Lab Validation Brief

NetApp StorageGRID 11.0

Cloud-scale Object-based Storage

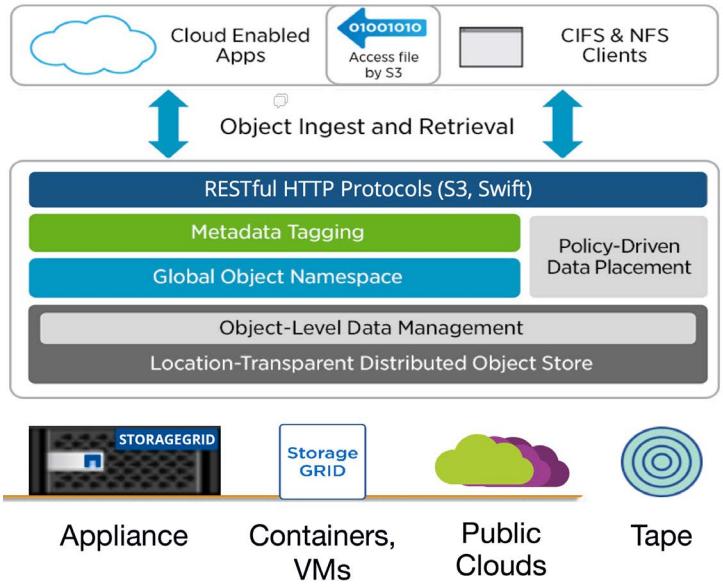
By Ritu Jyoti and Amita Potnis, IDC Infrastructure Team

Sponsored by NetApp | May 2018

IDC ANALYZE THE FUTURE

Lab Validation Brief Executive Summary

IDC validated key features/functionality of NetApp StorageGRID 11.0



Hybrid Cloud Platform Services

- CloudMirror Replication
- Simple Notification Service (SNS) Support
- Search Integration

S3 Compatibility

- New public metrics API
- Updated Object lookup interface
- Updated ILM
  - » Region support / location constraint
  - » AWS S3 Tags
  - » Information Lifecycle Management Policy UI enhancements
- Updated StorageGRID Appliance installer

IDC Opinion

IDC believes that NetApp StorageGRID is an easy to deploy, easy to manage, flexible, highly scalable and efficient software-defined object-based hybrid cloud storage platform. Hybrid Cloud Platform Services enable a tenant account to configure its S3 buckets to leverage external services and extend the functionality of StorageGRID to the public cloud. Intelligent policy-driven data management and integration capabilities of StorageGRID make it a robust solution for use by geo-dispersed enterprises contending with accelerated pace of digitization.

Validation Test Bed

IDC performed the validation at NetApp's labs in Raleigh, North Carolina.The Storage test bed consisted of StorageGRID clusters, and client workstations. The following table provides a summary of the test environment.

Site	Lab	Configuration	Validation specifics, if any
Data Center 1	RTP	• 4 Storage Nodes • 1 Admin Node • 1 Gateway Node	• Virtual nodes with NetApp E-Series storage • 24 GB RAM, 8 vCPU, VMware ESXi, 12.1 TB per node
Data Center 2	RTP	• 4 Storage Nodes	• StorageGRID SG5612 Appliances
Data Center 3	RTP	• 4 Storage Nodes • 1 Archive Node – Cloud Tiering	• Virtual nodes with NetApp FAS storage • 24 GB RAM, 8 vCPU, VMware ESXi, 12.1 TB per node

Notes

All the nodes ran the same version 11.0.1.2 of StorageGRID – although during upgrade they can be running different versions. The client workstations were running Windows 10 and Ubuntu 16.04.

Hybrid Cloud Platform Services

Objective: Demonstrate usage of Hybrid Cloud Data Pipelines with AWS.

Feature Category	Features Tested	Key Findings	Why it Matters
• Hybrid Cloud Platform Services	• CloudMirror Replication	• CloudMirror allows the contents of a StorageGRID bucket to be replicated to an external cloud bucket. Customers can enable CloudMirror replication for any StorageGRID bucket in their tenant account by associating replication configuration XML with the bucket. • S3 API used to configure the service is PUT bucket replication. • The configuration is fairly simple and intuitive. • StorageGRID software automatically and asynchronously replicates specified objects added to the bucket to the destination bucket or buckets named in the configuration XML.	• Empowers customers to use adjacent services (e.g. Amazon EMR, Hive or Spark) available on the Public cloud.

IDC Inference

StorageGRID platform services are simple and powerful to help extend the functionality of StorageGRID. They can help customers implement a hybrid cloud strategy. Any combination of platform services can be configured for a single StorageGRID bucket. For example, one could configure both the CloudMirror service and notifications on a StorageGRID S3 bucket so that one can mirror specific objects to the AWS Simple Storage Service™, while sending a notification about each such object to a third party monitoring application to help one track their AWS expenses.

Hybrid Cloud Platform Services

Objective: Demonstrate usage of Hybrid Cloud Data Pipelines with AWS.

Feature Category	Features Tested	Key Findings	Why it Matters
• Hybrid Cloud Platform Services	• Serverless Compute (Lambda function) enabled by Notifications.	• Bucket event notifications send notifications about actions performed on StorageGRID objects to an external Simple Notification Service (SNS). • Customers can enable event notification on StorageGRID buckets so that StorageGRID software will send notifications about specified events to a destination Amazon Simple Notification Service™ (SNS) • S3 API used to configure the service is PUT bucket notification. • Lambda function can be executed upon ingestion of an object to the bucket on AWS.	• Enable users to exploit the power of Serverless cloud compute Lambda function. • Enable users to avail public cloud services (e.g. deep learning capabilities provided by Amazon Rekognition to easily add video and image analysis to their applications).

IDC Inference

StorageGRID platform services are simple and powerful to help extend the functionality of StorageGRID to the public cloud. They can help customers implement a hybrid cloud strategy. Any combination of platform services can be configured for a single S3 bucket. For example, one could configure bucket event notifications on a StorageGRID bucket, that can send notification to Amazon SNS and invoke Lambda function (e.g. Amazon Rekognition) when a StorageGRID bucket is replicated to AWS.

Hybrid Cloud Platform Services

Objective: Demonstrate usage of Hybrid Cloud Data Pipelines with AWS.

Feature Category	Features Tested	Key Findings	Why it Matters
• Hybrid Cloud Platform Services	• Search Integration with Elasticsearch	• The search integration service sends StorageGRID object metadata to an external Elasticsearch index service. • Customers can enable search integration for StorageGRID buckets. • This feature sends object metadata for objects stored in a StorageGRID bucket to a configured Elasticsearch index. • S3 API used to configure the service is PUT bucket metadata notification (This is a StorageGRID custom S3 API). .	• Empowers customers to use an external search and data analysis service for their object metadata. • Elasticsearch provides users a distributed, multi-tenant-capable full-text search engine with an HTTP web interface and schemafree JSON documents. • Kibana provides visualization capabilities on top of the content indexed on an Elastic-search cluster.

IDC Inference

StorageGRID platform services are simple and powerful to help extend the functionality of StorageGRID. They can help customers implement a hybrid cloud strategy. Any combination of platform services can be configured for a single S3 bucket. For example, one could configure search integration service that enables enrichment of a StorageGRID object metadata with an external Elasticsearch index service.

Improved UI & Operations

Objective: Demonstrate StorageGRID 11.0 simplified Information Lifecycle Management (ILM) functionality – new public metrics API, updated object lookup interface, Updated ILM (location constraints, AWS S3 tagging and policy UI enhancements) and Updated StorageGRID Appliance installer.

Feature Category	Features Tested	Key Findings	Why it Matters
• Improved UI & Operations	• New public metrics API	• StorageGRID supports public metrics API for real-time monitoring and queries data collected internally from node exporter and service exporter (storage).	• Supports real-time reporting and exports enable customized dashboards for better visualization for insights.
	• Updated Object lookup interface	• StorageGRID's updated UI significantly improves Object lookup functionality.	• Ease of use and management
	Updated ILM • Region support / location constraint	• Location Constraint features allows administrators to define ILM rules and policies keeping regions as a requirement. This feature will help organizations define ILM policies around data placement as it relates to regions.	• Assists in compliance and data protection schemes in a granular way that includes region as a criteria.
	Updated ILM • AWS S3 Tags	• StorageGRID supports Key-value pair tagging similar to AWS	• Enables data categorization
	Updated ILM • ILM Policy UI enhancements	• Policies can be defined based on Location Constraint for regional data placement & support.	• Assists in compliance and data protection schemes in a granular way that includes region as a criteria.
	• Updated StorageGRID Appliance installer	• The new installer is significantly simpler to use and makes the task easier to manage and done through a single interface.	• Simplicity & ease of use allows resources to focus their time on high value tasks.

Validated: Hybrid Cloud Platform Services

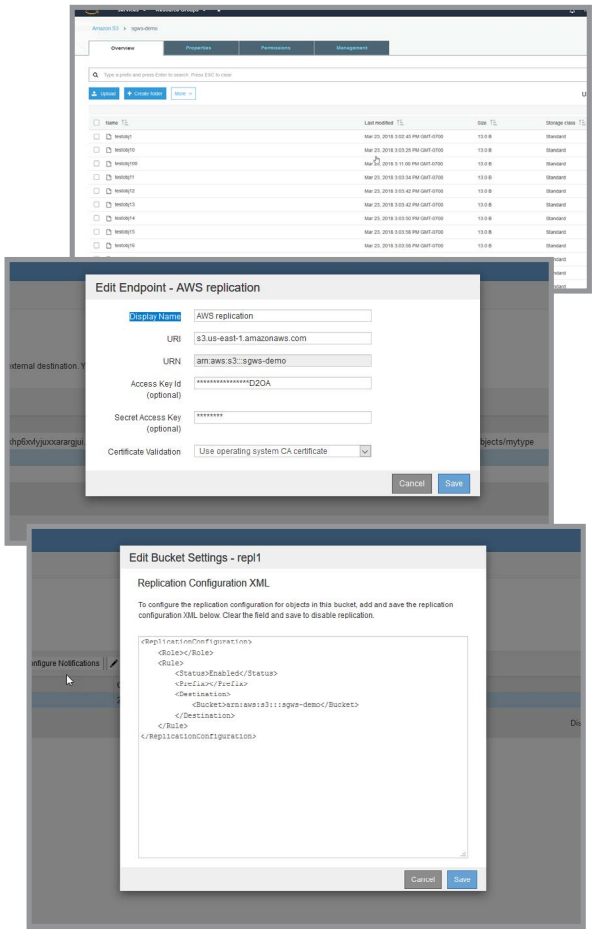
Cross-cloud Replication: Cloud Mirror Replication to AWS

Feature/Validation Summary

Cloud Mirror allows the contents of a StorageGRID bucket to be replicated to an external cloud bucket. StorageGRID 11.0 Hybrid Cloud Platform Services are configurable at the tenant level. Endpoints enable platform services to target an external destination. One needs to configure an endpoint for every platform service they plan to use. Multiple buckets can access the same endpoint.

Validation Process

- To validate CloudMirror Replication to AWS S3, IDC executed the following tasks:
- Logged into Amazon S3 and created a new bucket "sgws-demo".
- Logged into the Grid Management Interface (GMI) as the root of a tenant.
- Configured the endpoint – "AWS Replication".
  - Specified the Universal Resource Identifier(URI) as "s3.us-east-1.amazonaws.com" and Universal Resource Name (URN) as "sgws-demo"
  - Entered the access key id and secret access key for the AWS S3 Bucket.
- To configure the replication configuration for objects in the bucket "repl1" on StorageGRID, created and saved the "Replication Configuration XML"
- Uploaded an image "glasses.jpg" to "repl1" bucket on StorageGRID.
- CloudMirror replicated "glasses.jpg" to AWS.
- Using the Amazon S3 interface, we verified that as per the endpoint setting, "glasses.jpg" was uploaded to "sgws-demo" bucket on AWS.



Validated: Hybrid Cloud Platform Services

Bucket event Notification: Serverless Cloud Compute – Lambda enabled by Notifications

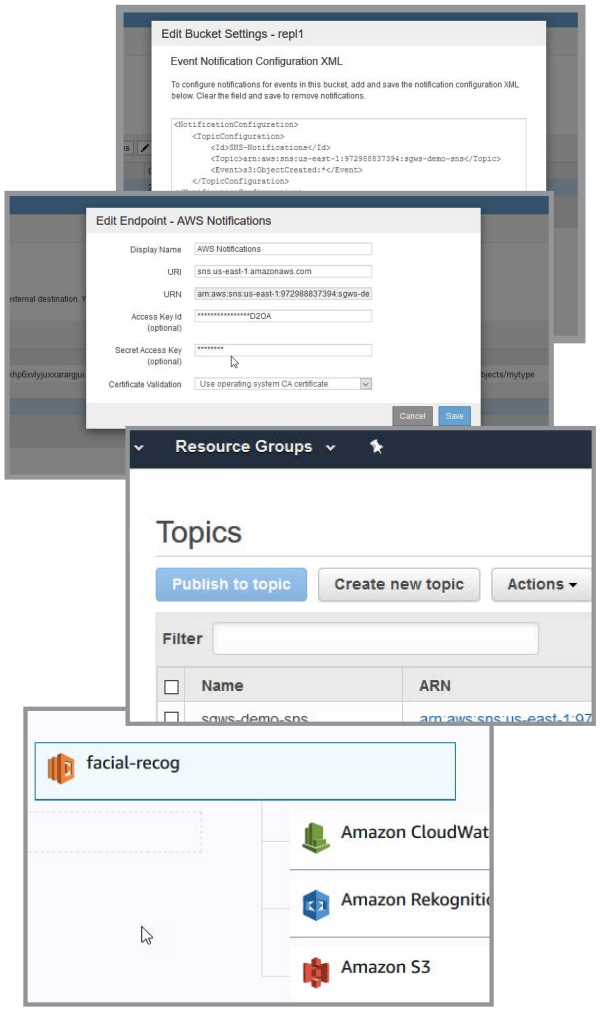
Feature/Validation Summary

Bucket event notifications send notifications about actions performed on objects to an external Simple Notification Service (SNS). StorageGRID 11.0 Hybrid Cloud Platform Services are configurable at the tenant level. Endpoints enable platform services to target an external destination. One needs to configure an endpoint for every platform service they plan to use. Multiple buckets can access the same endpoint.

Validation Process

To validate the execution of Amazon's Serverless compute Lambda function and Simple Notification Service (SNS) using bucket event notifications, IDC executed the following tasks:

- Configured the endpoint – "AWS Notification".
  - Specified the Universal Resource Identifier(URI) as "sns.us-east-1.amazonaws.com" and Universal Resource Name (URN) as "sgws-demo-sns"
  - Entered the access key ID and secret access key for the AWS SNS service.
- Next, using AWS SNS interface, created a topic – "sgws-demo-sns"
- Then, two subscriptions were created for that topic to support email and sms notifications - "email-json" and "sms"
- Next a workflow event was tied to "sgws_demo" bucket on AWS, to initiate the lambda function(facial-recog) which invoked Amazon Rekognition service as soon as an object was ingested to it.
- On upload of "glasses.jpg" to StorageGRID in the CloudMirror verification process, we noted that the lambda function was executed and the facial recognition service in AWS added 4 new pieces of metadata describing the image
- It also triggered an sms and email notification, which was verified in the test inbox.



Validated: Hybrid Cloud Platform Services

Search Integration: Elasticsearch – Kibana Visualization

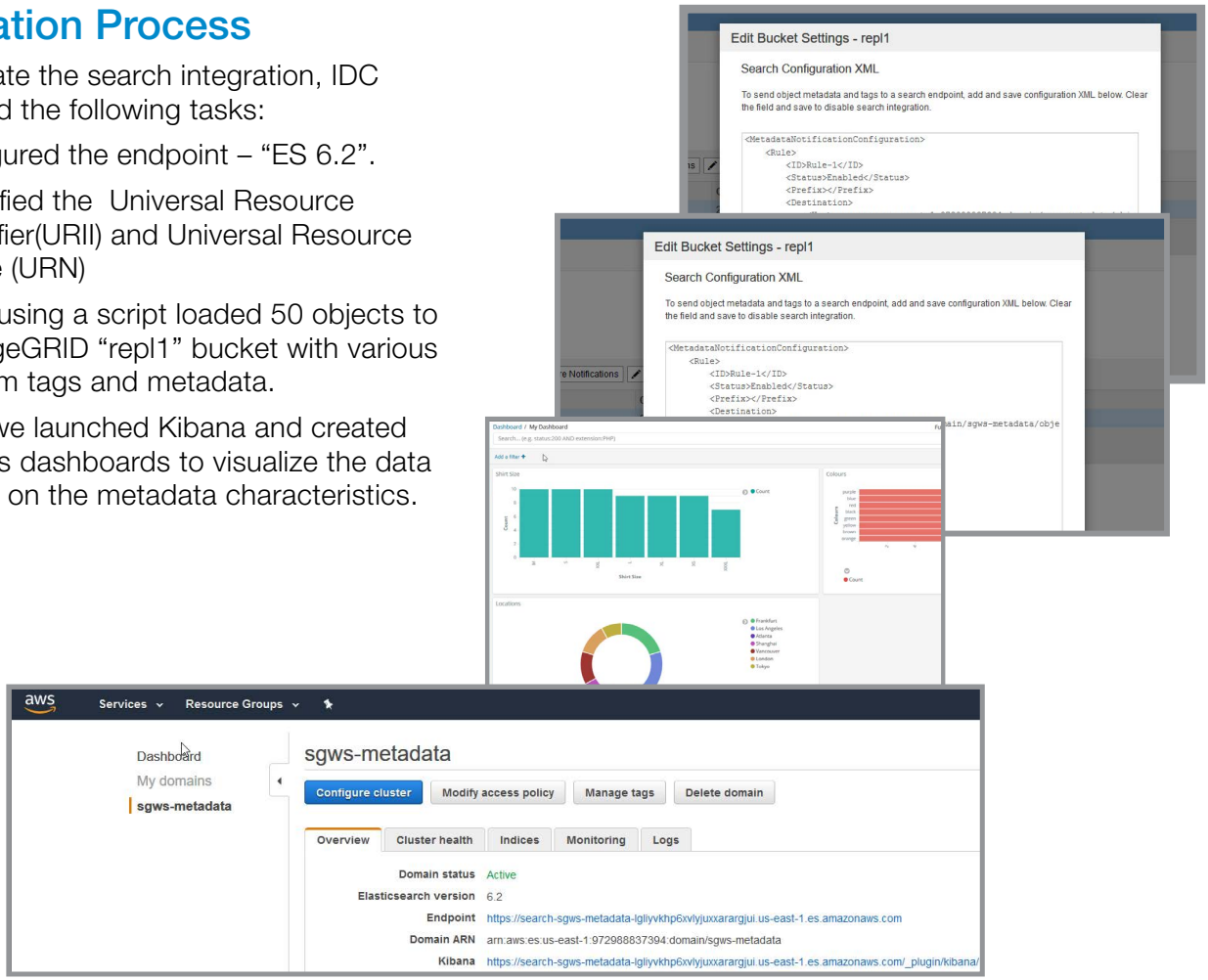
Feature/Validation Summary

The search integration service sends S3 object metadata to an external Elasticsearch index service. StorageGRID 11.0 Hybrid Cloud Platform Services are configurable at the tenant level. Endpoints enable platform services to an external destination. One needs to configure an endpoint for every platform service they plan to use. Multiple buckets can access the same endpoint.

Validation Process

To validate the search integration, IDC executed the following tasks:

- Configured the endpoint – "ES 6.2".
  - Specified the Universal Resource Identifier(URII) and Universal Resource Name (URN)
- Next, using a script loaded 50 objects to StorageGRID "repl1" bucket with various random tags and metadata.
- Next we launched Kibana and created various dashboards to visualize the data based on the metadata characteristics.



Validated: New Public Metrics

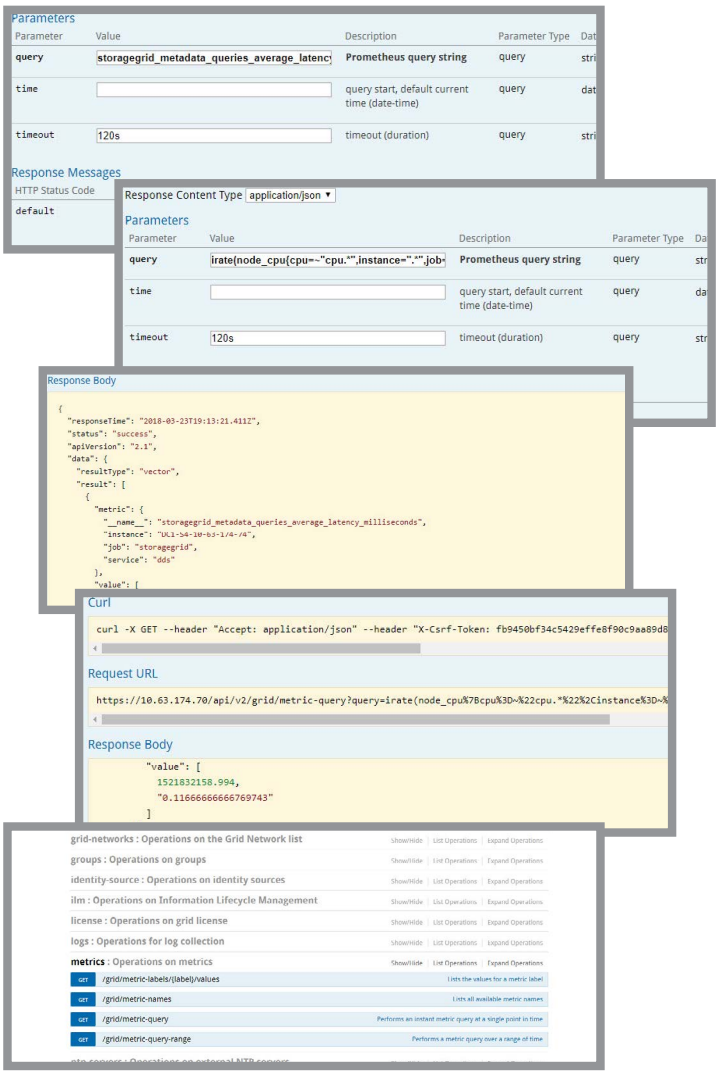
Feature/Validation Summary

StorageGRID 11.0. introduces a new Public Metrics API for real-time monitoring. , and querying data collected internally. This new metrics reporting is exposed via the Management API, and tracks both system level metrics from the operating system/node as well as Services level metrics from various StorageGRID services including ILM, metadata repository, etc. The Metrics Engine scrapes data at the OS/Node level from the Node Exporter and service metrics from StorageGRID Exporter. The Management API exposes the metrics and can integrate with external apps for customized dashboards to extract value.

Validation Process

To validate Public Metrics API, IDC executed the following tasks:

- Logged into the Grid Management Interface (GMI) as the root of the tenant.
- Access Management APIs classified under Help API Docs (Swagger Interface)
- Next, we query the node level metrics and the StorageGRID Exporter level metrics
- For Node Level metrics
  - IDC created a generalized query to show node level CPU utilization of all instances across nodes
  - Click “Try it Out!”
  - Reponses body populates the values of each of the CPUs running across all nodes which can be exported to integrate with external apps for customized dashboards
- For Services Level metrics - Metadata Latency
  - IDC created a generalized query to show metadata query latency level across StorageGRID nodes
  - Click “Try it Out!”
  - Response body populated the values of metadata latency across instances



Validated: Updated Object Metadata Lookup

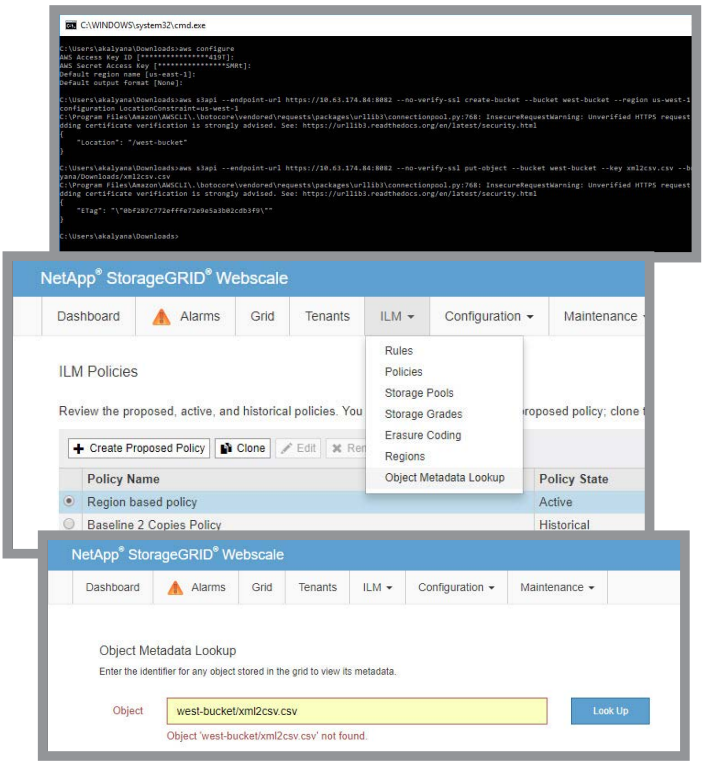
Feature/Validation Summary

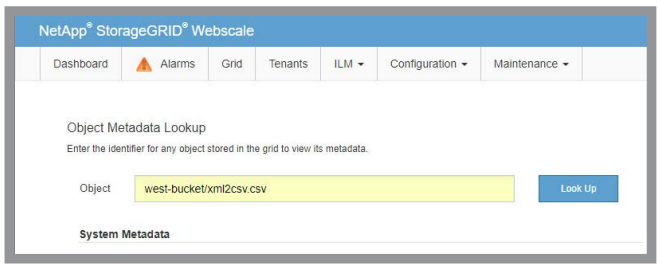
StorageGRID 11.0 has significantly improved Object Metadata lookup functionality making user experience better. In earlier versions of the software the user carrying out this task had to use multiple screens to search for objects which has been simplified in this version.

Validation Process

To validate improved & integrated Object Metadata Lookup, IDC executed the following tasks:

- Logged into the Grid Management Interface (GMI) as the root of the tenant.
  - Selected on ILM Object Metadata Lookup
- IDC first searched for an non-existent object“xml2csv.cvs”
  - StorageGRID UI returned an error
- IDC then created a bucket (east-bucket) with location constraints using AWS command line interface and created an object called “xml2. csv.cvs”
  - IDC searched for the object ““xml2.csv.cvs” again and this time the object was found
  - StorageGRID UI detailed metadata information along with replicated copy information for the object
- IDC then created a bucket (“west-bucket”) in us-west-1 region using location constraint and ingested object “xml2.csv.csv”
  - IDC searched for the object ““xml2.csv.cvs” again and this time the object was found
  - StorageGRID UI detailed metadata information along with replicated copy information for the object





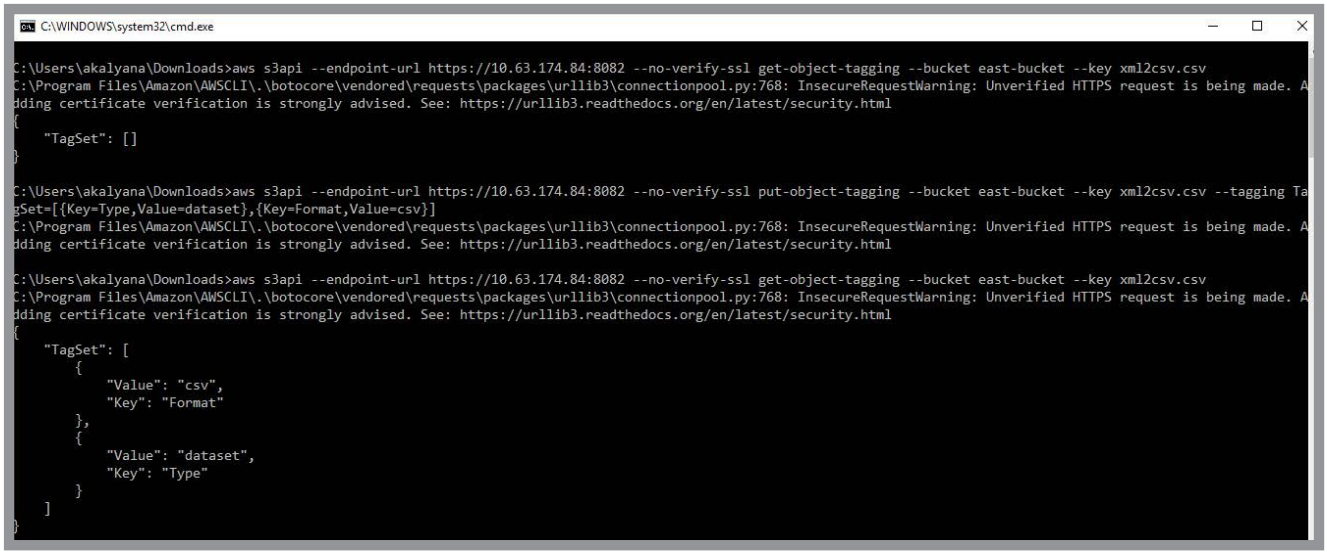
Validated: Updated ILM – AWS S3 tags

Feature/Validation Summary

StorageGRID 11.0 now supports AWS S3 object tagging. These tags can be used to categorize and manage customer data, set up ILM policies to defined data placement rules and analyze using Elasticsearch.

Validation Process

- To validate AWS S3 tags, IDC executed the following tasks:
- Using AWS CLI, IDC ran a get-object query to check the tag set on the object "xml2.csv.csv" in the east-bucket
- IDC found no tags associated with this object
- IDC then ran a put-object query for tagging for the object stated above
- IDC re-ran the get-object query to check the tag set on the same object
- The script returned Value = CSV for Key = Format and Value = Database for Key = Type



Validated: Updated ILM - Policy UI Enhancements

Search Integration: Elasticsearch – Kibana Visualization

Feature/Validation Summary

StorageGRID 11.0 introduced UI enhancements by specifically incorporating the Simulate feature as part of the UI. The Simulate feature allows users to test ILM policies before implementing it and makes the task easier for the user/administrator as it is completely integrated in the StorageGRID UI. StorageGRID 11.0 also introduces support for S3 Regions which allows ILM policies to be set based on the location constraint metadata.

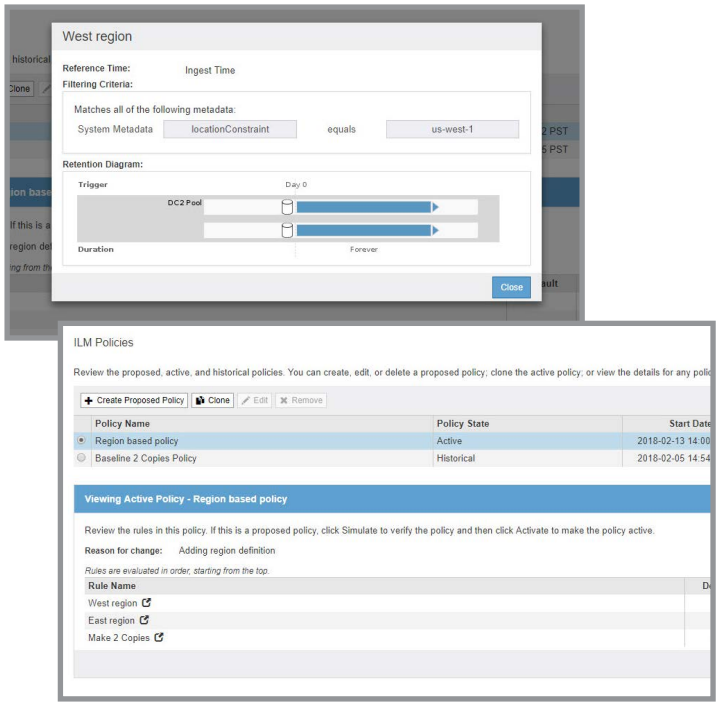
Validation Process

To validate the ILM Policy UI enhancements and S3 Region support, IDC executed the following tasks:

- Logged into the Grid Management Interface (GMI) as the root of the tenant.
- Selected on ILM Policies
- IDC created a policy called “Region based policy” specifying placement of the objects based on location constraint, the policy included the following rules: West Region, East Region and Make 2 copies

– IDC tested the object placement for the “Region based policy” using the ILM simulator in the policy editor

– The ILM simulator allows a user to test a policy before it is made active.



Validated: Updated StorageGRID Installer

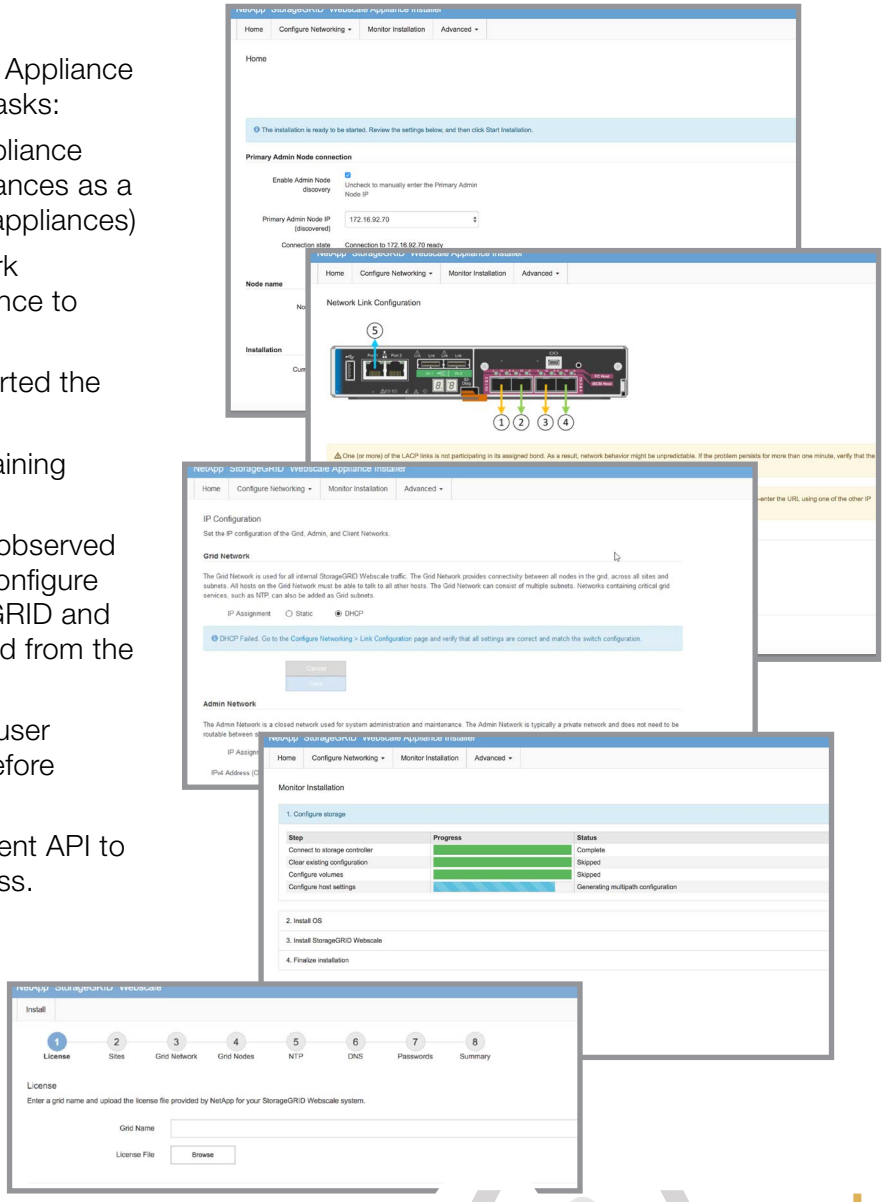
Feature/Validation Summary

StorageGRID 11.0 now utilizes a new and improved installer. The UI is now standardized across the management interface and the initial deployment UI of appliance. The new installer UI is intended for ease of use and simplification of deploying StorageGRID

Validation Process

To validate the updated StorageGRID Appliance Installer, IDC executed the following tasks:

- Logged into each StorageGRID Appliance Installer and installed the new appliances as a part of the StorageGRID cluster (4 appliances)
- IDC configured the links and network connections and pointed the appliance to primary admin node
- After naming the appliance, IDC started the installation
- IDC repeated the steps for the remaining three appliances
- As the installation progressed, IDC observed that all four stages of installation: Configure Storage, Install OS, Install StorageGRID and Finalize Installation can be monitored from the same UI for ease of use
- All nodes were discovered and the user was asked to approve the nodes before proceeding
- Users could also use the management API to fully automate the installation process.



Conclusion: StorageGRID 11.0 Key Differentiators

Feature Category	Features	Key Differentiators
Hybrid Cloud Platform Services	• CloudMirror Replication • Simple Notification Service (SNS) Support • Search Integration	• The target location for platform services is typically external to StorageGRID deployment. As such, Hybrid Cloud Platform Services give customers the power and flexibility that comes from using external storage resources such as public cloud or hyperscalers, notification services, and search or analysis services for their data. • Any combination of platform services can be configured for a single S3 bucket. • Platform Services could be used for a breadth of use cases and across location. Example Use Cases: – Customers could configure CloudMirror Replication to mirror specific customer records placed in one bucket to a sister organization that owns its own instance of StorageGRID, essentially creating a cross-organizational hybrid grid. – Customers could configure alerts to be sent to administrators about each object added to a bucket, where the objects represent log files associated with a critical system event. – Customers could configure their buckets to send S3 object metadata to a remote Elasticsearch service. They could then use Elasticsearch to perform searches across buckets, and perform sophisticated analyses of patterns present in their object metadata.

Conclusion: StorageGRID 11.0 Key Differentiators

Feature Category	Features	Key Differentiators
Improved User Interface and Operations	• New public metrics API • Updated Object lookup interface • Region support / location constraint • AWS S3 Tags • ILM Policy UI enhancements • Updated StorageGRID Appliance installer	• The ability to run real-time analytics is key to gain insights into existing datasets, strategize for the future and drive revenue. StorageGRID's new public metrics API enables users to extract value from metrics collected at the node and service level for better insights. • In general ease of use when managing and maintaining storage infrastructure is important to all users. StorageGRID's updated object lookup interface allows storage administrators to focus on higher value tasks. • Updated ILM features gives users the added leverage of controlled data placement and access policies by region using location constrain. Such features are beneficial from a compliance and governance standpoint allowing organizations to gain better control of their datasets. With these changes, organizations are now equipped with tools to carve out detailed lifecycle policies and tagging that enables efficient analytics. • In general ease of use when managing and maintaining storage infrastructure is important to all users. StorageGRID's updated appliance installer allows storage administrators to focus on higher value tasks.

Essential Guidance: Advice for Buyers

As businesses become data driven to survive in the new economy, they will seek more data sources, collect more data, and look to analyze and store this data in a decentralized manner. In many cases, they will look to perform real-time analytics on this data as it is generated and where it gets captured. Many others will seek to create on-demand OpEx-driven cloud environments for internal and external consumption. Non-traditional use cases, especially for highly scalable and decentralized semi-structured (machine generated) and unstructured data storage, will require nontraditional storage solutions like OBS.

OBS is an innovative approach to storage, and its procurement needs to be preceded by careful planning. Unlike traditional SAN or NAS arrays, OBS solutions are not built the same way nor are they built to suit all use cases and workloads equally.

IDC believes that object-storage platforms will become the core part of enterprise storage infrastructure strategy. Object-based storage platforms will enable existing workloads, as well as be future proof to seamlessly support new applications and workloads.

IDC concludes that the NetApp StorageGRID 11.0 system possesses the enterprise- grade attributes such as client connectivity via known protocols (NFS/CIFS, Cloud protocols like S3 and Swift), massive scale, global namespace across sites, comprehensive and automated ILM/Dynamic Policy management, flexible data protection methods (Replication and Erasure Coding), fault- tolerant architecture, security, audit capabilities, and integration with cloud and archiving solutions.

IDC Validation Methodology

This Validation InfoBrief provides a summary of an extensive validation process performed by IDC in collaboration with the supplier's teams. IDC relied on the supplier's equipment, facilities and their configuration to perform this validation. All of the tests were conducted during the presence of one or more IDC Analysts.

This InfoBrief is meant to provide a quick set of inferences and insights for IT professionals and business decision makers seeking to perform further due diligence on the capabilities of the product and/or services that have been validated in this InfoBrief. However, the goal of this InfoBrief is not to supply detailed hands-on test plans and validation jobs. It is not meant to replace the evaluation process that most businesses will conduct before making any decision to purchase the product and/or services.

It is for this reason that this InfoBrief is not designed to be an all-inclusive document on all the capabilities of the product, but rather as a concise document that highlights features/functions of products, their relative performance with respect to a traditional environment and the value these features bring to businesses looking to solving certain problems using the evaluated product.

Finally, even though this InfoBrief is a sponsored document, it is not meant to be an IDC endorsement of the product, service or the sponsoring supplier. IDC's opinions are its own and not influenced by the production of this document.