EBOOK

The Big Book of S3 Data Protection

Protecting and restoring data in Amazon S3

EBOOK

Commvault

Table of Contents

CONTENTS

Commvault

Introduction

Amazon S3 (Simple Storage Service) is a popular choice for storing data due to its low cost and high reliability. It is designed to be easily integrated with other AWS services, making it a versatile tool for a variety of use cases. Amazon S3 started off as a datastore for noncritical data, but today, it's used for business-critical and mission-critical applications. Amazon S3 is used as the storage layer for websites, mobile apps, big data analytics, and dev/test environments by millions of companies.

Because Amazon S3 started as a storage solution that was used primarily for non-critical data, many people believed it was therefore unnecessary to backup. This was no doubt reinforced by S3's 11 9's of durability, and a pervasive misconception that this durability was essentially the same as data protection. Today, data in Amazon S3 looks very different from the old days of a cheap dumping ground for non-critical data. As companies have uncovered the power of unstructured data, S3 increasingly houses business-critical and sensitive data, adding focus and scrutiny to this service.

There are numerous ways data in S3 can be compromised. Given the ever-expanding nature of data, admins perform frequent cleanup and cost-optimization sweeps. Sometimes the wrong data is accidentally deleted, leading to the harsh realization that data that was not backed up is not recoverable. There have been several instances of employees with access to certain buckets accessing sensitive customer data or maliciously deleting important data. And developers have accidentally publicly published private encryption keys on GitHub, potentially exposing them to bad actors.

Data in S3 has a unique challenge when it comes to scale. Individual objects can range in size from a few KB to several TB, and buckets can contain a handful of objects or tens of billions. Multiply that by hundreds to thousands of buckets across multiple accounts, and things start to get complicated. Protecting all the right data while avoiding the excess costs of protecting the wrong data could be difficult. And for enterprises with buckets in the tens of billions of objects, finding a solution to reliably protect and restore at that scale may seem impossible.

This eBook will lead you through all the ways Clumio helps you overcome the many challenges of protecting data in Amazon S3. Read on for section-by-section explanations on how to protect, automate, restore, and secure your S3 data.

WHY PROTECT AMAZON S3 DATA: BEST PRACTICES

There are several important details to consider when thinking about protecting your data in Amazon S3. It's important to understand the uses of backup vs. replication, data layout, versioning, and restoring.

Data durability does not eliminate the need for backup

Amazon's S3's famous 11 9's of durability does not protect your data from deletions or compromises. Moreover, even things like object-level versioning do not provide full protection, since a user can (unintentionally or maliciously) delete versioned objects.

Replication and backup have different uses

The purpose of replication is to keep a mirror copy of the primary bucket in a second location in order to fail over in case the primary data becomes unavailable. This is usually because of data loss or service outage.

In general, replication:

- Mirrors the data structure/layout of the primary bucket into the secondary bucket.
- Copies every single operation made to the primary bucket to the secondary bucket immediately. The secondary bucket is a hot/active copy and is consistently updated.
- Requires versioning to be enabled, which creates unwanted copies of objects and significantly increases costs.

Backup is similar to replication in the sense that a second copy is maintained, however backup is not directly accessed by the application. This secondary backup data is used to restore back to a known good copy.

- Backup is able to change the data structure/layout to optimize for cost efficiency, making it cheaper to retain the copies. Your application will not directly access the backup data, so it doesn't need to be structured the same way as your primary.
- Backup does not constantly copy every single update made to an object. Instead, backup copies your data at discrete recovery points, allowing you to perform PITR (Point in Time Recovery) while optimally consuming storage for backups.

Data layout and efficiency

When considering replication or backup, decisions about the way data is structured and stored will impact the copied data.

- Creating a lot of small objects is not ideal for S3 in terms of cost.
- Large numbers of small objects results in high PUT cost.
- Except for S3-Standard tier, small objects are penalized by either a minimum size requirement (e.g., S3 Infrequent Access) or fixed metadata per object (e.g., Glacier).
- The application drives the structure of the primary S3 bucket and the replica bucket (which mirrors the primary).
- Conversely, for backup it is possible to organize the data and optimize cost; for example merge several small objects into one large object and reduce multiple PUTs into a single operation.

Until very recently, bucket replication was the only available data protection option for Amazon S3, so understandably, many organizations still use replication today. Considering that using replication in place of backup often results in inefficiencies and higher costs, organizations should take a look at their data protection needs and make sure they are using the right tool for the job. The decision to have continuous recovery points or discrete recovery points should be based on your organizational requirements.

The restore process

It is very easy to forget about the recovery process, but the point of backup is to restore your data when something goes wrong. You want data restoration to be quick to meet your Recovery Time Objectives (RTO) and maintain continuous business.

- It is easy to set up S3 replication but it is another story if you want to restore the data back.
- Depending on where and how you stored the replica bucket, you will have to:
- Select the right version (if restoring to a specific version/point-in-time)
- Retrieve objects (if stored in Glacier)
- Copy objects to the destination
- Make sure that the retrieved objects are deleted

While this is only a few steps, you may need to do this with millions or billions of objects. You will need to track progress, handle or retry any failures and execute the restores in the most efficient manner. You want the restore process to be as quick, easy and painless as possible. Configuring AWS S3 replication is easy, but restoring from it is a big undertaking.

Object versioning

Many developers use object versioning in Amazon S3 as a fallback option should something go wrong with an object, like an unintentional change or deletion. While versioning can be helpful in a number of situations, it comes with inherent drawbacks. For example, for buckets with high change rates, versioning can quickly become expensive. Also, since versions are stored in the same account and region, they are not protected from malicious attacks. Additionally, object versioning only captures PUT changes. Certain changes, like changes to tags, do not generate a new version.

AWS Backup

AWS Backup offers data protection for S3, but this solution has limitations. AWS Backup for S3 requires versioning to be enabled, and operates at the bucket level only – you have to back up the entire bucket, even if it contains only a few objects you want to back up, which also adds cost. Finally, AWS Backup protects buckets containing a maximum of 3 billion objects. This falls short of many large enterprises' requirements.

Even though we have a number of native data protection features built into S3, deletions are customer driven, or they can be machine driven, and we don't know the intent of a customer request to delete an object. Is that a malicious delete? Is that an accidental delete, or is that a correct delete? We have to honor them regardless. So that's why we're starting to see the real need to protect this data.

Amazon S3
Principal Product Manager

Experience Clumio Protect for Amazon S3

Clumio Protect for Amazon S3 helps solve data protection challenges, delivering SaaS simplicity with intuitive protection, flexible recovery, infinite scale, and industry-leading performance. These features will be discussed in more detail throughout this ebook. For now, we'll consider Clumio optimizations that impact cost, performance, and scale. Clumio performs numerous storage optimizations in order to backup data in the most cost-efficient way possible.

In Clumio Protect for Amazon S3, versioning is optional, which is helpful in optimizing retention cost. Another way Clumio helps you save even more is to provide you with the right framework to tell us what you want to protect and what you don't. For instance:

- Do you care about all versions or specific recovery points?
- Do you want to back up specific prefixes or exclude others?
- Do you want to back up only specific types of objects or storage classes?
- While the amount of production data in S3 continues to increase, it is still used to store a variety of data that is not critical to back up. So it is important to control costs by selectively backing up your S3 data.

We implemented workflows to optimize efficiency when restoring buckets to enable the fastest possible recovery time. Despite these optimizations, restoring millions of objects can still be time consuming, so Clumio provides global search-and-restore capability. You can search and restore objects based on their:

- Prefixes
- Sizes
- Classes
- Objects
- Tags

Some large enterprises deal with data on a truly enormous scale, with buckets containing many petabytes to exabytes of data, and tens of billions of objects. Clumio is the only S3 data protection solution that can handle protecting and restoring data at that scale. With Clumio, you can protect, search and restore buckets containing:

- Up to 80 billion objects
- Up to 1 exabyte of data

By quickly recovering specific objects during a disaster or cloud outage, Clumio reduces your RTO, helping you meet your business continuity objectives.

In summary

- If you have critical data in Amazon S3, you need to protect it. 11 9's of durability means your data is very safe from server issues, but this does not protect you from other types of data loss.
- Replication is not backup, just like backup is not replication. Each has a distinct purpose.
- When considering replication and backup, think about the restore process and think in the context of hundreds of millions of objects.
- Building and maintaining everything yourself is possible, but probably isn't your most efficient option.

SETTING UP DATA PROTECTION FOR AMAZON S3

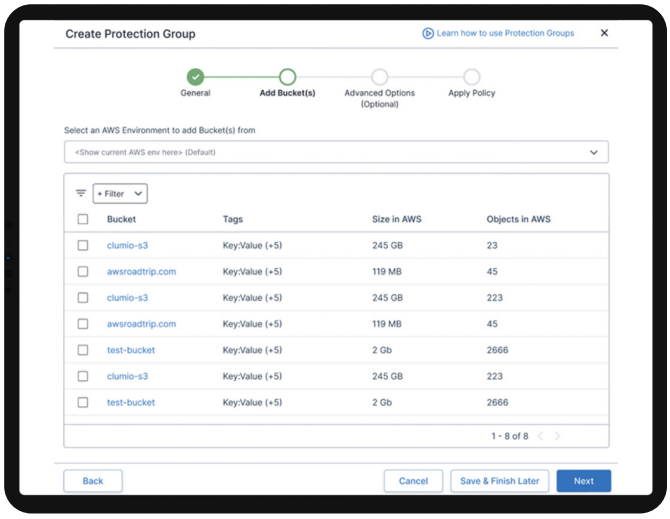
Using Protection Groups

Customers face several challenges in order to successfully differentiate between critical and non-critical data and protect only the critical data. This data classification challenge is solved by Clumio's innovative Protection Groups feature, which can be used to classify and protect critical data while producing tremendous cost savings. Protection Groups provides an abstraction layer to manage buckets and prefixes across all your AWS accounts. It provides a mechanism to classify data across buckets in all of your different AWS accounts to help protect critical data in accordance with business requirements.

Configuring Protection Groups is a simple 3-step process

Step 1:

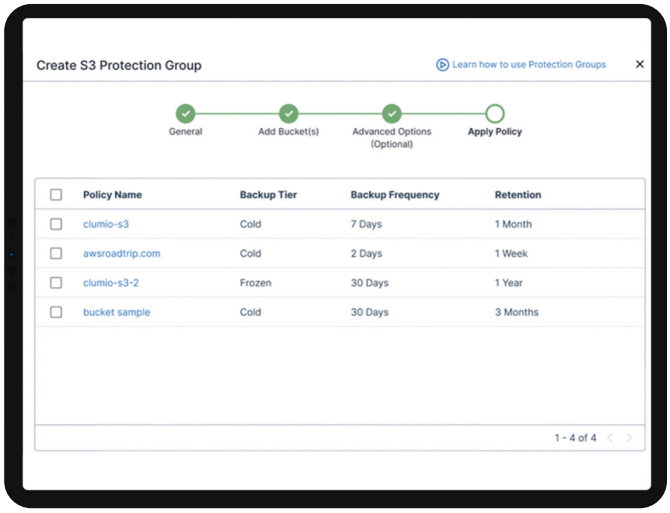
After giving the Protection Group an intuitive name, decide what buckets to add to it. These buckets could belong to either a specific AWS account or could be across all your AWS accounts. You can also specify Tags to be included, so that current and future buckets with those Tags are added to the protection group automatically.



Step 2:

Decide whether the entire bucket or only a subset is added into the Protection Group. You can use 3 different criteria to select which data is protected:

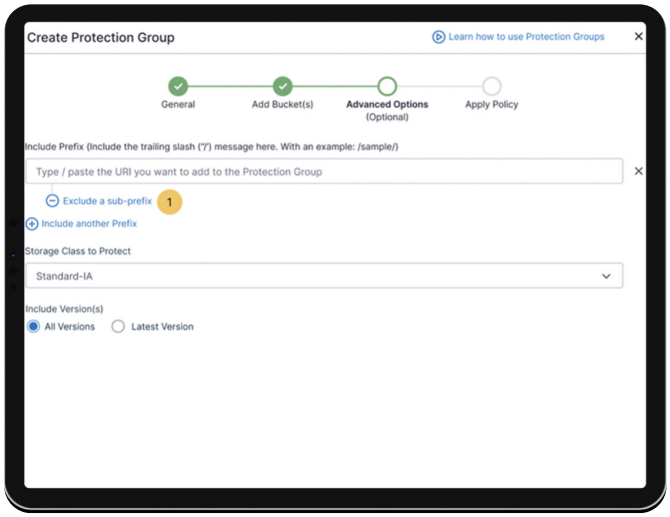
- Prefix: You can configure the Protection Group to include or exclude specific prefixes depending on what you want to protect. For example, several customers dump their DB logs into a specific prefix and want that data to be protected. They can configure /dblogs/ to protect all objects sitting inside that prefix to be protected. Conversely, if that data should not be protected, they can choose to exclude that prefix.
- Storage Class: You can configure what objects to backup depending on their Storage Class. For example, you can back up objects in Standard and Infrequent Access only, while not protecting objects in Glacier. This helps reduce time and cost significantly as objects stored in colder storage require time to unthaw and are expensive to pull out.
- Version: You can configure whether to protect all versions or just the latest versions of selected objects.



Step 3:

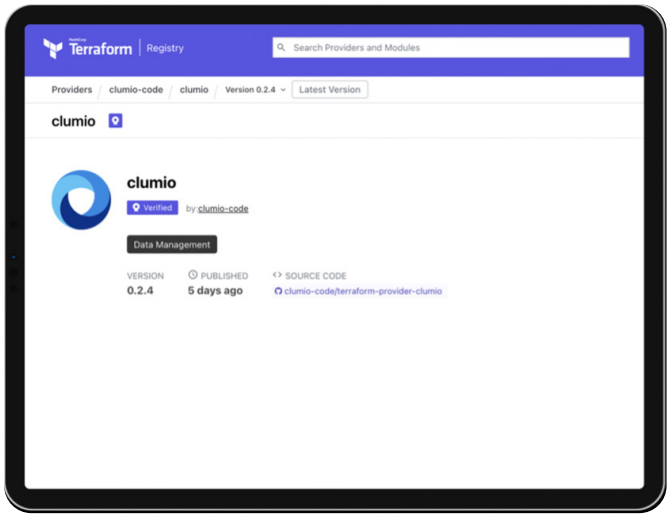
Apply a policy to the Protection Group. This applies backup tier, frequency, and retention to help make sure that all grouped data is protected in line with business requirements.

And that's it! Your Amazon S3 data is protected in an air gap environment, safeguarding against accidental or malicious data loss.



AUTOMATING END-TO-END DATA PROTECTION

A fundamental principle in recent years for IT and DevOps has been to consider infrastructure as code. Similar to how application code has defined syntax and formatting as well as a reproducible binary, a company's infrastructure should be managed and provisioned in analogous fashion. One of the most well-known infrastructure as code tools is Terraform which provides key syntax and structure to help companies minimize environment drift and automate end-to-end reproducibility of cloud-based environments.



To help our customers achieve these benefits, Clumio created its own Terraform Provider and became a Hashicorp (the makers of Terraform) Technology Partner. Clumio's Provider exposes a set of rich resources as well as a configurable module that abstracts the use of the Clumio backup as a service for AWS. From connecting multiple AWS accounts and regions, to setting up policies and protection rules, to adding users and creating organizational units, the Clumio Provider supplies customers with an easy to define, reproducible data protection environment.

How to get started with the Clumio Terraform Provider

As the Provider uses APIs to abstract the use of the Clumio cloud, you should create an API key from the Clumio UI or retrieve an existing one. For help with creating an API key, please refer to the Clumio documentation. The subsequent steps assume that such an API key is available to you.

Preparing your Terraform automation

Start by setting up the following environment variables to allow the Clumio Provider to interact with the Clumio cloud on your behalf. For allowed API base URLs, please refer to the Clumio Provider documentation:

```
> export CLUMIO_API_TOKEN=<CLUMIO_API_TOKEN>
> export CLUMIO_API_BASE_URL=<CLUMIO_API_BASE_URL>
```

The AWS Terraform Provider is used by the Clumio module to provision the resources required to perform data protection in the AWS account to be protected. As such, set the following additional environment variables:

```
> export AWS_ACCESS_KEY_ID=<AWS_ACCESS_KEY_ID>
> export AWS_SECRET_ACCESS_KEY=<AWS_SECRET_ACCESS_KEY>

# If a session token is required ...
> export AWS_SESSION_TOKEN=<AWS_SESSION_TOKEN>
```

The following starter Terraform configuration sets up for the required Clumio and AWS Providers. Download the providers with **terraform init**:

```
terraform {
  required_providers {
    clumio = {
      source = "clumio-code/clumio"
      version = "~>0.2.4"
    }
    aws = {}
  }
}
```

```
> terraform init

Initializing the backend...

Initializing provider plugins...
- Finding clumio-code/clumio versions matching "~>
0.2.4"...
- Finding latest version of hashicorp/aws...
- Installing clumio-code/clumio v0.2.4...
- Installed clumio-code/clumio v0.2.4 (signed by a
  HashiCorp partner, ...)
- Installing hashicorp/aws v4.14.0...
- Installed hashicorp/aws v4.14.0 (signed by HashiCorp)
...
```

Connecting data environments

Next, add the following to the Terraform configuration to instantiate a Clumio connection to the AWS account associated with the AWS environment variables setup during **Preparation. us-west-2** is specified as the region in which to install the Clumio module.

```
# Instantiate the AWS provider
provider "aws" {
  region = "us-west-2"
}

# Retrieve the effective AWS account ID and region
data aws_caller_identity current {}
data aws_region current {}

# Register a new Clumio connection for the effective AWS
account ID and region
resource "clumio_aws_connection" "connection" {
  account_native_id = data.aws_caller_identity.current.
account_id
  aws_region = data.aws_region.current.name
  description = "My Clumio Connection"
}

# Install the Clumio Protect template onto the registered
connection
module clumio_protect {
  providers = {
    clumio = clumio
    aws = aws
  }
  source = "clumio-code/aws-template/clumio"
  clumio_token = clumio_aws_connection.connection.token
  role_external_id = "my_external_id"
  aws_account_id = clumio_aws_connection.connection.ac-
count_native_id
  aws_region = clumio_aws_connection.connection.aws_re-
gion
  clumio_aws_account_id = clumio_aws_connection.connec-
tion.clumio_aws_account_id

  # Enable protection of all data sources.
  is_ebs_enabled = true
  is_rds_enabled = true
  is_ec2_mssql_enabled = true
  is_dynamodb_enabled = true
  is_s3_enabled = true
}
```

Confirm your work thus far with **terraform init** to download the Clumio module and then **terraform plan** to inspect what resources will be provisioned. NOTE the above Terraform configuration enables support for data protection on all AWS data sources. When ready run **terraform apply**:

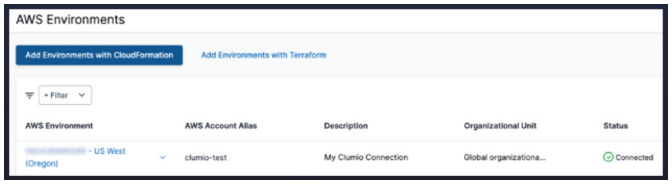
```
> terraform init
Initializing modules...
Downloading clumio-code/aws-template/clumio 0.0.4 for
clumio_protect...
- clumio_protect in .terraform/modules/clumio_protect
...

> terraform plan
...
Plan: 52 to add, 0 to change, 0 to destroy.
```

```
> terraform apply
...
Do you want to perform these actions?
  Terraform will perform the actions described above.
  Only 'yes' will be accepted to approve.

  Enter a value: yes
...
Apply complete! Resources: 52 added, 0 changed, 0 de-
stroyed.
```

Your AWS account and region are now onboarded. You can confirm this from the AWS Environments page on the Clumio UI:



Automating data protection

To get started with backup, include the following in the Terraform configuration to create a Protection Group for S3, define a policy for it, and associate the two together. As a result, any S3 bucket with the tag key-value **clumio:blog** will be protected:

```
# Create a Clumio protection group that aggregates all S3
buckets with the tag "clumio:blog"
resource "clumio_protection_group" "protection_group" {
  name = "My Clumio Protection Group"
  bucket_rule = "{\"aws_tag\":{\"$eq\":{\"key\":\"clu-
mio\", \"value\":\"blog\"}}}"
  object_filter {
    storage_classes = [
      "S3 Intelligent-Tiering", "S3 One Zone-IA", "S3
Standard", "S3 Standard-IA", "S3 Reduced Redundancy"
    ]
  }
}

# Create a Clumio policy for protection groups with a
7-day RPO and 3-month retention
resource "clumio_policy" "policy" {
  name = "S3 Gold"
  operations {
    action_setting = "immediate"
    type = "protection_group_backup"
    slas {
      retention_duration {
        unit = "months"
       value = 3
      }
      rpo_frequency {
        unit = "days"
        value = 7
      }
    }
    advanced_settings {
      protection_group_backup {
        backup_tier = "cold"
      }
    }
  }
}
```

```
# Assign the policy to the protection group
resource "clumio_policy_assignment" "assignment" {
  entity_id = clumio_protection_group.protection_
group.id
  entity_type = "protection_group"
  policy_id = clumio_policy.policy.id
}
```

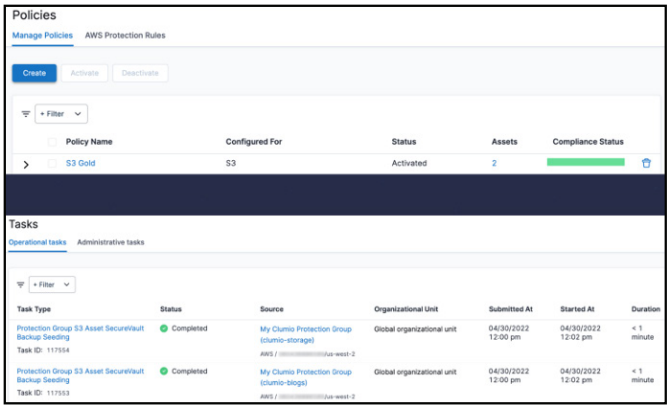
Again, confirm your work with **terraform plan (terraform init** is not required) to inspect what resources will be provisioned. When ready, run **terraform apply**:

```
> terraform plan
...
Plan: 3 to add, 0 to change, 0 to destroy.

> terraform apply
...
Do you want to perform these actions?
  Terraform will perform the actions described above.
  Only 'yes' will be accepted to approve.

  Enter a value: yes
...
Apply complete! Resources: 3 added, 0 changed, 0 de-
stroyed.
```

... and that's it! Any S3 bucket with the tag key-value **clumio:blog** will start to seed and subsequently back up every 7 days.



More resources on automating data protection

With these steps, you can take your data protection infrastructure and start to manage it as code. While the above walks you through a simple data protection setup, you can find more examples in our Clumio provider documentation. This includes how to connect and protect multiple AWS accounts and regions as well as how to organize and manage multiple users and organizational units. Documentation for each custom resource supplied by the Clumio provider can also be found.

Additional improvements and plans for the provider are continuously in discussion, and support for new features and AWS data sources will be added in subsequent releases. If you happen to already be using Terraform to manage your infrastructure, the Clumio provider is the perfect complement for your data protection needs (and if not using Terraform, this is a great chance to give infrastructure as code a try). We certainly welcome additional feedback from the community as we look to improve upon the provider. Better yet, if you want to contribute to our repository, we're happy to take pull requests. Happy provisioning!

RECOVERING AMAZON S3 DATA

Clumio helps make finding and recovering S3 data fast and easy. Here's how it works:

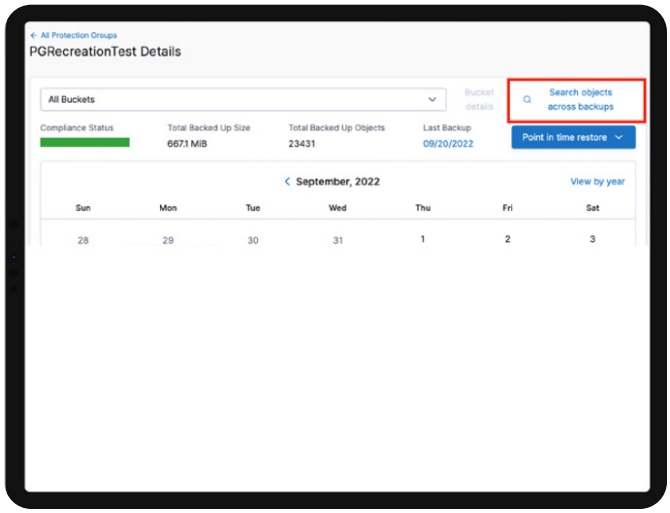
S3 data recovery using global search

A backup is only as good as its recovery. If you're unable to recover exactly what you need from any Point In Time, to any location you choose, your solution has missed the mark.

Most recovery use cases we hear from customers involve recovering a single object or multiple objects. Multiple objects could be an entire prefix, an entire bucket, or even multiple buckets at the same time. Everyone wants to avoid incidents within their production environment but if incidents do happen, the capability to restore multiple buckets becomes crucial. With Clumio, customers can either use the Global Search capability or select a specific Point in Time to recover objects. Let's start with Global Search:

Restoring from a Protection Group using global search

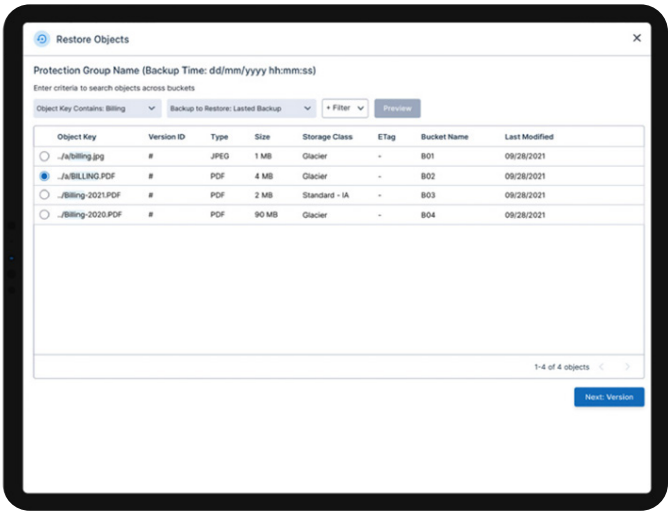
From your Clumio Protect dashboard, on the Protection Groups tab under S3, clicking on the name of a Protection Group will bring you to its asset details page. On this page, click on Search objects across backups as shown:



When searching across backups, you can specify parameters for a range of search options like:

- Object Key Contains
- Object Key Begins With
- Selected Backup (latest backup or all backups)
- Version ID
- Object Type
- Object Size (minimum and maximum)
- Storage Class
- ETag
- Bucket Name

Once the search criteria are specified, click Preview to see a quick preview of the first 100 objects that match the search criteria. If fewer than 100 objects match, you will see all of them:

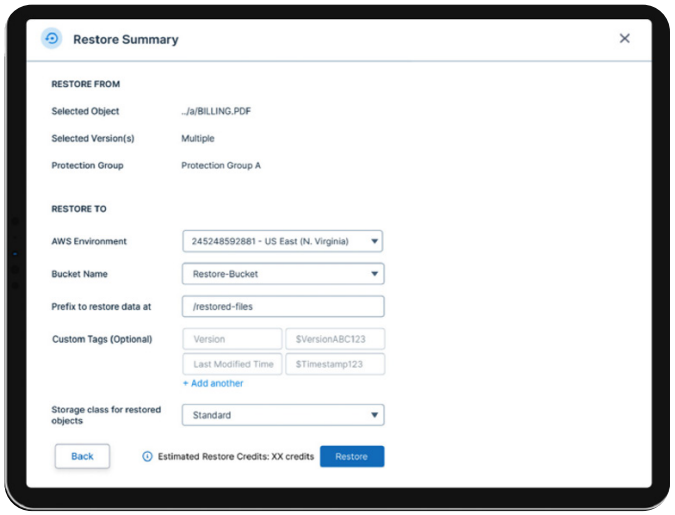


This is a Google Search-like experience where most people do not visit the second page but continue refining the search criteria to find that needle in the haystack of backups. Once they find what they are looking for, they can select the object or objects needed and click on Next: Version.

By default, the latest version that matches the search criteria will be selected, but you can select All Versions, which will populate all versions that match the search criteria. From there, you can select multiple versions of the same object to be restored. Click on Next: Summary to configure the restore options.

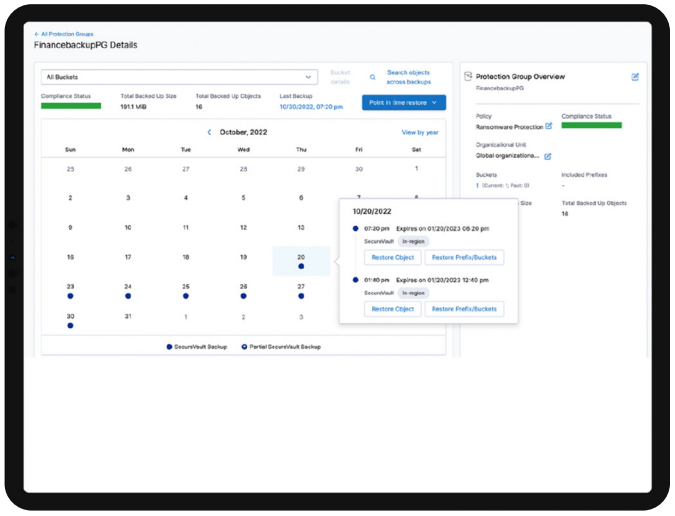
Selecting restore options

From this configuration, specify which account, region, and bucket the object (with or without multiple versions) should be restored into. You can add a prefix to differentiate this restored data or leave it blank to restore the data to its original location. The entire object key will be restored so that the object goes back to its original location. Clumio adds tags like Version and Last Modified Time so that these objects can be differentiated from others. This is especially helpful when objects are restored to their original location. You can also specify the storage class for restored objects and receive an estimate of the number of credits that will be charged for performing this restore.



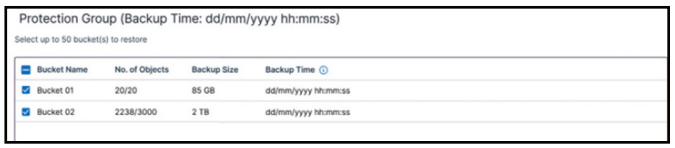
Restoring from a protection group using Point in Time

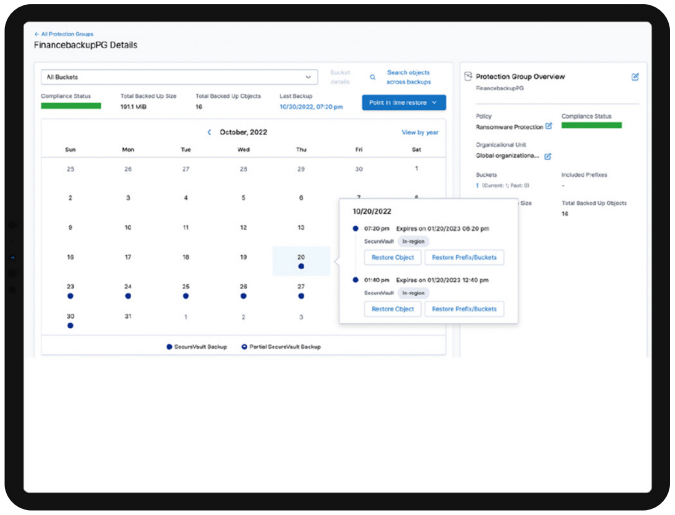
You can also restore a single object or multiple objects from a Point in Time using Calendar View.



In the calendar view shown above, pick a specific date and then select whether to Restore an Object or Restore Multiple Objects. Restoring an object is similar to the Global Search flow as covered above. Restoring multiple objects also flows similarly where you can specify search criteria. If you wish to restore entire buckets, no criteria are needed. Just select the buckets you would like to restore.

If you do need to configure search criteria, there's no need to select a particular object. All objects that match the search criteria are restored.





Preview the first 100 objects and confirm whether the objects look correct. If no search criteria are specified, then all objects that were present in these buckets will be restored.

> **Clumio was uniquely positioned to help us on dual fronts—protecting the data that we had just moved to AWS from our on-prem environment, as well as our rapidly growing native AWS assets. We explored several solutions, but Clumio's ease of use sealed the deal for us. As our cloud workloads increased, Clumio not only scaled smoothly but also made it effortless to manage the protection and recovery of large volumes of data across multiple AWS services. We have been thoroughly impressed with Clumio and are looking to fortify more of our data with their product.**
>
> **Jeremy Lee, Manager of Enterprise Cloud Storage**
> Warner Bros. Discovery

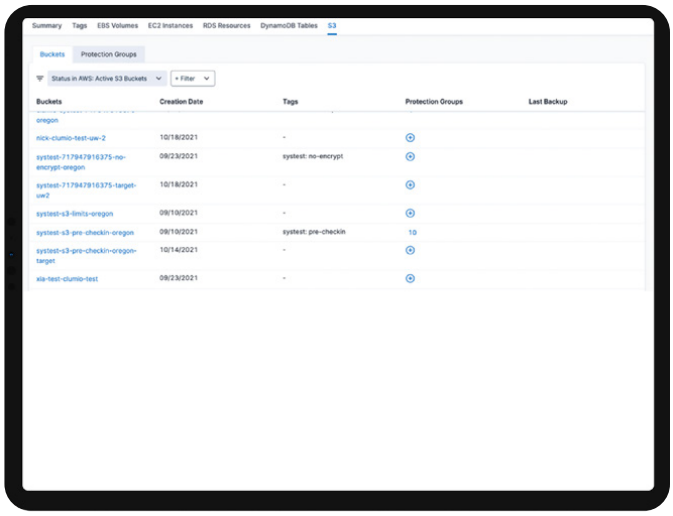
COMPLIANCE

Helping with compliance and audit readiness for your critical S3 data

Businesses in highly regulated industries run on data, but with uniquely strict requirements for security, data protection and retention. Organizations might be subject to regulations and standards such as and more. Having complete visibility and control over what gets protected and what doesn't and proving compliance when an audit occurs has become a basic necessity for Cloud Operations and Security Operations teams in the public cloud. This section will focus on how to get complete visibility and be ready to show compliance, as and when business needs arise.

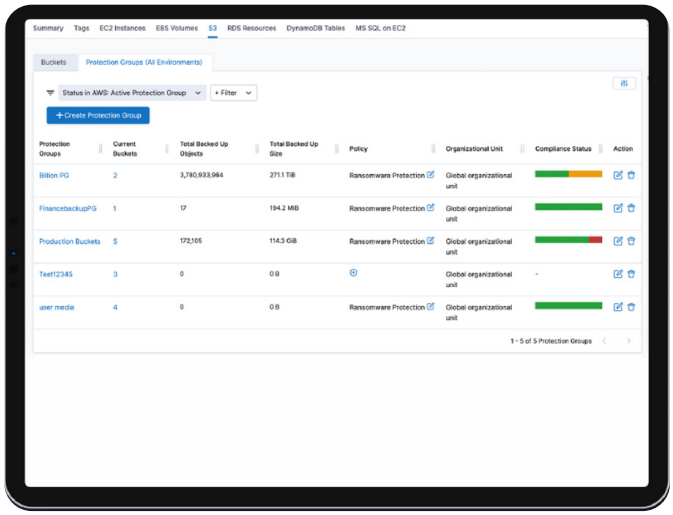
Visibility of S3 buckets with Clumio Protect for Amazon S3

As soon as you add your AWS Environment into Clumio, you get visibility into all of your S3 buckets present in that AWS Environment. You can see its name, creation date, tags, whether they are part of a Protection Group, and the last successful backup. This helps you meet your RPO requirements for your S3 buckets.



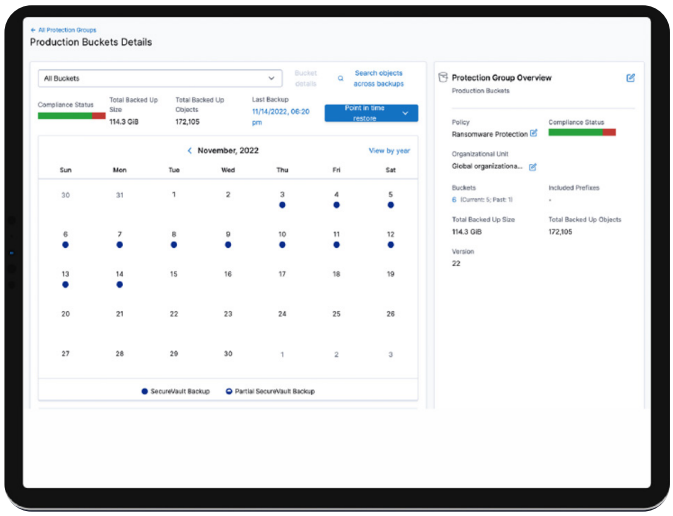
Verify critical S3 data is in compliance with Protection Groups

If you switch over to the Protection Group tab, you can easily see all the Protection Groups that have been created. You can also see the number of current buckets, the number of backed up objects, size of the backed up objects, policy that is currently protecting the Protection Group and the overall compliance status of the Protection Group. This information helps you quickly analyze whether the necessary and required number of buckets are getting protected or not and what is contributing to your overall backup costs. It is an easy way to understand if you are over protecting or under protecting S3 buckets and optimizing your backup strategy to strike the right balance between compliance needs, business continuity requirements (RPO/RTO) and cost.



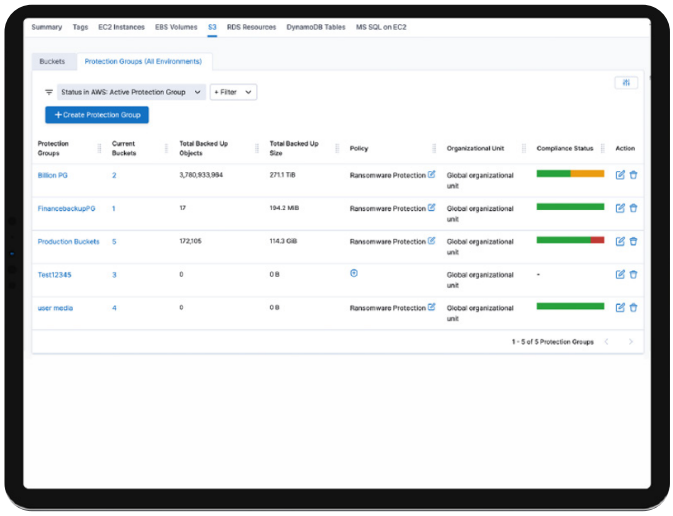
Calendar view visibility into S3 backup history

Once you click on the Protection Group, you arrive on its details page that lists a calendar with all the dates where the backups have happened in the past. This helps customers easily navigate to the state of the bucket on the particular date when the backup happened. In the calendar legends, you also see Complete Backup and Partial Backup. This easily provides visibility into whether all the objects from the bucket were backed up successfully or not. If Clumio missed any objects, then those details will be captured in the Tasks details page along with the reason on why those objects were missed. You should also be able to download a list of objects that missed a backup from this calendar view.



Get compliance reports at Protection Group levels

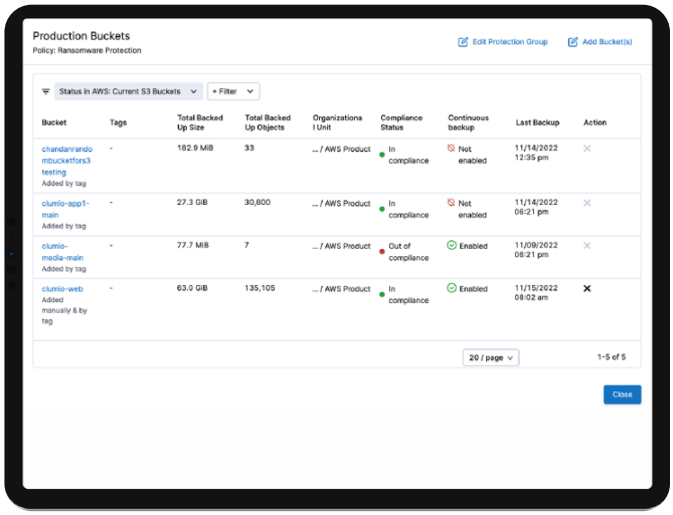
Let's switch gears to see how Clumio provides a compliance status for S3 data. Clumio Protect for S3 provides compliance reports at the Protection Group level as well as at the bucket level. As you can see in the Protection Group listings page, the compliance status is shown for all buckets inside it. It's essentially a horizontal stack bar graph of the compliance status of individual buckets inside it, and if all of them are compliant with the configured policy, then your entire Protection Group is compliant with cloud storage requirements.



Get compliance visibility at a bucket level

When you click on the number in the Current Buckets column, you get presented with a view to see all the relevant information, including Compliance Status, broken down to the bucket level. You also get other information like Backed up Size and Backed up Objects, but Compliance Status lets you know whether a key bucket is in compliance as per your backup policy.

Clumio Protect for S3 gives you complete visibility into the backups of your S3 data, helping you verify you are meeting your business continuity SLAs and prove compliance either at a bucket level or across multiple buckets through Protection Groups.



SECURITY

Security must-haves

Robust protection requires a multifaceted approach. Clumio employs these security must-haves:

- Air-gap: protected data is separated from your primary access controls
- Multi-factor authentication
- Single sign on integration
- Role-based access control
- End-to-end FIPS 140-2 encryption
- Bring-your own encryption keys

Most of these security features are self-explanatory, but why bring your own encryption keys?

Why BYOK matters

Doesn't Clumio encrypt all backups already, and if yes, why is BYOK needed? It's true that Clumio does encrypt all the backup data in its cloud with a customer-dedicated key, and that key is also rotated every 30 days. However, in some cases, certain customers have additional stringent security requirements:

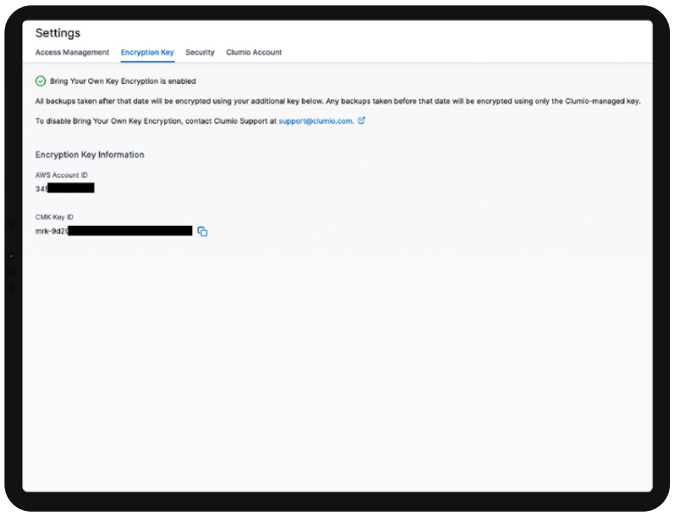
1. These customers have to make sure that their own customers' data is encrypted using that customer's own key
2. These customers want to make sure that they can readily revoke access to the backup data whenever needed.
3. These customers want to make sure that they can audit every time their data is accessed, along with the reason why it is being accessed.
4. These customers want to enhance their Zero Trust security posture by making sure that even if Clumio is compromised, attackers cannot access their data

How encryption works

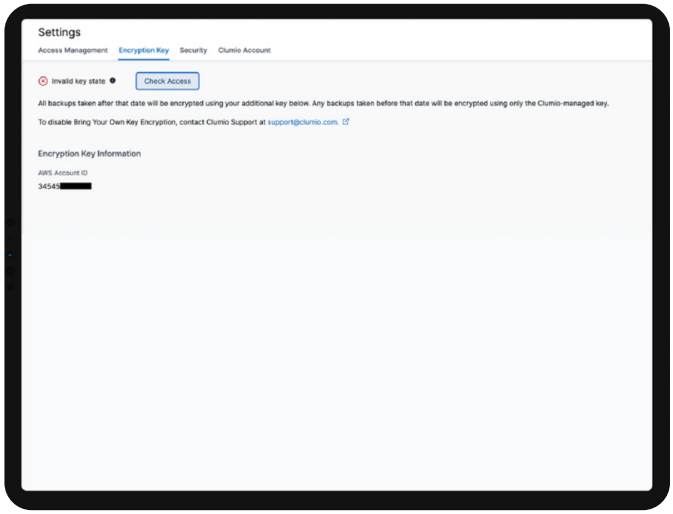
Amazon S3 bucket keys have reduced the cost of server side encryption by more than 99% and don't need to look up the key for every single transaction. Since Clumio backs up data with millions of transactions, having to look up the key for every single transaction was a no-go. With AWS' S3 bucket keys feature, Clumio can use the customer's BYOK key as the Amazon S3 Bucket Key and encrypts all the data landing in the S3 bucket using the customer's BYOK key.

How to enable the BYOK feature

Go to Settings and Security Features - Encryption Key. When you go to the page, Clumio provides details about the feature, along with its requirements and limitations. Customers can proceed by deploying the AWS CloudFormation StackSets inside any one of their AWS accounts where they want to use the BYOK key. The reason StackSets is needed is because Clumio will need to create a multi-region (global) key and use that for encrypting backups in all regions where the data sources are present. Once successfully deployed, customers can verify the connection in the Clumio UI by visiting the page at any time.

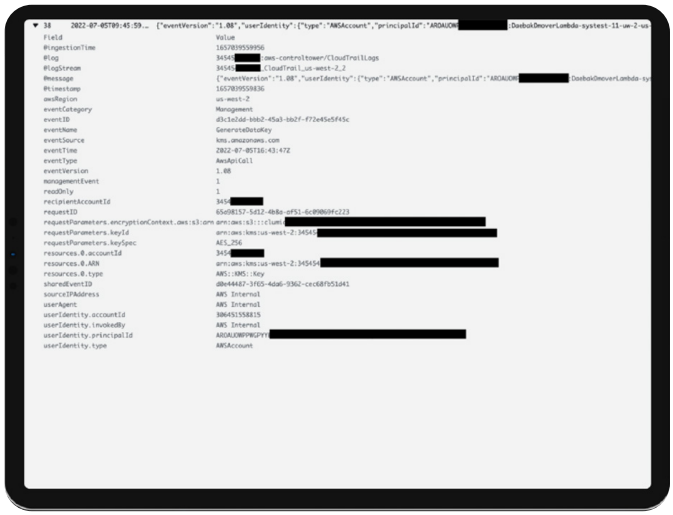


The green check at the top of the page indicates that everything is working as expected. For whatever reason, if the key is not accessible, it changes to a red X and allows you to Check Access again to see if things are resolved and working again.



How to verify and audit Clumio's access

One of the advantages of the BYOK feature is that customers can verify and audit any time Clumio has accessed their key. Customers can go to the CloudTrail logs and see all the details of each time Clumio accesses their key, including the reason for access. Customers can go to the CloudTrail Logs section in their AWS account where the AWS CloudFormation StackSet was deployed. For S3 bucket keys, since the keys are cached by the Amazon S3 bucket keys, access might not be present for every single transaction. However, keys are accessed for every single transaction for EC2/EBS, VMware, and M365 backups.



What happens when a key is no longer accessible

Remember this: BYOK gives you the power of encrypting all backup data, but on the flip side, if the key is lost, then you're in big trouble! Luckily, keys in AWS aren't like physical keys and even if someone deletes them, customers still get 30 days to recover them. For whatever reason, if the keys are not recovered, then backup data would be rendered useless. It is meant to be used as a fail-safe mechanism, but great care should be taken to use this feature.

WRAPPING UP

As discussed above, Clumio delivers data protection that is:

SIMPLE

Clumio is built with simplicity at the forefront, with intuitive UI, infrastructure as code integrations, automation, and incredible flexibility.

SECURE

Clumio air gaps your backups, securing them with end-to-end encryption, bring-your-own-key, multi-factor authentication, single sign on and role-based access control.

ECONOMICAL

Customers save an average 30% or more on TCO, thanks to competitive pricing and incredible work savings.

SCALABLE

You can protect buckets containing tens of billions of objects and an exabyte of data, search and restore it holistically or granularly. On the other end of scale, Clumio bundles small objects to keep your backups efficient.

FAST

Clumio's cloud-native architecture provides swift ingest, cataloging, search, and restore. Global search, calendar view, and granular restore shorten the path from need to delivery of restored objects.

To learn more about Clumio, schedule a personalized demo, or try Clumio free for 14 days, please reach out!

Try Clumio Free | Schedule Demo | Contact Us

To learn more, visit **commvault.com**