Software-Defined Storage for Private Clouds

Lightbits vs Ceph

December 2025
Rev. 2.0

Abstract
This document compares Lightbits and Ceph in terms of architecture, performance, Transactions Per Second, and Total Cost of Ownership, reflecting the current NVMe over TCP operating reality in private clouds.

Table of Contents

1. Overview.... 3
Lightbits vs CEPH Performance.... 4
2. Emerging Technologies.... 5
2.1 Evolution of Fast Block over Ethernet.... 5
2.2 Adoption and What “Mainstream” Looks Like.... 5
2.3 Benefits over Traditional Block Protocols.... 5
2.4 Implications for Architecture Choice.... 7
3. Comparing Architectures — Infrastructure, Cost Anatomy, and Trade-offs.... 7
3.1 Lightbits Infrastructure.... 7
3.2 Ceph Storage Infrastructure.... 8
3.3 Where the Costs Accumulate.... 10
4. Performance, TPS, and TCO.... 11
5. Conclusion.... 12
Appendix A. Testbed and Methodology.... 14
Appendix B. Sizing Example and Assumptions.... 14
Appendix C. Sources.... 14
About Lightbits Labs.... 15

1.Overview

Cloud computing, as an operating model, and cloud-native, as an application architecture, continue to redefine how IT is delivered and consumed. The strategy many data platform teams now follow is straightforward: standardize on software, automate aggressively, and lower total cost of ownership (TCO) without compromising performance, scalability, availability, or resiliency. That is where software-defined storage (SDS) has landed—especially in private clouds built with popular orchestrators such as Kubernetes, OpenShift, and OpenStack.

Ceph and Lightbits both deliver SDS on commodity x86 servers, but they employ different fundamental architectural schemas that impact performance and ultimately cost-efficiency at scale. This whitepaper will illustrate the differences and demonstrate how Lightbits is a superior solution for today's high-performance block workloads in popular orchestration environments, such as Kubernetes. Ceph is a unified, open-source platform that exposes object, block (RBD), and file services. Lightbits is a disaggregated, multi-purpose block platform natively designed with NVMe flash over standard Ethernet (TCP/IP) or NVMe/TCP, featuring seamless integrations with Kubernetes/OpenShift, and OpenStack.

Random 4K I/O per Lightbits server	
Workload	**up to IOPS**
Read Only	4.3M
70% Read / 30% Write	2.0M
50% Read / 50% Write	1.2M
Write Only	0.9M

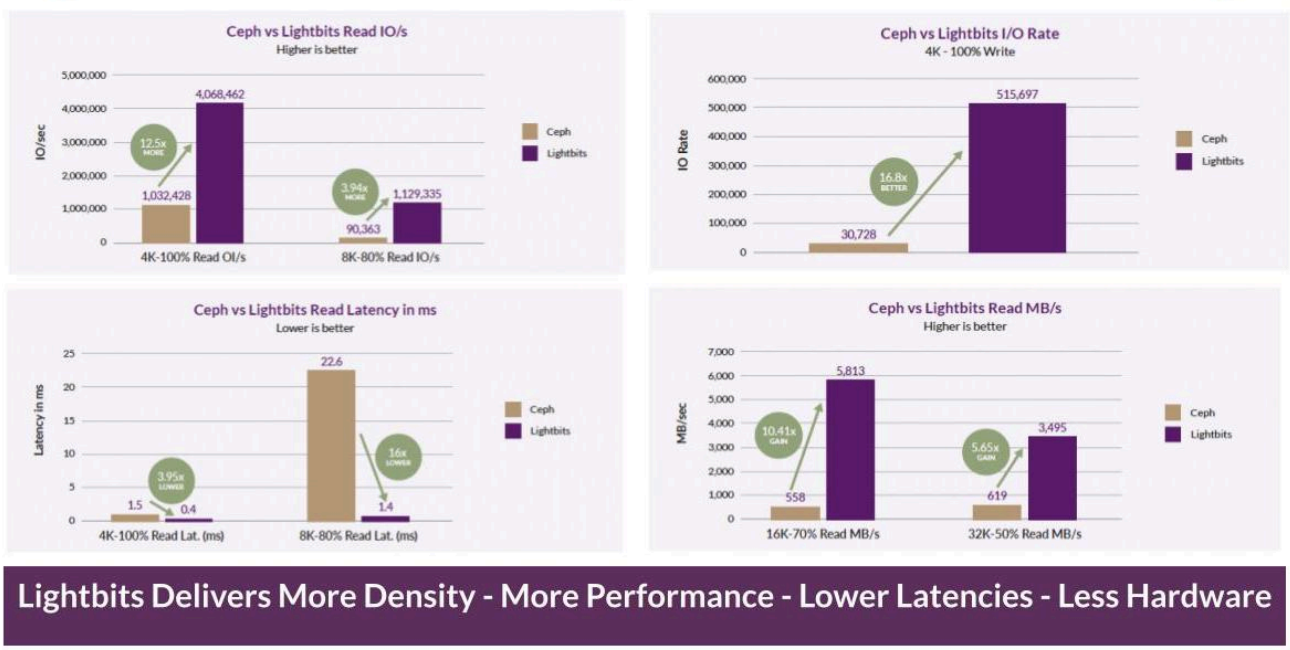
The architectural differences are evident in real-world workloads: on identical hardware and Kubernetes, **Lightbits achieved 4,068,462 IOPS per node** versus 1,032,428 Total IOPS for 4K reads, **515,697 IOPS per node** versus 30,728 Total IOPS for 4K writes, and **1,129,335 IOPS per node** versus 90,363 Total IOPS for an 8K 80/20 mixed read/write workload. For larger blocks, throughput likewise favored Lightbits software—5,813 MB/s versus 558 MB/s at 16K 70/30 and 3,495 MB/s versus 619 MB/s at 32K 50/50. These were similar servers with a significant difference in the number of drives being used. (Ceph drive count: 936. **Lightbits drive count: 36**)

A brief history of both solutions helps explain the differences in performance capabilities. Ceph storage began life as an open-source distributed object store and, over time, added block (RBD) and file services. It has steadily incorporated flash—using SSDs for critical metadata—yet many core design choices reflect an HDD-era origin. To reach NVMe/TCP hosts, Ceph added an

128K Bandwidth per Lightbits server	
Workload	**up to GB/s**
Read Only	22
Write Only	8

SPDK-based NVMe over Fabrics (NVMe-oF®) gateway that exports RBD images as NVMe namespaces. This improvement enhances performance but introduces a translation hop into the data path, resulting in added latency and CPU overhead as clusters scale. While IBM/Red Hat are major contributors to Ceph, there is a full alliance of contributors and individual contributors who update and enhance the platform. (https://ceph.io/en/foundation/members/)

Lightbits vs CEPH Performance



Lightbits Labs was formed to pursue flash-first performance over standard Ethernet and is credited with inventing the NVMe over TCP storage protocol. That origin is evident in the product's emphasis on the shortest possible data path (direct host-to-target NVMe/TCP), flash-aware durability (Elastic RAID with per-node erasure coding plus cross-node replication), and pragmatic integrations for Kubernetes (CNCF-listed CSI) and OpenStack (Cinder).

Since the original publication of this whitepaper, the NVMe/TCP ecosystem has matured. Discovery is predictable, multipath with ANA is broadly adopted, and security options, such as in-band authentication and TLS, are available across enterprise stacks. Ceph's NVMe/TCP path continues to progress through its SPDK-based NVMe-oF gateway, now featuring improved high-availability patterns and dashboard-driven operations.

Lightbits Labs has focused on customer-visible outcomes: steadier performance and effective capacity through production-ready TRIM/Deallocate; faster environment provisioning with Data Mobility Service thick clones; finer-grained scale control using per-controller I/O queue limits; and deployment flexibility through multi-network NVMe, single-server mode, multiple NVMe subsystems per cluster, API-driven key rotation, volume rename, and ongoing CSI compatibility for current Kubernetes releases.

2.Emerging Technologies

NVMe-oF—specifically NVMe/TCP—has become the pragmatic, fast-block transport for private clouds because it runs on standard Ethernet and scales horizontally. It maps parallel NVMe queues to TCP streams without specialized hardware. As the ecosystem matured, the operational experience improved: discovery is routine, multipath behavior is predictable, and security can be automated. That steadier foundation magnifies architectural differences above it. A direct host-to-target design, like Lightbits, benefits immediately; a gateway-mediated design, like Ceph's, inherits additional translation and hop costs as the load grows.

2.1 Evolution of Fast Block over Ethernet

The shift began when NVMe displaced legacy SCSI/SATA for local flash: the protocol is purpose-built for NAND, reducing command overhead and latency. NVMe-oF extended those benefits across a network, initially via transports that often required specialized NICs or switch features. NVMe/TCP removed that barrier by running on unmodified IP networks—no RDMA fabric requirements—so platform teams could scale flash performance over the same Ethernet they already operate. In practice, this means fewer constraints on topologies and an easier path to disaggregation for Kubernetes and OpenStack clusters.

2.2 Adoption and What "Mainstream" Looks Like

Analyst coverage has tracked NVMe/TCP's rise: the previous whitepaper version cites Gartner's view that by 2027, 25% of enterprises will adopt NVMe/TCP as a storage networking protocol, up from under 10% in mid-2023. The appeal is straightforward—higher performance than iSCSI and low-end Fibre Channel, with a simpler network story, which mirrors what we see in private-cloud builds. Real-world adoption is evident in transactional footprints. In an e-commerce case study, transitioning from Ceph to Lightbits on OpenStack and Kubernetes resulted in a significant increase in measured TPS, accompanied by a substantial reduction in server and SSD counts. [Read Global eCommerce Leader Boosts Transactions and Cuts Costs in OpenStack and Kubernetes]

2.3 Benefits over Traditional Block Protocols

NVMe over Fabrics—specifically NVMe/TCP—has become the pragmatic fast-block transport for private clouds because it runs on standard Ethernet, scales horizontally, and preserves NVMe's multi-queue design across the network. As NVMe matured, discovery, multipath, and security (in-band authentication and TLS) became routine in enterprise stacks—removing fabric friction and keeping the focus on the upper-layer data path. In that world, direct host-to-target designs like Lightbits benefit immediately, while gateway-mediated designs inherit translation and hop costs as load grows.

At the protocol level, NVMe was built for flash and massive parallelism. An NVMe controller can expose very large numbers of submission/completion queue pairs, and each queue can carry many outstanding commands. On Linux, the blk-mq storage stack maps NVMe I/O queues to per-CPU software queues, so I/O is issued and completed concurrently without a global lock—reducing head-of-line blocking and context-switch overhead. In practice, NVMe/TCP binds multiple I/O queues to CPU cores so reads and writes flow in parallel across connections, preserving the same parallel behavior hosts enjoy with local NVMe.

By contrast, SCSI-based transports (iSCSI and traditional SCSI/FC) originated in the disk era and conventionally funnel I/O through a single command queue per LUN with finite depth. You can increase concurrency with multiple sessions or paths, but the model centers on a shared queue with higher protocol overhead. At scale, the NVMe multi-queue model avoids that shared queue, lowers contention, and keeps latency more predictable under load. (Where NVMe/FC is used, it carries NVMe's multi-queue semantics across Fibre Channel; iSCSI and SCSI/FC retain SCSI semantics and their queueing limits.)

The practical payoff is what architects and operators actually care about: lower latency and higher throughput on the same Ethernet you already operate—without fabric-specific tuning or proprietary gear. Teams also gain predictable multipathing behavior for failover and path selection. This aligns with our earlier whitepaper's conclusion that NVMe/TCP delivers significantly lower latency and higher throughput than other TCP-based storage protocols while reusing existing IP networks, and with production results that show fewer servers are required to hit the same SLOs when the data path is direct NVMe/TCP.

Table 1: NVMe/TCP versus Traditional Block Transports

Dimension	NVMe/TCP	iSCSI	Low-End Fibre Channel
Transport & Fabric	IP over standard Ethernet; no special NICs/switch features	IP over standard Ethernet; often higher SCSI stack overhead	Dedicated FC fabric; proprietary HBAs/switches
Design Assumptions	Flash-native (NVMe commands/queues)	Disk-era SCSI semantics	SCSI over FC; optimized for SANs
Latency / Throughput	Lower latency / higher throughput on commodity networks	Higher latency under similar conditions	Good latency, but requires a dedicated fabric, which comes with additional costs
Operations	Standard routing/ACLs; predictable multipath/ANA behavior	iSCSI target discovery & tuning; legacy failover patterns	Fabric zoning, HBAs, and specialized ops
Adoption Drivers	K8s/OpenStack disaggregation, cost, simplicity	Legacy compatibility	Legacy SAN standardization

2.4 Implications for Architecture Choice

Because NVMe/TCP removes fabric friction, the dominant factor becomes the platform's data path. Direct host-to-target designs convert Ethernet speed into application performance with fewer moving parts. Gateway-mediated designs can expose NVMe/TCP semantics to hosts; however, the translation hop and the extra control plane show up as additional latency and CPU usage—and, ultimately, higher node counts as you scale.

Lightbits leans into the direct model: hosts talk straight to storage targets over NVMe/TCP on standard Ethernet with native multipathing and ANA for failover. Ceph, by contrast, brings NVMe/TCP to hosts via an NVMe-oF gateway that exports RBD images as NVMe namespaces. The gateway improves compatibility, but the added hop and protocol translation are real. As illustrated in the gateway diagram and discussion, that is the place where latency and CPU overhead accumulate under load.

You see the effect in scale math as well as on the wire. In the original like-for-like Kubernetes study, a Lightbits server delivered about 1.1M IOPS at 8K 80/20 while a Ceph server delivered about 90K. To reach ~4M IOPS, you are looking at ~4 Lightbits servers versus ~45 for Ceph. The ripple effects are predictable: more racks and switch ports, more power and cooling, and more operational surface area.

The production story mirrors the lab. In the global eCommerce case study, moving from Ceph to Lightbits increased TPS while shrinking the server and SSD footprint—yielding 68% less hardware and an 86% drop in cost per transaction over five years. That outcome is exactly what you would expect when you remove a translation layer from the data path and let NVMe/TCP's parallelism run end-to-end.

3. Comparing Architectures — Infrastructure, Cost Anatomy, and Trade-offs

Lightbits and Ceph are both SDS systems that run on commodity servers and Ethernet. Still, they differ materially in how I/O flows, how durability is achieved, and where costs accumulate as you scale. This section expands on the architectural description of each platform and then compares cost drivers side by side.

3.1 Lightbits Infrastructure

Lightbits implements a disaggregated, flash-native block architecture that presents NVMe namespaces directly to hosts over NVMe/TCP. The data path is intentionally short: application hosts connect straight to storage targets on standard Ethernet, using native NVMe multipathing with ANA for path selection and failover. Because compute, DRAM, and NVMe devices scale independently, the platform can be tuned for either per-node performance or aggregate capacity without over- or under-provisioning other resources.

Durability and media efficiency are handled by Intelligent Flash Management with Elastic RAID, which pairs per-node erasure coding with cross-node replication. This approach protects against device loss and node-level events while avoiding the capacity tax of blanket triple replication. Operationally, Lightbits integrates with Red Hat OpenShift via a CNCF-listed CSI driver and with Red Hat OpenStack via Cinder, making persistent volumes and snapshots first-class citizens in both environments.

From a cost perspective, the direct NVMe/TCP path reduces the number of servers required to reach IOPS and latency targets. Higher per-node performance translates into fewer NICs, fewer switch ports, lower power consumption and cooling requirements, and a smaller surface area to manage. Replication efficiency further reduces total media usage: in production case studies, Lightbits achieves target durability with two copies, whereas competitive designs typically assume three.

3.2 Ceph Storage Infrastructure

Ceph is a unified storage platform that layers block (RBD) and file services on a distributed object store. For block workloads, clients typically use the RBD protocol. When NVMe/TCP semantics are required, an NVMe-oF gateway exports RBD images as NVMe namespaces over TCP. This improves host compatibility but introduces a translation hop between initiators and the underlying RADOS cluster.

Historically optimized for HDD-era patterns, Ceph has adopted SSDs for critical metadata paths and continues to modernize; however, at larger scales, the combination of OSD processes, replication traffic, and gateway translation increases east-west bandwidth demands and CPU utilization. That pressure shows up as additional servers for the core cluster, potential gateway nodes for throughput and HA, and more switch capacity as clusters grow.

From a cost standpoint, the gateway fleet, higher node counts to reach performance targets, and triple replication for durability all expand CAPEX and OPEX compared with a direct NVMe/TCP design. These factors also increase the operational surface area—resulting in more processes to monitor, more failure domains, and more rebalancing during upgrades or drive events.

Table A. Architectural Contrasts and Cost Implications

Dimension	Ceph	Lightbits	Cost Implication
I/O data path	NVMe/TCP via NVMe-oF gateway exporting RBD images	Direct host→target over NVMe/TCP on standard Ethernet	Gateways add servers and a translation hop; higher latency/CPU at scale for Ceph
Durability model	Triple replication common for block durability	Per-node erasure coding + cross-node replication	Ceph uses more raw capacity and network for the same usable TB
Scaling pattern	Lower per-node IOPS; add core nodes and often gateways for throughput	Higher per-node IOPS; scale CPU/DRAM/NVMe independently	More servers, NICs, switch ports, power and cooling for Ceph at a given SLO
Kubernetes/OpenStack integration	RBD clients or NVMe-oF gateway to expose NVMe/TCP semantics	CSI (CNCF-listed) and Cinder driver (pre-installed from OpenStack Yoga)	Gateway layer increases operational surface area on Ceph
Network behavior	East-west traffic across OSDs plus gateway flows	East-west minimized by direct path; host multipath handles failover	More fabric bandwidth and operational tuning for Ceph at scale

Table B. Baseline I/O Performance (Higher is Better)

Workload	Ceph Cluster	Lightbits (4 nodes) Cluster	Lightbits Advantage
4K 100% Read (IOPS)	1,032,428	16,273,848	15.76X higher
4K 100% Write (IOPS)	30,728	XXX	XXX higher
8K 80/20 (IOPS)	90,363	4,517,340	49.99X higher
16K 70/30 (MB/s)	558	23,252	41.67X higher
32K 50/50 (MB/s)	619	13,980	22.58X higher

Tests performed using identical hardware setups in Kubernetes

Table C. Read Latency (Lower is Better)

Metric	Ceph	Lightbits	Lightbits Advantage
4K 100% Read (ms)	1.5	0.4	3.95X lower
8K 80/20 Read (ms)	22.6	1.4	16X lower

Tests performed using identical hardware setups in Kubernetes

3.3 Where the Costs Accumulate

Capital expenditure (CapEx) increases with the server count, the number of NICs, switch ports, and media types. In the 8K 80/20 profile from the baseline, a Lightbits server delivered approximately 1.1 million IOPS, while a Ceph server delivered about 90,000. To reach ~4 million IOPS, a Lightbits design required ~4 servers, while a comparable Ceph design required ~45 servers. The negative CapEx impact with Ceph is predictable: more racks and power circuits, more optics and cables, and a larger surface area to monitor and maintain.

Ceph's durability method compounds the negative CapEx impact. Lightbits meets availability targets with erasure coding and two copies across nodes, reducing write amplification and media footprint. Ceph commonly plans for triple replication on block workloads, which increases raw capacity, network traffic, and rebuild windows.

Operational expenditure (OpEx) increases with the number of moving parts. A direct NVMe/TCP model concentrates operations on the storage cluster and host multipathing. A gateway-mediated model adds another fleet of servers and software to deploy, scale, monitor, and tune, and introduces additional failure domains that must be covered with HA designs and runbooks.

Taken together, the architectural differences explain the performance and TCO deltas seen elsewhere in this paper. Lightbits' direct NVMe/TCP keeps latency and CPU overhead down and allows higher IOPS per node, while the gateway translation inherent to Ceph's NVMe/TCP path shows up as extra nodes—and extra dollars—at scale.

4.Performance, TPS, and TCO

The baseline IOPS and latency numbers translate directly into transactions per second (TPS) and cost. In production, one global e-commerce leader's transition from Ceph to Lightbits for OpenStack and Kubernetes resulted in a significant increase in TPS while reducing the server and SSD footprint. [Read Global eCommerce Leader Boosts Transactions and Cuts Costs in OpenStack and Kubernetes] What matters is not just peak throughput, but sustained TPS per node, replication efficiency, and how many boxes you need to hit service-level objectives during peak seasons.

Table 3. TPS Comparative Analysis

Metric	Ceph	Lightbits
Measured TPS (Mixed)	35,000	45,000
# of Storage Servers	40	4
# of SSDs Per Server	24	9
# of NVMe SSDs	960	36
TPS per Storage Server	875	11,250

The TPS per-node delta is the key takeaway: roughly 875 TPS per Ceph storage server versus 11,250 TPS per Lightbits server. **That efficiency meant four (4) Lightbits storage servers instead of forty (40) Ceph servers, and thirty-six (36) NVMe SSDs instead of 960 for the same transaction target.**

Replication efficiency also mattered; the Lightbits design met durability objectives with 2X replication, whereas the Ceph footprint assumed 3X.

Table 4. Performance-Based Cost Comparison (5-Year Hardware Only)

Metric	Ceph	Lightbits
# of Storage Servers	40	4
Total 5-Year Investment in Hardware (USD)	$1,393,920	$139,392

Table 5. Capacity-Based Cost Comparison (10 PB Project)

Metric	Ceph	Lightbits
# of Storage Servers (10PB project)	141	26
Total 5-Year Investment in Hardware (USD)	$6,204,000	$1,950,000

Table 6. Illustrative Sizing to ~4M IOPS @ 8K 80/20

Metric	Per-Server IOPS	Servers Required	Notes
Lightbits	~1,100,000	~4	Direct NVMe/TCP
Ceph	~90,000	~45	NVMe-oF gateway path

Translating performance into economics is straightforward. Fewer servers mean fewer racks, optics, cables, and switch ports; lower power and cooling; and less time spent on failure domains, rebalancing, and upgrades. That is, before considering the operational simplicity of direct NVMe/TCP versus a gateway that must be scaled, monitored, and tuned alongside the core cluster.

5. Conclusion

The choice to cost-efficiently support cloud-native, high-performance block workloads–Lightbits is the clear frontrunner. When examining the data path, Lightbits is designed for a direct NVMe/TCP host-to-target model, whereas Ceph delivers block storage through RBD and exposes NVMe/TCP via a gateway that translates between the two. That extra hop is where latency and CPU overhead creep in at scale—and it's why you end up deploying more nodes, more ports, and more people to keep it upright. The architectural delta isn't theoretical; it manifests directly in application behavior and in the number of hardware appliances you need to purchase.

On identical hardware with Kubernetes, Lightbits consistently outperformed Ceph across the workload spectrum. For larger blocks, Lightbits led on throughput and delivered dramatically lower read latencies. If you care about tail latency—and any architect running databases or microservices does—this is the difference between snappy and 'why is checkout slow?'

Efficiency follows the same pattern. In the 8K 80/20 profile, Lightbits delivered approximately 1.1 million IOPS per server, while Ceph delivered about 90,000. To hit ~4M IOPS, you're sizing ~4 Lightbits servers versus ~45 Ceph servers. That delta cascades into fewer racks, fewer NICs and switch ports, less power and cooling, and a smaller blast radius when hardware fails. Lightbits also meets durability goals with per-node erasure coding plus cross-node replication, rather than assuming blanket 3X replication—so you're not paying a capacity tax just to stay online.

The business impact shows up in transactions per second and in the bill. In a real-world scenario, migrating from Ceph to Lightbits for a global e-commerce platform increased PostgreSQL TPS while reducing the footprint. Specifically, the results showed a ~11,250 TPS per Lightbits storage node, compared to ~875 TPS for Ceph, with 4 storage servers versus 40, and 36 NVMe SSDs versus 960, to achieve the target. Over five years, that translated to 68% less hardware and an 86% lower cost per transaction. In other words: higher revenue capacity with lower infrastructure spend, and headroom for peak events without the scramble.

For the architect, a shorter, cleaner data path and flash-aware resiliency deliver more IOPS per node, improved latency, and faster rebuilds—on plain Ethernet, with CSI/Cinder integrations that behave as expected by OpenShift and OpenStack. For the C-suite: fewer servers and less replication overhead reduce CapEx and OpEx, lower power and space requirements, enhance sustainability initiatives, and higher TPS per dollar translates into real margin. Put simply, Lightbits converts the NVMe/TCP promise into measurable performance and TCO impact. At the same time, the gateway-mediated approach in Ceph adds cost and complexity exactly where you don't want it.

If you're deciding what to standardize on for the next five years, run a quick, like-for-like POC and size to TPS and SLOs. The baseline data and the production case study already point in the same direction: choose the direct NVMe/TCP architecture and reap the benefits of improved performance, resilience, and cost savings.

Appendix A. Testbed and Methodology

Both platforms were evaluated as storage for containerized workloads on Kubernetes, utilizing identical three-node (3-node) storage clusters and twelve (12) generator nodes. vdbench drove 96 container instances across the cluster, utilizing five workload profiles: 4K 100% read, 4K 100% write, 8K 80/20, 16K 70/30, and 32K 50/50.

Appendix B. Sizing Example and Assumptions

Sizing to ~4M IOPS uses the measured ~1.1M IOPS/server for Lightbits and ~90K IOPS/server for Ceph on 8K 80/20 from the baseline. The example is illustrative; always validate absolute numbers on the current kernel and distribution in a short POC.

Appendix C. Sources

Lightbits vs Ceph Whitepaper (March 2024): baseline performance table, architecture, and testbed. Global eCommerce Leader Case Study (2025): TPS measurements, server and SSD counts, and five-year hardware investment comparisons.

About Lightbits Labs

Lightbits Labs® (Lightbits) invented the NVMe over TCP protocol and offers best-in-class software-defined block storage that enables data center infrastructure modernization for organizations building a private or public cloud. Built from the ground up for low consistent latency, scalability, resiliency, and cost-efficiency, Lightbits software delivers the best price/performance for real-time analytics, transactional, and AI/ML workloads. Lightbits Labs is backed by enterprise technology leaders [Cisco Investments, Dell Technologies Capital, Intel Capital, Lenovo, and Micron] and is on a mission to deliver the fastest and most cost-efficient data storage for performance-sensitive workloads at scale.

www.lightbitslabs.com

info@lightbitslabs.com

US Offices
1830 The Alameda,
San Jose, CA 95126,
USA

Israel Office
17 Atir Yeda Street,
Kfar Saba 4464313,
Israel

LBWP06R3/2025/12