rubrik

Solving the Unstructured Data Challenge

eBook

ENTER →

Executive Summary

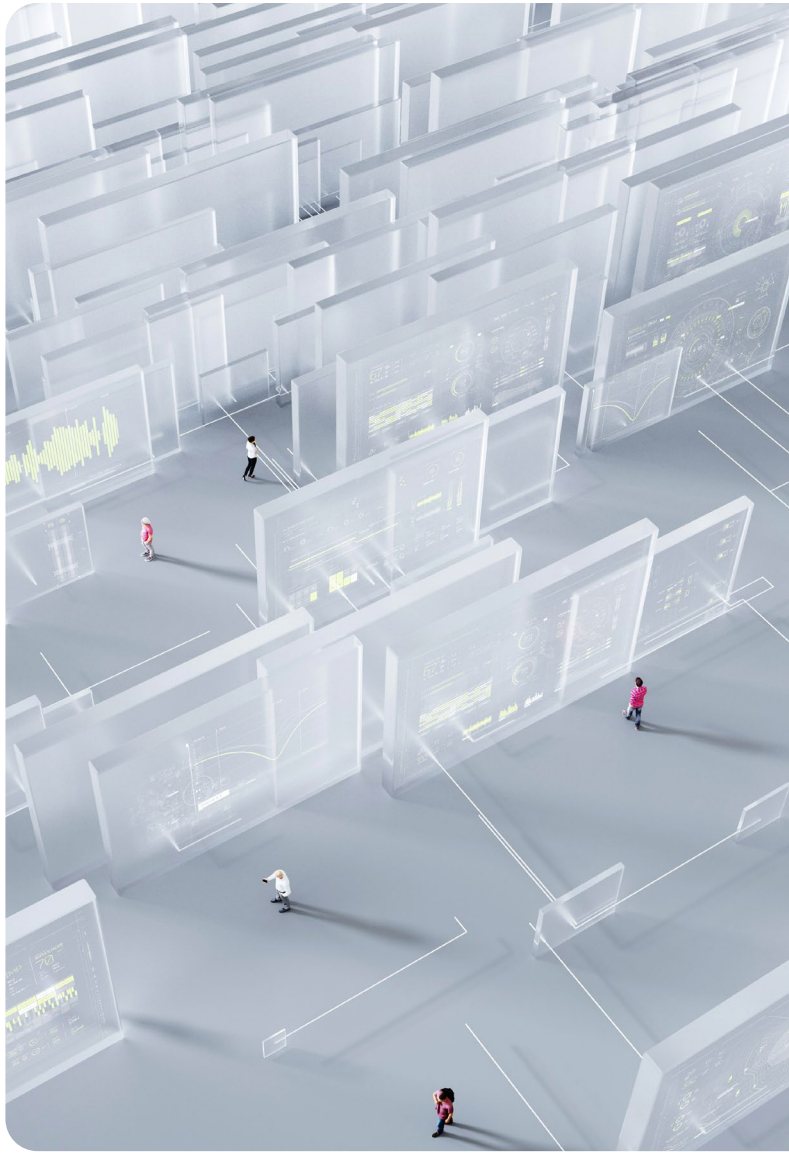
If you've got digital applications, you've got unstructured data. Petabytes of it. But do you know what you're actually collecting?

You likely already know the three Vs of data: **variety, velocity,** and **volume**.

In this ebook, we'll explore the forgotten fourth V: **value**. We'll go through five key steps to building a resilient unstructured data strategy. And we'll talk about how Rubrik can help you manage and protect your unstructured data.

Ready to dive into the murky world of unstructured data?

Let's go!



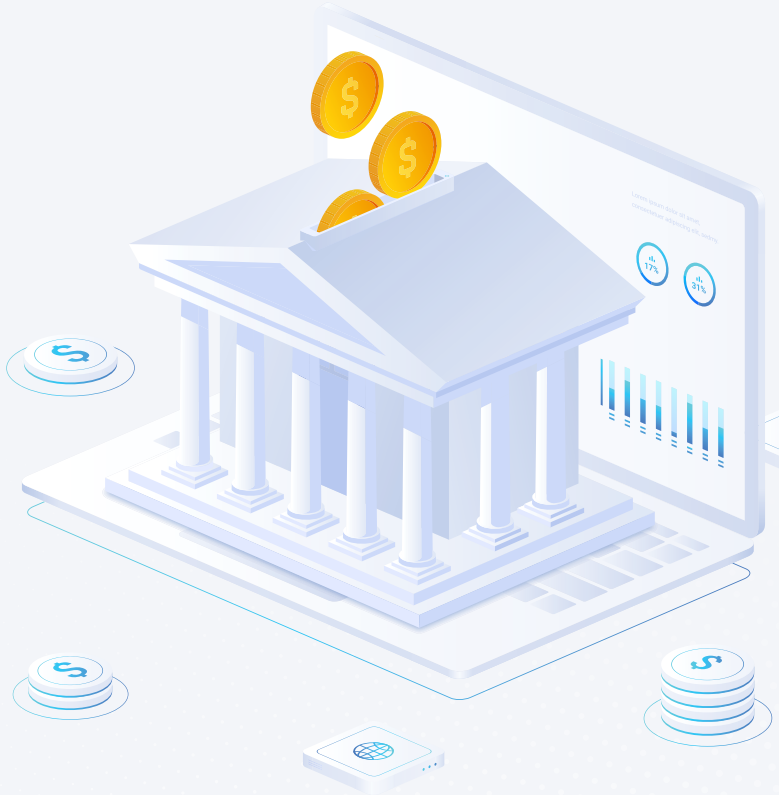
PREV NEXT

The Tale of MegaBucks Bank

You're the CISO of MegaBucks Bank, catering to ultra-high-net-worth individuals. You process a huge volume of data every day, much of it sensitive: customer account information, social security numbers and other PII, and more. This is your bread and butter.

You know how critical it is to keep this information safe, so you've invested heavily in your data security strategy: top-of-the-line databases featuring the latest protection, a solid Business Continuity and Disaster Recovery plan, and a team that's best in the business at thwarting daily cyberattacks. Your bases are pretty well covered, right?

Not so fast.



Today:

Two employees pulled a full account number from the database to facilitate their chat conversation.

Someone on the M&A team emailed a copy of a three-year-old contract containing proprietary information to a colleague.

Your CSRs fielded calls with customers, all of which were recorded, and many of which contained PII.

And you've just learned that a cyberthug has gained access to your systems.

Now you've got a problem. Because every one of these interactions generated a file that's now buried within your unstructured data: unorganized, difficult to manage, most likely containing sensitive information, and probably underprotected.

It's going to be a very long night.

Variety, Velocity, Volume, and...Value?

That was a nightmare scenario, right? Fear not. In this ebook, we'll help you develop a strategy to manage and protect your unstructured data so that you can avoid the headache that our fictional MegaBucks Bank CISO is facing.

Anyone who's been in the business in the past 20 years has heard of the three Vs: **Variety, Velocity, and Volume.** Defined by Doug Laney in 2001[1], the three Vs are properties of data an organization gathers.[2]

DATA VARIETY

Refers to the **different types** of data you're collecting. Having all kinds of information can give you a fuller picture, but you'll need a system that can manage things like:

- Structured and unstructured data
- Different formats
- Different nomenclature

DATA VELOCITY

Refers to **how quickly** you're generating and processing data. In business data, speed is the name of the game.

- A B2C e-commerce site that loads in 1 second has an e-commerce conversion rate 2.5x higher than a site that loads in 5 seconds.[3]
- About 1/3 of Americans are looking for good fraud protection when they're choosing a bank,[4] so financial institutions need to process data in real time in order to prevent fraudulent activity.
- Hackers are fast and relentless; 99% of IT and security leaders became aware of at least one cyberattack in 2022, with an average of 52 occurrences worked.[5]

DATA VOLUME

Refers to the **amount** of data you're collecting. Hint: it's a lot. Here are some truly astounding numbers from our Rubrik Zero Labs State of Data Security 2023 report:

- A typical organization averages 239.9 backend terabytes (BETB) of data.
- That number is even higher for companies in specific industries:
  - > Telecommunications – 442.6
  - > IT & Technology – 398.9
  - > Insurance – 301.5

So, you've got to manage tons of data, in all different formats, very quickly.

But that's not the end of the story.

We at Rubrik argue that there's a forgotten fourth V that you ***must*** have a handle on if you're going to be successful in managing and protecting your information: **Data Value**.

DATA VALUE

Refers to **how useful** the data you collect is to your business. You need to know what you've got, why it matters, and where it lives. Because you're collecting a ton of business-critical information – what we call your company's crown jewels. Remember that 239.9 BETB of data? A lot of it is sensitive.

- Globally, a typical company manages an average of about 24.8 million sensitive files.[6]
- The most sensitive data in a single Rubrik-secured organization is 1.3+ billion records.[7]

And your sensitive data footprint is only going to get bigger as your company grows.

[1] 3D Data Management: Controlling Volume, Velocity, and Variety
[2] Big Data: The 3 V's Explained; 3 V's (volume, velocity and variety)
[3] Site Speed is (Still) Impacting Your Conversion Rate
[4] New FICO Survey: Americans Value Financial Fraud Prevention More Than Banking Customer Experience
[5] Rubrik Zero Labs, The State of Data Security Report 2023
[6] Ibid.
[7] Ibid.

24.8 MILLION

average amount of sensitive files for a typical company

1.3+ BILLION

amount of sensitive data in a single Rubrik-secured organization

PREV NEXT

Bottom line: your data's value will dictate:

What you protect

How often you back it up

How long you keep it

How you secure it at rest to prevent encryption, exfiltration, and ransom

Now, you (hopefully) already have a good handle on the value of the structured data you have. You're doing all you can to manage and protect it well.

But what about your unstructured data? Do you have the ability – right now – to understand what's in your unstructured files and increase the resiliency of your data at rest to make you more exfiltration proof?

If you're like most data executives we encounter, your honest answer probably is…not really.

The good news: we're here to help.

5 Steps to a Robust Unstructured Data Strategy

We know that the vast majority of your unstructured data doesn't contain sensitive information, give you a competitive advantage, or enable you to run your business. Most of these files aren't your company's crown jewels.

KEY POINT #1 – But your unstructured data ***does*** contain the information you'd consider your crown jewels.

KEY POINT #2 – And hackers ***know*** that your unstructured data likely has juicy information, so they're intentionally targeting backups. Bad actors are ***logging in***, not hacking into the network. They now have use permissions to your unstructured data.

KEY POINT #3 – Plus, according to IDC, ***90% of your data*** is going to be unstructured in the next five years.[7]

So, if you're relying on the old thinking of snap-and-replicate or NDMP, you're setting yourself up for failure. You're in the dark about what information is where, and you could inadvertently be introducing malware into your backups and archives.

It's no longer wise to simply back up your unstructured data and hope for the best. Because that's a data ***continuity*** strategy, not a data ***resiliency*** strategy. And in the era of cybercrime, ***resiliency is key***.

The following five steps will help you understand, manage, and protect your unstructured data.

Let's dive in.

[7] IDC, Meeting the New Unstructured Storage Requirements for Digitally Transforming Enterprises

Step 1: Know Your Data

Data is growing at a mind-boggling rate; we at Rubrik estimate that the total volume of data a typical organization needs to secure will increase by **7x** in the next five years.[8]

This first step will help you wrap your arms around your unstructured data ecosystem so that you can properly value and protect it.

UNDERSTAND WHO IS CURRENTLY RESPONSIBLE for data collection and valuation. Many folks are tempted to think that this would be a storage and/or cloud administrator's job, but it shouldn't be. These administrators aren't in charge of knowing what specific data they're managing and where it's coming from. Collection and valuation should fall on data owners throughout your organization, but who is doing it now?

UNDERSTAND WHERE DATA IS COMING FROM. Do you have a handle on every application that's generating data across your business, or is shadow IT running free? Make a complete list. **Then ask:** Are all origin points free from malware? It's critical that, for all major buckets of data, the generation points are known, valued, and clean. **Also check:** Are data owners involved in and/or making decisions about their data's security? Or, are those decisions being made by storage and backup administrators?

[8] Rubrik Zero Labs, The State of Data Security Report 2023

Step 2: Validate Your Data

Now it's time to analyze what you have and set policies in place to move forward.

Broadly look at **what is stored** and understand its **criticality** to your company. Printer logs? Probably not crucial. But what about emails? What if they went down? Or what if you lost internal messages or sensor data? If you know which applications are producing important data, you'll be better able to figure out your protection strategy.

Establish **two key roles** in your organization and set/understand their workflows. Both of these roles will provide accountability and expertise in the validation process, and they'll also help you implement your unstructured data resiliency strategy.

C-SUITE CHAMPION

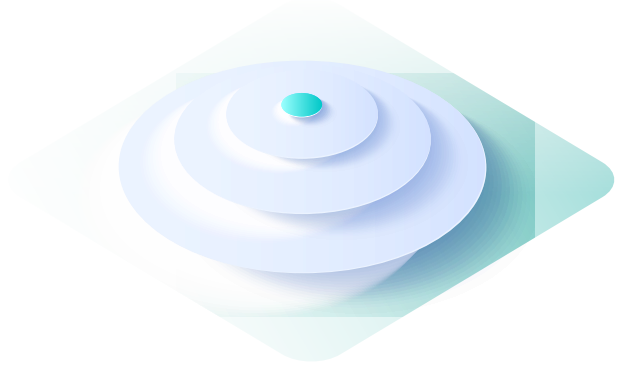
Only 54% of external organizations have a single senior executive responsible for data and its security, yet 98% of external organizations believe they currently have significant data visibility challenges.[9] If this sounds familiar, bridge this gap by elevating data to the C-Suite. Naming a C-Suite Champion will help you streamline and enforce your data protection strategy, and it will also demonstrate internally, to customers, and to stakeholders that you're serious about your data, no matter where it lives.

DATA CUSTODIANS

If storage and cloud admins are not data owners, who should be? We recommend data experts – Data Custodians – who are embedded in teams throughout your organization and are responsible for the entirety of their teams' Information Lifecycle Management (ILM); from the time data is generated to the time it moves to storage to the time it's deleted, they understand why it matters, ensure that it's clean, and safeguard it appropriately.

[9] Rubrik Zero Labs, The State of Data Security Report 2023

Step 3: Tier Your Applications



It's time to tier (rank) the applications that generate your data.

This will help you standardize a workflow and manage where and how data is stored and protected. How you tier your applications is specific to your company – you know what data is most crucial for your business.

HIGH TIER (CRITICAL)

The data produced by these applications is essential. Maybe these files aren't recreatable, like patient imaging files or point-in-time sensor and navigation data. They could be files you're required to keep for compliance or regulation purposes. Your emails also probably fall in this category – both because they may have sensitive information and because you need to keep your email running. This tier should have your strongest service-level agreement (SLA) policies, with robust protection and resiliency so that you can recover quickly to a very recent backup should this data be compromised.

Bottom line: If compromise of a specific app would cause you significant disruption, then that app belongs in this tier.

> Your unstructured data is coming from all corners of your company: email clients, sensors, social media programs, printers, word processing and presentation software...basically any digital program you're running.

PREV NEXT

MID TIER (IMPORTANT)

Data produced by applications in this tier is important to your business, but not critical. Think files like analytics that help iterate improvements, or (non-sensitive) reports that are useful to your teams. If you were victim of a cyberattack and lost this data, you could recreate it and continue with your work.

Bottom line: Apps in this tier produce data that's useful, but not critical to your day-to-day operations or enduring stability.

LOW TIER (EVERYTHING ELSE)

Low-tier applications produce data that has little or no business value to your organization. If you're like most companies, this tier is probably where your printer and social media apps belong. Data in this tier doesn't need robust protection, and your ILM strategy will probably archive it and remove it from the source.

Bottom line: If a hacker could compromise data from an app with little or no disruption to your business, then the app belongs in this tier.

Step 4: Determine and Implement Your Data Strategy

After all that groundwork, it's finally time to get to the meat of your strategy and implementation. Using the information you've gathered and organized, set your data rules. Standardize the people, processes, and technologies related to the application tiers you determined in Step 3. So, for High-, Mid-, and Low-Tier applications in each team across your entire organization, establish the following:

Who is each team's Data Custodian responsible for managing that team's application data?

How often should data from each type of application be backed up? When should data move to archives? When can it be retired (deleted)? Keep in mind that regulatory and legal requirements also might apply here and dictate your retention policy.

What is your recovery procedure going to be for each tier in the event of a cyberattack?

Where in your infrastructure should data from each application tier go?

- How much of it is backup data vs. archive data?
- Can you afford to lose the data you send to archives (which is usually not backed up and therefore would be unrecoverable if hit by a cyberattack)?
- Are there any regulations, legislation, or policies (such as HIPAA or CCPA) that need to be considered when determining where to send the data?

How is data protection for each type of application going to be handled?

> Remember: once you've set the rules, it's your Data Custodians who **OWN** the data lifecycle process of managing, protecting, and securing the data. They're closest to the data that's produced, and they're responsible for getting it safely to its appropriate resting place.

Step 5: Maintain and Monitor Your Strategy

You've done it! You understand where your unstructured data is coming from, who is managing it, and its value to your organization, and you've developed a procedure to keep it safe, secure, and appropriately housed in your infrastructure.

Only one thing remains: maintaining and monitoring your unstructured data strategy over the long term. This is a crucial step, and not one to be taken lightly. Because your data is growing exponentially. You're constantly adding new applications. People come and go from your teams. And cybercrime is only getting worse.

Here are four concrete actions you can take to ensure that your strategy remains effective and resilient:

LEVERAGE DATA VISIBILITY TO PROACTIVELY REVIEW SENSITIVE DATA ON A RECURRING BASIS

Make your life easier with data reduction. This could include removing sensitive data with no user access in the last year, finding and deleting duplicate data copies, or removing data in user shares for employees/clients/partners who left in the last year. This also applies to duplicative data across different data stores in a single enterprise.

KEEP AN EYE OUT FOR ABNORMALITIES TO ENSURE DATA RESILIENCY

Monitor for unusual data additions, deletions, or encryption using machine learning. Ensure that users have the right amount of access (e.g., read-only, editing) based on their need to know. Use a solution that will proactively alert you about potentially malicious activity in your backup data during a ransomware attack.

BE INTENTIONAL ABOUT DATA GROWTH

Do what you can to slow the explosion of data you have to manage. For example, set cloud growth to no more than 50% of the environment total, delete data based on set policies, or only place sensitive data in one enclave.

REVISIT YOUR STRATEGY OFTEN

Your strategy needs to be a living, breathing document. We recommend reviewing it as often as possible (at least twice a year) to ensure that your teams are in compliance with the policies you've set and you're actively keeping ahead of cyberthreats.

Congratulations!

You've successfully transitioned from an unstructured data **continuity** strategy to an unstructured data **resiliency** strategy.

PREV NEXT

Rubrik: Your Partner for the Journey

Now that you've read through the steps and are determined to avoid the problems that our MegaBucks Bank CISO faced, let's talk about how Rubrik can help.

Rubrik NAS Cloud Direct is the modern solution to manage and protect your unstructured data. This stateless VM lives in Rubrik's SaaS control plane, can be deployed natively in the cloud or in the data center, and has the ability to scan billions of files – any file workload, any time. Built on Zero Trust principles, it's ready for your biggest unstructured data challenges:

VOLUME

Got billions of files, with more being added every day? No problem.

- Protect petabyte-scale data across all NAS technologies with highly efficient scanning, indexing, and moving of data.

SPEED

Need to move files and generate backups without disrupting your production environment? We can do that.

- Scan, index, and move NAS data in parallel streams to maximize network throughput.
- Eliminate impact to users with dynamic throttling.
- Dramatically reduce backup windows and drive operational efficiency with truly incremental forever backups.

EFFICIENCY

Want to eliminate complexity and streamline your processes? We can help.

- Integrate with SecOps tools such as SIEM/SOAR to drive collaboration between ITOps and SecOps teams to quickly scope and identify threats.
- Archive data from any NAS source directly to any on-premises target, cloud, or private storage based on policy settings that you define.

MANAGEMENT

Looking for a way to manage your unstructured data, rather than leaving it an unruly mess? We do that, too.

- Search and locate specific files with ease and retrieve any previously protected version of your NAS file data.
- Identify the presence of sensitive data with Sensitive Data Monitoring and gain insights into its access security posture to limit the risk of exposure.
- Identify aged or growing datasets quickly with NAS CD Data Discover so they can be archived, migrated, or retired.

With Rubrik NAS Cloud Direct, you can better manage and protect your unstructured data with confidence.

Let's see it in action.

SECURITY

Want best-in-class protection? We're on the job.

- Secure NAS data with encrypted-at-rest, immutable backups, including credential isolation for increased cyber resilience.
- Leverage custom-developed NFS and SMB clients designed specifically for rapid data protection at scale.

RESILIENCE

Hit by a cyberattack and need to get back up and running quickly? We've got your back.

- Quickly identify and locate which applications and files were impacted by ransomware using Rubrik Anomaly Detection.
- Surgically recover affected NAS data with the help of Rubrik Anomaly Detection's impact analysis.
- Automate recovery workflows – such as mass recovery of NAS data – to production, including post-recovery tasks for faster recovery and less downtime.

PREV NEXT

Rubrik NAS Cloud Direct: Paying Dividends for a Financial Firm

Akuna Capital, a New York- and Chicago-based hedge fund, runs a high-frequency trading system that stores 400 TBs of data – that's two billion files.

In high-frequency trading, trades can happen in microseconds, so every moment counts. Before Rubrik, the backup solutions Akuna Capital used weren't able to scan and protect their data quickly and without disrupting performance. This slow and inefficient backup could have posed a serious risk to the company's business.

With Rubrik NAS Cloud Direct integrated into Pure Flashblade, Akuna Capital can now scan 400,000 files per second, completing backups in under two hours per day. Best of all, they're protected from cyberthreats.

Convinced? That’s what we like to hear!

Start your data security revolution today with Rubrik NAS Cloud Direct.

LEARN MORE →