

# Nyriad® UltraIO™ Storage System

Fundamentally Changing the Foundation of Storage

WHITE PAPER

---

UltraIO™, Nyriad's flagship product, is fundamentally changing the foundation of storage — simplifying how data is stored, protected, accessed, and managed. Our new erasure-code-based architecture utilizes patented algorithms powered by GPUs and CPUs to deliver extreme performance, resilience, and efficiency, enabling massive amounts of data to be managed in a single storage platform.



Achieving 20 GB/s sustained ingest capability with its intelligent data placement (which uses all of the available drives all of the time) delivers an incredibly dense and efficient solution. Nyriad delivers the first of a new generation of storage solutions that empower businesses to grow, adapt, and stay competitive in a data-driven world.

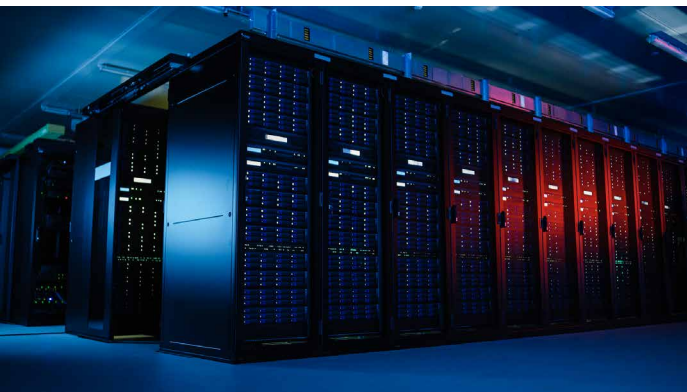
Let's discuss the problems the industry faces, and how we are alleviating them.

Global demand for compute, network, and storage resources are at an all-time high. This demand has caused a strain on storage infrastructures, particularly at Petabyte scale deployments. Compromises are being made to keep the data flowing and accessible. Instead of having a unified storage solution, many organizations have multiple iterations of file solutions, object solutions, and block solutions, all in less-than-optimal configurations. Capacity is sacrificed for performance and resilience, while in other cases performance is conceded for capacity needs. In some cases, risks are even taken with poor resiliency schemes because of desperation for more performance and capacity.



An ideal storage solution is based on a design that needs no compromises. The design delivers high performance, resiliency, and efficiency from a single solution. Today's storage solutions have been architected in such a way that they have forced customers to choose between one or two of these three attributes, while sacrificing at least one of them. Put simply, this means organizations are not realizing the true capability of the storage technology they are paying for or have purchased.

Until now, the industry has not achieved truly universal block level erasure-coded storage. Block level erasure codes haven't been widely leveraged due to the complexity of the algorithms needed and computational power required. File systems and proprietary software have been forced onto consumers when they acquired the storage. This of course makes the chosen vendor "sticky" and hard to remove. It's great for the storage vendor, but not necessarily good for the customer. A real universal storage system has block storage as its foundation and provides for customer-chosen file and object constructs to be built on top, with no artificial requirements for file systems and applications the vendor wants you to use for their benefit.



## What is causing current performance problems with data storage?

As the demand for data continues to grow, the industry is faced with the challenge of massive storage needs, driving the requirement for higher capacity devices with a lower cost per GB. This trend will continue as more and more data is being created, with denser media devices creating more efficient packaging per capacity point and thus lowering the cost for the device. Larger hyperscale data centers that use multiple data copies instead of RAID for data protection, coupled with proprietary data placement, have been able to take advantage of these denser drives. However, other smaller organizations have struggled to implement them, and the use of multiple copies is obviously highly inefficient.

## The RAID Compromise

On-prem data storage and storage at the edge have suffered from the limitations of RAID. In order to achieve sufficient performance, higher capacity drives are avoided in RAID powered systems to minimize the probability and scale of data loss and to reduce the period of degraded-array performance when a drive fails.

With this mounting pressure on storage performance, basic RAID 5 and 6 has been abandoned, in some cases, for more performant multi-data copy solutions, or RAID 50 and RAID 60 architectures. However, such solutions come with a significant loss of capacity efficiency as well as additional expense incurred by wasting purchased capacity.

It has been known for some time that erasure-coded solutions can improve storage efficiency and boost resiliency with the ability to fail many drives while not losing data. However, currently they require costly processing power, they don't provide high performance, and recovery times are lengthy. In addition, the storage efficiency provided by erasure codes is offset by inefficiencies caused by the additional compute and memory needed to implement.

## Data Duplication

Organizations seeking to maximize performance face a challenge in that traditional protection schemes are straining to keep up. This strain has been so great that, to reach desired performance density, many have been forced to adopt less than optimal solutions (i.e., multiple data copies) to escape parity/checksum slowdown or data protection schemes that leave large capacity solutions at risk and require more IT management.

If we double the capacity per device, connectivity throughput that ties the enclosure to the datacenter needs to be doubled just to deliver the same behavior as the previous generation of storage solutions. It should be noted that multiple copy solutions require a doubling or more of the needed network connectivity for writes since you must write everything two or more times. Further, the move to "scale-out" storage solutions has increased network complexity, cost, and storage latency. In some cases, check-summing is disabled completely as a desperate attempt to boost performance, even at the risk of data loss. Resiliency and capacity efficiency are also affected by these factors, as we shall see.



# What challenges are organizations facing with data storage resiliency?

Dense enclosures enable lower cost per GB media, resulting in cost-of-ownership reductions. However, for those looking to avoid the severe cost penalties of data replication, traditional RAID continues to be the choice for data protection and resilience. This is true despite only providing data loss protection in the event of a very few drives failing. Worse, when a drive does fail there is significant performance degradation while it is being rebuilt. The rebuild times, even for arrays with small drive counts, and small drives, can be substantial and the user feels the pain for the entire rebuild.

## RAID or RAIN

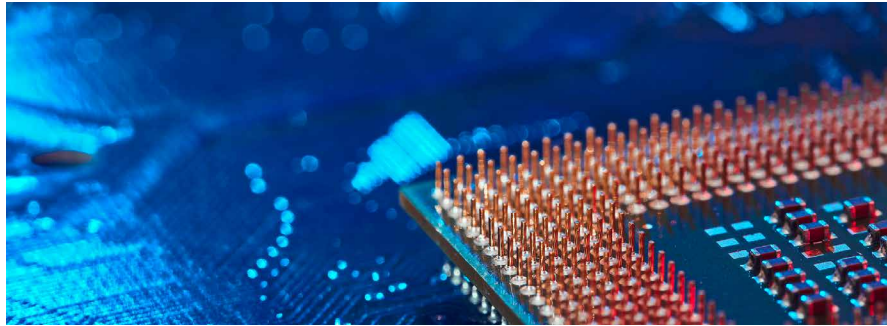
Excessive rebuild times measured in weeks have both constrained the use of the industry's larger storage drives and led to longer periods of operating with elevated risk of catastrophic data loss (while a previous failure is being rebuilt). Traditional computational methods for generating parity, especially the software-based methods, have struggled to keep up with the capacity density growth and the performance demands. This has resulted in a loss of both storage and performance density, as well as designs that handle parity at a file and object level or even higher. Those not willing to accept this performance loss have been forced into using multiple smaller RAID arrays or inefficient data copies. The vast majority of on-prem solutions use 8TB devices or smaller as a result of these limitations, even though the industry provides efficient 18TB and larger drives. In an attempt to alleviate performance issues, replication across individual nodes (RAIN) improves rebuild performance but adds network complexity, latency, and inefficiency.

## Erasure Coding

The use of erasure coding enables data resilience and availability without the need for replication. However, such benefits have been reserved for file and object stores and see limited use with block storage due to the intense amount of computation required. File and object solutions suffer from poor performance from intense CPU usage and network congestion in the case of RAIN configurations.

Erasure codes can provide a higher level of protection with respect to drive failure count, as well as enable the highest capacity media for the lowest total cost of ownership. What's needed is a way to bring these desired properties to block storage with intelligent erasure codes that won't lose performance benefits.

## Why is there inefficiency in the on-prem data center?



As stated previously, efficiency and cost savings are achieved almost entirely by increasing capacity density. Hyperscale companies have been taking advantage of the highest capacity media for over a decade because they jettisoned RAID in favor of replication and utilized proprietary device protocols. While replication works for the large distributed hyperscale infrastructure, in large part because the design requires multiple data copies for other reasons and because the use of over provisioning reduces the impact of duplication, it does not work for on-prem and edge data centers, where storage cost, space, and power are critical factors.

To utilize the capacity gains enjoyed by even standard storage devices, customers are faced with hard choices. They can be designed for the most efficient capacity use, which is likely an erasure code solution that doesn't allow for block storage and doesn't perform well. Or they can resort to using a RAIN configuration, but with a high cost of replicating capacity to boost rebuild performance. Prior to the development of Nyriad's UltraIO storage system, enjoying the properties of erasure coding at a block level has been elusive simply because of the required CPU utilization.

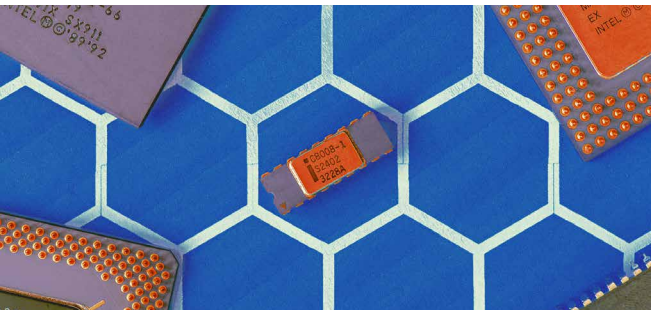
## Data Storage Costs

Efficiency extends beyond wasting money on unusable storage capacity. For example, the need for a certain level of data performance could force a user to make mirrors of that data, allowing reads to be pulled from two or more sources, which means they now have multiple devices to hold one device's worth of data. If these two devices use 12 watts each with an additional 6 watts of power consumed in a datacenter to cool that device, they are now consuming 36 watts of power or more for what could have been 18 watts, if no redundancy was needed.

This example of drive count and power has a string of negative effects throughout the data center. By doubling the drive count, we also double the enclosure count needed to house the drives as well as doubling the 'Top of Rack' switches and increasing the number of leaf and spine switches. All of this has the same power and cooling amplification described in the above simple drive example. If the additional architecture requires more servers to drive it, there is potential for software costs to increase as well if the software is licensed per CPU core. Additionally, solutions with predictable and repeatable performance requirements must over-design so that the performance is seen when drives have failed, as opposed to when everything is in an optimal state. This is the new minimum benchmark, and of course adds a layer of inefficient margin.

For all the aforementioned reasons, achieving all three fundamental principles of storage — performance, resiliency, and efficiency — simultaneously has been difficult if not impossible, until now.

**The solution is the Nyriad UltraIO™ storage system with a combined GPU + CPU processor architecture**



# What is a combined GPU + CPU processor architecture?

Nyriad's UltraIO storage system is fundamentally changing the foundation of storage by bringing the properties of erasure coding to block storage on a large scale and by implementing intelligent placement. Nyriad's software-defined, combined GPU + CPU processor architecture utilizes the combined computational strengths of GPUs and CPUs to generate impressive performance and value utilizing proprietary algorithms for erasure codes and data placement on block storage. We provide a balanced combination of performance, resilience, and efficiency.

This new combined processor architecture is a replacement for traditional RAID, allowing for performance up to 20 GB/s, providing a protected and redundant environment utilizing checksum creation, as well as efficient, intelligent data placement on storage. This solution is so resilient, you can choose to fail up to 20 drives in place with no intervention, while only experiencing approximately a 5% performance impact to your performance capabilities<sup>1</sup>. There is no longer a reason to overdesign to account for bad performance caused by drive failures or poor resiliency options with performance variability.

From what we have seen of the various challenges many organizations currently face, it is clear the principles of storage are tightly interlaced. Every action causes an opposite reaction. The common threads must work cohesively to enhance the solution.

<sup>1</sup> Based on Drive-pull IOR bandwidth on UltraIO v1.0 and BeeGFS 7.2.5. Test performed by System Fabric Works 7.12.2021

To deliver a superior combination of performance, resilience, and efficiency, the UltraIO architecture not only allows for high speed ingest of data, but also protects data with efficient erasure coding and high speed checksums, ensuring optimized data placement on the storage. Intelligent data placement results in lower IO activity on the storage devices—and fewer IOs wasted on such ‘behind the scenes’ activity means more IOs for customer use. The UltraIO system recreates individual blocks dynamically when required, for example, when a read or checksum validation fails. Crucially, this also allows for the largest capacity storage devices available to be used while still maintaining data resilience and performance. Denser storage devices equate to a smaller footprint and reduced Total Cost of Ownership (TCO).

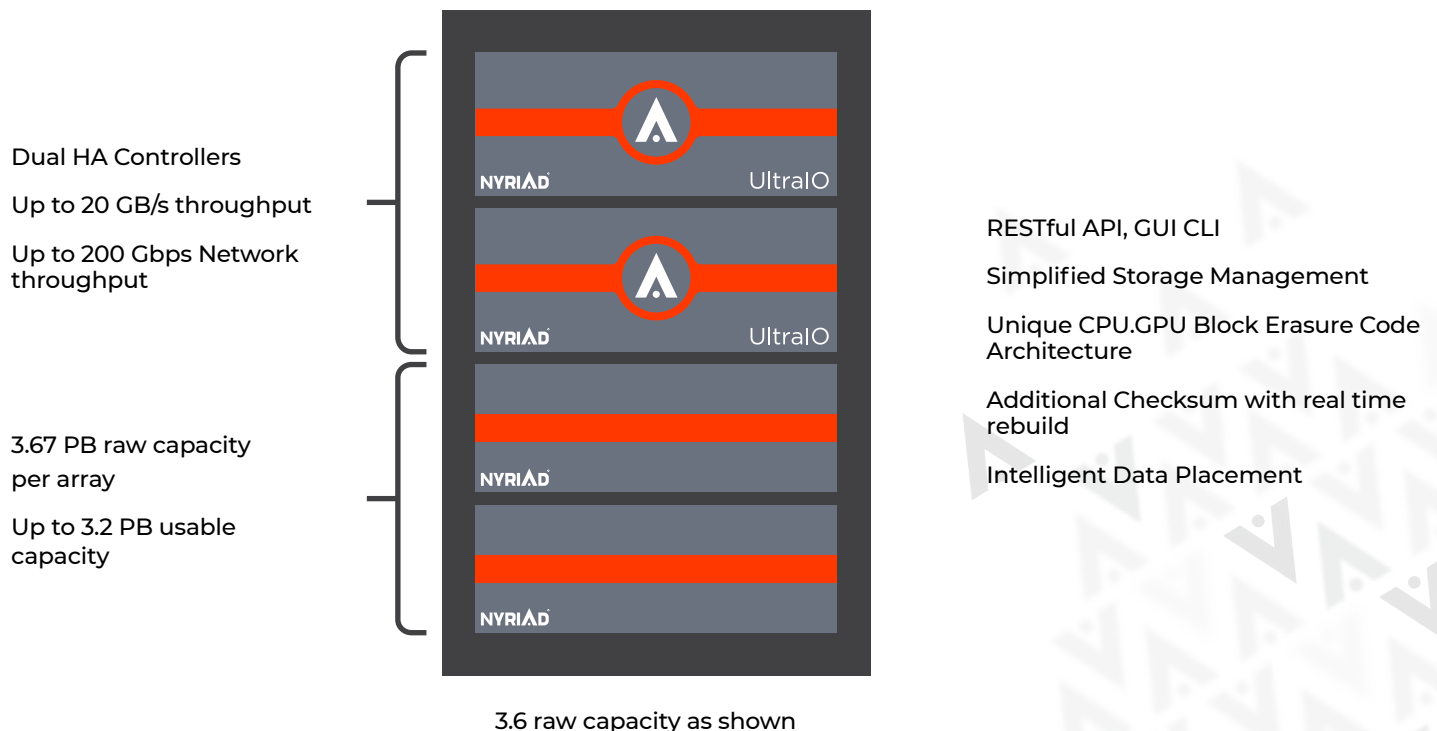


Fig.1: A typical configuration for 16u footprint

# Leveraging non-proprietary systems with cutting edge software for performance benefits

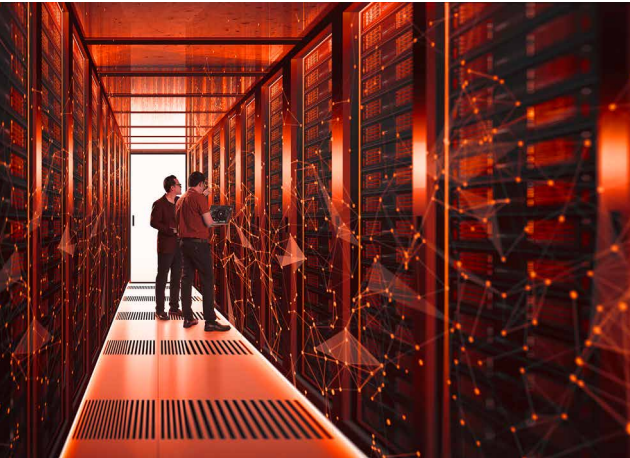
The combined use of GPUs and CPUs provides an extremely scalable storage engine. GPUs are highly parallel processors, originally invented to perform floating point calculation offloads essential to graphics applications such as video games and CGI rendering. Over the years, GPUs have evolved to become core processing elements in modern supercomputers and for machine learning applications. The GPUs intrinsic parallelism enables it to scale performance faster than any other CPU technology.

Having established the credentials of GPUs to handle massive calculations in a scalable manner, how can we use them in storage in a new and unique way?

As stated above, existing solutions based on erasure codes suffer from performance issues owing to the need for complex checksum calculation in real time with large arrays, coupled with write amplification tied to writing data to the array. UltraIO's combined GPU + CPU processor architecture not only handles the significant computational challenge of large scale, erasure-code-based arrays but does so at a pace that allows for block-level resiliency. This solution provides more value than simply being fast.

In an effort to have the efficiency of erasure codes without the computational problems brought forth by small files, many providers in the industry claim they deliver erasure codes at a file level. However, if the file size is 128k or smaller, those files get stored as triple copies. This of course is a waste of space utilized to desperately work around performance issues with small files. Nyriad utilizes erasure codes at a block level for any file size with no need to waste space on any triple data copies. We utilize erasure codes throughout our solution for the best performance and efficiency.

# Improve data resiliency



The UltraIO™ system is a foundational storage array that is built around block level data protection that is abstracted away from the actual data set, allowing for true universal resiliency.

This universal fit simplifies management of storage in the datacenter. Erasure codes implemented at a block level allows for much more effective data protection. Let's consider what this would mean for a 100-drive array.

A typical RAID 6 solution might have 10 arrays of 10 drives each. Each array would consist of 8 data and 2 parity drives worth of data stripes mixed within the array. The total parity would be 20 drives worth of data stripes. Theoretically, the user can sustain 20 drive failures. The reality is, if the wrong 3 drives fail (i.e., 3 drives in the same array), data is permanently lost. Moreover, blocks are bound to an individual RAID array, meaning seamless migration is either not possible or a lengthy manual operation is needed as each array is a separate entity. This can lead to hot spots causing significant performance degradation of user workloads. There is also a complication where one array fills up or does not have room for a large file being written when other arrays on the storage enclosure still have plenty of room. This leads to the customer thinking they don't have room for more data or a single large file, or it forces the administrator to create a control plane architecture to handle manual data movement and balancing when these events occur. Keep in mind, moving data to get around these problems also consumes precious resources needed for the normal workload's read and write operations. Building with a single large array removes the capacity issue but now in the previous example, 3 drive failures can destroy the entire 100 drive array. Adding more parity stripes in the design to correct that issue is challenging because of performance limitations due to computational struggles with existing industry solutions.

Alternatively, in a combined GPU + CPU processor architecture array, an erasure-coded pool of 100 drives can be created with 10 redundant drives. In this case, data would only be lost after the 11th drive failure, essentially having 5x the resilience with half the parity of the previous configuration. Moreover, the Nyriad UltraIO array will maintain near optimal performance<sup>2</sup> even after losing all 10 redundant devices (see Fig. 2). In the event of eventually rebuilding the array, we further reduce the user's pain by providing an extremely quick rebuild to reach an optimal state.

We extend this concept to account for as many as 20 drives out of 204 to fail with no data loss. For those not needing this high level of protection, they can choose a lesser number of parity segments and save capacity. The large data stripes allow for extremely fast rebuilds, as the data needed to rebuild is pulled from a large number of drives. This limits the rebuild effect on the array performance, in addition to speeding up the rebuild itself.

The UltraIO™ storage system allows the user to configure how much resilience they want on the array. So a user who wants to focus on absolute performance and capacity, might choose less. A user who wants to have the array run for many years without touching it, might configure for more.

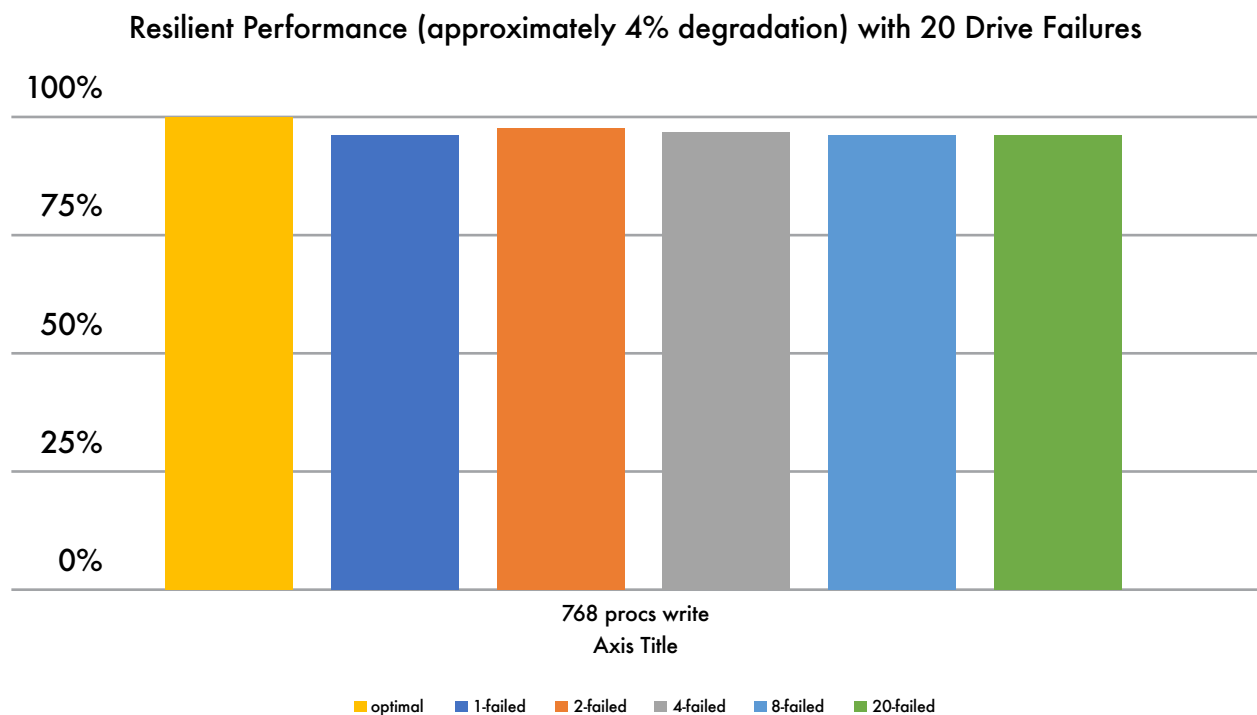


Fig. 2: Resilient performance of combined GPU + CPU processor architecture with approximately 4% write performance degradation with up to 20 drives in a failed state. Based on Drive-pull IOR bandwidth on UltraIO v1.0 and BeeGFS 7.2.5.

<sup>2</sup> Based on Drive-pull IOR bandwidth on UltraIO v1.0 and BeeGFS 7.2.5. Test performed by System Fabric Works 7.12.2021

# Ensure continual efficiency

The UltraIO™ system's combined GPU + CPU processor architecture eliminates the need to over-design a system to account for poor, degraded array performance. Indeed, it allows for a system to be built with the highest capacity devices without worry about long rebuild times, extended periods of poor performance, or the worst-case scenario of data loss. Storage device failures with no appreciable performance loss enable the use of 'fail in place' and 'scheduled storage device replacement' scenarios that reduce management cost by eliminating the need for emergency action 24/7, 365 days a year. The need to waste capacity with hot spares is also removed because the performance pain and data loss risk that drives a need for them have been removed. In addition, it enables remote storage placement for edge computing nodes that don't allow for immediate human interaction because of drive failure.



The combined GPU + CPU processor architecture enables intelligent data placement in a highly resilient, erasure-coded block environment that reduces write amplification to the storage devices, thus providing increased storage device performance. In a superior fashion, this intelligent data placement can enable users to take advantage of the highest capacity media. Previously, such ultra-high-capacity storage devices were built from the ground up to allow hyperscalers to save money, but they also required data placement schemes that the average user didn't have nor could use advantageously. In addition, reducing write amplification likewise reduces drive wear and therefore the chance of failure.

The efficiency is enhanced by the fact that intelligent data placement not only allows for large capacity drives, but enables the solution to be built on all hard disk drives with no need for an expensive Solid State Drive (SSD) tier. SSDs are frequently used to hide the standard hard disk drive performance most would observe, and the UltraIO system simply doesn't need this expensive abstraction.

Nyriad accelerates the UltraIO array by taking advantage of parallel access to many drives at once. Data is placed in intelligent patterns that ensure all of the drives are kept busy, thus ensuring a high level of overall performance. Data written in sequential patterns are usually read in sequential patterns and implementing write placement to account for this leads to both accelerated reads and writes.

## At a glance: the hunt for true performance, resiliency, and efficiency

Transform erasure codes from a resilient, efficient solution for file and object only, into a resilient, efficient, and performant solution for block, using an extremely powerful combined GPU + CPU processor architecture.

- Enable the use of highly efficient, large capacity storage devices, that currently only the hyperscalers can enjoy, by utilizing intelligent data placement
- Reduce write overhead in an erasure code environment by utilizing intelligent data placement
- Enjoy high performance throughput at 20 GB/s
- Remove the need for wasteful performance margin by having a solution that has minimal performance loss after drive failures
- Simplify management of the datacenter by reducing hardware footprint, all built around the most efficient erasure code options
- Build the datacenter around a common architecture designed at the block level, that can then be utilized for block, file, and object. This provides less complexity and less cost, as well as less operational risk

This extremely powerful combined GPU + CPU processor architecture from Nyriad has taken the next big step in storage, providing a solution that balances performance, resiliency, and efficiency!

- Sustained high performance<sup>3</sup>, even with multiple drive failures
- Ability to schedule replace and rebuild times without worry
- Extremely high performance that is also scalable
- Incredibly dense hardware footprint
- Ability to detect and recover from nearly all data corruption
- Lower power and cooling requirements
- Storage workload is enhanced and stabilized with both GPU and CPU involvement
- Built using off-the-shelf hardware
- Compatible with all existing POSIX based applications and file systems

## Sustainability

Nyriad's UltraIO™ system is ultra dense and efficient with respect to operational power and hardware footprint, and provides remarkable sustainability benefits. The power, cooling, and hardware footprints are very important aspects of a design with respect to cost. The cost to run a storage environment, as well as the power needed to cool the environment it is running in, is substantial. With that said, there are even more challenges. The location where the storage hardware resides has two main costs; the cost of the square footage needed to house the storage hardware and the cost of providing the power to run and cool the hardware. The data center incurs high costs for copper and power distribution to enable dense compute and capacity capabilities. These costs extend into network requirements as well as compute overhead. For example, a data center capable of servicing 2 million watts worth of hardware, versus a data center that services 1 million watts worth of hardware, requires twice the power capability, twice the copper wiring, and a double-sized service feed, in addition to HVAC cooling units at either twice the size or twice as many units as the smaller example. These aren't one-time costs, but costs that the user will be subjected to for as long as they own and run the data center with monthly power, cooling, and mortgage costs on the data center. When these operational costs are added in with the cost of acquisition, we get what we refer to as the total costs of ownership (TCO). This is usually viewed as a per month cost, and when looking at storage, it is usually referenced in a cost per TB per month.

<sup>3</sup> Based on Drive-pull IOR bandwidth on UltraIO v1.0 and BeeGFS 7.2.5.  
Test performed by System Fabric Works 7.12.2021

Nyriad provides a solution as described in this white paper with 3.6 PB of raw capacity in an ultra dense 16u space. This ultra dense solution is made possible by eliminating the negative effects of RAID and data replication, enabling the use of high capacity near-line drives. Nyriad's UltraIO system is capable of going even denser than this with additional storage enclosures and larger capacity drives in future configurations.

The current system described in this white paper provides a solution where less than 20% of the monthly TCO is operational costs versus acquisition costs. The user saves money and space while being extremely friendly to the environment.

Our use of non proprietary hardware means we can take advantage of industry standard improvements as they happen, with no need to wait for proprietary versions of those enhancements to be built.

## Universal Simplicity

Simplify your datacenter by leveraging a foundational storage array that provides an alternative to the inflexibility of proprietary file systems. Nyriad has built a storage architecture that works with your preferred non-proprietary file system(s) or application.

It is common for users to continue using solutions they don't really like because it is so complicated and painful to bring in a new file system and management scheme or remove an old one. This problem is compounded when they have different solutions for block, file, and object. Storage should be simple and built on a foundation of block storage, that is performant, resilient, and efficient — a foundation worth building upon.

The UltraIO™ system is placed below the software layers at the block level, enabling low disruption upgrades and an extremely easy path to adoption.

The architecture also gives organizations the flexibility to grow across multiple dimensions as needs dictate—quickly, easily, and without disruption. Nyriad arrays can scale from terabytes to petabytes either vertically by adding more storage to a controller or horizontally by adding additional controllers and storage. Furthermore, because the UltraIO system runs on industry standard hardware and uses off-the-shelf storage media, Nyriad can continue to take full advantage of performance improvements in GPUs, CPUs, communications connections, and storage media as they evolve.

# About Nyriad

Nyriad, Inc. is the developer of the UltraIO storage system, an all-new system that combines the processing speed of GPUs and advanced algorithms to deliver unprecedented performance, resiliency, and efficiency. The ground-breaking design enables UltraIO systems to support block storage media and block, file, and object data types in a single system for maximum flexibility. UltraIO systems run on industry-standard hardware, use the highest capacity, most efficient storage media, and simplify storage management to achieve low total cost of ownership. Headquartered in Austin, Texas, Nyriad empowers businesses to grow and adapt their storage to stay competitive in a data-driven world. For more information, visit us on the web at [www.nyriad.io](http://www.nyriad.io).

**NYRIAD**<sup>®</sup>

[www.nyriad.io](http://www.nyriad.io)

