Why Migrate?

Storage Solutions Are Not Created Equal.

Why You Need To Rethink Your SAN Storage Solution

Every organization makes a conscious decision about the type of storage solution to deploy. While many vendors provide myriad solutions from which to choose, those choices can be reduced to only a handful of options, from on-premise solutions such as direct attached storage (DAS), storage area networks (SAN), and network attached storage (NAS) to centralized cloud-based solutions. Many enterprises have distilled the choice down further to just SAN or NAS, finding both DAS and cloud-based solutions to be limiting. Making the choice between SAN and NAS architectures is not a selection of one equivalent solution over another. This white paper explores these two storage solutions, their architectural similarities and differences, and the performance capabilities and feature sets that set one solution apart from the other.

It's a Decision with Far-Reaching Consequences—Don't Make It Lightly

The choice of which type of storage solution to deploy for your organization is a critical one. The two predominant enterprise-grade architectures for on-premises storage are NAS (network attached storage) and SAN (storage area network). Each type of solution ultimately delivers data storage capabilities appropriate for most business needs, along with requisite peripheral features such as data access management, versioning support, backup and recovery, and a host of other capabilities. But don't let those similar outcomes fool you, though. The choice between NAS or SAN involves considerations ranging from speed and performance to deployment and supportability concerns. When you dive into the details, the right choice should be abundantly clear. NAS solutions such as those brought to market by OpenDrives bring superior performance and reliability especially for customers in the media and entertainment industry.

Basic Concepts of Storage

To understand the differences between SAN and NAS architectures and their differing performance levels, you need to have a basic understanding of how storage solutions operate generically. If you want to strip the concept of storage down to the bare minimum, it is the act of writing to or reading from disk. When an application makes a call to retain information outside RAM memory (a computer system's short-term memory), the storage system commits that data to disk. Similarly, when an application recalls that information from long-term storage, the system reads it from disk and sends it to the requesting application. All storage systems. at the most basic level, service these read/write requests.

Labeled another way, each of these actions is an input/output operation, also known as an IOP. IOPs are an important concept, because the totally number of concurrent IOPs a storage system can achieve, as well as the time it takes to complete an IOP (known as latency), directly impact how a system's performance is defined. As you try to determine resourcing for your storage solution, you have to factor in IOPs and the load that client applications put on the storage system. You have to ensure that your storage system has enough IOP resources for all of the client applications to read and write files without significant delays that degrade overall performance. The problem is that determining your needs based purely on IOP capabilities is not necessarily clear-cut, partially due to the fact that IOPs are not all the same. Some are actually more labor-intensive for the storage system than others. You can even look at IOPs in two ways: as a high-level transaction and a low-level transaction. A single higher-level IOP (such as a storage write) can actually trigger multiple lower-level IOPs (many disk writes).

The hardest IOP for a storage system is writing data to disk. Writing data to disk takes the most discrete steps. Let's break down the operation. First, the storage system has to receive the actual data. Then, it must figure out where to put it (determined by pretty complex algorithms), after which it creates chunks of data called blocks, calculates data integrity values known as checksums, writes the data and checksums to disk, recalculates the checksums to verify them, and then informs the client application of the success (or failure) of the write operation. Keep in mind that all this activity happens in a miniscule amount of time, a veritable blink of the eye, but the clincher is that the storage system puts in the same effort regardless of block size. This all happens whether it's a 4 KB block of data or a 1 MB block of data. Incidentally, some people claim that the truly hardest IOP is for the storage system to delete data—but consider the fact that a delete operation is technically a write operation!

As you can imagine, a lot more is going on within a storage system than just reading information from or writing information to disk. Depending on the level of sophistication, the storage solution has advanced data integrity features (a checksum or hash function), smart oversight and coordination of the read/write operations (or scheduling), and a host of other features and functions. Other peripheral concerns include capacity, scalability, and overall throughput of data across the transport medium (typically Ethernet or fibre channel). But at the end of the day, if you are tracking with the above basic description, then the rest of the discussion about SAN and NAS architectures and functionality should make sense.

SAN and NAS Architectures

If you look at storage solutions from a purely outcome-based perspective, they are very similar regardless of architecture. Storage systems retain application data persistently on some form of media (usually disks), service application requests for that data, and perform a variety of error-checking and data reliability functions. All cars move you from point A to point B, just in the same way that all storage solutions house data. But just as with cars, storage comes in a variety of architectures, forms, and most importantly performance. These are the areas where storage systems differ greatly depending on the vendor, the industry, and the uses to which the systems are applied.

In the enterprise-grade storage market, the two predominant architectures are SAN and NAS. While differing variations of each exist, you can understand them best when comparing them at this high level of understanding. Each architecture comprises a different device topology and therefore comes with strengths and limitations. Let's take a look at each one separately, beginning with SANs.

Storage Area Networks (SAN) follow a distributed architecture in which data storage as a whole is centralized, but separated from the servers and client workstations utilizing stored data. The topology of this architecture includes a variety of hardware devices, from the SAN storage arrays and metadata controller to the servers and workstations accessing the data. Interconnecting these various hardware devices is usually a fibre channel transport layer with appropriate switching equipment as well as a dedicated Ethernet transport layer to serve as the local interconnects. The following diagram illustrates a typical SAN deployment (Figure 1).

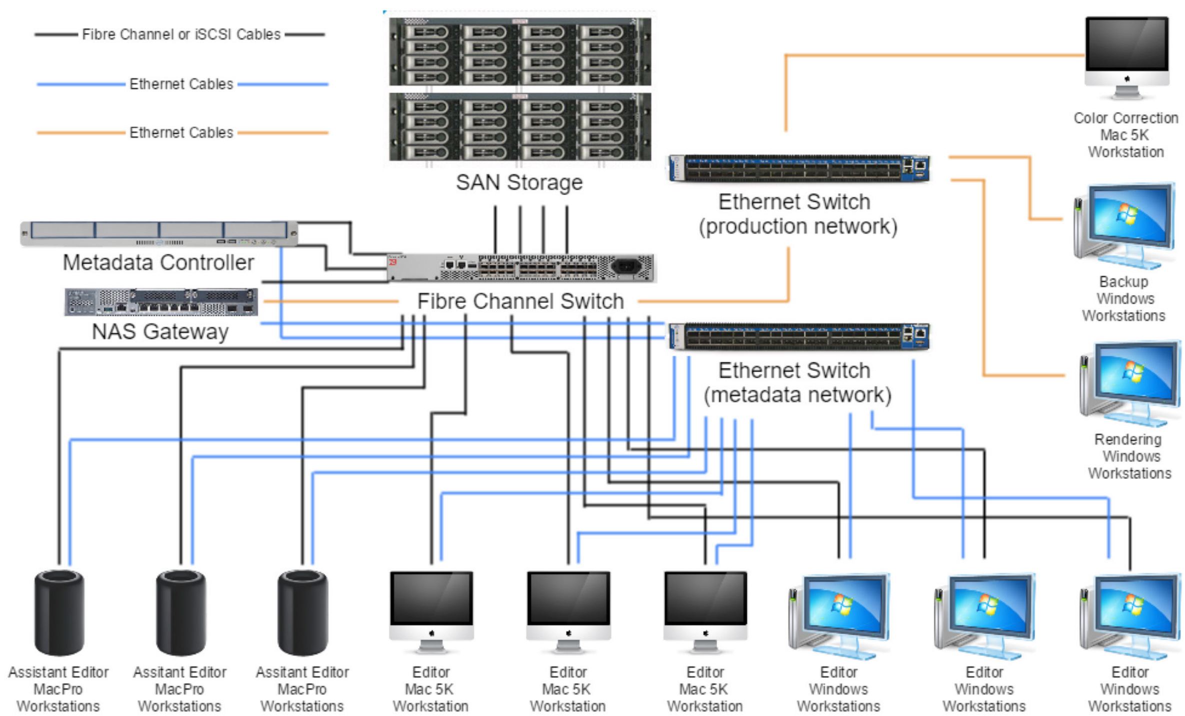

Figure 1. Typical Complex Storage Area Network (SAN) Architecture

By contrast, ***Network Attached Storage (NAS)*** consolidates many of the functions in the SAN architecture to a single device. NAS devices centralize the vast majority of storage-related activities and are usually controlled by a resident dedicated operating system. The topology of this architecture includes a much more streamlined deployment model, one in which client workstations interconnect with the NAS storage device through a common Ethernet switch. Ethernet, obviously, is the common transport mechanism for these solutions. The following diagram (Figure 2) illustrates a typical NAS deployment:

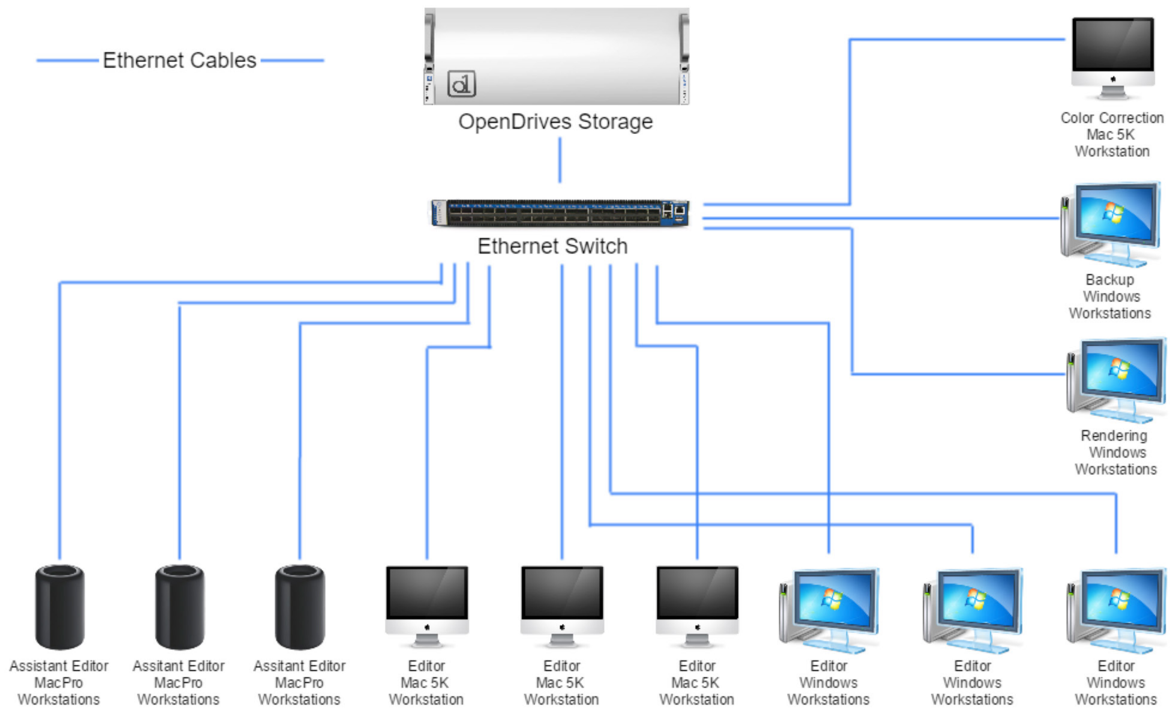

Figure 2. Typical Network Attached Storage (NAS) Architecture

Overall, both architectures depend on interconnection between one or more client workstations and the centralized storage solution through some form of network transport technology. They both service multiple users working simultaneously and often even collaborating with the same data in a shared system of data access. Lastly, both architectures build in scalability and extensibility features so that they can accommodate dynamic and evolving organizations and enterprise needs.

Based on the many ways in which these architectures resemble each other and attempt to perform the same functions for users, you might be led to believe that the decision is a toss-up between competing but equal architectures. You should not, however, draw that conclusion just yet. By sizing them up side-by-side with a focus on the points of differentiation, a clearer picture emerges.

Key Comparisons and Differientiators

Differentiating factors: complexity, transport mechanisms, performance

Let's take a closer look at the ways in which SAN and NAS solutions diverge into very separate and not necessarily equal solutions. Given the architectural models described in the previous section, the one major area of differentiation should be apparent to you already: complexity. When you look at the way SAN architectures are laid out, you can see many interconnections and moving parts. It is a very busy topology, like a confusing set of interconnected roads. This complexity is problematic from both a performance and an operations standpoint. Stated in a slightly more pointed way, SAN architectures are overly complicated, which in turn create many points of potential failure or bottleneck constraints.

The architectural complexity of SAN solutions derives from the evolution of the component technologies themselves. In the not-so-distant past, SAN storage vendors created the technology to allow centralized locality and administration of disks. This solution prevented administrators from having to hunt down workstations and access them individually to troubleshoot and fix. The only thing this didn't address was the ability to collaborate and share the centralized data without transferring between the siloed SAN components. That's where the clustered metadata components come into play. These metadata components ensured that all connected clients could see the data, but they also tried to make sure that only the first workstation to access it had the privileged to edit it. When it all works in unison, it can sustain performance.

Continued development was motivated by real performance challenges, and "pushing the bottleneck around" allowed for performance scalability in the way that devices scaled years ago—typically by adding more disparate devices or replacing devices with newer devices. Remember back to the days when you might have had a flip phone (for calls and texts), a Blackberry device (for corporate emails), a laptop (for work), a console (for gaming), a Zune media player (for music), and a portable DVD player, all dedicated to a similar purposes such as work and entertainment? These were all distributed and dedicated devices that all did their separate jobs and did them well, for sure. But was it easy to travel with them all and have them work in concert? Not so.

Complexity on the scale that most SANs present makes for difficult design, deployment, and operations. By contrast, NAS storage solutions simplify the topology through consolidation of hardware and centralization of control logic, while at the same time enabling even greater flexibility and scalability. A bit of technology history here: NAS devices actually came about after SAN solutions had matured a bit. Early NAS devices basically acted as beefy storage servers attached to the overarching SAN solution. Eventually, they became self-contained solutions capable of providing all the functions of a SAN solution. By relying on advances in all component technology areas, NAS systems can now outperform SANs in the same multi-user environment with far less bottlenecks, points of failure, or difficulty in operating and maintaining the solution. The real question is, when looking at the architectural diagrams above, which one would you—or your IT staff—prefer to maintain? Complex solutions are rarely fun to keep up and running and take a lot of care and feeding.

Another area of differentiation is the network transport layer. Network transport comprises the technologies and physical infrastructure used to shuttle data around the topology between storage and client workstations. Early on, to sustain transport speeds great enough to accommodate many clients in an enterprise environment, SAN vendors adopted fibre channel to interconnect the different systems of the storage network. At that same time, NAS solutions were in their infancy and integrated with the other clients on the slower Ethernet-based network. NAS were often beefy servers attached to those same SANs and were used as gateways associated with the slower performance of Ethernet transport. To put it bluntly, NAS got a bad rap because it was an early adopter of the Ethernet-only movement.

Today, fibre channel solutions are actually a shrinking market. The costs and complexity involved with the initial purchase and ongoing operations and maintenance make it a very expensive transport proposition. Major fibre channel switch manufacturers are seeing decreased revenues from the purely fibre channel part of their overall portfolio. Juxtapose those facts with the vibrant Ethernet industry. Ethernet now offers far higher speeds, north of 400G, with low latency and high-performance specifications. Ethernet is cheaper, more readily available, and is part of the vast majority of enterprise network environments. Seems as though the NAS vendors were a bit more foresighted than the credit given them.

All of these different factors lead to the overall issue of performance. NAS storage solutions now can outperform most SAN solutions without the unnecessary complexity, excess of hardware, potential bottlenecks, and outmoded transport options. What was once bursty and unpredictable, current NAS designs provide turnkey, resilient solutions offering many advantages over the SAN architectural

Benefits of implementing a NAS architecture heavily outweigh SAN.

Some Thoughts on Data Integrity

One of the primary concerns of data storage is the integrity of data which is written to and read from disk. You might think that it's a very straightforward process with little or no opportunity for data mismatching. The reality, however, is that many storage solutions implement different techniques to ensure the integrity of the data written to disk—and some of those techniques are not complete fail-safe measures against data corruption.

The primary way that a generic file system protects against data corruption is through checksum calculations. A checksum is an algorithmic way to ensure data integrity which results in storing both the data and the calculated checksum on the physical storage disk. Most of the time, this approach is effective. However, in certain instances a checksum calculation in and of itself does not guarantee the integrity of the data written to disk. Why? Because most storage solutions have file systems which write both the data and the calculated checksum physically in the same locale on disk. OpenDrives storage solutions prevent this situation and similar ones by physically separating on the disk the calculated checksum from the actual data it represents. This technique adds another fail-safe measure to data integrity.

As a matter of fact, OpenDrives implements a number of measures to ensure the integrity of data. We implement a much more resilient 256-bit method of calculating checksums, which leads to a much higher level of confidence in the checksum calculation than most solutions are able to carry out. Another method is a Copy On Write technique which means that we never overwrite data in place on the disk. Instead, to represent data changes, we write a completely new copy of the changed data to a different portion of the disk before ever removing the original data reference. Again, this measure adds a layer of data integrity that most storage solutions simply aren't able to replicate.

The goal of any storage system is to faithfully write data to and read data from the physical storage media embedded within the storage solution. You would think that, regardless of solution, this occurs 100% of the time. In practice, though, many situations distort or corrupt that process, leading to incorrect data. OpenDrives prevents this through a many-layered approach to data integrity both through the calculation of checksums and the physical manner in which data (and checksum references) are written to disk.

Media Technologies and the need for High Performance

The media and entertainment industry provides a unique environment—and many challenges—for storage solutions. We all know that advances in digital video technologies means higher resolutions beyond high definition and increased dynamic color ranges. The end result is nothing less than stunning visual media.

These elements present another outcome, one not so positive. Better resolution and increased color depth ultimately equate to huge demands on the backend storage solution. These advancements drastically increase the size of files. During post-production work, these much larger file sizes need to be moved, accessed, and delivered. This situation presents (and will continue to present) massive performance challenges for storage systems supporting workflows in the industry.

This is the context in which OpenDrives founders found themselves. Our founders were industry insiders with many years of post-production experience who had confronted these challenges time and again. To keep up with these technical demands, they struggled to get the best performance out of existing SAN solutions, especially as it started to fill up, but they finally came to the conclusion that a brand new storage solution, designed from the ground up, was what the industry needed. So they took it upon themselves to solve the problem first for their own purposes, then their friends and professional contacts, and finally the industry as a whole.

Fast-forward nearly a decade, and OpenDrives is still meeting these challenges by designing and bringing to market the highest-performing NAS solutions to meet these challenges. Each of our storage solutions exceeds the industry demands for greater performance, higher data reliability, and cost-effectiveness. On top of that, all our storage solutions are built around the streamlined NAS architecture. With a three-tiered product line-up, beginning with our active-archive Ridgeview solution through to our mid-range Spectre solution and finally our top-tier Avalanche storage solution, we have every type of customer need covered model. Of course, examining the unique needs of the organization is part of this comparison, too, so perhaps a little contextual industry information along with more specifics about NAS solutions would help.

SAN and NAS systems perform very similar functions—they retain, deliver, and move data within an enterprise. The similarities pretty much stop there, because the way in which they do these things—and their level of performance while doing them—are vastly different. OpenDrives is committed to our NAS architectural design which delivers speed, data integrity, and the ability to expand as an organization's needs grow. Most importantly, our NAS solutions are simple—to deploy, use, and maintain. Even if all other things are equivalent, isn't simplicity better? We think so.

We want to hear from you! Fast and flexible storage solutions are our passion, and helping our customers is our number one priority. Reach out to us if you want to talk about your challenges, see a demonstration, or just engage in tech-talk for a while. We're here when you need us.

www.opendrives.com