451 Research® | Advisory

# A Deeper Dive Into Hybrid Cloud Storage

**NOVEMBER 2016**

COMMISSIONED BY

**NetApp**®

## 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.
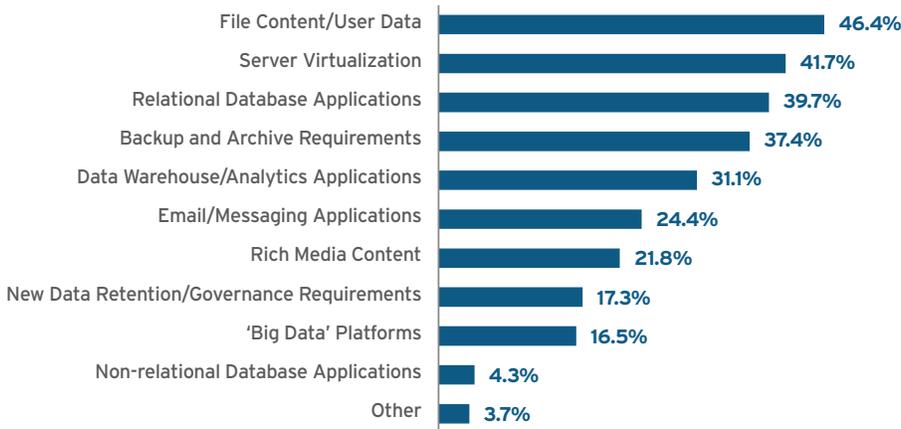
## Executive Summary

There's no doubt that the availability of cloud-based compute and storage services will continue to have a major impact on information management for businesses of every size. The combination of highly scalable, cost-effective and on-demand IT resources in the cloud – as well as the new tools and services that the cloud enables – is changing the economics of enterprise IT; however, the adoption of cloud, and cloud storage in particular, also raises a new set of concerns regarding the availability, protection, management and security of data that can now span multiple locations both on-premises and off. In traditional SAN/NAS storage, these issues are managed internally as part of a monolithic storage model within the firewall, but adopting a hybrid cloud approach means that customers are now being challenged to attain a similar degree of data protection, consistency and policy management across multiple, and often dissimilar, cloud storage offerings.

There are a number of drivers for adopting cloud storage, with the biggest being substantially reduced storage costs without the need to continually acquire new hardware. As seen in Figure 1, the top cause of data growth is File Content/User Data, although that category could easily be extended to include Backup/Archive, Email/Messaging, Rich Media Content and Data Retention/Governance when you consider these all to be unstructured data types. We believe that managing unstructured data will continue to be a key driver for storage customers to consider cloud storage, given the increasing growth of user-created document, audio, video and image files.

### Figure 1: Top Drivers of Data Growth – 451 VotE Storage Study, Q4 2015

*Q. What are the top three drivers of data growth in your Organization? n=647*
*Source: 451 Research, Voice of the Enterprise, Storate Q4, 2016*

| Driver | Percentage |
|---|---|
| File Content/User Data | 46.4% |
| Server Virtualization | 41.7% |
| Relational Database Applications | 39.7% |
| Backup and Archive Requirements | 37.4% |
| Data Warehouse/Analytics Applications | 31.1% |
| Email/Messaging Applications | 24.4% |
| Rich Media Content | 21.8% |
| New Data Retention/Governance Requirements | 17.3% |
| 'Big Data' Platforms | 16.5% |
| Non-relational Database Applications | 4.3% |
| Other | 3.7% |

The challenges in adopting a hybrid cloud storage strategy for all types of data stem from the fact that even though most popular cloud storage platforms are based on object storage, there are substantial differences in the way that top cloud storage providers utilize object technology internally, tier their various service levels, and deliver those services to their customers. The next-generation storage administrator will increasingly be challenged to leverage these new capabilities offered by hybrid cloud in order to extend the reach, ensure SLA delivery and improve the cost effectiveness of their storage infrastructure – as opposed to spending most of their time handling the routine storage allocation responsibilities of the past.

## Technology Discussion

Since 2010 a growing number of cloud-based storage services have been springing up that are capable of supporting a variety of enterprise storage use cases – extending from 'hot/warm,' database-optimized, high-performance block storage for cloud-hosted applications to 'cool/cold,' second-tier archival applications aimed at more static data storage applications. While this wide variety of public cloud storage offerings has brought a wealth of options for customers looking for on-demand capacity, negligible hardware investment and the promise of substantially reduced storage costs, many have come to find that it's rarely a matter of simply moving data to the cloud and carrying on business as usual. For many enterprise customers, there will always be a need for both on-premises and off-premises storage resources for a variety of reasons such as data security, industry compliance, or the absolute need for data availability that exceeds the SLA of public cloud offerings. Such flexibility is most easily accomplished with a combination of public and private storage services offered by a hybrid cloud environment; but 'hybrid' and 'simple' often prove mutually exclusive characteristics.

From a storage management standpoint, the separation between public and private capabilities can become far more complex than dealing with the mobile application workloads themselves when choosing between public and private cloud platforms. Even though virtualization and container technology has substantially leveled the playing field in terms of making application workloads more portable, the same can't really be said for their storage environments. Added to the traditional key factors like capacity, performance and latency that constituted the lion's share of on-premises storage management, decisions are now compounded by a new set of concerns over data visibility, protection, consistency and security. And then there's the fact that all of the most popular cloud storage platforms are based on object storage, rather than the block and file architecture that most legacy applications are built on.
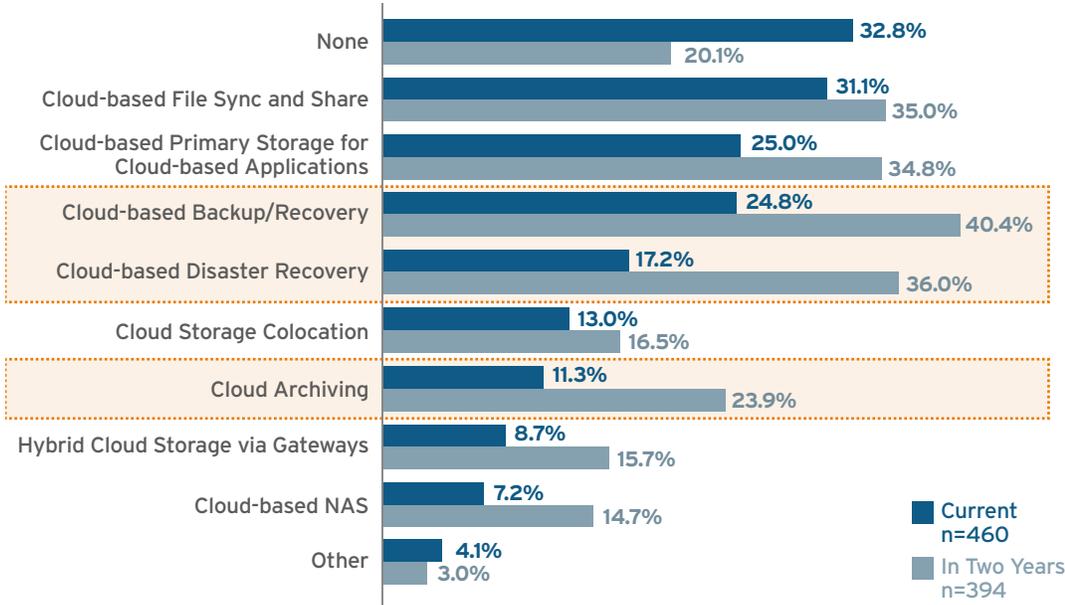
Object storage is well up to the task of addressing many of these challenges due to its rich metadata environment and its role as an abstraction layer that enables a wide variety of front-end delivery options, as well as supporting highly granular, back-end storage management and performance tuning. However, vendors in the public cloud storage business have chosen to adopt different object storage platforms for their own purposes, which raises yet another set of challenges for companies wishing to build a hybrid cloud strategy that includes the option to seamlessly utilize any public cloud vendor on demand, and without the associated time and expense of exporting, moving and importing large quantities of data between different cloud storage platforms.

As cloud usage has evolved, a substantial percentage of enterprise cloud adoption has occurred on an ad hoc basis, mainly because the cloud model presented a far simpler method of engaging easily consumable IT services. Developers only needed a browser and a credit card to spin up as much on-demand compute and storage resources as they required to get the job done, without having to invoke the support of an IT group that may have been saddled with a limited budget and personnel capabilities. This gave rise to the dreaded 'shadow IT' story line that vendors have been using to frighten IT managers for almost a decade – a concern that isn't entirely without merit because illicit cloud use for business applications can circumvent perfectly reasonable cost, security and management controls in the name of expediency. As indicated in our Voice of the Enterprise (VotE) storage study for Q2 2016 (see Figure 2), overall cloud service use has been growing steadily, and will continue to grow for a broad number of use cases into the foreseeable future – especially for the active archive, backup and disaster-recovery applications ideally suited for cloud-based object storage.

### Figure 2: Current Cloud Storage Service Usage & Growth Estimates

*Q. Which of the following public cloud and SaaS-related storage services and capabilities does your organization utilize today? And in two years?*
*Source: 451 Research, Voice of the Enterprise: Storage, Organizational Dynamics 2016*



| | Current | In Two Years |
|---|---|---|
| None | 32.8% | 20.1% |
| Cloud-based File Sync and Share | 31.1% | 35.0% |
| Cloud-based Primary Storage for Cloud-based Applications | 25.0% | 34.8% |
| Cloud-based Backup/Recovery | 24.8% | 40.4% |
| Cloud-based Disaster Recovery | 17.2% | 36.0% |
| Cloud Storage Colocation | 13.0% | 16.5% |
| Cloud Archiving | 11.3% | 23.9% |
| Hybrid Cloud Storage via Gateways | 8.7% | 15.7% |
| Cloud-based NAS | 7.2% | 14.7% |
| Other | 4.1% | 3.0% |

Current n=460
In Two Years n=394

## Common Cloud Data Types

There are a number of hybrid cloud use cases that present variations on legacy storage models, whether on-premises or not, so it is useful to review some of the most common types of cloud data in greater detail.

Production Data for Cloud Workloads – Persistent block storage that provides sustainable application performance and resides on the same cloud as the application instance it supports. This storage is a mounted volume that behaves like a local networked storage LUN, but is usually replicated at the cloud-provider level to assure 99.999% availability or greater; although it's also common to provide asynchronous replication to off-premises resources if desired.

**Large Data Sets for Analytics Purposes** – This refers to an extremely broad category of data, which could be derived from a number of sources both inside and outside the cloud. The highly scalable nature of cloud-based object storage and its attractive price point make data warehousing in the cloud an increasingly viable option.

**Sync & Share Business Data** – This has been a rapidly growing segment of the cloud storage model, in part because consumer-oriented sharing services gained so much attention and greatly simplified personal collaboration. Like many other cloud technologies, sync/share rapidly became a business problem because employees – whether intentionally or not – were able to circumvent data management policies by using a file-sharing service with no oversight from corporate IT.

**Unstructured Data** – As mentioned earlier, this form of data represents the fastest-growing segment of the overall data storage landscape. Media files like images, audio, video or other application-specific data files are easier than ever for end users to create, and are challenging to manage because they can be iterative and extremely difficult to classify. In addition, the growing need to store and deliver on-demand streaming media for consumers via the internet is ideally suited for cloud-based storage.

**Backup and/or Archival Data** – The role of data backup is changing, in part because data that was often locked away in nondescript backup sets is finding new life as a rich resource for data mining, as well as being a candidate for more effective data lifecycle management. The relatively low cost and grow-as-needed scalability of cloud-based storage are making it an appealing target for next- generation backup and archiving. Plus there is a growing interest in moving beyond traditional backup to focus on a model that protects workloads – as well as their data – to better provide disaster-recovery capabilities.

Of course all of the data types cited here are also candidates for backup/archival management, and the metadata capabilities of object storage both inside and outside the cloud offer the potential to significantly shift the traditional backup model into a more active archival platform that better supports long-term data accessibility, governance and automation.
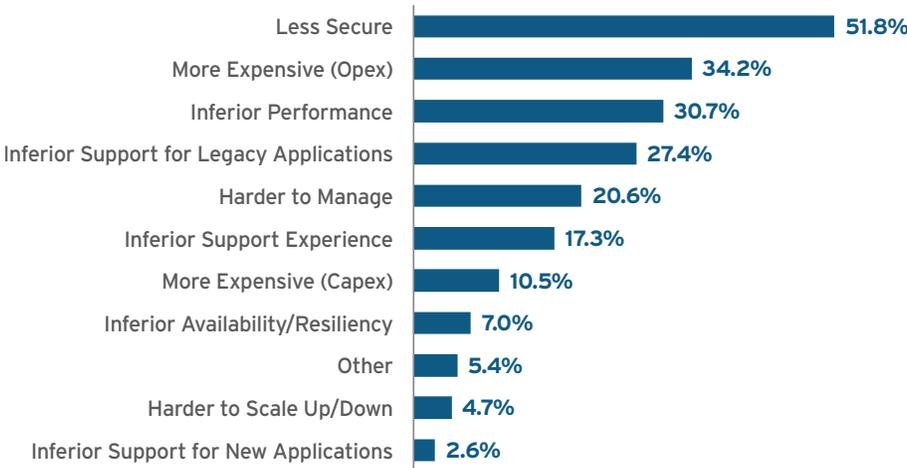
## Challenges to Hybrid Cloud Adoption

There are a number of differences between enabling hybrid cloud-based applications on- and off-premises and managing the associated short/long-term storage challenges, and the issues involved in developing a cohesive hybrid cloud storage management environment can be easily overlooked because of the relative simplicity in which cloud storage can be utilized. Current-generation cloud-based storage offerings have been able to abstract much of the underlying complexity required for allocating storage services in the past, as have many of the modern off-premises storage platforms. However, the adoption of hybrid cloud storage also adds a new set of concerns that never existed for storage environments that remained safely behind the corporate firewall. Figure 3 provides VotE storage data from Q2 2016 that highlights what customers believe are the major disadvantages regarding the use of public cloud storage services as part of a hybrid cloud environment.

### Figure 3: Chief Disadvantages of Cloud Storage

*Q. For storage specifically, what do you believe are the chief disadvantages of using public cloud and SaaS-based services compared to traditional on-premises storage? n=427*

*Source: 451 Research, Voice of the Enterprise: Storage, Organizational Dynamics 2016*

| | |
|---|---|
| Less Secure | 51.8% |
| More Expensive (Opex) | 34.2% |
| Inferior Performance | 30.7% |
| Inferior Support for Legacy Applications | 27.4% |
| Harder to Manage | 20.6% |
| Inferior Support Experience | 17.3% |
| More Expensive (Capex) | 10.5% |
| Inferior Availability/Resiliency | 7.0% |
| Other | 5.4% |
| Harder to Scale Up/Down | 4.7% |
| Inferior Support for New Applications | 2.6% |

For simplicity, we can break these down into a few key concerns that have consistently appeared as the major disadvantages of public and, by extension, hybrid cloud storage:

### SECURITY

Data security, even for traditional storage platforms within the datacenter, has always led the list of major concerns across all forms of enterprise storage. This makes it more than reasonable to expect that extending a storage environment outside the relative safety of the company datacenter via the internet involves additional risks. That being said, there are a number of ways that cloud storage environments aim to protect data both in transit and at rest, but those measures can vary from provider to provider, and may not be sufficient to satisfy the security concerns of specific vertical markets that are heavily constrained by compliance issues. This security problem is compounded by that fact that cloud storage adoption can occur on an ad hoc basis by well-meaning employees looking to solve individual business problems on their own while remaining unaware of the greater ramifications of offsite storage. Data security can often appear at odds with the accelerated development expectations in the age of the cloud, but it doesn't have to be if there's a hybrid cloud framework in place that can automate data security policies and ensure they are followed, while continuing to enable the on-demand capabilities of hybrid cloud storage.

### COST MANAGEMENT

Off-premises storage costs are relatively easy to track, but a hybrid cloud environment introduces a whole new set of pricing variables based on storage capacity, performance, location, movement and availability options – plus base pricing can fluctuate according to market pressure. This uncertainty is compounded by the fact that it remains a major challenge to track storage usage across multiple cloud platforms in the first place, much less to establish reasonable expectations for long-term costs and demand fluctuations.

### PERFORMANCE

Performance is certainly a greater issue for storage attached to cloud-based workloads such as analytics and databases, and there are a variety of differently priced tiers designed to accommodate specific application performance SLAs. But cloud storage performance is hardly reserved for application-specific traffic; cloud-based object stores are a key supplier of streaming media delivery services and web content, making consistent delivery performance an absolute necessity to ensure customer satisfaction. The growing adoption of flash and other high-performance storage technologies both on- and off-premises will continue to offer an endlessly changing variety of performance options over time. This presents an increasingly complex matrix of hybrid storage options that would clearly benefit from a unified storage management platform that doesn't exist at present, instead leaving these to the next-generation storage administrator.

### LEGACY APPLICATION SUPPORT

As mentioned earlier, the vast majority of cloud-based storage is based on object storage technology; however, most legacy applications must be connected to either a block or hierarchical file-based storage environment in order to function. The good news is that object storage is flexible enough to deliver practically any form of programmatic front-end interface such as CIFS/SMB, NFS, GPFS, HDFS and others, as well as common, object-specific APIs like S3 and Swift – even block, if necessary. It's very likely that the existing separation between object storage and traditional block/file architectures will continue to narrow through the increased adoption of native object storage by application developers, as well as through the increased availability of efficient block/file services running on top of any object storage core.

### MANAGEMENT DIFFICULTY

Whenever you spread the distribution of any system over numerous locations the management of that system becomes progressively more difficult. This is certainly true of the hybrid cloud storage, in part because cloud storage vendors have taken different approaches in the way they provide similar, but not identical storage services that are accessible by common APIs, yet vary in the variety of products and management methods available to their customers.

Ultimately, most of the concerns expressed in Figure 3 are directly related to the challenges of managing an increasingly complex hybrid cloud storage environment. Of course choosing to adopt a single public cloud vendor could reduce the complexity of managing a hybrid cloud environment, but this negates one of the key benefits of cloud adoption, which is the flexibility to move between cloud providers based on whatever combination of business needs, application requirements and cost considerations best suit the customer. And even though the rich metadata environment of object storage offers an ideal framework for unification between cloud storage offerings, there has been limited interest among public cloud storage providers in cooperating to deliver a unified cloud storage management platform. This is somewhat understandable, in part because diverse management capabilities and product feature sets can serve as compelling differentiators between service providers, but also because it just doesn't make good business sense to make it easy for customers to jump back and forth between competing cloud storage providers.

## Customer Considerations for Hybrid Storage Adoption

We believe that a hybrid cloud storage environment should offer the flexibility to utilize resources both on-premises and off, based on a framework capable of policy-based management and intelligent oversight to control costs, provide security, ensure SLAs and protect data – regardless of physical location. This framework should include the ability to automate data lifecycle management, minimize excessive duplication and provide long-term visibility into data.

### KEY CUSTOMER CONSIDERATIONS FOR HYBRID CLOUD STORAGE ADOPTION

- Storage analytics capabilities to monitor data placement and utilization – As hybrid storage environments become more and more distributed, it will become increasingly difficult to maintain awareness of the nature and location of storage resources. Effectively managing storage resources has been a perpetual challenge even within an internal storage infrastructure, and this challenge will only be compounded when faced with managing multiple cloud-based storage environments. This creates a very real need for an environment that offers unified visibility across all forms of storage in use.

- A common storage management interface – Cloud-based storage has embraced several API standards for data access, such as Amazon S3 and OpenStack Swift, but the same can't be said for management interfaces between cloud-based storage platforms. Customers should be aware of the challenges involved in managing a multi-cloud storage environment and look for vendors who can mitigate the management differences between cloud storage offerings, both now and in the future.

- Metadata-enabled data governance – Legacy block and file storage environments provide little information that can be utilized to control the increasing challenges of data lifecycle management and data sovereignty. Given the world-

wide nature of cloud storage, it will become increasingly important to be able to identify data and specify long-term treatment guidelines for corporate information to protect privacy and meet the changing requirements of local and international law.

- Policy-based automated tiering for both on-premises and cloud storage platforms – Part of the economic benefit of cloud-based storage lies in the ability to move data between performance-based storage tiers in order to reduce costs. The use of metadata supports the ability to make these tiering decisions based on more detailed, customer-specific information, rather than simply according to last-use statistics.

- Granular, policy-based data protection and security – Metadata also provides a framework for specifying far more specialized storage automation. Data can be tagged for customizable protection/security treatment based on a nearly limitless number of flexible parameters, such as legal status, intellectual property, high security and other business-specific considerations, as well as industry-standard compliance factors such as those for medical records and financial information, or other government-mandated specifications.

- File/cloud gateway capabilities – Because most cloud storage is based on object storage technology, any hybrid cloud environment that serves legacy applications will need to be able to deliver file services via NFS or CIFS/SMB. This can be accomplished in a number of ways, but most typically it is done using a separate hardware/software platform to bridge file/object protocols, or it is done natively as a service that's included as a feature within the cloud storage environment. Regardless, there is an ingest/export process that's part of any transition from traditional file-based storage to cloud-based object storage, and the efficiency with which that's accomplished should be part of the consideration process of any hybrid storage initiative.

- Verified, automated and low-impact data movement – Data movement is an obvious part of the cloud gateway function, but data movement is also an issue when it becomes necessary to change the physical location of data stored in the cloud. Data growth is a given, and if a company chooses to place most of its data within a single cloud vendor's environment, then moving that data to another vendor, or even repatriating it back to on-premises storage resources, can be a major challenge. Customers should take into consideration the difficulty of getting their data back from the cloud for whatever reason, and make sure there is a contingency plan in place should their cloud storage partners become unable to provide access to company data.

- Data availability and protection – There are a growing number of methods to help ensure that hybrid cloud customers maintain access to and control of their data, but most SLA guarantees offered by cloud storage providers only allow for discounts on billing based on the percentage of customer downtime. Ultimately this leaves the challenge of ensuring data availability and protection in the hands of the hybrid cloud customer and their vendor partners. Risk is simply a part of any technology initiative, so in the case of hybrid cloud storage it's important for customers to understand the differences that off-premises cloud storage entails and make accommodations for them as part of any hybrid cloud storage initiative.

## Recommendations

Hybrid cloud represents what can be the most effective method for incorporating the highly flexible capabilities and cost reduction afforded by cloud technology, while at the same time maintaining the security and internal control capabilities offered by on-premises infrastructure. But in order to make the hybrid cloud a truly dynamic production environment, it needs to be based on a framework that covers key security and policy management requirements, regardless of the physical location of mobile workloads and data.

Some of the challenges cited in this paper will be resolved with future advancements in hybrid cloud storage management technology, but as usual a number of the difficulties involved in adapting to next-generation IT practices will also require a more enlightened thought process for senior IT management, storage administrators and vendors alike.
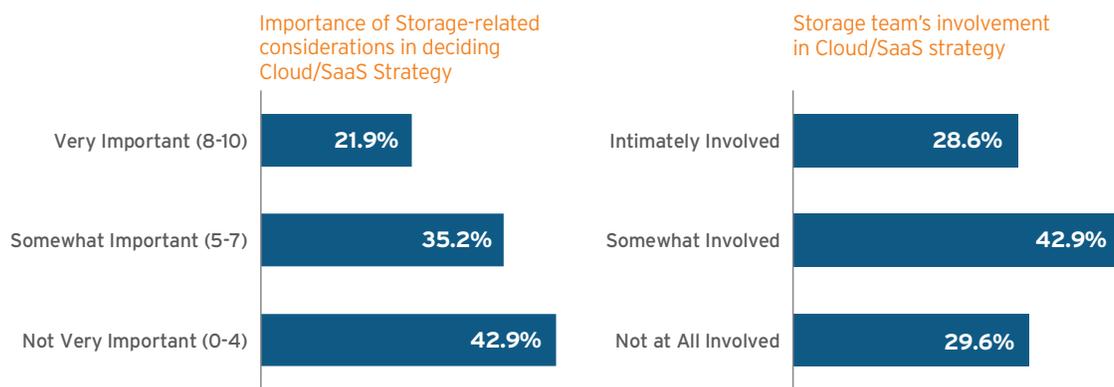
### FOR SENIOR IT MANAGEMENT

It's important to remember that cloud adoption doesn't have to be a binary decision. Cloud storage resources should be considered simply as another tool in the toolbox, and one that should be utilized based on the appropriate combination of business application suitability, security and management issues – as well as cost savings. There are a number of factors that affect the actual cost of public cloud storage services, and many companies have found that the volume-based savings offered by the efficiencies of cloud storage can easily be lost because of data movement, application performance and availability requirements particular to their production environment.

In our Q2 2016 VotE storage survey (see Figure 4), we were surprised to find that a number of companies may not include their storage administrators as part of the hybrid cloud decision-making process. Although it's true that cloud storage doesn't require the same type of oversight as on-premises legacy storage, storage administration teams have typically become extremely well-versed in the challenges of storage management, performance tuning and data protection – challenges that will only increase with a further distributed storage environment. It would likely be in the best interest of IT management to heed their advice.

## Figure 4: Storage Team Involvement in Cloud Initiatives

*Q. (left) How important were storage-related considerations in your organization's overall decision around use of public cloud and SaaS models. Please use a 0-10 scale where 0 is 'Not at All Important' and 10 is 'Extremely Important.' (Binned) n=525*

*Q. (right) How would you characterize the storage team's involvement in your organization's decision around use of public cloud and SaaS models? n=469*

*Source: 451 Research, Voice of the Enterprise: Storage, Organizational Dynamics 2016*



Importance of Storage-related considerations in deciding Cloud/SaaS Strategy

| | |
|---|---|
| Very Important (8-10) | 21.9% |
| Somewhat Important (5-7) | 35.2% |
| Not Very Important (0-4) | 42.9% |

Storage team's involvement in Cloud/SaaS strategy

| | |
|---|---|
| Intimately Involved | 28.6% |
| Somewhat Involved | 42.9% |
| Not at All Involved | 29.6% |

### FOR STORAGE ADMINISTRATORS

Storage technology is now undergoing a shift similar to the one that server environments experienced with the rapid adoption of virtualization over a decade ago. Many of the mundane day-to-day challenges of physical storage administration are being eliminated, but the growing interest in cloud storage capabilities will undoubtedly impact the type of work and expectations of storage technologists. Economic forces often drive technology adoption, so it will be important for storage administrators to become proactive in understanding the value of hybrid cloud storage, as well as the potential drawbacks. The capabilities of metadata-rich next-generation unified object storage present remarkable opportunities for managing the growth of unstructured data, as well as for building a dynamic storage environment that offers unprecedented policy-based capabilities to automate data management and facilitate long-term visibility for data throughout its entire lifecycle.

Storage technologists should approach hybrid cloud storage as an opportunity to explore capabilities and services that most IT environments could never afford to implement on-premises. But more importantly, storage professionals should look to provide valid and well-reasoned guidance in developing a holistic hybrid storage environment that balances data protection, availability, performance, security and cost while leveraging the next generation of storage capabilities that a hybrid cloud infrastructure can provide. Storage professionals should find it disconcerting that there is a trend (indicated by Figure 4) in which senior IT management doesn't necessarily perceive the importance of storage considerations in deliberations over cloud adoption. Whether this is based on a lack of understanding about storage issues or a perceived resistance among storage administrators to change, this information should serve as motivation for storage professionals to proactively evaluate hybrid cloud storage opportunities and prepare for a solution that provides senior IT management with a realistic and open-minded approach to sustainable hybrid cloud adoption that storage teams can also live with. It's already happening with or without input from storage administrators.

### FOR STORAGE VENDORS AND CLOUD STORAGE PROVIDERS

The responsibility falls squarely on the storage vendor's shoulders to make hybrid cloud adoption as painless as possible, by developing tools and capabilities that continue to support existing practices and storage platforms while offering the flexibility of adopting new technologies in the future. Vendors have been singing the praises of the hybrid cloud model for several years now, but for whatever reason this hasn't resulted in a hybrid cloud environment that offers a simple, common

policy framework – much less one that supports simplified data visibility, monitoring and movement between public and private cloud resources. We say this with the full understanding that this is a complex problem; but we also believe it is one that can only be solved by improved, industrywide cooperation between vendors, providers and customers.

The storage industry has struggled for years to address the differences between traditional storage platforms, and there have been efforts along the way to provide a common management environment for on-premises storage systems that were met with marginal success. Today, hybrid cloud presents an even larger set of challenges – now that baseline storage administration is compounded by the fact that data can reside in a far greater number of locations – which raises perfectly reasonable concerns over data protection, availability, security and management across on-premises and cloud-based storage resources. Part of the value proposition of the hybrid cloud model has always been the ability to seamlessly migrate between on-premises and cloud, or between clouds, but this has yet to happen. The truth of the matter is that the metadata-rich object storage platform has become a prerequisite for most, if not all, distributed cloud-based storage environments, and that same metadata framework can also be used to provide far more consistent and flexible storage management for hybrid cloud storage environments. We're not saying this will be easy, but we do insist that it's all but a necessity in order for hybrid cloud to deliver the freedom, automation and flexibility that's been promised.