SOLUTION SHOWCASE

Why You Should Look at the IBM Cloud for Archival Solutions

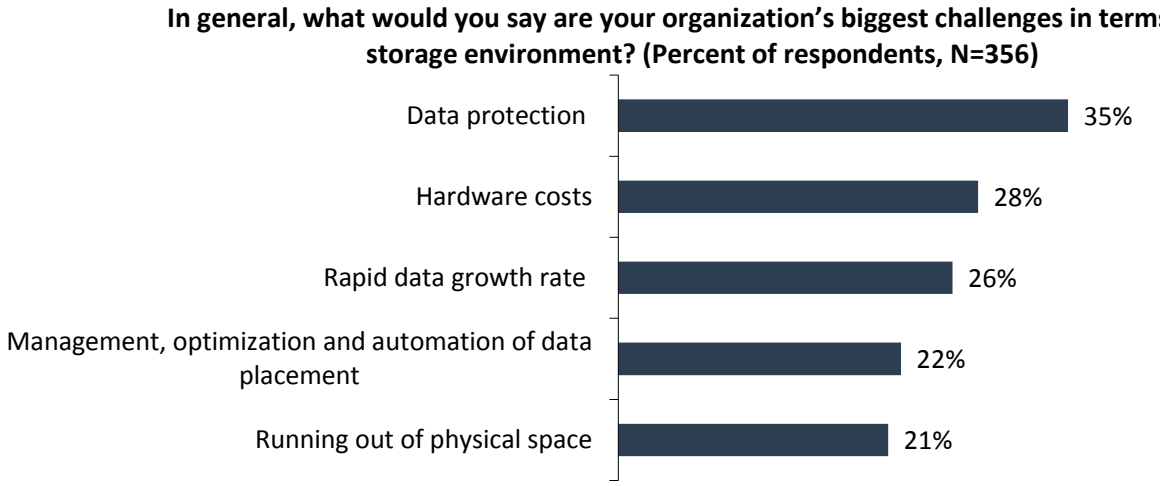
Date: November 2017 **Author:** Jason Buffington, Principal Analyst

Abstract: Organizations of all sizes continue to experience operational, economic, management, and technical struggles in how they are delivering IT services, often motivating them to embark on digital transformation journeys. For many organizations taking such journeys, the cloud is a big part of the answer. Cloud-based storage can provide a compelling alternative to on-premises archival-class storage and enable levels of agility not previously achievable.

Overview

Storage continues to be a common burden for IT organizations due to economic challenges (maintaining the pace of capacity growth with flat or declining budgets) and a lack of agility related to archaic approaches to data protection, data preservation, and data management (see Figure 1).[1]

FIGURE 1. Top Five Storage Challenges in 2017



Source: Enterprise Strategy Group

For many, the new data being created isn't the main challenge. It's the stagnant and non-necessary data that continues to consume a disproportionate amount of primary production storage. While regulatory retention pressures might be theoretically addressed by simply not deleting anything, reality requires that organizations rethink their archive storage strategy and implement additional types of content repositories that offer distinctive performance-versus-cost advantages.

[1] Source: ESG Brief, *2017 Storage Trends: Challenges and Spending*, August 2017.

The Need for Retention Repositories

Organizations of all sizes should be rethinking their broader storage strategy with the concept of a flexible and durable "*retention repository*" being a key aspect of the future across a variety of solution scenarios:

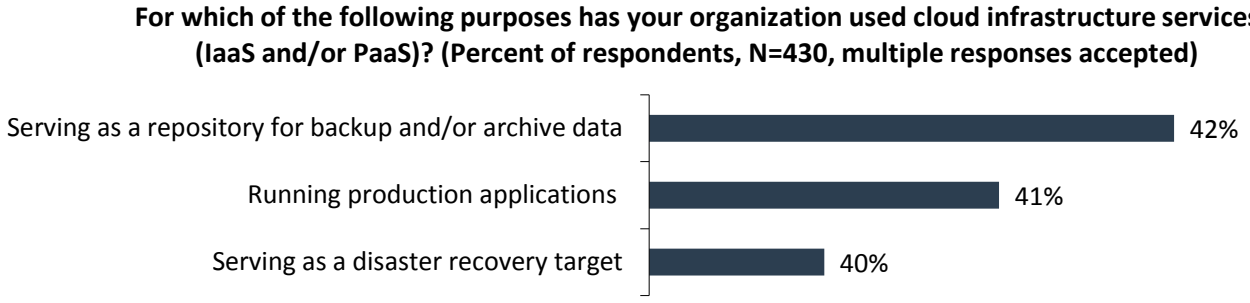
- Backups – as the foundation of most data management and data protection strategies, organizations should be looking for durable and efficient repositories for "protection storage."
- Archives – while backups create previous versions in preparation for recovery, archives often provide "the copy of last resort" for regulatory compliance, eDiscovery preparedness, and operational reference. That said, not all data needs to be preserved and as such, archival solutions usually retain only a subset of the overall production data, based on the business or governance implications of the data, with a variety of implementation strategies:
    - "Warm" or "Active" archives provide near-production levels of access performance to predominantly dormant data that needs to be immediately available upon request.
    - "Cold" or "Deep" archives typically have the longest retention requirements and the least expectations on responsiveness during retrieval, though many organizations still expect minutes (to hours) for even its coldest data.
- Storage Tiering – while archival strategies are typically driven by regulatory considerations or operational governance criteria of the data itself, many organizations simply want to manage all of their data within a unified storage stack that automatically and transparently moves data within different grades of storage (tiers), based on access patterns, so that the most recent, frequent, or critical data is on the fastest ("hottest") storage tier while lesser data seamlessly moves and is stored on more scalable and less-expensive "colder" tiers.

It is important to note that the above strategies and solution scenarios are seldom, if ever, achievable using only disk-based architectures. Instead, for organizations to be successful, a broader strategy that combines disks (HDD and SSD) with complementary media (tape and cloud) should be embraced.

Cloud-based Storage Can Be a Significant Part of the Answer

Digital transformation efforts must include IT modernization, which will almost inevitably utilize cloud services. In ESG's annual IT spending intentions survey research, cost reduction is a perennial business driver, with respondents citing the use of cloud-based infrastructure among their top planned ways to reduce costs.[2] Having a repository for backup and archive data has consistently been the most commonly cited use case for cloud-based infrastructure among survey respondents (see Figure 2).[3]

FIGURE 2. Top Three Cloud-infrastructure Use Cases in 2017



Source: Enterprise Strategy Group

[2] Source: ESG Research Report, *2017 IT Spending Intentions Survey*, March 2017.
[3] Source: ESG Research Report, *2017 Public Cloud Computing Trends*, April 2017.

Why the Cloud Should Be a Part of Your Solution

While economics may be why many organizations will start exploring cloud-based services, the real power in utilizing cloud-based storage as part of an archival or tiering strategy is tied to durability/availability, agility, and security.

Durability and Availability

Arguably, the most important aspect of an archival repository is its ensured survivability and accessibility regardless of circumstance. Unlike on-premises disk solutions or self-managed tape collections (which will be impacted or even destroyed by whatever calamity affected the production environment), cloud services are, by their very nature, regionally distributed and natively resilient.

As the copy of "last resort," cloud-storage used for archival or tiering purposes should be fault tolerant of server- or component-level failures within the cloud provider's data center. It should be replicated and resilient. Its availability should be assured transparently across multiple cloud-based points of presence. And most importantly, "last resort" data should be immune to gradual, silent deterioration and corruption—i.e., it should be invulnerable to bit rot.

On-premises storage will still require routine backups and other typical data protection behaviors, but cloud-based content repositories—ad-hoc storage utilized for non-production or interim usage—should remain unscathed and assuredly accessible for many years.

Agility

Agility in an archival or tiering repository can often mean flexibility of implementation and utilization, as well as responsiveness to user needs. Cloud-based storage can be adopted by organizations regardless of their size, without the delays of provisioning and enablement that typically plagued the enterprise data centers of yesteryear. New or additional repositories can be enacted in a matter of minutes with simplified access via a variety of protocols (e.g., S3) for use by backup/archive software, as well as via integration as a storage tier with seamless data movement, depending on the cloud service.

In terms of performance, legacy approaches to content repository preservation often relied solely on tape media. Modern tapes and libraries still have their place in scenarios that involve legal or regulatory considerations over extended periods, but many organizations might be surprised by the speed of ingest and egress that a cloud-storage tier within an archival or tiering solution can offer—without needing media management or other human intervention.

Security

The most common naysayer sentiment against embracing cloud services is the security concern when putting data in a facility outside the organization's direct control. In fact, 46% of IT professionals surveyed by ESG cited security among their concerns when choosing not to utilize cloud-based data protection services.

Cloud Archiving versus Cloud Tiering

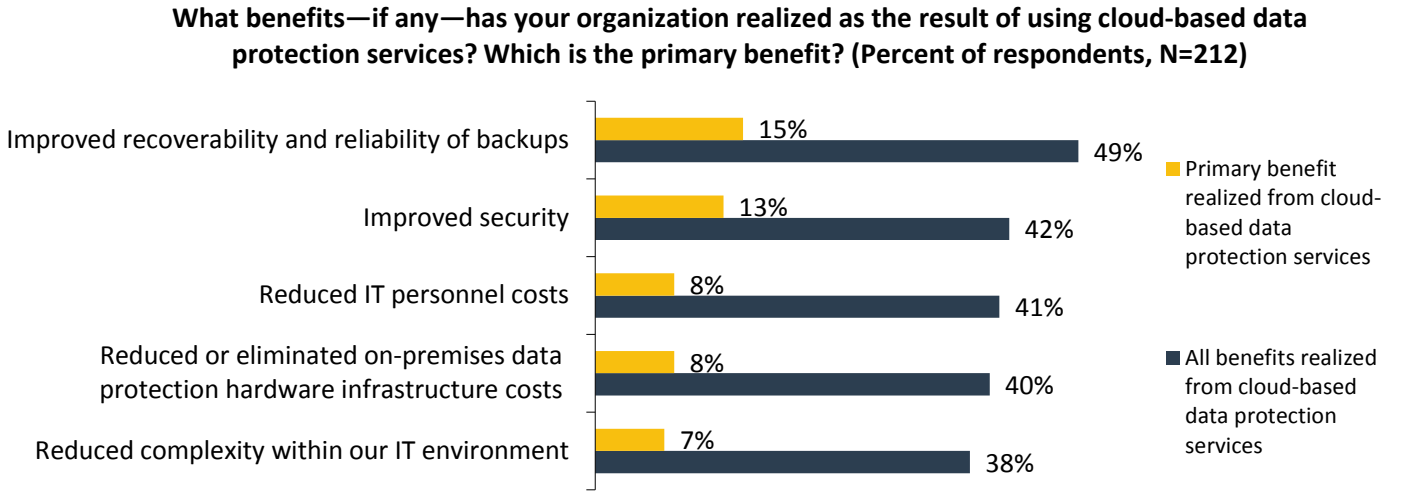
Sometimes, an organization will use an archive software product (or backup software with archive features) to do an after-the-fact assessment of stagnant data and then invoke policy to move that data off of primary storage.

But today, modern production storage platforms with tiering capabilities can deliver similar storage-optimization benefits simply based on access patterns. Cloud storage, when properly architected, can be a great candidate as a tier to be added to such a stack.

The point is, it's now possible for an organization either to pull its stagnant data "off the top" using its archive solution, or to let its stagnant data "filter out the bottom" to a cold tier; either of which is achievable using cloud-storage.

However, a significant number of IT professionals who have embraced cloud services actually cited *improved security* among the most common recognized ***benefits*** of utilizing cloud services (see Figure 3).[4]

FIGURE 3. Top Five Recognized Benefits of Using Cloud Services as Part of a Data Protection Strategy



Source: Enterprise Strategy Group

Many of the IT professionals who acknowledged improved security as a recognized benefit of cloud storage described improvements coming from data encryption in-flight across the network and at rest in the storage. Neither capability is offered by many on-premises backup and content repository solutions. In addition, physical access, damage, and theft-related concerns of traditional on-premises storage solutions (including backup/archive disks and tapes, as well as other primary/secondary storage systems) are mitigated by the use of remote cloud-based services.

How the IBM Cloud Addresses Data Protection, Data Preservation, and Data Management

For more than 100 years, IBM has been a technology innovator and accelerator of business modernization. It should come as no surprise to find IBM delivering cloud-based storage, namely IBM Cloud Object Storage, fueled by its acquisition of Cleversafe in 2015. Certainly, a "one IBM" solution pairing IBM Spectrum Protect (for backup) and Spectrum Archive software with IBM Cloud Object Storage (referred to by ESG as the IBM Cloud), along with IBM Resiliency Services and other expertise from IBM Business Partners, is exciting. But the IBM Cloud stands on its own as a compelling repository for a broad variety of ecosystem partners due to its durability-, agility-, and security-centric attributes.

The Durability and Availability of the IBM Cloud

As described, a vital characteristic of any copy of "last resort" is the assurance of survivability and accessibility—the data's integrity will remain pristine, even for decades, without suffering bit rot. In general, it will be there—still readable, still usable—even if all other copies have been destroyed.

Most cloud services have some level of resiliency built into their data centers' architectures (e.g., RAID), whereby the failure of an underlying commodity storage device does not affect access. They often also have replication between storage nodes across multiple cloud data centers. This approach delivers durability and availability, but it does so by significantly increasing cost and complexity due to a duplicated infrastructure. In contrast, the IBM Cloud uses erasure coding to distribute fragments across multiple servers and sites, thereby delivering durability and availability including geographic resiliency and component-level fault tolerance without duplication-related cost and complexity.

[4] Source: ESG Research Report, *Data Protection Cloud Strategies*, December 2016.

The Agility of the IBM Cloud

The IBM Cloud is recognized by many backup/archive software vendors as a "must integrate with" platform to ensure the success and delight of their joint customers. IBM's Cloud Object Storage service can be accessed using generic, industry-standard protocols (e.g., S3). But the thoughtful, deliberate integration with leading software solutions to ensure a better-together outcome should really interest large organizations and IBM Business Partners looking for an "enterprise-grade" cloud storage solution. Many will enjoy a combined experience imbued with ease of acquisition, ease of deployment, and ease of ongoing operation. They will have a flexible archive storage repository or storage tier that doesn't necessarily require backup/archive machinations or processing.

The Security of the IBM Cloud

As one would expect, IBM's commitment to IT innovation and its influence on and experience with cybersecurity products and services bolster the IBM Cloud's security controls and protections. The burden to ensure encryption and provide secure handling and storage of data was foundational to the Cleversafe architecture that now supports the IBM Cloud. There is built-in erasure coding, slicing, and all-or-nothing encryption that secures data at rest. In addition, you can also secure your data with "bring your own key" (BYOK) key management: BYOK allows organizations to retain complete control of their existing keys while benefitting from IBM's built-in advanced encryption for data at rest.

The Bigger Truth

The "simply add more" operating model of storage has never been more unsustainable than it is today. Primary production storage is burdened by growth barriers, economic pressures, regulatory challenges, and stagnancy that is forcing organizations to reimagine their storage strategy:

- Some organizations are adding storage tiers to their production stacks that transparently move data from high performance flash … to commodity hard disks with greater capacity … to scalable non-disk repositories.
- Some organizations are revisiting their archival/backup strategies in order to better protect non-stagnant data for shorter periods, while preserving the right subset of data for long-term retention based on regulatory mandates.
- Some organizations are doing both.

In all of these cases, the copies of data within these repositories are often the last remaining copies of the data within an organization, therefore requiring them to be active and agile enough to ensure on-demand usability, durable and available enough to be relied on without routine backups (like primary data), and secure and governed for the sake of organizational integrity and viability. Those characteristics describe the storage needed—and the traits of a few (not all) cloud solutions, with IBM's Cloud Object Storage being one example worth considering.