributed sites to ensure business continuity and rapid recovery in case of failures or disasters.
- **Advanced lifecycle management for automated data tiering** – Automates data movement between storage tiers based on access patterns and policies, optimizing performance and cost without manual intervention

Highlighted Features

- Bucket lifecycle management (LC) rules
- Multisite replication & disaster recovery
- Dynamic rebalancing
- QoS and multi-tenancy support
- End-to-end check sum for data integrity

Ceph Tentacle Tiered Storage with OpenFlex Data24 4000 and Ultrastar Data60 3000

The solution architecture integrates Ceph Tentacle object storage on industry-standard servers with OpenFlex Data24 4000 and Ultrastar Data60 3000 platforms, enabling a high-performance, cost-efficient tiered storage environment. In this design, OpenFlex Data24 4000 acts as the hot tier, storing and serving frequently accessed, latency-sensitive data, while Ultrastar Data60 3000 functions as the cold tier, hosting older or infrequently accessed data in a dense and cost-efficient manner. The Ceph platform intelligently manages data placement and movement across these tiers, ensuring transparency and operational simplicity.



This architecture leverages Ceph Tentacle's advanced tiering capabilities to seamlessly distribute data across high-performance NVMe storage and high-capacity HDD-based storage. Ceph OSDs are deployed following the Tentacle architecture, enabling independent scaling of compute and storage resources. OpenFlex Data24 provides shared NVMe storage, accessed over low-latency fabrics such as NVMe-over-Fabrics (RoCEv2/TCP), delivering exceptional performance for active workloads. In parallel, Ultrastar Data60 offers high-density storage enclosures, designed to deliver scalable, resilient, and cost-optimized capacity for colder datasets.

At its core, Ceph operates as a distributed, software-defined storage platform built on RADOS, where Object Storage Daemons (OSDs) manage data across both tiers. Within the Ultrastar Data60 environment, OSDs primarily reside on high-capacity HDDs, optionally enhanced with SSDs for BlueStore WAL/DB acceleration, enabling a hybrid storage approach that combines flash performance for metadata and hot data with economical disk storage for bulk data.

Ceph MON and MGR services run on the storage nodes to maintain cluster state, orchestration, and health monitoring, while RADOS Gateway (RGW) provides S3- and Swift-compatible APIs for seamless cloud-native application access. The architecture separates client (public) and cluster (replication) networks, ensuring optimal performance, fault isolation, and efficient data movement. Data placement is governed by the CRUSH algorithm, enabling intelligent, decentralized distribution without bottlenecks, while ensuring balanced utilization across the cluster and tiers.

OpenFlex Data24 provides shared NVMe drives accessed over a low latency fabric such as NVMe-oF (RoCEv2/TCP). This separation enables granular resource scaling, improves utilization, and removes local disk constraints from OSD nodes. The result is a highly resilient, cloud native object storage solution that delivers consistent performance, simplified operations, and enterprise grade durability, well suited for demanding workloads such as analytics. Each Ultrastar Data60 3000 enclosure contributes dense, high-capacity storage nodes, supporting petabyte-to-exabyte scale growth with linear scalability. The result is a highly resilient, cloud-native object storage solution that delivers consistent performance, enterprise-grade durability, and simplified operations with backup and tiered storage. This tiered approach is ideally suited for data-intensive workloads such as analytics platforms, AI/ML data lakes, backup repositories, and large-scale private cloud environments, where both performance and cost optimization are critical.



A production-grade deployment includes the following:

- Multiple OpenFlex Data24 enclosures (Hot tier): It provides High-performance, latency-sensitive workloads.
- Multiple Ultrastar Data60 enclosures (Cold tier): It provides Capacity-oriented, colder workloads
- Dedicated Ceph compute nodes (Servers running MON, MGR, RGW, and OSD services): It provides unified object storage layer with intelligent data distribution

The complete architecture eliminates Single Point of Failure by:

Distributing data across

- Multiple nodes
- Multiple drives
- Multiple enclosures

Using Ceph CRUSH (Controlled Replication Under Scalable Hashing)

- Data is automatically replicated across failure domains (rack, chassis, enclosure).
- Enclosure-level failures do not impact data availability.
- Recovery is automatic and self-healing.

Configuration Details

Specification	Value
Storage Product	OpenFlex Data24 4213 and Ultrastar Data60 3000
Ceph Version	Ceph Tentacle version 20.2.1
OpenFlex Storage Interface	12 x QSFP28 Interface (12 x 100Gbps per port)
Ultrastar Storage Interface	SATA
HBA Details	eHBA 9600-16e Tri-Mode Storage Adapter, FW: 8.17.1.0-00000-00001, Driver: 8.17.1.0.0
Server Details	PowerEdge® R550
Host OS	Ubuntu 22.04.05
Host NIC	CX6
Firmware Version	24.10-4.1.4 (22.32.2004 (MT_00000000437))
CPU	Intel® Xeon® Gold 5318Y CPU @ 2.10GHz or Higher CPU Configuration
CPU Core Details	Dual socket server with 24 core CPU each. 96 logical cores in total with HT enabled
Memory	256 GiB
Switch Details	NVIDIA® Onyx® 3700
OpenFlex Drive Details	Ultrastar SN655 3.84TB, WUS5EA138ESP7E4, RC610008
Ultrastar Data60 3000 Drive Details	Sandisk DC HC560 20TB CMR SATA Drives, WUH722020BLE604, PQGSWGC8
Local Drive Details connected to UD60 Storage servers	6x Sandisk SN655 7.68TB, WUS4BA176DSP3X4, R2210801 – 1 NVMe drive connected to each Storage server
Tool	Warp v1.4.0

Ceph Installation Details

There are multiple ways to install Ceph. We have used Cephadm tool for the current configuration.

Cephadm is a tool that can be used to install and manage a Ceph cluster.

- Data is automatically replicated across failure domains (rack, chassis, enclosure).
- Enclosure-level failures do not impact data availability.
- Recovery is automatic and self-healing.

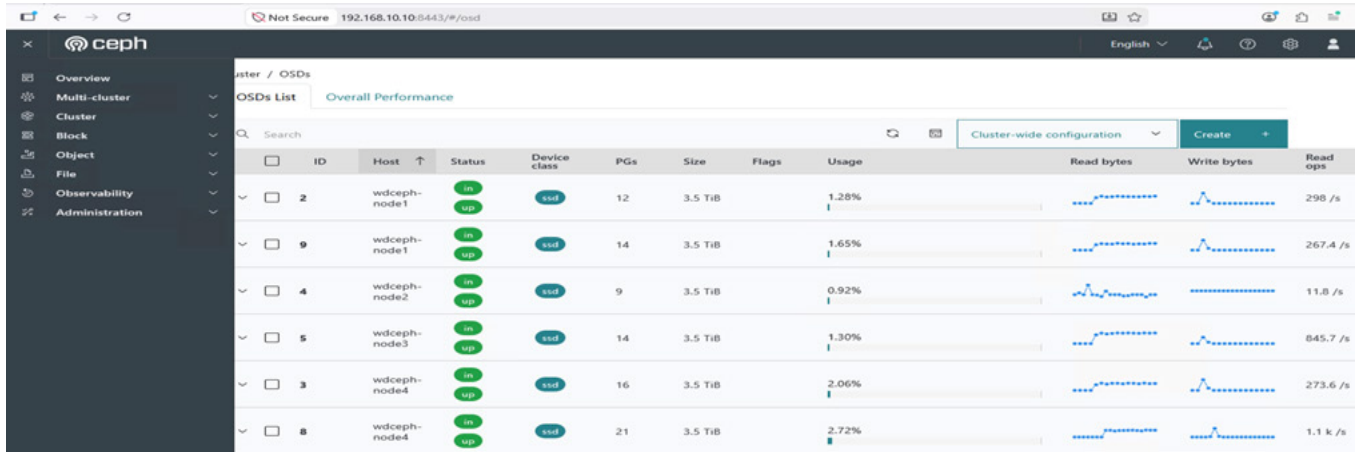
Cephadm creates a new Ceph cluster by bootstrapping a single host, expanding the cluster to encompass any additional hosts, and then deploying the needed services. Refer to [Cephadm](#) for more details.

Ceph Nodes

Ceph nodes are the physical or virtual servers in a Ceph cluster that run storage and control services such as MON, MGR, OSD, and RGW forming the distributed storage system.

Ceph OSD Details

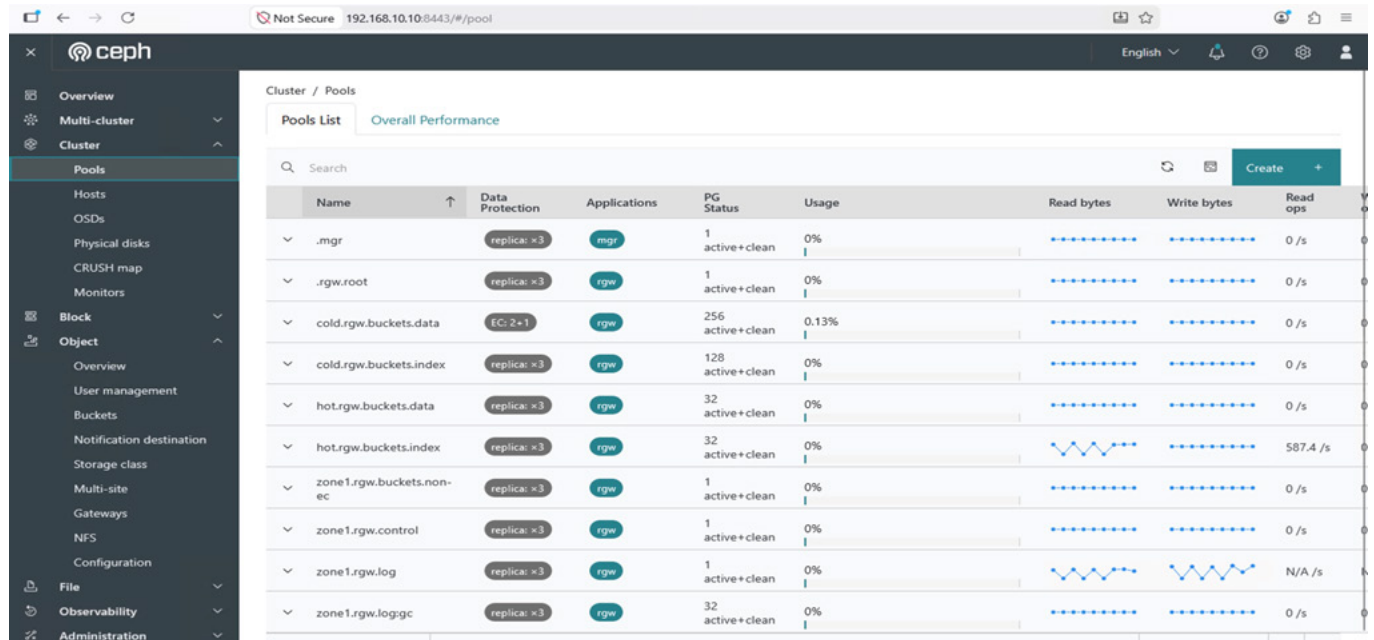
OSDs are the core data-handling components in Ceph that store, replicate, and recover data on disks while ensuring durability, balance, and fault tolerance across the cluster.



ID	Host	Status	Device class	PGs	Size	Flags	Usage	Read bytes	Write bytes	Read ops
2	wdceph-node1	up	ssd	12	3.5 TiB		1.28%			298 /s
9	wdceph-node1	up	ssd	14	3.5 TiB		1.65%			267.4 /s
4	wdceph-node2	up	ssd	9	3.5 TiB		0.92%			11.8 /s
5	wdceph-node3	up	ssd	14	3.5 TiB		1.30%			845.7 /s
3	wdceph-node4	up	ssd	16	3.5 TiB		2.06%			273.6 /s
8	wdceph-node4	up	ssd	21	3.5 TiB		2.72%			1.1 k /s

Ceph Pool Details

Ceph pools are logical groupings of storage within a Ceph cluster where data objects are stored, each defined with specific replication or erasure coding policies for performance and durability. They act as the core abstraction layer that maps application data to underlying storage devices while enabling tiering and fault isolation.



Name	Data Protection	Applications	PG Status	Usage	Read bytes	Write bytes	Read ops
.mgr	replica: x3	mgr	1 active+clean	0%			0 /s
.rgw.root	replica: x3	rgw	1 active+clean	0%			0 /s
cold.rgw.buckets.data	EC: 2+1	rgw	256 active+clean	0.13%			0 /s
cold.rgw.buckets.index	replica: x3	rgw	128 active+clean	0%			0 /s
hot.rgw.buckets.data	replica: x3	rgw	32 active+clean	0%			0 /s
hot.rgw.buckets.index	replica: x3	rgw	32 active+clean	0%			587.4 /s
zone1.rgw.buckets.non-ec	replica: x3	rgw	1 active+clean	0%			0 /s
zone1.rgw.control	replica: x3	rgw	1 active+clean	0%			0 /s
zone1.rgw.log	replica: x3	rgw	1 active+clean	0%			N/A /s
zone1.rgw.log.gc	replica: x3	rgw	32 active+clean	0%			0 /s

```

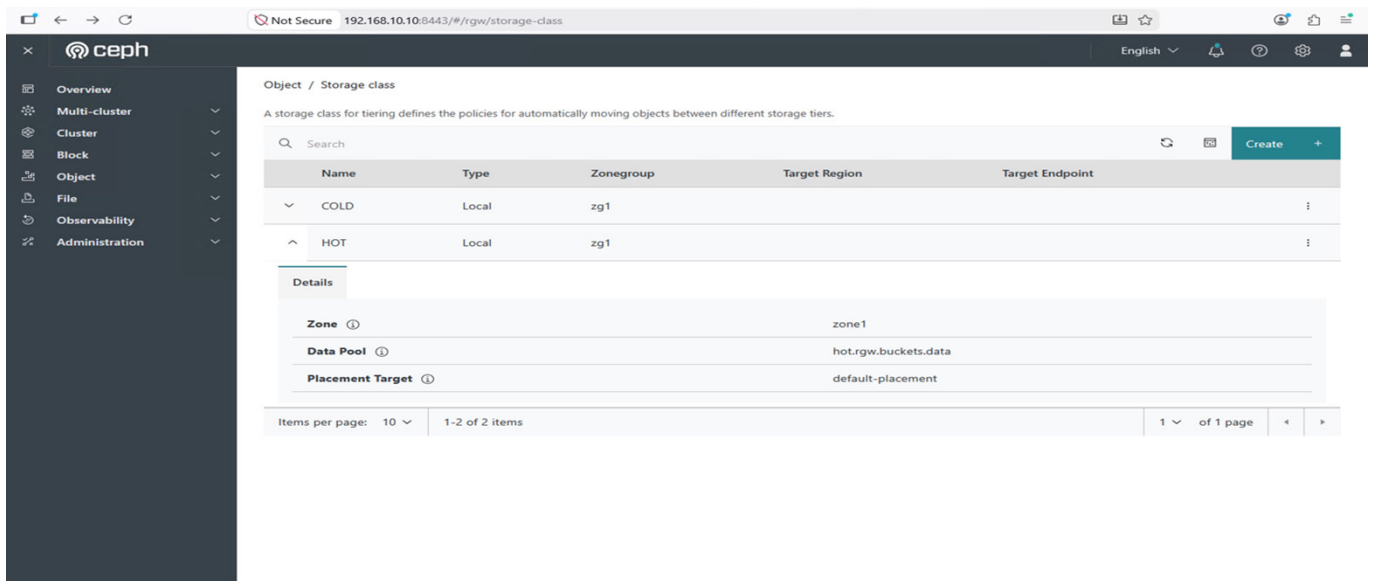
root@wdceph-node1:~#
root@wdceph-node1:~# ceph df
--- RAW STORAGE ---
CLASS      SIZE      AVAIL    USED    RAW USED    %RAW USED
hdd        1.1 PiB   1.1 PiB   13 TiB   13 TiB       1.22
ssd        84 TiB   79 TiB   4.6 TiB   4.6 TiB       5.45
TOTAL      1.1 PiB   1.1 PiB   18 TiB   18 TiB       1.52

--- POOLS ---
POOL ID PGS STORED OBJECTS USED %USED MAX AVAIL
.mgr 1 1 36 MiB 10 108 MiB 0 350 TiB
hot.rgw.buckets.data 2 1024 1.4 TiB 1.49M 4.3 TiB 5.44 25 TiB
hot.rgw.buckets.index 3 512 197 MiB 66 590 MiB 0 25 TiB
cold.rgw.buckets.data 6 256 1.2 TiB 1.04M 1.8 TiB 0.17 699 TiB
cold.rgw.buckets.index 7 128 1.0 GiB 99 3.0 GiB 0 341 TiB
.rgw.root 8 1 8.2 KiB 20 228 KiB 0 350 TiB
zone1.rgw.log 9 1 1016 KiB 244 3.7 MiB 0 350 TiB
zone1.rgw.control 10 1 2.2 KiB 9 6.5 KiB 0 350 TiB
zone1.rgw.meta 11 1 138 KiB 36 793 KiB 0 350 TiB
zone1.rgw.log:lc 12 32 21 KiB 0 63 KiB 0 350 TiB
zone1.rgw.log:gc 13 32 27 KiB 0 80 KiB 0 350 TiB
zone1.rgw.log:intent 14 32 34 KiB 0 102 KiB 0 350 TiB
zone1.rgw.log:usage 15 32 38 KiB 0 114 KiB 0 350 TiB
zone1.rgw.log:reshard 16 32 47 KiB 0 141 KiB 0 350 TiB
zone1.rgw.buckets.non-ec 17 1 4.4 KiB 0 13 KiB 0 350 TiB
root@wdceph-node1:~#

```

Ceph Storage class Tiring Details

Ceph storage classes define different data tiers (like HOT, COLD) mapped to specific storage pools, and lifecycle rules automatically move objects over time based on policies, enabling cost and performance optimization without manual intervention. Here OpenFlex Data24 4000 drives are configured as part of the hot tier pool and Ultrastar Data60 3000 HDDs are configured as part of the cold tier pool.



Object / Storage class

A storage class for tiering defines the policies for automatically moving objects between different storage tiers.

Name	Type	Zonegroup	Target Region	Target Endpoint
COLD	Local	zg1		
HOT	Local	zg1		

Details

Zone: zone1

Data Pool: hot.rgw.buckets.data

Placement Target: default-placement

Items per page: 10 | 1-2 of 2 items | 1 of 1 page

Ceph Bucket Life Cycle Rule for Storage Tiring and Life Cycle Progress Status

Ceph supports advanced lifecycle policies such as:

- Automatic data tiering (hot to cold)
- Expiration and archival
- Versioning and retention policies

How Lifecycle Works?

Ceph Object Storage (via RADOS Gateway – RGW) provides a policy-driven lifecycle management framework that automates how objects are stored, migrated, and retired across storage tiers. These lifecycle policies operate at the bucket level and are fully compatible with S3 lifecycle constructs, enabling seamless integration with cloud-native applications.

Lifecycle policies act as the intelligence layer of the tiering solution by:

- Monitoring object age, access patterns, and state
- Automatically transitioning data between tiers
- Enforcing retention, compliance, and deletion rules

This ensures that data placement aligns continuously with business value and usage patterns, eliminating manual intervention.

Lifecycle Policy Constructs in Ceph

Lifecycle rules are defined using a combination of conditions and actions:

1. Filters (Conditions)

Policies can be applied based on:

- Prefix (e.g., /logs/, /images/)
- Object tags (e.g., hot=true, archive=true)
- Object size thresholds (small vs large objects)
- Creation time / last modified time

2. Actions

Lifecycle rules support the following operations:

- Transition (Tier Migration): Moves objects between storage classes / pools:
Example: From hot pool (OpenFlex Data24 4000) → cold pool (Ultrastar Data60 3000)
- Can be time-based or policy-triggered
- Expiration (Deletion)
Deletes objects after a defined retention period
- Non-current Version Expiration
Applies when versioning is enabled
Removes older versions after a specified duration
- Abort Incomplete Uploads
Cleans up failed or incomplete multipart uploads

Storage Tier Mapping in this Architecture

In the current deployment design: Lifecycle policies orchestrate movement between the below pools.

Tier	Backing Storage	Ceph Pool Type	Use Case
Hot Tier	OpenFlex Data24 (NVMe)	Replicated Pool/EC optimized for performance	Active, frequently accessed data
Cold Tier	Ultrastar Data60 (HDD)	Erasure-coded pool	Archive, infrequently accessed data

Lifecycle Workflow Details

Phase 1: Data Ingestion

Objects are written to the hot tier (OpenFlex)

Benefits:

- Low latency writes
- Fast initial read performance
- Ideal for ingestion-heavy workloads (logs, sensor data, media uploads)

Phase 2: Aging and Monitoring

Ceph tracks:

- Object creation timestamp
- Access patterns (if integrated with analytics tools)
- Metadata and tags

Phase 3: Policy Trigger

Lifecycle rule triggers based on conditions:

- Age-based (e.g., after 1 day or 30 days)
- Tag-based (e.g., archive=true)
- Business-defined policies

Phase 4: Automated Tier Transition

- Objects are asynchronously migrated to Data60-backed cold pools.
- Migration is managed by Ceph RGW lifecycle worker processes.
- No application impact (transparent to users).

Phase 5: Long-Term Retention / Expiry

Optional archival period

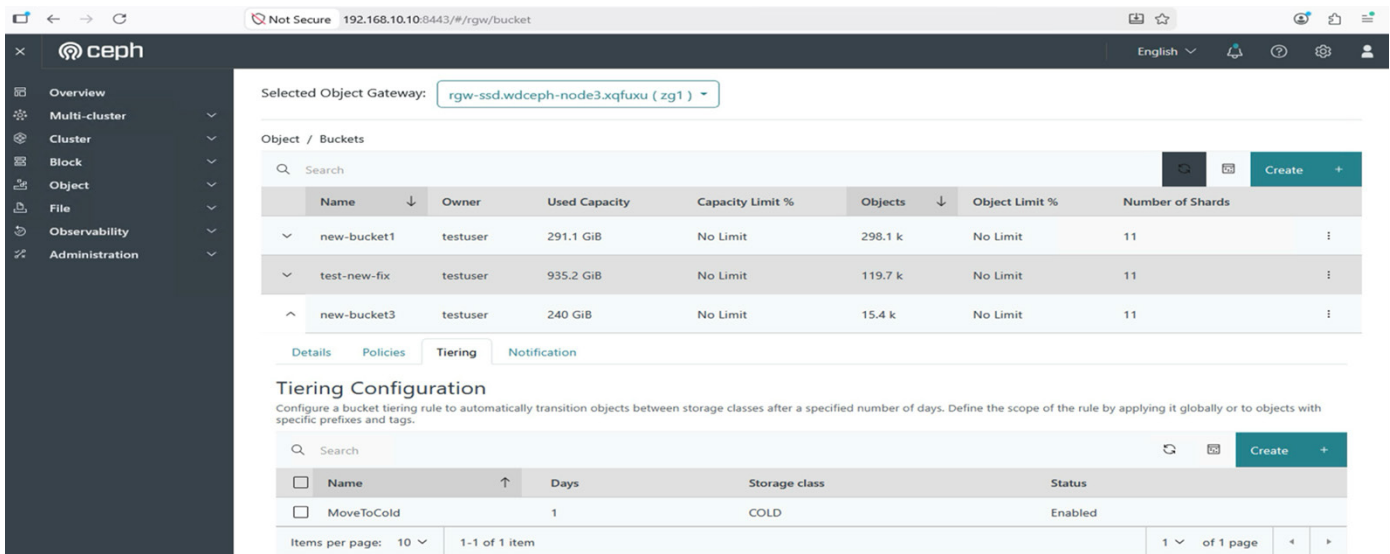
- Objects eventually:
- Remain in cold storage
- Or get deleted per retention policy

Example Lifecycle Policies: Standard Hot-to-Cold Migration

```
root@wdceph-node1:~# cat lifecycle.json
{
  "Rules": [
    {
      "ID": "MoveToCold",
      "Status": "Enabled",
      "Filter": {
        "Prefix": ""
      },
      "Transitions": [
        {
          "Days": 1,
          "StorageClass": "COLD"
        }
      ]
    }
  ]
}
```

A lifecycle transition in Ceph RGW is triggered by lifecycle policies and occurs when:

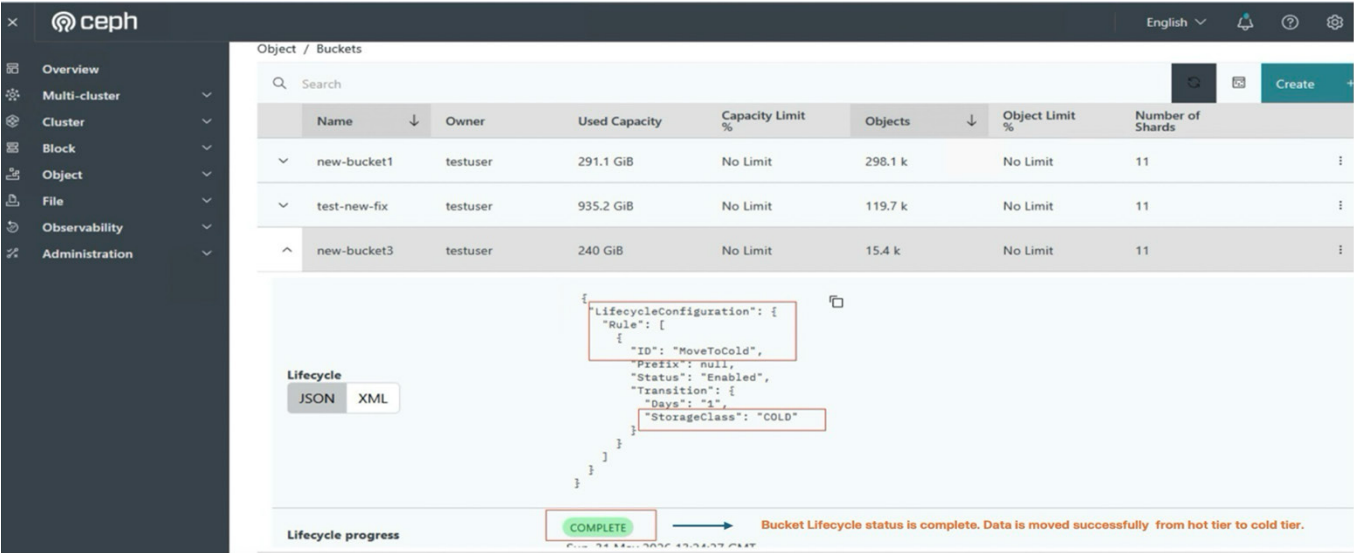
- **Object age condition is met:** The object reaches the number of days defined in the lifecycle rule (e.g., Days: 1).
- **Lifecycle (LC) engine execution:** The RGW lifecycle engine, a background process, periodically scans buckets and evaluates lifecycle rules.
- **LC worker scheduling:** The transition happens when the lifecycle worker runs and processes eligible objects during its scheduled cycle.



The screenshot shows the Ceph Object Gateway (rgw) web interface. The left sidebar contains navigation links: Overview, Multi-cluster, Cluster, Block, Object, File, Observability, and Administration. The main content area displays the 'Object / Buckets' view for the 'rgw-ssd.wdceph-node3.xqfuxu (zg1)' gateway. A table lists three buckets: 'new-bucket1', 'test-new-fix', and 'new-bucket3', each with details on owner, capacity, and object counts. Below this, the 'Tiering Configuration' section is active, showing a table with one rule named 'MoveToCold' that transitions objects to the 'COLD' storage class after 1 day. The interface includes search bars, pagination controls, and a 'Create' button for adding new rules or buckets.

Name	Owner	Used Capacity	Capacity Limit %	Objects	Object Limit %	Number of Shards
new-bucket1	testuser	291.1 GiB	No Limit	298.1 k	No Limit	11
test-new-fix	testuser	935.2 GiB	No Limit	119.7 k	No Limit	11
new-bucket3	testuser	240 GiB	No Limit	15.4 k	No Limit	11

Name	Days	Storage class	Status
MoveToCold	1	COLD	Enabled



Performance Details

The measured performance numbers are specific for the current configuration of Ceph Tentacle with OpenFlex Data24 4000 topology serving as the hot tier, in conjunction with the uses of 24 NVMe SSDs/osds with a multipath connection via 6 storage nodes. Warp tool is used to run the performance benchmarking on the system. The performance test is run using warp, simultaneously running tests from all the 6 Ubuntu clients.

Ceph Object Storage Replication Performance with OpenFlex Data24 4000			Ceph Object Storage Erasure Coding Performance with OpenFlex Data24 4000		
Policy Details	Profile	Throughput	Policy Details	Profile	Throughput
Replication (3-Copies)	GET	29.03 GBps	Erasure Coding (4+2)	GET	14.40 GBps
	PUT	6.37 GBps		PUT	6.38 GBps

Note: The performance results are captured after running multiple iterations with different data size (object size, objects, concurrency, policies).

Disclaimer: All the performance data are subject to change depending on the Drive model, capacity, workloads, servers, CPU, tunings and HA topology used.

Advantages

- Optimized Cost-Performance Ratio**

Achieves the right balance between performance and cost across storage tiers:

 - High-performance NVMe tier (OpenFlex Data24 4000) for latency-sensitive workloads.
 - Cost-efficient HDD tier (Ultrastar Data60 3000) for bulk and archival data.
- Seamless Scalability Across Tiers**

Enables linear and flexible growth as data expands:

 - Scale hot and cold tiers independently based on workload demand.
 - Add storage nodes without disruption or reconfiguration.
 - Supports petabyte-to-exabyte scale environments.
- Automated Data Lifecycle Management**

Delivers intelligent and policy-driven data placement:

 - Automatically moves data between hot and cold tiers.
 - Reduces manual intervention and operational overhead.
 - Ensures optimal resource utilization across storage layers.

High Availability with No Single Point of Failure

Built for enterprise resilience and continuous availability:

- Distributed Ceph architecture eliminates single points of failure.
- Self-healing and automatic data rebalancing capabilities.
- Supports replication and erasure coding for data protection.

Reduced Latency for Critical Applications

Optimized performance for demanding workloads:

- NVMe-based OpenFlex Data24 ensures ultra-low latency access.
- Accelerates analytics, AI/ML, and transactional operations.
- Minimizes response time for frequently accessed datasets.

Efficient Storage Utilization

Maximizes value from storage infrastructure:

- Stores hot data on flash and cold data on cost-effective disks.
- Optimizes capacity usage through intelligent data placement.
- Reduces unused or overprovisioned storage resources.

Simplified Operations with a Unified Platform

Streamlines storage management across environments:

- Single Ceph platform for object storage across tiers.
- Unified management with cephadm orchestration and automation.
- Reduced complexity compared to siloed storage systems.

Future-Ready Architecture for AI/ML and Big Data Workloads

Designed to support next-generation data-driven use cases:

- Supports AI/ML pipelines and large-scale analytics.
- Ready for cloud-native and containerized environments.
- Enables multi-site deployments with geo-replication support.

Hyperconverged Infrastructure vs Disaggregated Architecture

Aspect	HCI Setup - Hyperconverged Infrastructure	HCI Setup - Disaggregated Architecture
Architecture	Compute + Storage tightly coupled	1U compute nodes + Open-Flex/Data60
Scaling	Add full nodes (compute + storage)	Independent scaling of compute and storage
Cost	Higher due to overprovisioning	Optimized based on need
Performance	Limited tier differentiation	Tiered – NVMe + HDD
Flexibility	Low	High

Benefits with Disaggregated Architecture

- Performance: Up to 5–10x for hot workloads (NVMe tier).
- Resilience: Multi-enclosure fault tolerance.
- TCO: 30–50% reduction via capacity tiering.
- Scalability: Linear and independent.

Rationale for Disaggregation

HCI limitations:

- Forced compute scaling for storage expansion
- Resource underutilization
- Limited tiering capabilities

Disaggregated model solves:

- Enables right-sizing infrastructure
- Provides workload-aligned performance tiers
- Improves fault isolation

Why should organization choose OpenFlex Data24 4000 and Ultrastar Data60 3000?

- OpenFlex Data24 4000 provides NVMe-over-Fabrics capabilities enabling high-speed shared storage.
- Ultrastar Data60 3000 delivers industry-leading density (60 drives in 4U).

The Combined solution offers:

- Performance + Scalability
- Flexibility + Efficiency
- Cost + Capacity Optimization

Conclusion

The shift from HCI to a Ceph-based disaggregated storage architecture with OpenFlex and Ultrastar Data60 represents a strategic evolution in enterprise storage design. By decoupling compute and storage, organizations gain unprecedented flexibility, scalability, and cost efficiency.

This Architecture

- Addresses modern data challenges
- Aligns infrastructure with workload requirements
- Delivers enterprise-grade performance and resilience
- Significantly reduces total cost of ownership

Key Takeaways

- Tiered storage ensures the right data is stored on the right media.
- OpenFlex Data24 4000 accelerates performance-critical workloads.
- Ultrastar Data60 3000 reduces cost for long-term data storage.
- Ceph Tentacle provides the intelligence and scalability backbone.
- The architecture is highly adaptable for modern data-driven enterprises.

Ultimately, this solution provides a future-ready object storage platform capable of supporting next-generation data-driven workloads while maintaining operational simplicity and efficiency.

