Brought to you by:

CLUMIO

Backing Up Amazon S3

for dummies®
A Wiley Brand

Protect unstructured data at scale

Minimize data risks and automate compliance

Enable rapid and flexible data recovery

Lawrence Miller

Clumio 2nd Special Edition

About Clumio

Clumio is a secure backup as a service that provides comprehensive data protection against ransomware attacks and account compromises in AWS. Meet all your compliance needs, rapidly recover from any data loss, and get actionable insights to optimize your AWS backup spend. Try it for free at clumio.com.

Backing Up Amazon S3

Clumio 2nd Special Edition

by Lawrence Miller

for dummies®
A Wiley Brand

Backing Up Amazon S3 For Dummies®, Clumio 2nd Special Edition

Published by
John Wiley & Sons, Inc.
111 River St.
Hoboken, NJ 07030-5774
www.wiley.com

ISBN 978-1-394-22106-6 (pbk); ISBN 978-1-394-22107-3 (ebk)

For general information on our other products and services, or how to create a custom *For Dummies* book for your business or organization, please contact our Business Development Department in the U.S. at 877-409-4177, contact info@dummies.biz, or visit www.wiley.com/go/custompub. For information about licensing the *For Dummies* brand for products or services, contact BrandedRights&Licenses@Wiley.com.

Publisher's Acknowledgments

Some of the people who helped bring this book to market include the following:

Editor: Elizabeth Kuball

Acquisitions Editor: Traci Martin

Editorial Manager: Rev Mengle

Senior Account Manager: Cynthia Tweed

Production Editor: Saikarthick Kumarasamy

Special Help: Lindsay Piper, Ari Paul, Poojan Kumar, Woon Ho Jung, Kaustubh Patil

Introduction

Modern enterprises use Amazon Simple Storage Service (S3) to store ever-growing volumes of unstructured data. Application workloads and architectures have evolved to take advantage of the power and scalability of the cloud, but bring unique Amazon S3 data protection challenges and requirements.

With increased Amazon S3 adoption, digital data estates have grown exponentially. Naturally, as data surface increaes, so do the associated backup and disaster recovery challenges.

Traditional tools available for enterprises are complex and insufficient to deliver a secure, comprehensive data protection solution for data on Amazon S3, meet compliance requirements, ensure business continuity, and provide visibility into backup plans to understand risks or gaps in their data protection strategies. Instead, a modern, cloud-native data protection solution is needed to thwart ransomware attacks, ensure robust data protection, and address complex regulatory and governance requirements.

About This Book

Backing Up Amazon S3 For Dummies, Clumio 2nd Special Edition, consists of five chapters that explore the following:

- Understanding the challenges of backup and recovery for Amazon S3 data (Chapter 1)
- Different data protection methods in Amazon S3 (Chapter 2)
- Determining the best way to protect your data in Amazon S3 (Chapter 3)
- What makes a great Amazon S3 data protection solution (Chapter 4)
- Ten features that make an Amazon S3 backup solution better (Chapter 5)

Each chapter is written to stand on its own, so if you see a topic that piques your interest, feel free to jump ahead to that chapter. You can read this book in any order that suits you (though I don't recommend upside down or backward).

Foolish Assumptions

It's been said that most assumptions have outlived their uselessness, but I assume a few things nonetheless.

Mainly, I assume that you work in cloud operations, architecture, or security and are interested in learning about Amazon S3 data protection and compliance solutions. I also assume you have some understanding of cloud and data protection concepts, as well as cloud security and compliance challenges.

Icons Used in This Book

Throughout this book, I use special icons to call attention to important information. Here's what to expect:

REMEMBER

This icon points out important information you should commit to your nonvolatile memory, your gray matter, or your noggin.

TIP

Tips are appreciated, but never expected, and I sure hope you appreciate these useful nuggets of information.

WARNING

These alerts point out the stuff your mother warned you about. Well, probably not, but they do offer practical advice.

Beyond the Book

There's only so much I can cover in this short book, so if you find yourself at the end wondering, "Where can I learn more?," go to `https://clumio.com`.

IN THIS CHAPTER

- Understanding the shared responsibility model
- Staying resilient as an enterprise
- Defending yourself against ransomware
- Staying in line with industry standards
- Keeping costs down

Chapter 1
Understanding the Challenges of Backup and Recovery for Amazon S3 Data

We're in the midst of a data revolution. Data volumes are growing exponentially, and data has become more valuable than ever. Big data now powers customer experiences, employee productivity, process automation, strategic insights, competitive advantages, and new data-enabled products and services. Data has become one of the most valuable assets for modern organizations, and much of that data resides in unstructured data storage platforms like Amazon Simple Storage Service (S3).

Against this backdrop, data protection is more important today than it has ever been. Increasingly sophisticated cyberattacks, including ransomware, target data. There is the ever-present insider risk posed by accidental and malicious deletions, and software-based overwrites. As data increasingly becomes an organization's most valuable asset, the devastating business consequences of data loss grow in lockstep.

To keep pace with the cloud, data protection has to be born in the cloud. It needs to be designed for the massive scalability of the cloud, embrace the environment of constant change in the cloud, and leverage the economics of the cloud. Above all, it has to be simple, freeing organizations from the complexities of data protection.

As data increasingly moves to the cloud, legacy data protection solutions simply can't address many of the unique data protection challenges of the cloud. This chapter outlines these challenges.

The Shared Responsibility Model

Amazon Web Services (AWS) offers many strategic advantages to organizations, but moving to the cloud doesn't mean you can just "wash your hands" of all responsibility for your data environment. The shared responsibility model (see Figure 1-1) shows what Amazon is responsible for managing and what the customer is responsible for managing in AWS.

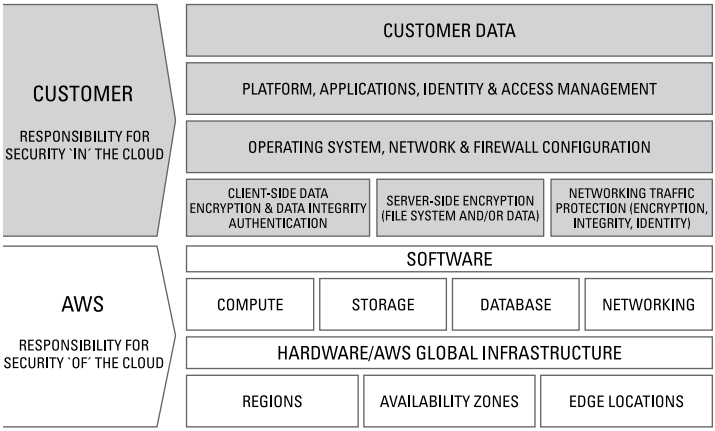


FIGURE 1-1: The AWS shared responsibility model.

Simply put, Amazon is responsible for the security *of* the cloud, including

- AWS data centers across regions, availability zones, and edge locations

- Hardware and infrastructure
- Compute, storage, databases, and networking
- Software

The customer is responsible for security *in* the cloud, including

- Encryption, authentication, and network traffic
- Operating system, network, and firewall configuration
- Platforms, applications, and identity and access management (IAM)
- Data

REMEMBER

You are always responsible for your data. This includes data security, privacy, and protection, including backup, recovery, and long-term retention.

Enterprise Data Resilience

Considering the implications of the shared responsibility model, you shouldn't rely solely on the cloud provider for backup and recovery. Public cloud provider service-level agreements (SLAs) aren't customized to meet your organization's unique recovery time objectives (RTOs) and recovery point objectives (RPOs), and cloud provider snapshots are not the same as regular backups. (We explore the differences among different modes of data protection in the next chapter.)

Given the exponential growth of data, it should come as no surprise that data loss is also increasing. Data can be deleted accidentally, it can be changed unintentionally by well-intentioned but flawed scripts, and insider threats are a very real concern.

Accidental deletions are a cause of data loss. These accidents happen all the time. For example, an Amazon S3 bucket may be accidentally deleted when migrating data to other services (such as Snowflake or Salesforce) or when performing cleanup to optimize costs.

BACKUP AND DISASTER RECOVERY: TWO SIDES OF THE SAME COIN

Backups and disaster recovery are related functions designed for different purposes. Backups protect your organization from data loss due to mistakes (such as a user accidentally deleting files), malicious events (such as a ransomware attack), or technical issues (such as a database failure). Backups are a critical component of disaster recovery (DR), but DR specifically addresses larger-scale events that may take many days or longer to recover from — such as an earthquake that causes a regional outage in an AWS data center.

Although AWS provides DR capabilities, such as availability zones and replication, to protect customers in the event of a large-scale disaster, you're responsible for ensuring these DR capabilities are properly configured to protect your data. For smaller-scale incidents that perhaps affect only a single customer, the customer is responsible for ensuring they have adequate disaster recovery capabilities — which includes backups.

Ransomware Defense and Cyber Resilience

Ransomware incidents are continuing to make headlines, and for good reason. Cybersecurity Ventures forecasts that ransomware costs will reach $10.5 trillion annually by 2025 and a ransomware attack now occurs every 11 seconds. So, having the right backup solution to help thwart such attacks by restoring data rather than paying a ransom is of paramount importance.

The right solution recommended by experts such as the Cybersecurity and Infrastructure Security Agency (CISA) is to have air-gapped backup data that is independently secured and saved, isolated from an organization's security sphere. This ensures that the hackers can't find the backup copy even after they've breached your network or cloud environment.

When organizations use AWS Backup to protect AWS data sources, the copies could be created in the same account, or a different account, but often reside in the organization's AWS security sphere. It's possible to store copies in other accounts, but those accounts still reside within the organization's AWS security sphere. The problem with this approach is that there is no separation, or *air gap,* between the primary data and the copies (see Figure 1-2).

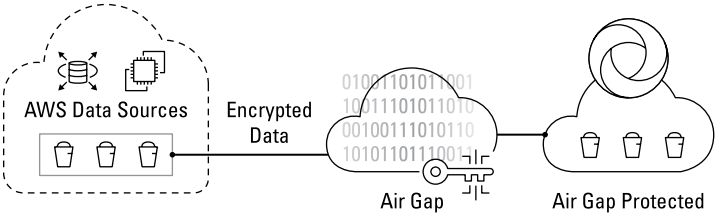


FIGURE 1-2: An air gap separates backup data copies from your primary data so that a problem in your primary data source doesn't affect your backup copies.

If a bad actor gets access to the primary account, they may compromise the replicas/versions/copies before compromising the primary data. This approach ensures that no valid usable backup copy exists and, thus, there is no way for the organization to recover its primary data. This is precisely the situation organizations don't want to find themselves in, and it's a big limitation in data protection mechanisms in the cloud.

Compliance with Industry Standards

The regulatory landscape for data protection is constantly growing and evolving, adding layers of complexity to an already confusing compliance environment.

Compliance mandates — such as the Health Insurance Portability and Accountability Act (HIPAA), Payment Card Industry Data Security Standards (PCI DSS), General Data Protection Regulation (GDPR), and System and Organization Controls (SOC) 2 Type 2, to name a few — put strict requirements around data retention, storage, and recoverability. The requirements are complex and vary across industries and regions. One common requirement is the ability to ensure that a backup policy is consistently

implemented after it's defined. Manual workarounds can lead to human errors and create process gaps, thereby increasing compliance risks.

Understanding your compliance requirements is imperative when considering a data protection solution for your cloud environment. Be sure to sort out the details before a compliance audit or assessment. Features such as data segmentation, entire copies stored outside of product accounts (also known as *air gapping*), and end-to-end encryption can help satisfy compliance requirements and provide much-needed security.


TIP

Successful compliance strategies require a holistic approach. Many elements — including people, process, and technology — can't be overlooked. A few potential problem areas include the following:

- Manual backup policy enforcement across multiple data sources and accounts
- Exposure to human error
- Rapidly changing data privacy requirements

Cost Optimization

Cloud costs are routinely identified as leading cost drivers for companies of all sizes, so it's no surprise that AWS cost optimization is an increasingly important focus. This typically requires a multifaceted approach, which includes

- Ensuring that compute instances are set up to balance optimal performance and cost
- Reviewing and updating data storage classes based on use, frequency of access, and performance requirements
- Reviewing and deleting data that is no longer needed
- Adding life-cycle rules to automate tiering and deletion
- Reducing backup and snapshot costs

Before working through cost optimization strategies, especially related to data stored in Amazon S3, it's vital to ensure critical data is backed up first. This way, you'll ensure easy recovery in case the wrong data is accidentally deleted.

WARNING

When you delete data from Amazon S3, it's deleted for good.

BACKUP VERSUS DISASTER RECOVERY

To understand backup for Amazon S3, it's important to understand the difference between backup and disaster recovery (DR). DR is a larger strategy and process for restoring systems and data in the event of a significant disruption like a regional outage, cyberattack, or other large-scale failure. DR strategies vary in line with companies' needs and SLAs, but they can include fully redundant systems for immediate failover and often include backup as an element. DR pertains not just to data, but to full applications, sites, and infrastructure.

Backup is an element of a DR strategy, but it isn't a full DR strategy itself. Backup is specific to data that resides in storage services, databases, and applications. Whereas failover copies are structured the same way as production data, backup copies are stored more efficiently. They generally must be restored to the production environment to be used. This strategy typically comes with cost benefits, but recovery is usually not as immediate. Specific to Amazon S3, data from multiple buckets may be grouped together in a backup, based on shared requirements such as retention or RPOs.

IN THIS CHAPTER

- Looking at backups
- Reviewing replication
- Understanding versioning
- Exploring Object Lock

Chapter 2
Understanding Different Amazon S3 Data Protection Methods

This chapter details the different methods of protecting Amazon Simple Storage Service (S3) data: backups, replication, versioning, and Object Lock. Each of these methods comes with unique uses, strengths, and drawbacks, and understanding these will help you make the best decision possible for your data protection requirements.

Backups

The term *backup* is often used to describe any manner of duplicating data for redundancy's sake, but it's not always an accurate moniker. By definition, a backup is a copy of data that is stored for use in case of data loss, corruption, or unavailability. This definition could describe a lot of things, until you consider what factors make that backup truly useful.

What qualities of a backup ensure you'll be able to recover data in any situation? Here are a few to consider:

- **Separate from production data:** If data loss or corruption occurs because of a regional outage or malicious activity, backed up data is only valuable if it's stored separately from production data, remaining available and secure in these situations.
- **Able to be restored where needed:** In situations where the original production environment is not available or not secure, backups need to be directly restorable to a different, secure, and available environment.
- **Time and cost efficient:** Backups are supposed to improve business recovery efforts. Therefore, a backup is only useful if it's less expensive and takes less time to restore than rebuilding production data manually.
- **Recover to any historical point in time:** Backups can be looked at as a time machine for your valuable data. If data is deleted, corrupted, encrypted, or overwritten, backups can help you restore your data back to a consistent state or point in time.

Many other factors further differentiate backups from other data protection methods and even other backup options (see Chapter 3), but the preceding factors are foundational to a true backup.

Amazon S3 Replication

Replication in Amazon S3 refers to creating one-to-one copies (also known as *replicas*) of Amazon S3 objects. The copies are stored in or a different bucket, in the same or a different region, and in the same or a different Amazon Web Services (AWS) account. Replication has several use cases:

- **Same-region replication** is used to copy Amazon S3 objects from one bucket to another within the same AWS region. It's used to create redundancy in case the source bucket experiences a failure or corruption, or to archive data into another storage class for long-term retention.

- **Cross-Region Replication** can be used to improve data access, reducing latency for globally distributed users and helping to comply with data sovereignty regulations. It's also used as a hedge against regional outages due to natural disasters or unexpected infrastructure failures.
- **Cross-account replication** is most commonly used to duplicate data to be safely used for multiple purposes. For example, production data could be replicated into a development account. Cross-account replication can also be used for security and disaster recovery (DR) purposes. It allows replicated data to be stored in a separate account, which may have separate security policies, passwords, and so on, adding a layer of safety against cyberattacks. It should be noted, however, that a separate account still lies within the company's security sphere and is, therefore, not a guarantee against data loss by malicious means.

Consider the following before you choose replication as your data protection method of choice:

- **Cost:** Replication increases the amount of data stored, and, therefore, the cost of data storage. If data is being replicated across regions, egress charges will also apply. Replication also requires versioning to be enabled, further adding to its cost.
- **Limited flexibility:** Replication is enabled via user-defined replication rules, and those rules are applied based on prefixes. So, depending on your data structure and taxonomy, you may not be able to precisely specify the group of objects to replicate.
- **Complexity:** Setting up replication rules is a manual process, so the complexity of the task grows with the size of your data environment, especially if replicating across multiple accounts and regions. Be sure to account for the associated management time.
- **Delete marker versus delete:** When a source object is deleted, its replicated copy is not actually deleted; instead, a delete marker is created and replicated to the destination bucket for a certain period of time.

Amazon S3 Versioning

Object versioning in Amazon S3 is a way to track changes and retain objects in their previous states. When enabled, a new version of the object is created every time the object is changed. Users can define how long previous versions are retained, which creates the ability to revert back if a problem is discovered with a new version. Versioning can also be used as part of object life-cycle management, wherein the user can set rules to automatically move old versions to less expensive storage tiers.

Read through the following before you choose versioning as your default data protection method:

- **Cost:** Versioning is enabled at the bucket level, rather than on individual objects, and in that way it's something of a blunt instrument. Even if you only want to version a few objects in a given bucket, all objects will be versioned, adding to the bucket's overall data volume and cost. This can be mitigated through use of rules with tags, Amazon S3 Lifecycle, Amazon S3 Intelligent-Tiering, and the Amazon S3 Glacier storage classes.
- **Complexity:** Because new versions are created at every change, special care should be taken in setting versioning rules based both on time and number of versions. Finding the right balance between these attributes can be challenging. Getting it wrong can mean excess expense on one hand or being unable to revert to the desired version on the other. The higher the change rate, the greater the challenge. Finding the right balance can require a lot of experimentation.

Amazon S3 Object Lock

Amazon S3 Object Lock can be enabled to prevent malicious or accidental deletion of objects. Object Lock is enabled at the bucket level, but retention periods can be configured for individual objects. When Object Lock is enabled and an object is within its

retention period, that object is immutable, meaning it can't be deleted or overwritten, even by users with administrative or root user permissions. Object Lock is most commonly used to comply with regulatory requirements for data retention. Object Lock can be applied using a specific retention period, or as a legal hold, which remains in place until explicitly removed.

Because Object Lock fully blocks any changes or deletions of objects, users must take care not to apply Object Lock too broadly, or they risk accidentally locking objects that need to be changed.

Please consider the following before using Amazon S3 Object Lock:

- **Specific use:** Because Object Lock could be applied to the primary or only copy of an object, it runs the risk of blocking productivity if the object is later found to need changes. This limits the types of data to which it can be applied.
- **No separation:** Object Lock keeps objects from being deleted or overwritten, but the objects remain in the user's security sphere. As such, in the case of a compromised or unavailable environment, the locked objects may not be available.

There are several data protection options for Amazon S3, each with its own uses, benefits, and considerations. Carefully assessing your unique data environment and needs before selecting one or more can help you avoid future headaches.

IN THIS CHAPTER

- Knowing what data you need to backup, when, and how to do it
- Defining data recovery and retention policies
- Keeping your backups secure
- Controlling data backup costs
- Ensuring visibility, control, and compliance

Chapter 3
Determining the Best Way to Protect Your Data in Amazon S3

This chapter explores some of the many aspects to consider when deciding how to back up your Amazon Simple Storage Service (S3) data. Things like data classification, compliance needs, and customer service-level agreements (SLAs) all factor into defining and meeting (or beating) your organization's recovery time objectives (RTOs) and recovery point objectives (RPOs) for business continuity. Additional but equally important considerations are security, costs, and visibility of your data estate.

What to Backup, When, and How

When backing up a vast and variable unstructured data store like Amazon S3, data classification is the key that unlocks your ability to find different types of data and protect each one correctly.

Consider the following factors when considering the what, when, and how of Amazon S3 data backups.

What

Amazon S3 houses a vast array of data that varies in use, format, size, change rate, criticality, and sensitivity. All of these attributes work together to mean different things about if, when, and how that data should be protected. Consider the following questions about each type of data:

- **Is it sensitive data that is subject to privacy or security regulations (for example, personally identifiable information (PII) or medical records)?** If so, the regulation will dictate aspects of the backup, such as whether it must be air-gapped and/or immutable and how long it must be retained.
- **How critical is the data to your applications or operations?** If the data is lost or corrupted, would it impact business function? How long could you wait to get it back?
- **Is this data owned or used by customers?** Have you made promises to customers about its availability, redundancy, or security? What does the SLA dictate?
- **Is the data frequently overwritten or does it remain static?**
- **Is it production or development data?**
- **Are there a lot of objects or only a few? Are they small or large in size?**

REMEMBER

Understanding what kind of data you have first informs whether it should be protected. The need to protect critical data seems obvious, but just as important is understanding what *not* to protect. This is especially true in Amazon S3, which can house data all along the spectrum of importance and at enormous volumes. Backing up too broadly can get needlessly expensive!

Of the data that does need to be backed up, now that you understand what you have, you can decide when and how to back it up.

When

Consider the time element of a backup and recovery plan from a few angles:

- **Will the backups be created on a one-time basis or ongoing?** In most cases, the simplest option is to automate the ongoing incremental backup of data that changes. This ensures that the backups always reflect the most up-to-date versions. However, there are some instances in which you may want to take a manual, on-demand backup:
  - Data that is backed up on an automated schedule, but receives an important change in between backups that you want to capture immediately
  - Outlier data that is important, but doesn't fall within your existing parameters for backup and is unlikely to change
- **How often do you need to capture the backups?** Another term for this is RPO. The more frequently your data changes, and the more critical that data, the lower the RPO you're likely to want, so you can capture all those historical points in the data's life. Schedule backup frequency to align with these requirements.

Now that you've determined what to back up, when, and how often to back it up, the last aspect to consider is how you'll back up your data.

How

The what and when of data backup involve a lot of considerations. Instead of manually managing all these aspects of backups on an ongoing basis, your best bet is to automate. Automating your backups consists of two parts: logically grouping like data and applying a backup and retention policy to that group:

- **Grouping data:** There are several means by which Amazon S3 data can be grouped together:
  - You can apply tags to Amazon S3 objects. Many organizations employ common tagging standards as a best practice. Tagging data for retention and compliance access are particularly useful in backup.

- Buckets may contain various mixes of data you do and don't want to backup, but they can still be a good starting point for grouping data, especially if you refine what's included in a later step.
- If you have versioning (see Chapter 2) enabled, you may want to back up all available versions or just the most recent. This could come into play on data that changes more frequently than the RTO.
- Prefixes uncover the granularity level in between bucket and object, and offer an excellent indicator of what kind of data is there, making them a great tool for grouping.

» **Applying policies:** Having grouped data with the same protection requirements, you would then apply a backup and retention policy to that data. The policy should have three facets:

- *Frequency:* How often backups should be taken, from continuous to monthly
- *Retention:* How long backups should be kept (from months to years)
- *Backup tier:* Standard for faster recovery or frozen for lower cost

How to Define Data Recovery and Retention Policies

RTO and RPO serve as the heart of every enterprise's business continuity plan (BCP). In today's business landscape, it's just a matter of time until an organization experiences a ransomware incident, a disruption, downtime due to bad actors, or accidental deletions. With so much at stake regarding an enterprise's applications and data, you need a disaster recovery (DR) plan in place to ensure business continuity in case any of those events happens.

Assessing your enterprise's tolerance for data loss and recovery time is an important step to understanding the impact of a potential disruption to its mission-critical applications and databases, and any effects it may have on infrastructure, end users, and customers.

Whereas RPO primarily informs backup decisions, your RTO requirements will inform the way you go about the recovery process. Data takes time to restore, and the amount of time is influenced by several factors:

- **Architecture of the restore process:** Like any software, there are numerous ways to build backup and recovery, and those choices impact its performance. The traditional way was to size the processor being used to optimize it for typical restore sizes, but this method resulted in bottlenecks that increased restore time, especially for large jobs. A more modern approach is to use parallelized ephemeral units of compute, such as lambda functions, to scale the restore effort to each job, thereby reducing RTO (see Figure 3-1).
- **Amount being restored at once:** If a restore job is very large, even the fastest, most advanced solution will be unable to restore all the data immediately. In this case, you have a couple of options:
  - *Prioritize objects or prefixes, and restore them in smaller groups, with the most critical data being restored first.* This can help you get the most critical items back up and running while the rest are being restored.
  - *Use Instant Access Restore.* This is a Clumio innovation that allows you to immediately access a read-only version of a backup bucket from your command-line interface (CLI), enabling you to temporarily point applications here while you wait for the full restore.

REMEMBER

Amazon S3 storage classes are great for optimizing costs and performance. With that optimization comes a reduction in restore speed if you're backing up to the frozen tier. As such, you should only back up data to the frozen tier if you don't mind a longer restore time.

TIP

What if you need to restore from a particular point in time? Point-in-time restore is an important tool in restoring high change rate objects or buckets to their last known good state.

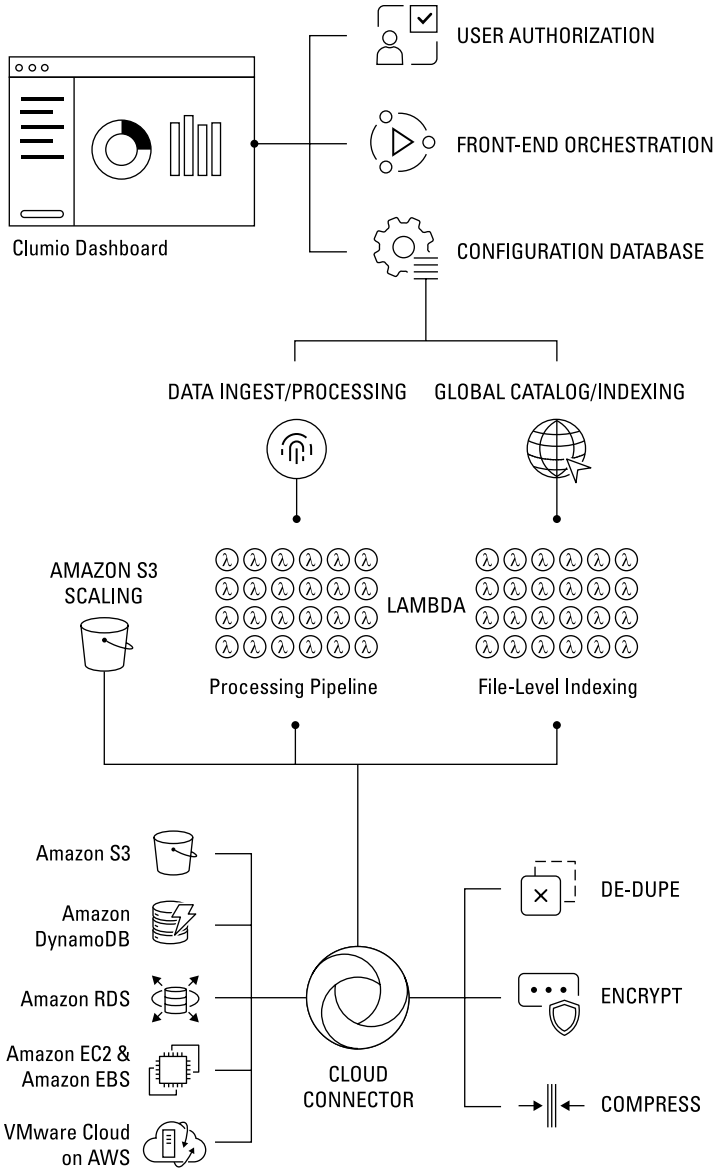


FIGURE 3-1: A modern backup and recovery architecture leverages ephemeral compute capabilities to provide infinite scalability.

When setting retention policies, you should consider a few things, such as:

- **Compliance requirements:** If your data is subject to any compliance regulations, check to see whether retention is specified. Regulations like the U.S. Health Insurance Portability and Accountability Act (HIPAA) and Family Educational Rights and Privacy Act (FERPA) require certain types of data to be retained for up to ten years.
- **Organizational requirements:** Organizations often have their own standards for data retention. Data related to things like intellectual property, personnel records, or legal documents could have long-term retention requirements.
- **Data use case:** If no compliance or organizational requirements are in place for a given type of data, consider how it's used. How critical is it to business function? How often does it change? When will it become outdated or no longer needed? Let the answers to these questions guide your retention decision.

WHAT YOU NEED TO KNOW ABOUT RPO AND RTO

Both RPO and RTO are at the center of a business continuity plan, but each serves its own unique, essential purpose.

RPO is primarily focused on data and an enterprise's resilience to the loss of that data. It's the maximum period of data loss that an organization is willing to accept in the event of a disruption. For example, if your organization is willing to accept losing a full day's worth of changes to its data, the RPO is 24 hours, and a daily incremental or full backup may be sufficient to meet the RPO requirement for your organization. An RPO of four hours may require more frequent incremental backups to meet the requirement.

RTO is the maximum amount of time an organization has to recover from a disaster before the downtime causes a break in business continuity, resulting in significant repercussions. In other words, this is the time window your organization has to fix things before your

(continued)

(continued)

customers, operations, and end users are significantly affected. RTO has a broader focus, and includes the entirety of the enterprise's operations and applications. It identifies the length of time the enterprise can function properly in the event of downtime before its operations are significantly compromised.

When calculating RPO and RTO, there are a number of factors to consider. The following steps can be used as a general outline to identify the baseline RPO and RTO for your enterprise:

1. **Create a comprehensive list of each system and application, including storage and databases, used for critical business operations.**
2. **List the teams and end users that would be disrupted by outages to these critical systems and applications.**
3. **Estimate losses in the event of downtime to these systems and applications, including revenue losses and additional expenses that would be incurred without access to them.**
4. **Consider whether your organization is responsible for overseeing customer data. If so, consult your service agreements to identify the precise time window you have for recovering customer data.**

After you've accounted for all your critical systems and applications, identify the system or application that would generate the most severe loss in the event of downtime. The recovery point and time requirements for this system or application is your organization's baseline RPO and RTO.

How to Secure Your Backups

It's just as critical to ensure your backups are properly secured as it is to secure your primary data. A backup's whole purpose is to be available to restore in case of a data loss event, so you need to make sure you can rely on it to be there when you need it. With this in mind, you should ensure that several security features are in place on your backups:

- **Air gap:** The concept of an air gap is also sometimes referred to as *offsite.* It refers to the best practice of storing

your backups in a totally separate security sphere from your primary data. This is a greater separation than simply storing them in a separate account. Using an air gap places an extra layer of security around your backups, so they would be unaffected by a breach of your primary security sphere.

- **Immutability:** Data that is immutable can't be deleted by anyone, maliciously or accidentally. Immutable backups are required by several compliance standards because they help eliminate certain risks related to data loss, even in the event security is breached.
- **Encryption:** Encryption ensures data remains indecipherable to unauthorized individuals. Encryption can be applied to data in flight and at rest; it's an important security feature, especially for sensitive data such as PII, health records, or financial data.


TIP

Backups should always be encrypted. In addition to the encryption keys provided by the backup software, also consider bringing your own key (BYOK), using the AWS Key Management System (KMS) to add an additional layer of security and control.

- **Multifactor authentication (MFA) and single sign-on (SSO):** MFA and SSO are quickly becoming security must-haves for organizations of all sizes and for all web-based applications and services. They help ensure logins are valid while simplifying the sign-in experience for authorized users.
- **Role-based access control (RBAC):** RBAC is key in maintaining the principle of least privileged access and in defending against insider attacks. RBAC allows enterprises to create separate organizations under their umbrella, dividing access and creating hierarchies that improve security while also simplifying reporting. Access controls also allow setting permissions for individual users.

How Not to Break the Bank

The Amazon Web Services (AWS) consumption-based pricing model enables enterprises to easily scale up their data environments as their needs grow, fundamentally changing the way data is managed. One challenge that has come along with this is that as data environments grow over time, so do the associated costs.

Protecting data increases your data footprint and, therefore, your costs, so it's important to understand and manage these costs. Backup costs come in two main forms:

- **Data costs:** On their face, data costs seem simple; whatever the data costs, it costs. But the reality is more nuanced than that. Different methods of data protection will come with different per-gigabyte, per-month prices, so a smart place to start is to simply compare the prices of potential solutions.

  In addition, consider the following data cost drivers:

  - *Data classification:* Coming back to a theme mentioned earlier in this chapter, robust data classification capabilities help reduce your backup spend by allowing you to granularly select and back up only the data that requires it.
  - *Backup tier selection:* When backing up data that has long-term retention requirements, but for which restoration speed would not be critical, you can reduce costs by backing up to a colder-tier backup such as frozen. This will come with a lower cost and slower restore times.

- **Management costs:** A sizable contributor to costs that is often overlooked is management cost. Consider the amount of time needed to set up, manage, and troubleshoot your backup solution. Depending on the size and complexity of your data estate, this can add up to a significant cost in terms of worker hours. This is both a real monetary cost and an opportunity cost. Look for automation features that allow you to set up parameters for protecting different types of data, capturing not only your environment's current state, but changes and additions, without requiring manual updates.

REMEMBER

The cloud is supposed to offer cost efficiencies across your data estate, including backup and storage. However, choosing the wrong cloud data protection solution can result in the opposite effect due to charges that you didn't anticipate, such as egress, compression, and storage.

There's also the issue of racking up expenses from inefficient creation, storage, and usage of copies. To make matters worse, many data protection vendors offer confusing pricing models that are overly complicated, making it hard to predict what the total cost of ownership (TCO) should even be.

TIP

You can avoid excessive cloud spending by choosing a cloud backup vendor with a simple, straightforward pricing model that clearly spells out the ongoing charges.

Many cloud backup users also tend to assume that the cloud platform will make the backup management process easier, but it's never a given. Aside from the actual managing of the backups (scheduling, running recovery tests, and so on), you may still have to account for the planning of data usage, upgrades, patching, and security.

Opt for a single cloud-native data protection service with an intuitive interface that provides clear visibility into all the processes you're responsible for managing while eliminating any guesswork and as many manual processes as possible.

WARNING

Management complexity in legacy backup infrastructure creates numerous challenges for organizations, including the following:

- Infrastructure design requires significant time, skills, and costs. Professional services can offload some of this effort, but this convenience comes at a price.
- Capacity planning involves forecasting many variables such as storage usage, data change, and deduplication rates.
- Hardware and software maintenance, upgrades, and patching require ongoing costs, planning, and testing.

REMEMBER

Cloud providers ensure infrastructure maintenance and security, but the onus of data protection is always your responsibility.

How to Ensure Visibility, Control, and Compliance

Data protection in the public cloud is evolving, and users don't always feel confident in their data's compliance and protection status. Enterprises not only need a solid data protection platform, but also need visibility and insights to help verify their data is being correctly protected.

Setting data protection policies across all applications is not a daily task. In fact, policies should typically be created once with the right attributes and then left alone. Thus, admins need tools and services that automate backups and track their history and compliance status.

It's important to have visibility into this type of information at their fingertips when needed. For example, they should be able to:

- Easily prove compliance during audits
- Easily find the right data during the restore process
- Select the right policy to protect a newly added application or data source
- Do all the preceding across hundreds of accounts

A cloud-native data protection and compliance solution can provide organizations with the following visibility and control capabilities and benefits:

- **Gain global visibility across accounts/regions.** Examine a satellite view of your entire AWS asset and data protection footprint for services like Amazon S3, Amazon Elastic Compute Cloud (EC2), Amazon Elastic Block Store (EBS), Amazon Relational Database Service (RDS), and Amazon DynamoDB across all accounts and regions.
- **Ensure the correct data is protected.** Highlight assets that are unprotected and vulnerable to potential data loss, including historical analysis and inventory information.
- **Optimize your AWS spend.** Optimize and reduce wasted expense by acting on orphaned snapshots or snapshots being saved outside your retention period.
- **Identify top data consumers and outliers.** See reports of the largest consumers and percentage change rate spikes.

REMEMBER

Keeping control of your data in the cloud is fundamental, but it's even more important to do it with the right tools that keep costs down while simplifying operations.

IN THIS CHAPTER

- Deploying with ease
- Scaling to meet your business needs
- Maximizing speed and flexibility
- Keeping your backups secure
- Improving efficiency with simplicity and automation
- Ensuring compliance
- Reducing stress

Chapter 4
Defining What Makes a Great Amazon S3 Data Protection Solution

In this chapter, you find out about some important features and capabilities you should ensure are present in your Amazon Simple Storage Service (S3) backup solution. Seeking out these features will help make your job easier, save you money, and keep your data secure.

Quick and Easy Setup

With any new product, there is always an initial setup period, which can range from minutes to months, depending on factors like the complexity of the product, integration with infrastructure, and scope of implementation. One of the many benefits promised by modern software as a service (SaaS) offerings is simpler setup without requiring customer-managed infrastructure or lengthy onboarding phases.

The best Amazon S3 backup solution will skip the complexity and allow you to start backing up data immediately. This is particularly important in backups, where every moment you're not backed up is another opportunity for a cyberattack or data loss to catch you off guard. Here are some best practices that ensure setting up your Amazon S3 backup solution is quick and easy:

- **Sign up through Amazon Web Services (AWS) Marketplace.** Instead of requiring a lengthy sales engagement, trying a backup solution through AWS Marketplace is a quick and hassle-free way to get started. Because it's connected to your AWS account, free trials are automatically engaged, and billing after the free trial is seamless. Signing up is as easy as clicking a button.
- **Connect multiple resources with AWS CloudFormation StackSets.** Most enterprises have multiple AWS accounts, and this number can range into the hundreds or more. In such cases, the ability to connect multiple accounts and regions at once saves immeasurable time. Integration with StackSets enables this feature.
- **Create protection policies in a few clicks.** Enabling automation should be easy. Instead of writing complex scripts, a few simple clicks are all it takes to set backup frequency, retention, and tier.
- **Use intuitive protection groups.** Just as in setting policies, writing scripts creates a huge speed bump in the logical grouping of data. Instead, the quick and easy method is via a few clicks to narrow down just the right data to be automatically protected by a given policy.

TIP

Understanding the factors affecting the speed and ease of setting up your backups can mean the difference between a monthslong effort and a 15-minute task.

Infinite Scale

Data scale exists on an almost infinite spectrum in Amazon S3. Enterprises' Amazon S3 environments can reach incredible scales with multiple exabytes of data. Buckets can contain anywhere from one object to tens of billions of objects, and each of those objects can range in size from less than 1 kilobyte to as large as 5 terabytes. The best backup solutions will be able to handle both ends of the scale spectrum with ease, speed, and efficiency.

The first scale consideration is whether a given backup solution fits your data's scale today, but the best-case scenario is a solution that will also scale automatically to fit your most ambitious projections for years in the future.

For enterprises with the largest Amazon S3 environments, maximum scale is not just a consideration, but a requirement. If your Amazon S3 buckets contain multiple billions of objects and hundreds of petabytes to exabytes of data, most solutions simply won't work. Clumio believes that the best solution for this case not only will dynamically scale to handle your environment's extreme scale, but will do so with enough speed to effectively meet customer-facing and internal service-level agreements (SLAs).

On the other end of scale considerations are small objects. Many objects stored in Amazon S3, such as text files, documents, and more, are as small as just a few kilobytes. The best backup solution will handle these small files efficiently and affordably, without imposing a minimum object size.

To enable seamless business growth, make sure the scale of your backup solution is well matched to your data environment's needs, now and in the future.

Fast, Flexible Data Ingestion and Recovery

Every stage of a backup's life is important, but the two stages most impactful to the user experience are ingestion and recovery. These are also the times when the volume of data and the performance of the backup solution most readily differentiate themselves.

Ingestion speed is important in part because the ingestion process is repeated on a regular basis. Slow speeds can result in excessively long ingestion time, missed SLAs, and failed backups. The larger the volume of data, the more severe the potential impact.

Data recovery is a less frequently needed process, but it's almost always highly time sensitive. Slow recovery speeds can mean lengthy business and service disruptions, lost revenue, and lost customer trust. In these critical moments, the benefits of a speedy recovery solution are incredibly valuable. Speed can come in the

form of processing power, but that power has limitations if it's within the confines of a monolithic data pipeline. When restoring data, especially at significant volumes, parallelized processing wins the day, allowing the restoration of greater volumes of data, faster.

Related to speed, flexibility is a key influencer of backup ingest and recovery. As the volume of data being ingested or recovered directly impacts speed, it follows that the flexibility to back up or restore only the needed data is an important performance optimization capability. In addition to the right data, the ability to restore directly to the desired account, region, bucket, and storage class speeds the process even more, instead of having to restore locally and then copy the data. This capability is especially important when restoring due to a security breach, in which the original environment may be compromised.

Air-Gapped and Secure

As discussed in Chapter 2, the term *backup* is sometimes used to describe a number of different data duplication strategies that don't quite meet the definition of a true backup. One important distinction is that backed-up data should be secure and separate from production data, within a totally separate security sphere. This is called *air gapping.* An air gap shields backed-up data from ransomware attacks and insider threats, ensuring that copies are readily available to be restored when needed.

Another term used often when referring to backups is *immutability.* Data that is immutable cannot be changed or deleted. This is an important layer of security, especially for data that is subject to retention requirements. Having immutable backups also ensures that even if a bad actor were to access your backups, they would be unable to delete or change the data.

Because air-gapped backups reside in their own security sphere, that security sphere needs to be robust, using security best practices like multifactor authentication (MFA), single sign-on (SSO) integration, end-to-end encryption, and role-based access controls (RBACs). These features help to ensure that your backups are secure and only accessed by authorized users.

Simple to Use

Simplicity is more than just a nice-to-have. The most efficient cloud teams are able to run lean in part because they rely on SaaS products that streamline their workflows, allowing them to focus on the projects and tasks that most closely align with core competencies. Backup is a must-have for all data-focused businesses, but it's almost never a core competency or differentiator. As such, the simpler a backup solution is, and the less time your team has to spend managing backup and recovery, the more beneficial it is to your business. A truly simple solution will be quick and easy to onboard and navigate and will allow you to automate your backups. When restoring data, it should be easy to find the data you need from the date and time needed. A truly simple solution can reduce management time by as much as 80 percent (`https://saas.clumio.com/rs/666-JFU-768/images/ESG-Economic-Validation-Clumio-June-2022.pdf`).

Compliant and Discoverable

One important thing a great backup solution can do is help you comply with regulatory standards and easily provide evidence in audits. Organizations that are subject to standards like the International Organization for Standardization (ISO) 27001, ISO 27701, Systems and Organizations Controls (SOC) 2, Health Insurance Portability and Accountability Act (HIPAA), Payment Card Industry Data Security Standards (PCI-DSS), Children's Online Privacy Protection Act (COPPA), and Family Educational Rights and Privacy Act (FERPA) have the added challenge, beyond just protecting their data, of protecting it in a compliant way.

Each standard is different, but there are three main components that are common across many such standards:

- Air-gap requirements
- Immutable backup requirements
- Retention requirements

As a first priority, make sure your backup solution can deliver in these areas. However, a truly superior solution should go a step further by offering its own certification for the standard in question. That extra layer of compliance comes in handy at audit time.

Speaking of audits, a backup solution should make them easier. Scheduled and on-demand compliance reports are helpful in preparing for audits, as well as being useful references to provide your auditor. During an audit, an auditor may perform a spot check in which they select some data and ask you to prove your ability to recover it. In these cases, the ability to search for objects across backups saves an incredible amount of time and stress.

Compliance is an ongoing requirement for many organizations. Make sure your backup solution makes it easier to accomplish.

Peace of Mind as a Service

The best backup solution will put your mind at ease, rather than add to your stress and workload. Automation is your friend in this endeavor, allowing you to "set it and forget it" as much as possible. Going one step further than automating backup tasks, solutions that offer automated proactive service offer true peace of mind. Instead of having to remember and set aside time to monitor the progress and success of backup and restore jobs, these solutions allow you to rely on automatic alerts to inform you of any issues. Getting automated daily summary reports adds an extra layer of certainty and awareness. The best solutions don't just let you know about issues, they strive to solve them automatically, without your involvement when possible.

A high level of automation across task types can not only help you reduce labor costs, but also provide an added level of confidence-boosting transparency through automated reporting.

MAKING THE CLOUD SAFE FOR CLIENT DATA

Frogslayer, a custom product development and commercialization firm, helps growing companies leverage the cloud as they digitally transform their businesses. "We not only design custom software for our clients, we host and manage the services as well," explains Brian Cahill, Director of Technology and DevOps for Frogslayer. "We offer them a solution in which they don't have to worry about managing the software. From backup to security to monitoring, we take care of it all for them."

The challenge: Supporting a growing, dynamic client base

Taking care of it all, though, was becoming increasingly challenging for Cahill and his team as Frogslayer expanded and added new clients. "The stuff that keeps me up at night is just making sure we've got our business continuity and disaster recovery plans in place," says Cahill. "As our client base grows, how do we make sure that we're protecting everyone's data? How do we guarantee their backups are reliable and that they can retrieve data when necessary?"

As Frogslayer's clients accelerate their move to the cloud, they've been increasingly concerned about security, demanding confident answers to questions about encryption, key management, and ransomware protection. Driven in part by industry regulations and standards such as the International Organization for Standardization (ISO) 27001, the National Institute of Standards and Technology (NIST), the Health Insurance Portability and Accountability Act (HIPAA), and the California Consumer Privacy Act (CCPA), these security concerns have only grown.

Like many firms, Frogslayer had chosen to create its own backup scripts for AWS. Although that may have worked in the beginning, the process was becoming painful, requiring new scripts for every upgrade of cloud application programming interfaces (APIs), and it started taking up more of their time.

"With our increased use of AWS services, as well as the need to use other clouds such as Microsoft Azure, we realized our code footprint was getting too big and the effort to upgrade the scripts was getting unmanageable," Cahill says. Simply put, Frogslayer couldn't sustain the data protection model it had in place; it had hit a wall. Cahill knew it was too complex and, ultimately, too expensive to maintain a home-grown approach.

The solution: Finding an innovative approach in Clumio

As the firm embarked on its search for a new data protection solution, the focus was on finding a service that would provide the robust backup and recovery needed for both Frogslayer and its clients, while significantly simplifying the management of the overall system.

Unfortunately, as Frogslayer's search for a solution progressed, it found that most data protection vendors were simply selling what Cahill was already doing, and without any real innovation. "The majority of the solutions that we were looking at were basically AWS snapshot managers. They ran mostly the same APIs I was running," Cahill says. "I would just be paying somebody else to do it."

(continued)

(continued)

The problem with snapshot managers was they forced Cahill to choose between low cost and high risk, or high cost and low risk. Although they provide a mechanism to back up data, their approach is cumbersome and doesn't provide adequate security for stored data. Plus, snapshot managers don't easily address data recovery. Their data retention is expensive and their restores are complex and tedious. Instead of solving Frogslayer's data protection issues, snapshot managers only introduced new ones.

That all changed when Cahill discovered Clumio. "Clumio is doing it differently," Cahill says. "They're doing something nobody else is doing." Clumio's architecture, which is built on and for the cloud, impressed Cahill from the start. Clumio's unique ability to dynamically handle data ingestion at scale, and store and secure the data in a separate and protected location, was a major factor in his decision to go with its service.

The results: No more trade-offs

With Clumio now managing his backup and recovery, Cahill no longer has to compromise between protection, complexity, performance, and cost. He can have high-performance data protection without unreasonable costs or significant demands on his team. He can scale to meet the needs of his clients, and he can do it in a way that ensures their data has the best possible protection. Other key benefits include the following:

- **Enhanced security:** Organizations are susceptible to the ever-increasing threat from ransomware attacks, as well as unintentional deletion of backups. AWS's resilient infrastructure is bolstered by Clumio's air gap solution that separates backups from primary data. This provides a reliable recovery from data loss, whether malicious or accidental.
- **Ease of management:** Clumio didn't just offer Frogslayer a full-featured backup solution — it also made it simple to administer with a very concise and easy-to-click-through user interface. And it took only a one-hour phone call for the company to get up and running with the Clumio service.
- **Exceptional support:** Instead of just sending daily status emails as some backup vendors do, Clumio proactively and immediately logs a ticket and notifies Frogslayer if something in the system isn't right.
- **Cost-effectiveness:** Internally, Frogslayer saw the biggest cost savings in being able to free up the hours Cahill's team had been spending working on the scripts and backups. The Clumio service allows Cahill and his team to spend more time focusing on the actual business of providing service to Frogslayer's clients.

Chapter 5
Ten Features That Make an Amazon S3 Backup Solution Better

Here are ten important considerations for you to think about as you're evaluating vendors to deliver an Amazon Simple Storage Service (S3) data protection and compliance solution for your organization.

Radically Simple Setup and Use

Computing in Amazon Web Services (AWS) comes with plenty of benefits, but one of the most impactful is speed to value achieved through the use of software as a service (SaaS) products, especially those that are available for immediate self-onboarding through AWS Marketplace. However, not all solutions are created equal. There are "SaaS" backup offerings that require users to deploy software on Amazon Elastic Compute Cloud (EC2) instances, monitor capacity, manually set up backups, and watch for failures and issues.

When moving to an Amazon S3 data protection and compliance solution like Clumio, you get up and running quickly while minimizing your ongoing operational overhead. By self-onboarding through AWS Marketplace, you can be protecting assets in as little as 10 to 15 minutes. Easy-to-setup backup automation means that you can set your backups and walk away, with no need to monitor or manage ongoing tasks. Simple, intuitive recovery workflows make it easy to find the data you need and recover it quickly.

Infinitely Scalable Protection

Whether your Amazon S3 data footprint today is small or enormous, it's most likely growing. Meanwhile, although some objects may be quite large, others are mere bytes. When comparing solutions, infinite scale is not a given.

Clumio believes it is the most scalable Amazon S3 data protection solution available. Amazon S3 protection groups (more on these later) automate ongoing protection of existing and new assets so your backups scale automatically with your data.

For the largest Amazon S3 environments, Clumio is an ideal solution. Because many enterprises' Amazon S3 environments are expanding well beyond a few billion objects or a few petabytes, Clumio was architected to protect and restore buckets containing beyond 30 billion objects, at exabyte scale.

Intuitive Protection Groups

Amazon S3's vast and unstructured nature presents challenges for protecting data. When the total volume of data is high and unimportant data is mixed in among critical data, it doesn't make sense to protect everything with the same broad brush. But given the lack of structure, how do you efficiently group and protect the right data?

Clumio's Amazon S3 protection groups provide a simple and intuitive mechanism to do just that. You can group objects that require the same type of protection (frequency, retention, and

storage tier) by buckets, tags, or prefixes. Go further with prefixes by specifying which to include and which to exclude. Not only is your existing data within these parameters captured and protected, but any future data you add with the same parameters will be protected automatically. Similarly, if you already use a robust tagging taxonomy to identify different types of Amazon S3 data, they can easily be captured in your protection groups to automate protection.

Protection groups make it easy to group and protect like data, while you avoid backing up noncritical data.

Continuous Data Protection

For critical data that changes often, daily backups may not be enough. Backing up much less frequently than data changes can introduce unwelcome vulnerabilities. For instance, say an object that receives daily backups is intentionally overwritten early in the day, but later a bad script causes an unintentional, problematic overwrite. Restoring the most recent backup would mean you lose that day's early progress.

For data with the most aggressive recovery point objectives (RPOs), Clumio offers continuous protection for Amazon S3, which backs up selected data on an ongoing basis, with an RPO as low as 15 minutes. This feature can be enabled on protected buckets with a single click, making it easy to comply with the most aggressive service-level agreements (SLAs).

Flexible Recovery in a Few Clicks

Don't wait until you experience a data loss incident to think about your recovery experience. Look for an Amazon S3 backup solution that enables granular recovery of a single object, multiple objects, a bucket, or a protection group. This level of granularity helps ensure you can find and restore exactly the data you need and meet recovery time objectives (RTOs) and regulatory requirements.

Of additional concern is recovering in the event of your data environment being compromised. In such cases, you may want to restore data to a new, secure AWS account instead of its original,

compromised account. Make sure you're able to restore directly to another account, instead of having to restore to the original account and then copy the data to the new account. At best, this wastes time, and at worst it may not be possible at all.

Clumio makes it easy to recover exactly the Amazon S3 data you need, utilizing global search, filtering, and calendar view options. When recovering data, you can select what account, bucket, and storage tier to restore into, reducing time and complexity.

Automated Compliance and Visibility

Today's compliance landscape is complicated and fluid due to the emerging location-specific laws regarding user data retention within certain jurisdictions. In simpler terms, a lot is at stake, and you have to be sure you're in compliance with everything from data privacy laws to proving the security and frequency of your backup during insurance audits.

The keys to meeting compliance are threefold:

- Understanding your specific compliance requirements
- Having clear visibility into your compliance status
- Ensuring robust data security

It's best to choose a backup solution that will offer a single view into your compliance status, notify you when backups are out of compliance, and provide features that can quickly and easily prove compliance in the event of an audit by pulling specific, relevant data. Your backup solution should also be able to store backups outside of production environments and provide end-to-end encryption to meet compliance security requirements.


TIP

When selecting a data protection solution, consider SaaS solutions that address your compliance requirements such as the Health Insurance Portability and Accountability Act (HIPAA), International Organization for Standardization (ISO) 27001, and Service Organizational Control Type 2 (SOC 2). Look for vendors that can provide a single view of your compliance status and provide proactive insights when compliance is at risk.

Undeletable Data

Your backup data is your ultimate safety net in all kinds of data loss incidents, so its integrity is critical to recovery and even business survival. Backup and recovery solutions need to have multiple layers of safety to ensure that data integrity.

One such layer that comes up often in discussions around compliance and data resilience is immutability. Immutable data can't be deleted, and it's an increasing focus of many compliance standards and cyber insurance carriers.

All data backed up with Clumio is immutable, automatically. Clumio has no delete button, so it's impossible to singlehandedly delete data, whether it's you, a bad actor, or a Clumio staff member, by accident or maliciously.

A RANSOMWARE DEFENSE CHECKLIST

Protecting your data against ransomware attacks requires a multilayer defense that covers everything from thwarting such attacks to recovering quickly in the event of a breach. Use the following checklist to help you protect your organization's cloud data against ransomware.

Vendor and supplier management

- ❑ Ensure that your organization's systems and applications are up to date and continuously patched.
- ❑ Evaluate — at least annually — all your existing and future vendors to determine the risk their services add to your organization.
- ❑ Meet with your vendors' security teams and go over their practices for information security.
- ❑ Confirm that your vendors have the appropriate certifications and reports in place, including International Organization for Standardization (ISO) 27001 and System and Organization Controls (SOC) 2 Type 2.

(continued)

(continued)

- ❑ Determine whether your vendors perform penetration tests from qualified security vendors (QSVs), and verify they have attestations available.

Training and awareness

- ❑ Set up an ongoing security and privacy training program for your employees, and track the results.
- ❑ Establish a regular cadence of internal communications around security and the dangers of ransomware, keeping employees aware of the issues and what needs to be done to ensure business continuity.
- ❑ Conduct annual company meetings dedicated to overall security awareness and training.
- ❑ Conduct quarterly employee meetings addressing specific subjects such as business email compromise (BEC), subscriber identity module (SIM) hijacking, Short Message Service (SMS) text scams, and so on.
- ❑ Implement a comprehensive set of drills to test your organization's ability to manage risk and response in the event of a ransomware attack.

Technical mitigations

- ❑ Identify and prioritize the assets — systems, data, and people — that are most vulnerable to ransomware attacks.
- ❑ Develop and implement an overall plan for defending against ransomware attacks.
- ❑ Reduce your attack surface by ensuring that your vulnerable end points are optimized for security.
- ❑ Secure your email system against ransomware bait using phishing filters.
- ❑ Prohibit the use of removable storage devices, and restrict unnecessary open ports.

Privilege and authorization management

- ❑ Limit the "blast radius" of what an attacker can do by managing the level of permissions with groups or roles.

- ❑ Ensure you have strong authentications for all services, with no shared credentials.
- ❑ Develop strict access control policies, such as a Zero Trust framework that protects and authenticates user access for your systems.
- ❑ Adopt and implement the principle of least privilege.
- ❑ Review privileged access frequently.
- ❑ Establish a quarterly review of users who have admin-level access and confirm with human resources if there is any change in their status.
- ❑ Monitor all access logs, checking for access anomalies.

Disaster recovery

- ❑ Establish a comprehensive backup and recovery strategy in case of a ransomware attack.
- ❑ Ensure that your backed-up data is stored securely off site or in the cloud and allows for at least seven days of incremental rollback.
- ❑ Protect your backed up data in an air-gapped solution that is separated from your production environment.
- ❑ Make sure your backup data is immutable and as secure as possible.
- ❑ Validate that your data is encrypted in flight and at rest with keys that cycle frequently.
- ❑ Make sure you have the ability to use your own encryption keys.
- ❑ Ensure that data can be quickly restored to an account/region that is different from the compromised account/region.
- ❑ Periodically test your ability to recover data from backups.

Always-On Security

The importance of securing data can't be overstated. Ensuring data security often comes with many complicated decisions to make across multiple components of hardware, software, networks, and cloud environments. The exposures of not securing data properly are real and come with significant consequences.

Consider SaaS providers that take responsibility for securing data to protect your data in the cloud. Key security areas to consider include the following:

- **Air-gap backups for ransomware protection** (see Chapter 4)
- **Secure infrastructure that is industry certified** (for example, ISO 27001 and SOC 2 Type 2)
- **Data security,** including secure access via multifactor authentication (MFA) and single sign-on (SSO) integration, data segregation, and encryption for data at rest with the option for customer-managed keys
- **Access controls** like role-based access control (RBAC) and organizational units, utilizing identity and access management (IAM) roles
- **Network security** consisting of end-to-end encryption for data in-flight over Transport Layer Security (TLS) connections
- **Penetration testing** to facilitate in-depth source code analysis and identify new and existing platform and application vulnerabilities

Data Protection That's Also a Budget Win

Storing data costs money. This simple truth underscores a common concern about backup: its cost. Critical data needs to be protected, but there are several ways to go about it in Amazon S3, and each way comes with benefits, drawbacks, and associated costs (see Chapter 1).

The most common and easiest-to-compare aspect of Amazon S3 backup cost is the simple per-gigabyte-per-month price, an area in which Clumio is highly competitive, with the ability to save 50 percent or more depending on the alternative. But particularly in Amazon S3, the ability to be very selective about which data gets backed up can have even greater impact. The difference between having to back up the entire contents of a large bucket versus the 25 percent of its contents that are actually critical is, well, it's about 75 percent cheaper, and that's pretty impactful.

Another unexpected impactor of cost is simplicity of use and management. We cover simplicity at the beginning of this chapter, but not its cost implications. Saving labor can translate directly to saving money, particularly when it frees up staff to work on other projects or removes the need for increased headcount.

TIP

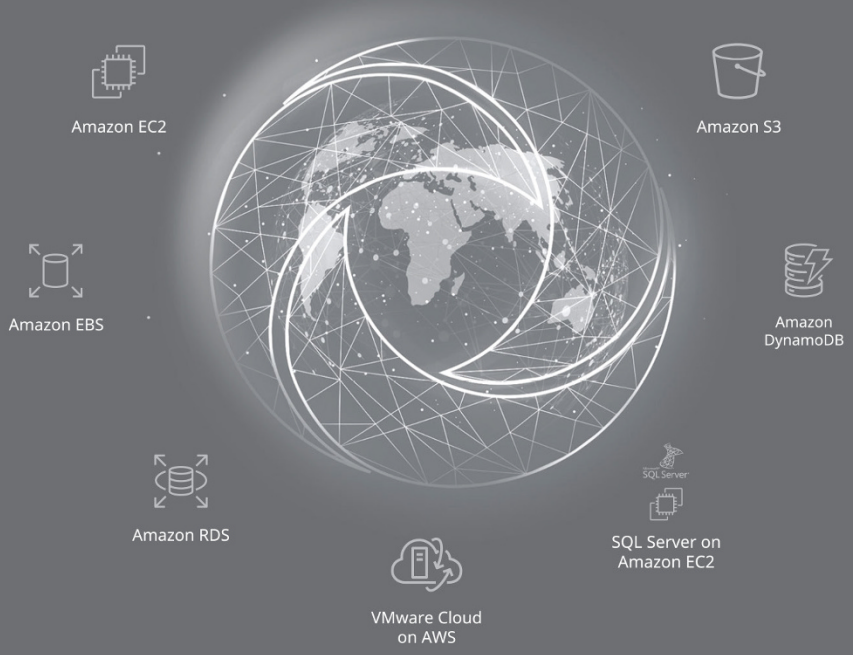
Choose a vendor with a simple, straightforward pricing model that requires little or no complicated cost estimating.

Automated Proactive Support

Support is another aspect of SaaS that buyers often fail to consider up front but ultimately experience when an issue arises. A common model is to provide a basic level of customer support free with the product, which will typically come with minimal hands-on support and multiday response SLAs. Faster, more intensive support often comes with a premium price tag. Either way, customer support is often customer-led, meaning it's the customer's responsibility to reach out if they need help with a problem.

A new and innovative way of handling customer support is to offer automated proactive support. In this model, the SaaS provider automatically monitors customer accounts for anomalies and failures, proactively reaching out to the customer while starting the resolution process immediately. Clumio provides this service at no additional cost, resulting in customer issues being resolved quickly, with more than 80 percent of support tickets being opened by the automated process.

SIMPLE, SECURE BACKUP FOR AWS



SEE IT IN ACTION AT:
clumio.com/demo

CLUMIO

Build resilience into your data lakes, applications, and sensitive information

Organizations are increasingly leveraging Amazon Simple Storage Service (S3) to build modern applications, data lakes, analytics, and AI. But protecting, recovering, and auditing giant volumes of in S3 is proving to be a challenge for many companies. Modern data use cases powered by S3 require a fundamental rethink of backup and recovery. *Backing Up Amazon S3 For Dummies* introduces a better way to protect data in the cloud.

Inside...

- Learn to swiftly recover from ransomware and cyberattacks
- Master data compliance
- Understand different data protection methods in AWS
- Know what, when, and how to backup
- Gain a deeper understanding of backup costs

Lawrence Miller served as a Chief Petty Officer in the U.S. Navy and has worked in information technology in various industries for more than 25 years. He is the co-author of *CISSP For Dummies* and has written more than 200 *For Dummies* books on numerous technology and security topics.

Go to Dummies.com™
for videos, step-by-step photos, how-to articles, or to shop!

ISBN: 978-1-394-22106-6
Not For Resale

for dummies®
A Wiley Brand

WILEY END USER LICENSE AGREEMENT

Go to www.wiley.com/go/eula to access Wiley’s ebook EULA.