Virtual SAN:

The "New" Storage Meme for Virtual Storage Environments

StarWind Software's Complementary Technology for Realizing the Virtual SAN Vision

NOVEMBER 2014

WHITE PAPER BY JON TOIGO

Managing Principal Toigo Partners International LLC

Chairman Data Management Institute LLC







Introduction: Virtual SAN Definitions Coalesce

"Software-Defined Storage" (SDS) is a term with little agreed-upon substance. Many vendors offer different definitions. To some, it is a product, to others, an architecture. With the introduction of a "Virtual SAN" by VMware earlier this year, and given the weight of that vendor in the realm of server virtualization technology, the definition of what will be classified as SDS became a bit clearer. In fact, definitions of SDS or virtual SAN are beginning to coalesce around three "centers of gravity," as represented by the diagram below.



At the top of the diagram is a class of virtual SANs that might be termed "hypervisor-dedicated SDS." Simply stated, in this architecture, the software for managing storage services is integrated directly into the server hypervisor stack, typically by the vendor that sells the server hypervisor software itself. This is VMware's Virtual SAN offering in a nutshell.

At the bottom of this chart are two alternative definitions of SDS, both of which are distinguished from the hypervisor-dedicated approach by the fact that they are "hypervisor-agnostic." These software products work with more than one hypervisor and, in some cases, facilitate the shared use of storage

resources by workloads operating under multiple hypervisors - or even by workloads that run on physical hardware with no hypervisor at all. As illustrated in the diagram, there are two variants of hypervisor-agnostic SDS: virtual SAN software products that virtualize storage services (enabling administrators to apply services selectively to the data stored on the hard disks and solid state storage deployed behind each server node; or alternatively, storage virtualization software products that virtualize not only the value-add storage services, but also the capacity of the storage itself. Both are alternatives to hypervisor dedicated virtual SANs, but this paper will focus on comparing SDS alternatives that do not involve storage virtualization wares.

As a virtualization administrator or planner, it is important to understand the alternatives for building storage cost-efficiently and in a performance-optimized way behind virtual server platforms. This paper is intended to offer criteria for framing comparisons of the benefits and drawbacks of hypervisor-dedicated and hypervisor-agnostic virtual SAN architectures.

VMware Virtual SAN: Latest in a Line of Storage Experiments



VMware's Virtual SAN, announced earlier this year, is the latest in a line of storage experiments from the hypervisor vendor. By 2010, it was clear that users were experiencing difficulties in making their legacy Fibre Channel and SAS fabrics, as well as their network attached storage and iSCSI SANs work together with the server hypervisor. User surveys demonstrated that storage inefficiency was stalling server virtualization programs.

One problem was linked to the proprietary nature of storage arrays themselves. Each array had its own array controller operating its own value add software and management utilities. As in pre-virtualization days, proprietary storage was creating management headaches and adding expense to infrastructure and administration.

In the virtual server world, "legacy" storage was said to be creating obstacles to desirable hypervisor functions like rapid workload hosting and vMotion. Whenever workload was moved to another server, applications needed to be reconfigured with the new paths or routes to their data in the storage infrastructure. This added delays to the rapid provisioning model promised by workload virtualization.

These issues, however, were dwarfed by the inherent problems with the handling of hosted application I/O by hypervisor software itself. Problems with I/O handling were evident using simple performance monitors that showed CPUs running at top speeds while queue depths (data waiting to be written to storage) were nearly non-existent. Such findings proved that storage slowdowns were not caused by a lack of storage responsiveness, but by a chokepoint created by the hypervisor that was brokering I/O operations.

VMware tried to address its poor I/O handling capabilities, first, by introducing vStorage API for Array Integration (VAAI) in 2010. VAAI consisted of nine non-standard SCSI commands, mostly intended to offload certain storage operations from the hypervisor to array controllers. The company said that this technique would offload up to 80% of storage I/O from the hypervisor's control.

In 2011, the vendor changed its command set a second time, again without advanced notification of either the storage hardware vendor community or of the standards group tasked to oversee the SCSI command language. After some concerns were raised by vendor partners and customers, the hypervisor vendor relented and sought formal approval from ANSI T-10, the committee that oversees the SCSI command language.

In the same year, VMware flirted briefly with building their own storage appliance for use with the ESX hypervisor. The appliance sounded, depending on vendor's speaker one heard, like network attached storage appliance or a simplified storage area network in a crate. However, the actual vSphere Storage Appliance delivered no NAS or SAN services. It was merely a repository for storing virtual machine disk files (VMDKs) and was quickly panned by reviewers.

In 2013, the company began working on a set of software and APIs intended to facilitate the use of commodity storage hardware behind clustered servers. The idea was to create an easily deployed, readily-scaled, and clustered storage architecture similar to the highly scalable storage infrastructure used with high performance supercomputing clusters. Project Mystic produced EVO:RAIL specifications for building hyper-converged virtual computing appliances including an embedded storage architecture called Virtual SAN.

In 2014, Virtual SAN was announced by VMware as a breakthrough storage architecture. (In fact, virtual SAN architecture had existed prior to VMware's adoption of the term. See below.)

At its introduction, VMware spokespersons described Virtual SAN as an integrated part of its uniform hypervisor stack, ESX, where it delivers all of the value of enterprise storage, but at a fraction of the cost of legacy storage and with far less administrative hassle. They claimed that their latest approach, which makes vSphere the broker of storage services, is superior to the legacy NAS and SAN storage that have been used for over a decade.

In the VMware model illustrated below, the Virtual SAN software layer is part of the hypervisor, where it is deployed on a minimum of three hardware/software nodes. The nodes replicate or mirror their data to identically-configured direct-attached storage arrays behind each physical node. Flash storage, usually mounted inside the server chassis, is also included in the Virtual SAN model where it plays the roles of read and write caching. Capacity can be scaled up, by adding more disk to each node, or scaled out, by adding more nodes to a cluster. All storage services, including mirroring and thin provisioning, are controlled by the VMware hypervisor.



While the illustrations seem simple and orderly, the capabilities and limitations of this hypervisordedicated approach to building storage need to be considered more closely. It is at both the macro and micro levels of analysis that clear differences emerge between the value of hypervisor-dedicated and hypervisor-agnostic approaches to building software-defined storage behind virtual servers.

The decision whether to implement storage using the latest experimental approach from VMware must begin with an evaluation of macro level strengths and weaknesses of such an approach. At a macro level, hypervisor-dedicated virtual SANs are a two edged sword. On the plus side,

- Coherent management from a uniform hardware/software stack: a multi-node virtual SAN topology comprising a uniform hardware and software stack – including the hypervisor itself – might well provide more coherent management than a hodge-podge of heterogeneous storage gear in a conventional SAN. Single point of management goes to the heart of lower OPEX costs: a good thing.
- Simplified resource and service allocation: By centralizing all of the storage services that in a legacy SAN are spread out over many array controllers on proprietary storage hardware, you gain the advantage of simplified resource and service allocation to workload.
- Greater efficiency when configuring mirroring and replication operations: With centralized storage software services and uniform hardware, the amount of time required to provision storage can be reduced and data mirroring functions can be implemented more simply.
- Improved availability and fault tolerance: The reason for the (minimum) three node cluster requirement is to build availability and fault tolerance into the infrastructure. With a virtual SAN, workload data and virtual disks are copied from one node's storage to the other nodes in the cluster. That way, if a node fails, the workload can be re-instantiated on a surviving node where a copy of the data already exists, thereby eliminating the need to reconfigure an application so it knows where to find its data if it is restarted on another node.
- Reduced Need for Storage-Savvy IT Staff: Finally, though vendors are reluctant to say it publicly, this arrangement reduces the need to have any storage expertise on staff. Server-side storage is simpler to configure and tools to allocate it are part of the hypervisor. So, your virtualization administrator has all of the skills and knowledge required to operate storage.

The above is a faithful recounting of the key advantages of Virtual SAN articulated by VMware evangelists. It should be noted that each of these "strengths" are actually potential strengths, and their realization is dependent on certain conditions that need to be understood in advance. Lower cost and greater efficiency of operation and administration are by no means guaranteed. And, of course, there are some aspects of this architecture that may be considered weaknesses at the macro level. These include:

 Lock-In (Dependency on a Single Vendor): IT professionals have seen "uniform hardware/software stacks" before and experienced hands are cautious about the consequences. Mainframe data centers in the 1980s, for example, were dominated by a single vendor, producing both the orderliness and discipline that often goes hand in hand with such arrangements, but also some significant downside resulting from a lack of freedom and choice.

- Higher cost products and services: In a lock-in environment, products and services tend to cost more, reflecting the lack of alternatives to dominant vendor proscribed solutions or products. A trade press report in 2014 placed the cost of the software for each node of a VMware virtual SAN at roughly \$2500 per CPU core. For a high-end cluster of three nodes, each having a multi-core processor, the price for Virtual SAN software licenses alone would be in excess of \$20,000, according to the report, while combined nodal hardware would add an additional cost of between \$10,000 and \$20,000. Such an expense may well price the architecture outside of the budget available to smaller companies or to medium and larger firms seeking to implement Virtual SAN in their Remote Office/Branch Office settings.
- Not necessarily best of breed: A proprietary hardware/software stack, by definition, eschews heterogeneous technology, so administrators are limited to the gear that the vendor "pre-certifies" for use in its storage infrastructure. This list may not include best of breed hardware to solve I/O problems. A recent trade press column by a well-known consultant and author noted that Virtual SAN originally touted the lack of a need for a RAID host bus adapter on its hardware kit, but customers later noted that the absence of a well-designed HBA caused minor storage failures to cascade into catastrophic interruption events. Considerable delays accrued as the vendor sought to certify an alternative HBA for use.
- Questionable fit with the realities of the data center: Simply put, data center workload or perhaps better stated as workloads may have varying performance and storage requirements that a uniform, single-source hardware-software stack from your hypervisor vendor may not be able to meet. A homogeneous hardware/software stack will likely create isolated storage islands, storage infrastructure that can only be used by workload virtualized by the dominant hypervisor. That may constitute a significant mismatch with data center realities as analysts see the business IT planners embracing multiple hypervisors for use in the same data center going forward (and also some non-virtualized workload as well). The complexity of managing multiple storage islands will add cost and reduce agility going forward.

From Macro to Micro Issues

In addition to the macro issues that impact the overall efficacy of hypervisor-dedicated virtual SANs, planners also need to consider the specific capabilities and limitations of the solution – its suitability to the work to which it is being tasked. Some of the questions that need to be addressed are listed here.

When evaluating a server-side or software-defined storage architecture for his or her environment, the IT planner will need to see what gear – what types of flash and disk devices – the virtual SAN supports. Are the products that the data center is currently using on a precertification list? Do hard disk drives need to be specially formatted or signed by a certified vendor before use, as was the case in a lot of proprietary disk arrays? Can administrators buy gear off the shelf?

Planners also need to know what interconnects are supported. Do external storage arrays need to be connected via Fibre Channel or SAS or iSCSI, or do they use network sharing protocols like NFS or SMB, or internet protocols like HTTP, FTP or even proprietary cloud protocols like S3? Interconnects may determine the cost and the performance of the virtual SAN solution.

Planners also need to examine into the overall efficiency of the solution, its fit to workload and durability over time. Does the virtual SAN technology use devices smartly, without undo wear? Can the architecture leverage tiered caching, write coalescence, or other techniques for prolonging the life of flash storage? Does the virtual SAN impose constraints on the use of flash or limits on storage capacities, drive types, etc.?

It is also critical to understand what services, exactly, the software-defined storage layer is providing. Recall that a contemporary enterprise storage array usually contains a motherboard that serves as a controller and hosts a number of value-add software features. Depending on the price, these storage services may range from thin provisioning and RAID-based data protection to mirroring and replication and continuous data protection functions. The list can be quite extensive and may include some services that, in your mind, justify the high cost of what is otherwise just a cabinet of commodity disk drives and other parts. Software-defined storage moves these "value add" software features off of array controllers and into a software layer. So, planners need to investigate what functionality the softwaredefined storage controller is actually providing and whether the suite of services covers all of the functionality that they think their workload data will require.

Of course, to attack the high cost of storage meaningfully and to realize the promised agility of softwaredefined storage, best of breed storage resource management is also required. That is, the virtual SAN storage hardware should be instrumented so administrators can get status information at any time on the health of the storage infrastructure, how replication and mirroring operations are performing, how efficiently disk and flash resources are being utilized, and so forth. A real time dashboard is better than a report, W3C RESTful management is superior to management via Simple Network Management Protocol (SNMP), and proactive storage management is superior to reactive. It could be argued that SDS is made or broken by the management it delivers, not only of allocation of software value-add services to workload, but over the hardware kit as well.

Usually, all of the above questions – these micro considerations of the specific features and limitations of the specific software-defined storage solution you are considering – will provide the information necessary to discover actual cost of ownership. Planners need to strive to ensure that the total cost of ownership of the selected software-defined storage architecture is in line with their budgets, so it is advantageous to start calculating known costs early on.

Are We Comparing the Right Thing?

The above discussion of hypervisor-dedicated virtual SAN architecture and its potential benefits reflects what can be called a "straw man" argument that VMware and other hypervisor vendors are using to frame the value case for their wares. "Legacy SAN or NAS" are portrayed as the rival to the software-defined storage paradigm.



Where there is legacy storage in place, some of the vendor arguments have some meaning. For example, trying to do vMotion with a back end legacy SAN will likely be time consuming and perhaps less efficient than simply copying the workload virtual disk file between multiple server-side nodes. Similarly, storage provisioning effort requirements may be greater with a legacy SAN, lengthening delays when provisioning storage to applications and reducing agility. And, management tools are often a bit less coherent in a heterogeneous SAN environment. So, a fairly compelling case can be made for using

hypervisor dedicated software-defined architecture rather than legacy FC or SAS fabrics or even generic NAS for storage behind virtualized servers.

Finally, the legacy storage environment creates the need for skilled storage administrators – IT functionaries who understand the nuances of storage and have the requisite judgment and skill to adjust the knobs and turn the levers in the storage infrastructure to tune storage to specific application needs. But, in many companies, the move to virtualize infrastructure has been driven in part by a desire to reduce IT labor cost and software-defined storage is supposed to enable the elimination of positions like storage administrator by simplifying storage.

However, just comparing virtual SAN to legacy SAN doesn't necessarily support the choice of a hypervisor-dedicated SDS architecture. In some companies, virtualized infrastructure is being deployed green field – meaning without precedent or prior investments in legacy storage at all. In such cases, and one could argue in all cases, the real comparison that needs to be made is between hypervisor-dedicated and hypervisor-agnostic SDS solutions.



The real comparison software defined data center planners should be making today is between the hypervisor-dedicated and the hypervisor-agnostic SDS architectures. There are actually a couple of different hypervisor agnostic approaches to building SDS architecture, including one that simply virtualizes storage resources and serves them up as virtual volumes that can move with workload from machine to machine. However, most of the products in this category instead deploy a technology that is similar to the hypervisor-dedicated approach, but in a hypervisor-agnostic way.

The essential difference between dedicated and agnostic approaches is that the hypervisor agnostic model complements the functionality of the server hypervisor software, and provides a different way to

deliver storage and storage services to workload. In most agnostic implementations, storage is treated like another hosted application, virtualized and running above the hypervisor kernel.



As suggested by the illustration, you still set up nodes with ongoing data replication between them. Only, with the hypervisor-agnostic approach, there may be some micro-level differences that affect what gear and interconnects you can use, how efficiently the gear is used, and how readily storage can be deployed, scaled and managed. The different price tag on the agnostic approach, both from a CAPEX and OPEX standpoint, might actually be more in-line with budget realities, too.

For the record, hypervisor-agnostic virtual SAN architecture predates the 2014 offering from VMware by about five years. In 2009, StarWind Software had already developed a virtual SAN storage stack from scratch that was designed to co-exist with a server hypervisor on the same server hardware where the hypervisor was running. Apparently, they failed to get to the trademark office and left the door open for others to use the "virtual SAN" moniker, but they did offer the technology as a product commencing in 2011. Since StarWind has bragging rights to the original concept of virtual SAN, it may be worthwhile to compare their wares to the VMware virtual SAN – which is a dedicated approach to delivering SDS services.

At a Macro Level

VMware Hypervisor Dedicated



- Single hypervisor support
- Creates storage environment only for VMware data
- Expensive to deploy: unfriendly to smaller users, ROBOs
- Optimized for disk storage

StarWind SW Hypervisor Agnostic



- Support for multiple hypervisors
- Creates block and file storage environment that can be shared between different hypervisors
- Significantly lower cost
- Flash-friendly

At a macro level, some of the features touted by VMware around their Virtual SAN technology viz a viz legacy SANs are a bit less compelling when you compare them to a hypervisor agnostic SDS architecture like the one offered by StarWind Software.

For one thing, the unified and integrated support of VMware's Virtual SAN technology actually seems a bit brittle next to the StarWind Virtual SAN® given the capabilities in the StarWind solution to support the storage requirements of multiple hypervisors. Thinking in terms of our overall environment, it is generally preferable to avoid the problem of creating and having to manage multiple "islands of storage" – each dedicated to a particular hypervisor. While some IT shops have selected VMware as their hypervisor software standard, it may still be advantageous to keep options open to enable the deployment of alternatives that better meet the needs of specific applications going forward. StarWind's Virtual SAN provides multi-hypervisor support and complements server hypervisor operations with a block and file environment that can be shared by different workloads running under different hypervisors.

Moreover, the VMware approach, as mentioned earlier, carries an expensive software licensing cost burden, combined with a minimum three-node configuration requirement leveraging pre-certified hardware components. In light of these facts, many commenters have suggested that VMware has priced itself out of small offices and remote or branch office environments where cost-per-node for the vSphere approach would be a show stopper. By contrast, StarWind's Virtual SAN facilitates the use of less expensive, consumer-selectable hardware and is less expensive from a software licensing standpoint than VMware. Moreover, the greater hardware utilization efficiency of StarWind – especially with respect to flash storage – would improve return on investment and help contain costs in a manner that would appeal even to larger shops with dreams of an all-silicon data center.

Applying the "micro" level criteria suggested earlier, StarWind Software's Virtual SAN delivers significant value as an alternative to VMware's proprietary Virtual SAN technology.

Hardware support: VMware claims to support a broad range of disk and silicon-based componentry, but administrators need to confirm that the products they prefer are on a certification list. Even then, reports are coming in of problems getting even the company's "Ready Node" products to work with servers and their host bus adapters. Moreover, support for flash devices is somewhat limited and scaling the capacities of nodes has more to do with hard disk, than flash storage. It could be argued that VMware was disk-centric when it was defining its software-defined storage play, an observation that is further born out by the inefficient way that it uses flash memory in write caching.

By contrast, StarWind Software's Virtual SAN supports a broad range of storage devices including SAS and SATA disks, PCIe and NVM flash devices, and can scale using either or both.

Lack of Efficient Utilization: VMware's Virtual SAN uses flash very inefficiently as a write cache. Depending on the virtual machine I/O patterns, it is entirely possible for flash memory devices used as write caches to be hammered constantly with a lot of small writes from hosted apps. This reduces flash storage life, given the erase-write nature of flash. Specifically, each small logical block write requires a whole (and comparably big) page of flash memory devices, which have a limitation of approximately 250,000 writes per cell. When a cell dies, a group of cells must be retired, and ultimately the entire device must be replaced at considerable expense.

An alternative to writing a lot of small blocks to the flash device is to stage them first in a wear resistant memory cache. StarWind Software can do this using Dynamic RAM (DRAM) and a technology called write coalescence. StarWind leverages log structuring technology with the DRAM cache, enabling small block data to be aggregated into bigger block pages that are written much less frequently to flash, resulting in less wear. This write coalescence and log structuring is part of the I/O handling built into StarWind's Virtual SAN architecture and is a significant advantage over VMware, especially for firms that are pursuing all silicon data centers.

The Right Set of Storage Services: As noted above, planners also need to evaluate their SDS architecture option for the value-add storage services that it provides and its adequacy given the needs of hosted applications and overall system performance and economy. StarWind Software's virtual SAN offers a robust set of services via its software-defined storage controller. Worth mentioning is its in-line

deduplication technology, which is on par with some of the best on-hardware based implementations of de-duplication in the market at present, providing a way to reduce the volume of data to be replicated between nodes and the space occupied by copies of data made for backup purposes.

Conclusion, for Now

With its hypervisor-agnostic Virtual SAN, StarWind Software provides a potentially more efficient and less costly solution to server-side, software defined storage than does its VMware, with its hypervisor-dedicated technology. Planners working toward the realization of an agile software-defined data center should include an evaluation of the StarWind alternative that can facilitate the realities of the contemporary enterprise data centers in terms of workload diversity.

For the smaller firm, or the remote office/branch office, Starwind Software's virtual SAN needs to be given close consideration for its greater flexibility in node hardware and lower cost software licenses.

While some IT planners like the idea of "one throat to choke" or single vendor solutions, the cost of uniformity in terms of lock-in and poor fitness to purpose should give pause. What is needed is a business-savvy storage solution in the virtual server environment. StarWind Software is worth a closer look.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior written consent of StarWind Software.

StarWind Software gives credit to Jon Toigo and confers rights to use the current document as it is, unmodified, for public purposes.

Contacts

US Headquarters	EMEA and APAC
 1-617-449-7717 1-617-507-5845 	 +44 20 3769 1857 (UK) +49 302 1788 849 (Germany) +33 097 7197 857 (France) +7 495 975 94 39 (Russian Federation and CIS) 1-866-790-2646
Customer Support Portal: Support Forum:	https://www.starwind.com/support https://www.starwind.com/forums

sales@starwind.com

info@starwind.com

In 2016, Gartner named StarWind "Cool Vendor for Compute Platforms".

Sales:

General Information:

Gartner does not endorse any vendor, product or service depicted in its research publications, and does not advise technology users to select only those vendors with the highest ratings or other designation. Gartner research publications consist of the opinions of Gartner's research organization and should not be construed as statements of fact. Gartner disclaims all warranties, expressed or implied, with respect to this research, including any warranties of merchantability or fitness for a particular purpose.

\approx StarWind

https://www.starwind.com StarWind Software, Inc. 35 Village Rd. Suite 100 Middleton, MA 01949 USA ©2014, StarWind Software Inc. All rights reserved.