ESG Economic Validation

NetApp StorageGRID

Maximizing Economic Value Via Software-defined Object Storage for Data in the Hybrid Cloud

By Jack Poller, Senior ESG Validation Analyst; and Alex Arcilla, ESG Validation Analyst

September 2018

This ESG Economic Validation Report was commissioned by NetApp and is distributed under license from ESG.

Contents

ESG Validation Reports

The goal of ESG Validation reports is to educate IT professionals about information technology solutions for companies of all types and sizes. ESG Validation reports are not meant to replace the evaluation process that should be conducted before making purchasing decisions, but rather to provide insight into these emerging technologies. Our objectives are to explore some of the more valuable features and functions of IT solutions, show how they can be used to solve real customer problems, and identify any areas needing improvement. The ESG Validation Team's expert third-party perspective is based on our own hands-on testing as well as on interviews with customers who use these products in production environments.

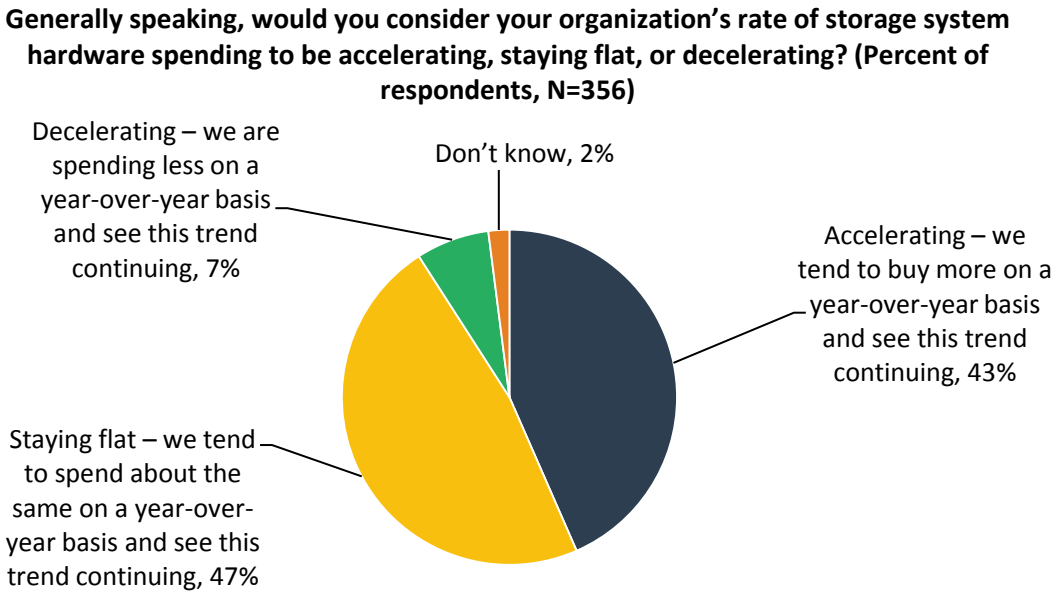
Introduction

ESG evaluated the NetApp StorageGRID solution to validate the economic value that an organization can achieve when using the solution to manage object data across its hybrid cloud. We also evaluated how StorageGRID helps an organization set up data management policies that apply to all object data regardless of physical location and manage those rules via policies, while minimizing both capital and operational expenses. Finally, we examined how an organization can employ StorageGRID to leverage Amazon Web Services (AWS) serverless applications while minimizing network transfer and overall cloud storage costs.

Background

In a recent ESG research survey, 54% of organizations indicated that their spending on storage system hardware will either remain flat or decelerate in the foreseeable future (see Figure 1).[1] As a matter of fact, 23% of respondents indicated that the primary reason for this slowdown in spending is the fact that they are using more cloud applications.[2]

Figure 1. Rate of Storage System Hardware Spending



Source: Enterprise Strategy Group

These trends make sense considering the huge growth in unstructured data that organizations are generating and collecting. With multimedia, customer service recordings, social media, and the Internet of Things (IoT), organizations want to analyze and extract value from that data to contain costs or increase revenue. However, this rise in unstructured data can make on-premises storage cost-prohibitive, while increasing storage administration costs, prompting organizations to consider hybrid solutions.

Employing hybrid storage raises issues for organizations, especially when both on-premises and cloud storage are geographically dispersed. They must consider how they can easily apply storage policies in real time so that users can access data regardless of time or place and archive data when not accessed frequently. At the same time, organizations must manage storage and retrieval of the data without incurring additional management costs. They also have to maintain performance regardless of where or when users access data. Finally, they must determine how they can best leverage a

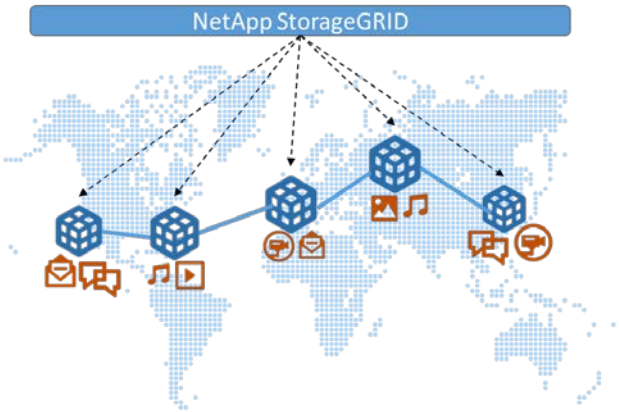
[1] Source: ESG Master Survey Results, *2017 General Storage Trends*, November 2017.
[2] Ibid.

cloud's compute services on their unstructured data without incurring excess transfer costs between on-premises and cloud storage.

The Solution: NetApp StorageGRID

An organization can use NetApp StorageGRID, a software-defined object storage solution,to specify how and where object-based data is stored. StorageGRID enables an organization to manage on-premises storage equipment in geographically dispersed locations as nodes within one single storage system. The solution can be deployed via a combination of containers, virtual machines, and appliances.

Using StorageGRID, an organization can configure rules and policies that define how long and where object data and its replicas (or copies) can be accessed and stored amongst storage nodes. Replicas ensure consistent performance with user access, regardless of where the data physically resides. When an organization changes an Information Lifecycle Management (ILM) rule or policy, StorageGRID applies those changes to both new and stored object data, ensuring that storage is used efficiently such that data is stored to balance availablity, durability, and cost.



For an organization wishing to implement hybrid storage, NetApp designed its latest version of StorageGRID to integrate with Amazon Simple Storage Service (S3). CloudMirror automatically ensures that select data is mirrored from on-premises storage to S3 and vice versa. With CloudMirror, organizations can leverage AWS resources to access data without first having to manually copy that data to the cloud.

NetApp also added support for pub/sub messaging using AWS Simple Notification Service (SNS). Application developers can use SNS to automate the processing of StorageGRID objects regardless of physical location. For example, an application can publish the creation of an object—a photo—on StorageGRID. Another application can subscribe to such messages, and upon receiving the notification, can run a function such as AWS Rekognition to implement automatic facial recognition. Thus, StorageGRID expands the on-premises environment into the cloud.

Because StorageGRID enables an organization to set up policies quickly and dynamically, automate policy execution, and leverage AWS S3, the solution also offers economic value in terms of lower storage costs, lower administration costs, and increased end-user productivity.

ESG Validation

ESG evaluated the potential savings in both capital and operational expenses that organizations could achieve using an example deployment scenario. We also performed hands-on evaluation and testing of the NetApp StorageGRID solution to determine how organizations can set up and manage their hybrid storage using existing on-premises storage and Amazon S3, create ILM rules and policies to be applied to objects entering the StorageGRID, and leverage Amazon's microservices to analyze data within the StorageGRID on an as-needed basis. This evaluation also revealed how the solution's capabilties can specifically help to decrease expenses.

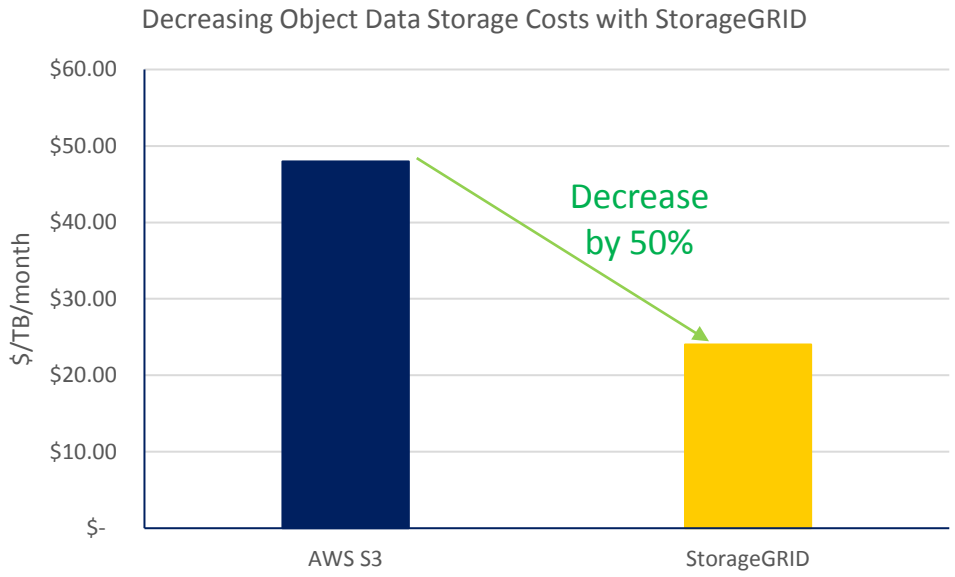
Total Cost of Ownership

ESG first assessed the potential economic value that an organization can gain from using StorageGRID over AWS S3 exclusively. We evaluated the comparative costs of using StorageGRID and AWS S3 exclusively over a five-year period. We also audited existing models and verified available pricing to conduct our analysis. We wanted to verify that StorageGRID

provides an economically viable alternative to using S3 when storing object data. Note that these are modeled estimates for a simplified case. As requirements may vary drastically, ESG suggests that organizations conduct their own TCO analysis to compare solutions and estimate potential savings.

For our analysis, we assumed 1000 TB of usable capacity for the StorageGRID. The average object size is 100 MB. We also assumed that objects will be erasure-coded, which impacts the amount of storage reserved. We compared this with the use of standard-class S3 (implying that users will consistently be accessing data on a frequent basis) across two AWS regions. With StorageGRID, we considered costs related to storage and network hardware and software, installation and support services, facilities, and IT administration/management. Conversely, we considered costs associated with monthly standard S3 usage and AWS support. Figure 2 compares the expected cumulative costs incurred over five years.

Figure 2. TCO Comparison – Usage of StorageGRID vs AWS S3 over a Five-Year Period



Source: Enterprise Strategy Group

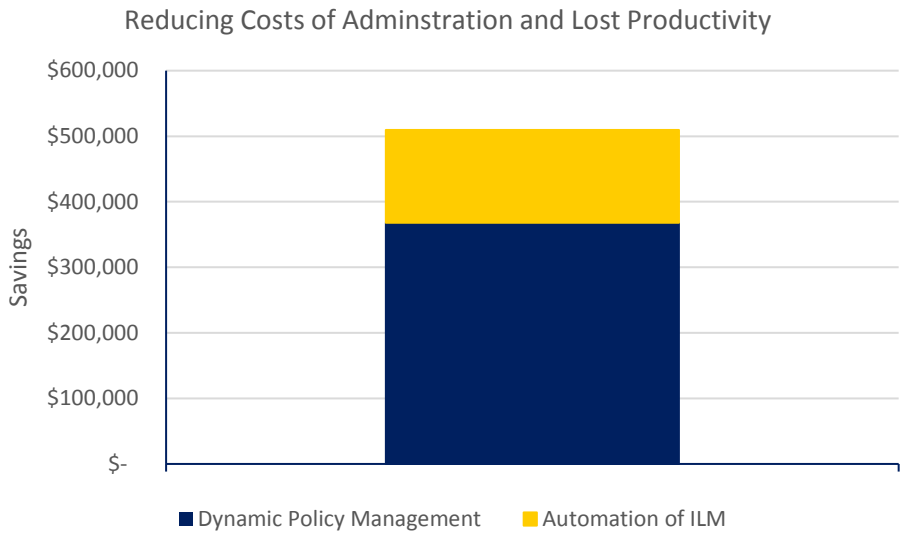
Assuming a 30% discount, ESG found that the cost for using StorageGRID is $27.03 per TB per month, compared with $47.98 per TB per month over five years for using S3 exclusively, representing a a 50% savings. We found that an organization can achieve savings primarily through the elimination of cloud ingress, egress, and inter-region transfer costs, specifically when object data stays on-premises on the StorageGRID system while calling on AWS microservices for additional analysis. This eliminates the need to copy data onto S3, which would incur those network transfer costs. By configuring StorageGRID to span both on-premises and cloud storage, an organization can achieve savings in cases such as controlling when data is replicated and sent to the cloud for analysis via AWS computing services or cloud bursting when a data center is at peak load.

ESG also examined the potential operational savings to be achieved when using StorageGRID. As mentioned in previous sections, the time spent on creating and deploying ILM rules and policies can reduce the amount of manual intervention when managing how and where data is accessed. Typically, an IT administrator would periodically examine how often users access available data, decide if data needs to be tiered from more expensive to less expensive storage tiers, then develop migration plans. In turn, this will impact end-user productivity as migration entails some downtime.

Figure 3 shows the estimated savings that can be achieved over a five-year period. Based on our previous scenario, we assumed that this organization has 1000 end-users. ESG estimated the costs associated with monitoring, planning and execution tasks that must be done by administrators when not having the dynamic policy management and automation

capabilities offered by StorageGRID. We also account for the loss in end-user productivity when an administrator must manually migrate data across tiers, thus incurring downtime.

Figure 3. Estimated Savings in IT Administration and End-User Productivity



Source: Enterprise Strategy Group

ESG estimates that the total operational savings over five years will be approximately $509,000. Costs associated with dynamic policy management and ILM automation include the time spent by an administrator to plan and execute data migrations that balance data access needs with minimizing overall storage costs. Because the organization will ingest data many times during the year, these tasks are completed according to business needs. Conversely, the time that an administrator will have to execute on data movement plans will result in end-user downtime, translating into lost productivity.

Why This Matters

While organizations are aware of cost savings that can be achieved using cloud storage, it is critical to consider the cases in which cloud storage can be leveraged without spending unnecessarily, especially when object data access may be infrequent. Employing a hybrid storage option may be a viable alternative to going "all-in" with cloud storage.

ESG validated cost estimates associated with using StorageGRID over AWS S3 over a five-year period. Based on our modeled scenario, we found that the cost per TB per month of using StorageGRID is 50% lower than that of using AWS exclusively. We saw that StorageGRID achieved savings primarily with cloud ingress and egress costs. Should the organization use cloud storage for archiving purposes, cloud bursting, or for accessing compute services on-demand, StorageGRID provides an alternative way to leverage S3 without incurring unnecessary network transfer costs. In this same scenario, we also estimated that an organization could potentially save approximately $509,000 in administration costs and gained end-user productivity.

Dynamic Policy Management

ESG proceeded to test how an administrator can create ILM rules and policies specific to certain types of object data. These rules define:

- How the data is protected (i.e., how many replicas are created and where they are stored).

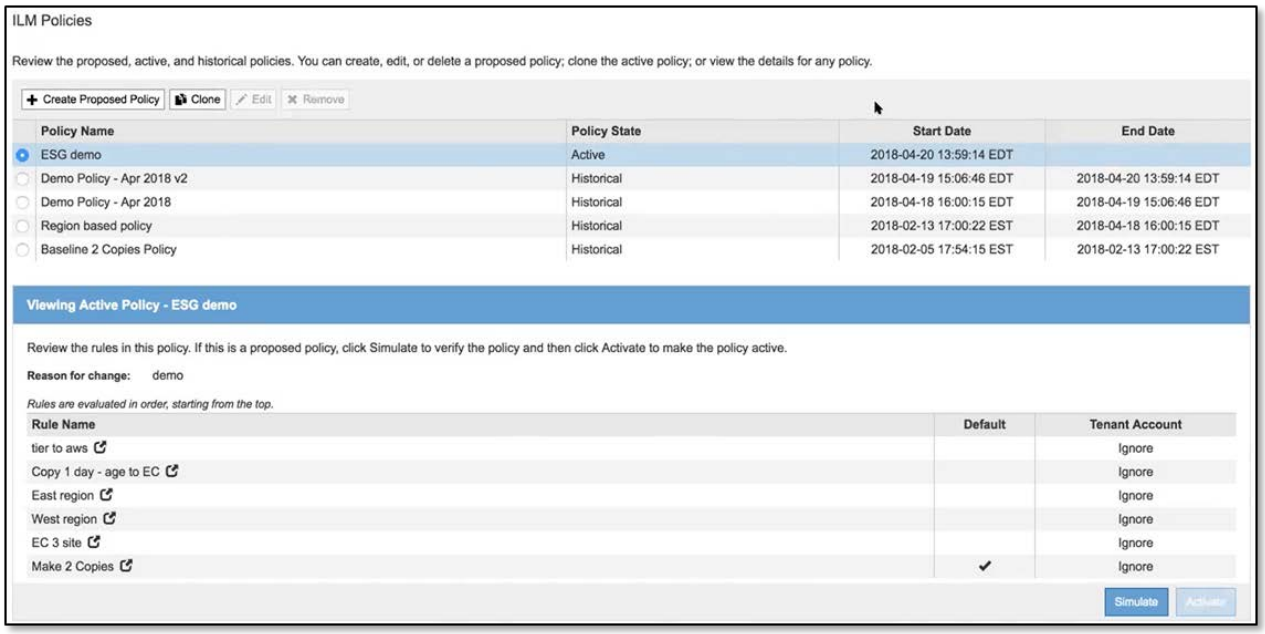
- How long the data will be kept.
- At which data centers the object data will be kept.

These policies ensure that data is stored to satisfy user needs while making the most efficient use of the available storage, balancing both access performance and storage costs. Storage administrators can implement policies managing data over multiple sites.

ESG Testing

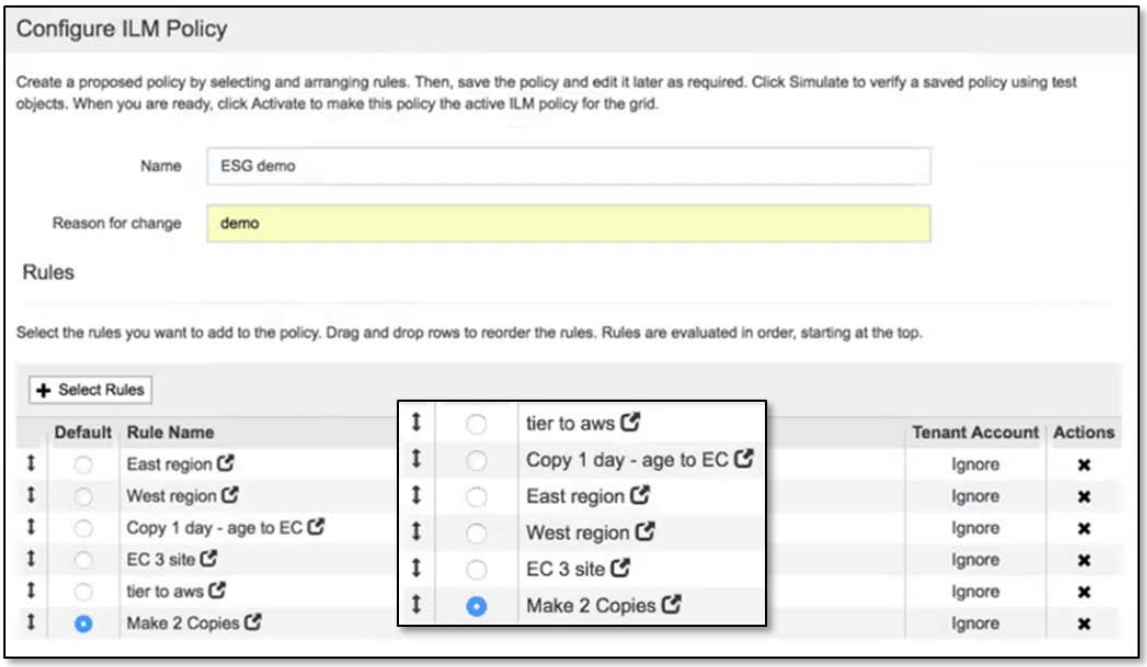
ESG viewed the list of existing policies by navigating to the *ILM* menu and selecting *Policies*. We saw both active and inactive policies with start and end dates (see Figure 4). We then selected the "Demo Policy – Apr 2018 v2" policy and viewed its rules and the order in which they will be applied to object data. We also noted that StorageGRID always includes the active default rule "Make 2 copies," which makes two copies of an object and stores them in any two nodes.

Figure 4. List of Existing ILM Policies with 'ESG Demo' Rule Detail



From the main **ILM Policies** screen, ESG clicked on the "Create Proposed Policy" button. We arrived at the **Configure ILM Policy** screen and typed in "ESG demo" (Figure 5). All rules inputted prior to creating "ESG demo" appeared, and we selected those rules to be included. We changed the order of the selected rules, either by clicking the double-headed arrow located to the left of each rule or dragging and dropping rules within the list (new arrangement shown in Figure 5). Changing rule order is important to a storage administrator to ensure that data is accessible, retained, and/or archived according to business needs (e.g., archive object when frequency of user access decreases). An ILM policy can help in using available storage efficiently, thus minimizing storage costs.

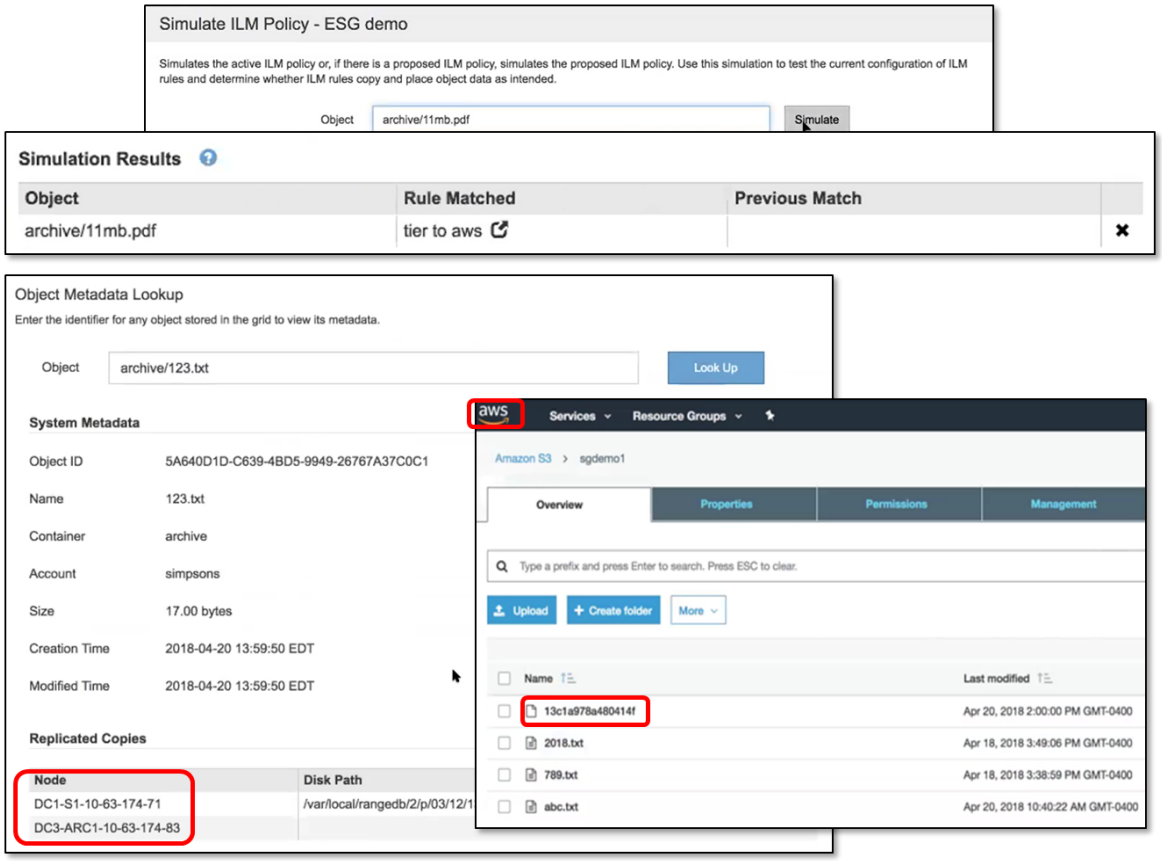
Figure 5. Configuring an ILM Policy by Rearranging Order of Rules Applied



Clicking ***Save*** marked the policy as "proposed" but not yet implemented. Administrators can test proposed policies before activation. ESG proceeded to test the "ESG demo policy," which has "tier to aws" as the first rule to apply. We uploaded a file named "11mb.pdf" to StorageGRID. ESG navigated back to the **ILM Policies** screen, selected the "ESG demo" policy, then clicked on the "Simulate" button in the lower right-hand corner. The window "Simulate ILM Policy – ESG demo" appeared (top portion of Figure 6). We typed the filename "archive/11mb.pdf" into the *Object* field and clicked on the "Simulate" button. We noted that the filename "11mb.pdf" matched the criteria in the ILM rule "tier to AWS," and that the file was replicated to AWS S3.

ESG conducted another test by searching on the object metadata. We uploaded another file named "123.txt" to S3. We then navigated to the *Object Metadata Lookup* option under the **ILM** menu and entered "archive/123.txt" in the Object field, then clicked the "Look Up" button (bottom portion of Figure 6). The results at the bottom of this screen showed that the default rule, "Make 2 copies," in the "ESG demo" policy was applied; one copy appeared in a storage node; the other in S3. To verify that the object was copied to S3, we navigated to the S3 console and saw the object ID of "123.txt" at the top of the list. Note that this object ID represents the object format that can be read by AWS S3. (If we used the CloudMirror feature, objects would appear as the same filename regardless of whether they are in on-premises storage or S3). Pretesting a policy helps the administrator verify that it has the intended effect on the object data of interest. The administrator can then create the policy once and apply that policy repeatedly across the organization's object storage.

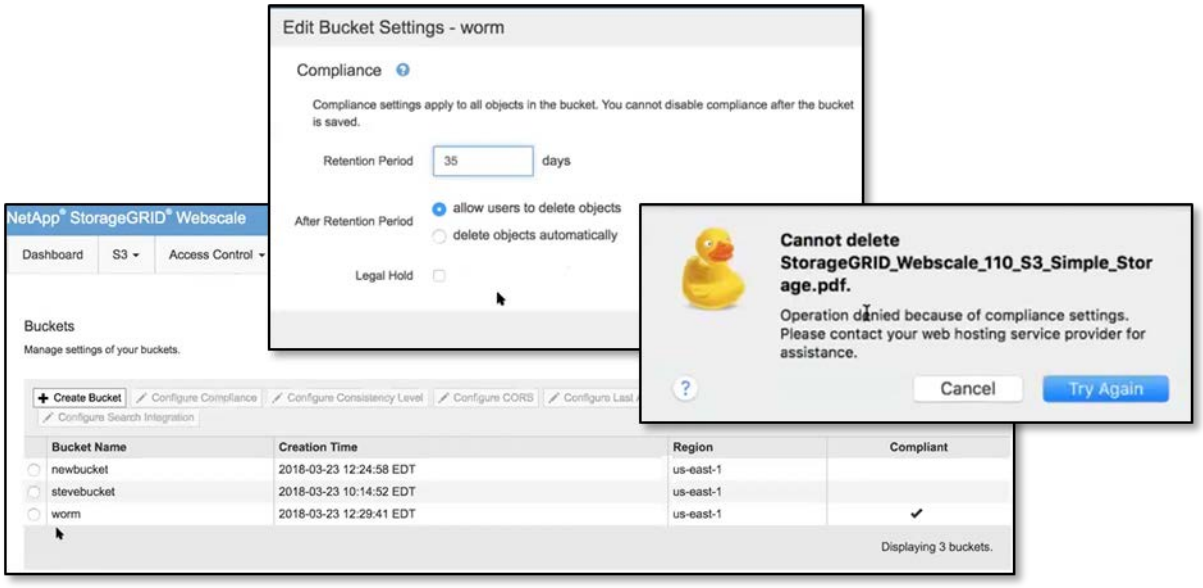
Figure 6. Testing ILM Policies Before Implementation



ESG also noted the new compliance feature included in this release of StorageGRID. The feature allows the storage administrator to create compliance rules that govern how users comply with business standards or regulations when accessing data (e.g., users cannot delete data saved for legal reasons). We used the Tenant Admin UI to validate how the compliance feature works. Under the S3 menu, we selected the Buckets option and saw a list of S3 buckets (shown in Figure 7). We chose the "worm" bucket and clicked on "Configure Compliance." We then edited the length of the Retention Period and chose actions to perform after the Retention Period expires.

We then tested this rule by uploading "StorageGRID_Webscale_110_S3_Simple_Storage.pdf" to the "worm" bucket and attempted to delete the file. We received the error message shown in Figure 7 since we specified that the object cannot be deleted for 35 days.

Figure 7. Enforcing Compliance Rules on a Storage Bucket



Why This Matters

Managing how object data is copied, stored, and deleted becomes cumbersome when data is located across multiple physical locations, across an organization's data centers and the cloud. Organizations need to set up, modify, and apply the object data consistently across all storage especially as the amount of data scales quickly.

ESG validated that a storage administrator can create and apply policies to manage the usage, storage, and retention of data across on-premises storage and S3. We saw how an administrator can choose existing ILM rules to be included in a policy, then order those rules to comply with business needs. We also saw how these policies can be tested before they are activated. Enabling an organization to create and test policies across the StorageGRID easily can decrease daily administration and management time. The organization also can minimize storage costs as policies ensure that storage is used efficiently without having to move, replicate, or delete data at each storage node.

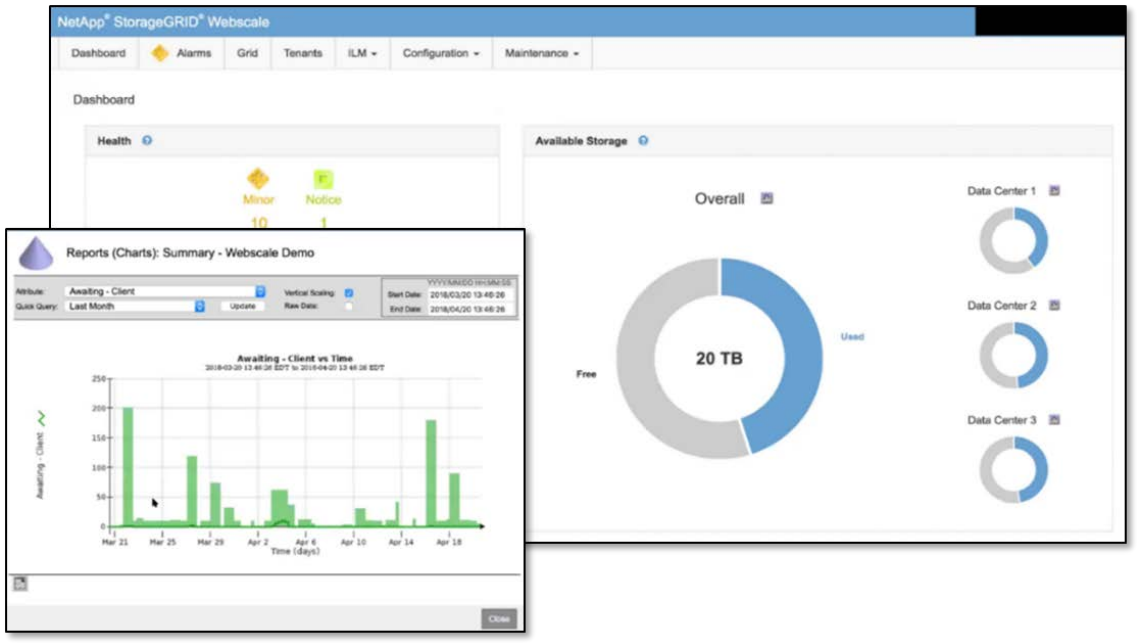
Automation of Information Lifecycle Management

To keep pace with the growth in object data, organizations must manage storage dynamically yet cost-effectively. Administrators need to set up ILM rules that determine how data is stored, copied, and moved automatically when required. These rules must also determine how long and where the data is retained, while ensuring that the data is secure from unauthorized access. StorageGRID allows users to configure ILM policies on per-object level regardless of where the data is ingested, negating the need for constant administrator intervention.

ESG Testing

ESG began with the StorageGRID administrator user interface (UI), which opens to a dashboard (as displayed in Figure 8), showing overall available storage as well as free and used storage. The UI also indicated the overall health of the organization's storage, noting any issues that may require further attention. We observed how an administrator can monitor operations over time with pop-out charts, such as wait times of users accessing objects. The ability to monitor operations can help the administrator to proactively detect and resolve issues in a timely manner, while minimizing issue resolution time. For example, if wait times of users accessing object data were increasing, the administrator could rebalance or add storage or create replicas closer to specific data access points.

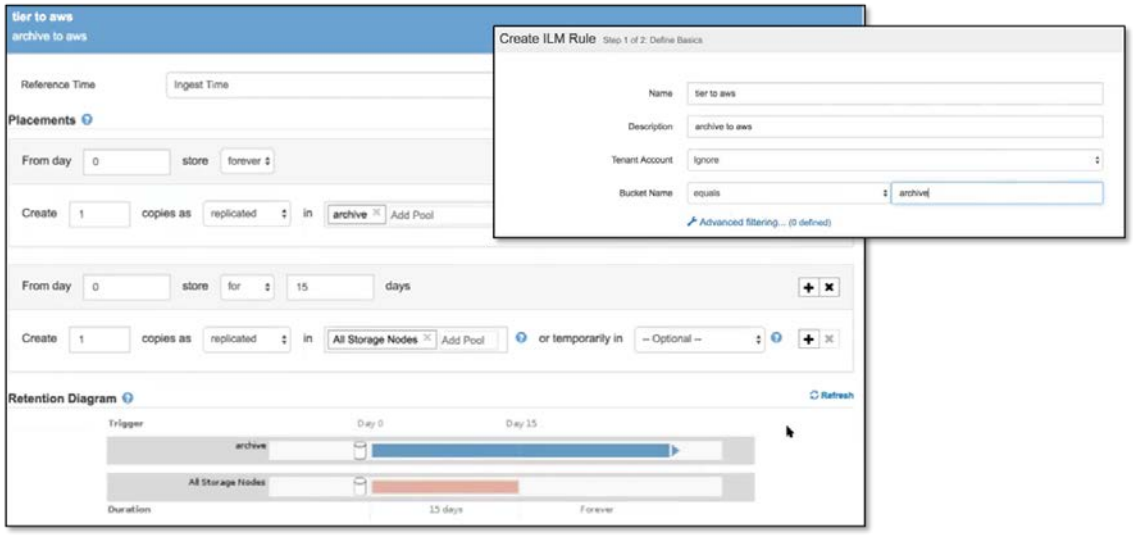
Figure 8. StorageGRID Dashboard with Pop-Out Chart



From the dashboard, ESG then navigated to the ILM menu and chose the *Rules* option. After seeing a list of existing rules, we clicked on the button "Create ILM Rule." We proceeded to create a rule that will tier object data to S3. First, we entered details such as the rule's name and description (as shown in Figure 9). We clicked on the "Save" button and were presented with the screen to input rule parameters. Here, we defined how long the data is stored in the grid, the number of object replicas created, where the replicas reside, and how long they are retained. We noted the Retention Diagram that graphically shows how long the object data and its replicas will be kept.

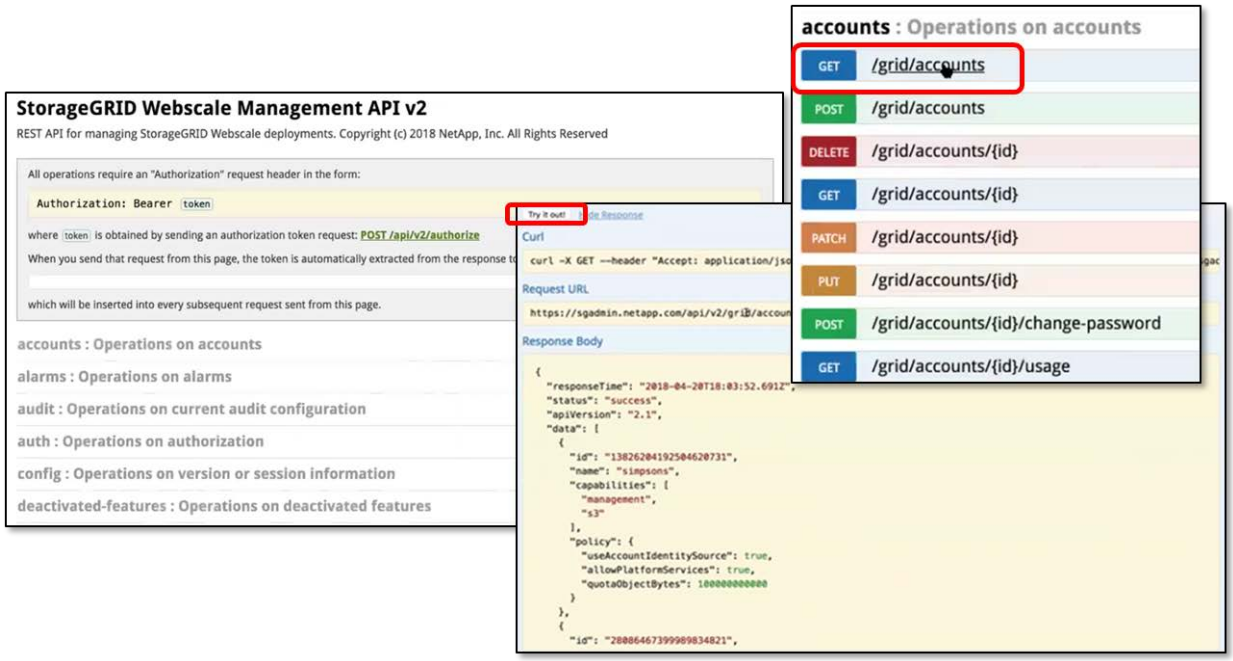
We also defined the type of replicas that are created with this rule. When ingested by StorageGRID, object data can be stored as a full copy at multiple sites to ensure fast access for a predetermined time. After that time expires, the administrator can set up the rule to create erasure-coded copies of that same data. Erasure-coded copies use available storage more efficiently by consuming less storage space. If accessed again, the data and their full copies will be restored in their original physical locations according to the original rule until the predetermined time expires again.

Figure 9. Creating ILM Rule to Tier Object Data to S3



In addition to managing ILM rules from the UI, StorageGRID also provides REST APIs, to help the administrator automate operations. ESG right-clicked on the toolbar to bring up the list of APIs NetApp has provided (as shown in Figure 10). We clicked on "accounts: Operations on accounts" to see its associated APIs. We then clicked on the "get" button to examine a sample code snippet demonstrating how to use the "get" operation. The administrator can modify and test the code snippet. We tested out the code associated with the "get" operation by entering the ID of an object and clicking on the "Try It Out" button in the upper left-hand corner of the code window.

Figure 10. APIs to Automate Execution of ILM Rules



Why This Matters

Organizations must ensure that users can access object data regardless of the data's physical location while considering where to store data to maximize performance access and minimize storage costs. Automating policies that control the location and number of copies increases storage system efficiency and eliminates the need for manual intervention.

ESG validated that StorageGRID helps an organization to automate the execution of ILM rules and common administration taks via API, addressing which data needs to be easily accessible, how to protect that data, and when and where to move data when not accessed frequently. We also observed that an organization can employ APIs to automate ILM rules. Both methods ease the task of managing object data regardless of where data is physically located. The setup time for ILM rules decreases, as StorageGRID applies the rules regardless of where the data is ingested or how much data scales over time. Thus, administration time and costs for ongoing storage management can decrease.

Integration with AWS

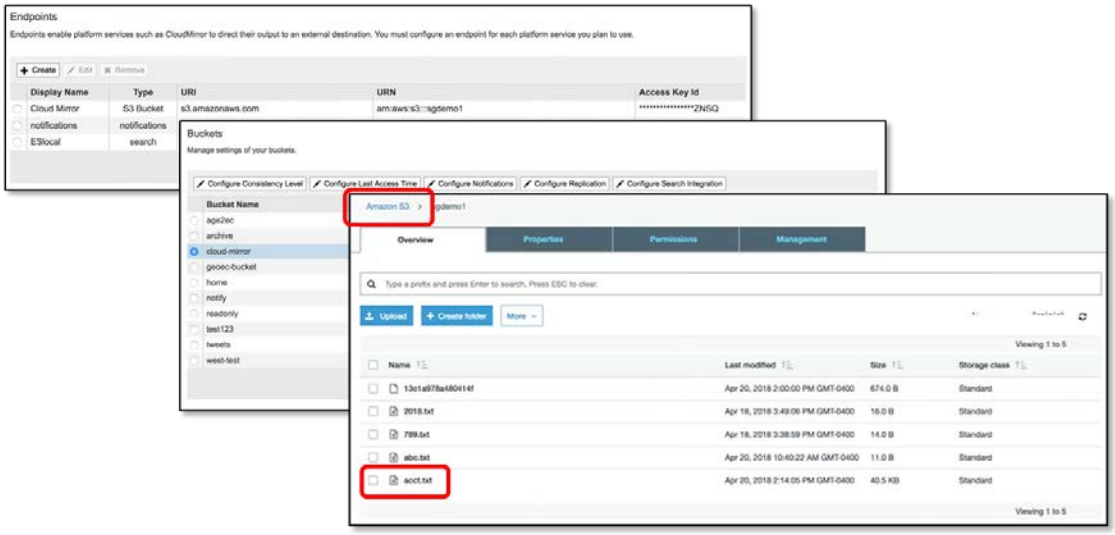
As organizations augment their on-premises storage with S3, they can utilize AWS computing services on-demand, while not storing object data in S3 in the long term. StorageGRID's support for the S3 API enables these organizations to connect applications developed for S3 web services with on-premises object storage. Using StorageGRID's integration with AWS, a storage administrator can transfer data to S3, allowing users to employ select AWS microservices for analysis, while minimizing both cloud storage and transfer costs.

ESG Testing

ESG began by observing how CloudMirror works. First, we selected an endpoint in the Tenant UI of StorageGRID. In this UI, we viewed existing endpoints by going to the **S3** menu and choosing the *Endpoints* option. We then selected the "CloudMirror endpoint" (shown in Figure 11).

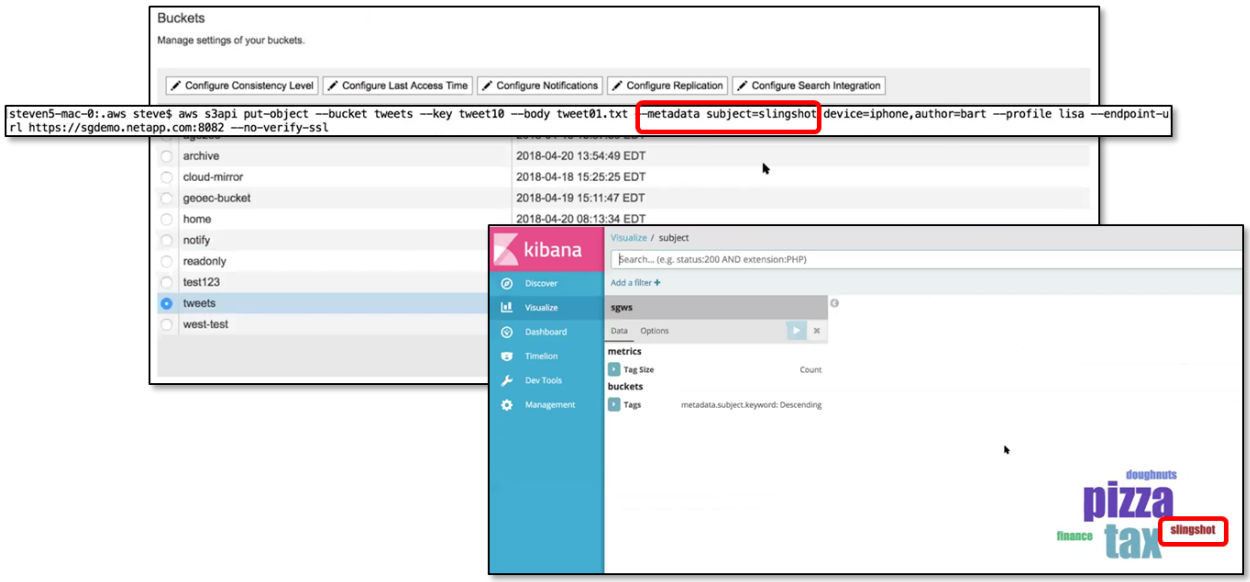
Next, ESG navigated to the ***Buckets*** option under the **S3** menu. We selected "cloud-mirror" as the target bucket for the chosen endpoint. We clicked on the "Configure replication" button and code appeared. We edited the code to define which object data is replicated to the bucket. We noted that the administrator can define specific objects that appear in the endpoint to be replicated into the target S3 bucket. We uploaded the file "acct.txt" to the endpoint and checked the AWS S3 console to verify that the file was replicated. Thus, an organization can minimize both network transfer costs and S3 costs while leveraging AWS compute services for a limited time.

Figure 11. Using CloudMirror to Replicate File to S3 Bucket



ESG also explored the use of Elasticsearch on select data from a StorageGRID S3 bucket. We navigated to the list of endpoints (shown in Figure 11) and selected the "ESlocal" endpoint. Next, we chose the "Buckets" option under the S3 menu and selected the "tweets" bucket. We clicked on the "Configure Search Integration" button to modify the metadata of the object data we wanted to locate via Elasticsearch. We changed the metadata so that the local ES server only searched on object data containing "subject = slingshot" (shown in Figure 12). After uploading an object into the "tweets" bucket, search results were replicated to the endpoint. We used Kibana (an open source visualization plugin for Elasticsearch) to view our search results and found that the local ES server keyed onto object data containing "slingshot." For an organization that wishes to leverage Elasticsearch, this AWS integration will ensure that only data matching the selected object metadata will be replicated, minimizing cloud egress costs.

Figure 12. Leveraging Elasticsearch Integration



Why This Matters

A public cloud provider may provide compute services that an organization wants to employ for specific purposes. Yet, the organization may not want to store its object data in S3 and incur unnecessary storage costs. A viable approach is to establish a pipeline between on-premises and cloud storage that is activated when certain object data is ingested. The organization can automatically transfer data to cloud storage and leverage compute services only when needed. This can minimize both network transfer and cloud storage costs.

ESG validated that an organization can use StorageGRID to set up rules that copy or transfer select object data between on-premises storage and S3 when specific criteria are met. We saw how CloudMirror enables an administrator to copy select data from on-premises storage to S3. We also examined how select data stored in S3 can be transferred to on-premises storage based on preset Elasticsearch parameters, helping to minimize network transfer costs.

The Bigger Truth

With the growth of object data increasing, an organization must consider how to use and manage storage efficiently while fulfilling users' expectations for data access and minimizing overall storage costs. This is especially relevant as organizations' spending on storage will remain flat or decrease in the foreseeable future.[3] The organization must consider how to apply ILM rules and policies in real time to keep pace with data growth. It must be able to create and automate ILM rules and policies to process the myriad of object data that an organization can ingest. The organization must also create policies that will change how data is stored and protected when not accessed regularly, thus minimizing storage costs. Finally, an organization that employs hybrid cloud storage must determine how to leverage a cloud's compute services on object data without incurring excess transfer costs between on-premises and cloud storage.

The NetApp StorageGRID solution enables an organization to define how and where object-based data is stored and archived in a hybrid cloud storage environment. Organizations that use StorageGRID can also save on both capital and operational expenses via efficient storage usage, simplified administration, and increased user productivity. StorageGRID enables the organization to provide user access of object data from any geographic location. The organization can employ StorageGRID to set and automate ILM rules and policies that maintain acceptable levels of access performance while using storage efficiently. Because StorageGRID can help in managing storage efficiently, an organization can minimize its on-premises storage costs. StorageGRID's integration with AWS S3 allows replication and transfer of object data between on-premises storage and S3, minimizing both cloud storage and ingress/egress costs. ESG audited current storage and cloud costs and verified that an organization can potentially save up to 50% in overall storage costs over a five-year period.

ESG validated that an administrator can use StorageGRID to create and automate ILM rules and policies to manage the usage, storage, and retention of data across all geographically dispersed nodes located in an organization's data centers and S3 (when hybrid storage is employed). By enabling the administrator to create and automate ILM rules and policies, StorageGRID can help an organization decrease daily storage management and administration costs as well as increase end-user productivity. Our modeled scenario use-case of a 1,000-user organization revealed that we could achieve operational savings up to $500K when using StorageGRID.

If your organization is facing huge unstructured data growth and needs to efficiently store and easily manage such data while minimizing your capital and operational expenses, especially within a hybrid storage environment, ESG suggests evaluating the NetApp StorageGRID solution.

[3] Ibid.

Enterprise Strategy Group is an IT analyst, research, validation, and strategy firm that provides market intelligence and actionable insight to the global IT community.

www.esg-global.com

contact@esg-global.com

P. 508.482.0188