MIT Technology Review Insights

Produced in partnership with

CLUMIO

A modern data resilience strategy starts with a cloud-native backup architecture.

Architecting cloud data resilience

Key takeaways

1 **Businesses that rely on public cloud infrastructure increasingly recognize the need for a simple, reliable cloud data resilience solution to protect their operations.**

2 **Traditional backup architectures aren't well suited for the unique qualities of cloud data. A cloud-native backup approach, by contrast, is designed to scale, built to prioritize simplicity and reliability, and optimized for speed of recovery.**

3 **Best practices for cloud data security include air-gapping, data encryption, and data immutability, as well as assurance that the tools used meet industry- and business-specific requirements for compliance and recovery times.**

Cloud has become a given for most organizations: according to **PwC's 2023 cloud business survey**, 78% of companies have adopted cloud in most or all parts of the business. These companies have migrated on-premises systems to the cloud seeking faster time to market, greater scalability, cost savings, and improved collaboration.

Yet while cloud adoption is widespread, **research by McKinsey** shows that companies' concerns around the resiliency and reliability of cloud operations, coupled with an ever-evolving regulatory environment, are limiting their ability to derive full value from the cloud. As the value of a business's data grows ever clearer, the stakes of making sure that data is resilient are heightened. Business leaders now justly fear that they might run afoul of mounting data regulations and compliance requirements, that bad actors might target their data in a ransomware attack, or that an operational disruption affecting their data might grind the entire business to a halt.

For all its competitive advantages, moving to the cloud presents unique challenges for data resilience. In fact, the qualities of cloud that make it so appealing to businesses – scalability, flexibility, and the ability to handle rapidly changing data – are the same ones that make it challenging to ensure the resilience of mission-critical applications and their data in the cloud.

"A widely held misconception is that the durability of the cloud automatically protects your data," says Rick Underwood, CEO of Clumio, a backup and recovery solutions provider. "But a multitude of factors in cloud environments can still reach your data and wipe it out, maliciously encrypt it, or corrupt it."

> "A widely held misconception is that the durability of the cloud automatically protects your data. But a multitude of factors in cloud environments can still reach your data and wipe it out, maliciously encrypt it, or corrupt it."
>
> Rick Underwood, CEO, Clumio

Complicating matters is that moving data to the cloud can lead to reduced data visibility, as individual teams begin creating their own instances and IT teams may not be able to see and track all the organization's data. "When you make copies of your data for all of these different cloud services, it's very hard to keep track of where your critical information goes and what needs to be compliant," says Underwood. The result, he adds, is a "Wild West in terms of identifying, monitoring, and gaining overall visibility into your data in the cloud. And if you can't see your data, you can't protect it."

The end of traditional backup architecture

Until recently, many companies relied on traditional backup architectures to protect their data. But the inability of these backup systems to handle vast volumes of cloud data – and to scale to accommodate explosive data growth – is becoming increasingly evident, particularly to cloud-native enterprises. In addition to issues of data volume, many traditional backup systems are ill-equipped to handle the sheer variety and rate of change of today's enterprise data.

In the early days of cloud, Steven Bong, founder and CEO of AuditFile, had difficulty finding a backup solution that could meet his company's needs. AuditFile supplies audit software for certified public accountants (CPAs) and needed to protect their critical and sensitive audit work papers. "We had to back up our data somehow," he says. "Since there weren't any elegant solutions commercially available, we had a home-grown solution. It was transferring data, backing it up from different buckets, different regions. It was fragile. We were doing it all manually, and that was taking up a lot of time."

Frederick Gagle, vice president of technology for BioPlus Specialty Pharmacy, notes that backup architectures that weren't designed for cloud don't address the unique features and differences of cloud platforms. "A lot of backup solutions," he says, "started off being on-prem, local data backup solutions. They made some changes so they could work in the cloud, but they weren't really designed with the cloud in mind, so a lot of features and capabilities aren't native."

Underwood agrees, saying, "Companies need a solution that's natively architected to handle and track millions of data operations per hour. The only way they can accomplish that is by using a cloud-native architecture."

Key features of data resilience solutions

Business leaders will want to consider the following features of their data backup and recovery solutions.

Air-gapping

An air-gapped backup solution separates copies of data from the originals, on a secure network isolated from all others.

Data encryption

Encrypted data requires decryption keys to read, keeping backup data secure and confidential, especially in the case of data breaches. Encryption of sensitive data is a requirement under many data protection regulations.

Data immutability

Immutable data backups use technological means to make it impossible to change or modify data once written. This prevents backups from being modified by bad actors or accidentally corrupted or deleted, ensuring long-term data reliability and integrity.

Data segmentation

Categorizing data by type and usage during backup lets organizations tailor their resilience strategies, including by providing additional security to the most sensitive data, assigning a higher priority for recovery to the most operationally important data, and streamlining costs by choosing not to back up noncritical data.

Recovery point objective (RPO)

RPO is an organization's maximum acceptable amount of data that can be lost due to a system outage, measured by the point in time to which data must be recovered (e.g., to 1 hour ago). A shorter RPO results in less data loss but requires more frequent backups.

Recovery time objective (RTO)

RTO is an organization's maximum acceptable time it can take to restore data and applications after an outage, measured in time (e.g., 1 day of downtime). A shorter RTO means systems come back up faster but has additional costs in resources for recovery.

A modern solution

Today, a cloud data resilience strategy is more than simply a precautionary measure; it's a business necessity. According to Bong, for example, "AuditFile was initially marketed to CPA firms. Then we started working with government agencies that required more rigorous infrastructure requirements. We would not be where we are right now without a solid data resilience strategy."

Gagle concurs that partners, suppliers, customers, and regulators are taking a strong interest in businesses' backup strategies – and making business decisions accordingly. "We get a lot of questions about our data, disaster recovery, and data resilience strategies," he says. "Nowadays everyone is asking about it – it's a common question."

BioPlus works within a regulatory environment that requires compliance with stringent industry-specific requirements for data protection. The company is licensed in 50 states and must comply with regulations ranging from those set by the National Association of Boards of Pharmacy to the United States' Health Insurance Portability and Accountability Act (HIPAA).

AuditFile is also deeply familiar with the data protection demands of a changing regulatory environment. As AuditFile continues to expand internationally, Bong says, "we have to comply with every regulation you could possibly imagine." Beyond the well-known regulations like HIPAA and General Data Protection Regulation (GDPR), he adds, "every single region seems to now have their own unique requirements, and they change all the time. So we have PIPEDA, International Traffic in Arms Regulations, SOC 2 compliance, ISO 27001, FedRAMP, FIPS 140-2."

Data backup and resilience solutions providers lke Clumio have focused on supporting regulatory compliance. "I cannot overstate how important it is for companies to design for compliance right from the first time they ingest customer data," says Underwood. "Data resilience can help you not only retain information for long periods of time, but also classify the information, retain the right information, make sure that it's secure, and make sure you have what you need when FINRA (the Financial Industry Regulatory Authority) comes calling and says, 'Can you show me that you've been compliant for the past seven

> "A lot of backup solutions started off being on-prem, local data backup solutions. They weren't really designed with the cloud in mind, so a lot of features and capabilities aren't native."
>
> Frederick Gagle, Vice President of Technology, BioPlus Specialty Pharmacy

Data resilience still a work in progress for many organizations

In case of a successful ransomware attack:

41% of organizations plan to restore from their standard data protection/backup solution

39% plan to restore from public cloud services

37% plan to restore from air-gapped storage

36% plan to restore from a disaster recovery service provider

35% plan to restore from an immutable backup ("gold copy")

Source: Compiled by MIT Technology Review Insights, based on data from "The Long Road Ahead to Ransomware Preparedness," Enterprise Strategy Group, March 2022.

Data resilience as a service

Seamless data resilience appeals to today's resource-strapped IT teams. "We didn't want to spend time fooling around with technology," says Steven Bong, founder and CEO of AuditFile. "Our data resilience system had to work seamlessly, it had to be easy to implement, and if it required three programmers and 12 months to implement, we knew it wouldn't work for us; we're a small startup."

Frederick Gagle, vice president of technology for BioPlus Specialty Pharmacy, finds great value in third-party managed backup of critical business data. "It takes the day-to-day management of backups off our shoulders, which can free up a lot of time," he says. In fact, because their provider also helps his team with time-consuming tasks such as file integrity and disaster recovery testing, BioPlus engineers have been able to focus more time on projects supporting the company's growth.

Increasing demand for data resilience managed services is no surprise to Clumio's Rick Underwood. "Today's architects want to build applications," he says. "They don't want to spend their time maintaining backups. If your applications are backed up by a cloud-native solution, you don't have to worry about your backups. It happens automatically in the background, whereas an on-premises approach to the cloud would require oversight and management."

years?' Forward-looking data resilience practice will ensure that you don't have any of those penalties, fines, or hiccups, seven or 10 or 15 years down the line."

And while some companies fear that they might run afoul of mounting data regulations and compliance requirements, they also must also be on the lookout for bad actors that might target their data in a ransomware attack, as well as for systems or logistics disruptions that might grind the entire business to a halt. The US experienced a **20% uptick in data breaches from 2022 to 2023**, reaching an all-time high.

"Everything that you do with your data in a business needs to be resilient so that your business can survive an outage, a ransomware attack, an operational disruption, or any unforeseen event," warns Underwood.

Critical functionality for data resilience

As the demand for cloud-native data protection grows, companies are learning that not all data resilience solutions are created equal. Cloud providers' default backup tools, for instance, often come with substantial limitations. Backups may be stored in the same security sphere as the original data, apply default recovery standards, lack industry-specific compliance, and circumvent a business's own people and processes – missing ingredients for a holistic data resilience strategy.

Instead, savvy business leaders are demanding data resilience strategies that segment data by use and importance; encrypt data and create immutable copies; audit backup data for accuracy; and create simple, fast, and tested systems for data recovery.

Air-gapping and data immutability top Gagle's list of desirable features in a backup solution. "We wanted a technology that isolates and keeps the data separate and won't allow it to be altered," he says. "Some backup solutions simply replicate whatever is on site and send it over to another location. But if that data becomes corrupted, the corruption could be propagated to the replicated copy. Having immutable point-in-time copies ensures we always have a recovery point available, and having those backups air-gapped adds additional peace of mind."

Underwood explains how air-gapping and immutability serve as a powerful line of defense at Clumio: "Any time a piece of customer data is being backed up, it goes straight into a different account in a different region that is outside of the customer security sphere. It's immutable, meaning that even if a hacker gets to it, there's nothing they can do with it. They can't delete it, they can't modify it, they can't corrupt it."

> “We treat it as an awesome responsibility that our customers have put their trust in us. So there is no plan B with data resilience. It’s everybody’s responsibility.”
>
> Steven Bong, Founder and CEO, AuditFile

A robust data resilience strategy should involve backup processes that don’t disrupt a production environment, contribute to additional IT complexity and technical debt, or require frequent manual intervention that could slow the recovery process. “With a cloud-native backup platform, you can restore your data about 10 times faster than a traditional backup platform could,” says Underwood. “It’s more fault tolerant, it’s less failure-prone, and it restores that much faster.”

An enterprise-wide responsibility

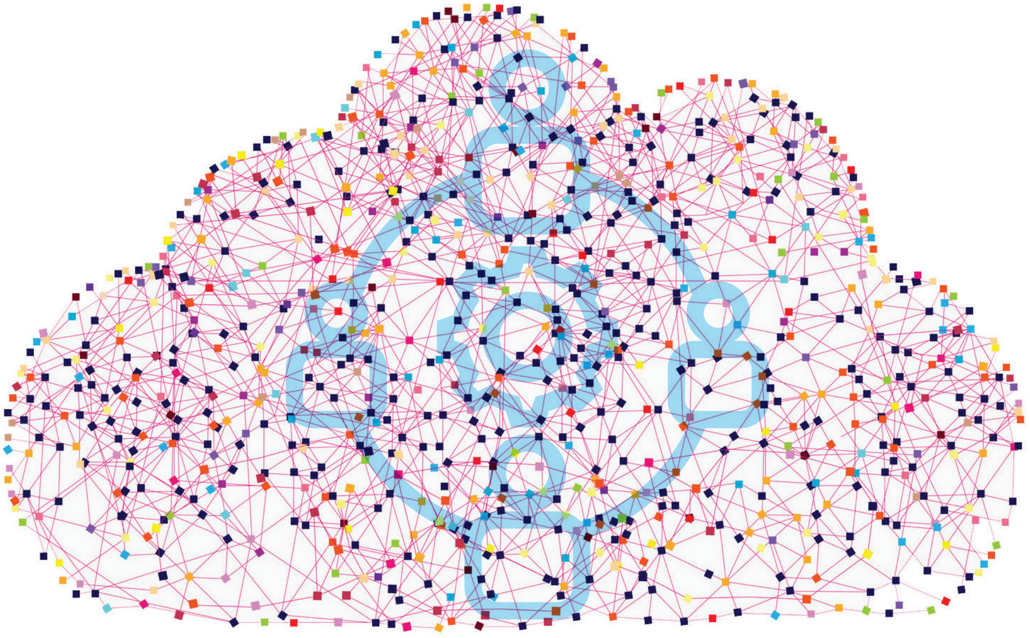
The massive proliferation of data in recent years, accompanied by the rise of AI applications that can derive insight from it, has made the business value of data clearer than ever. And there is no doubt that traditional backup architectures are no longer suited for the unique qualities of cloud data. A cloud-native backup approach – designed to scale, built to prioritize simplicity and reliability, and optimized for speed of recovery – is critical to survival in today’s digital world.

“Data is changing very rapidly and is being used in many different ways,” says Underwood. “So if something happens to your data today, and you have to go back three days to retrieve your data in its last operational state, it’s tantamount to three days of lost business, which is unpardonable and can result in millions of dollars of untracked revenue.”

Architecting for data resilience requires an understanding of security and reliability best practices, as well as meeting industry- and business-specific requirements for compliance and speed of recovery. But it also entails enterprise-wide ownership of data resilience strategies.

Driving enterprise-wide accountability isn’t always easy, especially when “backup and recovery often operate in the shadows,” says Underwood. “But a cloud data resilience strategy supports the entire infrastructure stack, whether you’re building apps, building machine learning models, or building analytics platforms. It touches everything.”

Bong says a collaborative approach to data resilience is critical to success. At AuditFile, he says, “everyone, all stakeholders, customers, management, the developers, they all understand what’s going on, where the data is, how it can be restored.” He adds, “We treat it as an awesome responsibility that our customers have put their trust in us. So there is no plan B with data resilience. It’s everybody’s responsibility.”



“Architecting cloud data resilience” is an executive briefing paper by MIT Technology Review Insights. We would like to thank all participants as well as the sponsor, Clumio. MIT Technology Review Insights has collected and reported on all findings contained in this paper independently, regardless of participation or sponsorship. Teresa Elsey was the editor of this report, and Nicola Crepaldi was the publisher.

About MIT Technology Review Insights

MIT Technology Review Insights is the custom publishing division of MIT Technology Review, the world’s longest-running technology magazine, backed by the world’s foremost technology institution – producing live events and research on the leading technology and business challenges of the day. Insights conducts qualitative and quantitative research and analysis in the US and abroad and publishes a wide variety of content, including articles, reports, infographics, videos, and podcasts. And through its growing MIT Technology Review Global Insights Panel, Insights has unparalleled access to senior-level executives, innovators, and entrepreneurs worldwide for surveys and in-depth interviews.

From the sponsor

We backup, so you can build.

Clumio helps the world’s leading enterprises automate the backup, recovery, and compliance of critical data that powers their business. Clumio customers, such as Warner Bros. Discovery, Atlassian, and LexisNexis, enjoy hands-off manageability, inherent security, and a 30% reduction in their cloud backup costs. For more information, visit **clumio.com**.

CLUMIO

Illustrations
Cover art by Adobe Stock, spot illustrations created by Chandra Tallman Design LLC, compiled from Adobe Stock and The Noun Project.

While every effort has been taken to verify the accuracy of this information, MIT Technology Review Insights cannot accept any responsibility or liability for reliance by any person on this report or any of the information, opinions, or conclusions set out in this report.

MIT Technology Review Insights

www.technologyreview.com

insights@technologyreview.com