White Paper

EMC Elastic Cloud Storage Offers Resilient Scalability for the New Generation of Workloads

By Scott Sinclair, Storage Analyst

October 2015

This ESG White Paper was commissioned by EMC and is distributed under license from ESG.

Contents

Introduction

This paper will present the technical and business benefits of EMC's Elastic Cloud Storage (ECS) solution. It will offer details on how ECS offers the potential to solve a number of data storage challenges associated with the rapid rate of data growth and the rise of a new generation of IT workloads.

The cumulative impact of years of data growth has created a cascading set of challenges for IT data center storage environments. Additionally, traditional SAN and NAS storage array architectures were not typically designed to store or protect data at large multi-petabyte capacity levels. This new era of massive content storage environments requires a new storage architecture, object storage. With object storage technology, IT organizations can not only keep up with rising capacity levels, but also store these new capacity levels at a manageable cost point.

The sheer volume of data, however, is not the only element that has changed since the traditional storage architectures were conceived. Organizations create, consume, and digest data differently. Developers are designing applications to serve data out to millions of mobile devices globally, or to store data generated from potentially hundreds or thousands of sensors worldwide serving the Internet of Things. Storage architectures are required to serve content at any time to anywhere in the world while providing the ability to centralize all of this data for business intelligence analytics.

While originally designed for archive storage of inactive data, the architecture and the capabilities of object storage have evolved in recent years. The more advanced object storage solutions offer a number of characteristics that help make it an ideal, or near ideal, platform for massive content and even cloud storage. Object storage technologies enable organizations to scale to near infinite capacity points in a single flat address space, eliminating the idea of a storage silo. Object Storage also introduces new data resiliency techniques to ensure data protection at massive capacity levels. Additionally, with multiple public cloud storage providers offering access to object storage as a service, deploying on-premises object storage solutions can enable a consistent experience for hybrid on- and off-premises environments. Finally, object storage solutions often leverage commodity hardware, helping to control storage infrastructure costs.

EMC, a market leader in storage, understands the potential for object storage and, as such, offers its Elastic Cloud Storage (ECS). EMC's ECS solution is designed for a number of next-generation workloads such as globally distributed content storage repositories, data analytics, modern or cloud application development, data lakes and cold archive, as well as storage for the Internet of Things (IoT).

Object Storage

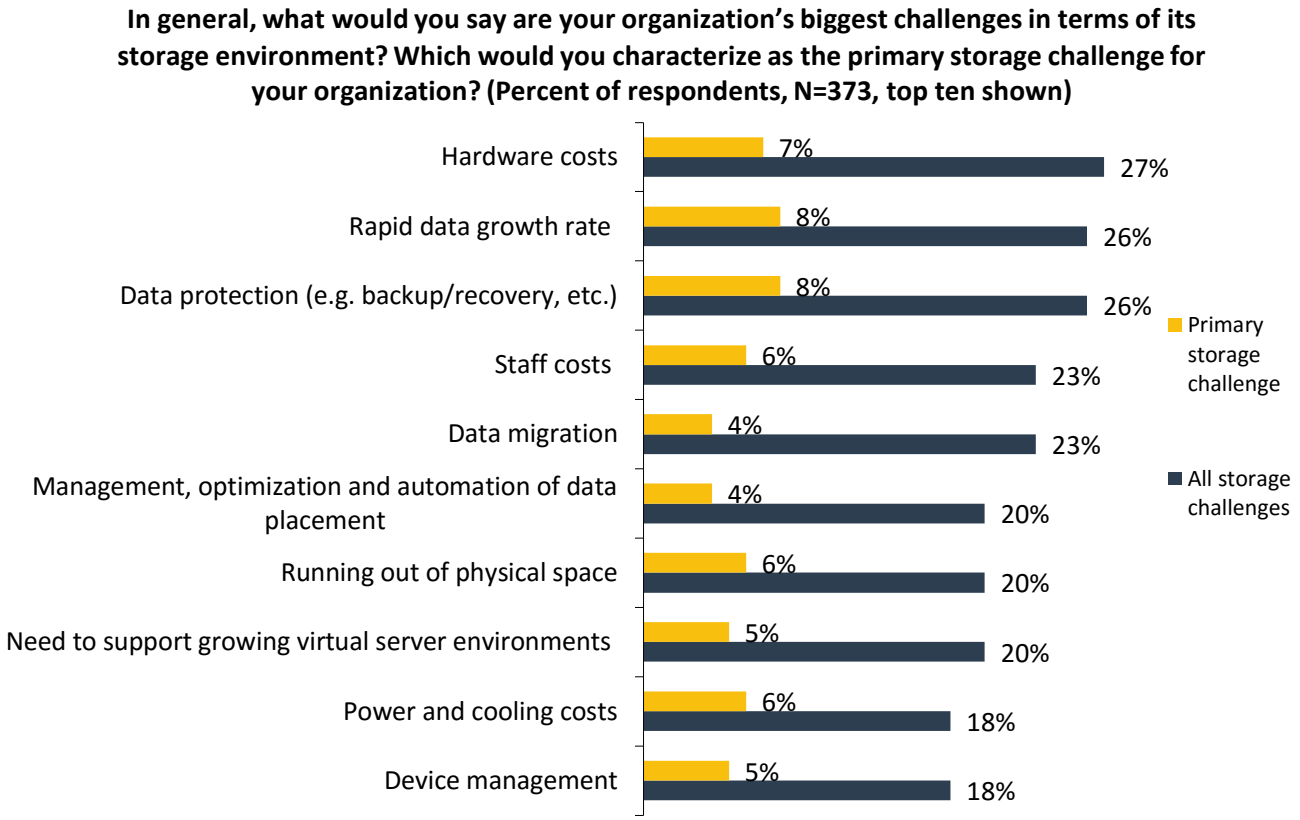
Object storage represents an evolution in the ability to store unstructured data, and as such object storage is sometimes considered an alternative for NAS storage. That idea, while technically accurate, can be a little misleading. While object storage often provides support for file protocols, such as NFS and SMB, the scalability of its architecture, along with some of its more advanced capabilities, greatly differentiate it from traditional NAS. For example, object storage offers automatic geo-dispersed data protection and the ability to leverage commodity hardware, making it an ideal storage environment for a new generation of IT workloads.

The scalable flat address space of object storage differs from the hierarchical structure of file systems. With object storage, file access is provided via a unique identifier. While for file systems, however, file access is provided via a path to the file. 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. While additional technical differences exist between object and file technologies, this paper will focus more on the resulting benefits, rather than the technical nuances. The takeaway, however, is that the design of object storage offers many differentiated benefits, such as managing the rapid rate of unstructured data growth and ensuring multi-site survivability.

The Challenges of Data Growth

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 environment. One of the questions asked respondent IT decision makers to identify their organization's biggest storage challenges. While the rapid growth of data being cited as a top challenge should not be a surprise, what may be somewhat surprising is that the rest of the challenges within the top ten most-cited responses can all 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.[1]

Figure 1. Organization's Biggest Challenges in Terms of Its Storage Environment



Source: Enterprise Strategy Group, 2015.

The takeaway is twofold: First, data growth continues its seemingly perpetual dominance as a top storage challenge. Second, the data growth generates a cascading effect that spreads across the data center and impacts multiple physical and operational aspects, further increasing costs.

The Key Benefits of Object Storage

Object storage solutions were designed to solve many of the challenges identified in Figure 1 that stem from massive levels of data growth. Protecting data in massive content storage environments while keeping storage costs under control requires a number of innovations, including:

- **Replica and/or information dispersion algorithmic protection:** Object storage commonly leverages object replication, information dispersion algorithms, or a combination of the two to provide data protection. These technologies provide several advantages over traditional RAID, including the ability for faster data

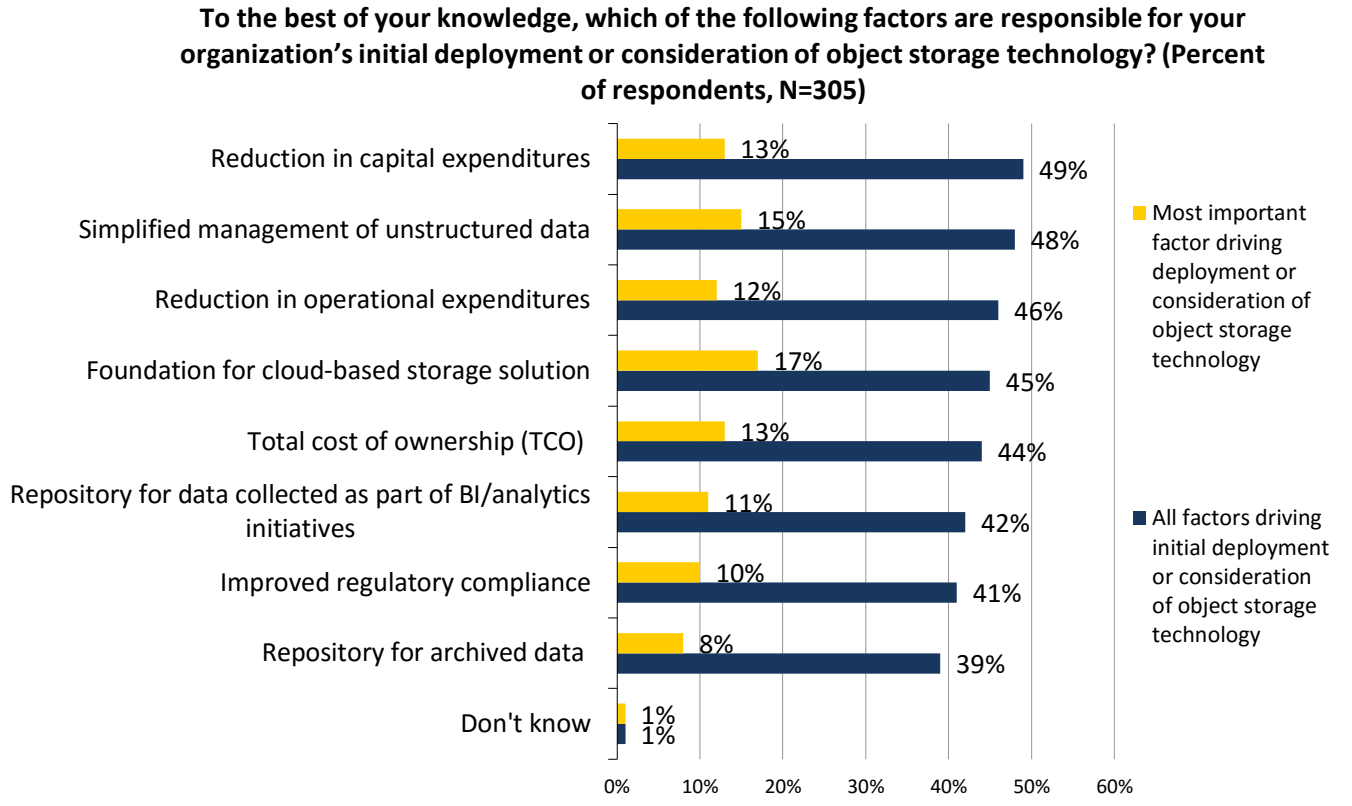
[1] Source: ESG Research Study, *2015 General Storage Trends Survey*, conducted in May 2015. All ESG research references and charts in this white paper have been taken from this research study.

recovery. By rebuilding in free space instead of having the rebuild process constrained by the performance of a single hot spare, recovery from a component failure in an object storage solution is often significantly faster. The net result is that object storage can handle a higher number of failures while still protecting data.

- **Multi-generational commodity hardware support:** Another benefit of both replica- and information dispersion algorithm-based protection schemes is the ability to provide resiliency while leveraging commodity server hardware infrastructure, rather than requiring a dual controller storage design. This allows object storage software the ability to leverage lower cost hardware. Additionally, many object storage architectures support the ability to mix and match different commodity hardware offerings. The net result allows data in place hardware generation upgrades to occur without requiring a data migration event. This is a necessary capability since object storage is designed for capacities often deemed too large to migrate as a whole.
- **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 (for example, 16TB). As capacities scale, providing multi-site protection for larger storage environments can result in the unwieldy management of scores of replication policies, if not more. Object storage's replica and information dispersion algorithmic protection schemes can often automatically expand their scope to spread protection automatically across data centers and geographies. The result greatly simplifies multi-site protection and greatly reduces the associated management costs.
- **Enhanced Metadata:** Another differentiated element of object storage is the option for more advanced metadata. Metadata is a term used for the data that stores information about particular digital data assets. File systems only support simple metadata information: basic file type and location information. Object storage, on the other hand, provides the ability to store advanced and/or custom metadata, which allows for more information to be stored with specific content. For example, object storage can store protection and retention requirements or search-related information with the object itself, improving the ability to run analytics or customize protection schemes

While the preceding list does not represent the full extent of object storage's capabilities, the overview provides some insight into how those technical advantages translate into benefits for IT organizations. To better understand how these capabilities translate into benefits, storage decision makers who expressed interest in object storage were asked to identify the factors driving that interest during ESG's 2015 research into the storage industry. The responses are shown in Figure 2.

Figure 2. Factors Driving Deployment or Consideration of Object Storage



Source: Enterprise Strategy Group, 2015.

Figure 2 reveals two narratives about the driving forces behind the interest in and potential of object storage. Focusing solely on the responses where organizations were allowed to choose all of the factors driving their initial deployment or consideration of object storage, IT decision makers often identified the potential of the technology to reduce the TCO of the storage environment, use less expensive hardware, and simplify management.

The other story the data reveals comes from focusing on the primary, or most important, factor driving consideration. This data set identifies the most popular factor driving consideration as a foundation for cloud storage. This data suggests that across the wide variety of potential use cases for object storage, organizations are turning to the technology to help control TCO in the wake of rising digital content levels. But of those potential use cases, more organizations are looking to object storage as the foundation for the next generation of modern and cloud-based workloads.

EMC Elastic Cloud Storage

To serve as a solution for both high-capacity and next-generation workloads (such as cloud), EMC has developed Elastic Cloud Storage (ECS), an offering based on object storage architecture. ECS is designed with the differentiated tenets of object storage discussed previously, such as massive scalability, data resiliency, and the ability to leverage commodity server hardware, to help ensure that the cost of storage stays reasonable. Additionally, for greater deployment agility and flexibility, EMC's ECS solution is, at an architectural level, software-defined storage (SDS). While the term SDS is used by multiple offerings and providers and can have multiple definitions, for ECS, an SDS architecture means that the solution can be procured as a purely software solution and deployed on the commodity server hardware infrastructure of the organization's choosing, or as an appliance with the software pre-installed on commodity hardware from EMC.

This flexibility not only simplifies the deployment of ECS, but it also allows ECS to provide a single pool of data that can spread data across a variety of underlying infrastructure components including commodity or ECS appliances, and even EMC and third-party storage arrays. In addition to the flexibility offered by the software-based

architecture, ECS enables multi-head access, allowing different protocols, such as object and HDFS, to access the same data concurrently.

Features of EMC's Elastic Cloud Storage

In addition to the core capabilities and benefits of object storage discussed, ECS offers a number of additional features including:

- **A strongly consistent, global scale-out namespace:** ECS presents a single global namespace across all the nodes within the solutions regardless of where those nodes reside, even when distributed across multiple sites around the globe. While other object-based solutions can provide the single presentation of data, ECS is able to do this in a strongly consistent manner. The result is that data can be read and written from any site, with the system designed to always provide the latest version. In other words, geographically distributed nodes can process write requests for the same object simultaneously, and write to different sets of disks. ECS also incorporates WAN optimization when data is distributed across sites to better improve performance.
- **In-place Hadoop analytics with HDFS support:** ECS simplifies the enablement of big data analytics with the ability to run Hadoop on data in place via HDFS support. When leveraging direct-attached storage (DAS) solutions, storage silos are likely to emerge. Storage silos can often result in higher data management costs and can limit the amount of data available for analysis. Additionally, ECS is able to offer data resiliency without requiring three copies of the data, unlike traditional HDFS replication environments. The net result can provide a significant savings in storage infrastructure costs while improving the time to results. While public cloud environments can provide access to larger volumes of data, they can also make accessing off-premises data storage both slow and costly.
- **Support for S3 and OpenStack APIs:** With a rising interest in both public and private cloud storage, organizations are developing more applications that leverage cloud-based API protocols, such as S3 and Swift. ECS provides support for both of these interface options, providing development flexibility not only for today but for the future as well. Additionally, for organizations seeking to leverage a hybrid cloud data center environment, ECS's support for S3 and Swift allows applications consistent data accessibility, easing development efforts.
- **Software-defined architecture leveraging commodity components:** As mentioned previously, one key benefit of object storage is the ability to leverage low cost commodity hardware. The resulting infrastructure flexibility allows for organizations to keep capital infrastructure costs under control while supporting massive content storage environments. For organizations that do not wish to procure storage hardware separately, EMC offers the ECS appliance. The appliance provides many of the benefits offered by pure commodity hardware-based solutions while also providing a tested and validated combined software and hardware storage solution. The solution is also completely supported by EMC's global service network.
- **Combined erasure coded and replication-based protection:** Object storage solutions often leverage either erasure codes or replicas to provide data availability and resiliency, whether failures are at the disk, node, rack, or data center level. With ECS, globally distributed protection is automatic and doesn't require complex configuration, while failover and recovery, according to EMC, is seamless. Erasure codes also offer more efficient use of storage, while replication can provide better performance, especially across sites. ECS leverages both to provide availability and resiliency across sites with a low storage overhead of around only 1.8 times the effective capacity. This combination of information dispersion algorithms and replication also allows for recovery from local hardware failure (the most common hardware failure), without requiring data to be read over a WAN connection.
- **Built-in journaling, snapshots, and versioning:** ECS offers automatic version protection for files to protect against changes. This capability can be critical for compliance-based workloads. ECS can ensure that as content evolves over time, the record of changes is maintained. Additionally, as an extension to protection, ECS offers built-in snapshots, allowing for specific data sets to be read as they existed at a specified time.

Finally, to ensure that the integrity of the digital content is maintained, ECS stores a unique checksum for every chunk of data written.

- **Multi-tenant architecture:** To better support cloud infrastructure environments, ECS offers a multi-tenant architecture. As such, ECS is able to apply quota limits on specific sections within the object store. ECS also provides metering capabilities with reporting on capacity, object count, and bandwidth. For organizations that have more complex or specific needs, ECS offers a REST API for more advanced monitoring and auditing capabilities.
- **Flexible deployment models:** For deployment, ECS is available via either an appliance or as software. As mentioned earlier, flexibility of deployment options offers numerous advantages. The ability for ECS to be deployed as software allows organizations greater choice in commodity hardware selection and can speed up the deployment process considerably. While some organizations may prefer choice, others may prefer to select a validated hardware and software appliance. ECS offers both options, and allows organizations to mix and match models within the same deployment.
- **On-premises data security:** By providing many of the benefits of public cloud while retaining data on-premises**,** organizations can have better visibility into the data security measures being applied to the storage infrastructure. Organizations can then in turn better control who can access the data and the storage hardware. This capability can be critical for organizations that store sensitive customer information or in industries with compliance or regulatory requirements. Additionally, the ability of ECS to support standard cloud APIs, such as S3 and Swift, can also ease the integration of content search and management functionality, allowing the organization to retrieve critical data when needed in a timely fashion.

These advanced capabilities expand on the differentiated elements of object storage, and provide a more protected, more manageable, and more flexible pool of storage that aligns with a number of emerging workloads.

Workloads Enabled by Elastic Cloud Storage

The core capabilities of object storage along with the added innovations offered by EMC's ECS solution provide a number of advantages for a variety of workloads and IT environments. This section will walk through a number of environments experiencing benefits from object technology. This is by no means an exhaustive list, but does highlight some of the areas where ECS can help improve TCO and ROI for organizations.

- **Universal object storage platform for Internet of Things (IoT) with in-place analytics:** The strong consistency model of EMC's ECS is an ideal fit for IoT environments. With support for strong consistency, data stored on ECS can be read or written to as soon as it is applied from any one of multiple globally distributed sites. This capability offers a variety of benefits. Application development can be easier. As the data pool grows and evolves, applications do not need to be updated to adjust for data locality. This is increasingly important for the IoT environments. IoT-based workloads often require the ability to access and comprehend data collected from multiple sources and locations. The strong consistency offered by EMC's ECS provides the ability to collect and store sensor data in locations closer to the actual "things," which in turn improves performance and saves cost.
- **Global content repository**: For content distribution workloads, ECS offers a single scalable, resilient, and cost-effective pool that can serve multiple application and content types, enabling all the content within the repository to be globally accessible by web, mobile, and cloud applications. ECS capabilities such as geo-protection and policy-based retention are critical aspects of ensuring resiliency and accessibility across multiple globally dispersed sites while maintaining compliance and governance. EMC claims that via the ability to leverage commodity hardware, ECS solutions can offer up to a 65% lower cost than that of public cloud.
- **Foundation for a cloud storage solution**: As seen in ESG research, the top primary reason respondents identified for considering an object storage solution is using it as a foundation for a cloud solution (see Figure 2). A key element of deploying a cloud infrastructure is the ability to dynamically allocate

infrastructure resources and report on what resources are being leveraged. ECS's multi-tenancy and metering capabilities help in this respect, offering improved control and manageability. ECS's potential for nearly infinite scalability along with the ability to leverage commodity hardware make the solution well suited for helping to keep infrastructure costs low. In multi-petabyte environments, any price per capacity decrease, or increase, can dramatically impact storage costs.

- **Data lake for business analytics:** ESG's research into general storage trends also revealed that business and intelligence and data analytics was the most common response when IT decision makers were asked which initiative would drive the most storage growth over the next 12 months. ECS offers the ability to pool all of that content into a single repository. The net result reduces the number of storage silos and copies required for business analytics, and reduces the need to move data back and forth for analysis. The globally distributed and multi-protocol nature of ECS allows analytics applications to access the data seamlessly wherever it resides. The ultimate result can improve efficiency and can save TCO.
- **Modern application development:** As mentioned previously, organizations are developing a new generation of applications. Isolating data in storage silos can impede development activities. Modern application development can often require geographically distributed development organizations. For these environments, the geo-distribution capabilities of ECS can offer significant efficiency advantages. To this end, the scale and automatic geographic accessibility of ECS can provide a significant benefit, especially when this new breed of applications requires access to a large pool of read-only content.
- **Cold storage archive:** Content archives are sometimes confused with backup, but are very different. An archive is a primary data store, not a redundant copy. In many cases, archives serve as a means to control the growth of unstructured data by migrating "cold" data off of high-performing and more expensive storage. The demands of an archive align with the capabilities of object and ECS. Archives must offer massive scalability along with high levels of resiliency, while allowing for the data to remain online and accessible.

While these are just a few of the areas where ECS can benefit IT organizations, they represent many of the more modern and data-intensive IT initiatives underway in the industry today. The bottom line is that the next generation of workloads is looking to be even more dependent on larger levels of digital content than ever before, and therefore needs the next generation of storage.

The Bigger Truth

As applications and IT workloads evolve, data storage infrastructure must be able to support that evolution. Older, less scalable, and less manageable infrastructure can hinder modern workload development by making storing and accessing the necessary data too costly. Object storage solutions, such as ECS, offer a resolution. Much like how solid-state storage is designed for and therefore ideally suited for low latency transactional workloads, object storage is designed for and ideally suited for large content storage. While this paper has already covered the technical benefits of object along with some of the advanced functionality of ECS, there is another key benefit of ECS, or namely EMC, that is also important to discuss.

EMC's ECS solution, ultimately, offers a foundation upon which IT organizations can build out the next generation of applications and workloads.

With its design for massive capacity storage, the selection of an object storage partner can create somewhat of a headache for IT organizations. The nontrivial nature of moving large amounts of digital content, sometimes to the level of multiple petabytes, adds a degree of lock-in. If a partner can't deliver on its promises for whatever reason, the impact to the business of switching solutions can be significant. For organizations where this is a concern, EMC's position as a market leader in storage can offer a degree of comfort. EMC has a long history of providing enterprise storage solutions, and will likely continue to do so far into the future.

EMC's ECS solution, ultimately, offers a foundation upon which IT organizations can build out the next generation of applications and workloads. Whether for mobile content, cloud infrastructure, or business intelligence, the demand

for consistent, flexible, and even global access to data will likely only continue to increase. For these environments, traditional storage array silos face a losing battle. As the level of content increases, more traditional storage architectures will likely be too costly. The next-generation data center will need a next-generation storage solution, like ECS.

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