ılı LOADBALANCER

The evolving role of load balancers in scale-out NAS environments

How intelligent load balancing can optimize the performance of high-volume data storage environments

Executive summary

The volume of data generated and stored worldwide continues to grow exponentially. IDC predicts that the global datasphere will expand from 33 zettabytes (ZB) in 2018 to 175 ZB by 2025.[1] Scale-out network-attached storage (NAS) is an effective way for enterprises to manage this escalation in data storage requirements. More scalable than block and file storage architectures, scale-out NAS systems enable organizations to expand their capacity on demand by adding more storage clusters to their core environments. Scale-out NAS architectures are ideal for organizations that need to regularly access, share and update data.

Load balancing functionality is essential for scale-out NAS environments. Almost all NAS vendors embed basic load balancing technology within their products to enable them to function. These entry-level load balancing capabilities fulfil a purpose, but they do little to optimize the performance of the entire NAS environment, particularly as it grows through the addition of more and more storage arrays.

Organizations can significantly improve the performance of their NAS architectures by coupling their NAS with a load balancer product from an independent load balancer vendor. Modern load balancers can count connections on individual storage nodes within scale-out NAS architectures – no matter how large these storage environments become – and dynamically direct traffic to different nodes to optimize performance for users.

The role of the load balancer in scale-out NAS architectures is evolving. In the near future, load balancers will be able to play an ever greater role in delivering high performance. Collaborating with NAS vendors, Loadbalancer.org is encouraging them to make data more accurate and available, in real time, which load balancers can use to direct traffic around NAS architectures more intelligently, not just based on traffic type and load. This development will enable load balancers to function more effectively and allow organizations to get best value from their investment in NAS solutions.

Organizations investing in load balancers for a NAS architecture should bear the evolving role of load balancers in mind. It is important to speak with shortlisted load balancer vendors about their partnerships with NAS vendors and how they are optimizing their products for NAS environments.

Finally, look to your shortlisted load balancer vendors for guidance on NAS deployments, how responsive their support is and any licensing restrictions that might limit the scalability of the NAS - in the end, it's not just the load balancer that will make the difference to the performance of a scale-out NAS.

---

1 IDC White Paper, sponsored by Seagate, 'Data Age 2025: The Digitization of the World from Edge to Core', Document #US44413318, November 2018.

What is scale-out NAS?

Simply put, scale-out NAS is a way of storing large amounts of data on a computer network. It is an especially appropriate storage architecture for IT environments where the volume of data is growing rapidly, because the total amount of disk space can be easily expanded, over time, as and when needed. Whenever the storage limit is reached, another array can be bolted onto the existing arrays to expand the system capacity.

Scale-out NAS are growing in popularity around the world. Indeed, according to Data Bridge Market Research, the global market for NAS solutions is expected to rise by 23.10% between 2019 and 2026.[2]

Scale-out versus scale-up

Scale-out NAS solutions differ significantly from more traditional, scale-up block and file storage platforms. In scale-up architectures, each storage system relies on a single controller head or pair of storage controllers. When more storage capacity is needed and the capacity limits for the controllers are reached, users have to upgrade to a more powerful controller or buy a whole new storage solution. In contrast, users of scale-out NAS solutions can continually add more clusters to their existing NAS architecture, as each array added has its own self-contained disk, processors, memory and networking resources.

NAS versus Object Storage

Like NAS, Object Storage also has a scale-out architecture. Object Storage is most commonly used in cloud computing and is the basis for leading online storage services such as Dropbox, Amazon Cloud and YouTube. The scale-out Object Storage approach adds metadata to files, allowing stored data to be retrieved more quickly, and is ideal for storing large volumes of files that will be accessed but not changed. In contrast, scale-out NAS architectures are more suited to enterprise environments where files may need to be accessed, shared and edited dynamically by many users. NAS can be used for email servers, virtual machines, databases and file servers.

2 Data Bridge Market Research, 'Scale-out NAS Market – Global Industry Trends and Forecast to 2026', [cited 2020 Jun 18]. Available from: https://www.databridgemarketresearch.com/reports/global-scale-nas-market

> *According to Data Bridge Market Research, the global market for NAS solutions is expected to rise by 23.10% between 2019 and 2026.*

Such is the similarity between NAS and Object Storage systems today that organizations that previously only considered Object Storage may now wish to also consider NAS. While they used to operate in separate spaces, there has latterly been quite a lot of convergence between the two technologies. For example, the protocol S3 (Amazon Simple Storage Service) used to be exclusive to Object Storage but now S3 is becoming available on NAS platforms. Some vendors are even deliberately blurring the differences between NAS and Object Storage to offer their customers the advantages of hybrid systems.

What role does load balancing play in scale-out NAS architectures?

Load balancing plays an absolutely essential role in NAS environments. As NAS systems comprise multiple connected storage arrays, it is vital for there to be some mechanism for sharing traffic across all the storage nodes in the cluster. If there was no load balancing technology in place to fulfil this function, then all the traffic would be directed to a single node, the node would become overloaded and the NAS system would fail.

The bare essentials for an operational NAS

As load balancing is such an integral part of how NAS architectures function, all NAS vendors include some basic load balancing capabilities embedded within their products. Generally speaking, these vendor-supplied load balancers are simple Domain Name System (DNS) solutions that mostly work by providing what is known as 'round robin' functionality. Each end user (client) gets presented to one node and then the next node, and then the node after that, and the cycle goes around and around, in a loop.

On the whole, vendor-supplied load balancing functions do work; if they didn't, the NAS solutions themselves wouldn't work. However, they do not deliver the optimal storage performance for customers. For example, the round robin method of load balancing doesn't take into account what is actually taking place on the network at any point in time and will assume that each node can handle a similar load. If some nodes have less capacity, they could overload, while other nodes could be underutilized, resulting in inconsistent performance for users.

The key component for optimal performance

In most instances, organizations will supplement the vendor-supplied load balancing functionality with a load balancer from a third party, specialist load balancer vendor. After all, an investment in a scale-out NAS architecture warrants the small additional cost of a product that will enable the storage solution to meet the performance expectations of the business.

Offering more sophisticated functionality than the embedded vendor solutions, third party load balancers distribute traffic across storage nodes in the scale-out NAS environment more logically. They work by counting the number of current connections on nodes, in real time, and directing the load to the resources with the fewest client connections. This approach leads to more even utilization of all storage clusters in the scale-out NAS architecture and more optimum performance for all users.

> *Offering more sophisticated functionality than the embedded vendor solutions, third party load balancers distribute traffic across storage nodes in the scale-out NAS environment more logically.*

How is the role of load balancers evolving in NAS environments?

It has long been recognized that today's load balancers have sophisticated functionality that could be used far more extensively in scale-out NAS environments to deliver ever greater storage performance for end users. Rather than just routing traffic by counting connections, load balancers could be used to intelligently direct the traffic based on an understanding of the actual network load and traffic type, in real time.

The load balancing technology to do this exists now. However, users of NAS solutions are not able to take advantage of it because NAS vendors do not routinely generate the accurate statistical data about connections in NAS environments that load balancers need to 'see' to route traffic around the storage solution.

The future is likely to bring greater collaboration between NAS vendors and load balancer vendors, as experts in both fields work together to optimize storage environments for their customers.

Some vendors openly state that their network throughput data is only collected up to every 10 seconds, so any data produced may create invalid readings for load balancers. Other vendors cannot provide data on what type of connections they have on each node. Consequently the load balancers cannot distinguish between Network File System (NFS), Common Internet File System (CIFS) or S3 connections and cannot use this added insight to more intelligently route traffic based on network load and optimize the performance of the NAS system.

Evolution through partnership

The future is likely to bring greater collaboration between NAS vendors and load balancer vendors, as experts in both fields work together to optimize storage environments for their customers.

Loadbalancer.org is working with a number of leading NAS solution vendors to help them identify ways to develop their NAS architectures so that their customers can make better use of the intelligent features in Loadbalancer.org solutions. With some vendors, Loadbalancer.org has phase two developments in progress and phase three plans in the pipeline, which will inject greater intelligence into the load balancing function in scale-out NAS environments in the near future.

Regular API calls allow the load balancer to adjust load distribution based on real-time node stats
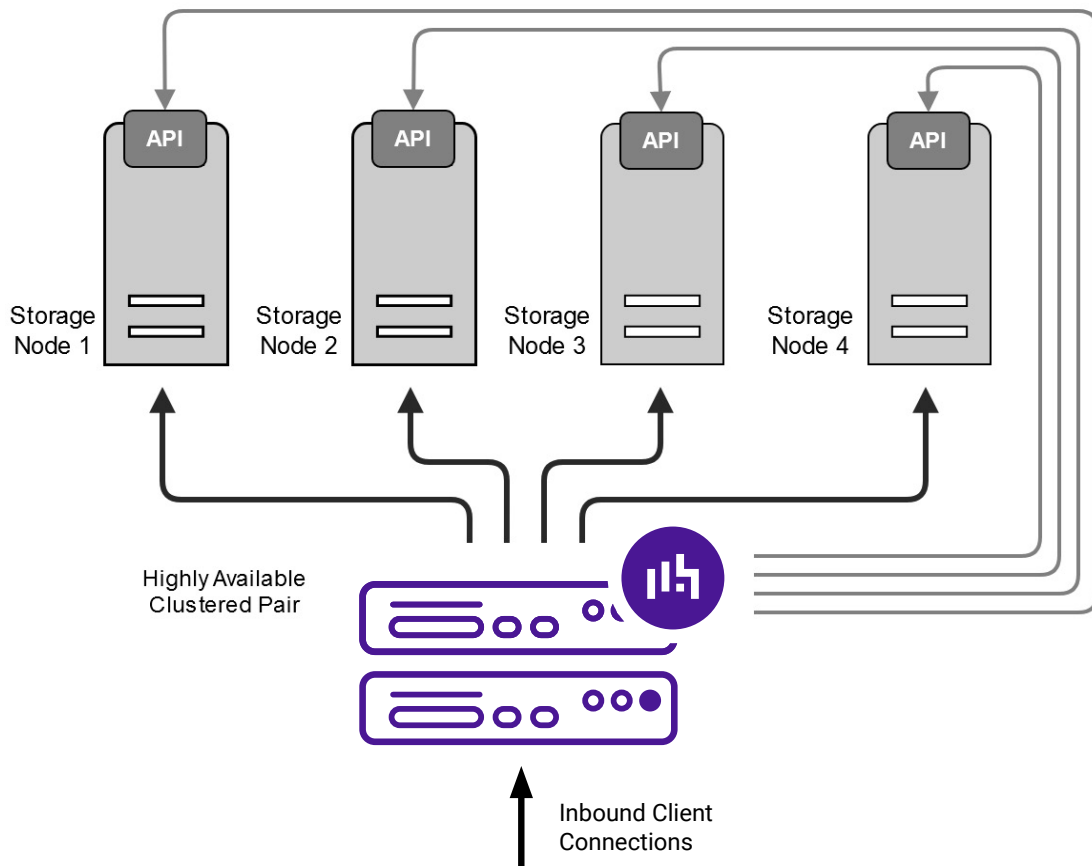
Detailed node statistics accessed in real-time using the NAS vendor's built-in API enables intelligent load balancing decisions to be made

Which criteria are important when choosing a load balancer for a NAS environment?

Organizations that are planning to build, extend or upgrade a scale-out NAS environment should give careful consideration to their choice of third party load balancer. There are many different load balancers on the market – often at vastly different price points with very similar functionality. It can be very easy to spend too much money on a product that doesn't offer the scalability, flexibility and reliability that is essential for scale-out NAS architectures.

Key criteria to consider when choosing a load balancer vendor include:

Partnerships with NAS vendors

With an eye to the future, organizations that are investing in NAS systems today should look for load balancer vendors who have existing partnerships with NAS vendors and a willingness to work closely with more NAS vendors. A collaborative approach will pave the way for the exchange of accurate statistics from the storage nodes to load balancers and enable more intelligent load balancing. For example, load balancer vendors ideally need to work with NAS vendors to help them develop their application programming interface (API), so that the capabilities of the load balancer are not limited by the API's ability to provide real-time cluster and node data. It is important to note that while all load balancers can theoretically be configured to work with any statistics supplied by NAS vendors, not all load balancer vendors would be willing to do this adjustment without a significant additional consultancy fee.

Adaptable health checks

As NAS vendors begin to make more and more statistics available to support intelligent load balancing, load balancer vendors will need to be able to adapt their health check functionality to enable the load balancer to interact with the API of the scale-out NAS. For example, Loadbalancer.org is developing flexible health checks that can make API calls to the nodes to get real time node statistics. This allows Loadbalancer.org to adjust client distribution to nodes based on node characteristics to optimize user performance. These characteristics range from standard CPU and RAM statistics to more sophisticated statistics on the number and type of connections or protocols to ascertain load at a more granular level.

Product-specific deployment guides

Product-specific deployment guides can be invaluable during the implementation of scale-out NAS architectures, helping to accelerate deployments and reduce delays from manual installation errors. If load balancer vendors have product-specific deployment guides for object-storage solutions, this is a strong indication that they will have – or will be producing – a similar suite of product-specific deployment guides for NAS solutions too.

Capacity for storage growth

Given that one of the advantages of scale-out NAS architectures is their scalability, organizations should be careful to make sure that they don't inadvertently restrict the growth of their NAS environment by purchasing a load balancer with throughput limits. Some load balancers have restrictive licensing agreements, which impose artificial limits on throughput, and these kinds of products can become costly or even obsolete in just a few years. Organizations should therefore seek products without throughput limits and size them to take into account the likelihood of rapid, significant data storage growth.

Reliable support and highly quality professional services

A scale-out NAS environment is likely to be a key business system and, as such, uptime will be critical. As well as evaluating product features, organizations should carefully consider the vendor's support credentials and review testimonials. Having a vendor that the IT team can trust to provide a rapid, effective and friendly response in an emergency will provide enormous peace of mind. In addition, many organizations will benefit from guidance from a team of professional services experts who are experienced in delivering customer solutions. They can architect a load balancer solution to optimize scale-out NAS environments, provide remote load balancer installations and deliver customer training to maximize your return on investment.
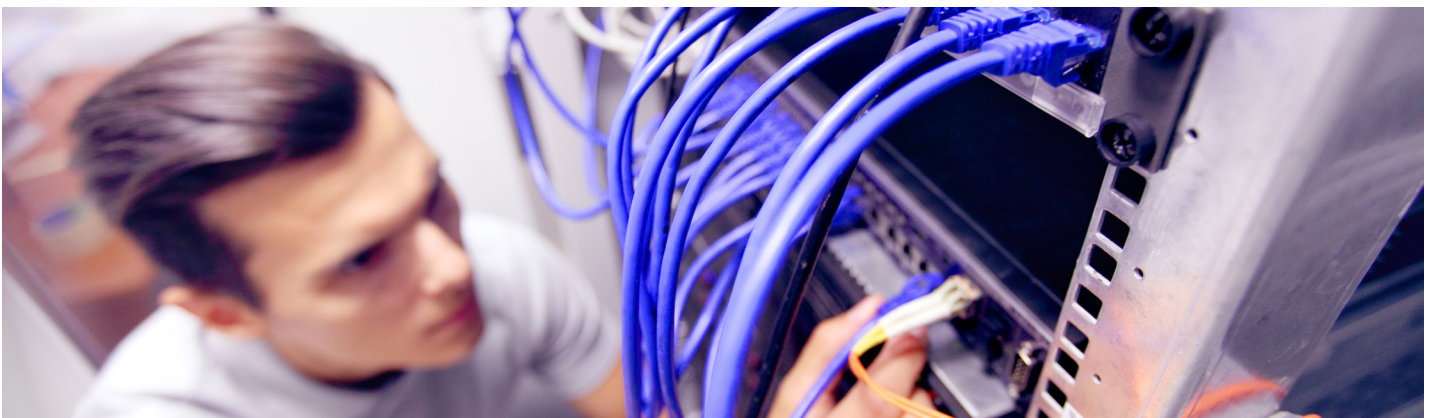
About Loadbalancer.org

Loadbalancer.org's mission is to ensure that its clients' businesses are never interrupted. The load balancer experts ask the right questions to get to the heart of what matters, bringing a depth of understanding to each deployment. Experience enables Loadbalancer.org engineers to design less complex, unbreakable solutions - and to provide exceptional personalized support.