IDC
Analyze the Future

I D C A N A L Y S T C O N N E C T I O N

Ashish Nadkarni
Research Director, Storage Systems

Software-Defined Storage: A Fundamental Shift in How Storage Is Delivered and Used

June 2014

The commoditization of hardware platforms — and the relatively slower time to market for (and higher development costs associated with) storage products based on custom hardware components — has compelled suppliers to shift the focus of their innovation to delivering their value proposition via software-defined solutions. This change initially started in the computing domain and is now sweeping the storage and networking domain. The era of the software-defined datacenter has a profound impact on how storage platforms are classified — and more importantly how this classification alters the landscape, creating a new breed of software-defined storage (SDS) platforms.

The following questions were posed by HP to Ashish Nadkarni, research director of IDC's Storage Systems practice, on behalf of HP's customers.

Q. What is software-defined storage, and how is it different from hardware-defined storage?

A. First, software-defined storage is a set of platforms that deliver a full suite of storage services with a software stack that uses, but is not dependent on, commodity hardware — that is, server hardware built with off-the-shelf components. For anything to be software defined, the system should not contain proprietary hardware like custom ASICs, chipsets, memory, CPUs, or anything custom. That would disqualify the platform.

Second, software-defined storage software should not make any underlying assumption about the customized hardware components. The storage software should not force you to buy a specific type of customized server hardware and/or internal components, even if it is commodity.

Third, it should be able to run on different server hardware platforms. By not making that software assumption, you're essentially allowing the system to run on multiple, different instances of hardware not prescribed by the vendors themselves. Buyers should be able to procure their own hardware and then install and run it.

Fourth, the solution itself should be autonomous or standalone, meaning that it's a complete system in itself. It should provide both northbound storage services and southbound data persistence without requiring any special type of hardware or software. A true software-defined storage platform is a standalone platform in its entirety.

Q. What are the benefits of SDS?

A. There are several benefits, starting with the fact that SDS decouples the software from the hardware life cycle. By that I mean, in a hardware-defined platform, typically a mix and match of multiple generations is not allowed. Hardware upgrades are an all-or-nothing proposition, and furthermore such upgrades are intricately tied to the software that is supported on that platform. By decoupling the software from the hardware life cycle, customers can stretch the value of their investment, by running, mixing, and matching server instances and generations. In other words, they can run and federate newer software on multiple older hardware instances as necessary, for greater investment protection. Customers can liberate themselves from the need to upgrade all of the hardware every time there's a software upgrade.

Another benefit is that customers have flexibility in procuring the storage platform via multiple delivery models. They can run it procured as hardware instances (appliances), as software only (which means the customer chooses the hardware instance), or as virtual machines, etc. Customers are not forced to run the system in a certain way, which gives them more flexibility, more agility, and more efficiency, too. They're able to leverage existing investments and gain cost savings. SDS allows customers to homogenize the datacenter at the software layer.

Q. What are the features that buyers should look for when evaluating SDS solutions?

A. Buyers should look for scalability because they don't want to create storage islands. They should also look for federation capability that allows them to create a large-scale solution that aggregates disparate storage sources into a single pool and has data mobility within that pool. You want the ability to start small and grow linearly as needed.

You also want the ability to procure your own hardware and hypervisor from the vendor of your choice, as well as the ability to run the software platforms you want.

Another feature to look for is the ability to run a truly heterogeneous configuration (in which multiple code versions can be configured to run as one) or the ability to perform rolling upgrades. You should have the ability to mix and match hardware types, as well as the ability to access data via file, object, or block interfaces. This provides the ability to mix and match applications on a single platform.

Q. What are the different delivery models for SDS that buyers can choose from?

A. Suppliers have come up with multiple delivery models for their storage platforms. These include virtual storage appliances (that run as virtual machines in a hypervisor environment), cloud-based pay-as-you-go models, software-only models, and hyperconverged software or appliances (in which the compute and data services are adjacent to each other).

In a virtual storage appliance model, the SDS platform is delivered as a virtual machine to the buyer. Most suppliers will support commercially popular hypervisors.

A cloud-based pay-as-you-go model is just what it sounds like, with the supplier often hosting the SDS platform on a public cloud infrastructure and offering it as a service to the buyer. This is becoming a popular delivery model as more and more businesses leverage the cloud for on-demand scalability. The service can also be delivered as a virtual private cloud.

In the software-only model, the supplier makes the SDS platform available as software that can be downloaded or purchased. The buyer then installs it on a hardware platform or virtual server of its choosing.

Hyperconverged software or appliances refer to a relatively new delivery model based on the premise of running the storage environment in a hybrid manner — meaning that it's a hybrid compute and storage environment. By using file-, block-, or object-based data organization in a creative manner, these platforms can allow workload localization on the storage node, minimizing the need to move data to the compute layer. In order for a hyperconverged platform to be software defined, it should not make use of any custom hardware inside the appliance. In other words, the appliance-based delivery model is designed for selling software-defined storage in a manner in which businesses are used to procuring storage.

Q. What should buyers look for in an SDS supplier and/or service provider?

A. First, you want to look for platform openness. Gone are the days of proprietary APIs and customized applications. For an SDS supplier to be successful, it must support open standards and interfaces (e.g., OpenStack).

Second, you want a supplier that offers robust data services specific to storage independent of underlying hardware capabilities, such as integrated snapshots, thin provisioning, multisite replication, and disaster recovery. You also want investment protection, with the ability to change underlying hardware and hypervisor at a later time. Similarly, you should be able to easily upgrade to dedicated storage arrays if needed.

Third, you want a supplier that offers a fast SDS platform. Many older-generation platforms were not designed to be fast. And by fast, I mean that the SDS solution must match or exceed hardware-based platforms in performance.

Finally, you want your supplier's solution to support workload adjacency. We are fast entering a world where data and compute cannot reside separately, no matter how fast the network is that connects the two. In the era of big data and analytics, it is often necessary and efficient to make the analytical workloads run very close to where the data lives. This means your supplier's SDS platform must accommodate compute workloads in a way that the compute workloads have data affinity or localization.

ABOUT THIS ANALYST

Ashish Nadkarni is a research director within IDC's Storage Systems research practice. He provides detailed insight and analysis on evolving industry trends, vendor performance, and the impact of new technology adoption. Mr. Nadkarni is responsible for producing and delivering timely, in-depth market research with a specific focus on file- and object-based storage, storage for and in the cloud, and storage for Big Data and Analytics. He also coleads IDC's Big Data Global Overview program, in addition to managing the infrastructure component of that program.

ABOUT THIS PUBLICATION

This publication was produced by IDC Custom Solutions. The opinion, analysis, and research results presented herein are drawn from more detailed research and analysis independently conducted and published by IDC, unless specific vendor sponsorship is noted. IDC Custom Solutions makes IDC content available in a wide range of formats for distribution by various companies. A license to distribute IDC content does not imply endorsement of or opinion about the licensee.