ActualTech MEDIA

CLUMIO

# Compliance and the Increasing Challenges of Cloud Data Lifecycle Management

Dan Sullivan

## IN THIS PAPER

Regulatory compliance is not a trivial task, especially when retaining data and demonstrating compliance. With more workloads moving to the cloud, businesses are learning that much of the responsibility for compliance still rests with them. Too often, an initial response is administrators write narrowly focused scripts that solve one piece of the compliance puzzle. This leads to more scripts being written and more scripts means more complexity. Systems are dynamic and a script written today can quickly become non-functional when the system changes. It becomes quickly apparent that complexity breaks compliance.

It's essential for organizations using the cloud to have access to simple, air-gapped data protection mechanisms that support compliance across the enterprise.

It's essential for organizations using the cloud to have access to simple, air-gapped data protection mechanisms that support compliance across the enterprise.

As information technology becomes more important to the operations of businesses and governments, there's a corresponding increase in regulations that specify the "rules of the road," and the adoption of public clouds is adding a new dimension to the challenges of staying compliant. Regulations typically require organizations to demonstrate they're in compliance with certain rules. These requirements include demonstrating backups are performed so that in the event of data loss essential information can be recovered, detecting problems so that issues are discovered preemptively, and proof of a remediation plan for when failures occur. Although you may have well-established practices for on-premises operations, they likely do not easily or completely map to requirements of public clouds.

Organizations must also be able to demonstrate rapid retrieval of data so that recovery operations can occur quickly to minimize service disruptions. Companies should have a disaster recovery (DR) plan in place, as well as specific and quantitative Recovery Point Objectives (RPOs) and Recovery Time Objectives (RTOs), and these need to be tailored to the way workloads are run in the public cloud.

In addition to these explicit regulations, companies are facing three common challenges with staying in compliance, which we'll explore in detail in this tech brief.

## CHALLENGE 1: DATA VOLUMES ARE GROWING

As businesses continue to focus on growth strategies, they're generating an increasing amount of application data and much of that data is stored, transformed, and analyzed in the cloud. For example, applications can capture each time a user interacts with a web page or mobile device and record their path through various parts of an application and that data is immediately streamed to a pipeline running in the cloud that performs real-time analysis. Telemetry data about application performance and resource utilization is being generated by applications and infrastructure on a continual basis and needs the scalability and availability of cloud infrastructure.

The scope of captured data is widening as well. Digital media, including audio, images, and streaming video, are now standard components in business data. Finally, machine-generated data from Internet of Things (IoT) devices, sensors, and mobile devices are creating orders of magnitude more information than traditional transaction processing data. Again, these changes in the way businesses are generating and using data are driving the adoption of cloud storage and analysis services.

> Organizations are finding that many on-premises, legacy systems are difficult to adapt to increasingly demanding regulations and do not efficiently scale to the volumes of data businesses are now generating in the cloud. This requires new cloud-native data protection solutions and procedures to remain compliant.

Much of the data businesses collect and generate are subject to regulations, such as personally identifiable information (PII), financial records, and healthcare information, as well as information about how that data is stored, processed, and used. Companies are engaging in digital transformation initiatives that enable customers, clients, and patients to self-serve and a side effect of this is that new types of data, including PII, is being collected and stored in the cloud.

Organizations are finding that many on-premises, legacy systems are difficult to adapt to increasingly demanding regulations and do not efficiently scale to the volumes of data businesses are now generating in the cloud. This requires new cloud-native data protection solutions and procedures to remain compliant.

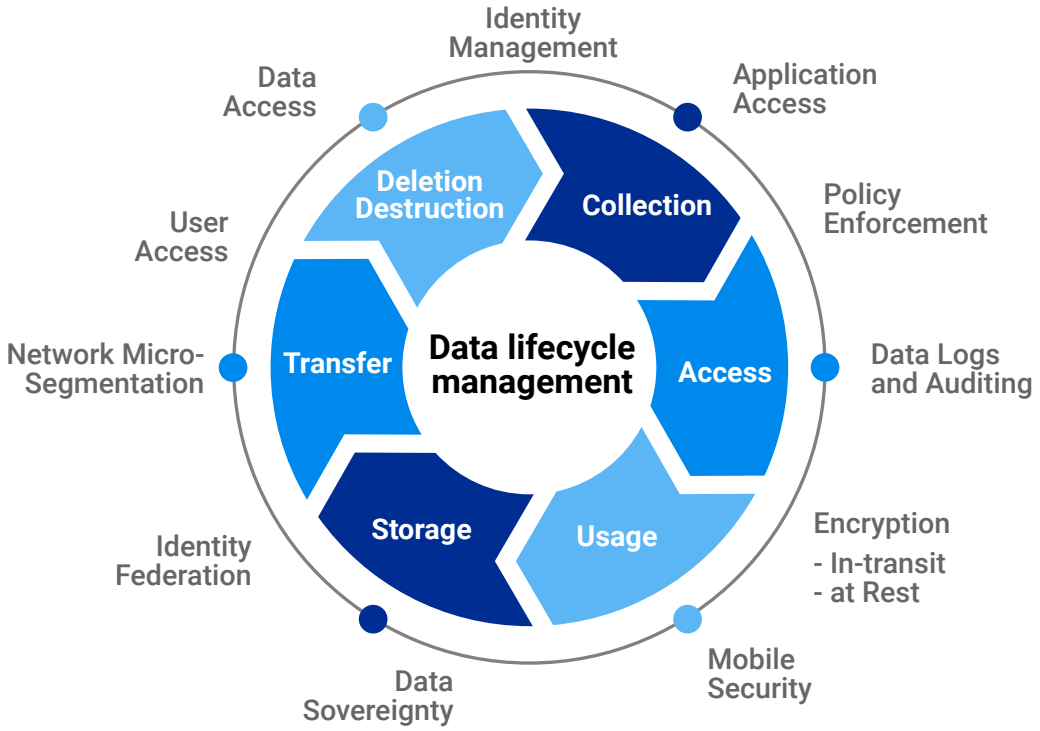


**Figure 1:** Data lifecycle management helps an organization define and organize the cycle through which data must go to get the most out of it until it's deleted

## CHALLENGE 2: DATA LIFECYCLE MANAGEMENT COVERS MULTIPLE DIMENSIONS OF CONTROLS

Data management is more than making sure there's sufficient storage infrastructure to store data (see **Figure 1**). It requires you to ensure that data is available, protected, secured, and stored efficiently.

Data must always be available when users want to access it, but failures do occur. Unfortunately, data loss due to account compromises, human errors (someone accidentally deletes a critical dataset, for example), and bad actor-related issues are all too common. Organizations need data protection, and how quickly one can recover from such issues and how much data loss can be tolerated dictates the frequency of backups.

One important decision is determining how to efficiently perform the backups to support your availability and long-term data retention requirements without exploding your budget. It's also important to keep in mind that different types of data require different levels of access controls. For example, PII is subject to stricter access controls than publicly available information. So, you need a data protection solution that helps you perform data classification and apply backup policies based on that classification.

Different kinds of data also have different compliance requirements. Some data might need to be retained for a set number of years or deleted after some amount of time. This is typically different for different classes of data adding to the complexity of data lifecycle management. Typically, taking advantage of different classes of storage helps from a performance and cost perspective, but compliance typically dictates length of retention.

> One important decision is determining how to efficiently perform the backups to support your availability and long-term data retention requirements without exploding your budget.

In most cases, data is used soon after it's created. As data becomes older, the chance of it being needed typically decreases. However, you may want to warehouse it for ML purposes or archive it for compliance requirements.

In these cases, you'll need policies to control data migration from hot, to nearline, to archive storage. Ideally, all aspects of data lifecycle management mentioned in this section should be automated to make scaling possible while remaining cost-effective.

Also, compliance dictates who can and cannot access each class of data. This in turn drives the need for Role-Based Access Controls (RBAC).

## CHALLENGE 3: DEMONSTRATING COMPLIANCE WHEN USING OPERATIONAL BACKUP AND RECOVERY WITH AWS SNAPSHOTS

Two challenges that constantly confront businesses adopting the public cloud are ensuring they can quickly demonstrate compliance, and the complexity of automating data compliance management. Common practices, like using built-in data recovery cloud services, are only a partial solution to these challenges.

> Two challenges that constantly confront businesses adopting the public cloud are ensuring they can quickly demonstrate compliance, and the complexity of automating data compliance management.

Data security and compliance in the cloud is the shared responsibility of cloud providers and their customers. While cloud providers will attend to physical infrastructure security and provide security building blocks, like access control systems and encryption, much of the responsibility for compliance rests with cloud customers. In particular, cloud customers create policies around data protection and choose the platforms and services that implement security and compliance controls.

Businesses often must comply with multiple regulations. Some, like SOC 2 and PCI, cover a broad range of industries, while others, like HIPAA/HITECH are targeted to specific industries such as healthcare. In addition, industry-oriented regulations, local regulations such as the California Consumer Privacy Act (CCPA), and Europe's Global Data Protection Regulation (GDPR) specify protections for citizens in their respective jurisdictions.

Compliance-based long-term data retention is mainly achieved via backup solutions. AWS customers often use snapshots for creating those backups. While snapshots help to recover from operational errors, they are not designed to implement or provide a comprehensive data protection solution that includes compliance and long-term data retention. IT teams could develop custom scripts to implement compliance-based data protection with snapshots, but this approach has significant drawbacks: Scripts need to change over time to accommodate changes in cloud services or application requirements, there may be errors in the script that aren't discovered until they complicate operations, and scripts may be used inconsistently by IT admins.

It's also important to keep in mind that the complexity of these scripts tend to escalate over time, which also leads to an increase in maintenance costs. A lesson that's learned repeatedly by organizations is that complexity breaks compliance. Compliance solutions should be simple, policy-driven, and cost effective. Obviously, they need to be secure, too. Snapshots are subject to tampering if an admin account is attacked so compliance practices should be built on an air-gapped platform that is resilient to tampering.

Another basic challenge for AWS customers is understanding if their assets comply. There are no reports, dashboards, or similar reporting services available that make it easy to get this critical information. It is, however, essential that businesses can report on how their assets are in compliance.

There are additional operational challenges to keep in mind when planning for data backup and recovery. The cost of long-term data retention can be quite steep, and determining cost-effective retention policies can be

complicated when it's based on data classification (for instance, HIPAA-covered data versus personal data/PCI-covered data). Data in backups can also be quite difficult to retrieve selectively without an index or catalog. This difficulty can slow recovery, which is already a complex operational process.

## THE CLUMIO ADVANTAGE

Fortunately, solutions to the ever-growing challenges of data management and compliance *do* exist. One such solution is Clumio, a data protection as a service that provides comprehensive and centralized backup plus reporting in AWS, allowing its users to demonstrate continuous compliance (see **Figure 2**).

Clumio not only provides support for file indexing, cataloging, and searching, but does so at significantly lower cost than off-the-shelf tools such as snapshots. With Clumio, you can easily search and find what you want to restore, whether to demonstrate compliance or perform recovery from a disaster situation.

Clumio's solution combines the ability to automate compliance-based backups and retention services, support for necessary reporting and monitoring to demonstrate compliance, and recovery services to restore critical data within RPOs and RTOs.

Clumio's data protection services bring enhanced security to your data protection operations, beginning with secure, fully air-gapped backups that provide protection against account compromises such as ransomware attacks and bad-actors. Clumio backups are immutable, and there's no ability to delete data once it has been stored in Clumio.

> A lesson that's learned repeatedly by organizations is that complexity breaks compliance.

In addition, Clumio uses strong encryption for both data at rest and data in motion. Access to Clumio services and data is restricted by a combination of multi-factor authentication and authorizations based on RBAC and Entity-Based Access Control (EBAC).

Clumio's solution is cost-effective and predictable. The SaaS pricing model is based on OpEx charges, so there are no upfront capital expenditures to implement a data protection service that can apply to all assets in an enterprise.

One of the most challenging aspects of data protection is that new data is constantly generated, sometimes

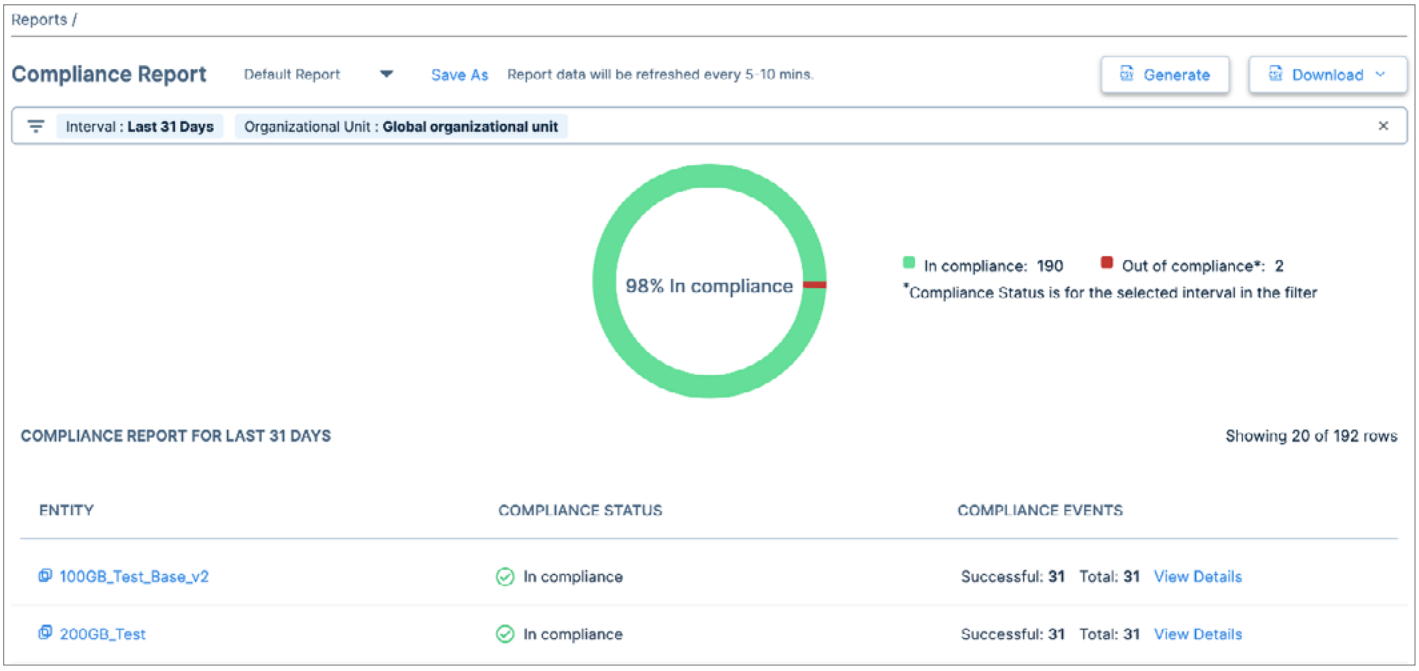


**Figure 2:** The Clumio centralized compliance dashboard

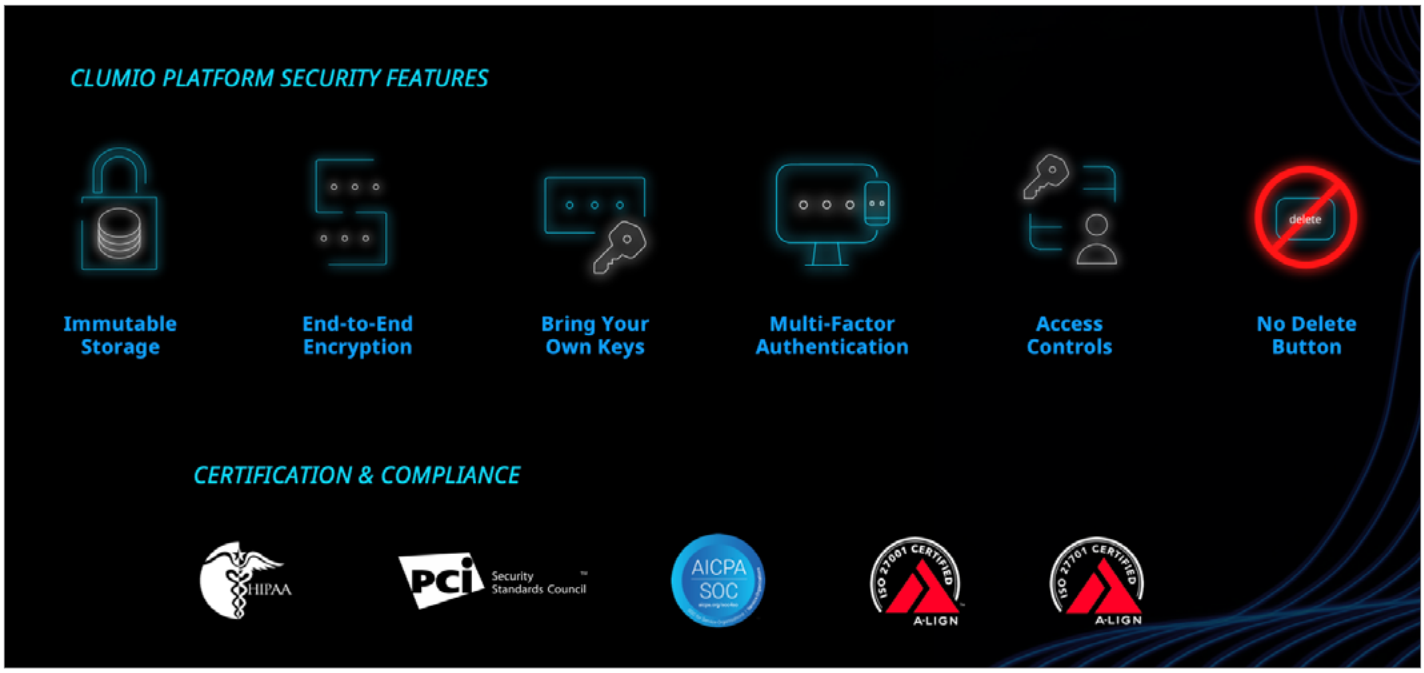


**Figure 3:** Clumio provides ultimate security

under multiple AWS accounts. Clumio automates compliance-based data retention with global policies. In addition, Clumio can help with discovery and addition of new AWS accounts and start applying already established global compliance-based data protection policies. These newly discovered assets are listed in real-time dashboards, so you'll always know what assets are protected. This ensures that as data keeps growing, you'll continue to protect your assets without gaps.

> With Clumio, you can easily search and find what you want to restore, whether to demonstrate compliance or perform recovery from a disaster situation.

Working with multiple AWS accounts can increase the complexity of data management, but Clumio addresses that complexity by seamlessly working across multiple accounts and regions. You can, for example get customized compliance reports for all assets across the entire AWS environment, as well as receive instant alerts when compliance may be at risk. Clumio enables granular and rapid recovery of data, which is a key factor to proving compliance and passing audits. This also helps ensure you can meet low RTO SLAs.

Clumio's commitment to protecting your data is validated by multiple rigorous certifications including ISO 27001, ISO 27701, SOC 2 Type 2, HIPAA, and PCI DSS (see **Figure 3**). This rigorous testing makes Clumio one of the most secure platforms in AWS and ensures that customer data is stored adhering to their compliance requirements. Efficient storage follows from eliminating redundancies and unwanted replication of data in backups.

Learn more about how Clumio can help you manage the increasing complexity of maintaining compliance in the cloud while meeting challenges of cloud data protection by visiting clumio.com.