THE STATE OF DATA SECURITY

THE JOURNEY TO SECURE AN UNCERTAIN FUTURE

Rubrik Zero Labs

CONTENTS

- BEGIN YOUR JOURNEY 04
- ABOUT THE DATA 06
- UNDERSTANDING RISK 08
- CURRENT DATA REALITIES 12
- DATA PREDICTIONS 17
- SENSITIVE DATA CHANGES 19
- DATA SECURITY EVALUATION 25
- RECOMMENDATIONS 30
- SUMMARY 38

Data sources

(RT) RUBRIK TELEMETRY (WR) WAKEFIELD RESEARCH

This is a story about data

How much we have right now, how much we're going to have, and how it will impact our ability to protect our data.

It's also a story about

optimism

How it hurts us and how it can help us navigate an uncertain future.

ABOUT THE DATA

ORIENTING THE MAP

Rubrik Zero Labs strives to deliver actionable, vendor-agnostic insights to reduce data security risks. To that end, we incorporated findings from two independent sources.

Rubrik Telemetry ®

We utilized Rubrik telemetry in an effort to stay as close as possible to the ground truth of a typical organization's environment and the threats facing them. Using this approach also helped us create actionable future models from actual datasets. Speaking of transparency, here's what makes up this dataset and how it influences our perspective. Rubrik telemetry includes:

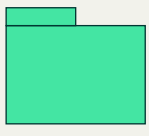
5000+
Clients

22
Industries

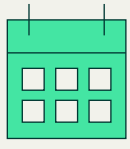
35+ EB SECURED
with 24+ billion sensitive data records

67 countries in 3 regions

Total volume of data secured:
35+ exabytes of logical storage
817 backend petabytes (BEPB) of physical storage



24+ billion sensitive data records



Data covers January 2022 through July 2023

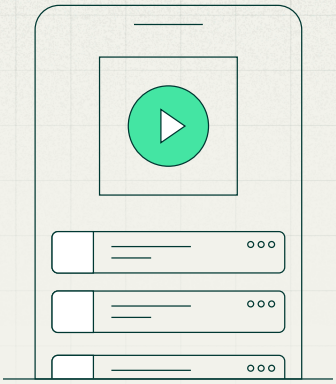
Hey nerds!

When most of the world hears "data," they think of logical storage, also known as frontend storage. Those of us in the data business focus on backend storage.

Rubrik takes the entirety of an organization's data and performs a number of different techniques—including deduplication and compression—to reduce the amount of frontend data to backend storage. We'll use backend storage throughout this journey.

A sense of scale: How much data is 35 EB?

What's a journey without a killer playlist? A typical high-quality song is about 4.8 MB. Our #2023DataJourneyJamz playlist of 35 EB would contain more than 7.3 trillion songs and play for more than 48 million years[1][2][3]. Look on the Mr. Brightside, we're not going to have repeats.



Wakefield Research (WR)

Perspectives from 1600+ IT and security global leaders

10
countries

49%
CIOs or CISOs

We commissioned a survey conducted by Wakefield Research to supplement our Rubrik telemetry and provide a broader view of the data security landscape. We also intentionally chose to engage IT and security leaders to study the difference between the leaders' point of view and what we're seeing on the ground.

1600+
IT and security leaders

49%
CIOs and CISOs

16%
VPs

38%
senior directors or directors

Three regions
(United States, EMEA, and APAC)

10 countries
(United States, United Kingdom, France, Germany, Italy, Netherlands, Japan, Australia, Singapore, India)

Research time frame included
1 June 2022 through 1 June 2023

1 https://math.ucr.edu/home/baez/timeline.html#:~:text=50%20million%20years%20ago%20%2D%20India,of%20all%20species%20died%20out!
2 https://ucmp.berkeley.edu/tertiary/eocene.php
3 https://www.whistleout.ca/CellPhones/Guides/How-Much-Data-Does-Spotify-Use-Canada

UNDERSTANDING RISK

QUICK THINKING AND QUICKSAND

Before we dive into your data journey, here's a little primer on how we make decisions. Some pro tips, if you will, for the choices you'll make in the future.

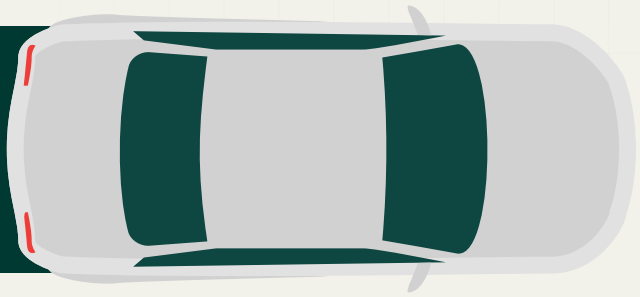
Humans are inherently optimistic.

We have to be. If we truly had to face the amount of hardship in today's world, many of us would never get out of bed. We need optimism to survive.

But that same optimism can also blind us from important realities.

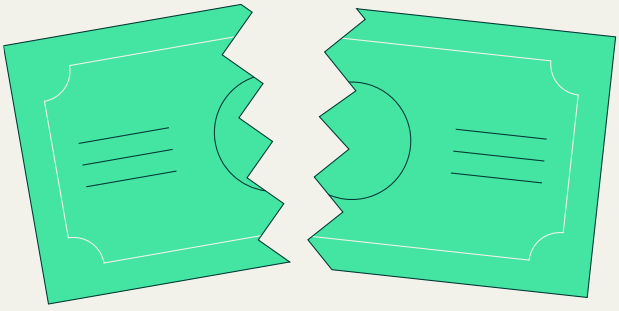
Somewhere in the back of our minds, we know every time we hop in a car we're taking a risk. In fact, insurance companies plan for the average driver to be in a car accident...

once every 18 years



Still, almost all of us overestimate our ability to beat the odds. We Text, Dial, Eat, Talk, Read, Groom... all while driving.

We know something bad could happen. We're just counting on it not happening to us, not today.



2007

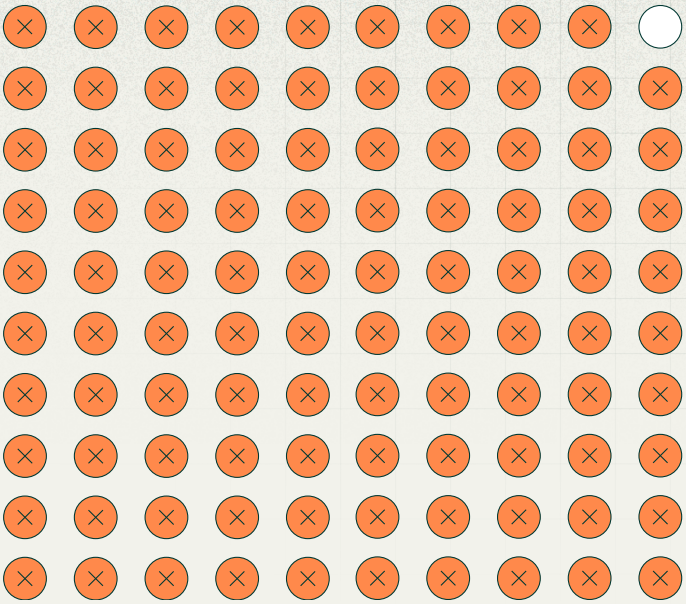
Our optimism also applies to how we see things in larger groups. In 2007, the signs of the impending global economic crisis were available to see. Some did.

Michael Burry, founder of hedge fund Scion Capital, detailed how he tracked the warning signs of the housing market crash beginning in 2003 in an op-ed for The New York Times, titled "I Saw the Crisis Coming. Why Didn't the Fed?"[1]

Most of the foremost experts either didn't see or discounted these important signals.

Somewhat ironically, organizations often approach cybersecurity—a topic notorious for its association with fear, uncertainty, and doubt—with more optimism than you might think.

99%



This spring's Rubrik Zero Labs report[2] revealed that 99 percent of IT and security leaders were made aware of at least one attack in 2022 with an average of 52 occurrences worked. These same leaders guide teams and are often under-resourced and ill-prepared to address and recover from those same attacks.

Organizations know they might be attacked. But they hold onto a certain amount of "but not us, not today" optimism.

That Gut Feeling

Every time we hop into a car, sign a mortgage agreement, or arrive at any of the other 35,000 big and small decisions[3] we're estimated to make every day, we take a calculated risk.

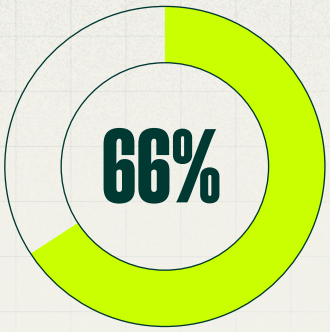
According to researchers, this risk evaluation happens in one of two ways: **intuition or analysis.**

Economists and psychologists debate the usefulness and accuracy of each kind of thinking, but generally, intuition is great for quick decisions with little effort[4], and analysis helps us with more complicated decisions[5], albeit more slowly and with more effort.

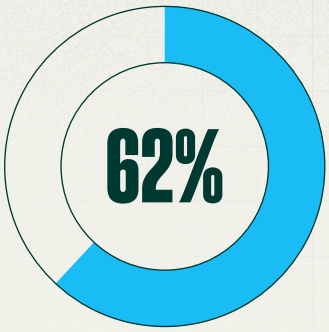
Both types of thinking are subject to error, but intuitive thinking is considered less accurate. However, accuracy in certain kinds of intuitive thinking can be improved by experience. That's why skilled emergency personnel can "sense" danger even before it happens.

We already know organizations are optimistic–at least in their actions–when it comes to cybersecurity.

Let's examine some intuition points:



believe their data growth has already outpaced their ability to secure data and manage risk.[2]



The majority of external organizations believe employees are accessing data in violation of established data policies.[2]

There's a clear disconnect between how organizations behave and how their experts feel about their data security approach.

Does this same disparity exist in the hard data? Are these experts canaries in a coalmine or is this intuition run wild? *Let's find out!*

1 https://www.nytimes.com/2010/04/04/opinion/04burry.html?searchResultPosition=3
2 https://www.rubrik.com/zero-labs#hero
3 https://edition.cnn.com/2022/04/21/health/decision-fatigue-solutions-wellness/index.html#:~:text=Whether%20you're%20making%20breakfast,put%20your%20finger%20on%20why.
4 https://thedecisionlab.com/reference-guide/neuroscience/automatic-thinking
5 https://thedecisionlab.com/reference-guide/philosophy/system-1-and-system-2-thinking?utm_campaign=TDL+Dynamic&utm_medium=ppc&utm_source=adwords&utm_term=&hsa_mt=&hsa net=adwords&hsa_ad=564666141034&hsa_src=g&hsa_cam=14567061057&hsa_kw=&hsa_grp=127713121155&hsa_tgt=dsa-19959388920&hsa_ver=3&hsa_acc=8441935193&gad=1&gclid=Cj0K CQjw2qKmBhCfARIsAFy8buJQVuDOqwKnaCjbZ1gPhagxEBoYo6z8q6VxIf0_thiI3IfcDPUDozcaAknUEALw_wcB

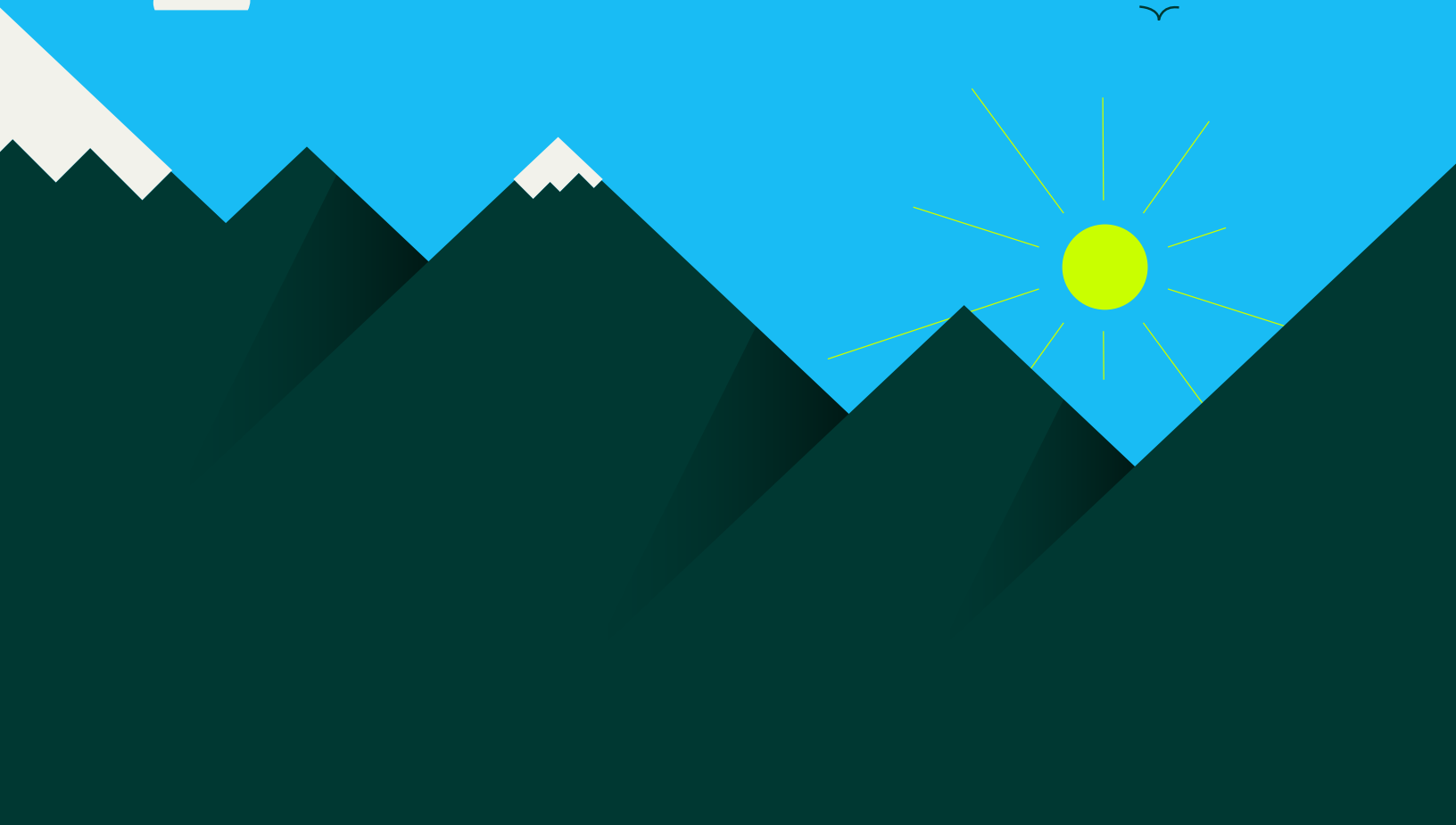
CURRENT DATA REALITIES

TODAY'S TRAIL MARKERS

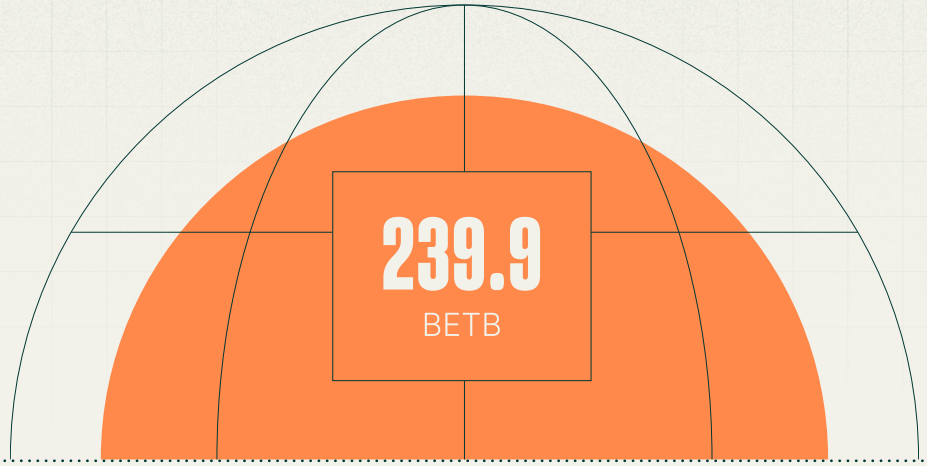
Let's switch our thinking from intuition to analysis. Hard data with concrete takeaways. Without analysis we're nothing but vibe merchants.

Vibe Merchant: noun.

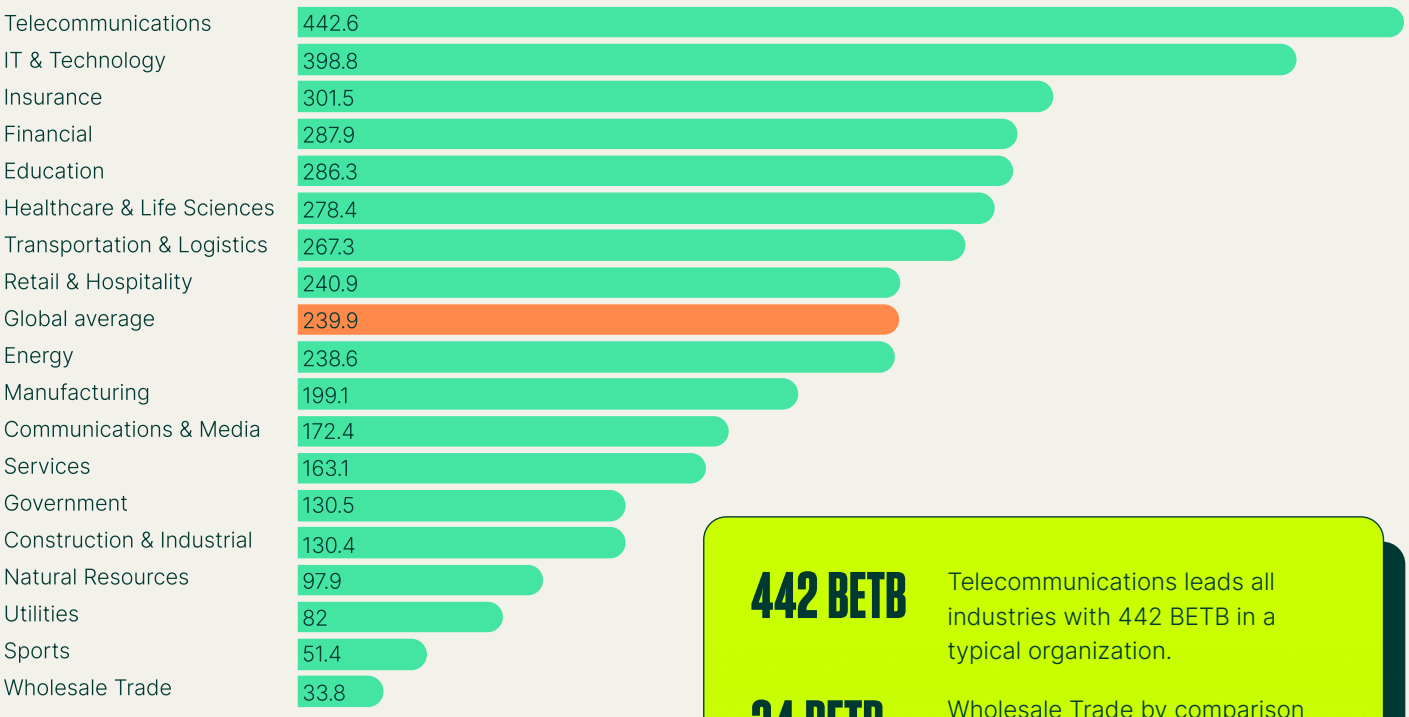
A person or thing making all decisions based on gut feelings or perceptions of what's popular/trendy. Similar terms are: Vibe Vendor, Data Influencer, or anyone involved in the Pontiac Aztek. Example: The Pontiac Aztek was 100% created by vibe merchants.[1]



OK with that out of the way, we're ready for analysis. Here's the typical data volume for a global organization:



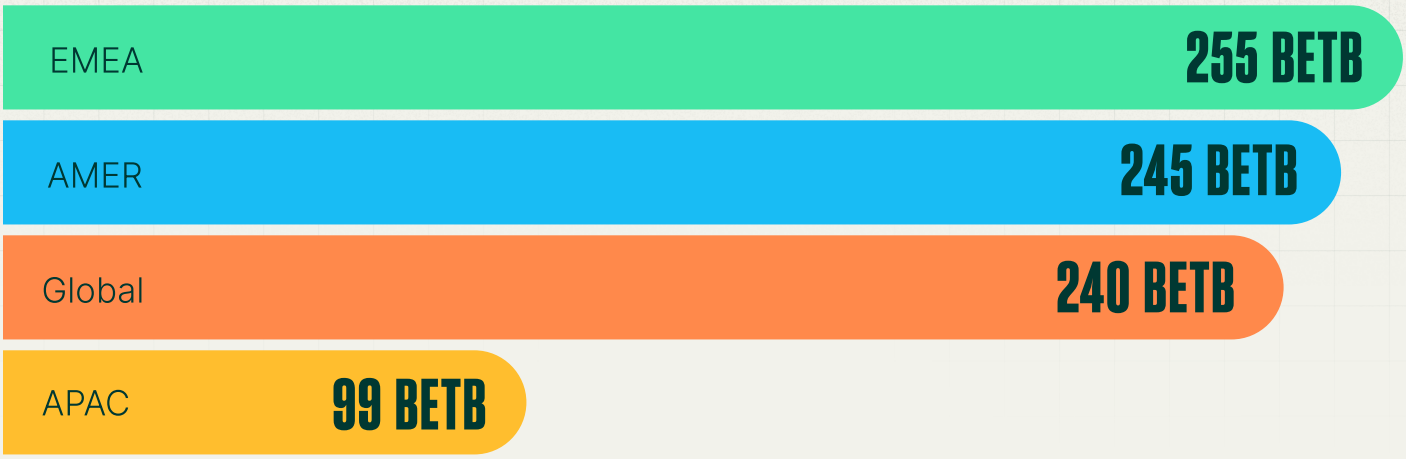
Now Let's compare all industries and regions: [1]

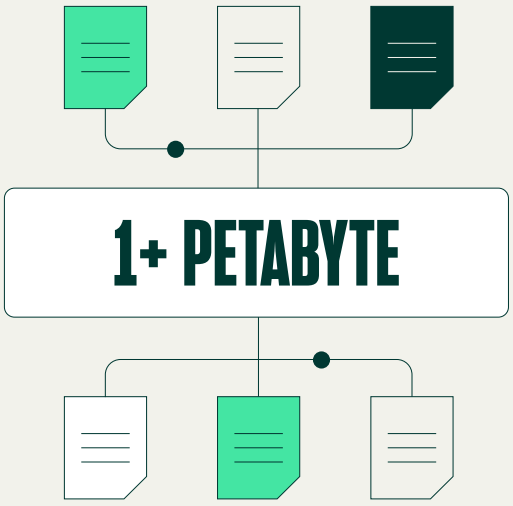


442 BETB Telecommunications leads all industries with 442 BETB in a typical organization.

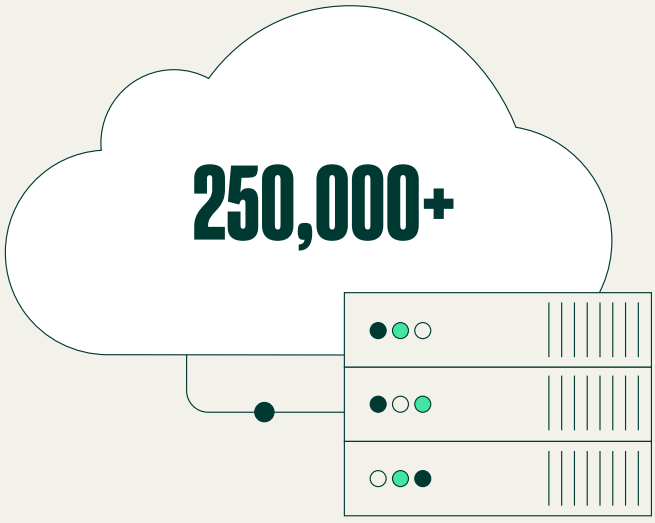
34 BETB Wholesale Trade by comparison comes in the smallest at 34 BETB – about a thirteenth the size of a typical Telecommunications organization

A typical APAC organization has about 60% less data volume than a similar organization in EMEA or the AMERICAS:





Three Rubrik-protected organizations contain more than a petabyte of backend data storage.



The highest number of Rubrik-protected virtual machines in a single organization to date is 250,000+ VMs.

SIDEQUEST!

IT and security teams have been thinking about the growing influx of data for a while.

2008

Wired Magazine declares the beginning of **"The Petabyte Age."**[2]

2010

Former Google VP Marissa Mayer notes three big changes to internet data—**real-time data, unprecedented processing power, and new kinds of data**. She claims user-generated data alone is growing at 15x, which is faster than Moore's Law.[3]

2011

Data industry reporting indicates emerging **CRM pushes will lead to more data** about organizations, a growing number of **"Internet of Things" devices will exponentially raise data levels**, and pending database improvements will expand data use cases.[4]

2013

Discussions center around **a data "tipping point"** resulting from data moving from analog to digitized and increasing "sensor" data.[5]

2014

Analysts argue social media, mobile, and web advertising will drive a **"big data explosion."**[6]

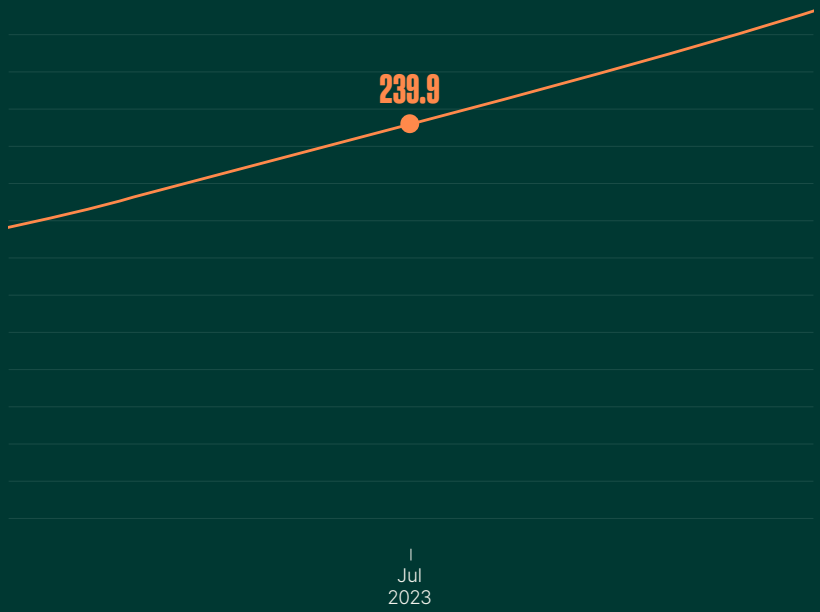
2016

IBM estimates that poor quality data costs US businesses $3.1 trillion annually.[7]

How We Got Here

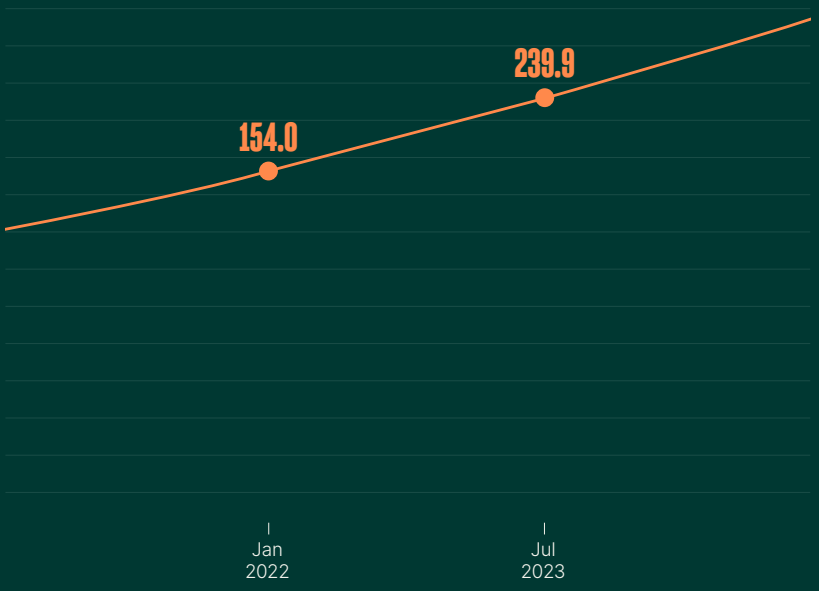
We've been worrying about data for as long as it's been around. But we ain't seen nothing yet. Let's look at how we got to this spot.®

Total BETB Avg



Rubrik Zero Labs analyzed growth patterns from Jan 2022 through July 2023.[1]

Total BETB Avg



A typical global organization's data is growing at the following rates across the last 18 months:

TOTAL: 42%

ON-PREMISES
20%

CLOUD
73%

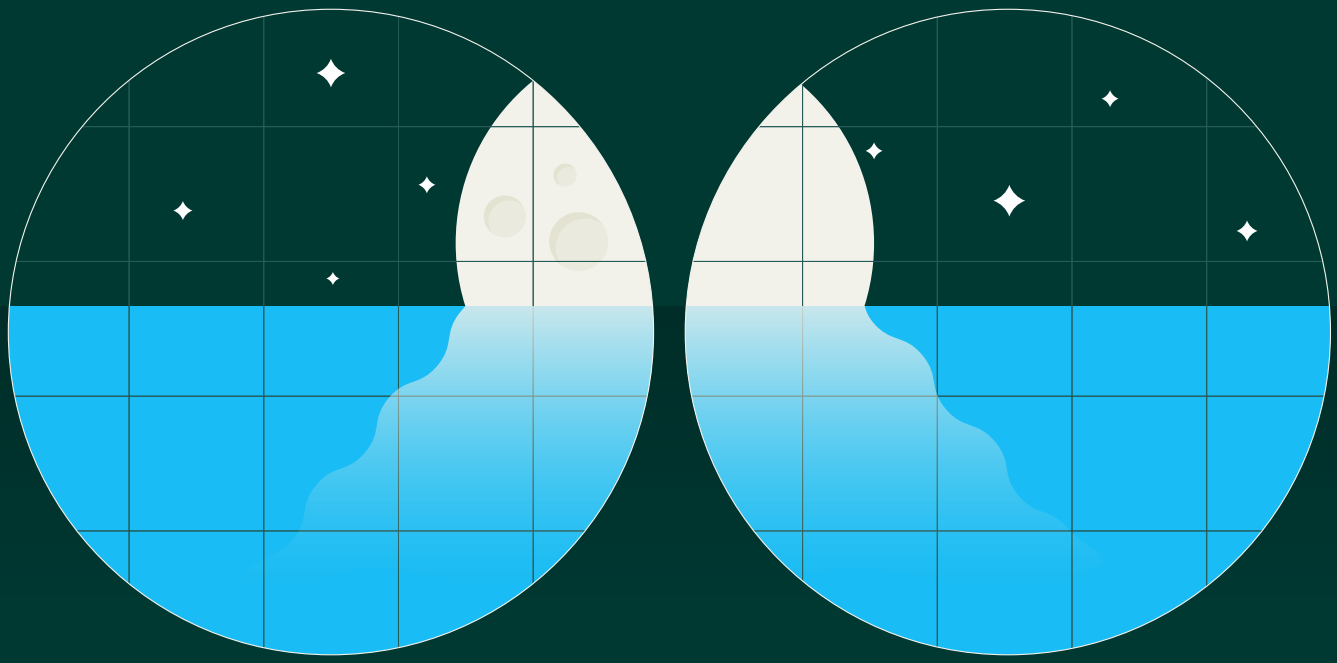
SAAS
145%

1 https://www.caranddriver.com/features/a14989657/pontiac-aztek-the-story-of-a-vehicle-best-forgotten-feature/
2 The End of Theory: The Data Deluge Makes the Scientific Method Obsolete
3 The Coming Data Explosion - The New York Times
4 https://www.smartdatacollective.com/where-did-data-explosion-come/
5 https://news.microsoft.com/2013/02/11/the-big-bang-how-the-big-data-explosion-is-changing-the-world/
6 https://www.informit.com/articles/article.aspx?p=2238298&seqNum=3
7 https://hbr.org/2016/09/bad-data-costs-the-u-s-3-trillion-per-year

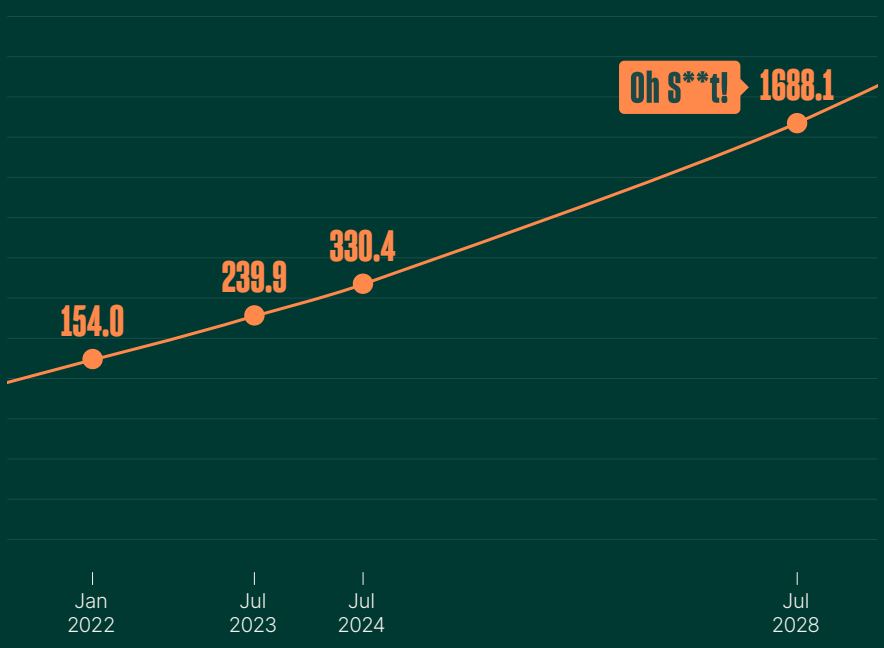
DATA PREDICTIONS

LOOKING AT THE HORIZON

Now let's look forward instead of backward. Rubrik Zero Labs applied the data growth trends and extrapolated to find what organizations are likely to see in the future.



The total volume of data a typical organization needs to secure increases by almost 100 BETB in the next year and by 7x in the next five years.[R]



WOOOOAH, RIGHT? WE KNOW! SAID THE SAME THING!

Big changes rarely happen overnight. By the time US banks and mortgage companies started failing in mid-2007, the first mortgage backed security was already more than 35 years old, and adjustable-rate mortgages had been gaining in popularity since the 1980s.[1] [2]

Nobody really knows how much data we'll have in the next five years. We've regularly outpaced data-growth predictions. Fairness in advertising: This truth applies to Rubrik Zero Labs as well. Our previous research, based on data solely from 2022, indicated a 25% data growth rate. But if we extrapolate other areas based on today's growth rate... It's about to get real.

1 https://www.investopedia.com/terms/m/mbs.asp
2 https://predatorylending.duke.edu/business-analysis/evolution-of-mortgage-lending/subprime-lending/

SENSITIVE DATA CHANGES

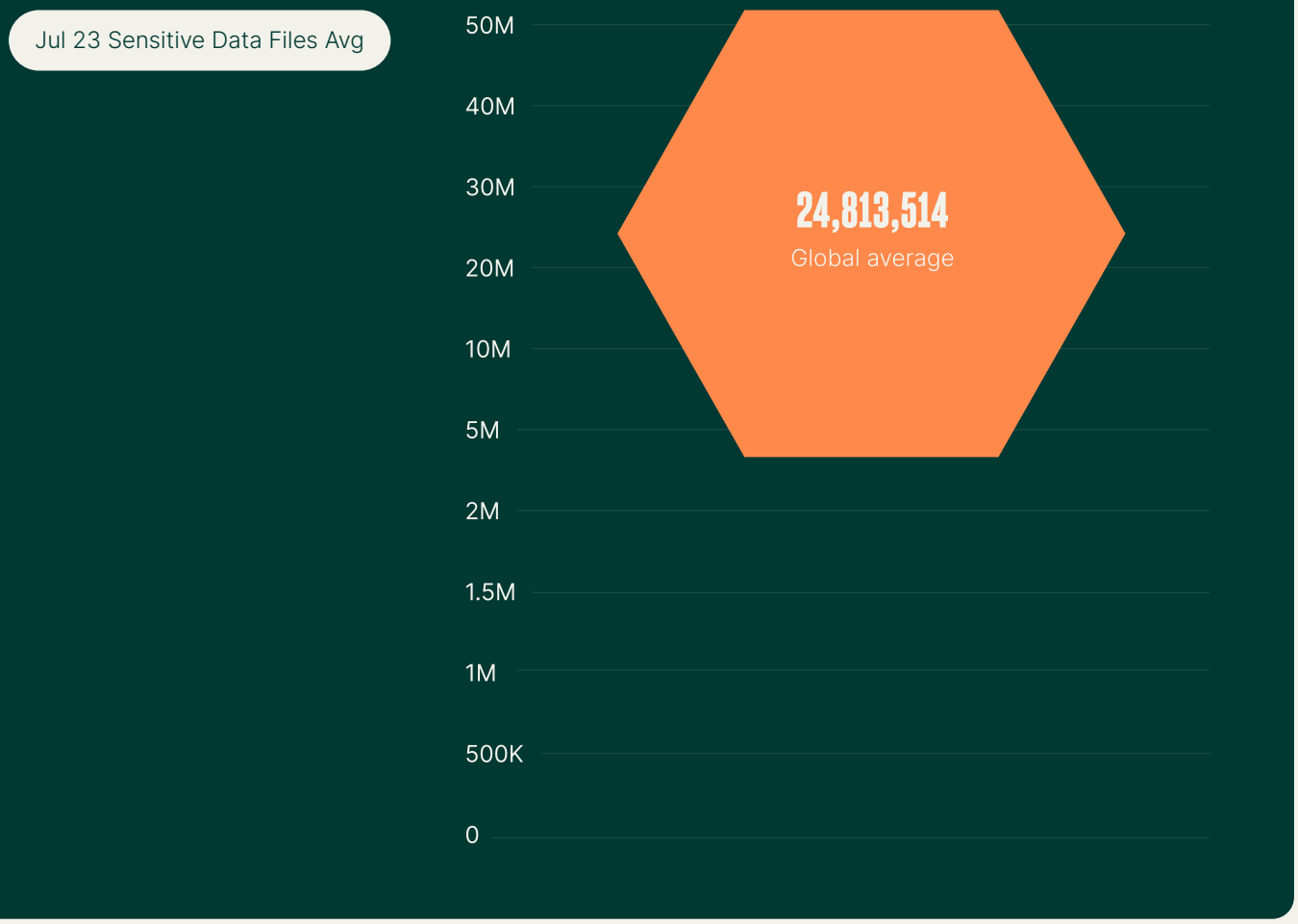
PACK ONLY THE ESSENTIALS FOR THE JOURNEY

You can see how much cargo you're scheduled to carry on this trip, but does the load need to be this heavy?

We care about almost all of our data, but some matters more. For example, no one really cares if a desktop folder full of past blogs is compromised. But things like ID numbers, health records, and business plans... We care a bit more about those.

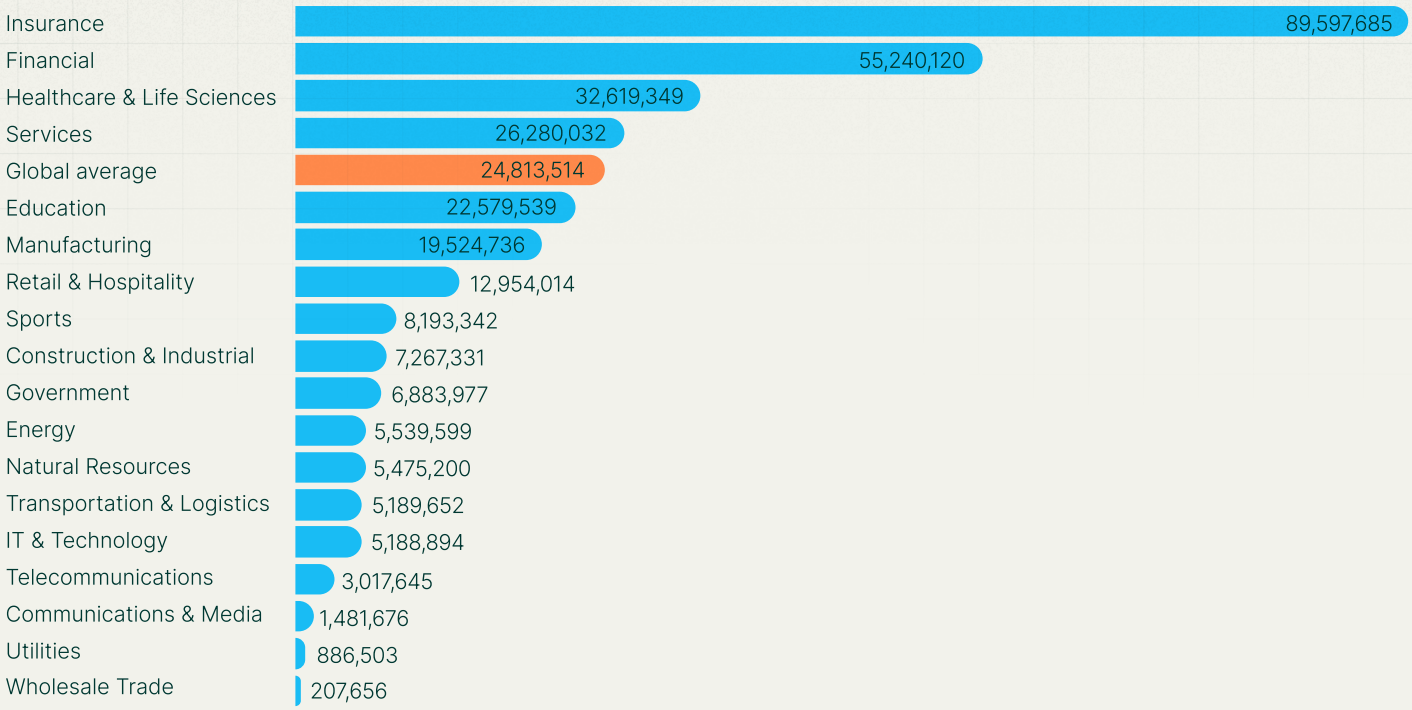
Sensitive data in this report is defined as data elements flagged sensitive by Rubrik technology solutions using parameters derived from various industry standards or regulations—such as PII, HIPAA, GDPR, and CPAA.[1234] Additionally, a given organization can tag data as sensitive using Rubrik technology based on any factor of their choosing (source code, legal hold, etc). That data is also included in these numbers.

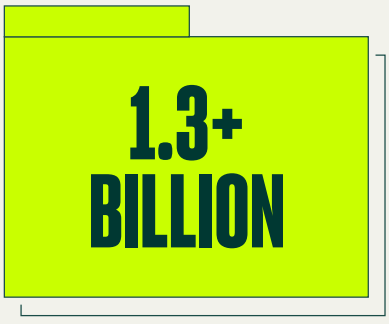
Here's the global average for a typical organization:



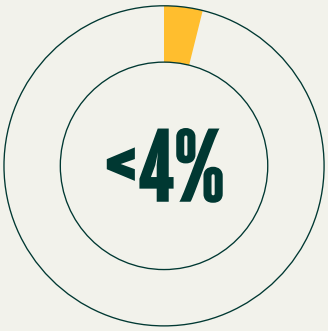
If we step back, there's significant variation. Here's all industries.[RT]

Jul 23 Sensitive Data Files Avg

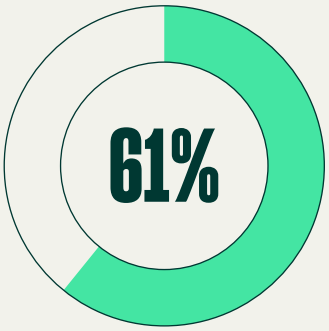




The most sensitive data in a single Rubrik-secured organization is 1.3+ bllllon sensitive data records.



Less than 4% of external organizations have a dedicated sensitive data storage location.[MR]

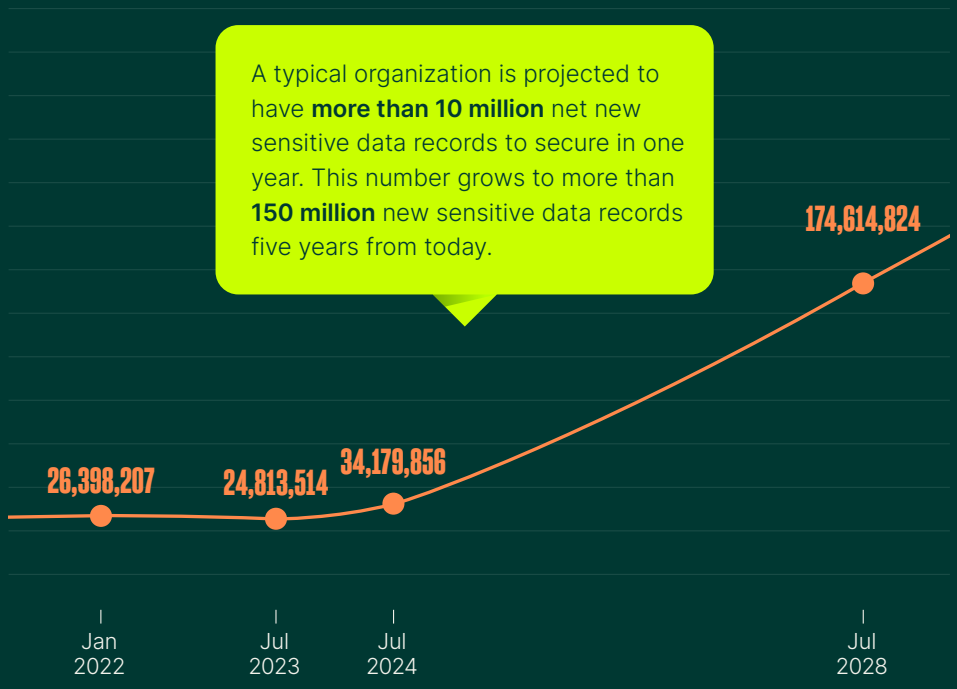


61% of external organizations store sensitive data in multiple locations across cloud, on-premises, and SaaS environments.[MR]

Time to see the projected growth of sensitive data you'll have to manage in the future:[RT]

Sensitive Data Files Avg

A typical organization is projected to have **more than 10 million** net new sensitive data records to secure in one year. This number grows to more than **150 million** new sensitive data records five years from today.

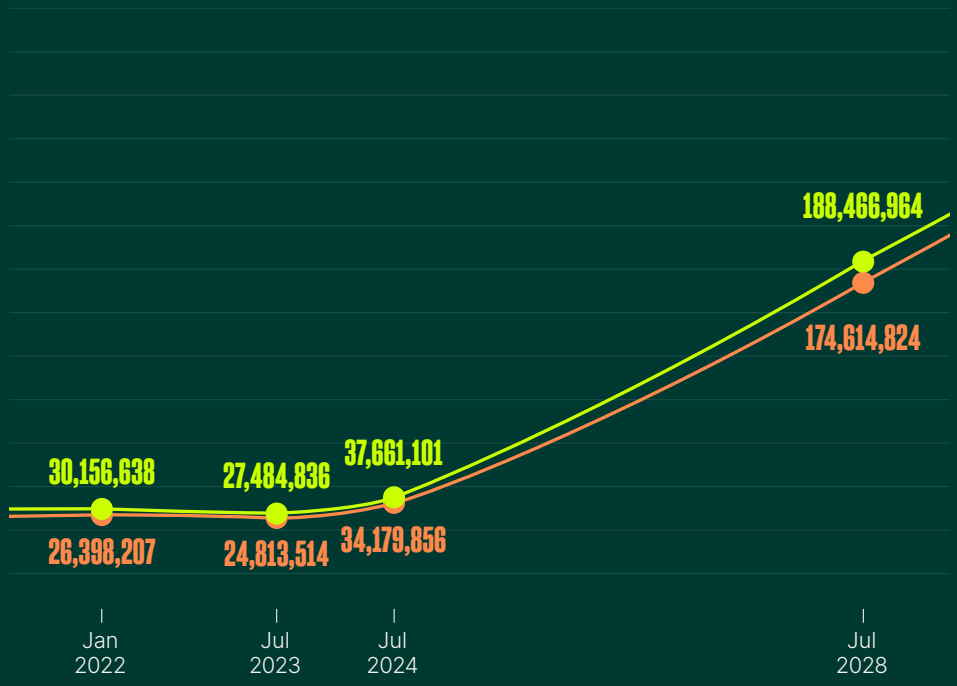


Global average

REGIONAL BREAKDOWN

AMER

Sensitive Data Files Avg



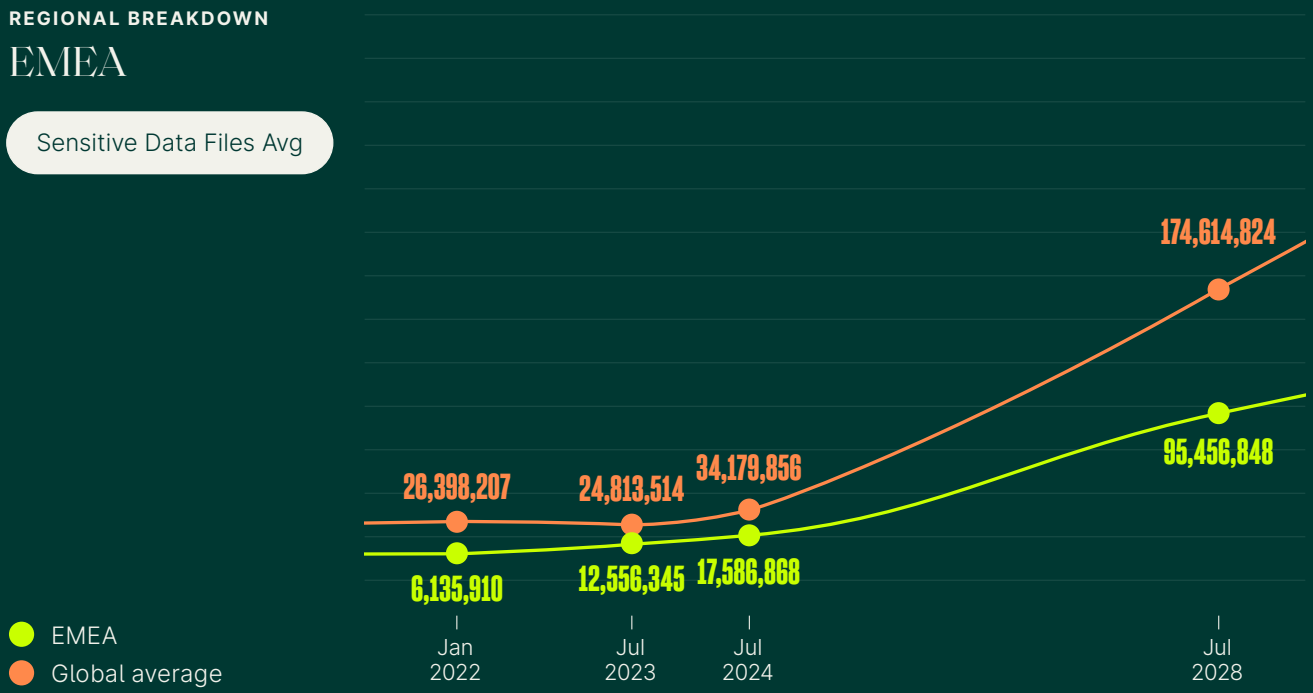
AMER

Global average

REGIONAL BREAKDOWN

EMEA

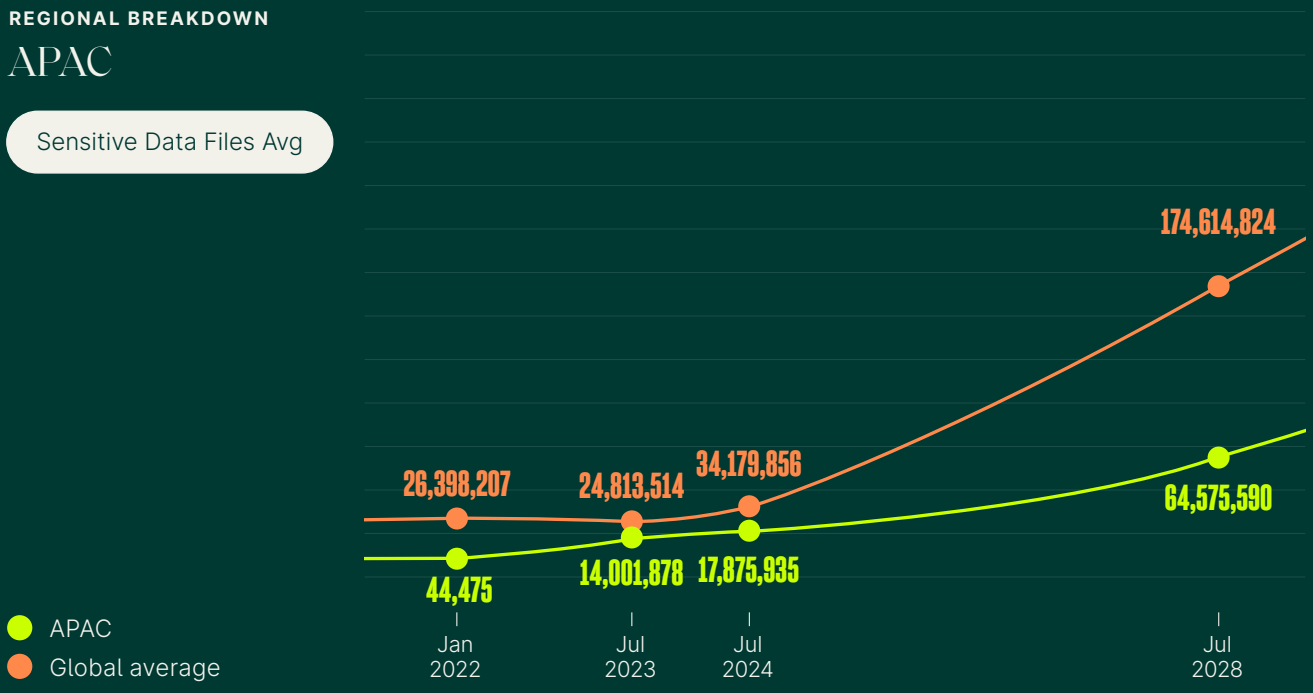
Sensitive Data Files Avg



REGIONAL BREAKDOWN

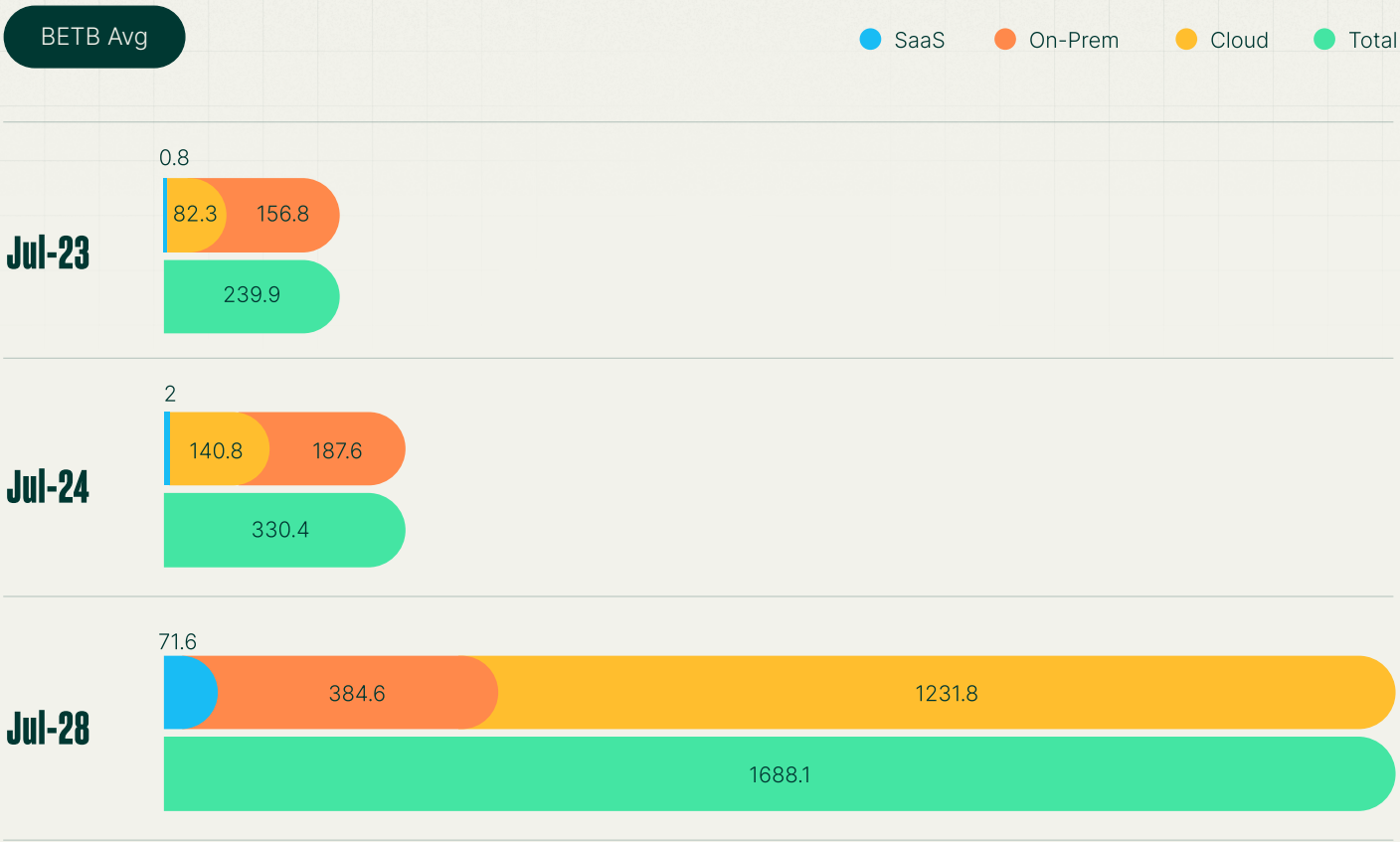
APAC

Sensitive Data Files Avg



Shifting Hybrid Environments

It's not just about how much data you have. It's also about where that data lives.®



An important trend for the next five years is the dynamic nature of hybrid environments (a mix of on-premises, cloud, and SaaS). Every indication points to hybrid remaining the dominant environment type, but the ratios will dramatically shift. Cloud will easily surpass on-premises as the preferred storage medium. SaaS will experience the highest growth rate over the next five years and will equate to what organizations conduct in the cloud today for perspective.

1 https://gdpr-info.eu/art-4-gdpr/
2 https://www.cdc.gov/phlp/publications/topic/hipaa.html
3 https://www.dol.gov/general/ppii
4 https://oag.ca.gov/privacy/ccpa#:~:text=The%20right%20to%20limit%20the,personal%20information%20collected%20about%20them.

DATA SECURITY EVALUATION

HERE THERE BE MONSTERS

OK, so there's a lot of data to secure and we're shifting where it lives. How are we doing when it comes to securing our data today? Let's use Data Security Scores for some data-driven ground truth.

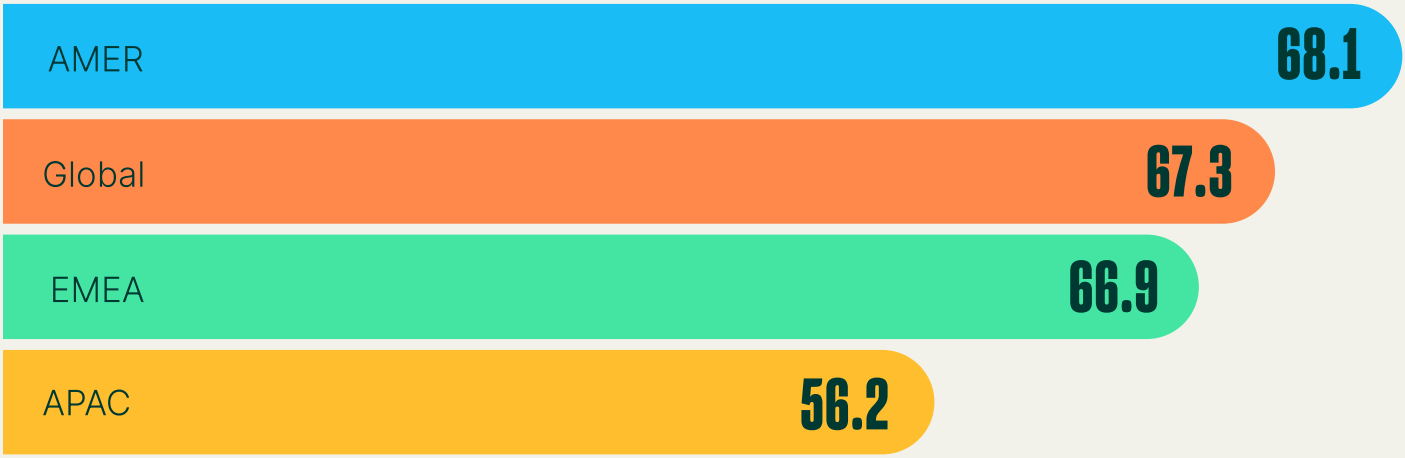
'HERE THERE BE MONSTERS' WAS A PHRASE WRITTEN ON MAPS TO INDICATE UNCHARTED WATERS OR SUSPECTED TREACHEROUS AREAS.

About Data Security Scores

Data Security Scores are calculated every 24 hours based on the following categories:

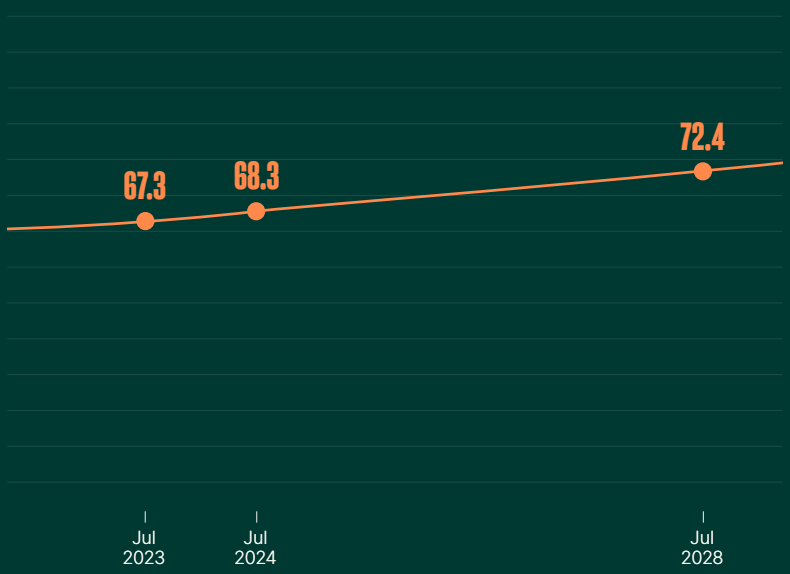
1. **Platform Security:** Measures the effectiveness of infrastructure security where data is stored and includes factors like user controls, administrative authentication, audit logs, etc.
2. **Data Protection and Recovery:** Analyzes how well the backup data is secured, if a clean copy of the latest backup is available, and other related factors.
3. **Ransomware Investigation:** Determines quality and frequency of ransomware threat monitoring and if data can be recovered after an encryption event.
4. **Sensitive Data Discovery:** Measures how much sensitive data is being protected, evaluates access controls for this data, and determines if sensitive data is prioritized for recovery.
5. **Scores are assessed as following:**
   - 0-50: Unsatisfactory
   - 51-75: Needs Improvement
   - 76-90: Satisfactory
   - 91+: Excellent

A typical global organization has a data security score of 67.®



Here's how the future will look based on analyzing these trends and vectors.®

Data Security Score Avg



Global average

8%

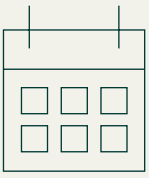
Data Security Scores only increase by a projected 1% in the next year and 8% within five years. ®

16%

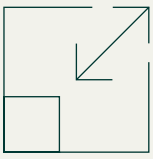
These predictions show a marked shift away from an overall increase of 16% in 2022. ®

10 of 18

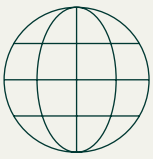
Only 10 of 18 Industries and a single geographic region end up in the Satisfactory or Excellent range five years from now. ®



Data security scores begin to rapidly differentiate the haves and have nots in the next five years. ®

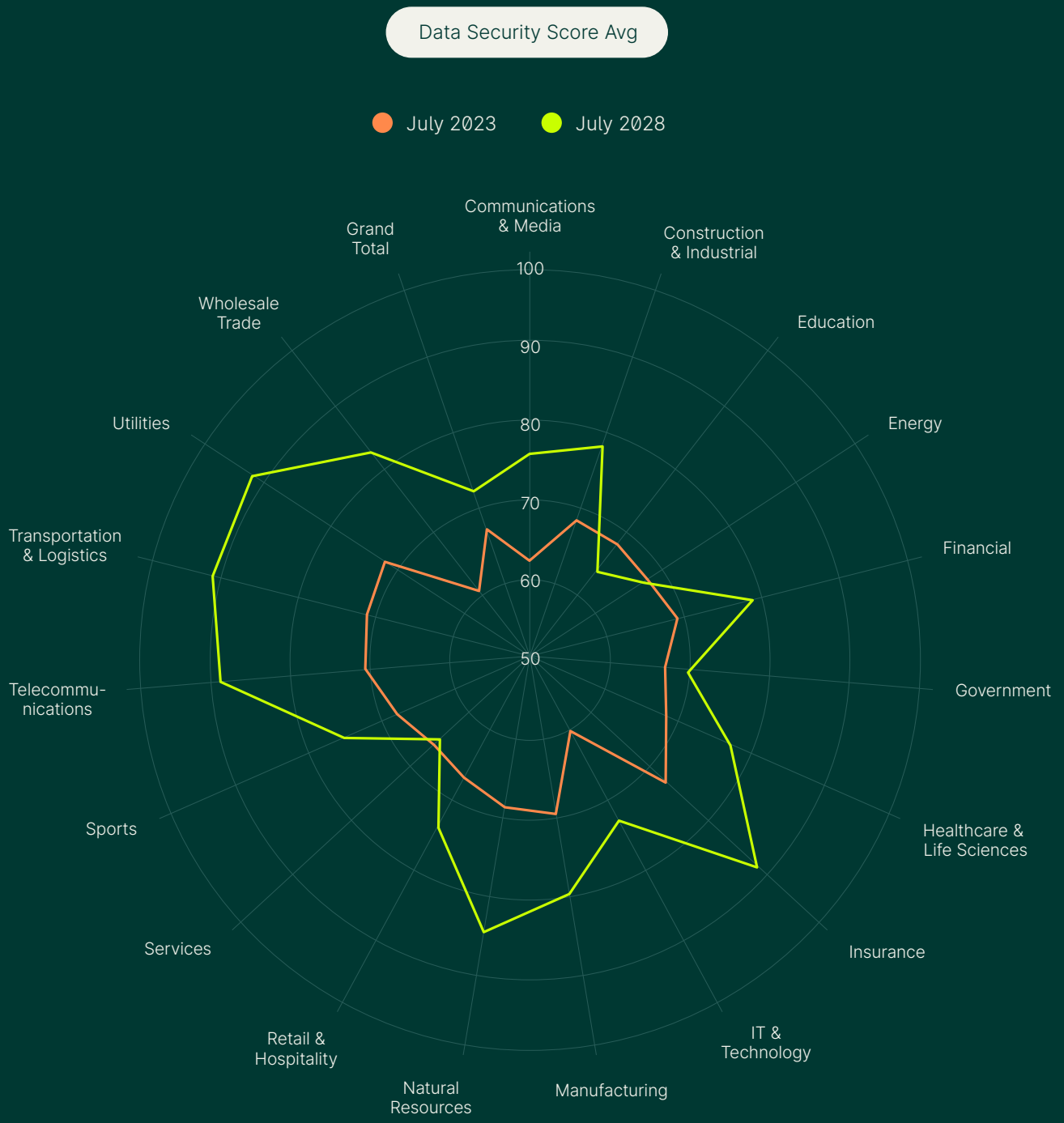


Two industries increase to the 90th percentile (Utilities and Transportation/ Logistics), while three other industries worsen their average in the same timeframe (Education, Energy, and Services). ®



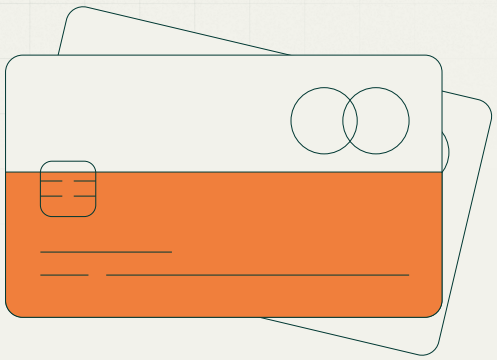
While Americas and EMEA remain close to each other and similarly close to the Global average, APAC projects to look demonstrably different in most major categories. ®

We've examined several data security trends over time. Here's how they look side by side for a bit of a different perspective.®



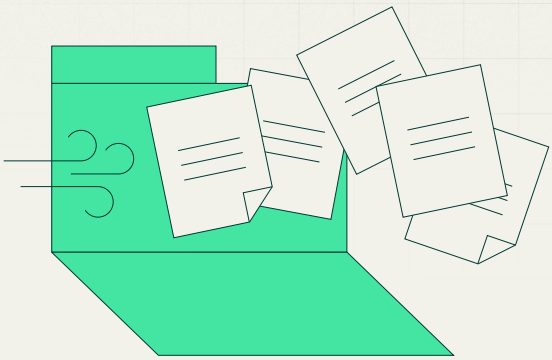
Data Loss Realities

So far, we've used intuition and analysis to look at data as a whole. What about threats to this same data? Lets flip back to intuition and hear from your peers.



53%

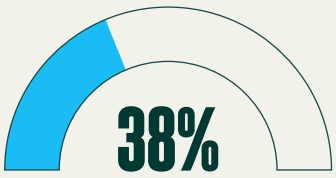
53% of external organizations experienced a material loss of sensitive information last year [VR]



16%

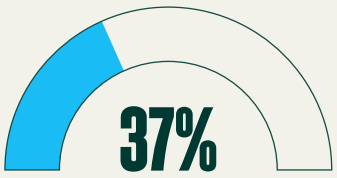
Approximately 1 of every 6 external organizations (16%) experienced multiple material losses of sensitive data in 2022 [VR]

Data Types Compromised in External Organizations Last Year: [VR]



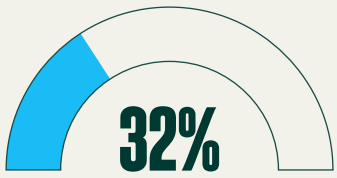
38%

Personally identifiable information



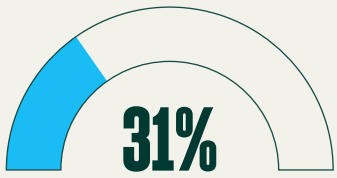
37%

Corporate financial data



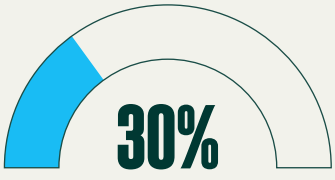
32%

Authentication credentials/Keys



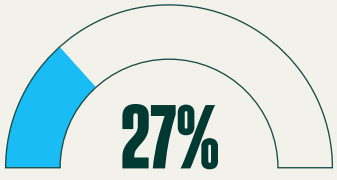
31%

Intellectual property



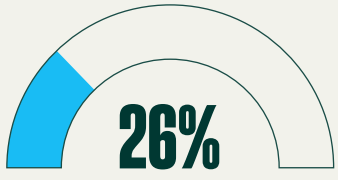
30%

Payment card data



27%

Account numbers



26%

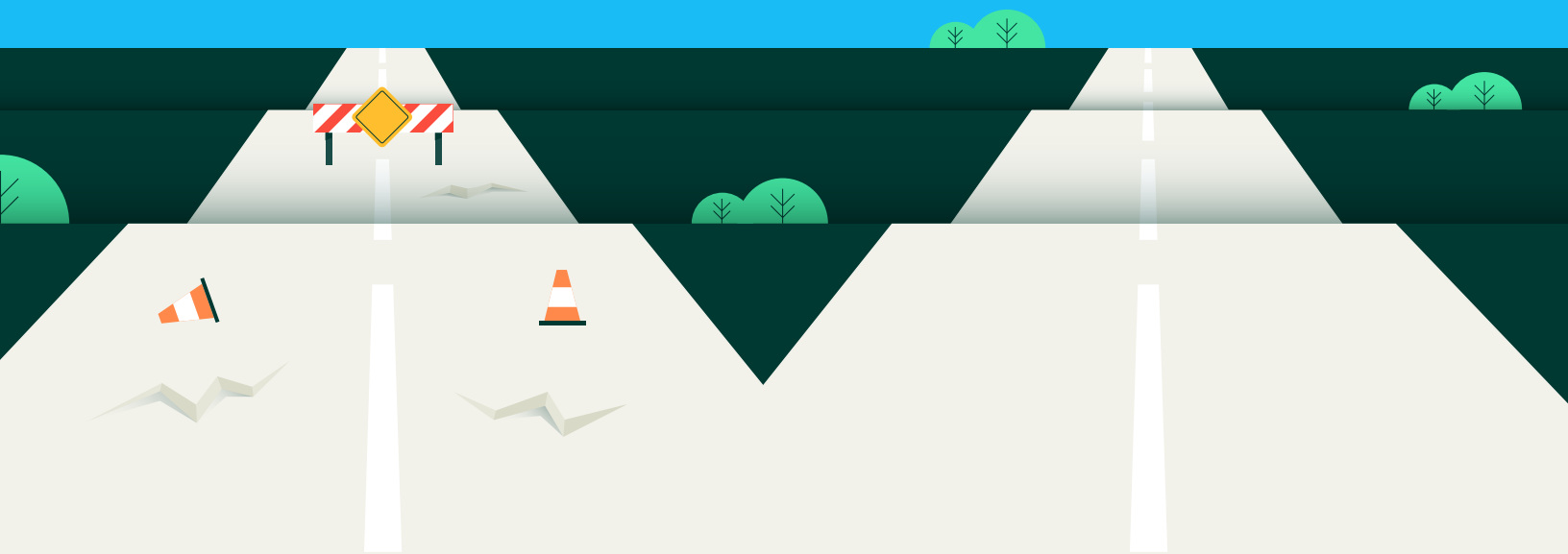
Protected health information

RECOMMENDATIONS

THE ROAD LESS TRAVELED

We might not be great at predicting our futures. But now we have a guidepost to use in charting a new path.

Let's examine some fundamental, proven methods to improve data security. These may sound basic. That's because they are. They also happen to work really well. **Don't trust us–we'll show you.**



98%

The vast majority of external organizations (98%) believe they currently have significant data visibility challenges.[69]

54%

of external organizations have an appointed single senior executive responsible for data and its security.[69]

This future is not written in stone. Our choices today impact our future tomorrow. Let's look at a few examples.

Rubrik Zero Labs Top Three Recommendations

1.

RECOMMENDATION

Leverage data visibility to proactively review sensitive data on a recurring basis.

If an organization implemented a 20% reduction of last year's total data, multiple risk reduction measures improve immediately. Data reduction 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.

2.

RECOMMENDATION

Be intentional about data growth.

Total size, environment choice, and total number of environments are all potential levers. Examples can include setting cloud growth to no more than 50% of the environment total, deleting data based on set policies, limiting total number of storage locations to less than four across all of the organization's environment, or only placing sensitive data in one enclave.

3.

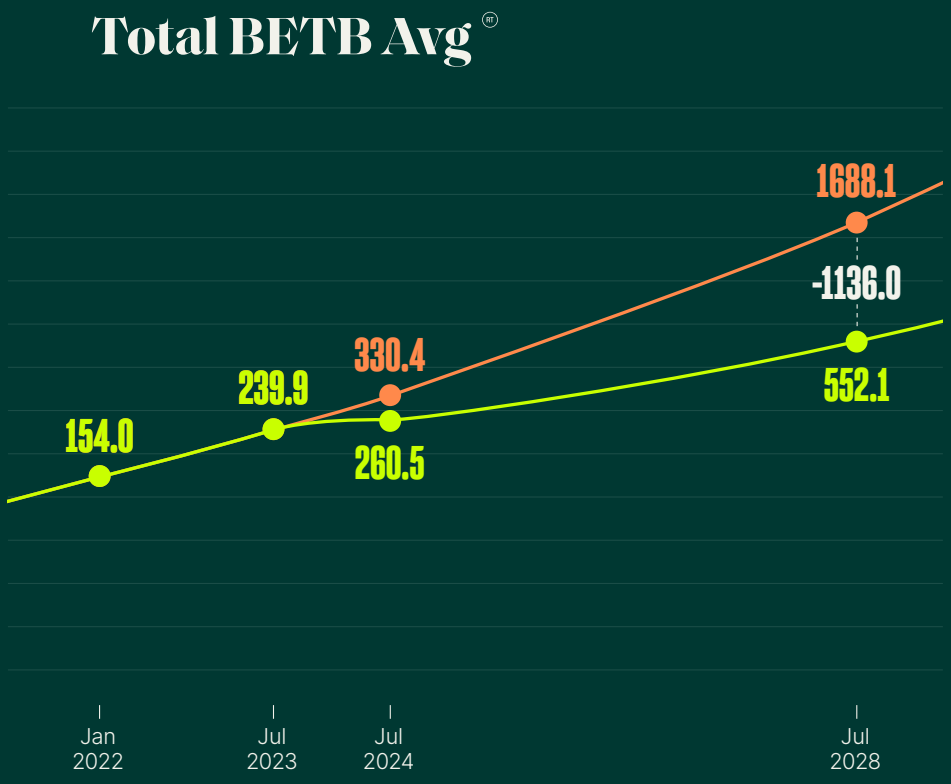
RECOMMENDATION

Make data security a specific focus area at the C-suite level.

A named responsible owner, expected policies, and practice/enforcement of organizational best practices will all produce profound positive impacts and shared responsibility.

Applied Recommendations

Let's see what happens if these changes were implemented today.



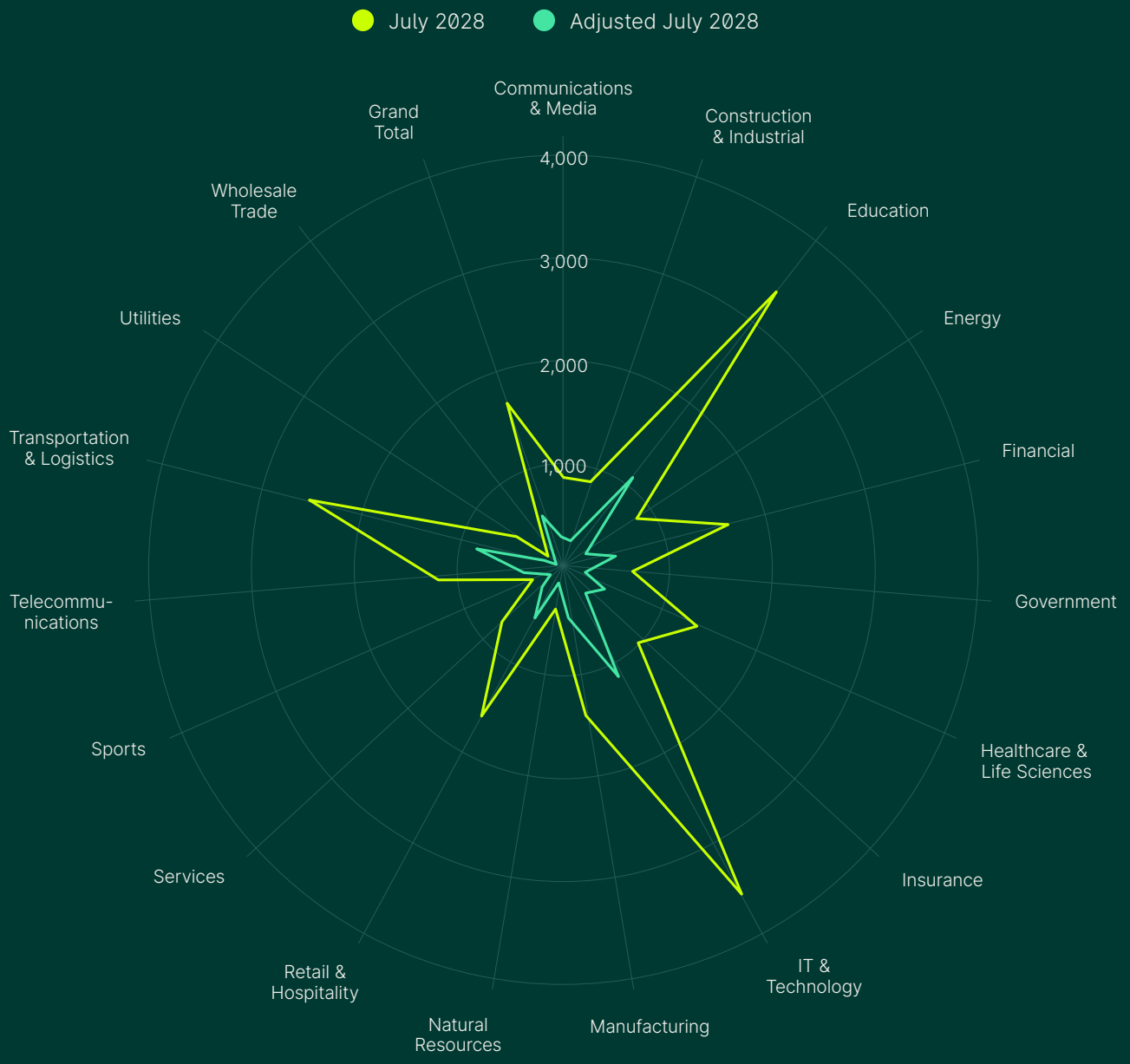
Applying these recommendations would lead to these differences in five years:

1100+ BETB LESS

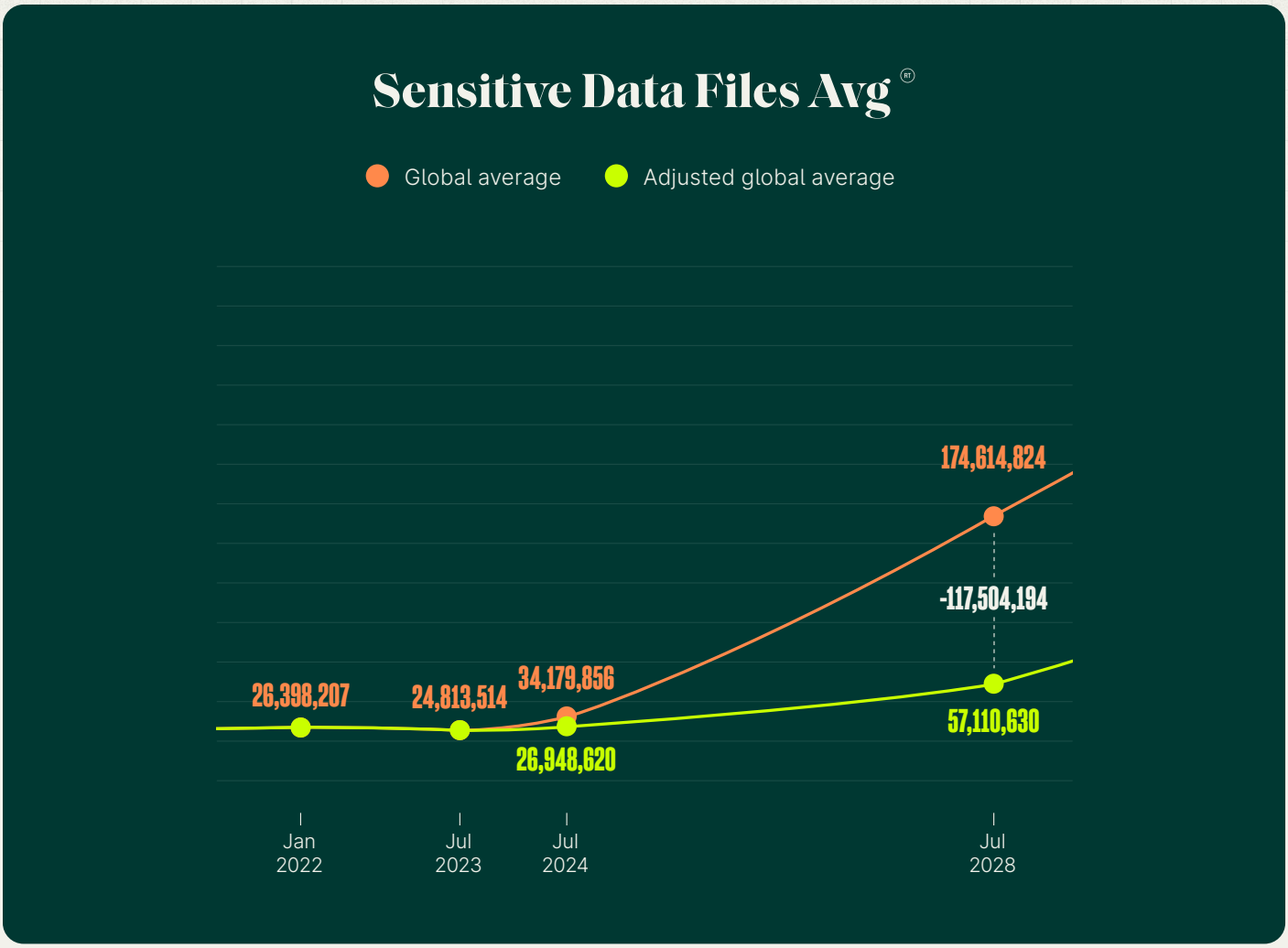
data storage needed for a typical organization

Total BETB Avg®

We can already see demonstrable shifts based on these recommendations. Let's get nuts and compare all of them at the same time for a comprehensive view:



Applied Recommendations

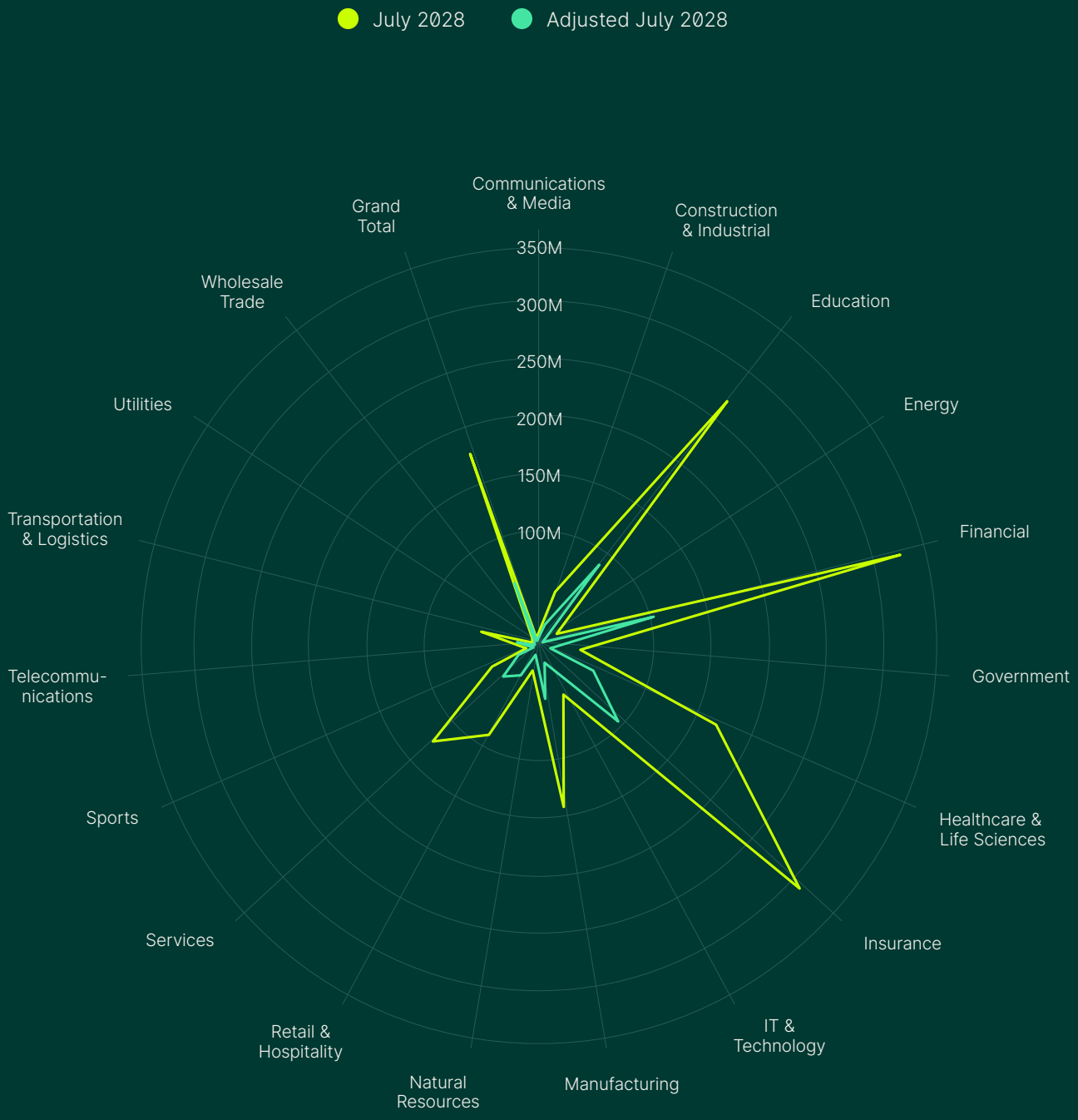


Applying these recommendations would lead to these differences in five years:

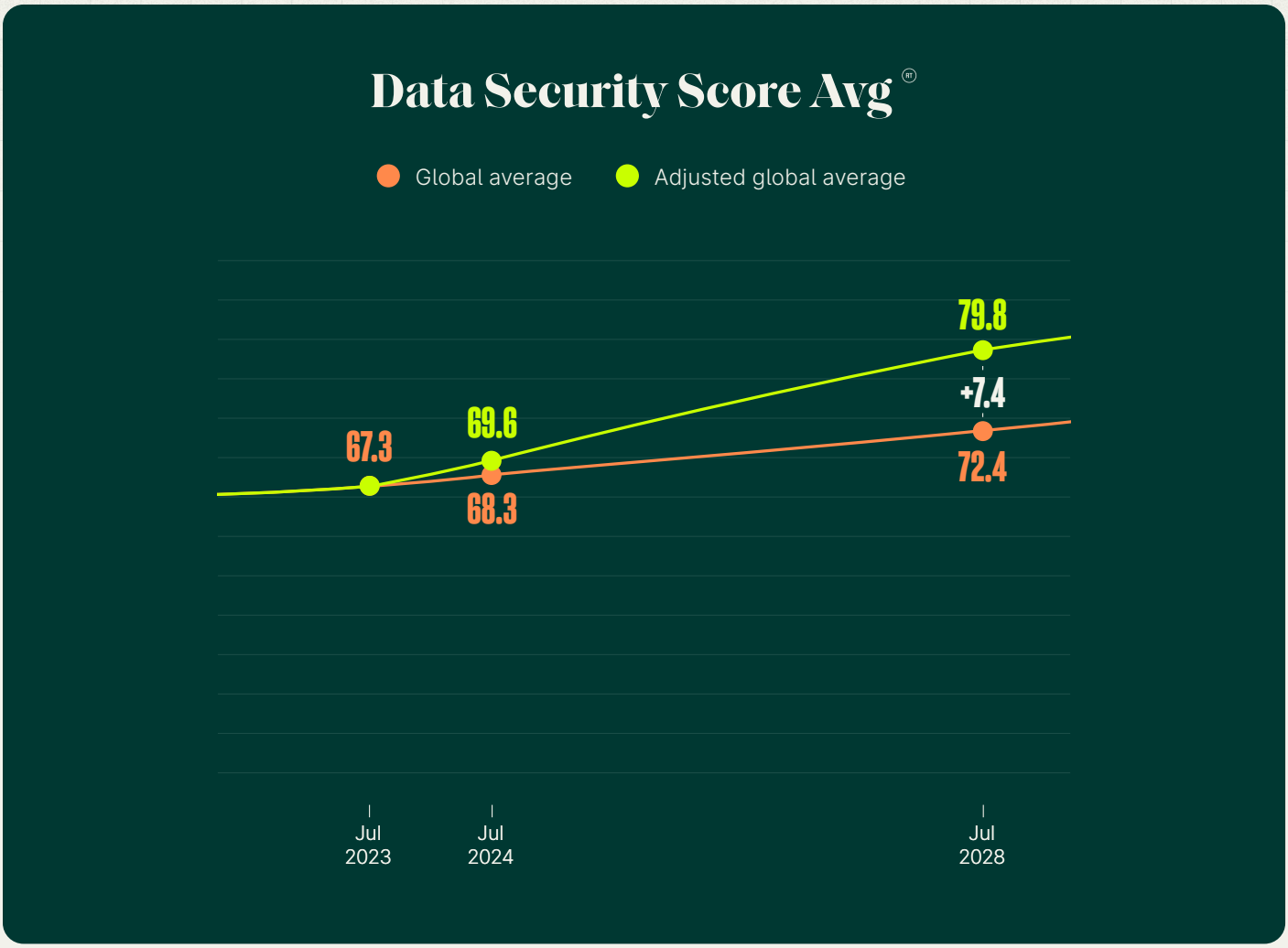
A REDUCTION OF MORE THAN 117 MILLION

sensitive data records in a typical environment

Sensitive Data Files Avg [RI]



Applied Recommendations

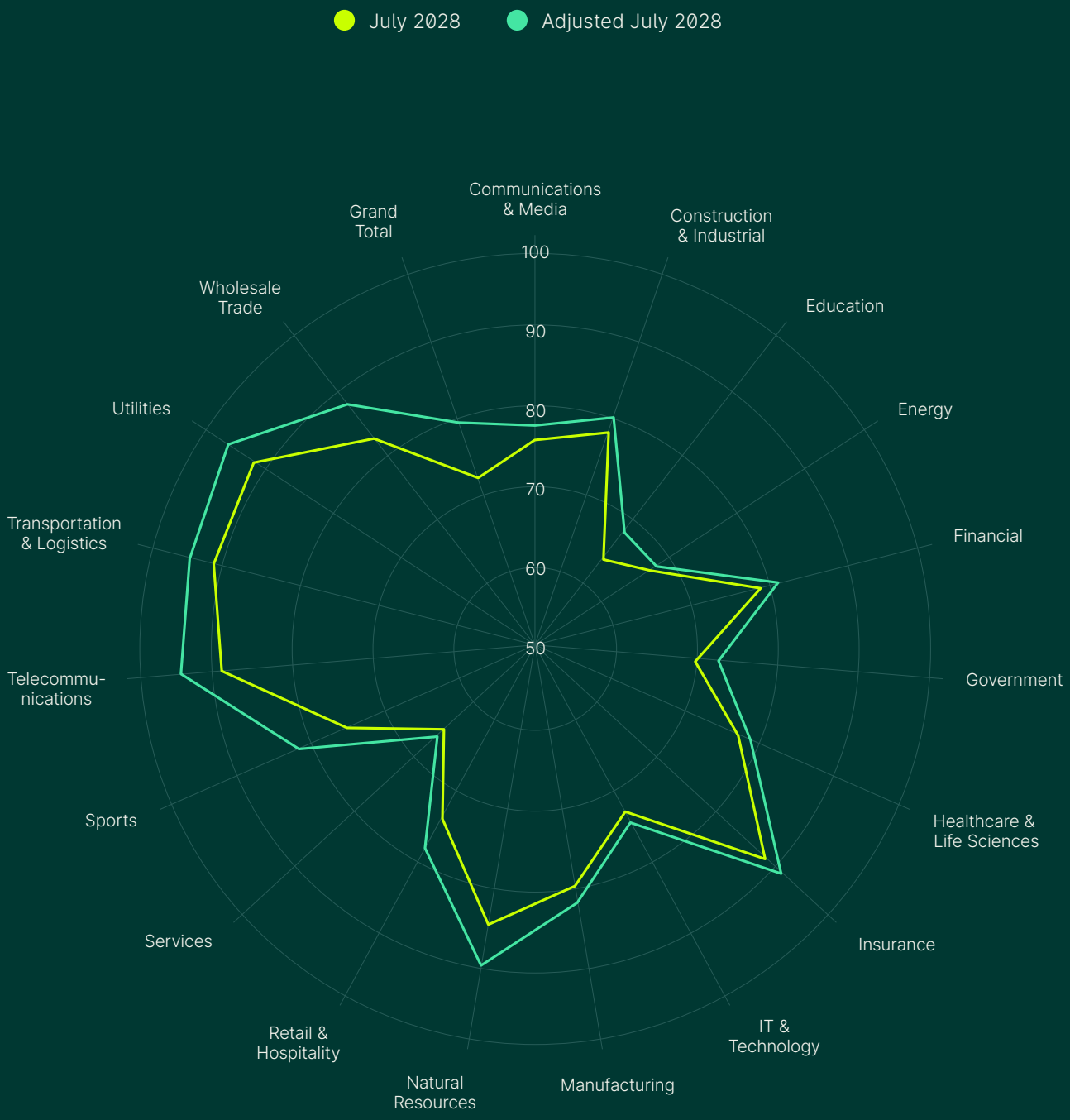


Applying these recommendations would lead to these differences in five years: RT

10% IMPROVEMENT

in Data Security score–more than double the predicted increase

Data Security Score Avg

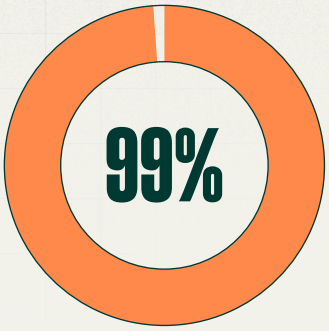


SUMMARY

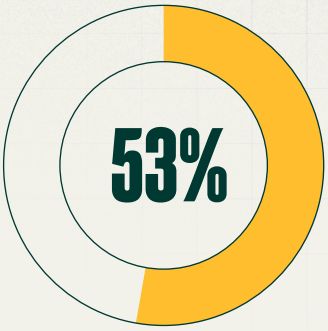
JOURNEY'S END (FOR NOW)

We started this journey by saying humans are inherently optimistic. After this trip together, we're optimistic you have a more actionable understanding of your data and whether or not you've been looking at it through rose-colored glasses.

But is optimism so wrong?
Yes and no.



Number of organisations that experienced an attack last year



Number of organisations that experienced a material loss of sensitive information in the last year

If 99 percent of organizations experienced an attack last year and more than half admit to having a material loss of sensitive information in the last 12 months, the numbers are clearly against you. No right-minded IT or security professional would step into this fray without a certain level of optimism. It's a good thing we're optimistic because we need all hands on deck to tackle the challenges ahead.

But as we've seen, optimism, taken too far, can have dire consequences.

If you take away anything from this report, we hope it's this:

- **Stay optimistic.**
- **Do the analysis.**
- **Heed the signs.**
- **Make informed decisions.**

Acknowledgements

Rubrik would like to end this journey by expressing our sincere gratitude to those responsible for bringing this work to light. Wakefield Research provided data to make this research as objective as possible. Shaped By (www.shaped-by.com), yet again, found an incredible way to take an idea and bring it to life. Finally, many Rubrikans worked hard to provide capability, context, and guidance. It's impossible to thank them all, however we'd like to extend our appreciation to Amanda "Danger" O'Callaghan, Linda Nguyen, Lynda Hall, Ajay Kumar Gaddam, Ryann Goss, Derek Morefield, Josh Burns, Gunakar Goswami, Prasath Mani, Ethan Hagen, Kevin Nguyen, Caleb Tolin, Kelly Cooper, and Olivia Howard.

Rubrik Zero Labs