

Software-Defined Storage

A comprehensive guide to software-defined storage technology, how it works, and why you should use it to modernize your data storage infrastructure.

What Is Software-Defined Storage?

Software-defined storage (SDS) is an intelligent virtualization software layer that manages and unifies every SAN or NAS device inside your data center, while providing enterprise-level functionality with zero dependence on proprietary hardware.

Just as hypervisors virtualize server hardware for VMs, SDS virtualizes storage area network (SAN) and network-attached storage (NAS) hardware as virtual disks.

The software-defined model is the modern way of re-architecting your storage infrastructure by separating storage software from storage hardware. By switching to the SDS model, you will attain the following benefits:

- **Availability**
- **Scalability**
- **Flexibility**
- **Interoperability**
- **Management**
- **Performance**

In the software-defined storage model, the disks, enclosures, and networking components are all interchangeable, but the software intelligence managing the hardware does not need to be replaced.

Storage virtualization software innovation is always ahead of the hardware. This is why SDS solutions continue to evolve with the development of new functionalities like erasure-coding, deduplication & compression, encryption, cloud integration, object storage, and container storage. The software layer is the secret sauce that makes all the

magic happen. Software-defined storage truly provides an unfair advantage over traditional NAS and SAN devices. When software is decoupled from hardware, customers win. It's that simple.

Why Switch to Software-Defined Storage?

Most companies today continue to struggle with major storage challenges:

- **High costs**
- **Latency issues**
- **Storage downtime**
- **Complex migrations**
- **Consolidation**
- **Managing silos**
- **Exponential data growth.**



That's exactly where storage virtualization through software-defined storage comes to save the day.

SDS is the most modern solution that addresses all of the pain points your organization may be experiencing today or tomorrow. Right now, SDS is the most talked-about storage technology, and everyone is eager to jump onboard.

Regardless of what storage hardware you own today, the software is the brain and layer of intelligence that determines which functions, features, services, and benefits you will be able to offer to your hosts, applications, and end users.

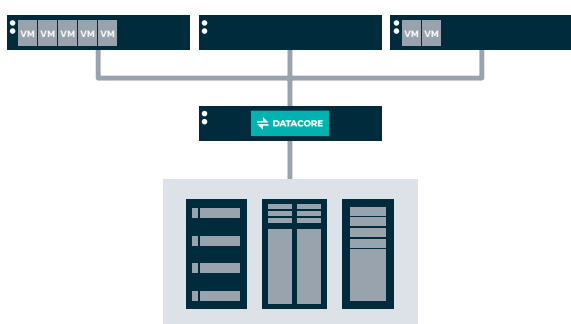
If you find yourself losing sleep at night due to existing SAN/NAS struggles and challenges, then it is time to evaluate software-defined storage. The number of benefits are too valuable to ignore.

Your organization cannot afford to continue depending on traditional storage. It is time to modernize your storage architecture.

How Software-Defined Storage Works

One important reason why companies are switching to a SDS model is the ability to consolidate and manage all existing storage under a single management console, while also facilitating a set of diverse features and functionalities.

SDS is a software controller that virtualizes and manages your physical storage. Essentially all of your disk arrays will be consolidated into a single “virtual pool” and thin-provisioned for maximum capacity. Once the virtual pool is created, “virtual disks” are also created in the pool and then presented to your host servers as raw LUNs to store data.



Imagine mixing arrays from your HPE 3Par, EMC VNX, IBM v7000, or any SAN inside your data center and consolidating them into a single virtual pool. What if you could stripe your application data across all of them regardless of how fast or slow the arrays may be?

You can even configure your SDS solution to automatically move cold and hot data on-demand across the arrays by simply assigning a tier number to each of array. How cool is that?

Natively, the most popular traditional SANs do not support data auto-tiering across competing hardware vendors' arrays. In other words, a VNX cannot auto-tier with a 3PAR or V7000 because they are not able to communicate with one another.

The quick remedy to this challenge is solved by investing in an SDS solution that supports every possible SAN in the world. An added benefit: if you ever need to replace any of the expensive traditional SANs, you will have the freedom to find more cost-effective alternatives.

When it comes to migrating your data from the old SAN to the new one, your SDS platform makes this challenge painless. With just a few clicks, you can start migrating your data from the old clunkers to the new arrays, with minimal downtime to the applications and host OS.

Software-Defined Storage Deployment Models

While there are numerous features and functionalities built into each SDS platform, not every solution executes functions in the same manner, nor are they packaged equally. Some are based on open source code, while others use proprietary software code.

The other difference is the internal storage structure used to manage the data, whether it is based on Distributed File Services (DFS), object, or block-level storage.

That's where you need to do your homework and dig deeper to fully understand how each software-defined software solution works.

In addition, some SDS platforms are offered as physical or virtual appliances. Others can be easily downloaded and installed on your own x86 servers or hypervisors.

This brings us to the next important topic: the types of SDS options. Here are the 3 most popular deployment models:

- **Software-only: Download your SDS platform and install it on any x86 Server or even run it as a VM. This option will allow you to unify and manage any internal- or externally-attached storage.**
- **Appliance: Both the server and software are bundled together with pre-configured specs at a premium price tag. You can also manage both internal and external storage.**
- **Appliance + storage: This is a fully-integrated system that includes SDS, server, and internal storage. Typically these appliances do not support external multi-vendor storage.**

It's important to understand the pros and cons of each of these SDS models. Most end users prefer to go with the option that offers them the most flexibility and freedom to pick and mix the hardware that best fits their data requirements and budget.

COMPARING SOFTWARE-ONLY VS. APPLIANCE MODELS

OVERVIEW

SOFTWARE-ONLY MODEL

By default, the software-only model is the most cost-effective and can adapt to any budget or infrastructure requirements. It's as simple as downloading the software, installing it on an x86 server, connecting your storage, connecting your hosts, and starting the management of your data.

The other use case is the ability to install your SDS software in a VM running on any hypervisor flavor. This enables you to run it as a "virtual SAN" without requiring you to purchase a dedicated x86 server to host the SDS platform.

You still get all the same functions and features found in an enterprise SAN but at a massive discount. This option truly helps you avoid vendor lock-in and gives you the most leverage when it comes to negotiating any future hardware refresh.

Watch how hardware vendors outbid each other for your money and you get to decide which is the most convenient and cost-effective deal on the table. Now that is true freedom from vendor lock-in.

APPLIANCE MODEL

Both the appliance and appliance + storage options are extremely similar. The main differences are that one already comes with storage capacity bundled in, and the bare appliance has no disks. (You must attach your own SAN.)

Each appliance model provides different tiers of compute power, capacity, and connectivity. The various models' options are designed for small, medium, and enterprise-sized companies. Your requirements determine which one is best for you.

Once you decide on a specific appliance model, you can't modify it, which hinders your ability to scale out at a component level without compromising your budget. If you want more compute power or more capacity, then you need to upgrade to a higher appliance model.

COMPUTE POWER

SOFTWARE-ONLY MODEL

With the SDS software-only model, you have full control in choosing a server vendor, whether it's Dell, HPE, Cisco, Lenovo, Super Micro, Fujitsu, Hitachi, or any other brand. You also get to decide how much compute power you want from your CPU and memory.

This gives you an opportunity to start small and gradually scale up as your I/O demands increase. Just pop open the server's cover and add more memory modules to increase your I/O caching performance.

If you need more CPU power, you can simply call Dell and ask them for the upgrade options available for your server and immediately boost up your CPU cores for higher multithreading performance.

APPLIANCE MODEL

The appliance options already come with predefined CPU specs based on the model you choose. If you ever need more CPU power, you will need to upgrade to a higher appliance model that meets your new performance requirements. The disadvantage here is needing to replace the whole appliance instead of upgrading the CPU only.

Memory also comes pre-configured with a fixed amount of RAM that is utilized for different internal tasks, including a reserved amount dedicated to data I/O caching. Even if you were allowed to add 2 more memory modules to boost I/O caching performance, the software may have a built-in limitation that will negate the benefit of upgrading the DIMMs.

STORAGE CAPACITY

SOFTWARE-ONLY MODEL

You decide how much capacity you want inside the box for data storage and how fast you want the drives to be. Maybe you have the funds to go all PCIe Flash and SSDs, or maybe you have a limited budget and must go with a hybrid design by mixing SSDs and HDDs (SAS or SATA).

You can also mix internal disks running inside the x86 box and external disks that are attached via SAS cables, iSCSI, or FC connections. The possibilities are endless.

APPLIANCE MODEL

As previously mentioned, there is an appliance option with storage already included, such as the HPE StoreVirtual SVA appliance. There are also appliances without storage, like IBM's Spectrum SVC appliance.

If you choose a SDS appliance with storage already included, you will need to scale in nodes. This means that each appliance has a fixed amount of disks. Let's say 10TB per appliance, that's considered a node. If you need a 40TB solution, you will need a minimum of 4 nodes to meet your capacity requirements.

This is without taking into consideration the data replica factor. If you want high availability (HA) with 2 replicas of your data, then the requirements for a 40TB solution automatically increase to 80TB or 8 nodes.

The SDS appliance without storage already included gives you more freedom to scale out without requiring you to buy additional appliances. Instead of adding nodes, you simply add iSCSI or FC connected SAN or JBOD enclosures. This is more cost-friendly and easier to upgrade.

REDUNDANCY

SOFTWARE-ONLY MODEL

Industry's best practices are always to have redundant devices and redundant components inside your devices. This is also applicable to your SDS solution. It is wise to get multiple RAID controllers in case one fails.

It is also best practice to have multiple NICs and HBAs to avoid single-point-of-failures (SPOF) and keep your data available even if one of your cards fail.

APPLIANCE MODEL

Pre-configured appliances are limited by the number of redundant components inside the box. Most of them only have a single RAID controller. If you lose it then you will have to wait for the part to be replaced before getting your SDS appliance back online.

The same SPOF risk is applicable to the initiator and target ports where all the I/O flows through. If one port fails, it will take the node offline as well even if the disks are available.

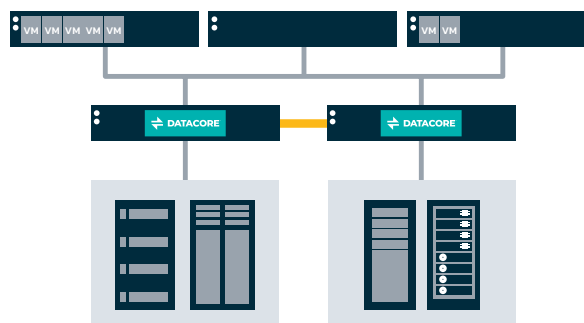
Benefits of Software-Defined Storage

So far, we have identified a fair amount of pros and cons with the 3 most common SDS deployment models. Now we will elaborate a bit more on the general benefits you will gain by adopting a software-defined architecture in your infrastructure.

Availability

Always-on data access is a top priority for every organization whose core revenue stream depends on mission-critical applications' availability around the clock. Software-defined storage meets these requirements by offering high availability (HA) benefits within the same data center or across 2 separate data centers communicating via dark-fiber.

This type of N+1 architecture will require a minimum of



2 nodes. If one node experiences a failure, the second node will continue to host production IO without causing any downtime to core business applications. This type of protection is like an insurance policy that pays off whenever a SAN array fails.

Having 2 active/active software-defined nodes will reduce hardware-related downtime by 100% and will provide a fully-automated failover infrastructure for your hypervisors, databases, and applications.

Scalability

As your application data and I/O performance demands continue to grow, your ability to scale on the fly will be crucial to stay ahead of your competitors. This can only be achieved if you have an adaptive infrastructure in place.

SDS enables you to scale out or scale up depending on the shifts from your business demands. If you need more capacity, you can add new storage arrays to the existing virtual pool and expand your storage capacity immediately.

For more performance, you can add more memory, CPU, or target ports to your SDS nodes and see an immediate performance boost. If your server's hardware specs are maxed out, you can add a 3rd or 4th node, attach some storage, and gain both performance and capacity simultaneously.

Flexibility

Host OS and application IO requirements change faster than most IT departments can keep up with. It is mostly due to the lack of flexibility offered by their storage solutions. Whether it's higher IOPS requirements or

lower latency thresholds, the adjustments must be met at some point.

Businesses want to max out the ROI on their infrastructure investments and are always looking for solutions that can act like a Swiss Army knife of technology. Software-defined storage provides that level of leverage.

With SDS as your primary layer of intelligence, you don't have to continue buying the same storage arrays from the same vendor. You have the flexibility to acquire arrays from different vendors and mix up SSD, SAS, and SATA disks for hot and cold data.

If you need more capacity for hot data, you can invest in SSDs from HP, Dell EMC, IBM, or NetApp. If you need additional capacity for cold data, you can deploy SAS/SATA arrays from Cisco, Super Micro, or Lenovo. You are in control.

Interoperability

Acquisitions and mergers have created nightmares for storage admins when it comes to merging data hosted in multi-vendor SANs. The task to integrate HPE storage, IBM arrays, and Dell EMC storage with each other is impossible because the SANs are not compatible with each other.

This is similar to placing two people together who speak different languages. They simply won't be able to communicate due to a language barrier. Likewise, multi-vendor storage compatibility and interoperability is not a realistic goal unless you add a translator.

SDS is the translator or bridge that can unify heterogeneous storage and place them in a consolidated virtual pool. What once seemed like an impossible task is now possible with SDS storage pooling technology.

Lower Costs

Show me one IT executive that does not care about costs, and I will show a thousand that make their final buying decision solely on costs. Everyone wants to save on CapEx and OpEx, which makes software-defined options very attractive in the long and short terms.

You have learned about the many benefits this software technology offers, and cost-savings is a key factor this solution excels in. How much you can save totally depends on your unique situation.

If you want to repurpose existing storage and avoid hardware rip-and-replace, you will probably see the

biggest savings as you only need to invest in the SDS nodes that will virtualize your storage.

If your situation calls for a storage refresh, you can opt to buy a few x86 servers and fill them up with SSD and SAS drives and add the storage virtualization software. This design will give you a super-fast SDS appliance using commodity disks instead of buying a traditional SAN.

If you have slow storage and need to boost performance, you can buy an x86 server and add a couple SSDs. The CPU and RAM will increase performance, and auto-tiering will shuffle your data between the SSDs and your slow storage. That's an easy way to fulfill performance needs without breaking the bank going all-flash.

Management

The evolution of storage solutions has provided IT admins with a long selection of attractive options. You have new players such as SolidFire, Nimble, Tegile, and old players like EMC, HP, NetApp, and IBM. By the way, the 3 new players I mentioned have been acquired by the bigger players.

As great as feature sets all those storage solutions can offer, they all present one common challenge: you cannot manage them from a single pane of glass. You literally have to open up several management consoles to do your storage administration tasks.

The answer to this problem once again is software-defined storage. Take an old NetApp box and connect to your SDS node. Then connect a Nimble array, add a VNX, throw in a 3PAR, and add them into an SDS pool.

Now you can manage all 4 storage arrays from a single console. You can create snapshots, full clones, or create a DR copy of your data. You can execute every enterprise-level function supported by your SDS node, even if the managed arrays are not licensed for those features. Problem solved.

Performance

Many IT administrators are turning to all flash arrays (AFA) to gain more performance. Although this is a viable option, not everyone can afford it. Even if you get started with AFA right now, it is not a sustainable approach most companies can carry on for long.

SDS solutions can exceed AFA's performance by leveraging 3 core features:

1. I/O parallelism with multi-core CPUs
2. Read/write data caching with RAM
3. Hot/cold data auto-tiering

I/O parallelism enables multiple cores to dynamically participate in processing I/O requests from hosts. If workloads begin to peak, more cores are called on to help handle the heavy lifting. AFA solutions do not have this dynamic parallel I/O processing power.

Read/write caching is the key benefits found in traditional RAID controllers and SAN controllers, yet they support a very limited amount of memory for caching, ranging between 4GB to 32GB. SDS nodes, on the other hand, can support up to 8TB of RAM dedicated for data caching.

SDS auto-tiering provides massive benefits both from performance and costs perspectives. When it comes to the average hot-data changes per day, studies have shown it is less than 5% of your total used capacity. This means that you really don't need flash for all of your data.

The alternative to an AFA is hybrid arrays, which are less expensive and employ auto-tiering to move the hot/cold data up and down the disk tiers. However, you can only auto-tier within the same storage enclosure, not external arrays.

Software-defined storage shines in this area as it enables you to make smart decisions when buying flash storage. Most SDS users invest in 5% to 10% of flash-based tier 1 storage, and the other 90% is a mixture of SAS and SATA arrays.

Related Software-Defined Storage Concepts

Hyperconverged Infrastructure (HCI)

This new concept is founded on SDS technology. It consists of integrating storage virtualization, networking, and the hypervisor—all in a box. It is quickly gaining popularity with small- and medium-sized organizations because it is cost-effective and easy to scale out by simply adding more boxes as you grow.

Hyperconverged can come in different flavors as well. You can use a SDS software-only approach and buy the hardware and hypervisor separately. This gives you maximum flexibility, and there is no vendor lock-in. Perpetual licensing is another huge benefit you gain with this option.

Your second option is to buy an HCI appliance that comes bundled with the SDS software, hypervisor, disks, and networking services. It seems to be easier to deploy, but you lose other benefits, such as maximizing servers hardware configuration. (Appliances usually come with limited hardware configurations.)

Ideally, you want to deploy 24 drives in a 2U x86 server and max out your memory DIMMS up to 1.5TB of RAM capacity. It is completely up to you how much power you want to pack in your HCI server, if you pick the software-only option.

Storage Virtualization

The terms software-defined storage and storage virtualization tend to be used loosely and are normally treated as equal. In reality, storage virtualization is only one part of the whole SDS intelligence stack.

The main benefit of storage virtualization is the ability to place all physical disks into a virtual pool and create thin-provisioned virtual disks. Each virtual disk's data is spread across all physical disk arrays.

Scale-out Storage

One of the initial limitations with SDS solutions was their inability to expand beyond a 2-node architecture. Today, that limit has been removed, and you can scale out to 64 nodes (depending on the SDS product you choose). It is designed to make scaling out easily as your data requirements grow. You can expand your storage capacity by adding one node at a time to the existing SDS cluster.

Secondary Storage

There are many types of secondary storage, but when it comes to SDS, the main types are HDDs, SSDs and flash. By definition, secondary storage refers to any non-volatile storage device that is internal or external. Memory (RAM) is usually considered primary storage.

Primary storage is typically faster and can be directly accessed by the CPU. Secondary storage requires cables or ports for the CPU to access and use it as permanent storage. Disks alone are not fast enough to provide enough performance for today's high I/O workloads created by databases and hypervisors. SDS helps accelerate I/O workloads and data throughput sent to secondary storage.

For additional information, please visit datacore.com or email info@datacore.com



© 2018 DataCore Software Corporation. All Rights Reserved. DataCore, the DataCore logo and SANsymphony are trademarks or registered trademarks of DataCore Software Corporation. All other products, services and company names mentioned herein may be trademarks of their respective owners.