Neuralytix™
Strategies for tomorrow

SDS: The Road to a Hybrid Cloud Data Strategy

Table of Contents

Introduction

In this *Insight*, Neuralytix analyzes the characteristics of software-defined storage (SDS) and its impact on the data-driven economy. In particular, this Insight will provide a deep dive of IBM Spectrum Software Suite, a SDS solution, and how it addresses the business needs of the modern enterprise.

Neuralytix estimates that through 2020, 30% of net new IT projects will be built on some form of converged infrastructure with SDS software. Our position on SDS is a strong one – we believe that all organizations to plan to include SDS into their IT strategy, period.

What is SDS?

A common misconception with respect to software-defined storage (SDS) is that it is a singular *product*. Software that one installs to provide the features and functions expected from SDS. To an extent, this is correct. SDS must relieve the enterprise from having to use multiple point solutions, that often come from multiple vendors.

Neuralytix research suggests that close to 100% of all organizations have some formal (or shadow IT) in-cloud storage.

We further estimate that roughly 95% of all large organizations are looking for a hybrid storage strategy that includes cloud-based storage, while roughly 80% of mid-market customers are looking at how to integrate cloud-based storage with on-premise storage.

In either cases, our estimates suggest that roughly 55% of organizations are actively seeking to start implementing a SDS strategy within the next 12 months, with roughly 25% of organizations desiring to start implementing an SDS strategy within the next 36 months.

As organizations look to the cloud and reducing the dependency on dedicated, physical storage, SDS provides a flexibility and consistency to span different deployment models with a consistent operational experience However, SDS is, in reality, more of a framework. Within that framework there a many moving parts and dependencies.

That framework include quite a number of functions:

- Storage virtualization of on-premise and in-cloud storage capacities;
- Orchestration of all available resources, including cloud based resources;
- Scalable file system that can stretch into the cloud;
- Data services, such as, presenting storage in various protocols. For example, iSCSI, S3, HDFS, NFS, SMB, etc.;
- Data tiering within and across on-premise and in-cloud storage;
- Integration of on-premise and in-cloud data; and
- Quality of Service.

Tying all these functions together to create a cohesive, and perceived "single piece of software" requires a vendor that has had experience in all aspects of the SDS framework.

One simple way of considering what SDS does is to think of it as *data virtualization*. Essentially SDS "liberates" the data from the hardware infrastructure. In doing so, the business can focus on the value of the information that can be derived from the data.

Figure 1 illustrates the basic interactions between SDS and the hardware infrastructure around it:

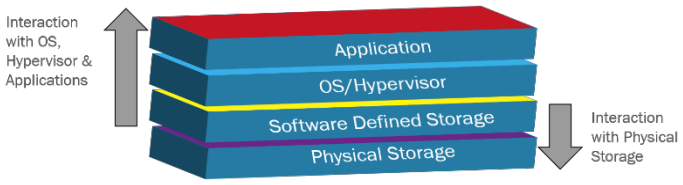

Figure 1: Block Diagram of SDS Interactions (Neuralytix 2016)

One key element that is consistent across all SDS software is that it can, through a single "pane of glass," manage all the interactions between the application(s) and the operating system or hypervisor; and the physical storage capacity.[i]

This makes data management immeasurably simpler. No longer does an administrator have to associate specific applications with specific storage systems that leads to data silos, and a suboptimal use of the overall storage capacity. Additionally, in more advanced SDS software, analytics can help to ensure that data placement is such that SLAs such as guaranteed IOPS, or data redundancy (such as erasure coding) to protect the data.

In other words, rather than IT being *infrastructure* technology, it can truly be about *information* technology. The business results should be simple: growth and increased competitive advantage.

Where is SDS applicable?

In Neuralytix's paper entitled *A Buyer's Guide to Software-Defined Storage (SDS): Establishing a Sustainable Data Strategy* (May, 2016), Neuralytix advised customers that

> *"SDS **must** be part of any IT organization's strategy."*

We emphasize this advice again here.

Apart from what we noted earlier about "liberating" data; SDS has other equally critical functions. Being software-defined (aka software-driven), SDS can enable the faster provisioning and movement of virtual machines and bare metal hardware.

The ability to support multiple data access protocols, allow for data movement between traditional access methods such as SMB and NFS, and next generation methods such as HDFS and S3. The latter are critical for applications that require massive scaling such as Hadoop and other Big Data applications.

Data movement is not limited to on-premise to on-premise transformation, but with protocols such as S3, it enables data easily flow between on-premise and in-cloud instantiations for storage capacity.

Data liberation (which is the result of being able to have free flow of data movement) can accelerate time-to-fail! The adage of "fail often, fail fast" applies here too. The adage suggests that through experimentation, many of which will not yield the desired result, it does allow the enterprise to test as many scenarios as possible to arrive at the desired state.

Additionally, through experimentation, enterprises have found new opportunities that they had not previously though existed, and can result in net new innovation, that may include new or improved products and solutions, better customer service, higher service levels and productivity.

As such, liberated data, when used properly, will enable the enterprise to accelerate its time-to-insight, and time-to-action. This results directly towards the business goals – whether it is business growth, or competitive advantage. The ultimate objective is to improve the overall health of the business.

Key Performance Indicators (KPIs) for a successful SDS implementation

Neuralytix believes that like all modern information technologies, the measures of success any technology is broken down into three KPIs:

- Predictability;
- Repeatability; and
- Scalability.

The IT department must focus on value creation through Information. As such the infrastructure must exhibit attributes that lends itself to value creation, competitive advantage, and growth.

- The optimal infrastructure is one in which the outcome and performance must always be **predictable**. No matter whether a new application, new compute or storage capacity, a new node, etc. is added or decommissioned.
- The processes involved with scaling must be **repeatable**. Same as above, no matter what changes are made to the infrastructure, it should follow a repeatable (preferably automated) process, that yields predictable results.
- The infrastructure should be ideally infinitely **scalable**, or at least foreseeable for the next five to 10 years in every dimension – application, compute, storage, etc.

Neuralytix believes that the IBM Spectrum Storage Suite addresses all these three KPIs, to help customers ensure a successful SDS, and to help drive business growth, innovation, and competitive advantage.

The SDS Framework

Earlier in this Insight is noted that SDS is a framework, not a single software product. This framework has within it a set of interdependent point solutions that, when brought together, make SDS what it is.

At the most basic level, SDS has to have a storage virtualization and scalable storage pool. For SDS to bring any value to a datacenter, it is necessary that it be in a position to virtualize all the storage resources in the datacenter. But simply providing storage virtualization would simply add an unnecessary layer in an already complex storage stack. Therefore, in order to bring value to the storage stack, SDS *must* be able to present a file system that scales across all the storage resources available. Ideally, this is done without any data movement. The challenge for SDS is whether the scale-out storage system is mature, dependable and enterprise featured..

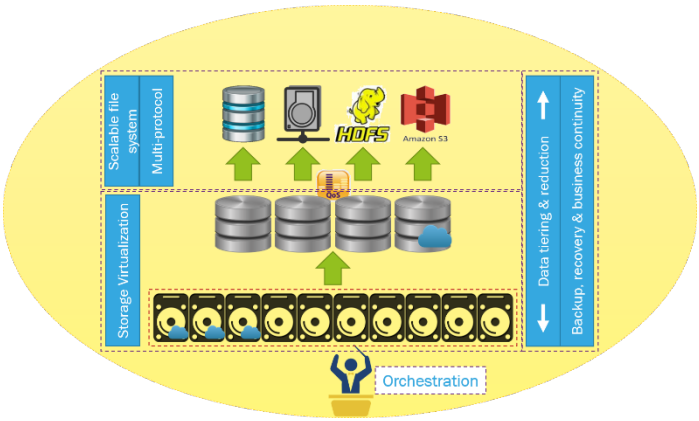

Figure 2: The SDS Framework (Neuralytix 2016)

Added to the storage virtualization and scale-out file system, SDS needs to be able to present the data in the same manner as applications demand. As application diversity increases, so too must the number of protocols supported by SDS.

So, if the virtualized data is of a block nature, it must be able to present it as block; and the same goes for file data, object data, etc. The data services the SDS software must provide at the most basic level include the presentation of data in block, file, S3, and HDFS protocols.

These features are the most basic of what should be expected from SDS software. One way of considering these set of features is that they come together to form the first generation of SDS software.

The maturation of SDS software brings a whole new, more capable set of features.

Other data services that SDS should ideally provide include data tiering, data deduplication, data compression, backup and recovery, archiving, and business continuity.

All data services should be presented with a level of Quality of Service compatible with the applications that are accessing the data.

The orchestration and management layer of the SDS software is a critical identifier of whether the SDS software under consideration is truly a next generation SDS solution.

The orchestration and management of data and storage should not be confined to on-premise solutions. It needs to address the trend towards building out hybrid clouds. In other words, the SDS orchestration and management layer needs to be able to integrate, control, and manage on-premise and in-cloud (whether it is a private or public cloud) through a singular interface, and treat cloud-based storage as a tier within the overall storage strategy.

Management should ideally be accessible from the Cloud, to manage both on-premise and in-cloud storage. Data must be able to travel freely and without restrictions between on-premise storage targets and in-cloud storage targets.

SDS is the stepping stone to the Hybrid Cloud

When all is said and done, all the layers of abstraction, management, orchestration, etc. the end goal is to abstract data away from the infrastructure.

This is often called data mobility. Data mobility allows data to travel dynamically from the end-user to the core systems, out to mobile devices to reporting and analytics systems. It also allows data to travel to and from the cloud as needed or desired.

As more and more processing takes place in the Cloud, and as an increasing amount of data is ingested from data collection points based in the Cloud, SDS solutions need to incorporate data mobility not only within the datacenter, but also across datacenters, leveraging public or private clouds, and also between on-premises and in-cloud storage targets.

Apart from the processing of data, data mobility also serves to enable organizations to take advantage of the economics of on-premises and in-cloud data storage targets. For example, temporal data may be best stored in the cloud, where it is cheaper over a short amount of time, as opposed to long term archiving, which may be best stored on linear tape, offline, off premises.

IBM Spectrum Storage Suite

IBM Spectrum Storage Suite is a unique SDS offering. While most of the commercial SDS offerings are focused on delivering point solutions focused around either block, file, object or virtual machines, the IBM Spectrum Storage suite of solutions provides flexibility and agility to the customer, by allowing them to acquire and deploy only the components relevant to their needs, as well as allowing customers to bring together all the multi-application, multi-protocol, multi-workload, multi-location (on-premise, and in-cloud) support that enterprises should demand from their SDS software.

The IBM Spectrum Storage Software Suite is made up of six components:

- **Spectrum Accelerate** – provides consistent data performance an an expandable grid of servers ideal for quickly deploying and growing databases, supporting VMs and transactional workloads.
- **Spectrum Archive** – allows data stored on linear tape to be addressed in the same way as data stored on magnetic or flash disk drives.
- **Spectrum Control** – provides unified management of SDS as well as traditional physical storage with data tiering and data reduction technologies that can reduce storage costs by up to 73%[ii].
- **Spectrum Protect** – provides rapid backup, snapshot, and recovery capabilities that scales across the organization and the hybrid cloud.
- **Spectrum Scale** – brings a mature, high performance, highly scalable file system with object and HDFS support.
- **Spectrum Virtualize** – providing the ability to virtualize any storage resource. It protects and enables existing investments to be brought into a SDS framework.

Neuralytix research suggests that close to 100% of all organizations have some formal (or shadow IT) in-cloud storage. In reality, all organizations are seeking to find the balance between on-premises and in-cloud storage of data, and arriving at a hybrid cloud strategy.

Absent IBM Spectrum Software, most organizations will find themselves in a situation where they might be moving to a SDS on-premises solution, but a totally separate in-cloud solution. This is not ideal.

Referring back to the idea of data mobility, data must flow freely and with minimal friction between on-premises targets and in-cloud targets. This requires a solution that can abstract on-premises and cloud-based storage targets with a single management and orchestration platform.

The IBM Spectrum Storage Software Suite achieves this goal.

Example: Leveraging cloud storage as a "tier"

In order to have a cloud "tier" in any storage strategy, one needs to have a compatible scalable file system that stretches from on-premises storage to in-cloud storage. IBM is unique capable at this. IBM Spectrum Scale, one of the components of the IBM SDS framework, is a high performance, highly scalable file system for on-premises or cloud SDS deployments. This allows a customer to define cloud storage as a tier within the overall data tiering strategy.

If customers prefer a public cloud, the same Spectrum Scale software can also can also take advantage of the S3 or Swift protocols to move data into public clouds such as Amazon Web Services (AWS) S3. In fact, if desired, on-premises file based storage can be converted to object based cloud storage (S3) on the fly to enable previously datacenter-bound data to be leveraged in the cloud.

This is true data mobility.

Example: Protecting data across multiple media

Traditionally, protecting data is done by way of backing up to a local target for quick retrieval, and ultimately off to linear tape.

Since IBM has been able to leverage cloud storage as a tier, using Spectrum Protect, data can be backed up to multiple targets including local disk targets, the Cloud, and linear tape. This gives flexibility and agility to customers by allowing them to take advantage of the most cost effective backup medium available, given their needs.

Example: Building your own storage cloud, becoming your own Service Provider (SP)

The goal for many IT organizations is to become an SP to their users (customers). Providing capacity to storage precious data for the users to use leverage is a critical operation.

As a critical operation, the provisioning of storage must meet the highest level of KPIs outlined earlier in terms of predictability and repeatability, with scalability being assumed.

In order to do that IBM Spectrum Accelerate can provide IT organizations to take bare-metal servers and storage and present storage capacities in a matter of minutes.

Spectrum Accelerate can also operate in-cloud on Softlayer, IBM's IaaS cloud. So, this means that storage can be provisioned in minutes for on-premises operations and in-cloud for cloud-based applications – all with the same interface, process, orchestration, and management.

SP's and IT can also leverage Spectrum Accelerate to create backup targets for external customers, since Spectrum Accelerate was designed multi-tenancy in mind.

Guidance

As the examples illustrate, IBM Spectrum Software can be deployed to many diverse use cases. In many instances, multiple use cases can be combined together with the appropriate IBM Spectrum Software components to achieve a highly agile, high performance fully SDS environment.

Neuralytix believes that what sets the IBM Spectrum Software suite of solutions apart from other SDS offerings is the completeness and agility of the suite of software.

IBM has addressed all the necessary elements that one should expect from any storage solution for an agile and scalable enterprise.

The IBM Spectrum Software suite of solutions address the multi-protocol, multi-application, multi-location, and multi-workload needs of the modern enterprise. It brings together a suite of solutions that are mature and packages it in a way that allows agility of deployment, and completeness in capability.

The ability to address both on-premise and in-cloud storage is a feature that is rapidly being demanded by enterprises as they understand the economics and benefits of in-cloud storage capacity.

SDS should not be a point solution. It needs to deliver an all-encompassing suite of storage and data services. IBM does this, and does this very well.

Neuralytix guides our Clients to make IBM Spectrum Suite a serious contender when considering their storage strategy.

Endnotes

[i] Excerpted from *A Buyer's Guide to Software-Defined Storage (SDS): Establishing a Sustainable Data Strategy* (Neuralytix, May 2016)
[ii] Based on IBM's calculations

Contact Us

Neuralytix®
755 Sansome Street
Suite 360
San Francisco, CA 94111

- Call us at (415) 919-1188
- Email us at info@neuralytix.com
- Follow us on Twitter at @Neuralytix

Global Offices:

- Australia
- Brazil
- Hong Kong, SAR
- Germany
- United Kingdom