A SOLIDFIRE PAPER

Getting it Right
OpenStack Private Cloud Storage

NetApp®

Table of Contents

Introduction

The amazing agility, flexibility, and power of cloud computing has executives everywhere trying to figure out how to get cloud functionality and economics inside their own data centers. Beyond basic infrastructure improvements, many are also looking to run their IT departments like internal service providers. This shift represents a fundamental transformation in how IT services are delivered.

Transitioning from legacy data center practices and architectures to those of the next generation cloud is no small undertaking. The benefits of cloud are indisputable. Commodity servers and myriad networking choices have made choosing the right infrastructure relatively painless.

Storage, however, remains a complex, expensive aspect of setting up a cloud — one you can't afford to get wrong. When determining what goes into your next generation data center, you should be asking yourself the following questions:

- Is private or hybrid cloud computing right for me?
- How do I automate and orchestrate my infrastructure?
- What kind of cloud storage is best suited to my needs?

In this article, we:

- Discuss the values of, use cases for, and features you should expect from your private cloud
- Highlight the reasons OpenStack is right for your private cloud
- Outline the critical considerations for selecting the right storage for your cloud

Private cloud: The benefits of cloud computing, inside your firewall

In today's age of global business, the importance of IT infrastructure cannot be understated.

Reliable and flexible infrastructure is critical to any business, and the era of cloud services has raised the bar on what it means to deliver IT services.

With increasing demands being made on enterprise IT departments to keep up with the accelerating pace of business, infrastructure has to enable more dynamic, efficient application and IT service delivery. Unfortunately, more often than not, IT provisioning can be a bottleneck. Traditional means are often inefficient. When developers plan their projects and estimate the server, networking, and storage resources they need, they often overestimate and pad their requests to ensure that what actually gets provisioned is adequate. The formal request comes via an IT ticket, and then the waiting begins. Days, possibly weeks go by before they get the resources they need, and all the while, they're unable to iterate on and evolve core business offerings.

"Overall, there are very real trends toward cloud platforms. …. [This enables] individuals and businesses to choose how they'll acquire or deliver IT services, with reduced emphasis on the constraints of traditional software and hardware licensing models."

–Chris Howard, Research Vice President, Gartner

Gartner Newsroom. Gartner Says Cloud Computing Will Become the Bulk of New IT Spend by 2016. Available at http://www.gartner.com/newsroom/id/2613015. Accessed March 17, 2015.

When finally provisioned, the resources may not be agile, flexible or easily scalable to allow for changing requirements, which can lead to frustrated and unhappy developers.

Cloud services have shown just how responsive and self-service oriented IT delivery can be, and more and more companies are looking to emulate this flexibility, scale, and automation inside their firewalls. In fact, private cloud is the most popular cloud deployment model and is expected to grow at twice the rate of public cloud.[1]

Gartner highlighted five key attributes of private clouds.[2] These distinct characteristics make private cloud a compelling infrastructure delivery model to many organizations:

- Resources are provided as services
- Services are scalable and elastic to meet consumer demands
- Resources are shared to build economies of scale
- Services are tracked with usage metrics to enable multiple payment models
- Cloud is delivered using internet identifiers, formats, and protocols

1 QuinStreet Enterprise. 2013 Cloud Computing Outlook. Available at http://www.itbusinessedge.com/slideshows/cloud-adoption-trends-favor-private-cloud-two-to-one-04.html. Accessed March 17, 2015.

2 Gartner & Co. Five Refining Attributes of Public and Private Cloud Computing. Gartner #G00167182. May 5, 2009.

5 keys

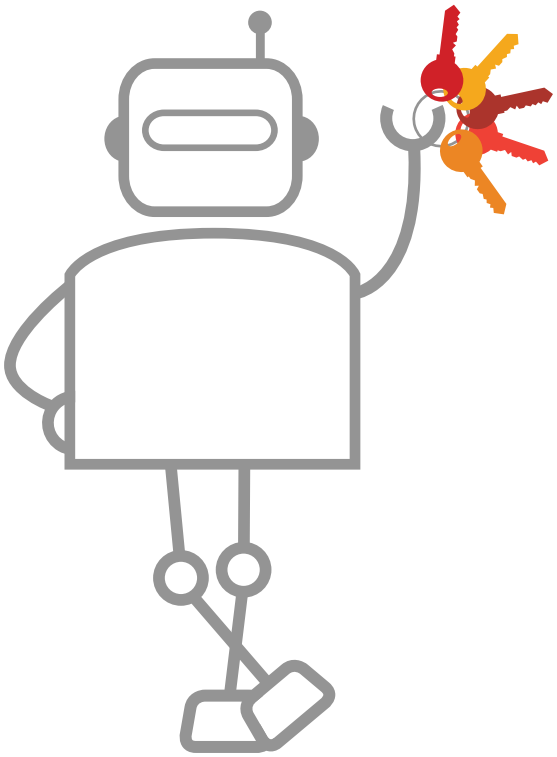
Resources as services
Private clouds enable a service model to end users, giving them greater flexibility over obtaining IT resources and enabling those resources to be obtained programmatically. End users are abstracted from the underlying technologies used to provide the services, allowing them to focus on development needs rather than how best to carve off an adequate piece of IT infrastructure.

Flexibility and scale
When services are scalable and elastic, resources can be added or removed as needed to better enable dynamic business demands. This eliminates complex advanced capacity planning because resources can be spun up and used as needed.

Resource sharing
Because developers often overestimate the resources they need, they may inadvertently under- or over-provision their resource request. This can lead to inefficient resource usage, where provisioned infrastructure sits idle waiting to be used even though other development groups may have an acute need for it. When IT silos are consolidated into a private cloud, they can be shared, helping build economies of scale. Fewer islands of infrastructure reduces capital outlays and minimizes the operational expenses associated with managing disparate systems. Infrastructure can be used most efficiently and can serve the needs of multiple users simultaneously.

Measurement and payment
Private clouds enable showback (a quick way for companies to understand how infrastructure is being used by business units and how that usage is translating into cost). Those companies doing actual chargeback can adjust infrastructure expense based on usage, helping drive costs down and increasing resource utilization.

Use of internet protocols and technologies
In non-cloud environments, interaction with infrastructure is via nonstandard, proprietary ports and protocols, which often require third-party management. Cloud deployments are maintained over internet protocols, such as HTTP. Because clouds are delivered via internet protocols, code can be written to them via APIs. All you need is a standard API client to maintain your cloud infrastructure.



Orchestrating your IT infrastructure

The right cloud computing software is a critical component in the success of your private cloud. When automating and coordinating the tasks involved with configuring, coordinating, and managing servers, networking, storage, software, etc. across disparate systems (often located in dispersed data centers), choosing an orchestration platform that best meets your business requirements and desired outcome is paramount.

There's no shortage of cloud orchestration operating systems to choose from: OpenStack, CloudStack, and VMware, to name a few. With significant developer and vendor momentum behind it, OpenStack has managed to position itself as one of the most attractive data center operating systems in the market.

RightScale's 2015 State of the Cloud Report indicates 63% of respondents have deployed a private cloud,[3] and 451 Research predicts OpenStack revenue will reach $3.3B by 2018.[4]

The opportunity for OpenStack in the enterprise is significant. Large-scale enterprises like Comcast, Bloomberg, and Best Buy are already using it in some capacity. Some of the world's largest cloud infrastructures today are built on a combination of open source software, commodity hardware, and a bunch of PhDs to make it all work. But those unable to afford the cost, complexity, or extra brain power on staff want their PhDs built into the infrastructure, not standing next to it. This is the void OpenStack and its ecosystem are looking to fill.

3 RightScale. 2015 State of the Cloud Report. Available at http://www.rightscale.com/lp/2015-state-of-the-cloud-report?campaign=701700000012UP6. Accessed March 17, 2015.

4 PRWeb. OpenStack Revenue Expected to Hit $3.3 Billion by 2018. Available at http://www.prweb.com/ releases/2014/09/prweb12144923.html. Accessed March 17, 2015.

Why OpenStack?

Reasons for adoption

Deploying orchestrated, shared infrastructure is no small undertaking. Because OpenStack is open source, it provides the flexibility to explore an orchestrated private cloud with little risk. The reasons for choosing OpenStack vary, but OpenStack users highlight four key reasons why they chose the software.[5]

Ability to innovate

This business driver jumped from sixth to first from May 2014 to November 2014, highlighting the value users place on orchestration solutions that enable innovation. When infrastructure maintenance stops consuming spare cycles, time can be focused instead on innovating features and functionality that differentiate your business.

Open technology

Open technology has many benefits beyond reduced cost, including greater flexibility, interoperability, and the green light to try it out before buying. OpenStack's large community constantly contributes code, improving the project's features and functionality. As a result, the overall project becomes more customizable for users.

Cost savings

The low cost of open-source technology such as OpenStack is extremely compelling to users looking for lower-cost options. Proprietary software costs include the initial purchase cost, licensing fees, support renewals, and the hidden costs of being locked into a proprietary technology. Open-source technology eliminates most, if not all, of those costs.

Avoiding vendor lock-in

By not using proprietary software, you are no longer beholden to one vendor for products, services, proprietary APIs, or onerous switching costs. Lack of lock-in results in greater solution flexibility and the freedom to pick the technology that works best, not the one you're tied to.

OpenStack use cases

Many organizations dip their toes into OpenStack with test, development, and quality assurance (QA) use cases. However, the percentage of production deployments has increased from 33% to 46% between May 2014 and November 2014[5], indicating users have embraced OpenStack for their business-critical production workloads.

Scalable web applications and dev/test environments are workloads well-suited for OpenStack. In fact, the top three OpenStack workloads are:[5]

1. Web services: Web services are good fits for cloud. As site traffic or request volume increases, cloud-provided infrastructure can quickly scale horizontally to accommodate growth and vertically to accommodate a temporary burst in requests. More systems can be added, or existing systems can be made bigger, enabling business agility.
2. QA/test environments: Shared infrastructure for QA/test environments lowers the cost of operations through self-service. Infrastructure can be provisioned without IT involvement, used as needed, and spun down when testing is done.
3. Database-driven applications: Databases can be run from an OpenStack cloud, enabling users to focus on developing applications instead of underlying infrastructure. MySQL and its variants are the most popular databases in OpenStack environments, but other nonrelational ones such as MongoDB also enjoy notable usage.

The use cases for cloud are myriad and also include continuous integration: enterprise applications, benchmarks and stress testing, management and monitoring, and more.

OpenStack benefits

OpenStack users may value individual benefits such as ability to innovate, lower cost, or no vendor lock-in. But the broader OpenStack project and community also provide noteworthy benefits.

Despite being an open-source technology, OpenStack has the support of many notable companies, including Red Hat, Rackspace, and Dell. Many supporting companies have their own OpenStack distributions, providing support and services to facilitate installation, easing troubleshooting, and taking some of the DIY out of deploying OpenStack. A robust ecosystem around OpenStack accelerates and helps ensure successful deployments.

OpenStack is supported by a large community committing code and contributing to projects. It also counts a large list of companies as users, including Wells Fargo, American Express, Best Buy, Comcast, Disney, AT&T, and Cern, to name a few.[6] OpenStack can develop new features quickly. Companies that either lack a desired data service or find the services supported in their OpenStack deployment immature can simply innovate and create new services that meet their needs. OpenStack's module-based framework makes it easy for contributors to make a direct and significant impact.

OPENSTACK

CONTRIBUTORS	USERS
Dell	American Express
Red Hat	Wells Fargo
Rackspace	Comcast
	Best Buy
	Disney
	AT&T
	Cern

5 OpenStack SuperUser. OpenStack User Survey Insights: November 2014. Available at http://superuser.openstack.org/articles/openstack-user-survey-insights-november-2014. Accessed March 17, 2015

6 OpenStack Foundation. OpenStack Powers Demanding Production Workloads Worldwide. Available at http://www.openstack.org/user-stories. Accessed March 17, 2015..

You've picked OpenStack. Now what?

Robust automation is critical to successful orchestration
There is a fundamental difference between automation and orchestration, and getting the automation portion of your infrastructure "wrong" can materially impact your ability to orchestrate successfully.

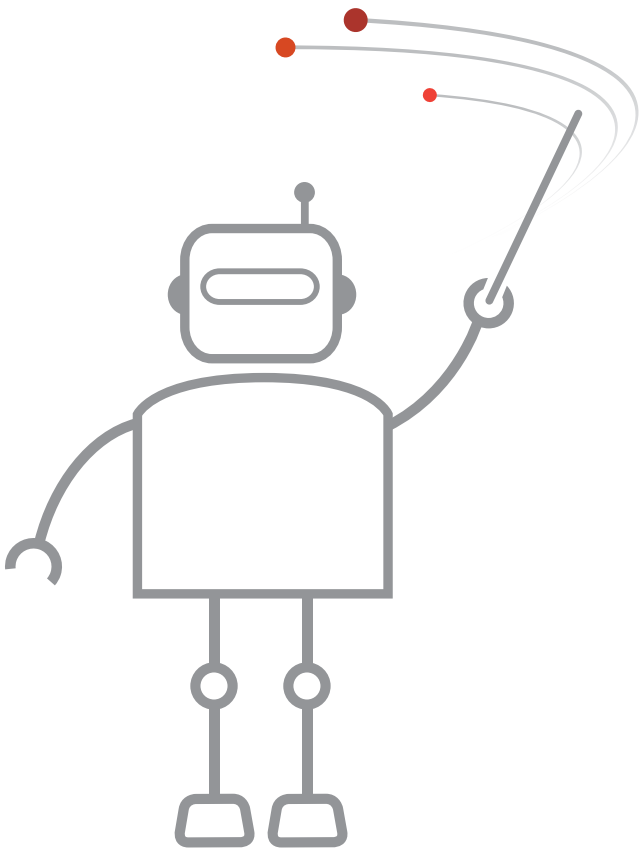
Automation refers to tasks: installing an operating system, configuring a server, provisioning volumes on storage, deploying code and system updates, user provisioning tasks, stopping a service, etc. Orchestration builds on automation; it refers to stitching together a series of automated tasks into a workflow. By streamlining these automated tasks, orchestration enables you to more quickly deploy your infrastructure and focus on other higher-value tasks. It also affords an opportunity to streamline – to optimize – those processes for even greater gains in speed of deployment.

Complementary to streamlining process is that those processes can become consistent and repeatable. Consistent workflows result in IT stability; the process is proven to work and can be repeated extensively with confidence. Not only can you accelerate deployment, but you can ensure deployment goes off without issues.

Infrastructure with complete APIs enables comprehensive automation. If your underlying IT doesn't or can't support comprehensive automation, it's actually a giant roadblock to successful orchestration; it will become an operational burden and prevent you from moving toward a private cloud. When your infrastructure supports full automation and thereby can enable orchestration, you have a platform that you can innovate on. Not having to manually run the infrastructure all the way to the bottom of the stack allows you innovate and focus instead on what distinguishes your business.

Object or block storage?
OpenStack has subprojects that deliver both object (Swift) and block (Cinder) storage. A variety of performance-focused primary and optimized secondary storage solutions are on the market, and they provide flexible, highly scalable storage services for OpenStack.

Optimized secondary storage has clear value for object-based, large-scale storage, where spinning disk still maintains a $/GB advantage over flash and performance is not a significant concern. Features such as erasure coding and a lack of compression or deduplication (which don't provide much benefit in object storage environments where data is already compressed) further highlight the focus and suitability of these systems for storage of large-scale cold data, such as backups, images, video, content archives and more.



Swift

Swift provides a distributed scale-out object store across nodes in an OpenStack cluster. Swift data resilience depends on "eventual" consistency when replicating data. Running virtual machines on Swift storage can be troublesome because the storage only reads and writes entire objects without ensuring consistency. Swift's ability to provide scale-out storage on commodity hardware may make it a more attractive option to external storage such as a SAN for the use cases listed.

Performance-focused primary storage can deliver cost-effective capacity using compression and deduplication that work well for primary data. This architecture is best suited to performance-sensitive primary storage workloads, such as VM disks, SQL and NoSQL databases, and data processing/analytics. Using external Cinder block storage in an OpenStack environment enables both availability and performance to be managed from the storage.

Initial small or one-off OpenStack projects may lend themselves to optimized secondary storage, where more comprehensive array features, data optimization, ease of scale, and resiliency aren't as essential. However, mission-critical workloads often need the reliable performance and data services that external block storage provides. Swift object storage often can complement performance Cinder block storage, functioning as a backup location for primary data files or longer-term archive.

Cinder

Cinder provides block storage as a service in OpenStack. Cinder uses a plug-in architecture, allowing the use of various supported storage devices as backend stores. Additionally, Cinder includes a fully contained reference implementation built using logical volume manager (LVM). Deployers are able to choose one or more backends and may start with one backend today and scale out to other devices as needs change.

The role of the Cinder Project is evolving very quickly and, in many ways, is quickly maturing through community contributions. The rate of change is fast and furious and the OpenStack community has moved beyond simply breaking block storage out from Nova compute; they have started exploring how to make block storage in OpenStack the best it can be.

Cinder's mission is to:

- Implement services and libraries to provide on-demand, self-service access to block storage resources
- Provide software-defined block storage via abstraction and automation on top of various traditional backend block storage devices

So, in other words, Cinder aims to virtualize various block storage devices and abstract them into an easy, self-serve offering to allow end users to allocate and deploy storage resources on their own quickly and efficiently. It simply allows the dynamic creation, attachment, and/or detachment of disks to Nova compute instances.

Cinder is a plug-in architecture. You can use your own vendor's backend(s) or use the default LVM:

- Backend devices invisible to end user
- Consistent API regardless of backend
- Differentiating features exposed via custom volume types and extra specs

Cinder implementations vary by storage platform. Some do not have simple implementations and require a lot of manual work, while others are completely API driven, making the Cinder driver implementation relatively simple. Cinder's plug-in architecture enables storage choices and the mixing together of different storage vendors' arrays. As of the Juno release, there were 31 different Cinder drivers to choose from, covering a wide range of storage vendors.

What does the right OpenStack block storage look like?

Merely having a Cinder driver while enabling block storage to be accessed in OpenStack is not in and of itself a differentiator. Each vendor that has a Cinder plug-in has its own merits, and some excel at specific use cases. Some questions to ask yourself when considering Cinder block storage:

- Does it meet the performance needs of my cloud workloads?
- Can it provide consistent performance, not just high performance?
- Does it account for noisy neighbors?
- If it uses flash, is it architected to optimize for flash's benefits?
- Is it cost effective and able to maximize density?
- Does it enable full automation?
- Is it tested and supported? Will it really work in OpenStack?
- Is it DIY? Does it have professional services? Support? All of the above? None?

When it comes to storage for OpenStack, one size doesn't fit all, and you need to choose the right tool for the job. The idea that a single pooled storage system can capably consolidate all three tiers of storage will likely end badly. Any discussion about distributed storage solutions for cloud should include commercial options alongside open-source ones. In the case of cloud storage for performance-sensitive applications, the options provided by open source as well as legacy storage vendors are significantly lacking.

Very few production-quality distributed storage systems are available today. Some popular open-source storage solutions were architected for optimized secondary storage use cases, such as file servers and unstructured data. When it comes to performance-optimized workloads, however, these solutions were simply not built with this use case in mind. To help identify the right tool for the job, here are some key considerations for anyone evaluating performance-optimized distributed storage for cloud infrastructures:

Consistent performance
Tier 1 applications generally expect consistent latency and throughput from storage systems. Achieving this in a multi-tenant legacy storage system is challenging enough, but in a complex distributed system, it becomes an even larger problem.

Performance control
Without the ability to provision performance separate from capacity and dial in performance allocations for each application, a storage system will quickly be dominated by the "noisiest neighbor," which starves resources from more critical applications. Ensuring that your storage provides consistent performance (and the ability to control and guarantee that performance) is critical in large-scale cloud deployments.

Flash aware
With cost rapidly declining, flash is now appropriate for a large percentage of Tier 1 use cases, particularly when combined with data reduction. However, plugging SSDs into a storage system designed around disk is a recipe for problems. Disk-based architectures can't deliver the maximum IOPS from flash, while wear and endurance are real concerns due to write amplification. Only native flash storage architectures can deliver both the performance and endurance required for most Tier 1 applications.

Data reduction
By definition, Tier 1 storage is going to use faster, more expensive media – either fast disk or preferably SSDs. In-line deduplication and compression, without impacting performance, are critical for making the system cost-effective and achieving maximum density in a cloud environment.

Manageability

APIs are an often overlooked component of block storage in cloud environments. A robust API that lends itself to automation of all aspects of the storage system is imperative to achieve the promised operational benefits of cloud. Having APIs alone isn't enough. Are they robust and complete? Can your chosen storage solution withstand all the API calls you're going to make?

Professional testing and support

Tier 1 applications are called mission critical for a reason. Ensuring that the storage hardware and software you use is thoroughly tested and supported helps minimize the downtime and errors encountered when these platforms are deployed in production environments.

Qualified hardware

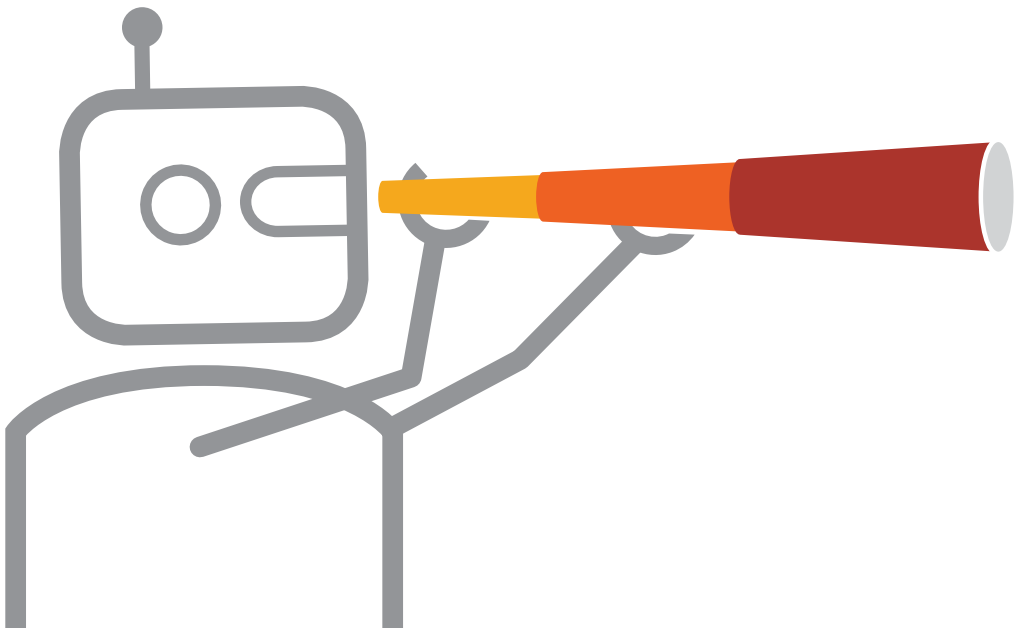
Consuming storage in an appliance form factor has real, measurable benefits. Vendors bear the burden of ongoing qualification of the hardware and software while providing a single resource for support without finger pointing. Firmware bugs in commodity storage controllers and drives are a very real problem, and system vendors are in the best position to identify and correct or work around these issues. Why resource an effort so far outside of your core competence when your vendor will aggressively ride the hardware cost curve for you?

Is it DIY?

Do-it-yourself storage solutions can save a lot of money, but don't forget the hidden costs of time. If no services are available to help you install and configure the system, deployment can be slow and complicated. If something goes wrong, who provides support? Will you have to spend cycles triaging what broke or possibly figuring out who you can escalate it to if/when you're unable to address the issues yourself? Services and support come with a dollar cost but can often cost less than lost revenues or productivity when you're trying to install and troubleshoot issues yourself.

Availability and scalability are notably missing from this list: These traits should be viewed as table stakes in any tier of storage. The key attributes listed above are unique to Tier 1 storage and are seldom delivered by optimized secondary storage systems. It is clear that certain tools, while good for other things, simply weren't intended for high-performance use cases in a cloud storage infrastructure.

As the IT industry shifts from the classic monolithic and static operational models into the new era of dynamic, agile workflows of cloud computing, it's time to look beyond traditional storage hardware architectures and consider products that are built from the ground up for the next generation of applications and operations.



Conclusion

Private clouds will continue to grow and gain adoption as IT looks to keep pace with business demands. Jonathan Bryce, in his opening keynote at the Atlanta OpenStack Design Summit in May 2014, spoke about the necessary shift to IT infrastructure that can "match the velocity of your software development." With increasing demands being placed on enterprise IT departments to keep up with the accelerating pace of business, the infrastructure – storage in particular – needs to enable more dynamic, efficient application and IT service delivery.

Visit **solidfire.com/OpenStack** and learn how SolidFire's OpenStack solutions can power your private cloud.

References

1. QuinStreet Enterprise. 2013 Cloud Computing Outlook. Available at http://www.itbusinessedge.com/slideshows/cloud-adoption-trends-favor-private-cloud-two-to-one-04.html. Accessed March 17, 2015.
2. Gartner & Co. Five Refining Attributes of Public and Private Cloud Computing. Gartner #G00167182. May 5, 2009.

   All statements in this report attributable to Gartner represent SolidFire's interpretation of data, research opinions or viewpoints published as part of a syndicated subscription service by Gartner, Inc., and have not been reviewed by Gartner. Each Gartner publication speaks as of its original publication date (and not as of the date of this article). The opinions expressed in Gartner publications are not representations of fact, and are subject to change without notice.
3. RightScale. 2015 State of the Cloud Report. Available at http://www.rightscale.com/lp/2015-state-of-the-cloud-report?campaign=701700000012UP6. Accessed March 17, 2015.
4. PRWeb. OpenStack Revenue Expected to Hit $3.3 Billion by 2018. Available at http://www.prweb.com/ releases/2014/09/prweb12144923.html. Accessed March 17, 2015.
5. OpenStack SuperUser. OpenStack User Survey Insights: November 2014. Available at http://superuser.openstack.org/articles/openstack-user-survey-insights-november-2014. Accessed March 17, 2015.
6. OpenStack Foundation. OpenStack Powers Demanding Production Workloads Worldwide. Available at http://www.openstack.org/user-stories. Accessed March 17, 2015.