Backblaze

WHITE PAPER

Why Neoclouds Should Not Build Their Own Storage

Executive summary

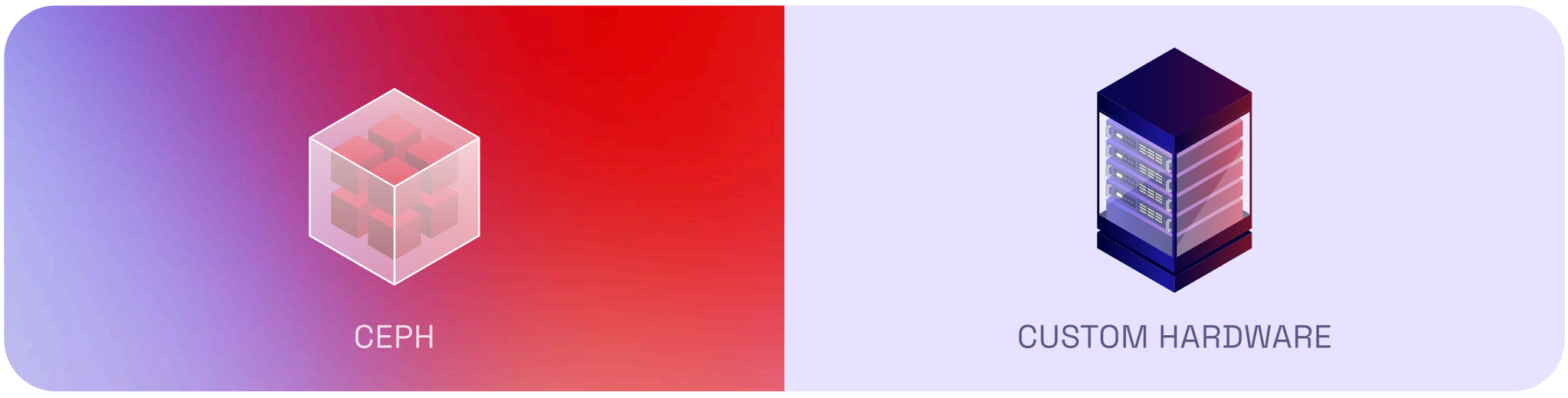
The unbundling of the cloud stack has created a massive opportunity for neocloud providers. Deep specialization in best-of-breed compute, AI, and database services is now possible without taking on the operational sprawl of hyperscalers.

However, the storage layer remains a critical bottleneck. Without an independent, S3-compatible object storage backbone, neoclouds face the split-stack problem, where customers run compute on the neocloud but leave data in a major cloud provider and incur large and unpredictable egress fees.

To solve this, engineering teams often default to building their own storage layer, assuming the choice is simply between open-source software (Ceph) or custom hardware design. This white paper argues that both paths introduce unpredictable technical debt and force organizations to inadvertently launch a second, internal storage operation that diverts critical R&D resources from your core product.

Solution in brief

A practical guide for neoclouds evaluating whether to build or partner for object storage. Building storage through Ceph or custom hardware could take 6 - 18 months and will pull engineering focus away from core compute and AI differentiation. An independent storage layer supports a more open, economically viable cloud stack.



Solution elements

THE CHALLENGE

The split-stack problem, where neocloud compute runs independently while data remains in a major cloud provider, creating egress costs and architectural friction.

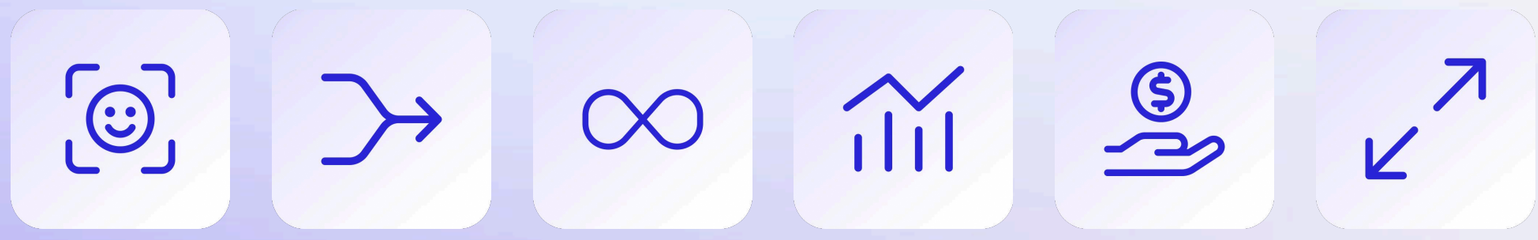
THE REALITY

Building storage through Ceph or purpose-built hardware introduces ongoing operational responsibility, infrastructure complexity, and long-term capital exposure.

THE APPROACH

Backblaze is an independent, S3-compatible storage backbone designed to integrate cleanly with neocloud platforms, including partner architectures that avoid punitive egress costs.

	Benefits
FOCUS	Keep engineering effort directed toward core compute, AI, and platform differentiation rather than storage operations.
SIMPLICITY	Remove the need to manage CRUSH maps, OSD tuning, rebalancing events, or high-friction data migrations.
FLEXIBILITY	Avoid long-term hardware commitments and supply chain constraints tied to custom storage builds.
PERFORMANCE	Avoid storage throughput ceilings and latency bottlenecks that can constrain high-performance compute and AI workloads.
ECONOMICS	Reduce or eliminate egress-driven cost friction and support a more open cloud architecture.
SCALE	Access exabyte-scale storage immediately without hardware procurement, deployment, or expansion cycles.



Two paths that pull neoclouds off course

1. The software trap: The real cost of Ceph

Ceph stands as the industry's de facto standard for open-source, software-defined storage because of its ability to unify object, block, and file workloads on commodity hardware. While other options exist, Ceph is the default choice for most engineering teams attempting to build S3-compatible architecture from scratch.

On paper, Ceph offers an attractive, open-source entry point for commodity hardware. In production, however, it represents a significant operational liability for neoclouds scaling beyond the petabyte range.

Entropy in day-2 operations

Deploying Ceph is trivial; managing its entropy is not. Real-world performance requires the constant, manual tuning of CRUSH maps and OSD (Object Storage Daemon) parameters to match shifting workloads. This is not a task for generalist sysadmins; it requires a dedicated team of specialized storage engineers to manage manual replication rules and rebalancing events that saturate network bandwidth.

Throughput ceilings

Generic Ceph implementations on standard clusters frequently hit throughput ceilings that purpose-built architectures are engineered to bypass. For high-performance use cases like AI model training, these limitations become a hard bottleneck for your compute product, constraining the value of the very services you are trying to differentiate.

Migration lock-in

Data gravity creates high-friction architectural lock-in. This is because Ceph relies on the CRUSH algorithm to mathematically scatter data shards. Migrating out requires re-hydrating every object through the gateway. This massive read operation creates a “noisy neighbor” effect, where resource contention can crash live workloads during the extraction process.

Two paths that pull neoclouds off course

2. The hardware trap: The risks of purpose-built R&D

Designing custom storage servers can look appealing to neocloud teams that want predictable performance and close control over how hardware and software work together. But in practice, this control comes at the cost of ongoing operational complexity, rising costs, and systems that are difficult to adapt or unwind.

Supply chain & drive heterogeneity

Unlike hyperscalers, most neoclouds lack the storage-focused engineering depth and operational maturity to guarantee consistent component availability. For example, if an architecture is designed for 14TB drives but the supply chain shifts to 20TB, engineers are forced to deploy mismatched hardware. This leads to stranded capacity, where the software layer cannot efficiently utilize the additional storage density—so you pay for hardware you’ll never use.

Erasure coding math

Managing data durability is a constant mathematical trade-off between capacity and safety. As hardware ages, you may need to shift parity ratios—for example, moving from 17+3 to a safer 15+5 scheme. This reduces your usable

capacity and inflates your cost-per-terabyte. It could also trigger a re-encoding workload that consumes CPU and I/O cycles, degrading performance for customers at the exact moment reliability matters most.

Infrastructure black boxes

At exabyte scale, object storage requires a network layer to load-balance and route massive ingress and egress flows. Teams must decide between dedicating resources to tuning open-source alternatives or buying expensive load balancers. These hardware systems act as opaque “black boxes” that are difficult to debug when connections drop during large file uploads. Either path creates a critical point of failure that demands niche storage networking expertise.

CapEx rigidity

The purpose-built model ties up capital in depreciating, low-margin assets on multi-year depreciation cycles. This locks the business into a cost structure optimized for yesterday’s demand forecasts, reducing the ability to pivot as workloads, pricing pressure, or customer expectations change.

The Decision Shift

From Infrastructure Ownership to Specialization

At scale, the storage question stops being a technical challenge and becomes an organizational one. Every hour spent tuning replication policies, managing rebalances, or planning hardware refreshes is an hour not spent improving developer experience, workload performance, or customer-facing innovation. .

The neocloud ecosystem thrives on interoperability. It functions best when providers connect with independent, open layers rather than building every component in-house.

The Solution

Storage as a Utility

When neoclouds partner with Backblaze, they eliminate the operational liabilities of Ceph and the capital rigidity of hardware R&D. Storage becomes a resource that is instantly available and infinitely scalable, rather than an internal engineering project that drains resources and slows time-to-market.

Technical advantages for neoclouds

S3 COMPATIBILITY	Integrate your existing neocloud compute, AI, and container workflows (including Terraform, CLI, and SDKs) with a drop-in, API-compatible storage layer.
MANAGED DURABILITY	Replace the complexity of erasure coding, bit-rot protection, and fleet management with a simple, reliable SLA backed by proven operational scale.
FREE EGRESS	In Backblaze-powered and colocated partner environments, move data freely between your compute instances and Backblaze storage without punitive egress charges.
INSTANT ELASTICITY	Eliminate the provisioning lag of hardware procurement. Scale from terabytes to exabytes without managing supply chains, data center expansions, or refresh cycles.

Key specifications

Throughput	Data Reliability	Security & Compliance	Availability
Up to 1 Tbps aggregate throughput	>99.9999999 99% (11 nines) annual durability	SOC 2 Type 2, SSO/MFA, Server-Side Encryption (SSE), and Object Lock (Immutability)	99.9% uptime SLA

Build what matters, not storage

The strategic advantage of a neocloud lies in its ability to specialize in high-performance compute and AI. Having to build and manage storage infrastructure dilutes this focus. By partnering with Backblaze, neoclouds gain an instantly scalable, economically viable storage layer that allows them to double down on what actually differentiates their platform.

Focus on what makes your neocloud great. Let Backblaze handle the storage. Learn more about Powered by Backblaze, or reach out to our storage experts to start a conversation.

About us

Backblaze gives businesses the freedom to innovate without limits by removing the barriers of lock-in, complexity, and cost. Our high-performance cloud object storage accelerates AI workflows, powers data-heavy applications, streamlines media management, and protects critical data. As an award-winning independent cloud, we provide unparalleled levels of interoperability that enable over 500,000 of our customers to reach and serve hundreds of millions of end users in 175 countries around the world.

Headquartered in San Mateo, CA, Backblaze (NASDAQ: BLZE) was founded in 2007 and serves over 500,000 customers in 175 countries around the world. For more information, please go to www.backblaze.com.