WHITE PAPER

Software-Defined Storage: A Pervasive Approach to IT Transformation Driven by the 3rd Platform

Sponsored by: EMC

Ashish Nadkarni Laura DuBois
November 2015

IDC OPINION

IDC believes that the IT industry shift to the 3rd Platform will be remarkable but disruptive. The four pillars of the 3rd Platform — Big Data, social, mobile, and cloud — are already governing how firms conduct their business. This includes the manner in which they engage with customers, how they innovate, the speed with which they introduce new products and services, and how efficiently and reliably they run their operations. As the 3rd Platform evolves, a wave of innovation accelerators such as the Internet of Things, robotics, 3D printing, and cognitive computing are radically expanding the portfolio of next-generation (next-gen) applications. IDC predicts that virtually all new strategic IT investments made by enterprises through 2020 will go toward 3rd Platform technologies and solutions. The IT infrastructure that supports such next-generation 3rd Platform applications must be agile, resilient, scalable on demand, and manageable in a capex-friendly fashion. In other words, this infrastructure needs to be:

- **Service based:** Support software-centric control, orchestration, and automation of homogenized datacenter resources
- **Autonomous:** Operate independent of any hardware-specific features and be programmatically extensible to the application tier

As a core component of this software-defined infrastructure, the storage tier — which serves a crucial role of handling data persistence — has to be in lockstep with these design goals. Current-generation storage systems — which are based on controller storage software that is tightly coupled to custom hardware design — can be complex to manage and difficult to scale economically, which makes them not well suited as building blocks for 3rd Platform infrastructure. Software-defined storage (SDS) solutions, on the other hand, present a dramatic shift toward controller storage software that is decoupled from underlying hardware, runs on industry-standard (COTS) hardware, and delivers a complete set of storage services at "cloud scale." SDS solutions support various data organizational approaches such as file, block, and object as well as resource orchestration and delivery models such as software-only and hyperconverged appliances.

Suppliers like EMC, which truly embrace the principle of delivering software-defined infrastructure for the 3rd Platform era, understand the importance of taking a portfolio approach to delivering software-defined storage. They acknowledge the need to offer an SDS solution that is best suited for the use case (scale), application, or customer (deployment type). For EMC, it is about not just the development of new SDS products and offerings such as ScaleIO, ViPR, and Elastic Cloud Storage (ECS) but also the preservation of the legacy of its tried-and-trusted storage products such as Isilon and VNX by

reintroducing them in the form of software-defined offerings. While the adoption of 3rd Platform technologies may be disruptive to buyers, suppliers like EMC have risen to the occasion of providing them with necessary options to make the appropriate infrastructure-related decisions.

SITUATION OVERVIEW

The 3rd Platform era is already upon us, and before long, it will have completely overtaken the manner in which firms conduct business. The four pillars of the 3rd Platform — Big Data, social, mobile, and cloud — are already governing how firms conduct their business. This includes the manner in which they engage with customers, how they innovate, the speed with which they introduce new products and services, and how efficiently and reliably they run their operations. For example, in the new era:

- Line-of-business buyers are driving analytics at double-digit growth rates for real-time intelligence.
- Cloud and managed service providers are showing significant gains in infrastructure management.
- Customer experience and social technologies are helping drive an era of cognitive response.
- Mobile device interactions are becoming passive as users accept information surfaced from intelligent systems and machine learning.

In addition, a wave of innovation accelerators such as the Internet of Things, robotics, 3D printing, cognitive systems, and next-generation security are radically expanding the portfolio of next-generation applications. As the 3rd Platform evolves, IDC predicts that virtually all new strategic IT investments made by enterprises through 2020 will go toward 3rd Platform technologies and solutions.

Next-Gen Applications

The next-generation applications built for the 3rd Platform era are very different from current-generation applications. For example, next-gen applications are built using design principles in which the application components are stateless and horizontally scalable. The application and/or data layer processes do not assume infrastructure resilience. They leverage development methodologies such as DevOps, continuous integration, and agile development. In addition, many of these next-gen apps leverage:

- Data management and persistence mechanisms that include "key/value pair"-based storage and/or caching techniques (e.g., Memcached, Redis, HDFS, and/or NoSQL-oriented databases such as Cassandra, MongoDB, and Riak)
- Service-oriented architectures (e.g., message queues and RESTful APIs)
- Newer application frameworks (e.g., Rails, Django, Spring, and AngularJS)
- Programming models like MapReduce for processing and generating large data sets with a parallel, geodistributed algorithm on a cluster
- Multisource geodispersed data collection mechanisms from sources such as social media streams or Internet of Things sensors
- Built-in analytics capabilities intended to inform/help the development process and quickly feed information back into the business process in an iterative fashion

Infrastructure for the 3rd Platform Era

The "ground reality" is that much of the current-generation IT infrastructure falls short in its ability to support the design goals of these 3rd Platform applications. This necessitates a switch to newer service-oriented computing, networking, and storage paradigms.

For example, the compute tier for these apps has to move from hypervisors and bare metal to a PaaS layer and/or application containers (e.g., Docker) to facilitate infrastructure portability. Many of these apps are network chatty, requiring software-defined high-speed interconnects between the application nodes. Current-generation storage infrastructure suffers a similar fate. Current-generation storage systems that are based on controller software that is tightly coupled to custom hardware design are:

- **Complex to manage:** Managing the storage infrastructure as a silo makes it inherently complicated. Furthermore, the lack of visibility into data results in multiple data protection schemes and access protocols.
- **Limited in their ability to scale economically.** Architecturally, current-generation storage systems are inefficient, have a high overhead (especially at geoscale), and lack unified management capabilities.
- **Not cloud ready:** Current-generation storage systems are not architecturally suited to provide on-demand scalability for many of the next-generation applications.

It should come as no surprise that many of the hyperscalers, social media firms, and Web 2.0 firms — suppliers that are considered to be trailblazers in 3rd Platform adoption — have made the switch to a new paradigm in infrastructure known as "software-defined infrastructure."

Software-Defined Storage

Software-defined storage is a core component of a software-defined infrastructure, a building block approach specifically designed for an onslaught of next-gen applications in the datacenter. This infrastructure is meant to be agile, resilient, scalable on demand, manageable, and deployable in a capex-friendly fashion. In other words, this infrastructure is:

- **Service based:** Supports software-centric control, orchestration, and automation of homogenized datacenter resources
- **Autonomous:** Operates independent of any hardware-specific features and is programmatically extensible to the application tier

Like other building blocks of a software-defined infrastructure, software-defined storage is not a single product but an approach.

IDC defines software-defined storage solutions as solutions that deploy controller software (the storage software platform) that is decoupled from underlying hardware, runs on industry-standard (COTS) hardware, and delivers a complete set of storage services.

This approach includes different data organization models, both homogeneous and heterogeneous data persistence schemes, and different delivery models:

- **Data organization models** refer to how data can be organized in a block-based layout, in a unitary or distributed file system, or in a tenant-account-container-object-based model.
- **Data persistence schemes** refer to where the data resides — it could be on disk media inside servers, external JBOD, external storage arrays, tape, or even public cloud.

- **Delivery models** include appliances, software that can run on industry-standard (COTS) hardware, and hyperconverged solutions that run inside a hypervisor. SDS can also be deployed as standalone systems or part of a cloud framework (e.g., OpenStack).
- **Storage services** include data access via known protocols, API-based access/management, automation, resource federation, and data mobility.

Design choices such as scale-out metadata, shared-nothing architectures allow some solutions to be highly scalable and offer "cloud scale" economics.

Key Market Drivers for SDS

SDS solutions seek to address the fundamental shift in the architecture required for scaling 3rd Platform applications — solutions that are enabled by open APIs, open source software, and commercial storage software running on commodity hardware.

Hyperscalers were the trailblazers in embracing that dictum. Suppliers are now offering these solutions so that enterprises can move to a scalable and resilient infrastructure as they roll out 3rd Platform applications. The promise for enterprises is that the decoupling of SDS controller software from hardware and the ability for the controller storage software to run on industry-standard (COTS) hardware acquired from the vendor of choice will lead to a reduced capex burden. Additionally, SDS solutions are designed to simplify existing storage environments — and in the process save costs by reducing the management overhead. The key "success" metrics for SDS therefore are:

- **Deliver a path to the cloud:** SDS solutions should pave the way for converting the datacenter into a service-based "cloud scale" datacenter. By creating a management and federation layer across all datacenter storage resources, they should seek to homogenize them. Additionally, they should allow the use of industry-standard (COTS) hardware to scale the infrastructure on demand, with the appropriate blend of ephemeral, low-latency, and capacity-optimized tiers.
- **Reduction in capex and opex costs: SDS** solutions should offer tangible savings realized during the acquisition process as well as measurable savings in the ongoing management and maintenance. In other words, they should enable businesses to realize "cloud scale" economics within a three- to four-year TCO window.

EMC'S APPROACH TO SOFTWARE-DEFINED STORAGE

As a market leader in storage solutions, EMC was an early entrant into what was a fledgling market segment in an otherwise mature multibillion-dollar industry. EMC's vision for SDS is not a singular product but a concept that should span how storage will be delivered in the future. With this goal in mind, EMC has built a formidable portfolio of software-defined storage solutions — a portfolio by IDC's count that is one of the broadest in the industry. Key differentiators of EMC's portfolio include:

- **Open:** Vendor-neutral, open-standard APIs allow products to be used in a standalone fashion or as part of a cloud deployment (e.g., OpenStack). Open source community editions make it easier for buyers to evaluate software at no cost.
- **Diverse:** EMC's SDS portfolio includes SDS data organization and delivery models — file, block, object, HDFS, and hyperconverged. It also includes next-gen rack-scale, datacenter, and hyper scale-out architectures.
- **Flexible:** EMC wants to make its solutions available as appliances or downloadable software that can be installed on industry-standard hardware.

EMC's focus is to provide an industry-leading portfolio of software-defined storage solutions for:

- Orchestration and management of datacenter resources
- Storage, management, and in-place analytics of unstructured data
- Scalable storage for structured data and virtual infrastructure
- Infrastructure for cloud-scale, geodispersed workloads
- Repository for analytics-centric workloads
- Storage for hyperconverged infrastructure
- Infrastructure for workloads that require data adjacency

With this portfolio, EMC is seeking to deliver enterprise-class software-defined storage solutions for a comprehensive range of use cases, all of which deliver the economics and flexibility of industry-standard (COTS) hardware.

EMC's Software-Defined Storage Portfolio

EMC's SDS portfolio consists of new SDS products and offerings such as ScaleIO, ViPR, and Elastic Cloud Storage and also products like vVNX and the new software-only IsilonSD Edge, which are built on the legacy of the company's tried-and-trusted VNX and Isilon storage products (see Figure 1).

FIGURE 1

EMC's SDS Portfolio

Source: EMC, 2015

Orchestration and Management (ViPR)

EMC launched ViPR Controller in 2013 as an SDS solution for automated provisioning and management of heterogeneous storage resources in the datacenter. At the time of launch, ViPR Controller was an innovative software-defined solution focused on tackling multivendor multiproduct storage sprawl and bringing it under a common service-based management framework. ViPR Controller integrates with cloud stacks such as VMware, OpenStack, and Microsoft and offers RESTful APIs for integrating with other management systems. With ViPR Controller, firms can reduce provisioning times by as much as 63%, reduce provisioning costs up to 73%, and empower consumers to subscribe to new storage in five simple steps.

In an industry first, EMC is now planning to release ViPR Controller code – all the storage automation and control functionality – to the open source community as project CoprHD under the Mozilla Public License. EMC's goal is to leverage the open source community to accelerate development and increase support for storage arrays and data protection technologies. In turn, community-driven development provides customers and partners with choice, flexibility, and transparency.

Scale-Out Block and Hyperconverged (ScaleIO)

EMC acquired ScaleIO – an Israel-based storage start-up – in early 2014 for its innovative server-based storage technology, which uses industry-standard commodity hardware to create a software-defined scale-out block storage solution. The ScaleIO product is differentiated in the market by its ability to run as a hyperconverged storage platform and support multiple hypervisors, operating systems, and computing hardware (including ARM-based servers). Further, a key differentiator for ScaleIO is its unprecedented scalability – the ability to scale to thousands of nodes in a single federated cluster. These features have allowed EMC to position ScaleIO toward large enterprises and cloud service providers, for whom scale and heterogeneity are critically important. This is in contrast to VMware's vSAN, which is targeted more at the midrange markets, in firms that deploy vSphere-based virtual infrastructure.

Scale-Out Object (Elastic Cloud Storage)

EMC launched Elastic Cloud Storage as its next-generation hyperscale object-based storage solution. Delivered either as software only on a third-party commodity infrastructure or as an integrated appliance, ECS is used to store, archive, and access unstructured data at scale. ECS is designed to overcome some of the limitations of Centera (EMC's older-generation software-defined object-based storage) and Atmos (EMC's current-generation software-defined object-based storage), in addition to other object-based storage platforms. It is designed to enable businesses to deploy massively scalable storage in a private or public cloud. ECS enables customizable metadata for data placement, protection, and life-cycle policies. Data protection is provided by a hybrid encoding approach that utilizes local and distributed erasure coding for site-level and geographic protection. ECS can be accessed via a RESTful API with support for Amazon Web Services S3, OpenStack Swift, and EMC Atmos as well as CAS – the Centera API. According to EMC, over a four-year period, the TCO for storing data on ECS is lower than the TCO for storing the same amount of data in Amazon Web Services. EMC claims that ECS is the industry's most cost-effective HDFS platform.

ECS is being positioned as the next tech refresh for Centera and Atmos. In fact, the development road map for ECS Software and ECS Appliance adds new capabilities in a phased manner to enable a seamless transition for both Centera and Atmos customers.

Scale-Out File (Isilon)

EMC acquired Isilon in 2009 to add next-generation scale-out NAS capabilities to its storage portfolio. Powered by OneFS – a distributed file system – the Isilon NAS product family was one of the earliest entrants in the fast-growing scale-out file market segment. OneFS was one of the earliest examples of a software-defined file-based storage solution – it employs a shared-nothing architecture to deliver a wide variety of storage protocols on industry-standard servers. To meet the needs of traditional procurement channels, EMC chose to make Isilon available as fully integrated appliances.

With the recent introduction of EMC IsilonSD, a new family of SDS offerings, EMC has decoupled Isilon OneFS from the appliance-only delivery model. IsilonSD Edge, the first product in this new Isilon SDS family, runs the same OneFS operating system as the appliance version. It allows enterprises to easily extend their data lake from a central datacenter to enterprise edge locations, including remote offices and branch offices, to consolidate and distribute unstructured data used to support a wide range of 2nd and 3rd Platform applications. The company claims that IsilonSD Edge can be deployed in minutes on industry-standard (COTS) hardware. IsilonSD Edge is tightly integrated with VMware ESX and vCenter for simplified management and deployment with an existing virtual infrastructure.

Unified SAN and NAS (vVNX)

EMC positions the VNXe as a unified SAN (iSCSI) and NAS (NFS/SMB) entry-level array that is suited for midsize IT environments. It is now launching a software-defined edition of VNX that is suited for use cases such as development/test of vApps with virtual VNX storage and to test functionality such as data replication and snapshots/clones without the hardware overhead. IT administrators will appreciate the consistent look and feel across the VNXe series using Unisphere and feature parity with the appliance version.

In another industry first, EMC also plans to release the vVNX as a community edition. IT administrators can freely download it for nonproduction use. EMC initially plans to support the vVNX as a virtual machine running on vSphere but plans to eventually release editions for other hypervisors.

CHALLENGES/OPPORTUNITIES FOR EMC

For all incumbent storage vendors like EMC, the adoption of a software-defined storage product strategy presents a shift at several levels. It presents a shift in how the vendor develops, markets, and sells products and services. It presents a shift in how the vendor interacts with buyers that may have a preconditioned mind set on how storage is procured and managed. It presents a shift in how it supports these buyers once they have embarked on a software-defined infrastructure deployment. And finally it presents a shift in where and how revenue is recognized. In short, embracing a product strategy that is based on software-defined storage is massively disruptive to say the least. And yet for many such vendors, a status quo presents an existential threat: Embrace the SDS vision, be disrupted in the short term, and stay relevant in the long term, or stay the course and be completely irrelevant in the long term. The opportunity is clear.

IDC believes that EMC is one of the few incumbent vendors that sees this companywide transformation as an opportunity to be a leader in software-defined storage solutions. EMC has commanded its engineering resources to transition its existing product lines to be software defined. At the same time, it has used its M&A capabilities wisely to acquire newer products to inorganically boost its "3rd Platform storage" portfolio. EMC has embraced disruption voluntarily so it is not disrupted involuntarily. And yet, this is just the beginning of a long journey for EMC. Along the way, EMC will need to face several challenges — both internally and externally — as it changes the perception of its own employees and its loyal install base to embrace software-defined storage. EMC needs to ensure that its vision, product portfolio strategy, and execution for a software-defined infrastructure are in lockstep all the way to the destination.

CONCLUSION

IDC believes that as technology becomes the underpinning of digital transformation, new business models as well as products and services that seamlessly blend digital and physical data will alter business and customer experiences and generate additional revenue streams. These new business models will transform industries in completely new ways. Key defining features of this era will be immediacy, new buying centers, future of how business is conducted, measured efficiency, abundance of resources, and personalization of goods and services.

3rd Platform IT infrastructure — which includes software-defined storage — will move in lockstep with this transformation. As suppliers prepare to deliver the next-generation infrastructure, they'll need to focus on providing value via a portfolio of offerings — each designed to serve as a cog in a

"cloud scale" datacenter. Newer technologies like silicon photonics will allow the implementation of disaggregated compute, memory, and storage pools — bound by a seamless software layer.

The 3rd Platform is the new core of IT market growth — and software-defined storage (infrastructure) is a crucial element of that growth. The journey just started.

About IDC

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications and consumer technology markets. IDC helps IT professionals, business executives, and the investment community make fact-based decisions on technology purchases and business strategy. More than 1,100 IDC analysts provide global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries worldwide. For 50 years, IDC has provided strategic insights to help our clients achieve their key business objectives. IDC is a subsidiary of IDG, the world's leading technology media, research, and events company.

Global Headquarters

5 Speen Street
Framingham, MA 01701
USA
508.872.8200
Twitter: @IDC
idc-insights-community.com
www.idc.com