SOFTWARE-DEFINED STORAGE: Reshaping The Storage Landscape

Software-Defined Storage (SDS), a set of technologies that allow for greater management, agility, and efficiency in storage, is a key building block for future datacenters.

Enterprises should examine how SDS can be applied across the areas of management and orchestration, block storage, VSANs, and file and object storage.

Planning for software-defined storage can help organizations liberate themselves from the added burdens of legacy storage architectures while lowering costs and simplifying management.

The datacenter has changed fundamentally since the 1990s, when large processing and storage resources first became entrenched throughout the enterprise world. Two decades ago, that world was fairly homogeneous, with individual datacenters built around a single vendor or architecture. Today, enterprise storage is a much more robust, fine-grained, and complex patchwork of vendors and platforms. Viewing the modern landscape through the filter of 1994, seeing storage management as a single-platform, single-console affair is a recipe for poor performance and significant waste. The industry needs to find new ways to embrace the potential of new storage developments and craft tools to make these developments more intuitive and effective.

Today's top datacenters amply demonstrate that commodity, or off-the-shelf, hardware is quite capable of delivering greater ROI than legacy-oriented, proprietary storage solutions when implemented with appropriate architectures and management. Google arguably stands as the best-known example of a massive data architecture based on commodity components. As this trend away from proprietary systems continues, software-defined storage (SDS) is quickly emerging as the way for enterprises to meet their next-generation storage needs within the confines of current budgets and capabilities.

This paper provides an overview of SDS technology and aims to inspire readers toward beginning SDS deployments in their own operations.

What is software-defined storage?

The current pressure for more flexible and scalable datacenter resources is intense. A recent **Digital Universe study**[1] predicts that, in 2014, 1.7MB of data will be created each minute for every person on Earth. Behind this and many similar trends sit ever-growing ranks of datacenters, all of which need to support these solutions with storage.

Leading industry analyst firm **IDC notes**[2] that SDS "should offer a full suite of storage services and federation between the underlying persistent data placement resources to enable data mobility of its tenants between these resources." In effect, SDS virtualizes storage and delivers it as a service that can be quickly scaled, protected, and moved as changing needs dictate.

Like other facets of virtualization, SDS has few clearly defined boundaries and can become something of an empty buzzword in the wrong hands. Still, software-defined storage has a handful of key attributes that differentiate it from traditional hardware-based storage:

- Commodity hardware. Many proprietary storage solutions emphasize the benefits of their custom ASICs for high-speed storage processing and RAID computation. However, much of the need for this processing can be alleviated in well-planned SDS implementations, and modern CPUs are more than able to shoulder the processing loads of legacy architectures, including in virtualized environments. The use of commodity hardware can save both vendors and end-customers substantially on the cost of storage deployments.
- Assumed unreliability. Another perceived advantage of legacy NAS and SAN systems is their higher level of "enterprise-grade" reliability. Proprietary solutions typically feature top-caliber, expensive, thoroughly qualified hardware, and the software tied to those products assumes that underlying hardware is reliable. With SDS, Murphy's Law becomes the assumed norm, not the exception. Hardware is expected to fail, potentially in multiple ways, and the software architecture plans for this.

- Superior resiliency. Since SDS assumes failures will happen, software is designed to help self-heal the system. This results in fewer "emergency" hours logged in by IT and more ability to remedy issues in a time-efficient, planned manner. Building infrastructure for resiliency rather than reliability reduces operational burdens, as IT can address failures such as dead hard drives on a planned schedule without impacting application performance.
- Multi-tenancy and self-service. In situations where multiple users share a common infrastructure, such as four servers with one connected storage pool, management under legacy architecture can be cumbersome. An admin might need to log into the storage array's management console, create a new volume every time a new user is added, set access privileges, and so on. If the resource demands of these clients change, the admin must make manual adjustments. With SDS, though, APIs connect the storage pool to a common console, such as Windows Azure Pack UI, VMware vCloud Director, or OpenStack Horizon. The console leverages the APIs to help the admin create automated policies, automatic volume resizing, snapshot creation, and so on. This greater convenience and added automation capability with SDS-driven multi-tenancy further reduces costs.
- Agility. Because of its software-based, virtualized nature, SDS can be provisioned and deployed literally in seconds. In contrast, requisitioning and setup of legacy storage systems often takes weeks. Storage agility allows IT to respond quickly to changing business needs and user demands.

These points form the backbone of why at least **one analyst report**[3] pegs the software-defined datacenter (SDDC, a superset of the SDS market) at over $5 billion by 2018, up dramatically from 2013's $396 million. Properly implemented SDS more than pays for itself, and enterprises are taking notice. In fact, one **Forrester Consulting study**[4] found an 81% ROI over three years for one evaluated SDS solution.

The benefits of SDS are too compelling to ignore. Enterprises should begin considering how to start integrating SDS approaches within their operations. These plans are likely to encompass any of three main areas.

Management and Orchestration

From a single server to multiple server racks to entire datacenters, IT must manage storage. Managing storage is a significant operational burden to CIOs and IT administrators. Management covers everything from setting up the physical hardware to provisioning storage for applications, monitoring performance, and replacing failed components. On the management software side, management tools abound for IT workers. From VMware to Microsoft to open platforms such as OpenStack, all can provide excellent capabilities to admins.

The trouble is that while organizations may have one virtualization platform at its server foundation, the company may have accumulated several storage solutions from different vendors as time and shifting priorities have dictated. Not only will these multiple solutions each have different management mechanism but they may target storage functions that are misaligned with current business needs. Repurposing existing storage for new business applications is possible, but may require costly application rewrites, downtime, or require using storage that isn't ideally suited to a given task.

The industry sorely needs new tools which have the ability to unify these legacy and newer, software-driven storage solutions under a single management console. Different vendors are working toward products that will be able to deliver this functionality, but so far the niche remains immature and under development. When a vendor finally does nail down a universal management tool or standardized API, the advantages to datacenter operators will be enormous.

Also in this management vein, the industry needs standardized interfaces that can let storage hardware report its capabilities to higher layers for more optimized running of services while simultaneously allowing applications to send requests down to hardware. Today, this type of inter-layer functionality is very limited. But imagine if a latency-sensitive application, such as a transactional database, could "ask" the IT infrastructure for low-latency storage. With the right software-defined storage capabilities in place, the appropriate storage resources could be configured, provisioned, and delivered to a database, all without IT staff intervention. Again, the advantages for performance and ease of management could be very significant.

The point to remember is that software defines how hardware is managed and what services it can accommodate. But software can only do this if the underlying layer capabilities are exposed in the APIs. Avago and other industry partners are currently working to expand APIs tied to today's leading storage software and hardware to make such next-generation functionality a reality.

Block Storage and VSANs

Storage area networks (SANs) provide block storage technology, wherein data is stored in logical volumes containing data sequences of a given fixed length (the block size). For decades, block storage has been and remains the backbone of database management systems. The approach also remains very popular in virtualized and cloud-like environments.

Most SANs take the form of a physical storage array, often an appliance, connected to one or more servers. Based on Fibre Channel, iSCSI or Fibre-Channel-over-Ethernet, SANs generally offer high performance, but they are costly both to buy and to manage.

Alternatively, a virtual SAN (VSAN) is a piece of software that resides on multiple servers and ties the storage within each of them into a common pool exposed to the hypervisor running on those systems. Products such as LSI Syncro CS provide a quick upgrade path for existing servers, outfitting them with sharing and resiliency features while shedding network overhead burdens. A VSAN infrastructure driven by high-performance storage controllers will yield faster writes as well as more effective peering between flash and hard disk resources, often realizing even higher solution performance than a conventional SAN.

In essence, a VSAN offers the storage flexibility and utility of a conventional SAN but with the added cost savings and resiliency of virtualization. With a VSAN, there are no appliance costs or expensive proprietary solution licensing. While the VSAN approach might not be suitable for all block-based storage use cases, and traditional SAN-based deployments will continue to be leveraged for performance-driven applications, VSANs are gaining popularity and are expected to grow in the small-to-mid sized business and remote office segments of the market.

Object Storage

Many cloud and hyperscale datacenter applications rely on object storage. Most simply, object storage treats files as things to be dropped in buckets. There are no block sizes to mind. The admin doesn't have to worry about managing volumes. An object storage file system merely uses metadata to keep track of which objects went in which buckets and retrieves them when needed. This concept can be scaled from a single appliance to redundant datacenters. Also, because object storage is free from many of the metadata restrictions of block storage, there is much less total management overhead and object content can be much larger than block storage, easily spanning across many discrete drives or even logical volumes.

Object storage relies on a key/value method for placing and retrieving files. In turn, key/value functionality is enabled through an API able to expose object storage in a wide variety of scenarios. Across the storage industry, vendors are working to bring open storage APIs into play. The Storage by LSI division of Avago, for instance, is partnering with several object storage vendors to provide means to accelerate object storage by offloading complex storage-related tasks through its intelligent storage adapters — again, forming new bridges between storage layers and enhancing performance in the process. One of the opportunities now facing storage vendors is the need to consistently advance their data protection algorithms for key/value object frameworks.

Why Change is Needed

Amidst exploding storage demand, set in a digital universe **expected by some**[5] to reach 44 zettabytes by 2020, legacy storage solutions continue to sell. Unfortunately, the costs tied to legacy approaches, if they are to keep pace with demand, are quickly rising to points the market will not bear. The industry needs more efficient storage solutions that won't sacrifice performance or data safety.

This efficiency will be achieved in large part through decoupling, or abstracting, physical resources from the services they provide. We see this happening in converged infrastructure's segregation of storage, network, and compute resources. We see it in the liberation of storage capabilities from yesterday's static SAN and NAS environments. Decoupling of resources thrives on commodity hardware platforms running software that adds new capabilities across three areas: management and orchestration, block storage for virtualized deployments, and scale-out file and object storage.

These three vectors will help propel enterprises to make the most of their IT assets, and companies such as Avago are making the investments today to enable this more scalable, democratized storage future. With software-defined storage as a part of an overall software-defined datacenter, applications will be more responsive, business can respond more quickly to changing needs, and the ROI of datacenter storage will dramatically climb.

1. EMC Digital Universe Study, data by IDC, 2014. http://www.emc.com/leadership/digital-universe/index.htm
2. "IDC Brings Clarity to Software-Based/Software-Defined Storage Markets," http://www.idc.com/getdoc.jsp?containerId=prUS24068713
3. http://www.prweb.com/releases/software-defined-storage/sddc-market/prweb11551907.htm
4. http://finance.yahoo.com/news/forrester-research-market-study-confirms-160000541.html
5. http://www.emc.com/leadership/digital-universe/2014iview/index.htm