Cloudy Weather for Enterprise Storage
– an Enterprise Storage Veteran Looks Forward

July 2014
by Dave Roberson

Executive Summary

This whitepaper looks at the root causes of user demands for better, cheaper storage services and IT's options to meet those demands. In particular, it explains how the shift in value off proprietary hardware and onto open software stacks is happening in storage just as happened over the last decade with server virtualization.

Enterprise storage as we have known it for the last four decades is facing significant internal and external pressures. The rapid growth of inexpensive or even free public cloud storage for consumers has created a bring-your-own storage grass-roots revolution demanding IT deliver simple self-service storage at costs far below current chargeback levels.

At the same time the capacity requirements for user-generated content from sources like mobile applications is becoming a driving force in storage expansion. While public cloud options exist for enterprises they still have often unacceptable gaps in security and control despite their favorable economics compared to legacy on-premise storage.

All of these factors are making it both technically and economically indefensible to just keep expanding your reliance on existing legacy NAS and SANs that suffer from years of vendor lock-in and are burdened by costly proprietary hardware. Something new is needed and the option of pushing data out to public cloud providers just isn't feasible for organizations who need their storage to protect intellectual property and comply with regulations.

But there is another cost-effective alternative. A new approach called Software-Defined Storage delivers reliable and secure private cloud storage far cheaper than legacy NAS/SAN storage and less than half the cost of business-grade public cloud storage. Building on an open source platform like OpenStack Swift supplemented by advanced enterprise-grade management systems like SwiftStack lets you quickly deploy private cloud storage-as-a-service on inexpensive commodity hardware. Swift and SwiftStack can help you deliver on premise storage that satisfies both user expectations and the enterprise's needs for low cost, security, and availability.

Audience

This tech brief is intended for IT storage professionals working in enterprises and government. Anyone responsible for delivering secure and reliable storage to their organization's users, and to the applications used by those users, should find this material relevant to the challenges they face.

About the Author

Dave Roberson works with SwiftStack as an advisor and brings many decades of experience from all corners of the enterprise storage market. He has been a private investor, business consultant and served on several boards since 2011. Roberson is a member of the board of directors at Brocade Communications Systems, Inc., Quantum Corporation, and TransLattice, Inc. Roberson most recently served as senior vice president in the enterprise server storage and networking division of Hewlett-Packard and was responsible for managing the storage business unit of the company. Prior to that, he was the CEO at Hitachi Data Systems Corporation, a global storage systems and solution provider.

Why don't your storage users love you anymore?

Why is your Enterprise Storage becoming such a disappointment to your users? Why are your IT storage systems that looked ambitious and forward-thinking just last year no longer meeting their expectations?

Likely it is isn't due to any failing on your part or even on the part of your storage vendors. Instead what has happened is a series of small changes - changes in the applications and devices that generate and consume data, advances in storage technology, the rise of storage as a service, and most importantly, changes in user's expectations.

Consumerization of Storage Inevitable

Just as happened with desktop computing and mobile devices, the consumerization of storage is happening, perhaps even faster than in the other two cases. As shown by the prior waves of bring-your-own-computer and bring-your-own-device you can either get on the bandwagon or get run over by it.

While this consumerization of storage is occurring, the data needing to be stored, secured, and managed is growing exponentially. IDC has predicted a 50-fold growth in Exabytes of storage deployed between 2010 and 2020. This ubiquitous growth in capacity isn't just happening in the enterprise. Consumer-generated data is estimated to have been over 1,900 exabytes back in 2012 with roughly 70% of that overlapping with data held or used by enterprises.[1]

In the business-oriented portion of the public cloud, Amazon has grown its capacity at 50% per year[2] selling storage as a service to businesses on and off the web. Google, Rackspace, and Microsoft have similar offerings although they trail Amazon's dominant share of cloud revenue estimated at over $1 billion in the 4th quarter of 2013 alone[3]. According to Kevin Turner, COO at Microsoft, their commercial cloud services grew more than 100% year-over-year in 2013.[4] Even more importantly, however, public cloud storage serving consumers has accumulated over 625 million users and are growing at over 25% a year[5]. An estimated 30% of U.S. broadband users are using personal cloud storage services.[6]

Looking back historically, people now personally manage more storage capacity and data as consumers, spread across their social media sites, photo sharing streams, and growing cloud storage services, than a typical enterprise did 20 years. This alone has the potential to both demystify and devalue enterprise storage in the minds of a lot of users.

---

[1] THE DIGITAL UNIVERSE IN 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East, December 2012, By John Gantz and David Reinsel, IDC - idcdocserv.com/1414

[2] Amazon earnings call reported on GigaOm October 24th 2013 – link

[3] Synergy Research Group's (SRG) latest snapshot of the infrastructure as a service (IaaS), February 17th 2014 – link

[4] Microsoft Earnings Report for FY14 Q2 – link

[5] IHS iSuppli Mobile and Wireless Communications Service at information and analytics provider HIS, September 7th 2012

[6] nScreenMedia cloud storage usage study, April 16th 2014 – link

Storage Costs Keep Dropping

Cost is inevitably tied to value in most people's minds. When I first entered the market mainframe storage cost $1 million per megabyte while today you can rent a petabyte in the cloud for $100 per month. One can imagine a future where storage is eventually infinite and free[7] although how existing personal cloud startups can ever survive such a drop in prices is a big unknown.

Even now, the cost per GB for public cloud storage has gone down to the point where it is almost free, at least in the minds of consumers. Cloud storage services like Dropbox and Google Drive start out giving every user from 2 to 15 GB of free storage. Need more storage? Prices seem to drop monthly, but additional GBs cost consumers anywhere from $.10 to less than a penny per GB. Google, for example, now offers a TB of storage for only $9.99 per month[8]. There are other hidden fees with the business-oriented portions of the public cloud, such as transfer fees, but the total is still far cheaper than legacy on-premise storage.

Keep in mind that those low cost public cloud prices likely aren't based on their actual costs. Although the true economics are usually buried in some "Other" category in their financial reports, there is little doubt that large firms like Google and Amazon are subsidizing the land-grab prices of their public cloud storage services with their main revenue sources from web search advertising or e-commerce sales.

Public Cloud's Unfair Advantages

Economies of scale aren't in your favor either compared to the public cloud providers. For example, Google's Q1 2014 earnings report showed that they spent over $2.3 billion on data center infrastructure in just that quarter alone[9]. This was nearly double what they spent a year ago giving them cost scaling benefits that no other enterprise IT group could match.

In some cases, entrenched software vendors are leveraging the foothold they already have with your users to provide reasons to use cloud storage beyond just low prices. For example, Microsoft's most recent versions of its best-selling Office suite prompt users to store their files in OneDrive's cloud storage right within the software itself. According to Microsoft, over 250 million people already use OneDrive – that's one-fourth of the 1 billion people using Microsoft Office[10].

Another big advantage that the consumer cloud storage vendors have over legacy enterprise storage is how simple it is for users to get started. Like prior waves of software as a service (think Salesforce vs. on-premise CRMs) the current surge of consumer storage as a service bypasses the barrier of organizational red-tape. Consumer services use a self-service model where you just have to fill out a registration form and worst-case supply a credit card number. There are no POs, no contracts, no review and approval cycles, no budgetary issues. The charges are so low they aren't even worth expensing.

---

[7] Aaron Levie, CEO of Box on Box website blog, April 24th 2012 – link
[8] Google Drive pricing as of April 2014 – link
[9] GigaOm report on Google infrastructure spending in FY2014 Q1, April 17th 2014 – link
[10] Microsoft by the Numbers as of April 27th 2014 – link

Performance Isn't Everything for Users

Performance might seem like a key differentiator for legacy enterprise storage. The reality is that for most users the performance of public cloud storage services is essentially "good enough". Users are generally happy with the speed of access to files stored in consumer public cloud services even though it is many times slower than on-premise enterprise storage. And the level of performance isn't just dependent on how well the consumer clouds are provisioned - most consumers, especially in the US, are still limited to Internet bandwidths well under 25 Mbit/second — well blow most and Wi-Fi and a far cry from Gigabit Ethernet speeds.

When used from PCs most cloud storage services use locally cached copies of files on the computer's hard drives to hide some of the access delays. This is a clever hack but opens a significant security risk for data theft as PC hard drives are rarely secured using strong encryption and good security practices. Costly legacy enterprise storage that is able to deliver a lot IOPS clearly still makes sense for use cases like large transactional databases. However, for the relatively small and infrequently updated files typically stored on cloud storage services this level of performance is overkill and definitely not a cost-effective investment.

Cost ≠ Value

All of this leads to a drop, real or perceived, in the value your users see from enterprise storage. These changes in perception might be expressed as "I can go buy a 2 TB consumer hard drive for $85, so why does the same size drive for your enterprise array cost so many times as much?" Or it might be "I can get a TB of cloud storage for $10 so why are your monthly chargebacks to my department $75 per TB?"

For legacy storage systems based on proprietary hardware your costs, of course, are very real no matter what users perceive. Older legacy NAS or SAN storage is often already expanded to their limits, requiring very large up-front investments for new storage. Assuming they haven't hit their capacity limits adding an extra enterprise-grade hard drive storage might only cost $2/GB. This seems like a small amount even though it is roughly 40 to 50 times what consumer commodity hard drives cost per GB.

But beyond that one-time hardware cost, you're faced with ongoing costs for software licenses, facilities, power and cooling, maintenance and warranties, backup systems, plus quite costly skilled IT FTEs for break-fix and day-to-day monitoring and management. Using aging SAN and scaled-out NAS proprietary hardware to deliver cloud storage is not just technically challenging, it is economically disadvantaged from the start with all-inclusive TCO of over $3,000 per GB per month over its lifetime[11]. Even looking at only the incremental costs of expanding legacy NAS/SAN storage, it ends up failing to be competitive.

---

[11] Wikibon Data Storage TCO Study, March 13, 2014 – link

What are Your Options?

Clearly a new approach to enterprise storage is needed to address these gaps. If you don't embrace a new means of deploying storage to users, using a self-service model and at far lower cost, here's what will inevitably happen:

1. Your users, like any other consumers, will be using cloud storage services at home for their personal content, accessible from any of their PCs or mobile devices, mostly without paying anything at all.

2. They'll realize how useful these services are to them, and will come to work expecting you to deliver the same sort of easy to use cloud storage for their work files, at the same low (free!) price, and with the same sort of simple self-service model.

3. Not having your storage costs covered by web search advertising or venture capital investments, and being subject to both internal policies and external regulations, you probably won't be able to offer the same level of free cloud storage services on your existing storage infrastructure.

4. When the users don't get what they've come to expect, they'll just start storing their work files in the same cloud services they're already using for their cat photos.

5. The net result will be that your carefully managed and secured enterprise storage will be ignored. Instead, users will rely on consumer cloud storage that lacks any serious capabilities for security, control, or backup for their work files.

6. And even worse, you're still on the hook to pay for all of the expensive legacy storage hardware your users aren't using.

Here's What Won't Work

This gap between user's expectations and IT's delivery doesn't have to be an unproductive tussle but needs to be resolved. While you could just ignore the whole mess and hope it goes away that option can clearly be career-limiting as a long-term strategy.

You could try winning by fiat – issuing policies forbidding any enterprise data from residing anywhere other than in approved enterprise storage. But, as the prior bring-your-own-whatever revolutions have shown, control by fiat rarely works for any length of time.

You could try to explain all of the benefits of managed enterprise storage vs. public cloud consumer services. Perhaps the costs of legacy storage could be offset in your user's minds by its security, availability, and reliability. But likely that won't resonate unless there has recently been a widely publicized data theft or cloud service outage.

So, if you're not going to ignore or squelch the user demand for inexpensive ubiquitous cloud storage, what are your remaining options?

Is Public Cloud Storage Even an Option?

You could adopt the public cloud storage service that has a business-flavored variant, like Dropbox for Business, and try to reconcile its limitations with your requirements for transparency, control, and availability. Depending on what regulatory burden you face and your appetite for risk this sort of quick fix may work for a time. Or it might not.

Unfortunately, most of consumer public cloud storage services still lack comprehensive support for enterprise-level storage management features like auditing and encryption, backup and recovery, retention policies, usage quotas, monitoring and alerts, etc. Their origins as file synching tools or collaboration platforms bring along a lot of baggage that users might like but which aren't likely to mesh well with the data protection and privacy policies of most organizations.

For example, many of them have handy sharing features that let users toss access rights for a file across to anyone with an email address with very little effort and little or no options for oversight or control. This may be great for user productivity but for concerns like regulatory compliance, customer privacy, or intellectual property protection they might as well be putting the company's data up on Facebook. You could, of course, just order users not to use those handy features or even block their use if the cloud provider allows that, but those moves inevitably lead back to the same old failure-of-fiat problem, with unhappy users working their way around any restrictions.

Beyond these issues of data security and control, integrating public cloud storage into existing internal IT systems can be troublesome. The providers are starting to add enterprise features, but getting any true interoperability in even simple things like single sign-on or chargebacks may involve more compromises than benefits. You'll likely end up having to either place a lot more trust in service providers than warranted or put far too many IT resources into temporary workarounds.

However, there is a solution that can meet both user expectations and the enterprise's needs – private cloud storage. This approach can let you retain control of your company's intellectual property while still keeping your users happy and productive. Just as important, it can help you deliver reliable enterprise-grade storage to your users at costs well below what even the recently discounted public cloud storage services charge.

Providing Cost-Effective Private Cloud Storage

So if you've accepted that you need a new way to provide your users with cost-effective private cloud storage, how do you get it done?

In the past, the only way to get required levels of both reliability and performance from enterprise storage was to throw specialized proprietary hardware at the problem. CPUs and networks were slow, RAM was scare, and capacity on hard drives was orders of magnitude more expensive. Every byte and bit counted, every IOPS and CPU cycle had to be optimized and conserved.

Ensuring that data was protected and rapidly accessible used to require building that technology directly into the silicon in the form of very expensive RAID controllers, proprietary hard drives, and specialized SAN interconnections. Entrenched legacy storage vendors like

EMC have become large and profitable by locking in customers while making margins of 60% or more on required parts, maintenance, and services.

Fast, Cheap, Reliable — now you can Pick Three

Today things have dramatically swung in the other direction as technological advances have made computing and storage infrastructure fast, cheap, and reliable. The key, as in so many other areas of IT, is to do it in software. The steps in this ongoing change aren't anything new, even though they are very disruptive to legacy storage vendors. These should sound like familiar and very rational steps to anyone who has been in IT over the last two decades:

1. Consolidate from under-utilized, over-provisioned silos of expensive proprietary hardware onto highly-utilized shared nodes based on generic commodity hardware.

2. Scale it up linearly and infinitely just by adding more nodes rather than having to deploy whole new silos (or arrays) once limits are hit.

3. Deliver availability and redundancy via intelligent software design rather than overbuilt and soon to be obsolete proprietary hardware.

4. And manage it from anywhere using easily accessible browser-based tool suites.

This shift off proprietary storage hardware and onto intelligently designed Software-Defined Storage is going do to legacy enterprise storage hardware what server virtualization did to proprietary server vendors – i.e. make them largely irrelevant.

In storage however, the economic impact of this shift can be even more pronounced when the platform is based on open source software rather than the proprietary solutions like those from VMware and Citrix that captured most of the server virtualization market. While there may always be a place for legacy storage hardware vendors, the uses cases where capacity demands will grow the fastest – user-generated content and non-transactional "big data" archives – are better and more cheaply served by software-defined storage.

Do It Yourself?

However, even with the economic advantages of cheap fast commodity hardware and the flexibility of a software-defined, rather than hardware-restricted, storage infrastructure, designing and building your own private cloud service from scratch could be difficult and risky, draining IT resources from projects with better payback. Building your own management tools when there are commercial alternatives just doesn't make sense.

Despite their often serious limitations, public cloud storage providers do let you sidestep this potential pitfall. They all provide some limited form of management console and some even have made tentative first steps to integrate with enterprise authentication services at some limited level.

But it is fairly likely you'll need far more than they can provide. Sometimes it looks like doing it yourself is the only way to address both user expectations and IT's requirements. But for private cloud storage that really isn't necessary given the groundwork already done by open source storage platforms, some of which are already proven in production use at large IT service providers like IBM and HP.

Open source projects can have their limitations of course. Ideally, pairing an innovative open source platforms with a proven enterprise-grade management system will give you a solid and secure, but still low-cost, platform for private cloud storage services.

SwiftStack and OpenStack Swift — Enterprise Grade Private Cloud Storage

One of the most promising of these open source private cloud storage platforms is OpenStack Swift. Swift is an open source object storage platform written in Python, a modern high-level programming language. The object storage provided by Swift is considerably more flexible and scalable than file- or block-storage options from legacy NAS and SAN vendors. Swift provides both backward-compatibility for users via filesystem gateways and numerous integration options for application developers and IT. Key to its flexibility is a HTTP-based RESTful API, making access to Swift and integration of its storage simple for websites, mobile applications, and enterprise systems.

Swift's object storage uses an eventual consistency approach that doesn't require global locks on data being updated. While this level of consistency might not work well on database servers or other systems that need ACID (Atomicity, Consistency, Isolation, Durability) transactions, it is ideally suited to the current capacity growth areas of user-generated content that rarely if ever need to handle simultaneous concurrent updates.

Swift's approach to consistency makes it very effective at delivering high-performance storage on low cost commodity hardware through the use of redundant replicas of data. Its automated software-based data replication and distribution ensures that data remains available even in the face of multiple media, node, or network failures. Swift's automated replication combined with SwiftStack's storage policies let you easily match the investments in storage performance to the actual needs of users and applications.

Swift also uses a self-service provisioning model similar to that used by public cloud providers. This lets IT setup a storage pool for users and applications to tap as they need. Applications can define and extend their own storage without any overhead or delays required for provisioning by IT.   And costs for Swift storage in production deployments has ended up at less than half of business-grade public cloud storage from Amazon or Google, even after their recent price cuts.

So if users aren't satisfied with your existing enterprise storage offerings, there are real alternatives to putting things up in the public cloud and hoping for the best. Combined with advanced management suites like SwiftStack, Swift's open source object storage platform can help you take the next step towards Software-Defined Storage, providing both the simplicity that users want and the low cost and control IT needs in private cloud storage.

Want more information?

For more information on low cost private cloud object storage using OpenStack Swift, visit swiftstack.com/openstack-swift. For more information on SwiftStack's management capabilities, deployment automation, and intelligent software-defined scalability, please visit www.swiftstack.com.

Copyright © 2014 SwiftStack Inc. All Rights Reserved