451 Research® | Advisory

# Choosing a Dynamic Storage Foundation for OpenStack

MAY 2016

COMMISSIONED BY

redhat.

## About this paper

A Pathfinder paper navigates decision-makers through the issues surrounding a specific technology or business case, explores the business value of adoption, and recommends the range of considerations and concrete next steps in the decision-making process.

## About 451 Research

451 Research is a preeminent information technology research and advisory company. With a core focus on technology innovation and market disruption, we provide essential insight for leaders of the digital economy. More than 100 analysts and consultants deliver that insight via syndicated research, advisory services and live events to over 1,000 client organizations in North America, Europe and around the world. Founded in 2000 and headquartered in New York, 451 Research is a division of The 451 Group.

## TABLE OF CONTENTS

## STATE OF THE MARKET – CHANGE OR PREPARE TO BE LEFT BEHIND

"The lightning said to the oak tree, 'Stand aside, or take what is coming to you.'" So goes an old Afghan proverb that neatly summarizes the situation today for IT decision-makers when it comes to storage architectures and strategies. A combination of inexorable forces is shifting the ground on which the traditional tier-one storage market has been built – so much so that traditional storage environments are increasingly unsustainable, far too costly and, most of all, ineffective in the face of underlying challenges. In a nutshell, the monolithic storage frame arrays deployed for the past two decades do not cut it in a world where organizations want the IT stack to look and behave like a cloud. This is why we believe enterprise storage is entering a transformative period, both in terms of technology and in how storage is organized and managed as the next-generation datacenter evolves.

As a result of these underlying changes, alternatives to traditional storage environments are emerging. One is software-defined storage (SDS). The key difference between traditional storage and SDS is that SDS places the emphasis on storage-related services, not on storage hardware. With SDS, in essence the storage infrastructure is managed and automated by sophisticated software instead of hardware. In fact, SDS is by nature independent of the underlying hardware. Thus, the SDS offerings in the market today focus on the use of cheaper commodity hardware in place of rigid, more expensive appliances. These offerings also feature the use of catalogs and automation to prepare and provision storage resources without human interaction, creating even greater efficiencies. Not surprisingly, in 451 Research's Voice of the Enterprise, Storage Wave 19 study in September 2015, 54% of respondents already reported they are moving toward a software-defined datacenter architecture for their environments (see Figure 1).
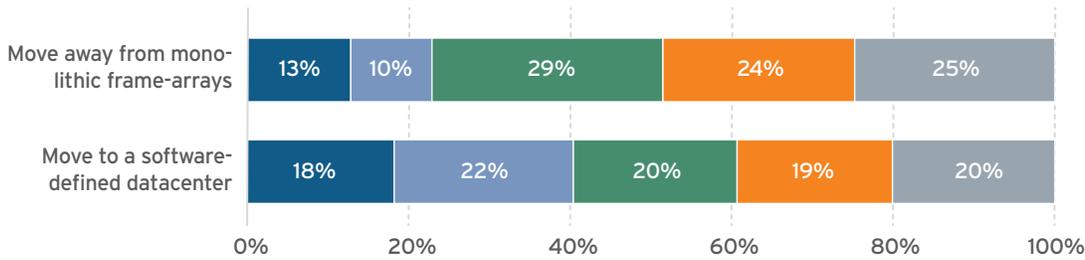
Meanwhile, sales of the monolithic frame arrays associated with the tier-one storage market have been dropping by double-digit margins over the past few years. Although 49% of respondents in the Storage Wave 19 study disagreed with the premise that monolithic frame arrays are going away, it is clear that the flagship products of storage vendors are vulnerable, and organizations are starting to look for alternatives.

Additionally, our research clearly shows that OpenStack is continuing its strong and steady growth as a platform for C-suite strategy consideration when it comes to cloud implementations. In our Voice of the Enterprise Cloud Computing Study from Q3 2015 a majority of the 635 enterprises surveyed already have OpenStack projects or pilots in use, or will by the end of 2016. As OpenStack has grown, so have the components and subprojects – including block storage. Again, our Voice of the Enterprise study showed that 57% of organizations surveyed either have an implementation of SDS in use or in pilot, or will by the end of the year, with Ceph being one such implementation. As a cloud platform, OpenStack has some storage requirements and challenges so complex that traditional storage solutions and appliances cannot easily mitigate them. By contrast, SDS offerings have already proved themselves as a highly suitable paring with OpenStack when it comes to provisioning storage services.

As cloud architectures such as OpenStack transition further into mainstream use cases, storage architectures will need to become more flexible and cost efficient to meet customer needs.  Ultimately, storage architecture shifts toward SDS are required to make on-premises infrastructure more agile and flexible in order to match the capabilities of cloud services.

### Figure 1: SDS Is Rising as Tier One Storage Slips



| | | |
|---|---|---|
| Move away from mono-lithic frame-arrays | 13% 10% 29% 24% 25% | |
| Move to a software-defined datacenter | 18% 22% 20% 19% 20% | |

0%   20%   40%   60%   80%   100%

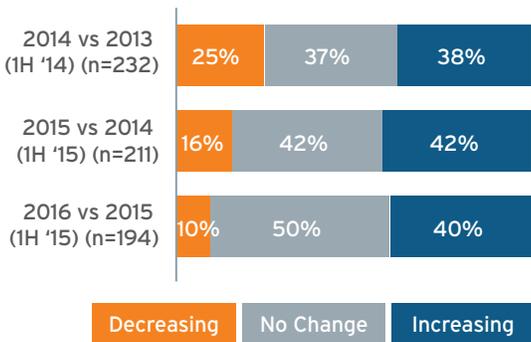# THE FORCES OF CHANGE IN ENTERPRISE STORAGE

As mentioned, several major forces are combining to provoke a vast rethink of cloud strategies and architectures. We believe organizations that don't change or are slow to do so risk being saddled with a rigid storage infrastructure at a time when flexibility and agility are table stakes for organizational success. Not only that, but an organization that is ill-equipped to leverage the ongoing changes in the nature of data itself will likely be outmaneuvered by competitors that recognize the value of exploiting the myriad data types available.

## STORAGE BUDGETS ARE FLAT OR DECLINING

In our Storage Wave 19 survey, 60% of respondents reported having flat to declining budgets. For the lucky few that were expecting a substantial increase in their 2016 storage budget, we found that only 15% of respondents expected to get an increase of 25% or more (see Figure 2). If all other things were equal and data volumes were growing modestly, this budget reality might not be so dire. But we believe that to achieve a sustainable storage environment, organizations must consider finding alternatives to legacy storage arrays to improve the efficiency of storage resource delivery and management.

### Figure 2: Most Storage Budgets Are Flat Or Declining

#### Storage Budget Changes



| | | | |
|---|---|---|---|
| 2014 vs 2013 (1H '14) (n=232) | 25% | 37% | 38% |
| 2015 vs 2014 (1H '15) (n=211) | 16% | 42% | 42% |
| 2016 vs 2015 (1H '15) (n=194) | 10% | 50% | 40% |

Decreasing   No Change   Increasing

#### 2016 vs 2015



>50% more
25-50% more
11-24% more
5-10% more
<5% more

<5% less
5-10% less
11-24% less
25-50% less
>50% less

-5%   0%   5%   10%   15%   20%

**40% plan increased spending**

*50% plan stable spending*

**10% plan decreased spending**
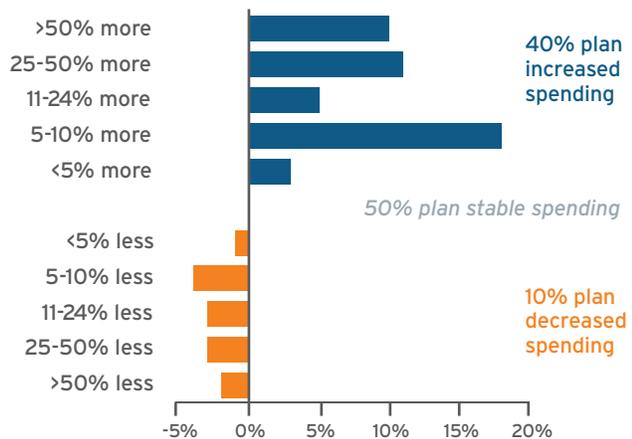
## DATA VOLUMES ARE EXPLODING

In talks with midsized and enterprise organizations for the Storage Wave 19 study, most reported annual data growth in the 30-40% range. At 40%, the data volume doubles in just two years. Continuing to operate traditional monolithic storage systems on largely flat or declining budgets is an obvious recipe for disaster given these data growth realities. Organic business growth continues to be the largest driver for storage growth at organizations (see Figure 3). Analytics (20% of respondents) and multimedia (11% of respondents) were also listed as major drivers for growth, along with server virtualization (19% of respondents).

Secondary storage workloads for data protection and archiving are also having an impact on data growth, given that copy data (11%), regulatory compliance (8%) and backup retention increases (7%) were all listed as significant drivers for data growth. We also project a 38% CAGR in 'big data' volumes through 2020. Plus the emergence of the Internet of Things will generate additional tidal volumes of data. Current systems, including storage architectures, simply cannot store, process and analyze such volumes. But in the future, business success may well depend on the ability to do so.

What's more, the nature of this data is rapidly tilting toward the unstructured type. This refers broadly to information that does not have a predefined data model and is not organized in a predefined way, such as in a data warehouse. Traditional storage models simply choke on data originating in email, video, social media, news feeds, and many other unstructured data sources. But we believe the potential business value of intelligently and efficiently storing, processing and analyzing unstructured data to be enormous.

## REGULATION AND COMPLIANCE ARE ROCKING TRADITIONAL STORAGE FOUNDATIONS

Regulatory compliance challenges are forcing many organizations to retain data for longer periods of time, and prevent storage professionals from simply deleting old data to make way for new content. Compliance and security concerns continue to be two of the biggest obstacles for organizations looking to offload their storage burden to public cloud services, and as such storage customers will have to look for efficient on-premises storage alternatives to shoulder the load when traditional storage arrays become too costly for the task.

> The importance of copy data and backup retention suggests that organizations will need to leverage storage reduction technologies such as de-duplication, compression and erasure coding in the future to boost the efficiency of primary and secondary storage systems (such as backup and archives) in order to control data growth. Future storage architectures must have the flexibility to deal with various workloads ranging from server virtualization to unstructured data storage for file sharing and archiving.

### Figure 3: Organic Growth Is the Leading Driver of Storage Growth

*What projects, technologies or initiatives are most responsible for capacity growth in your networked storage? n=244*

| Category | Value |
|---|---|
| Organic Growth | 23% |
| Analytics | 20% |
| Server Virtualization | 19% |
| New Business Applications | 13% |
| Multimedia (4K Video, Medical Images, etc.) | 11% |
| File Server/Share | 11% |
| Copy Data/Databases | 11% |
| Data Hoarding | 9% |
| Business Reorganization | 8% |
| Regulatory Compliance | 8% |
| Backup Retention Increases | 7% |
| Email | 5% |
| Support Technology Change | 5% |
| SAP | 4% |
| VDI | 3% |
| Private Cloud | 3% |
| Performance Improvement | |
| Mobile Workloads | |

## THE NUMBER OF STORAGE HOSTS IS ALSO GROWING RAPIDLY

Within the cloud space alone there exists public cloud, notably AWS; private cloud, which is unique to each organization; and hybrid cloud. The stark reality is that traditional enterprise storage architectures are incapable of efficiently managing storage across diverse platforms, owing largely to their proprietary nature. Thus, organizations increasingly struggle with the day-to-day maintenance of their complex storage infrastructures. Beyond high cost, the lack of skilled staff and the expense of keeping staff current was a major concern for storage professionals in our Storage Wave 19 survey, with 13% of respondents listing this issue as a top storage pain point. Disparate storage management was an issue that became worse in the past year, with 11% of survey respondents citing this as a major problem, up from just 4% a year earlier.

Storage professionals are dealing with more capacity and spreading themselves across more skill sets than ever. That means less time for researching new technologies, new architectures and new vendors. Simply staying up to date with the new vendors and offerings can be painful. As one healthcare storage professional put it, "There are too many storage vendors, and there is no clear path for any technology.."

## Figure 4: Storage Pain Points

*Q. What are your top storage-related pain points? List up to three*
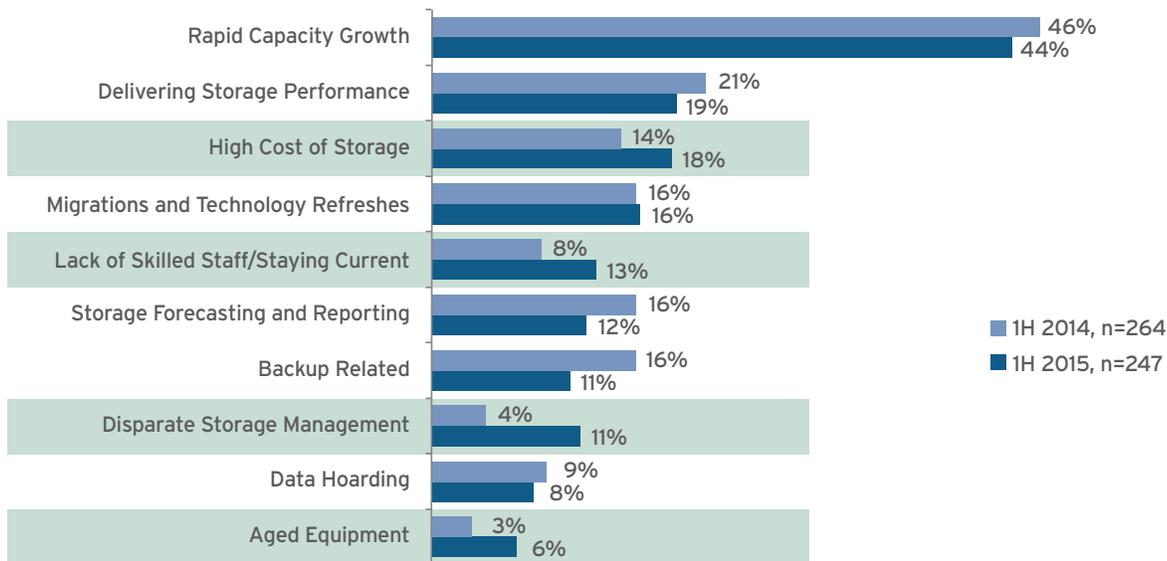


## MODERN STORAGE REQUIREMENTS FOR CLOUD ENVIRONMENTS

In order for cloud architectures such as OpenStack to succeed, your storage management architectures must transition beyond the proprietary tools sold by enterprise storage vendors that currently dominate the market toward standardized management platforms that can leverage storage resources from a wide range of vendors and provision them quickly to application workloads.

Respondents from our Voice of the Enterprise: Cloud Computing, Q2 2015 survey are gradually deemphasizing their non-cloud infrastructure, for modern private cloud architectures and services. While respondents currently have 53.8% of their application workloads running in on-premises infrastructure that is non-cloud, in two years' time these respondents claim that figure will drop to 37.8% of their workloads (see Figure 5). Meanwhile, the percentage of application workloads running in on-premises private cloud infrastructure will rise from 16.5% to 21% in two years. Hosted private cloud is also expected to increase, from 6.8% of application workloads to 11.6%, along with infrastructure-as-a-service (IaaS) public clouds, which will rise to 8.7% from 7% of workloads.

### Figure 5: On-Premises Infrastructure Is Shifting to Cloud Architectures

*Q. (Left) Approximately what percent of the Industry-specific Applications workload is deployed in the following environments today?*

*Q. (Right) Approximately what percent of the Industry-specific Applications workload will be deployed in the following environments in two years' time?*

**CURRENT**
n=91

| | |
|---|---|
| On-Premises, Non-Cloud | 53.8% |
| Off-Premises, Non-Cloud | 3.3% |
| On-premises Private Cloud | 16.5% |
| Hosted Private Cloud | 6.8% |
| Infrastructure-as-a-Service (IaaS)/Public Cloud | 7.0% |
| Software-as-a-Service (SaaS) | 12.7% |

**FUTURE**
n=87

| | |
|---|---|
| On-Premises, Non-Cloud | 37.8% |
| Off-Premises, Non-Cloud | 5.1% |
| On-premises Private Cloud | 21.0% |
| Hosted Private Cloud | 11.6% |
| Infrastructure-as-a-Service (IaaS)/Public Cloud | 8.7% |
| Software-as-a-Service (SaaS) | 15.8% |
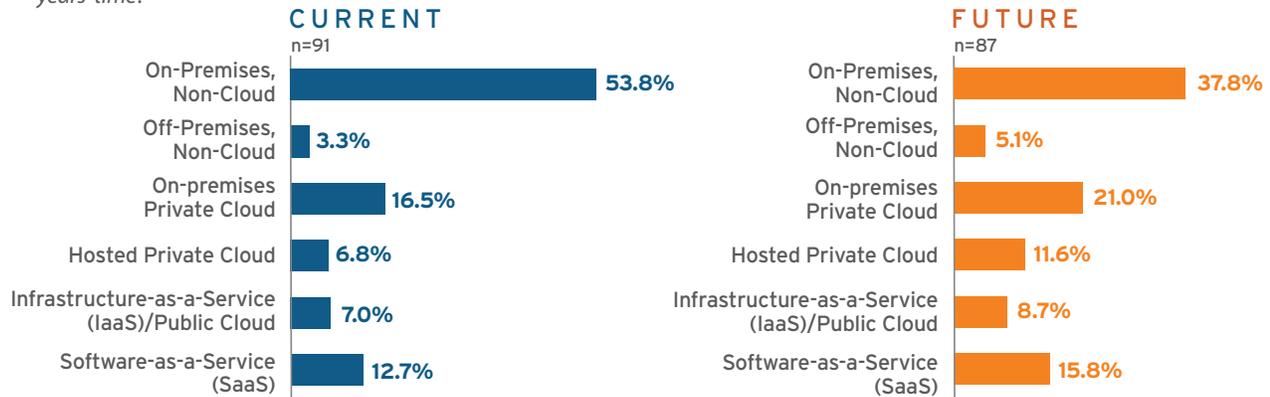
## MODERN STORAGE REQUIREMENTS FOR CLOUD ENVIRONMENTS

As data from our Voice of the Enterprise study shows, 40% of organizations surveyed will be increasing spending on OpenStack (see Figure 6). Technically, it is helpful to think of OpenStack as an IaaS framework consisting of subcomponents for storage and other functionality. The framework and components are written primarily in the open source Python programming language. All OpenStack software is open source, and vendors and users alike say OpenStack benefits from a clean, modular architecture. Components for OpenStack now include block storage. OpenStack controls large pools of storage and other resources, and manages them via a dashboard. This gives administrators the control they seek while at the same time empowering users to provision these resources through a Web interface.

It is little surprise, therefore, that we forecast the total OpenStack market will boom along at a healthy 34% CAGR into 2020. It is also not surprising that the OpenStack market has attracted many well-established vendors including Red Hat, Rackspace and Dell. The number of OpenStack-supporting organizations today approaches 600.

The increased spending on OpenStack as reflected in our research closely approximates 2016 spending intentions for SDS storage, including Ceph – an open source storage platform capable of storing data on a single distributed cluster. Ceph also provides the necessary interfaces for object-, block- and file-level storage. Ceph is touted as having the ability to be totally distributed, without a single point of failure, while being hyper-scalable to the lofty exabyte level. We believe this kind of massive scaling potential aligns well with the hockey-stick growth curve of enterprise data volumes noted earlier. Ceph is further touted as being both self-healing and self-managing, which can dramatically lower management and administration costs – a crucial feature in an era of flat or declining storage budgets. And what we feel is a tight integration between Ceph and OpenStack makes the combination of the two potentially very compelling.

### Figure 6: OpenStack Spending Plans

*Q. How would you describe your organization's spending plans in 2016 compared to 2015 for OpenStack? n=462*

| | |
|---|---|
| Increase | 39.8% |
| Remain the Same | 57.8% |
| Decrease | 2.4% |

## DRIVERS OF OPENSTACK

The drivers of OpenStack are many, and they are diverse. Major drivers include:

- **Both enterprise and service-provider demand.** OpenStack was originally intended for service providers. Today, enterprise demand has emerged as the main driver, largely owing to the continued surge in interest and spending on cloud deployments. Leading OpenStack vendors are ramping up efforts to serve both market segments.

- **Great fit for storage.** OpenStack currently includes a variety of APIs to manage and control object, block and file storage, which meshes well with the needs of the enterprise in the face of today's data and storage challenges.

- **Business benefits abound.** Many OpenStack users applaud key benefits such as flexibility (and the business agility it can engender), cost savings and the ability to avoid vendor lock-in.

- **Private cloud appeal.** Organizations increasingly want the scalability, elasticity and other advantages of public cloud services – only they want it under their own roof and under their own control. We observe that for a growing number of enterprises and service providers, the answer is OpenStack, which therefore can also be a key element for deploying more enterprise apps to the cloud.

- **An open alternative.** OpenStack presents an open alternative to proprietary cloud options, as well as to other open source options. Compared with other open source options such as CloudStack, Eucalyptus and Joyent, we believe OpenStack presents more pressure for openness.

Of course there are challenges with OpenStack as well. Perhaps the most notable is a shortfall in the supply of IT talent proficient in OpenStack. Given the growing demand and the reality that deploying OpenStack does require a high degree of know-how, talent poaching is rampant. Meanwhile, there are some concerns we have heard that the various OpenStack offerings are not necessarily compatible. And many enterprises that have deployed OpenStack are still cautious about using OpenStack for a production cloud.

But as noted earlier, these concerns are not dampening the overall enthusiasm for OpenStack – quite the opposite, considering the growth rate we have targeted. If it is true that 'what's past is prologue,' then the nearly 600 OpenStack support vendors out there are working feverishly to mitigate some of the technical and deployment issues. And IT talent, as always, will 'follow the money,' which in this case means burnishing their OpenStack skills and knowledge to begin bridging the gap between supply and demand of OpenStack talent.

## CORE STORAGE REQUIREMENTS REMAIN IMPORTANT

Beyond the basic preservation and retention of data, storage for cloud and SDS environments must retain the following enterprise storage traits:

- **Resiliency**. In the end, the most important attribute for any storage architecture is its ability to maintain and protect the integrity of the data that has been entrusted to that system. In the distributed storage architectures found in SDS and used by storage platforms such as CEPH, data integrity is ensured by creating replicas of data across multiple nodes to protect data in case of a system failure. Periodic scrubbing of the data is done to ensure that replicas stay consistent to avoid data corruption.

- **Performance.** The storage world is in the midst of a major transformation from conventional hard drives to flash-based storage. The SDS architecture you choose should be optimized for leveraging flash and other future solid-state technologies to keep up with the current and future performance needs of workloads. Flash management is a key capability for storage systems going forward, and should be a major consideration as you and your teams evaluate SDS offerings.

- **Data protection.** The common data-protection techniques used in enterprise storage arrays will still be needed to protect workloads. Snapshots should be available to create recovery points to roll back errors or data-corruption events, and this technology is also commonly used to create clones of data sets for test/dev. Beyond the local replication used to protect data within a datacenter, remote replication must be available to ensure that data is still protected in the event of a large-scale disaster at the primary data site. To increase storage efficiency for primary and secondary storage, compression and deduplication technologies should also be implemented whenever possible.

# RECOMMENDATIONS

Evaluate open-source storage offerings and development communities. The vast majority of storage system software on the market today continues to be closed-source and proprietary, although open-source platforms such as CEPH and Gluster have emerged in recent years. While open-source storage software is free to download and use, you must still evaluate communities and support options to ensure that problems can be rectified when they emerge. The vibrancy of an open-source storage development community is also important to evaluate, since this will largely determine the level of innovation going forward.

## WHAT TO FOCUS ON BEFORE BUILDING YOUR NEXT STORAGE ENVIRONMENT

**Consider going all-in for 'Open.'** The vast majority of storage system software on the market today continues to be closed-source and proprietary. But now open source platforms such as Ceph are here, and they offer capabilities to reduce storage costs through automated management. While open source storage software is free to download and use, you must still evaluate communities and support options to ensure that problems can be rectified when they emerge. The vibrancy of an open source storage development community is also important to evaluate, since this will determine the level of innovation going forward.

**Open source communities are great, but...** For a serious foray into SDS, it is prudent to consider a trusted third-party partner to assist by adding its deep knowledge and experience in helping establish things like support requirements and problem resolution techniques. The good news is that open source software has no up-front license fees – and that is important. It is equally important that organizations take a hard, objective look at the vibrancy and experiential depth of the various open source development communities, which are not all created equal. A trusted partner can help here as well, because that vibrancy will be a key element in an organization's ability to innovate going forward.

**Collaborate with business stakeholders before building your new infrastructure.** Your updated storage infrastructure must be able to satisfy a broad number of different workloads while delivering the benefits of storage resource consolidation and automated provisioning through global catalogs. To do this, there must be collaboration between business stakeholders and domain experts, to ensure that service classes can be defined according to workload needs. Without this step, you run the risk of building a menu from which no one wants to order.

**Don't forget Linux when it comes to OpenStack.** Underlying the OpenStack platform is the Linux operating system, and OpenStack relies on it for just about everything from service operation and access to hardware resources. Be sure the OpenStack vendor implementation you choose for your storage platform features an airtight integration of OpenStack and Linux.

**Go for experience.** Early releases of OpenStack were missing key technologies for enterprises to successfully implement the software. When considering OpenStack storage vendors, be sure they have added enterprise qualities to solutions based on software-defined storage. Enterprises will need the same high levels of production expertise, security, and integration they got with proprietary storage offerings when they deploy SDS-based solutions.

**Remember: Automate to reduce management burdens.** With relatively flat budgets and rapidly growing data stores, the only way cloud professionals will be able to stay ahead is by automating as many processes as possible. A good place to start would be with storage provisioning.