White Paper

The Rise of Object Storage

By Scott Sinclair, Storage Analyst

September 2015

This ESG White Paper was commissioned by Cleversafe, an IBM Company, and is adapted from the previously published ESG Market Landscape Report, *Object Storage*, and distributed under license from ESG.

Contents

What Is Object Storage?

If there is one universal truth when it comes to information technology over the past decade, it is that data growth is inevitable and unstoppable. Whether driven by personal or professional motives, individuals are simply creating more digital assets than ever before. Regardless of the industry, the success of corporate operations depends on the ability to utilize digital assets. Whether it's media and entertainment leveraging higher resolution video or developing more realistic digital effects, energy exploration firms capturing detailed 3D or 4D seismic data, security systems capturing higher resolution security footage, or online content distribution, the creation and the efficient utilization of digital assets is critical to staying competitive across nearly every line of business.

Along with this surge of digital content, and the fact that our corporate culture is focused on litigation, comes an increase in compliance regulations and internal controls and audit processes that require organizations to be more responsible than ever when storing and curating content. In some cases, organizations have responded to the possibility of future litigation by instituting a culture where data is kept for decades and potentially longer. With the emergence of machine data and with the Internet of Things (IoT) on the horizon, content creation promises to expand well beyond the previous limitations of content created by just the human populace. The net result is a future where nearly everyone and everything will be creating some form of data that will have to be stored and protected for some period of time (if not indefinitely). Object storage, more than any other storage architecture, is designed to store and protect the resulting massive content repositories.

Just as solid-state technology has emerged in response to the increased demand for performance, spinning media has begun to specialize in high-capacity and cost-optimized storage delivering larger capacities at a lower cost point. Despite these cost reductions in underlying storage hardware infrastructure, storage system capacities have reached a tipping point—a threshold where traditional storage and protection technologies, such as file systems, are no longer viable options. Object storage architectures provide the necessary capabilities critical to storing and protecting high-capacity content environments, and with the projected growth of digital content over the next few years, object storage looks to become a much larger portion of enterprise storage deployments.

Limitations of RAID and the Need for Object Storage

Object technology delivers a direct response to the challenge of storing and protecting large amounts of unstructured data. Traditional storage arrays, which often leverage RAID technology, hit serious limitations at scale.

In 2015, ESG conducted a research study investigating general storage industry trends. As part of this study, ESG surveyed 373 IT decision makers responsible for their organizations' data storage environments. One of the questions asked respondents to identify their organization's biggest storage challenges. Not surprisingly, the rapid growth of data was identified as a top challenge. What is interesting—though also not too surprising—is that the rest of the challenges within the top ten most-cited responses all can be considered symptoms of data growth. Challenges such as increased hardware costs, data protection costs, and staffing costs are all created by or exacerbated by the data growth (see Figure 1).[1]

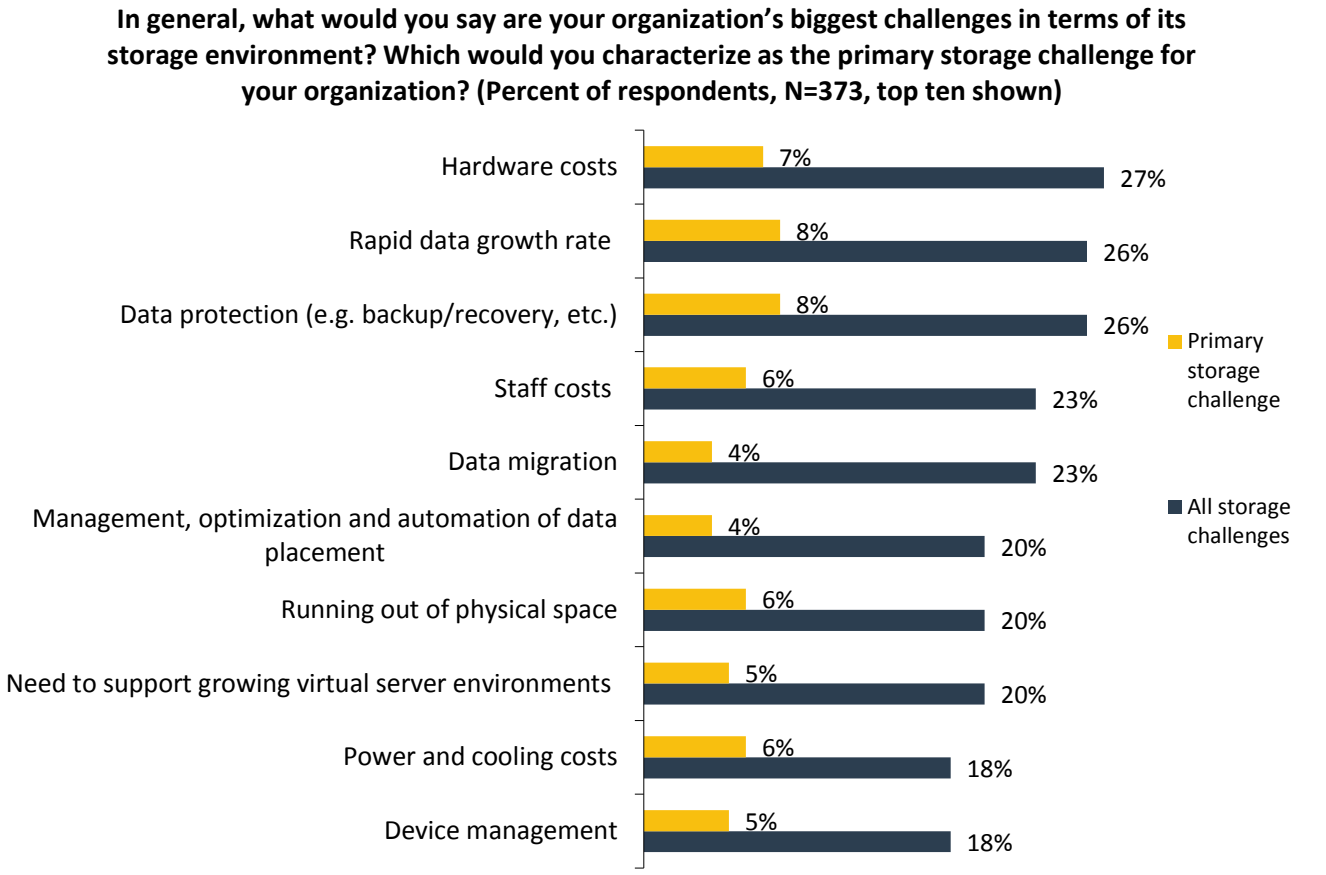
There are also nuances associated with protecting data at scale. While traditional storage systems with RAID were designed primarily to protect against single drive failures, larger hard drive capacities have translated into an increased likelihood of multi-failure events, which leaves the organization exposed to the danger of losing data should a second drive failure occur during the rather lengthy rebuild time associated with high-density drives.

Multi-failure scenarios arise in a couple of different ways: Additional drives could fail during a rebuild or bad sectors on a previously designated healthy drive could be uncovered during a rebuild and result in unrecoverable data. The likelihood of unrecoverable bits used to be low. However, with larger capacity drives, the chances of not being able to recover some data have increased significantly in recent years. As a result, the industry has shifted from recommending RAID 5 to recommending RAID 6, which protects data in the event of a dual drive failure. While the addition of a second parity bit has alleviated some concerns about finding an unrecoverable bit during the rebuild

[1] Source: ESG Research Study, *2015 General Storage Trends Survey*, conducted in May 2015.

process, the introduction of larger drives is also resulting in longer rebuild times, increasing the amount of time data is exposed to the possibility of experiencing a secondary or tertiary failure. And while the odds of a multi-failure event are still low, they can increase to undesirable levels for organizations that have hundreds or thousands of drives.

Figure 1. Organizations' Top Ten Biggest Storage Environment Challenges



Source: Enterprise Strategy Group, 2015.

These risks are further exacerbated by the fact that massive capacity environments are often too large to back up using traditional backup methodologies, meaning that in many cases, these repositories may store the only copies of the data available.

Object Storage Provides a New Level of Scale and Protection

This section intends to provide some technical insight into object storage and to provide some discussion on the background and history of object storage technology. Traditional file systems are designed to store content in a hierarchical manner, often in a tree of files and folders. In these systems, users access a file by following a path to a specific location. While this method can be intuitive for storing a few files, when content storage explodes to billions or even trillions of files, a hierarchical access method can create too much complexity and in some cases, overwhelm traditional file system storage architecture. Some file system-based solutions can run into limitations once a certain number of files and folders is reached. Additionally, some file system architectures are unable to effectively scale metadata, greatly impacting the performance of metadata operations once the system exceeds a certain amount of capacity or number of files.

Object systems are designed using an alternative approach with a single and massively scalable flat address space where file access is provided via a unique identifier. An analogy that can help describe this difference is that accessing a specific file in a file system is like following a set of directions to find a location—for example, "take the

first left, then the second right, etc." Object storage, on the other hand, is like using GPS coordinates. This more efficient manner of identifying specific content helps enable object storage to scale to higher capacities than can be achieved by traditional file systems.

Object storage solutions were designed to solve the challenges of protecting massive capacity environments where traditional backup is often not an option. Object storage architectures provide robust protection by expecting multi-failure scenarios to be common occurrences. Object storage provides several innovations to help accomplish a more robust and bulletproof ecosystem:

- **Replica or erasure coding versus RAID:** Instead of traditional RAID, object storage commonly leverages the use of object replication, erasure coding, or a combination of the two to provide data protection. These technologies provide several advantages over traditional RAID, including the ability to leverage commodity hardware, the ability to apply specific protection schemes to specific objects or groups of objects rather than on a volume level, and the ability to evolve the protection scheme over time, e.g., to reduce protection from four replicas down to three over time. Additionally, by using replicas or erasure coding, object storage can automatically expand protection across sites by spreading data across multiple drives, nodes, and even geographies.
- **Rebuild into free space:** The majority of traditional RAID storage environments are typified by the "hot spare," which is a free drive waiting to take over after a production drive fails. However, in the event of a failure, only one drive is the write target for rebuilding the data, creating a significant bottleneck that is further exacerbated as drive capacity increases. The vast majority of object storage solutions provide the ability to rebuild data from a failed drive into free capacity across many drives and nodes in the system, significantly speeding up the recovery process thanks to the massively parallel nature of leveraging many drives (versus one), and therefore greatly minimizing the time the system is in a degraded or vulnerable state.
- **Self-healing:** Another innovation designed to greatly speed up recovery time and nearly eliminate the risk of encountering unrecoverable bits during the rebuild is the ability to self-heal. Object storage solutions often provide background processes that read presumably healthy data, and verify that the contents are intact. When an unrecoverable bit is identified, the object is rebuilt in available free space, ensuring that data is always healthy and readable. Additionally, since it is common for drives to fail a few sectors at a time rather than all at once, with self-healing, it is likely that most of the data will already be rebuilt to healthy sectors on other drives when the drive in question finally is determined as failed.
- **Automatic geo-protection**: Traditional storage arrays often rely on file system or volume/LUN-based replication for multi-site protection. These traditional storage containers are also often bound by some architectural capacity limit, 16 TB for example. Providing multi-site protection for large capacity content storage environments with these limitations can result in the unwieldy management of scores of replication policies, if not more. Multiple object storage solutions support a massive, near-infinitely scalable flat address space, with the capability for automatic multi-site protection built in, greatly simplifying multi-site protection and greatly reducing the associated management costs.

Enhanced scalability and resiliency are not the only advantages of object storage infrastructures, but they are key architectural capabilities that separate object technology from more traditional file and block storage methodologies. Despite a number of advantages, however, object storage is not a panacea for all storage concerns.

Additional Considerations about Object Storage

Object storage systems are designed around storing and retrieving entire objects, as opposed to block-based or file system-based technologies that are designed to read and write individual blocks of data. The most significant impact of that difference is experienced in write performance, especially when a user or application wishes to modify a file. Instead of simply applying the modified blocks of data, object storage systems require the entire object be read before it can be modified, and, once modified, the object is rewritten in its entirety to the storage platform. This architectural difference led object storage systems to originally house workloads that were often

write-once-read-many (WORM) environments, such as medical imaging, energy exploration, or other workloads with image or video media content. These environments are typified by content that is created and rarely modified, but that needs to be kept online and at some point retrieved. For example, Cleversafe delivers immutable vaults of data so that data cannot be deleted and logs all changes to data so that it can be tracked at a detailed level.

Heavy transactional workloads are typically better suited for file or block storage environments. However, advances in processing and memory technology along with the integration of solid-state have improved the performance capability of object storage solutions considerably. Cleversafe, for example, has delivered a containerization model for improving utilization efficiency and performance of smaller objects. As a result, Cleversafe claims its customers report a 300% increase in operations per second with this model over the previous file-based mechanism for storing. In several cases, object storage solutions can be designed to serve transaction use cases, although perhaps not as efficiently as other storage technologies.

Another consideration and potential limitation for object storage is limited storage protocol support. Traditionally, access to object storage is available via programmatic access over the HTTP protocol. While HTTP access can offer some advantages, a lack of more popular storage protocols has limited application support. This challenge is discussed in more detail later in this report, but it is important to note that recently a majority of object storage solutions have expanded protocol support to include more common storage protocols such as NFS and CIFS. Additionally, S3 has emerged as a more widely supported object protocol.

Despite inefficiencies in high transactional workloads or limited protocol support, as organizations experience increases in data growth, object storage systems are becoming a more significant portion of their IT storage strategies. As such, a growing number of object storage offerings that look to solve the challenges of high-capacity content storage are available in the market.

Object Storage Use Cases

When evaluating object storage solutions, there are a number of considerations to include in the investigation. This section will review specific considerations and trade-offs for core object storage capabilities, as well as feature and functionality implications for specific use cases. The goal of this section is to provide a framework to use when evaluating object storage solutions in both general and use-case-specific evaluations. Included in this section are examples of functionality that can provide an advantage for specific environments. However, organizations must always ensure that the latest information is used when evaluating any object storage solution.

Object Storage: Core Functionality Considerations

Object storage is defined by the ability to scale and protect at scale. However, all solutions do not protect or scale data in the same way. These several categories of capabilities should be considered during a technology evaluation:

- **Replication versus erasure coding:** The vast majority of object storage solutions leverage a replication methodology, an erasure coding methodology, or some combination of both to protect data. The replication protection scheme replicates an object when it is written to the object store and stores multiple copies (often three or more) of each object across multiple nodes in the environment. It is therefore common for object storage systems to require at least three nodes in a pool to provide the necessary level of resiliency. By proving three separate physical nodes, the solution can lose two without risking data loss. The trade-off of this scheme is that the raw capacity for the system is required to be three times the capacity of the content being stored. While object storage systems often couple this protection scheme with the ability to leverage lower cost hardware, using only a replica-based protection scheme can lead to a substantially large hardware infrastructure deployment. While the addition of deduplication or compression can help alleviate the raw capacity concerns, solutions that only offer replica-based protection require a significant increase in the raw capacity when compared with similar erasure-coded solutions.

  Erasure coding, on the other hand, is similar to RAID where failure protection can be provided without a direct multiplication of the raw capacity by using parity information. The difference between erasure coding and RAID is that erasure coding is more flexible and can be distributed across multiple drives, nodes, or

even sites for multi-site failure survivability. Due to the natural alignment of massive content repositories and multi-site failure survivability, erasure coding saw its earliest implementations in the object storage. Cleversafe claims to be the first company to bring erasure coding to the market in an enterprise storage system. While erasure coding provides multi-failure protection with far less raw capacity than leveraging a replica-based protection scheme, erasure coding requires more processing power to calculate the coding schemes, which can impact performance, or may require more processing hardware to achieve the same level of performance. Additionally, reading an object requires accessing content from multiple nodes and possibly multiple sites, potentially impacting the read performance. However, erasure-coded schemes can provide the flexibility to protect against a large number of simultaneous failures, if desired, which may not be feasible when using a replica-based protection scheme. Of course, the more failures the system is configured to withstand, the larger the hardware investment, so it is important to understand the impact of higher resiliency on the cost of the solution and balance accordingly.

In response to the trade-offs, some solutions offer a combination of erasure coding and replica capabilities, allowing administrators to select between improved performance and improved capacity optimization. For example, Cleversafe offers the ability to select between replication and erasure-code-based protection.

- **Geo-dispersion and multi-site protection:** Replication, asynchronous or synchronous, has been a common feature in traditional storage systems for a while. While some applications may be able to get by relying on backup alone for protection, for massive capacity environments, backup may not be an option as the time to back up may exceed even the widest of acceptable windows. As discussed earlier, the chief concern as capacities increase is managing the multitudes of replication policies that may be required across an environment. Multiple object storage solutions provide automatic multi-site protection via either a distributed erasure coding scheme or a replication capability. It should go without saying that if your organization supports multiple sites, or requires multi-site protection, you should ensure that any object storage solution considered supports automatic multi-site protection.

  However, object storage solutions can offer multi-site protection in different ways. Whether the base protection scheme leverages a replica-based or an erasure-code-based model, a multi-site configuration may simply extend the base protection scheme to include nodes on other sites in addition to nodes on the primary site, or may apply a separate layer of replication on top of the base protection scheme. For example, if the system uses erasure coding to provide data protection, the architecture may span the erasure-coded objects across multiple nodes that reside across physically separate sites or the solution may simply replicate an erasure-coded portion of data to a secondary or tertiary site. Additionally, if the solution leverages replicas as the primary protection methodology, the system may simply always ensure additional replicas are placed on nodes in physically separate sites.

  For example, Cleversafe uses a combination of its SmartRead technology, where reads are always performed from the fastest responding node, and WAN optimization techniques, such as leveraging parallel concurrent connections, to improve performance. Additionally, automatic site protection can be very useful in content distribution or collaboration use cases, as content is automatically distributed to remote sites. While multiple providers offer solutions for geo-dispersion or multi-site protection, when selecting a vendor, it is important to have the vendor provide validation as to whether that vendor has existing and proven distributed or globally distributed deployments. There is a difference between simply claiming to support global distribution and actually delivering a globally distributed storage environment in practice.

- **Multi-generation architecture:** A side effect of massive scale environments is that they are very difficult to migrate. Large-scale content stores and the underlying object storage infrastructure can't exist in the continuous three-year upgrade cycles of traditional storage infrastructure, and, as such, object storage systems are designed with the ability to support multiple generations. The ability to incorporate multiple hardware generations into a single pool enables object storage to take advantage of new hardware technologies, while eliminating the need for big data migrations. As nodes need to be retired, migration is done only for the individual node on a rolling upgrade basis. Because of the massively parallel nature of the system, the retirement process is relatively quick and easy compared with traditional SAN and NAS

migrations. When a node is retired, the data from that node is simply copied to free space throughout the cluster, leveraging the ability for multiple drives to write data in parallel.

There are a couple of different ways that object storage solutions deliver multi-generational support. One way is to offer multiple appliance options, continually developing and offering new appliance solutions as hardware technology evolves. Each generation of the appliances is able to mix and match with previous generations in a single pool. This option can provide a simple deployment model, but is ultimately limiting in terms of the hardware options that are available. In some cases, appliance models may reach the end of their supported life, requiring at least a partial migration. While this method provides multi-generational support, it is less flexible than the second method, which is to simply provide the object storage technology as software and allow that software to support virtually any hardware option.

- **Software-defined versus appliance:** Multiple solutions across the storage industry leverage the software-defined storage nomenclature, each with different definitions. But one segment of software-defined storage technology provides deployment flexibility by providing storage functionality—in this case, object storage functionality—as a software-only package. Solutions in this segment can be deployed on commodity server hardware, providing flexibility in hardware selection that offers a number of benefits for IT organizations. By providing the ability to select hardware, IT organizations can directly lower either their capital or operational expenses. In some cases, an organization may choose to select lower-cost commodity hardware, while in other cases, organizations may choose to leverage server hardware from a familiar manufacturer in order to reduce the management complexity and associated costs.

  Also, while traditional storage systems recently have been leveraging more commodity components, server systems often offer a faster hardware refresh cycle than storage systems. In some cases, new technology is available as quickly as every 12 or 18 months as opposed to every three to four years for storage systems. The net result is the ability to leverage faster components, such as memory and processing, more quickly, resulting in increased capability for an equal or lower price.

  Finally, buying cycles can be made easier to manage by separating the software licensing from the hardware upgrade cycles. Traditional storage or appliance offerings can, however, provide benefits as well. With integrated hardware and software offerings, the entire solution is validated, ensuring the hardware will perform optimally with the software. Also, sometimes, traditional storage or appliance offerings can be less complex to deploy. As such, software-defined object storage solutions typically also offer an appliance deployment option. Cleversafe, for example, provides both software and appliance options.

- **Proprietary versus open source technology:** With the increased interest in cloud deployments and multi-tenant IT resource orchestration solutions, a variety of on-premises cloud solutions have arisen, some of which, such as OpenStack, are offered as open source solutions. Cloud solutions, like any IT environments, require a storage component. Desiring a storage solution that aligns with the core tenets of cloud, such as hardware independence, infinite scalability, and affordability, many cloud solutions have turned to object storage, offering an open source version of the storage technology. These solutions, such as OpenStack Swift or Ceph, are available in their pure open source form or as a bundled product with support from storage partners. The most notable benefit of an open source solution is the software capital expense, or in other words, the price. The open source technology is freely distributed. If an organization desires to purchase an enterprise license from a provider, there is a cost, but the license pricing is still generally low.

  When evaluating the benefits of open source technology versus proprietary offerings, it is important to note that there is more to storage costs than simply the software licensing expense. The costs of supporting and managing storage are also significant expenses. Open source technologies may lack some of the more advanced features and manageability capabilities of their propriety counterparts.

Object Storage: Considerations by Use Case

Object storage's ability to store and protect massive capacities of content applies to a variety of use cases. This section will attempt to highlight some of the more popular uses cases and some of the important capabilities that

apply to those use cases. This is by no means an exhaustive list, but an attempt to cover the more popular use cases and highlight the key capabilities to look for when evaluating solutions.

- **Active archive/content repository:** An archive is not a backup, though it is often confused for one. A backup is meant to serve as a secondary copy of production data, and to provide a level of protection in case a disaster occurs and the data needs to be recovered. An archive conversely is meant to serve as the primary copy of the data held for long-term preservation. In many cases, archives are put in place to some extent to reduce the pressure on backup infrastructure. Despite the fact that an archive is different than a backup in many organizations, tape had been the media of choice for deep archives, a way to retain data for long periods of time and remove it from the primary infrastructure. In the past decade, however, government compliance requirements and internal legal demands placed pressure on IT organizations to be able to retrieve this archived data in a more timely fashion, driving the need to retain archived content on demand and keep content online. The demand for content to remain retrievable led to the rise of active archives in which the data was kept active often via disk technology as opposed to tape.

  The shift to disk-based infrastructure, however, led to a new list of requirements for active archive storage infrastructures. Environments are often typified by few write and heavy read operations, often aligned with specific industry vertical solutions and coupled with a vertical industry software application, such as medical archiving, media and entertainment, or energy exploration. This use case is also starting to see another surge in growth due to the increase in machine-generated data for the Internet of Things. The large capacities typified by active archives also make them difficult, if not impossible, to back up. The result is an environment that must serve as the first and last line of defense for the storage and protection of archive content.

  As such, active archives must be resilient enough to survive multiple failures, and operate in an environment where multi-failure events are more common and expected. The core tenets of object storage technology, such as infinite scalability and the ability to provide high levels of resiliency at scale, along with affordability, make object storage an ideal platform for active archive deployments. As such, an active archive could be considered the base use case for many of the object storage use cases listed later in this document, where many of the additional use cases are simply variations of an active archive.

  Additionally, archives are meant to store data for long periods of time—five years, seven years, or, in some cases, even decades or longer. The period of time that organizations expect to store content will see numerous hardware revisions and upgrade cycles. The benefits of multi-generational support and software-defined storage discussed earlier resonate soundly for this use case. Another key characteristic to look for is integration with primary solutions to migrate content off of primary storage and onto an archive object storage platform, without impacting the application. Cleversafe actively partners with CommVault and Qstar and provides integrated solutions for active archive. Other capabilities to consider are policy-driven movement or tiering capabilities within the object storage system or to the cloud as well as integrated search and retrieval.

- **Compliance archive:** Similar to the traditional active archive use cases, compliance archives require the scale and protection of object storage along with additional functionality designed to meet government or industry requirements for content and records retention. Organizations are subject to multiple regulations such as Commodity Futures Trading Commission (CFTC), Health Insurance Portability and Accountability Act (HIPAA), Security and Exchange Commission (SEC) 17a-4(f), Sarbanes-Oxley Act (SOX), and the Model Requirements Specification for the Management of Electronic Records (MoREQ), to name a few.

  Content storage solutions should look to not only retain critical and sensitive information but also provide auditable reporting in order to protect organizations during internal or external audits, investigations, or litigation events. Some providers, such as Cleversafe, maintain an audit log in the user interface and the REST API. Several features are important to look for when deploying a compliance archive. Compliance workloads often require the ability to identify specific content as immutable and provide the necessary tracking and reporting in order to verify that that specific piece of content was kept immutable for a certain

amount of time. Often, the capability to make a certain piece of content immutable is referred to as write-once-read-many (WORM) or locking.

This feature designates a certain type of content as read-only for either a certain period of time or until the administrator changes the status of the content. Differentiation is often delivered via greater granularity of control where some solutions apply the read-only tag at a volume or folder level, and others have the ability to apply the designation to specific objects based on unique metadata information. Cleversafe, for example, offers the functionality to designate vaults as immutable, eliminating the ability to delete any contents and to log all changes to documents. While multiple solutions provide WORM functionality, they each have different ways of applying the read-only designation. Always ensure that the solution complies with the appropriate regulations or standards for your industry.

Solutions also differ based on the level of protection, specifically how the immutability tag can be removed. Look for solutions that can apply immutability based on a predesignated amount of time, such as five or seven years, removing the need for manual intervention. For many of these environments, the end goal is to prove that a specific object was not modified. If a litigation event occurs, the easier it is to prove that a file was not tampered with, the easier it will be for the organization. Also look for solutions that can provide immutability with versioning, where a file may be modified but the storage solution can save immutable snapshots or versions of that object at designated time intervals. Some providers, such as Cleversafe, offer content immutability along with versioning support via either copies or immutable point-in-time snapshots.

As with active archive solutions, strong search capabilities are critical for compliance archives. When a litigation action occurs, the easier it is to isolate the appropriate material, the less operations will be impacted. Solutions without search capabilities may be required to turn over far more information than actually required during a legal action if the organization can't isolate and verify that all the appropriate content has been provided. Finally, look for solutions that can provide auditable and verifiable content destruction when objects expire and are eligible for deletion as well as digital shredding capabilities.

- **Content distribution:** This third use case requires a combination of improved performance and multi-site distribution. Whether for online digital content distribution or content aggregation (such as video surveillance), this use case requires the features of object storage active archives, such as the ability to scale and protect at scale while controlling storage costs, but also requires significant amounts of content to be moved or updated across multiple physical sites. The solutions often require the ability to provide high levels of throughput along with the ability to scale throughput performance often by adding more storage nodes. Additionally, the ability to control the scale of performance with quality of service (QoS) capabilities in order to ensure that other operations do not interfere with available bandwidth is another aspect to providing performance.

  Content distribution network (CDN) solutions are designed to support large numbers of geographically dispersed users, such as scale-out web applications for digital media streaming. In addition to the necessary performance, these solutions also need the ability to serve content from multiple nodes and multiple sites to protect content delivery if a storage solution, a data center, or even an Internet service provider (ISP) fails. Cleversafe SmartRead technology, for example, predicts the optimal network routes and storage nodes to most efficiently return digital content. Digital content is then reassembled in segments, and Cleversafe constantly optimizes the path to return content in real time as it is streaming. If a connection failure or server slowdown occurs, which are common occurrences across the Internet, recreation of content is routed around the failures and content delivery occurs seamlessly. For content distribution architectures, also look for object storage solutions that provide multi-site protection with automatic geographic distribution of content along with the ability to deliver any piece of content from multiple nodes and locations. Some object storage providers, such as Cleversafe, have partnerships and validated solutions with multiple CDN partners.

- **On-premises cloud solutions:** As mentioned earlier, object storage solutions are designed to provide affordable massive content storage, with proprietary and open source object storage solutions advertising

cloud-level economics and scale to enable on-premises private clouds. With the rise of cloud orchestration layers, whether in the form of VMware, OpenStack, or Cloudstack, the capabilities of object storage align to a cloud infrastructure or hyper-scale model. One obvious feature to look for is support for S3 and Swift API protocols, along with certified cloud solutions, such as Openstack certification. Cleversafe, for example, supports the S3 and Swift API protocols for better cloud integration. In addition to the protocol support, a software-defined storage architecture that allows for hardware flexibility can allow the storage deployment model to align with the hardware cloud deployment model and help keep storage costs affordable.

The ability to provide metering and billing reporting capabilities can benefit organizations looking to divide and allocate cloud storage resources between multiple clients or tenants. Cleversafe is an example of an object storage provider that offers multi-tenancy. Look for solutions that can specify which tenant has access to which component of the infrastructure. Additionally, if your organization is planning or even considering expanding into public cloud storage, hybrid cloud support can be beneficial as well.

- **Data lake/hub:** The rise of big data analytics and the potential for competitive advantage derived from the resulting improved business insight have placed a recent focus on the proper method for storing and protecting the data required for these solutions. As mentioned earlier, a data lake or hub is required to scale to keep up with the creation of content, protect that content, and house a variety of disparate data types from multiple protocols.

  When evaluating an object storage solution, there are a few key capabilities, beyond the need to scale and protect at scale, which can help in a data lake or hub environment. Look for solutions that support multiple protocols to ensure the ability to serve as a repository for multiple application workloads. Data lake environments can also be better served by robust metadata tagging to ensure better and faster search capabilities. Some solutions, such as Cleversafe, support HDFS and offer integration into Hadoop environments. The ability to run analytics directly on the established storage environment can reduce or eliminate the need to physically move data off the storage system to run the analysis.

- **Collaboration**: As multi-site or global content collaboration increases in popularity, object storage solutions can provide a scalable, affordable, and multi-site architecture to support collaboration applications, such as file sync and share or video editing. While online file sync and share solutions, such as Dropbox, have become more popular, a large number of organizations desire the accessibility of public cloud solutions but require the content to remain onsite. During ESG's research study into next-generation storage architectures, several organizations that contributed to the study responded by saying they wanted the accessibility of the cloud, but security concerns precluded them from storing the content off-premises. As a validation of this finding, the report found multiple organizations deploying file sync and share software applications that leveraged on-premises content storage, such as Citrix Sharefile.[2] When looking for an object storage solution to serve as the back-end for a file sync and share solution, look for several key features: In addition to the object storage staples of scalability and the ability to protect at scale, collaboration solutions are often better served by automatic geographical distribution capabilities, allowing content to be automatically distributed across the global environment. As an example, Cleversafe messages their geo-distribution capability as designed for collaboration solutions. Additionally, with support for file system protocols, Cleversafe is able to increase the number of collaboration applications supported. As such, Cleversafe, offers certifications with application sync and share partners.

- **Backup target and file server consolidation:** Object storage has also found a home as a backup target. With backup software applications supporting NAS protocols and recently adding cloud protocol support such as S3, IT organizations are finding that object storage solutions provide the right blend of scalability and affordability to serve as backup targets. Features to look for include high bandwidth and throughput support, the ability to scale capacity and performance, affordability, and deduplication or compression capabilities. Multiple storage providers, such as Cleversafe, are also certified to work with a variety of backup software providers such as Symantec, Veeam, Commvault, and Unitrends.

[2] Source: ESG Research Report, *Next-generation Storage Architectures*, March 2015.

While file server consolidation is typically a NAS use case, the inclusion of NAS protocols in a number of object storage solutions, combined with the fact that a number of file system providers leverage erasure coding for protection, has blurred much of the line between NAS and object in recent years. Some providers have provided specific capabilities to serve traditional NAS workloads such the deployment of home and file shares.

Object storage has the potential to serve a growing number of use cases across multiple IT environments. While the technology still has a way to go before taking a dominant share of the storage market, multiple industry trends are moving in favor of object storage. It is possible that in the not too distant future, storage will no longer be defined by protocols, such as Fibre Channel or NAS, or even be defined as block and file. With the combined innovations of object storage and solid-state, the future of storage may shift to be divided between low-latency transactional storage and massive content repositories. If such a shift were to occur, object storage would likely emerge as the capacity platform of choice, offering a highly scalable and affordable storage foundation to support a wide variety of workloads.

The Bigger Truth

Whether for personal or professional use, our populace has an insatiable appetite for data. As drive capacities increase, existing storage architectures are at or past their limits. Object storage provides a superior method to protect and scale data storage at high-capacity points. With emerging trends such as on-premises cloud storage, the Internet of Things, and big data analytics, the demand for object storage capabilities will likely continue to increase. The simplest way to look at the situation is this: At current growth rates, the content within a typical organization will eventually reach the point where it can no longer be protected by traditional means. Traditional external storage solutions meant to consolidate data storage for easier management will be unable to keep pace with data growth and will become their own silos. The resulting isolated pools of data will be not only more difficult to manage and protect, but also more difficult to leverage for analytical analysis. Even for organizations able to make due today, a tipping point is likely somewhere on the horizon. When that tipping point is reached, a potentially painful migration event will occur. The earlier you can make the transition to an architecture that can scale, the better. With the right architecture, IT resources burdened previously by simply keeping pace can be freed to focus more on adding business value. In this ever-evolving world, organizations are consistently looking to leverage data to find a competitive edge. Eliminating the burden of infrastructure scaling and shifting focus to maximizing the value of data can be the difference between leading the pack and falling behind.

Enterprise Strategy Group | **Getting to the bigger truth.**

20 Asylum Street | Milford, MA 01757 | Tel: 508.482.0188 Fax: 508.482.0218 | www.esg-global.com