CLUMIO

THE DEFINITIVE GUIDE TO BACKING UP DATA IN AWS

Cloud is officially the present—and the foreseeable future

01

Cloud is the way customers operate today. Being shackled to a physical site in order to run day-to-day operations is no longer an option, as the new world requires the ability to efficiently scale up and down on demand. It has become the center of gravity for all computing innovation, enabling companies of all sizes to move faster without having to manage and worry about IT infrastructure.

The cloud is not just the modern computing stack, it is also the modern data stack. Companies are building petabyte-scale data lakes, data warehouses, data visualization and analytics engines, and powerful machine learning and AI applications in the cloud.

Amazon Web Services is at the center of this seismic shift in data-driven businesses, and it's hard to think of a more influential company in this space. It has enabled large ecosystems of innovation, birthed billion-dollar companies, and given rise to entirely new industries.

But this breakneck pace of innovation has left a key vulnerability in its wake. If entire industries are built on data, shouldn't we prioritize protecting that data? All cloud providers operate on the shared responsibility model: the cloud provider guarantees resilient infrastructure. But the integrity, safety, and resilience of your data? Well, that's your problem.

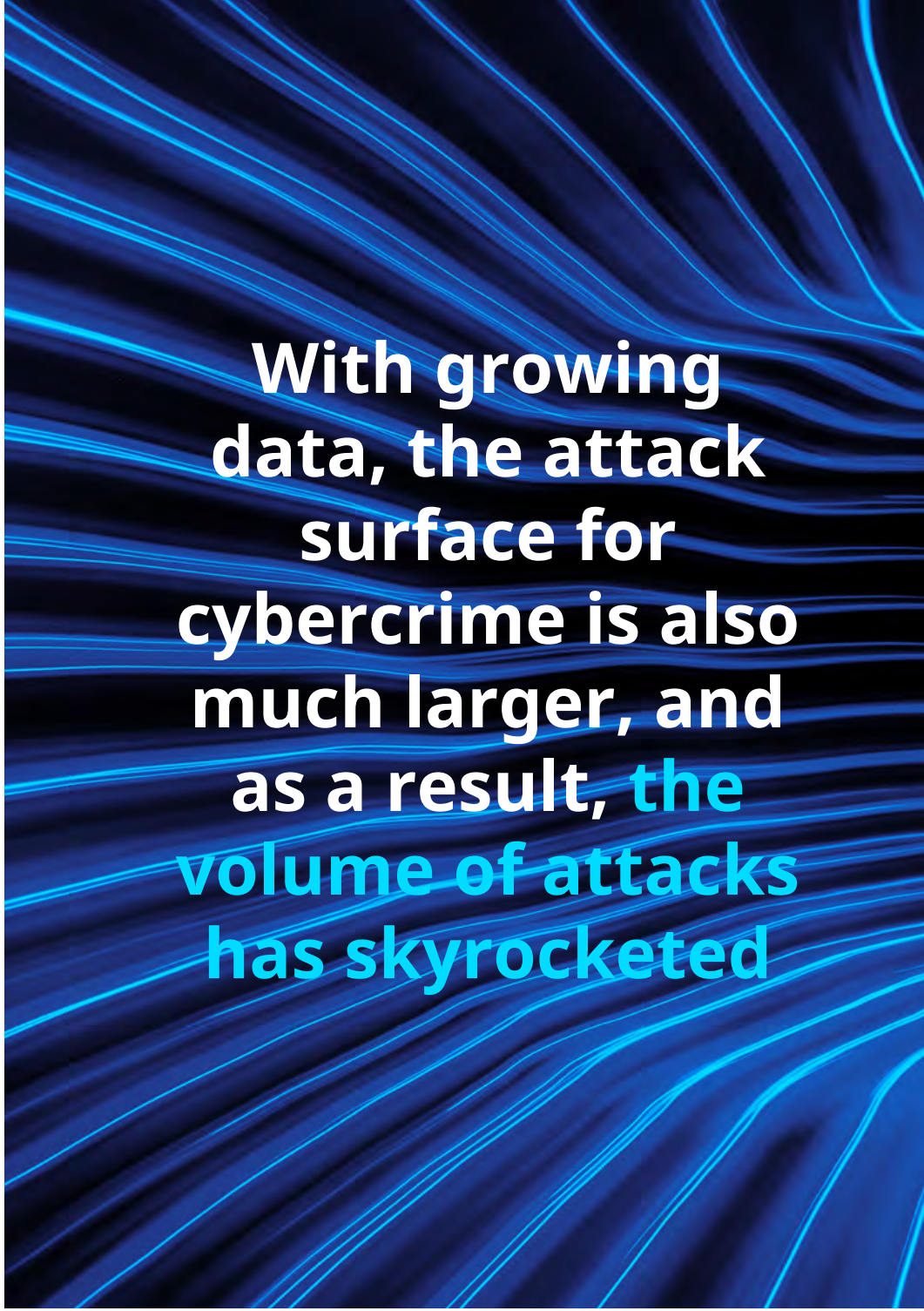
The challenges with protecting data in the cloud

Most companies shifting from on-prem appreciate this, but it has to be said for the cloud-natives: you absolutely need a data protection, backup and recovery strategy in the cloud. In some ways, it is more important in the cloud, because of the volume, usage, and velocity of data. But these same reasons also make it more challenging.

Cloud applications scale much faster, resulting in quickly ballooning production data that requires protection. Cloud data is also more scattered — across applications, accounts, regions, and platforms.

The rapid adoption of cloud data warehouses and data lakes has also massively increased the potential for something to go wrong with production data.

From a data protection standpoint, organizations today find themselves reacting to crises, breaches, and data loss, rather than proactively operationalizing a holistic cloud data protection strategy.



How do you backup your data in AWS?

Many organizations begin in the cloud with “shadow IT” projects or through digital transformation initiatives. Cloud migration is part of that strategy. Arguably, no company makes that easier than AWS, where you can spin up some instances and start migrating databases to RDS.

Well, that data which was safely protected with a traditional backup solution on-prem, is no longer protected. It can't be expected that data engineers migrating workloads are even thinking of backup and recovery — it's probably not even part of their vernacular! The assumption is that everything has 11 9s of availability (spoiler: your data doesn't). Despite AWS's developer-friendliness, when it comes to data protection, it can be difficult to know where to start in the cloud.

A seemingly convenient path at this point is to leverage AWS native services to protect your data. One of the most common ways organizations start implementing data protection for AWS is using its snapshot management service.

At first, snapshots appear reasonably priced and relatively easy to operate. Just to be sure, engineers go ahead and snapshot all EBS and RDS volumes in a few clicks. All good, right?

Well, maybe not.

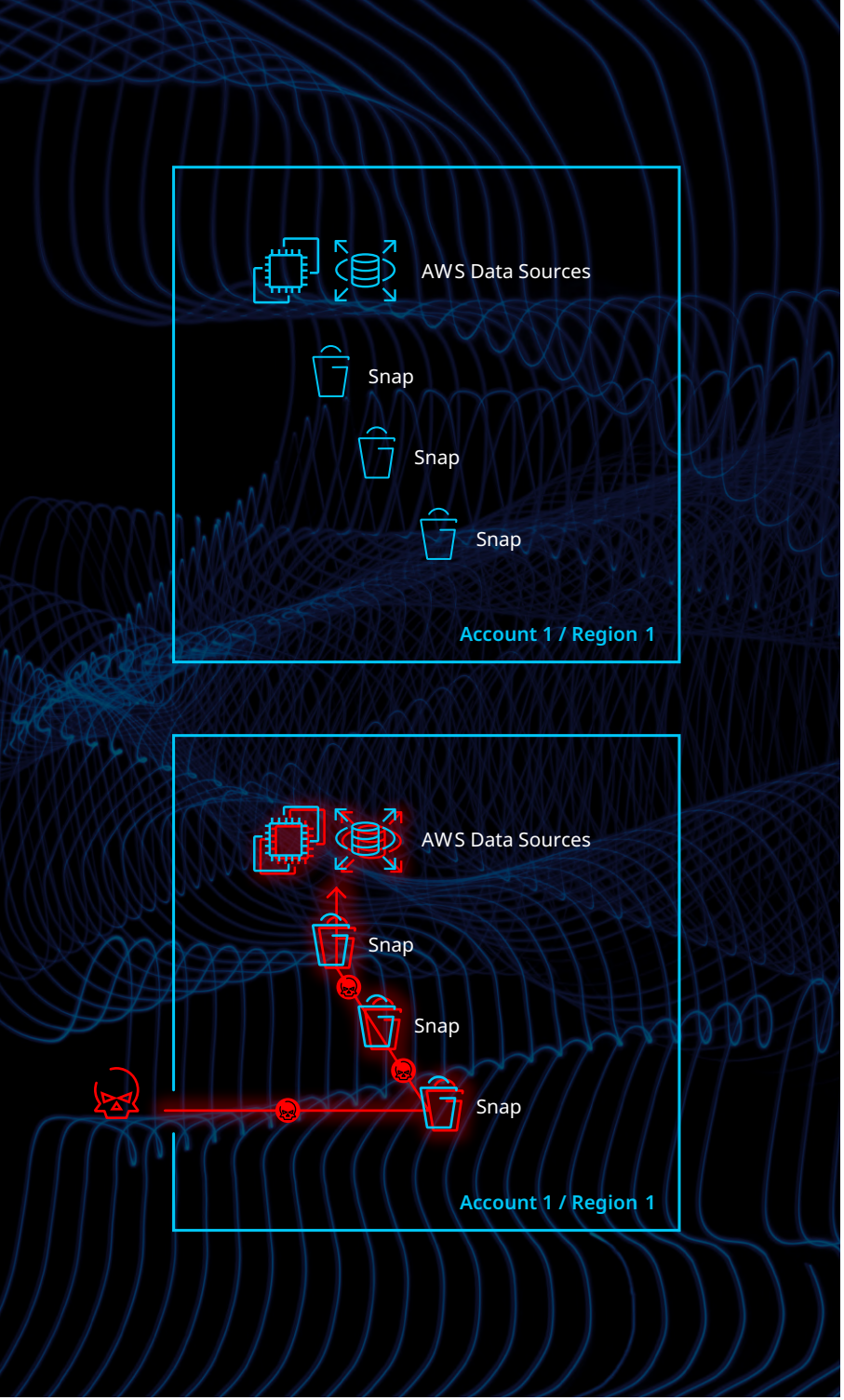
The challenges of backing up data in AWS

02

Unfortunately, it doesn’t take long for organizations to start seeing the limitations of using snapshots to protect their critical data in AWS.

While snapshots provide basic data protection capabilities—such as recovering from operational errors—recovering from snapshots in case of an emergency or attack is very, very complex. In the event of data loss, finding the right snapshot and getting dependent applications back up and running is nuanced and time consuming. Also, over time, snapshots can get very expensive, and the company has to make a choice between short retention and manageable costs.

Let’s take a closer look at the additional challenges of using snapshots as your only data protection tool.



Mind the (air) gap

Ransomware attacks are skyrocketing and every week we hear about enterprises with million-dollar IT budgets getting ransomed. For smaller companies, keeping up with data security and preventing ransomware attacks is even more difficult — so one can only imagine the state of affairs there. Ransomware costs are forecast to reach $10.5 trillion annually by 2025 and security experts estimate one ransomware attack to occur every 11 seconds. Therefore, it is of paramount importance to have the right backup solution to help thwart such attacks.

The CISA (Cyber-security and Infrastructure Security Agency) recommends air gapping your backed up data. This means independently securing and isolating it from an organization's security sphere. This ensures that the hackers cannot find the backup copy even after they have gained entry into the cloud account.

As shown, when organizations use native snapshots to protect data sources such as Amazon EC2, EBS, RDS, etc., the snapshots are created within the security sphere of the company — sometimes in the same account or region as the primary data sources! The problem with this approach is that there is no separation or air gap between the primary data and the snapshots.

If a hacker gets access to the enterprise security controls, they may be able to compromise snapshots along with primary data. Once that happens, there is no valid, usable backup copy that exists and hence no way to recover the primary data. This is precisely the situation organizations do not want to find themselves in and is a big limitation in data protection mechanisms in the cloud.

Long recovery times

Data protection comprises two fundamental functions: backup, and recovery. While it is important to ensure all critical data is backed up securely, it is equally important to have a solution that enables fast application recovery in the event of a data failure or compromise. Not being able to recover data results in disruption to operations, poor customer experience, and, in some cases, can even put organizations out of existence.

Recovery is about more than just restoring the backed up data; it is about the time it takes to getting the application back up and running. Recovering the right data and restoring the application from snapshots can take several hours if not days.

The recovery process is like finding a specific file in a box, with hundreds of such boxes in a warehouse. Sure, there may be some boxes labeled 'important files', but you have no idea what exactly is inside each box, forcing you to rummage through boxes and files. When you find the file that you think you want, you begin the restoration process by bringing out the whole box. Does this sound efficient?

STEP 1

Determine the file and time period for restoration

STEP 2

Locate Snapshot from that period for the EC2 Instance

STEP 3

Determine volume, size, region, and OS

STEP 4

Create EBS Volume of the same size/region from Snapshot

STEP 5

Create EC2 Instance with the same OS

STEP 6

Connect EBS Volume to EC2 Instance

STEP 7

Boot EC2 Instance

STEP 8

Mount Volume on EC2 Instance

STEP 9

Search for files to restore

STEP 10

Download files

STEP 11

Make sure to spin down, detach, and delete

Once you mount the 'box' to an instance, you load it and see if it truly is the file you wanted. If it isn't exactly the right version, you will have to repeat this process all over again. Overall, this can easily take several hours.

Let's look at a real-world scenario to illustrate this point: An organization is using DIY tools to protect its EBS volumes and now needs to recover a specific file from one of the volumes that is compromised. They will need to perform the following steps to recover the file.

Depending on how long it takes to find the right snapshot, identify the right OS, volume size, region, and finally create an EC2 instance that matches the original instance, this whole process can easily take several hours.

And these steps need to be performed for every file that needs to be recovered. RDS is even more complex as it requires the restore of an entire RDS instance to get access to the data, even if only a single record is needed. Finally, customers need to ensure that the resources that were spun up in the cloud to recover the file are spun down in order to avoid unwanted expenses.

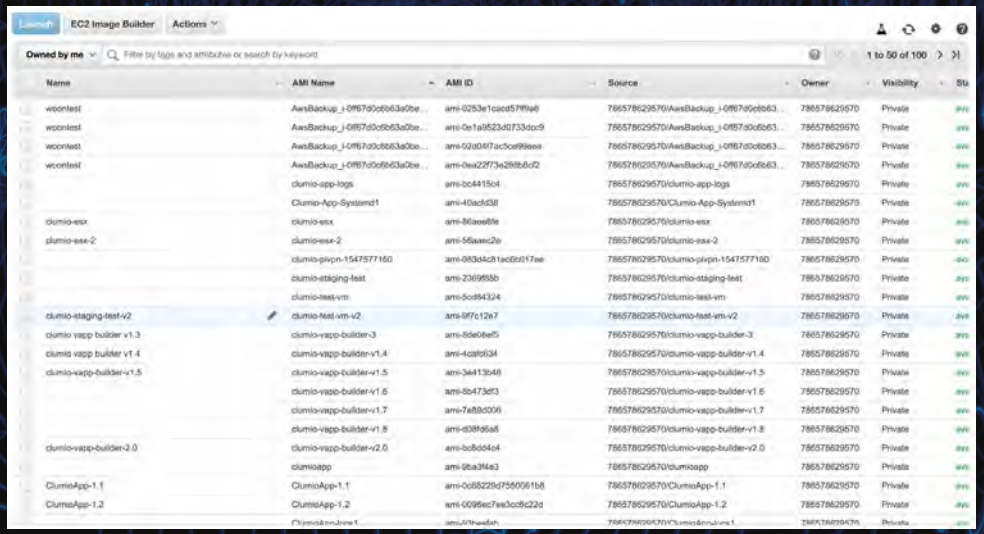
Limited visibility

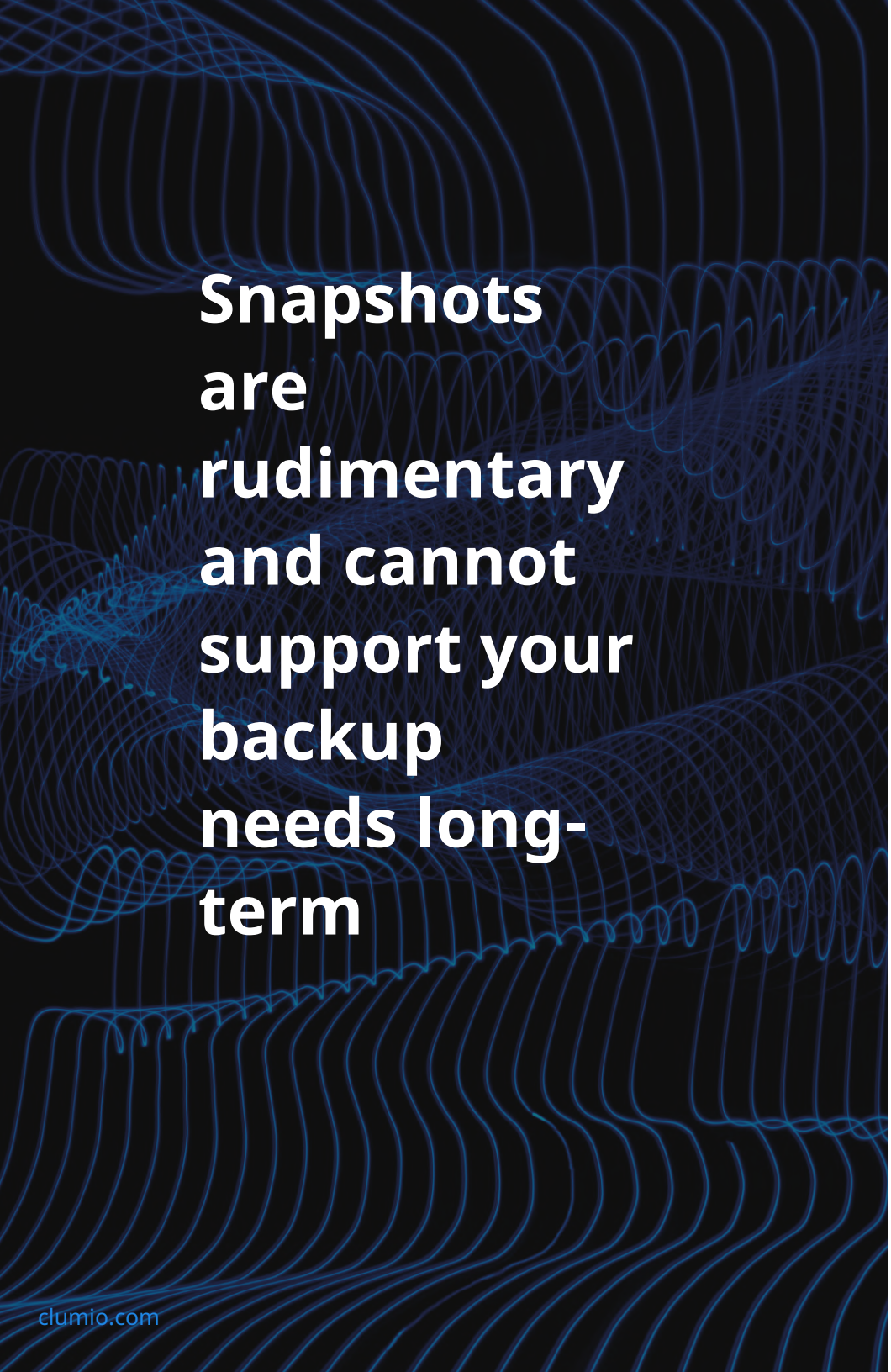
Setting data protection policies across all applications is not something you should do frequently. In fact, they should be set once with the right attributes and not tampered with to maintain compliance requirements. Naturally, it's not easy for cloud admins to remember every policy, backup history, and compliance requirement for each data source that is protected in their environment.

However, it is important for admins to have this information at their fingertips when needed. For example, they should be able to:

1. **Quickly prove compliance during audits**
2. **Easily find the right data during the restore process**
3. **Select the right policy to protect a new application or data source that gets added**
4. **Do all the above across hundreds of accounts**

Lots of snapshots to scroll through
It can get complicated and time-consuming to find the right snapshot to restore





Additional complexity

From these examples, it should be clear that snapshot-based backups provide very basic capabilities. Organizations that start using snapshots to back-up their AWS data soon run into these shortcomings and find themselves forced to respond.

They end up writing complex scripts to add missing functionalities into their data protection solution. They have to dedicate valuable IT resources to develop and maintain these scripts on an ongoing basis rather than having them focus on their core business. This is not in line with leveraging the cloud to add agility and enable faster innovation in your business.

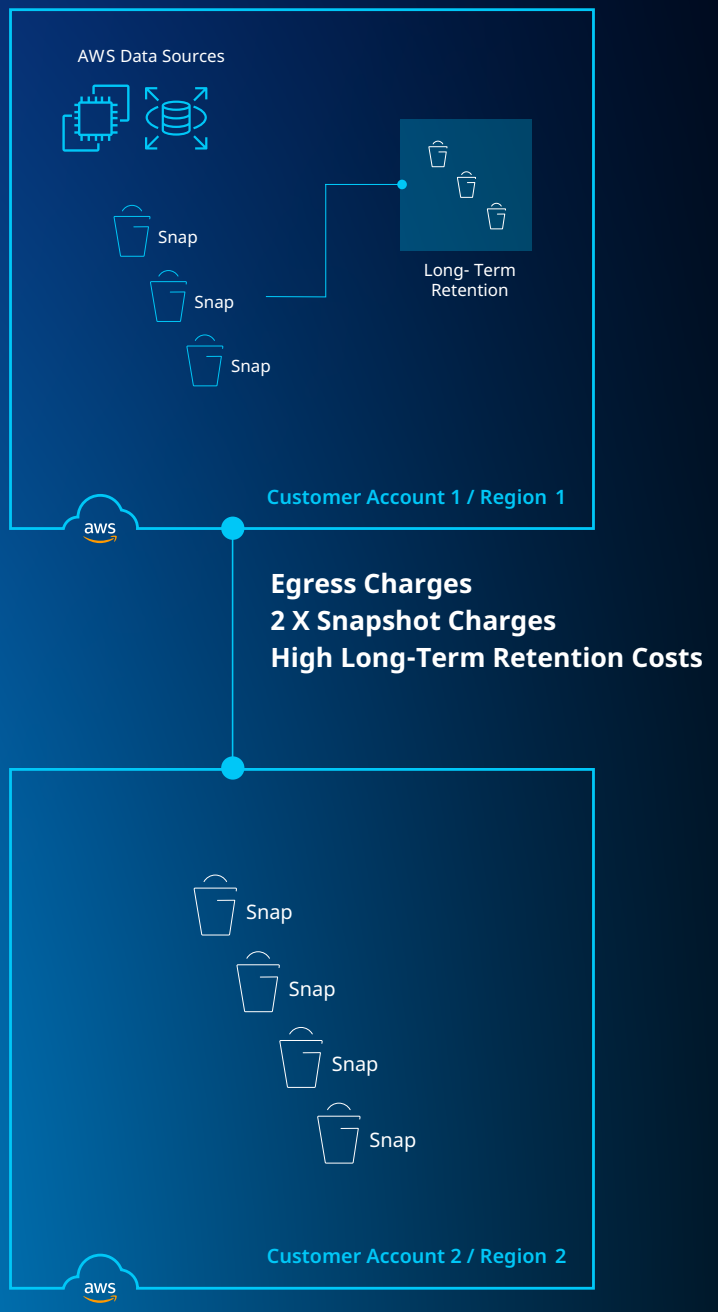
Watch out for a large backup bill!

As part of a data protection strategy, it is typical for organizations to set up long-term retention for production data as well as protection against account compromises due to attacks. To accomplish these two objectives, organizations do the following:

1. **Create snapshot based long-term retention that results in creating multiple snapshots per account in high-tier storage**

2. **Replicating snapshots from one account to another account to protect against account compromises that results in doubling the number of snapshots. There are additional egress charges for the cross-account transfers**

In a typical scenario, a few months go by and an organization is starting to realize a concerning trend. Their AWS bill has steadily increased and doesn't seem to be showing signs of leveling out.

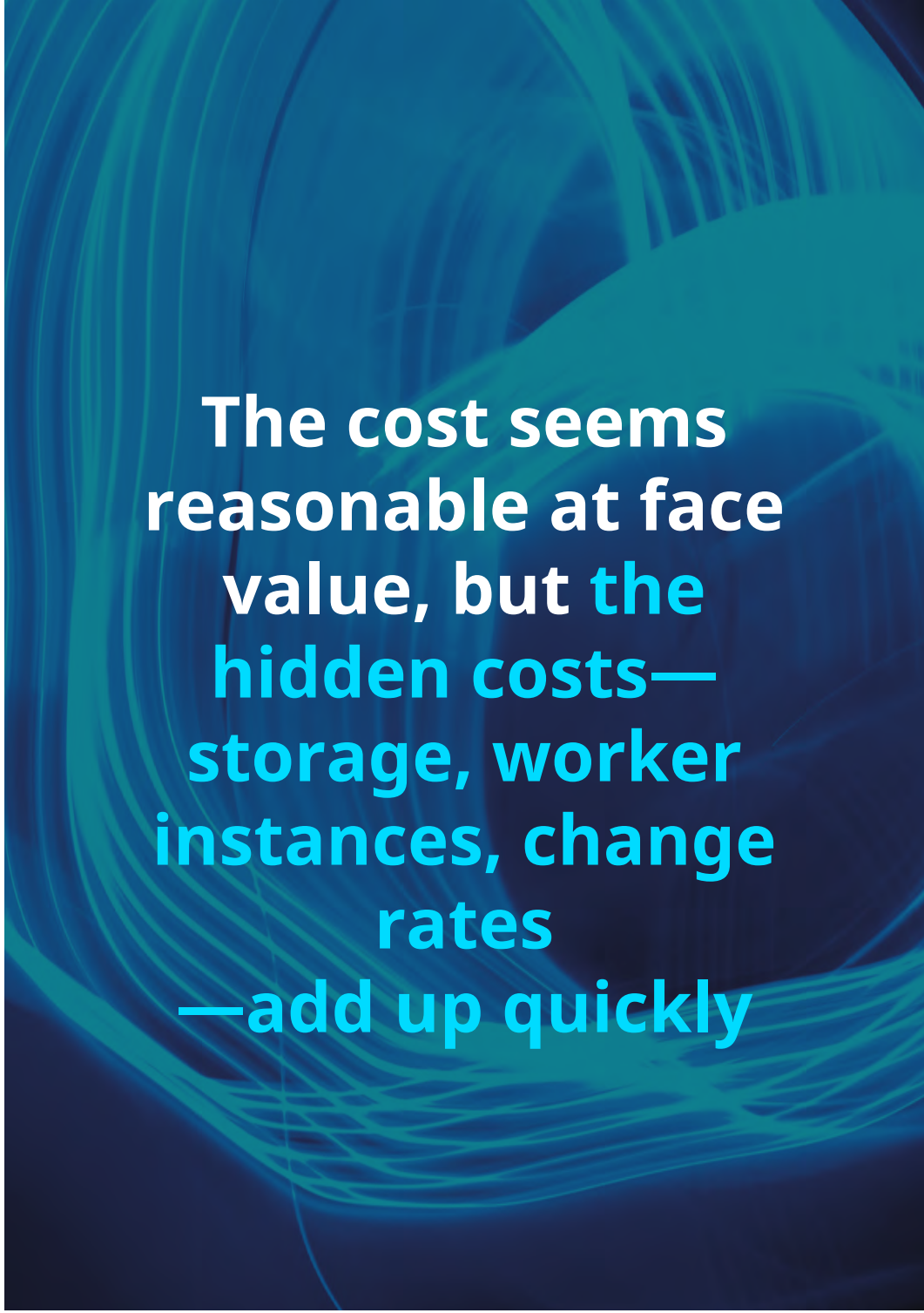


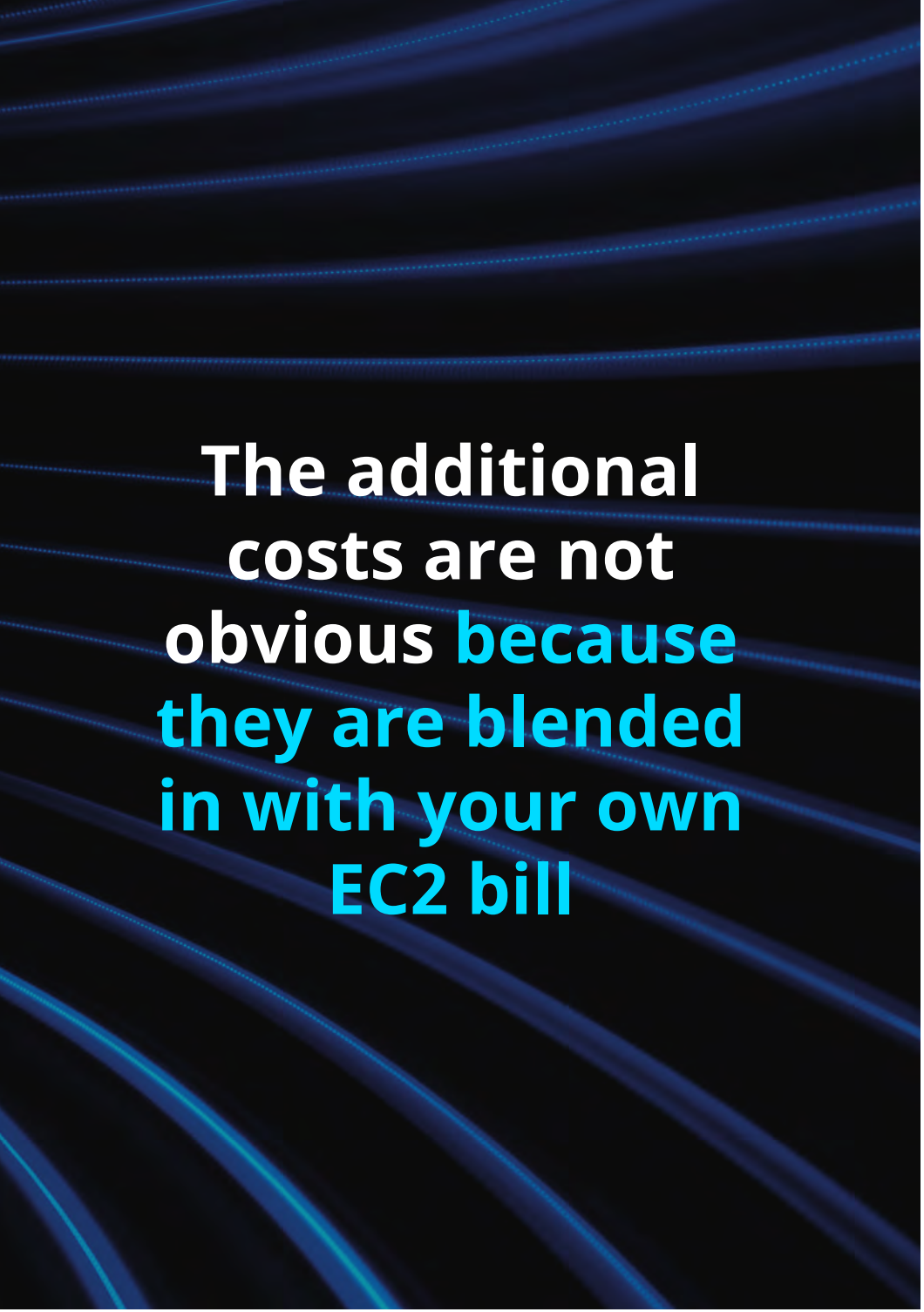
The challenge of using third-party snapshot managers

As organizations grapple with managing the cost of their backups in AWS, they start to look at third-party snapshot managers. These snapshot managers talk about helping reduce organizational costs significantly by being able to tier to Amazon S3.

The cost seems reasonable at a few dollars per protected EC2 instance per month, but the total cost is actually much higher, since it includes the additional cost that the organization must incur to store the data. To help understand the implications of this license fee for these solutions, let's look at an example. If an organization has an average of a 200GB EBS volume per EC2 instance at a face value cost of $5 per protected instance per month, that has a daily backup with a 3% daily change rate, it adds $0.013/GB/month to their first month's EBS Snapshot costs of $0.050/GB/ month. That's over a 25% premium over native AWS Snapshot costs.

And if their EBS volume is smaller, this premium effectively increases. On top of that they also run an in-account EC2 backup manager instance and temporary worker EC2 instances that add additional usage to the AWS infrastructure and are blended into the bill.





Even more hidden costs!

But do organizations actually save money on the tiering-to-S3 feature? In order to offer this capability, third-party snapshot managers spin up temporary EC2 instances that run for longer periods of time and add a hidden cost to your bill. The costs are not obvious because they are blended in with your own EC2 bill. The temporary EC2 instances have to traverse through a company's EBS Snapshots and copy chunks over to S3.

The break-even on just moving it to S3 tends to be around the three-month mark due to all these hidden costs. This means if a business has retention periods less than three months for their daily backups, they could actually be spending more than if they had just left it in EBS snapshots. Add in the licensing costs we talked about earlier and the cost climbs even higher.

And because EBS snapshots are incremental, it's hard for the S3 tiering to effectively move only what's needed. This is because daily snapshots (that shouldn't be moved to S3) may point to older blocks that are part of a yearly backup that were already moved to S3. Due to this complexity, it is likely the backup manager has a lot of your data in both EBS snapshots and S3.

What about custom scripts?

Management sees just how much is being spent on AWS backups, especially the third-party manager, which convinces them to dedicate a shared engineering resource to help develop custom backup scripts that can handle the organization's backup requirements.

The organization eventually gets what it needs up and running and they are license-free and finally feeling they have a handle on their backup policies. But as AWS changes their APIs and business needs evolve (like needing to backup cross-region or protect backups from ransomware), so must scripts. The shared engineering resource becomes a full-time role. And now the business has a not-so-hidden cost for managing, refining, and creating custom backup scripts.

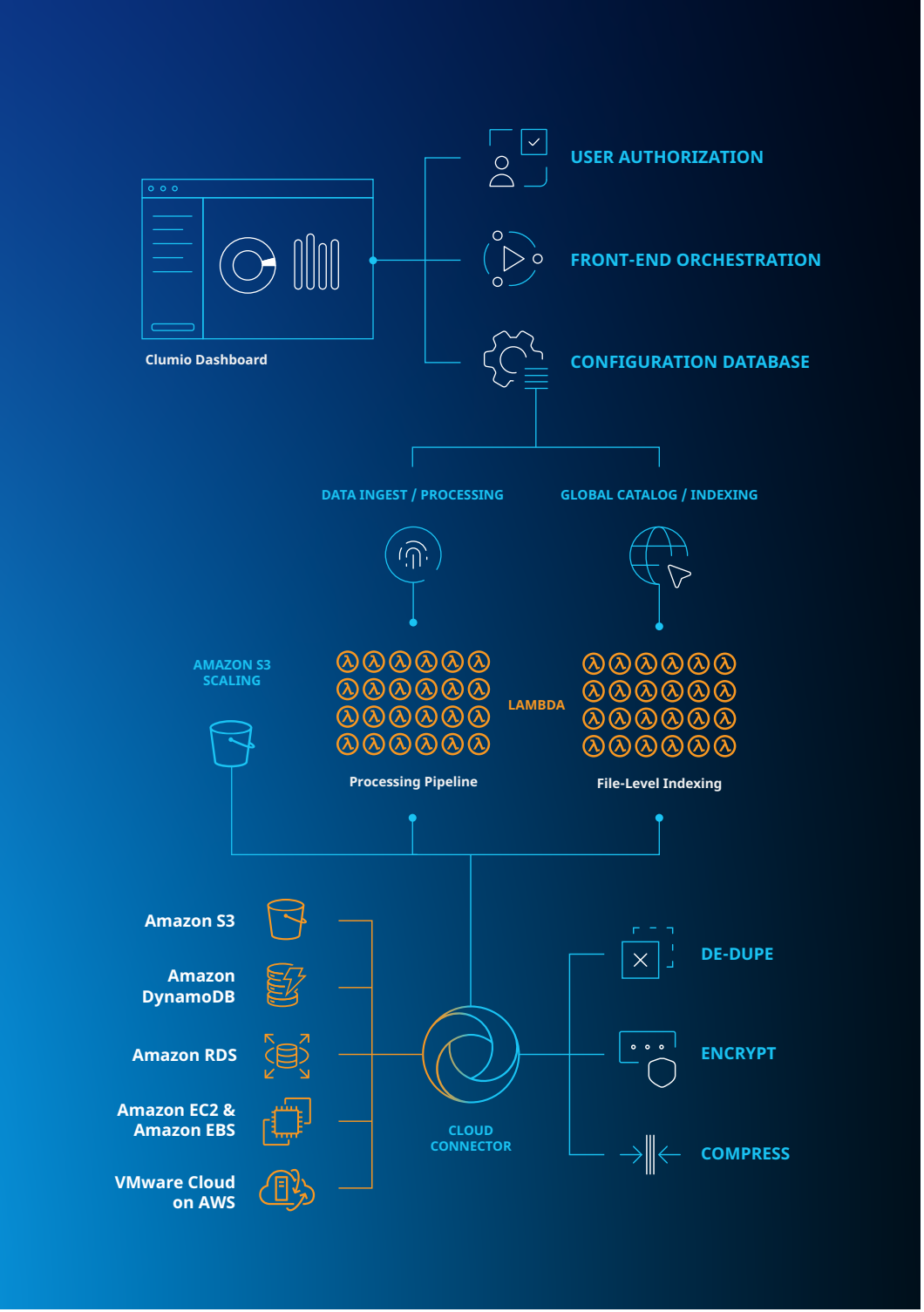
03

Clumio’s approach to AWS data protection

From all the challenges discussed in the previous section—long recovery times, no air gap, lack of visibility, mounting costs—one could conclude that effective data protection in AWS can be extremely challenging. Yet AWS provides undeniable advantages for businesses harnessing the power of the cloud.

It is increasingly critical that organizations find and implement the right data protection solution to fully leverage the benefits of AWS. To ensure it is done right, organizations need to implement a solution that addresses each of the key challenges of protecting AWS data today.

At Clumio, we have deeply evaluated the limitations of backing up cloud data and created a cloud-native solution that leverages cutting-edge AWS functions and addresses these challenges. Let’s take a look at how we do it.



Ultimate architecture for ultimate scale

Clumio protects a broad range of workloads — including S3, EC2/EBS, RDS, VMC, and DynamoDB. In addition, Clumio also has the highest scale for any backup technology in the market today. For example, Clumio Protect for Amazon S3 can scale to a staggering 20 billion objects, measuring hundreds of petabytes. No matter how much data customers need to protect, Clumio delivers.

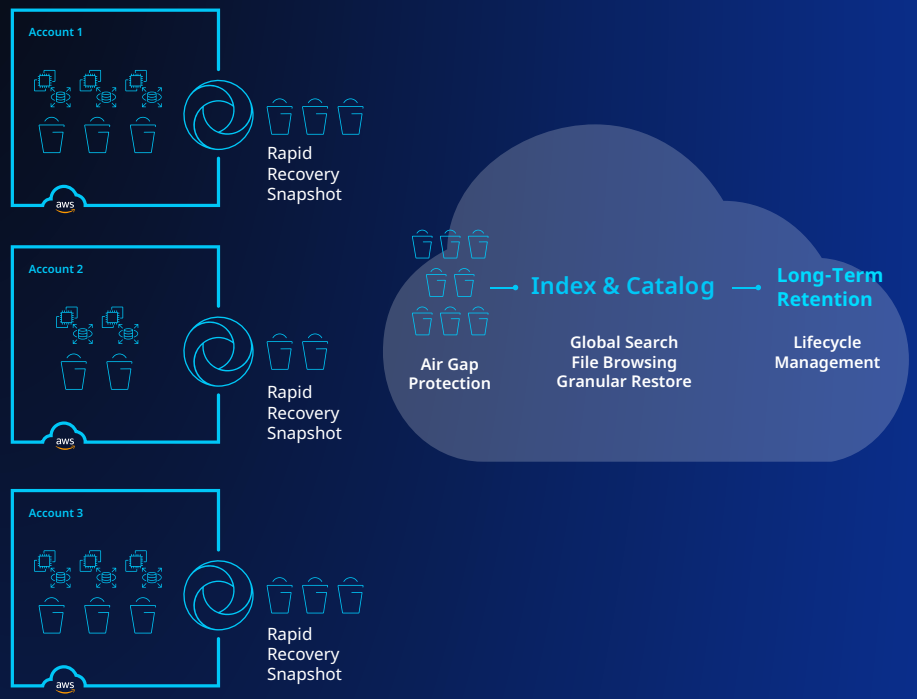
How is Clumio able to do this? Clumio has absolutely no dependencies on legacy backup constructs, and is built with decoupled cloud resources to scale dynamically with demand. Clumio's data processing pipeline is stateless and built on Lambda functions. During a recovery operation, Clumio automatically scales to increase restore throughput, and Clumio's intelligent indexing optimizes the restore process further. The speed at which Clumio can ingest and recover data is unmatched.

Always-on security

First, to guarantee that backups are valid and usable to initiate a recovery process when primary data is compromised, it is necessary to ensure that backups are saved outside of the security sphere of the primary data. This separation is called an air gap. By air gapping the backups, hackers or bad actors cannot access it, thereby leaving an open path to successful recovery from any account compromise.

Given the rise of ransomware attacks, combined with the large and decentralized attack surface of the cloud, organizations should pay extra attention to the security posture of their cloud data protection solution. The solution should deliver:

1. **Air gap backup**
2. **Immutable backups, so that the backup copies cannot be modified even if bad actors somehow get access to it**
3. **No "Delete" option for backup data. This combined with immutable backups ensures that the backup data is well secured**
4. **End-to-end encryption of user data, in transit and at rest**



Air Gap + Long-Term Retention

Ransomware and bad actor protection

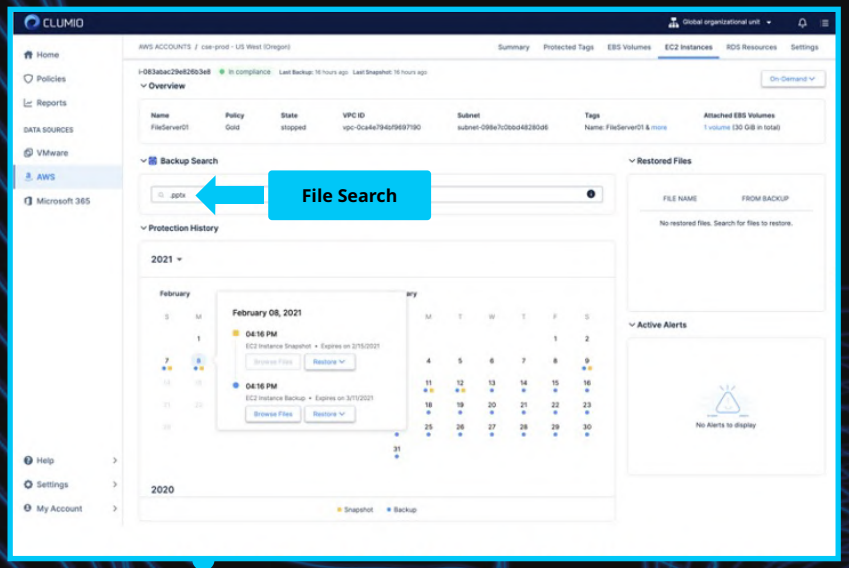
Immutable backups

No delete button

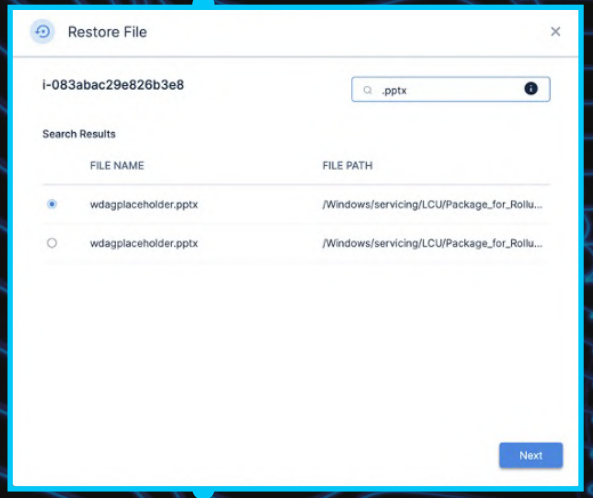
Extreme compression and archiving for cost savings

Rapid recovery

When it is time to recover from a data failure, one should be able to do so quickly to ensure business continuity. The right data protection solution must provide a quick way to find the data (snapshots, instances, files, records, etc.) that needs to be recovered and then restore it. Here is how Clumio enables rapid recovery for files in AWS EC2.



Step 1
Type in a search term for the file you want to recover

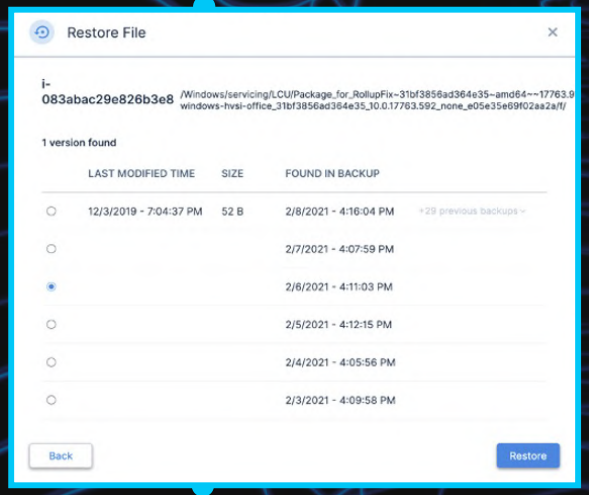


Step 2
Select the file you want to restore

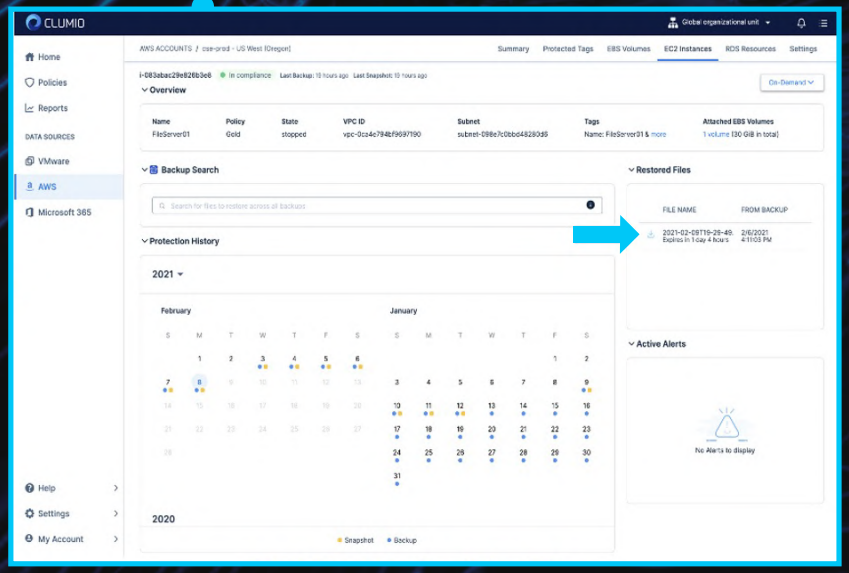
Clumio’s calendar view allows an organization to efficiently browse the entire file system to locate the exact data that needs to be restored, amidst petabytes.

Instead of having to load or restore a file and then confirm it is the correct one, an organization can identify the needed file just by typing in data parameters and conducting a search. The user will see all the different versions stored, with time stamps, and can easily identify and restore whatever file or files are needed. Clumio significantly reduces recovery time essentially by letting an organization search its entire warehouse of file boxes without having to open them up first.

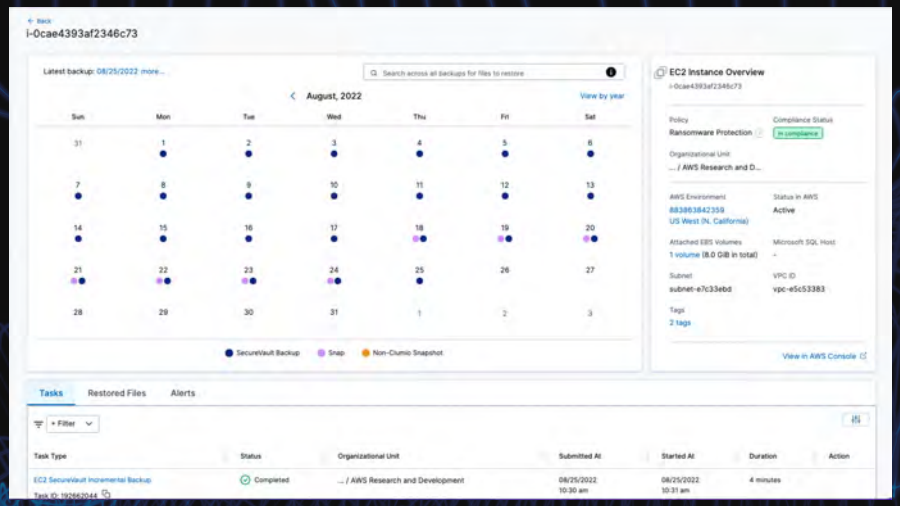
This easy to use rapid recovery process has helped Clumio’s customers reduce average recovery times from over 4 hours to less than 10 minutes.



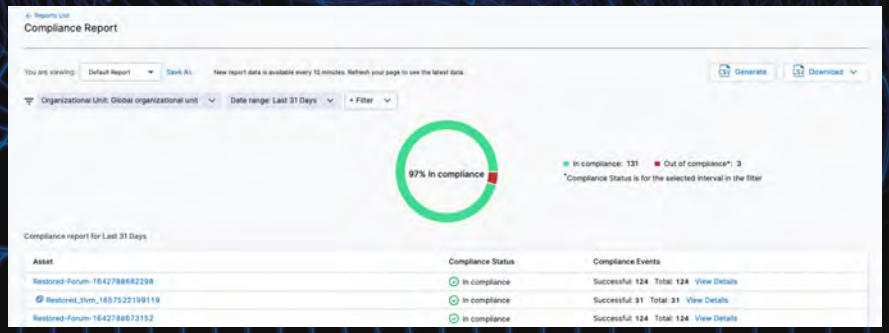
Step 3
Select a version you wish to download



Done!
Download the file



Backup History Calendar View
Find any backup quickly for fast restore at any point in time



Global Compliance Reporting
Single source of truth for audits and compliance

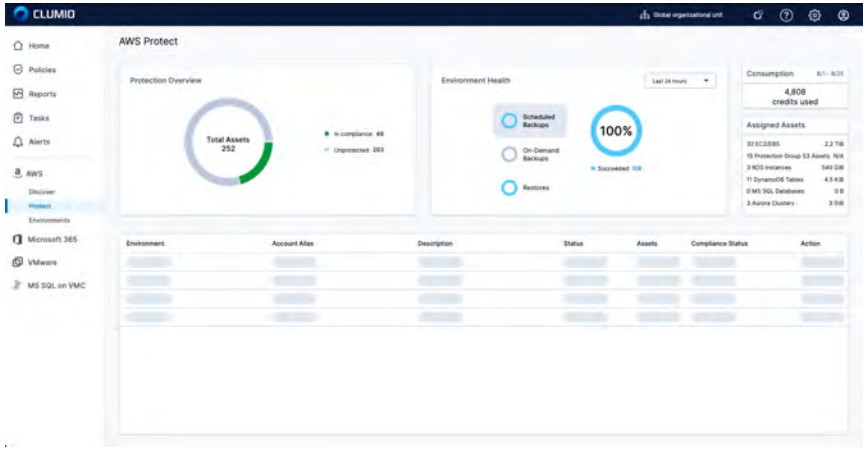
Better visibility

A key requirement for a good data protection solution is the ability to provide the necessary information about protected assets in a simple, consumable manner. Requiring the user to develop custom reports and dashboards to view assets protected by data source, policy, global cross-account status, or to understand whether they are meeting compliance requirements, is a recipe for introducing errors. This approach will end up consuming valuable resources on an ongoing basis. Instead, these tasks should be built into the data protection solution.

Clumio's calendar view presents a global understanding of all the backups created for an AWS data source. You can then easily perform a point in time restore of the entire data source or view individual files/records to perform selective restores. Clumio’s environment dashboard and the compliance report provide real-time compliance status for selected accounts or the entire environment. This ensures that the security team is on top of their data governance requirements and always audit ready when the need arises.

Simpler management

Organizations should be looking for cloud data protection solutions that simplify and automate backup orchestration. It should enable simple onboarding, quick backups, global policy setting and ease of recovery.



With a SaaS data protection service such as Clumio, it takes minutes to start protecting AWS data sources and you get complete control over your backup and recovery operations. This simplifies and automates day-to-day tasks for the IT and Operations teams, freeing up time to focus on their core business.

Instance Restore
Restore an entire Instance to any account

Browse and Restore
Browse the file system to restore a single file

Granular Record Restore
Scheme browser with Record recovery w/ SQL Query

Global Search
Search for any file across any Instance or Volume

Global Policies
Simplify policy setting across multiple accounts

Within 10 minutes an organization can click a button and back up their data in an air gapped environment

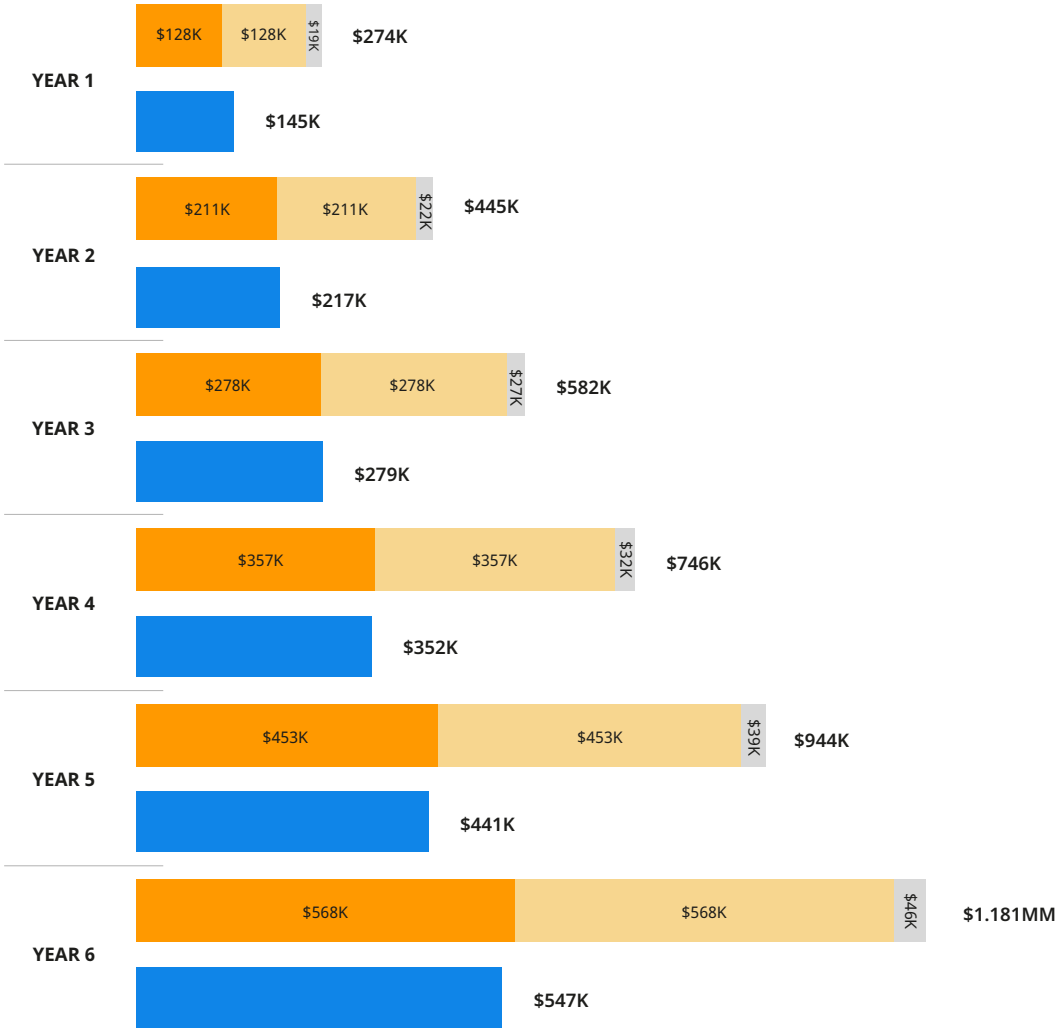
Lower TCO

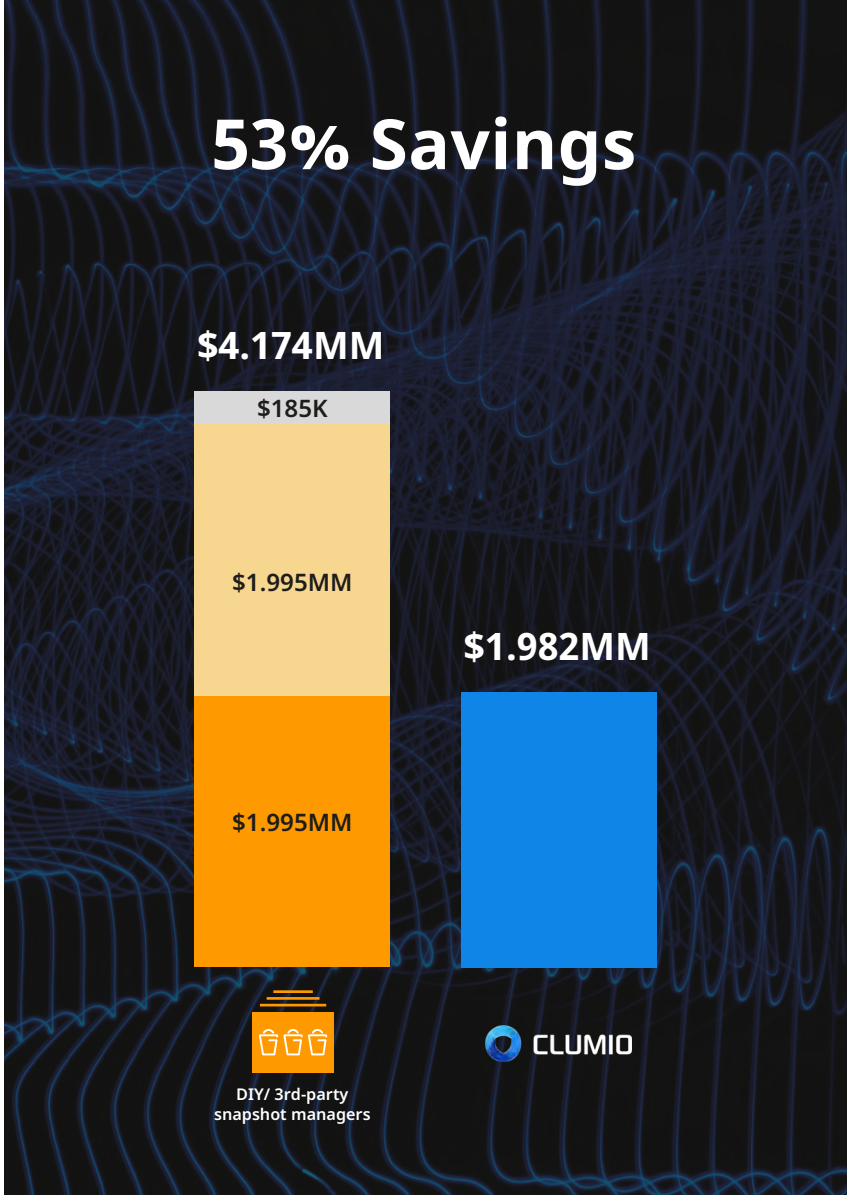
One of Clumio's significant TCO advantages is in how it implements its air gap. In a typical scenario, if an organization had $100,000 worth of snapshots sitting in their account and wanted to copy these files into an "air gapped" account (basically saving a copy into an account with a different password, the total investment would now double to $200,000.

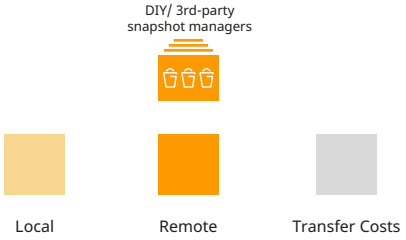
With Clumio, the process is automatic and fully managed. There is no tiering that the customer needs to worry about, it is all automatically taken care of, and the customer simply pays on one easy to understand per GB model. Within 10 minutes an organization can click a button and back up their data in an air-gapped environment. A business doesn't need to set up scripts, manage the process, or make mirror copies of their snapshots—thereby saving work hours and reducing backup costs by up to 50%.

The chart on the next page shows that by using a data protection service such as Clumio that has built-in air gap protection as well as lifecycle management for long term retention, organizations can avoid inefficient ways of backing up their data and save up to 50% on their AWS backup costs. Considering all the other benefits that are delivered by this solution, this is icing on the cake.

On top of that, we help our customers understand what's driving their backup costs. And for customers that want to easily protect their backups in a secure air gapped vault, customers can easily move their snapshots into Clumio Protect at a lower cost.







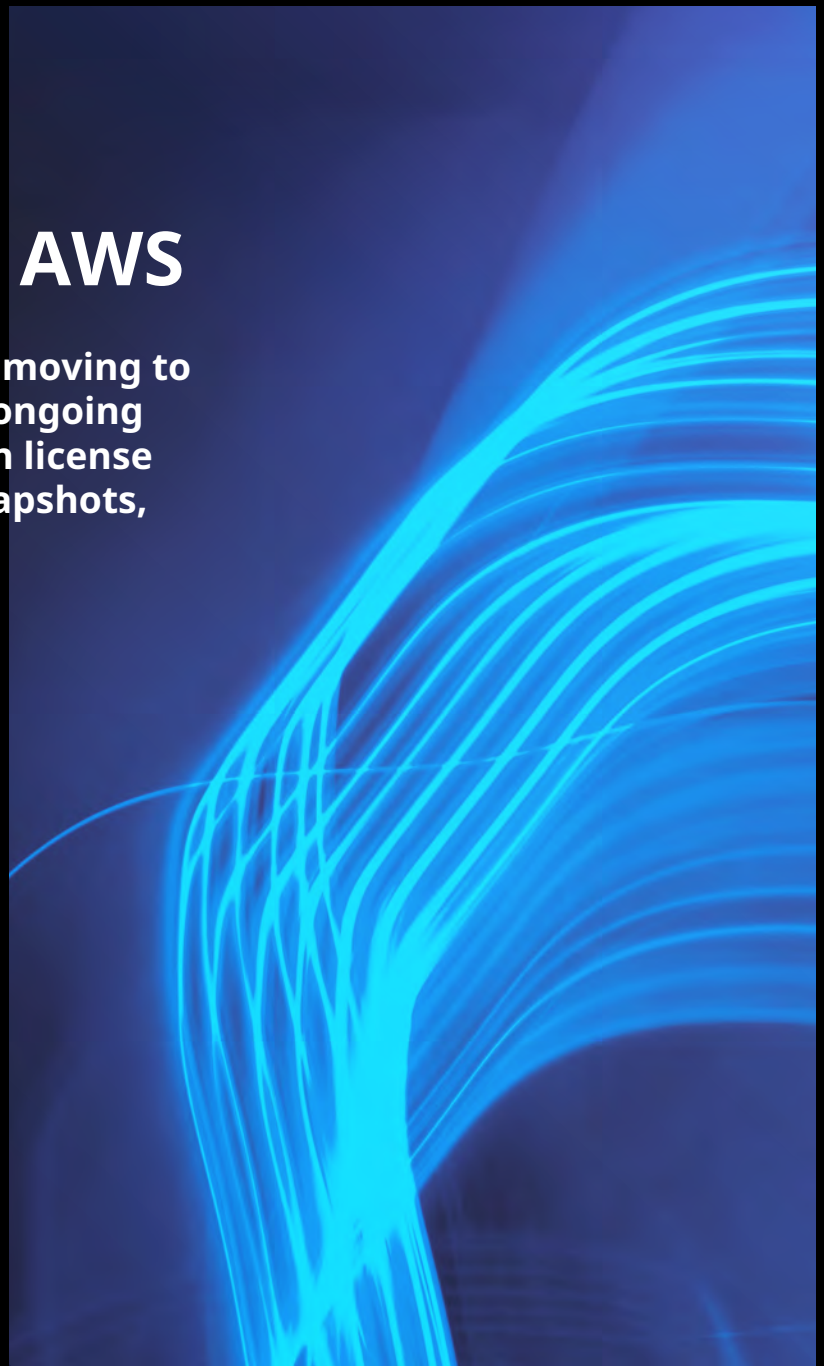
Assumptions

400 EC2 Instances
60TB EBS
30 daily snapshots
12 monthly snapshots
7 yearly snapshots
20% yearly data growth
3% / 20% / 50% data change rate
No productivity improvement

AWS Snapshot costs are local and remote vs Clumio air gapped backups
25% discount on Clumio, any in-account related to backup

PROTECT YOUR DATA IN AWS

By moving to Clumio, you get all the benefits of moving to a modern data protection solution without the ongoing cost of developing and maintaining it: no hidden license fees, easier administration, lower costs than snapshots, and better reporting.



CLUMIO

Ready to simplify backup for AWS?

Request a 1:1 demo: **clumio.com/demo**