451 Research® | PATHFINDER REPORT

Object Storage as Part of a Hybrid Cloud Protection Strategy

COMMISSIONED BY

VEEAM

APRIL 2019

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 THE AUTHOR

DAN THOMPSON

RESEARCH DIRECTOR - MTDC

As a Research Director for 451 Research, Dan Thompson provides insight into the Multi-Tenant Datacenter (MTDC) market space. Dan is particularly focused on MTDCs that are trying to move up the stack to offer additional services beyond colocation and connectivity. These services may include disaster recovery, security, various forms of cloud and other managed services. He also assists the 451 Research Information Security group when their interests overlap.

Executive Summary

Attempting to properly manage the ever-growing mountain of data inside most organizations around the globe can feel like an impossible battle, and trying to protect all that data is just as challenging. The good news, however, is that there are options available to companies of all sizes that were previously within reach of only the largest enterprises. One such option is cloud-based object storage, which can be leveraged as part of a tiered storage approach to better align the cost and performance of data that will be retained longer-term. This approach can be an easy win for IT organizations because it allows them to move backup data off-site and pay as they grow rather than spend a lot of money up front for a dedicated on-site solution that is vulnerable to the same catastrophic events as the live data.

Key Findings

- **Enterprises today are managing hundreds of terabytes to hundreds of petabytes of data, and that data is steadily increasing over time.** This onslaught of data is creating several pain points for organizations, not least of which is the high cost of storage options to house all that data.
- **As companies look for ways to protect all this data, the backups become part of the overall problem because they, too, inevitably increase in size.** In response, organizations must either purge data (which may not be an option depending on the industry and country of origin) or buy more storage options to house it all.
- **IT organizations should work with senior leaders/stakeholders and IT implementers to create a data management plan that outlines what data types need to be retained and for how long.** This plan should include a structure by which IT staff can easily identify the data earmarked for long-term data retention.
- **Thanks to the cloud, many options are now available to organizations for robust, scalable and multi-site data protection and storage.** Enterprises should investigate object storage and other storage tiering options to maximize the cost and performance of data resiliency. It is also worth investigating software platforms that allow for automatic and intelligent tiering of data to further bolster the data protection plan.

Introduction

Enterprises all over the globe are bearing the weight of an ever-increasing mountain of data. This growing mass of data is taxing for the business; it represents significant and ongoing capital outlays in addition to substantial risk because IT staff are tasked with keeping it available and protected for decision-making or regulatory needs. As the data grows, so do the challenges with keeping it backed up and readily available should a disaster arise, or simply an accident that requires a recovery. As IT leadership evaluates how to deal with these challenges, it should become clear that there are more tools available to them than were available even just a few years ago. It may also become clear that the tools traditionally leveraged by IT organizations for backup and disaster recovery are either no longer adequate or need supplementing with cloud-based products and other solutions for long-term data retention.

One option that is now available to organizations, but potentially underutilized or not yet explored, is cloud-based object storage. In terms of the challenges mentioned, the question becomes whether the cloud can in fact address these issues. And if so, how can object storage be leveraged for the best outcome? This report examines the following:

- Industry expectations of data growth and the most common storage challenges that this causes
- How cloud object storage is changing the possibilities for most organizations
- How hybrid data protection strategies can help organizations of all sizes
- Recommendations on what to consider when assessing your data protection goals and the use of cloud storage as part of your strategy.

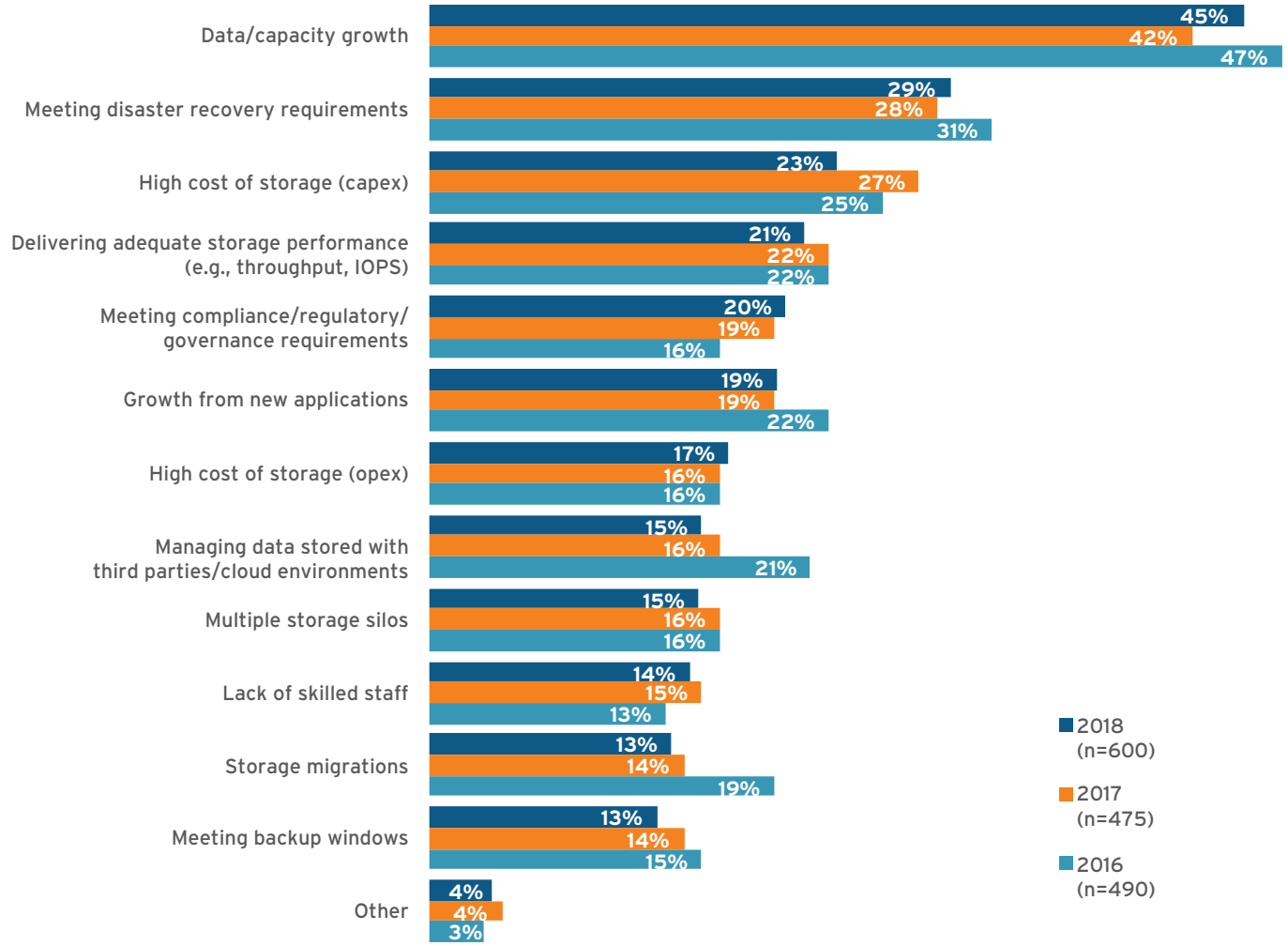
Although cloud-based storage and object storage are definitely not new, a path to leverage them as part of a data protection strategy has not always been straightforward. Fortunately, many backup and data protection vendors are making it easier to take advantage of this type of storage as part of an overall data recovery and resiliency plan. By using storage in the cloud, enterprises can add geographic separation for their backup locations from the primary locations of their workloads – something that is easily pushed aside, or expensive to implement, with more traditional architectures. Additionally, by its very nature, cloud-based object storage can be globally accessible, adding a capability that would be difficult, if not impossible, to replicate for tapes or other online or offline media. Finally, cloud-based storage, including object storage, allows companies to take advantage of different tiers, which operate at different price points, allowing customization of price, performance and availability, for a truly scalable, 'pay as you grow' model.

The State of Data

It is no surprise that the volume of data most enterprises are managing is rapidly increasing, and organizations are left to struggle with several problems regarding storage. Over the last several years, we've asked IT leaders to identify their top storage pain points as part of our Voice of the Enterprise: Storage, Organizational Dynamics study, and some interesting insight can be gleaned from looking at how those answers change over time. Data/capacity growth, meeting disaster-recovery requirements, and the high cost of storage (capital expenditures) are the pain points that consistently rise to the top.

Figure 1: Top Storage Pain Points

Source: 451 Research's Voice of the Enterprise: Storage, Budgets & Outlook 2018
Q: What are your organization's top pain points from a storage perspective? (Select up to 3.)

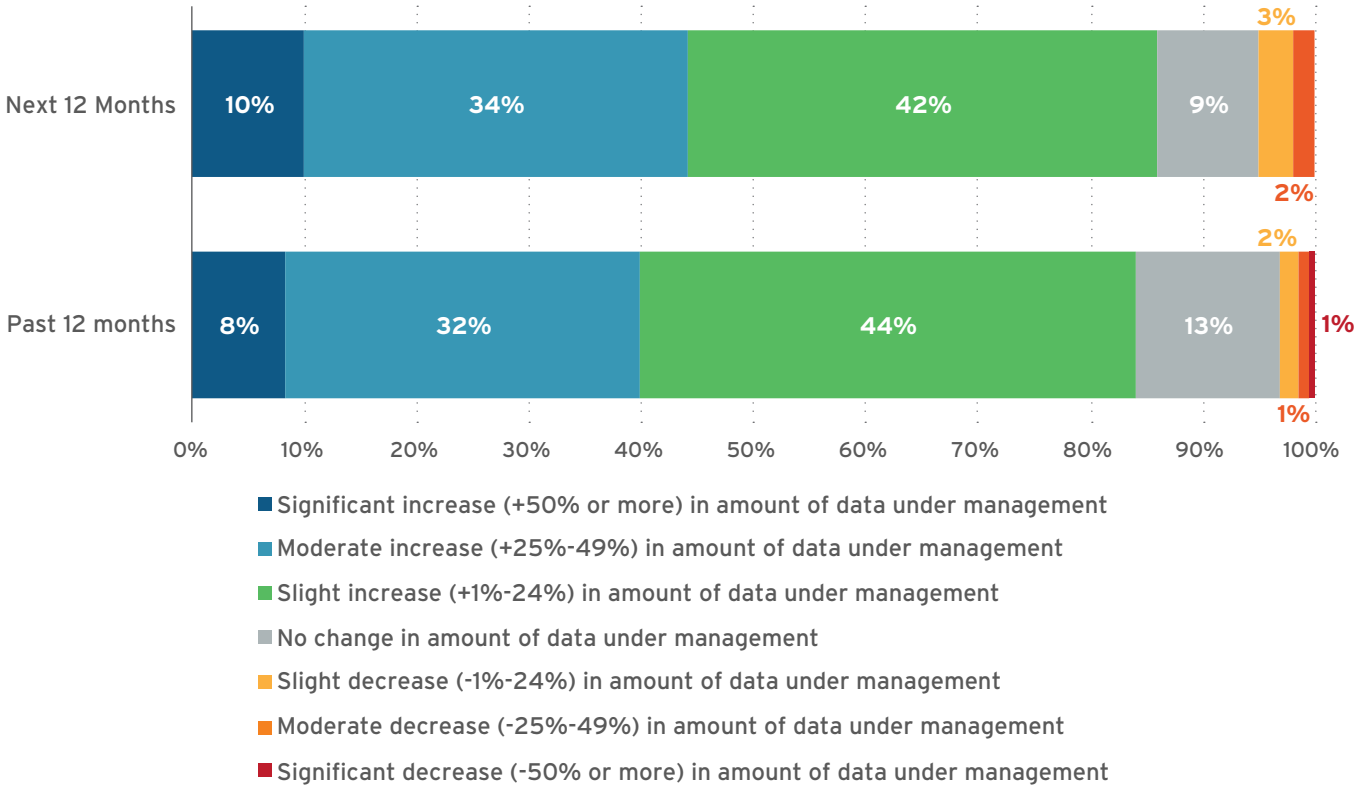


Regarding data growth, it's interesting to note that companies tend to buy storage for three-year periods. One can almost see this short-term lifecycle playing out in Figure 2 below. IT managers are out of space and need more in the year ahead, but they worry about the high cost of new hardware. Following a storage purchase, they then worry about how to keep the data growth at bay to avoid another storage purchase too soon. Rinse and repeat.

Figure 2: Expected Change in Data Under Management Within 12 Months vs. Past 12 Months

Source: 451 Research's Voice of the Enterprise: Storage, Organizational Dynamics 2018

Q: Over the past 12 months, has there been an increase in the amount of data your organization has under management, a decrease, or has there been no change?



The second major storage pain point cited by users is meeting disaster-recovery requirements. Typically, when we think of disaster-recovery requirements, we think of recovery time objectives (RTOs) and recovery point objectives (RPOs), but beyond these, we see a couple of self-perpetuating issues taking place. One is that as the data sets that need to be protected grow, the size of the backups themselves also grows. Second, as the backups grow, so does the portion of those backups that need to be considered for long-term retention (essentially duplicates of backups or partial duplicates). Since backup chains have traditionally seemed delicate, administrators will keep more copies than perhaps they need to just to ensure they've got something to restore from (if this one fails, I'll try that one), which of course just eats up more space.

Finally, when it comes to the high cost of storage, there is a very familiar conversation around the fact that the price per gig continues to fall. Storage vendors often reference this ('Our price per gig is less than our competitors'), but the truth is that the problem is now one of scale more than price. Sure, the price per gig has fallen, but many organizations are needing petabytes of storage now, where in the past, terabytes would have been enough. So while it is helpful that the price per gig has come down, the overall price of the storage capacity remains high.

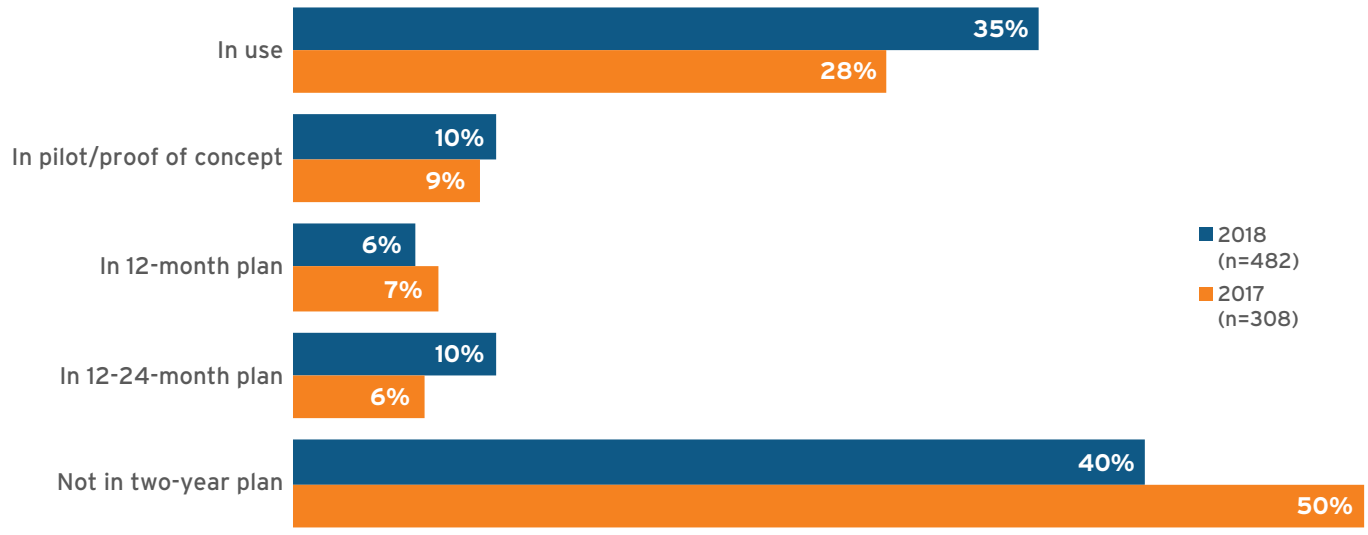
The Rise of Object Storage

For the last two decades, the storage industry has generally focused on delivering basic file- and block-level networked storage services. These systems have evolved over the years, and these days appear in the form of dedicated and mostly proprietary storage-attached network (SAN) and network-attached storage (NAS) systems. However, work began in the mid-1990s to create a storage architecture that identified a set of data as an individual entity (or object) rather than simply a location on a disk. These systems maintain an independent database of pertinent information about that data itself, making it possible to collect and categorize a relatively limitless amount of metadata about the stored information that can be used to index, classify and enable extremely flexible data management unlike any other storage platform.

Although enterprise adoption of object storage has been somewhat slow, we're now seeing it pick up year over year. Historically, vendors have struggled to articulate the value proposition of object storage over more traditional storage options. Many storage vendors adopted a model of creating devices that required additional software licenses to enable the object storage functionality, effectively raising the price on what should have been a lower-cost storage tier. That is, of course, until Amazon Web Services (AWS) came along with its S3 storage option.

Figure 3: Adoption Status of Object Storage, 2018 vs. 2017

Source: 451 Research's Voice of the Enterprise: Storage, Organizational Dynamics 2018



The object storage model is the underlying technology for all cloud-based storage systems, in part because of the platform's ability to dynamically scale across massive, multi-node storage systems, but also because the metadata capabilities of object storage don't place any theoretical limitations on the number or size of objects in storage. This metadata has also allowed major cloud storage providers to easily migrate objects between multiple tiers of storage, giving customers the flexibility to optimize a cloud storage environment based on whatever combination of cost, performance, availability and data protection best suits their data governance requirements.

While the AWS platform may have gained wide popularity first, many other vendors have built their platforms utilizing object storage as well. Offerings such as Microsoft Azure Blob, IBM Cloud Object Storage and Google Cloud Storage, just to name a few, now host tens of trillions of files using object storage – not only because object storage is directly accessible via HTTP or other storage interfaces, but because object is the only platform capable of providing the advanced data protection, automated tiering and immense scalability required for such a massive storage undertaking. Perhaps most important, the metadata capabilities of object storage provide the extended identification, classification and ownership information necessary to establish policy and automate functions such as billing, tiering and custom access management.

Beyond the public cloud providers, several others have adopted object storage. Storage vendors have worked to incentivize their channel partners to build services around these platforms, resulting in managed service providers (MSPs) and smaller cloud service providers (CSPs) having object storage in their portfolios. It is not uncommon to see MSPs and CSPs offering object storage tiers for production storage, long-term data retention, backup repositories (short- and long-term) and other use cases. We've also seen service providers combining the object storage capabilities of the various arrays available on the market today with other features such as replication outside a metro area, for example – a capability that would have been cost-prohibitive for smaller companies to pull off just a few years ago. The expertise on tap from the various service providers, combined with the overall usability at the public cloud level, means object storage can be within reach of companies any size.

Hybrid Data Protection Using Cloud-Based Object Storage

Storage use has stratified somewhat into two major use cases: flexible, high-performance block/file primary storage for cloud-based production workloads (on-premises and off); and object stores for secondary storage uses with more flexible performance needs. It's this second-tier object storage platform that's most rapidly affecting the very model of business continuity (BC) and disaster recovery (DR) because of the nearly limitless capacity and economy of scale that it offers to customers. Furthermore, the off-site nature of cloud storage addresses one of the key issues in BC/DR, which is geographical separation from the primary site. It is worth noting that storage, on-premises or in the cloud, has a somewhat greater need for resilience than the applications that utilize it. A failed application can simply be restarted, but lost data is lost forever, so data storage demands a different level of vigilance to ensure that data either persists or can at least be reconstructed in the event of a system failure.

Historically, the only way to protect systems from all forms of disaster was to build hot mirror sites. This costly process was only accessible to the largest enterprises, but this type of system redundancy is now theoretically available, on demand, to every company regardless of size in the form of cloud-based systems. Cloud-based storage and compute resources offer a completely new set of options that could almost never be delivered in an on-premises IT model. Both tapes and disk are limited in terms of their physicality. To access the data on the media, they must be physically connected, and accessible to whomever is needing access. By their very nature, however, backup repositories stored in the cloud are globally accessible and connected. This is not to say that cloud-based backup repositories don't have their own issues to contend with, but they are different issues, as discussed in the examples below.

What might a solution look like that takes advantage of the various options available today? First, consider that some data/applications are inherently more valuable to an organization than others, and thus are typically protected differently. This notion of tiering can be applied to backup frequency, just as it can be applied to the media on which those backups are stored. In fact, this practice is quite common among organizations today. To that end, it's worth considering the specific business unit's expectations for recovery time objectives and recovery point objectives, and ensuring that all the tiers match those expectations.

Let's look at a healthcare-specific example. In the US, the Health Insurance Portability and Accountability Act (HIPAA) of 1996 dictates that medical records be kept at least six years from their creation date or last effective date. From the start, we can see the need to properly identify such data. Healthcare data and systems will likely have a higher priority and perhaps a longer retention period than most 'ordinary' data. Whatever the case may be, all practices regarding this data need to be recorded in the company's data management policies and followed consistently. If we assume for this example that the infrastructure is all on-premises, there should be a window in which backups are stored on-site, such that restores can happen faster. Once that window passes, that backup set can be moved to a cloud provider's first tier of storage for another specified amount of time. Again, once that

window passes, the backup set could be moved yet again, to the provider's second tier of storage for long-term data retention, or it could be deleted, based on whatever the organization's policies dictate. In this way, there are always near-term backups available on-site for fast restore, and longer-term backups available in the cloud in case of a disaster.

It is important to point out that the movement of these backup sets should never be manual, but rather automated in some way. To that end, simply having cloud-based storage, object-based or otherwise, is not enough. This data-mover component and the related process automation is where software vendors can help create a complete platform.

As a second example, let's consider a business based in the EU, where the General Data Protection Regulation (GDPR) mandates that data be kept only as long as necessary, and only to serve the purpose for which it was obtained. In this example, let's assume company policy has stipulated a two-year retention policy, and that the company's infrastructure is hybrid in nature, with key components hosted in the public cloud. In this scenario, as in the last, it makes sense that the backups be retained near the workloads for fastest recovery time; the public cloud's object storage could be used for this near-term storage. After a predetermined period of time – say, seven days – those backups could then be moved to the cloud provider's next lower tier of storage, where it could be kept for the remainder of the two-year retention period, and then automatically deleted. If need be, valuable components not containing user data could also be downloaded on-prem and written to tape or disk for long-term retention.

As we think about these examples and countless others, there are a few key questions that must be answered for a data management program to be successful:

1. Is there a company policy that outlines the retention periods for various data types?
2. Can a storage administrator properly distinguish between a healthcare document, an accounting document and some other arbitrary file that a user created? The crux of this issue is that those creating the data are often not mandated to properly identify or classify it, and those managing the data have no way to know what is valuable and what is not, or what is regulated by compliance standards and what is not. The typical response? Keep everything.

"WE HAVE RETENTION PERIODS OF 10 YEARS, BUT WHEN THEY GET TO 10 YEARS, NOBODY EVER WANTS TO DELETE THEM…"

– INTERVIEW #13, >10,000 EMPLOYEES, FINANCE INDUSTRY, IT/ ENGINEERING MANAGERS AND STAFF

3. How can cloud-based object storage be used to supplement on-premises storage for more flexible recovery scenarios? Is 'recovery in the cloud' a requirement? If so, having the backups stored in the cloud would make a lot of sense.
4. What happens in the case of litigation? Can backups be surrendered to the legal team, with proper chain of command, if they are stored in the cloud?

5. Today, most tape vendors advertise that tapes can last anywhere from 15-30 years in storage. Should we trust that? If we store long-term data in the cloud, do we trust that these companies will still be around over that same time period? One thing is for sure, eliminating tape will reduce some amount of labor and storage costs associated with those technologies.
6. Does the company have a requirement for data to be stored off-site? Most regulated industries have a stipulation for this; however, even for unregulated industries, it is wise to do so. The cloud is an easy way to accomplish this separation of data location.
7. In the case of ultra-sensitive data, is cloud storage even an option? This question will likely need to be answered by the company's legal team through a deep understanding of the various regulations that affect the organization.

These questions underscore the importance of not only considering where backups are stored and for how long, but also that IT organizations have a policy that is consistently followed and can properly identify data within backup sets to ensure that each is protected properly and as regulated. For far too long backups have been viewed as 'just backups' – but it is clear in today's business environment that backups must be viewed as being every bit as critical as the related applications.

Conclusion/Recommendations

IT organizations are tasked with managing and protecting an ever-growing mountain of data. This mass of data often comprises some portion that is a candidate for long-term retention and perhaps some that must be deleted after a certain point, due to compliance mandates or company policy. All these data groups must therefore be managed differently, necessitating a multifaceted storage approach.

The good news is that once all these specific data groups are identified, there are various software platforms that allow IT administrators to properly handle each. Furthermore, companies have an opportunity to take advantage of storage tiering options in the cloud and on-premises as part of this approach to maximize the location, performance and cost of the various data groups. Based on this, IT organizations need to take the following actions:

- **The mountain of data must be protected, and its retention/recoverability assured by a good data protection program.** Part of the reason there's so much unchecked storage growth is that no one wants to be the one who deletes valuable data. Step one of any good data management program is a policy. If your company doesn't have a policy (and a data protection/data management platform that can enact the policy), then make one and get buy-in from senior leaders and stakeholders and IT implementers from the very beginning.
- **Remain current with the evolving compliance requirements of your business.** Many verticals like finance and healthcare have strict guidelines in place for data governance, but privacy laws and other initiatives are always evolving. Ensuring industry compliance must be at the top of any data management initiative.

- **Identify groups of data based on long- and short-term retention requirements.** This is admittedly the hardest part. Data identification can be tricky, but in today's business and legal environment it is necessary. To do this, take into account the regulatory and policy requirements put together in the previous two steps.
- **Evaluate how various tiers of storage can be employed for the greatest performance/cost impact.** Cloud-based object storage has some strong advantages, but it is not a one-size-fits-all solution. Some data will inevitably need to be stored local to the processing for fast access and overall availability. Some data, however, could be moved elsewhere for archival purposes and long-term storage. Consider policy requirements for data separation and evaluate how the cloud can help.
- **Create a system whereby the various data groups can automatically be moved between storage tiers as part of their lifecycle.** For consistency and accuracy, this 'data moving' should always happen automatically, and without human intervention where possible. It's worth investigating how software can be leveraged to accomplish this. A great solution should move data that needs to be retained long term from warm storage to cold storage automatically as the lifecycle dictates.

About 451 Research

451 Research is a leading information technology research and advisory company focusing on technology innovation and market disruption. More than 100 analysts and consultants provide essential insight to more than 1,000 client organizations globally through a combination of syndicated research and data, advisory and go-to-market services, and live events. Founded in 2000 and headquartered in New York, 451 Research is a division of The 451 Group.

NEW YORK
1411 Broadway
New York, NY 10018
+1 212 505 3030

SAN FRANCISCO
505 Montgomery,
Suite 1052
San Francisco, CA 94111
+1 212 505 3030

LONDON
Paxton House
30, Artillery Lane
London, E1 7LS, UK
+44 (0) 203 929 5700

BOSTON
75-101 Federal Street
Boston, MA 02110
+1 617 598 7200

451 Research®