ESG Technical Review

Conveniently and Securely Archive to Microsoft Azure with HubStor

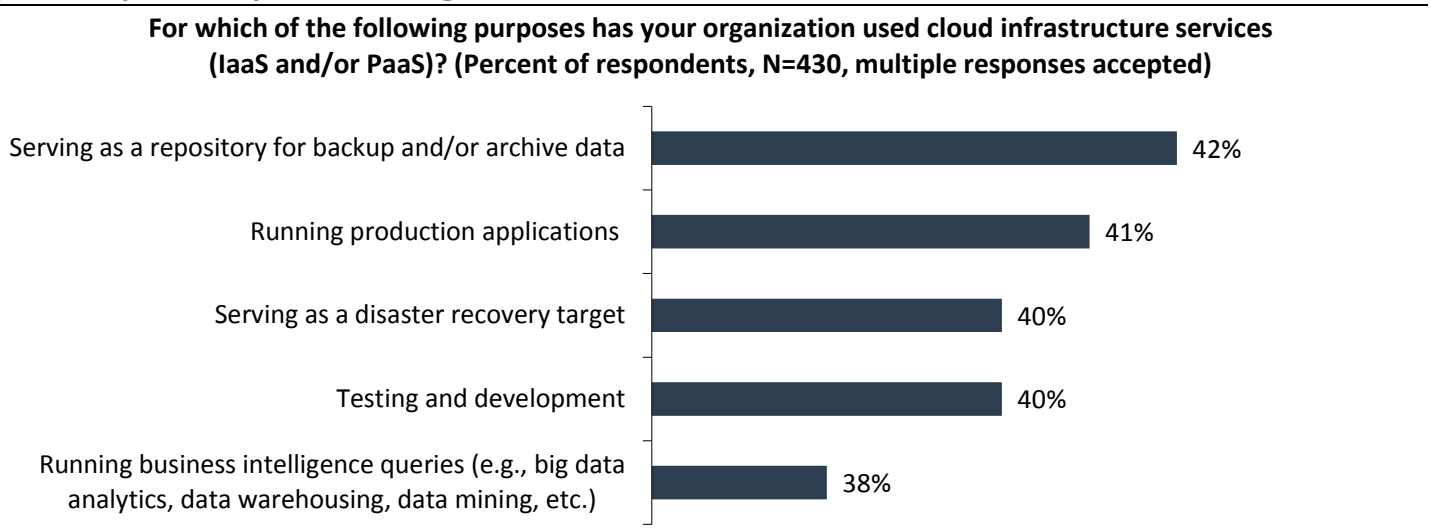
Date: November 2017 **Author:** Mike Leone, Senior Validation Analyst | Dom Amato, Associate Validation Analyst

Background

Cloud usage is becoming a mandate within organizations looking for ways to manage data growth and improve operational efficiency, while reducing costs. The challenge for organizations is understanding how to consume cloud storage because, by itself, the cloud is often viewed as just infrastructure or a platform-as-a-service. A software layer or software-as-a-service to make cloud storage consumable and relevant to an organization's needs and ongoing strategy is missing. The market wants an easy way to not only consume cloud storage, but also to put data in or pull it out without being locked in.

One such area where cloud storage makes a lot of sense is in the backup, disaster recovery, and archival space. The cloud serves as an ideal place for long-term storage of structured data that is low-touch or inactive. In fact, research shows that the most-often cited cloud infrastructure services use case for midmarket and enterprise-class organizations is using cloud as a repository for backup and/or archive data (see Figure 1).[1]

Figure 1. Top Five Purposes for Using Cloud Infrastructure Services



Source: Enterprise Strategy Group, 2017

Simply having data reside in the cloud requires additional features to help with security and protection. Sending data to the cloud has proved to be cost-effective, and the growing popularity of cloud-based backups runs parallel. Because the number of cloud-based SaaS applications and workloads such as email and CRM continues to swell, backing them up is often overlooked. The same ESG survey shows that 90% of organizations either use, have plans to use, or are interested in SaaS.[2] SaaS lends itself particularly well to private and public (or even private-to-public) cloud archiving since the data already resides in an Internet-based model. Rather than forcing data migration back to physical storage, legitimate cloud backup becomes a tempting offer with the increased focus on cloud-based software, infrastructure, and platform-as-a-service.

However, archiving to the cloud does not typically solve every problem. Logistical issues such as compliance and administration can lead to unforeseen complications. Further, without a smart software layer that provides a simple and efficient method of cloud backup and archiving, an organization can incur higher costs, whether because data is stored on the wrong type of cloud storage or inefficient recovery methods are used when something goes wrong in the production environment.

[1] Source: ESG Research Report, *2017 Public Cloud Computing Trends*, April 2017.
[2] ibid.

HubStor

HubStor offers autonomous data archiving software geared for Microsoft Azure. The technology enables customers to take full advantage of public cloud storage with the aim to streamline data analytics, data archiving, and indexed search at scale. Users can customize policies for offloading primary storage to the cloud, which can be tagged as private or sensitive for auditing or sharing later with partners and peers. This ability to expedite data archiving and retrieval provides a modern and efficient process that organizations require to benefit from the possibilities of leveraging the public cloud for unstructured data storage. Gone too are licensing fees and term commitments, as the HubStor subscriptions are based on actual monthly consumption rates.

HubStor's array of features not only simplifies primary storage archive, but also offers a refreshing approach to adjusting and consuming data. The search-as-a-service approach allows users to create search clusters that can be scaled up or down as needed to meet demand and control cost. Search policies offer control over scope, user access, and customizable metadata for future keyword searches and real-time analytics within HubStor. Options for version tracking detect any changes to data, and cost savings from typical reduction in storage volume within HubStor is automatically reflected in the monthly subscription price. The Connector Service can be installed in multiple instances to fit the required connection for data, while HubStor's admin and end-user web portals are similar to Office 365 for streamlined sharing. Furthermore, export and recovery functions that can select metadata, date, time, and individual items are not limited by HubStor, so customers maintain full control of data.

With low-touch data consuming IT resources and creating risk when unsecured, HubStor enables organizations to move this inactive data to the public cloud, alleviating IT budget and security concerns. HubStor archives primary file storage to a more cost-efficient cloud environment with secure policies, while built-in governance features simplify data compliance. Features such as real-time policies, analytics, and search-as-a-service allow organizations to easily connect and manage all archived data. Further, these abilities are wrapped in a convenient web-based UI where the HubStor software provides a unique, autonomous, and cost-effective method to archive, search, and share primary file storage.

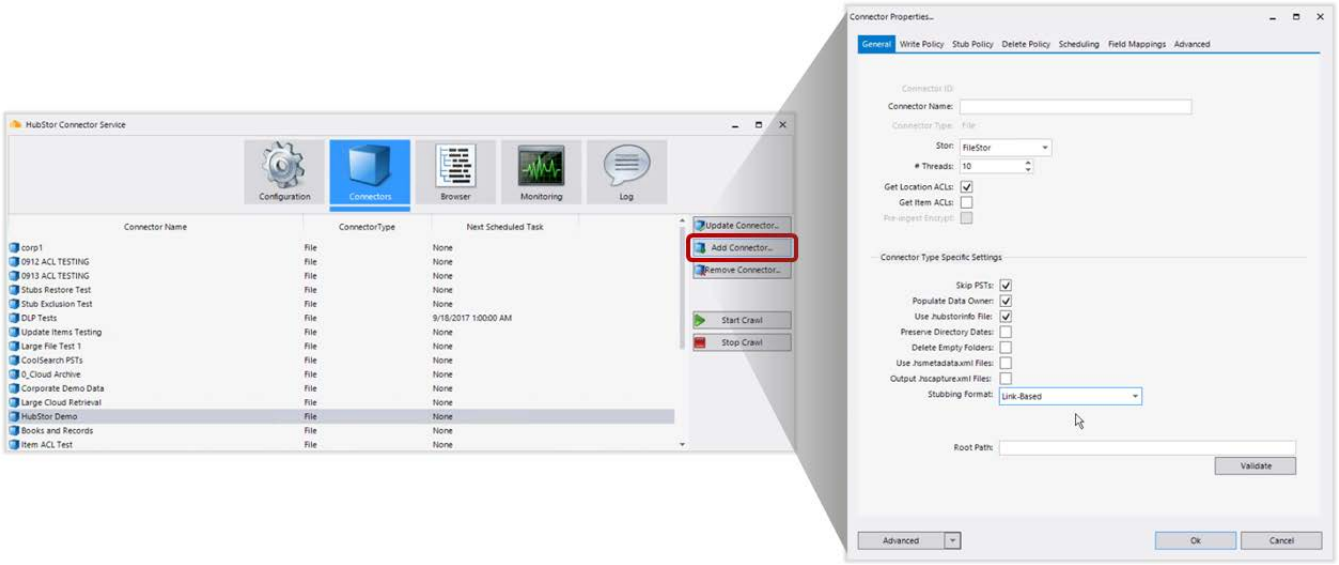
HubStor looks to deliver on increasing demand for an Azure-based archive sitting alongside Office 365 and take full advantage of its features. User authentication and identities are provided by Azure AD, and Azure Encryption secures data with the user's choice of at-rest in the cloud or pre-ingestion with on-premises key access scenarios. If that weren't enough, in the unlikely event of a regional failure, Azure helps HubStor configure warm secondary locations to ensure availability in only a few hours. On top of HubStor's own attributes, having Azure as a foundation can provide better content and durability to customers, including searchable cold archive, storage tiering, Office 365 backup, WORM compliance, and email archiving.

Simplicity, Convenience, and Smarts

Simplicity and convenience are core to HubStor's platform. HubStor takes advantage of the familiarity and convenience of Microsoft Azure to enable organizations to easily archive data to the Azure cloud where the configurable storage tiering policies make rehydrating data just as easy. And with a HubStor single-tenant storage account, organizations can interact with Azure the same way that Office 365 does, meaning Office 365 customers will have an easy time signing into HubStor. ESG walked through the simplicity and convenience HubStor provides as a robust software layer that enables archiving to the Azure cloud.

Data enters HubStor through its Connector Service, with data being sent directly through a LAN connection or via drive shipping through the Microsoft Azure Import/Export Service. ESG learned how organizations get started with selecting what data to archive and how to archive it. As displayed in Figure 2, from the HubStor Connector Service UI, ESG selected ***Add Connector***, which gave several options as to the type of data being archived and how that data should be archived and classified, the majority of which could be changed with a single click of a checkbox or drop-down menu option.

Figure 2. HubStor Connector Service



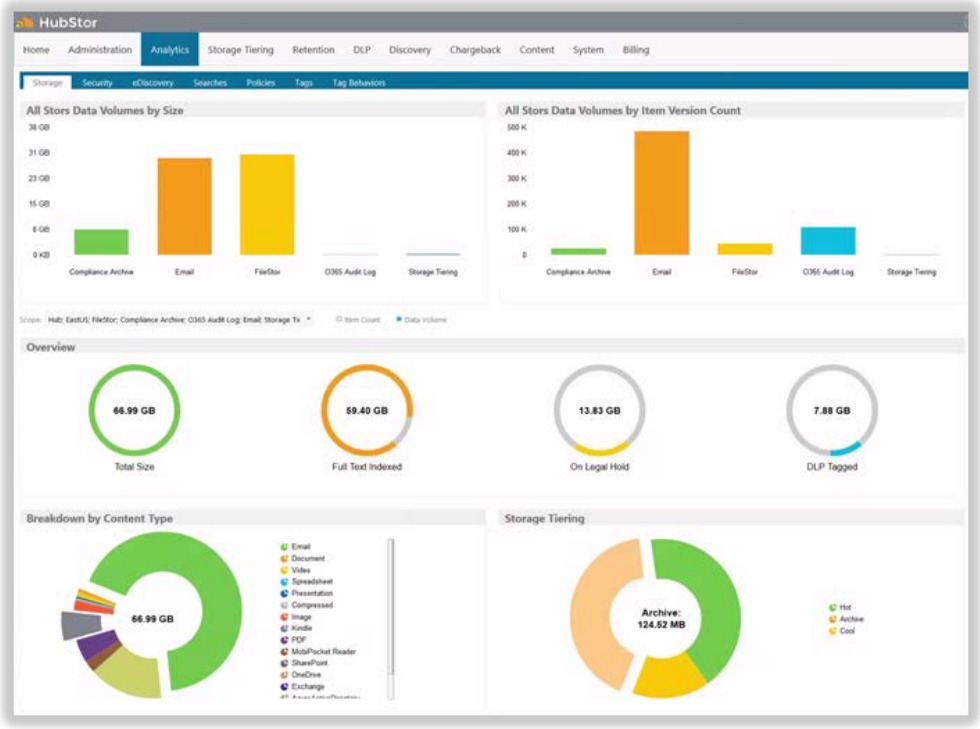
Source: Enterprise Strategy Group, 2017

Outside of deciding where to archive data and how to archive data such as in stubbing format, ESG could quickly change how HubStor could view and retrieve data. One particularly impressive function was how HubStor categorized and interacted with metadata, including how metadata could be classified to affect the searches, policies, and analytics. ESG witnessed several scenarios in which specific sets of popular metadata fields were prefabricated and automatically associated based on the type of connector, such as email, compliance, and Office 365 audit logs. It should be noted that HubStor offers the ability to completely customize how the connected data is recognized in the HubStor platform. The whole process highlighted the simplicity with which data could be taken from applications and document protected servers and placed into HubStor while maintaining all associated metadata. Further, the system was intelligent enough to know when and where to use that metadata across the entire HubStor platform.

Next, ESG turned to HubStor's policy creation, which focused on details related to data retention and storage tiering. While cloud archive is a cost-effective method, the cost of accessing cold data can be higher than expected if not properly configured. Creating policies with HubStor gives users the power to archive data to different tiers based on numerous factors, including the last data access time or whether the data is full text or a seamless stub. ESG created several policies in a matter of seconds with just a few clicks. One of the policies that was created monitored data within a folder for storage tiering purposes. After six months, if the data was untouched, the data was moved down to the cool tier. After twelve months, if the data was still not accessed, it was permanently moved to the archive tier. And it is important to remember that as data moves down tiers, the cost of storing that data reduces.

HubStor's policy creation tools can save organizations time and money associated with having to manually track usage and move data when appropriate, but the real power of HubStor is in its analytics. ESG viewed the analytics dashboards (see Figure 3), starting with simple breakdowns of data volumes, capacity consumed, specific data types, and tiering of the stored data. Each graph in the dashboard was interactive and could be drilled into, proving to be particularly effective in instances where custom metadata or policies were created. Deeper insights were also viewable through various analytics tabs for the purpose of digging into user activity around specific folders and files, any legal holds in place, and access controls for individual users or entire teams. The next level of granularity from HubStor's analytics is the content application, which provided insight into the structure of archived data and enabled ESG to open different blobs of data to see version history, permissions, and audit details. ESG was able to filter on a specific AD group and gain insight into what data that team had access to, as well as a complete structural overview of every folder and file.

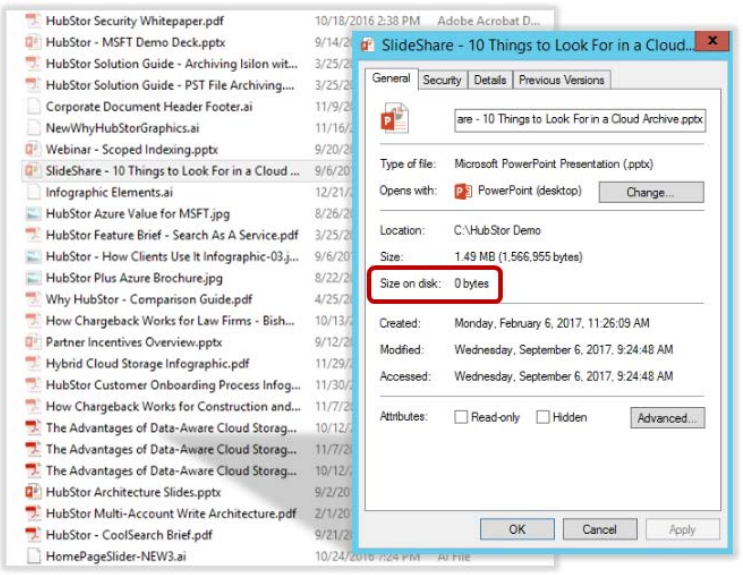
Figure 3. HubStor Analytics



Source: Enterprise Strategy Group, 2017

After understanding the process of getting data into Azure's cloud storage and efficiently managing and monitoring it through interactive dashboards, ESG turned to retrieval to understand how archived data can be accessed. With traditional archival storage tiering, IT managers are often forced to use complex and high-risk methods of data retrieval by building storage appliance agents or network layer interceptions to get at data faster. HubStor's retrieval service is provided through a separate application that directly integrates with Azure Active Directory to communicate with HubStor and enables organizations to set up local caches that optimize the retrieval of data. ESG was shown how users have the advantage of interacting with archived data as original files, seamless stubs, or link-based stubs. Items that are stubbed display in Windows as a grayed-out file type icon, and in CIFS shares they have a shortcut overlay. When selecting properties of a specific file, all the same metadata is displayed with one difference in that the size on disk is near-zero. When the file is opened, the retrieval service downloads the file into its cache, where it holds the file so any subsequent requests will open the file from the local cache, as opposed to downloading from the cloud again. IT has control of the size and location of the cache.

Figure 4. Seamless Stubs



Source: Enterprise Strategy Group, 2017

The last area ESG focused on was the ability to recover. A ransomware attack was simulated in which data was maliciously encrypted. HubStor, using Azure AD as the authenticator, enables IT administrators to easily egress data from Azure and restore all of it to its original location. Using seamless stubs, ESG forced an overwrite of the encrypted files to a point in time when the data was last clean, and the best part was that a portion of the data could be rehydrated as a seamless stub, making the entire process fast and lightweight because less active and older data could be recovered in a virtual manner.

Why This Matters

As adoption rates and sizes of cloud-based applications continue to grow, the prospect of storing that data in the cloud has become more attractive than ever. However, many traditional methods of cloud archive are difficult to digest and lack a software layer to assist IT managers with strategy implementation. The few cloud archive models in existence can become quickly expensive as a result of further complicating administration issues.

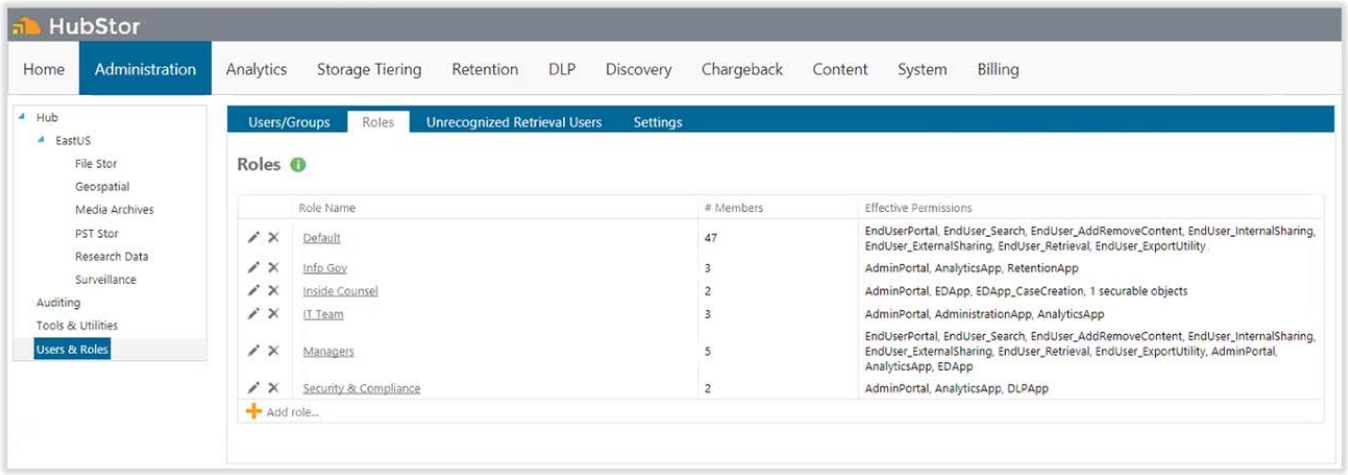
ESG validated HubStor's ability to provide organizations with an easy and efficient way to confidently offload primary file system data to secured Azure cloud storage through easy-to-configure policies. From connecting a new data source and smartly archiving to the most cost-effective cloud storage tier to managing the data based on customizable metadata mappings and interactive dashboards, ESG was impressed with the overall simplicity and convenience of applying HubStor's smart software layer to Microsoft Azure blob archive storage.

Access Control and Security

HubStor's cloud-based data archive solution offers organizations key advantages over traditional archive methods, particularly around the painlessness of moving data to the cloud. While the cloud can bring to light competitive pricing, some organizations remain hesitant to embrace it due to security concerns. HubStor is aware of these concerns and has purposefully designed its product with security top of mind using patent-pending technology designed to keep data managers in complete control without needing to sacrifice features or functionality.

HubStor leverages a single tenancy deployment model. Single tenancy eliminates concerns related to shared cloud storage spaces by providing every organization with a dedicated cloud environment with singular access and resources. Furthermore, all authentication must go through Azure AD before communicating with the cloud. This approach is similar to how Office 365 works, making HubStor a natural extension of its single-sign-on, multi-factor authentication, and user identity management settings. In fact, any form of penetration testing would first require penetrating Azure AD before getting the chance to penetrate HubStor. Once inside HubStor, IT administrators are able to apply additional user controls on top of Azure AD to fine tune control and access. Figure 5 highlights roles that can be assigned to users once inside HubStor.

Figure 5. Fine-grained User and Access Control

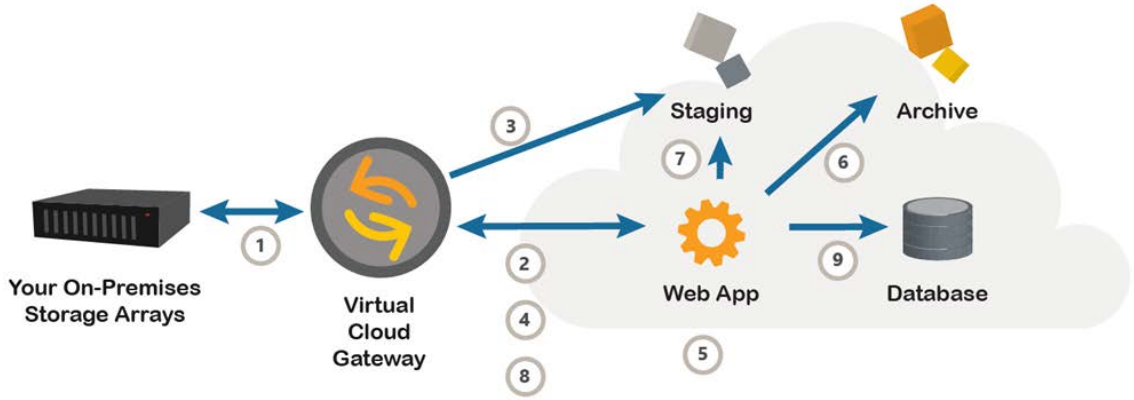


Source: Enterprise Strategy Group, 2017

Once the data is archived in Azure, several layers of data redundancy are applied to protect against potential data loss. Redundancy levels for local and geographically redundant storage create synchronous copies of data within one or multiple data centers. Azure AD already meets industry-standard compliance standards around ISO, HIPAA, and SOC, while HubStor provides additional tools that leverage Azure AD such as eDiscovery. HubStor provides specific capabilities for policy and search creation for all archived data. Users can be polled to monitor data traffic, individual access, legal holds, and any requested egress from data archived in HubStor. These types of features give HubStor the power to create a transparent environment for IT administrators who must control and monitor even the coldest of data to ensure compliance is maintained.

The final aspect of security where most cloud storage solutions (not just archive solutions) fall short is in the manner in which they request access to cloud storage—through an API key and password—before transferring data to the storage container. Though the keys are encrypted, they still force the customer to provide those keys to access the storage. Rather than forcing organizations to surrender this information to access the archive storage, HubStor's patent-pending Generic Connector Framework offers an alternative. The most important component of the process is that authentication through Azure AD and HubStor is done on the client's environment, meaning API keys and passwords never leave the building. The nine-step process is highlighted in Figure 6. First, the HubStor virtual cloud gateway uses defined archive policies and data connectors to sync data to be archived. Second, when HubStor detects a write to the cloud is required, a time-limited write-only token is obtained providing access to a temporary staging container in the web application. Third, data is written to the staging container. Fourth, once written to the staging area, the web app is notified, which triggers step five, where a deduplication algorithm is run. In step six, the unique data is written from the staging container to the archive container. Staging is then cleaned up, metadata information is sent back to the HubStor gateway, and the HubStor database refreshes, updating all the analytics and running any policies based on what data has changed.

Figure 6. Securely Archiving with HubStor



Source: HubStor

Why This Matters

Archiving data in the cloud has proven to be a cost-effective and logical next step when dealing with the growth of cloud-based applications, but it does not solve every problem. Many IT professionals and organizations wish to use the cloud for backup and disaster recovery, which means security has become the highest priority. Unfortunately, most current offerings have shortcomings in this area, as many are built on multi-tenant environments and require customers to give up the credentials to their data.

ESG Lab validated that HubStor solves many of these problems without asking customers to change anything. HubStor offers single-tenant environments to isolate customer data and has no need to externalize API keys or data passwords as it is integrated tightly with Azure AD. During a simulated ransomware scenario, HubStor was able to easily recover data as full items or stubs on a point-in-time basis that required only a few minutes of user time.

The Bigger Truth

Data growth is a problem for many organizations—it forces them to endure ongoing capital and operational expenditures to keep up with the scale of data, having to store both primary data and backups, and dealing with archival solutions that are both complex and costly. The cloud is a natural fit to help address these concerns, especially in the archival scenario, but cloud storage alone does not meet all that businesses require. For organizations to sign off on a cloud solution, security remains top of mind, and software is needed to help fill the gap and put IT administrators' minds at ease when it comes to compliance and user access.

HubStor is a data-aware cloud archival solution that eliminates traditional archiving complexity related to data growth and cost. Organizations can easily and confidently offload primary file system data to secured Azure cloud storage through easy-to-configure policies. HubStor not only gives users better access to their data, but also keeps that data searchable at scale, offering true storage convenience. Organizations can tag folders or individual files that contain private or sensitive data, helping them rapidly respond to audits or other legal claims, while data can also easily be shared externally with other entities. Further, with cost being top of mind, HubStor provides a simple, pay-as-you-go price structure based solely on storage consumption—no licensing fees, term commitments, or additional fees.

ESG was impressed with HubStor as it served as an ideal onramp to consuming cloud storage the way it was meant to be consumed—conveniently, securely, and cost-effectively. As HubStor continues to add powerful features, such as anomaly detection based on user access patterns, its already-robust feature set will continue to provide organizations with peace of mind while exceeding customer expectations. If you are looking to modernize your archiving strategy to benefit from the vast advantages of the cloud, ESG suggests considering HubStor.