CYBER-SECURITY Sourcebook 2020

Published by
Information Today, Inc.

Publisher of
database TRENDS AND APPLICATIONS
BDQ BIG DATA QUARTERLY

AEROSPIKE

Next Generation NoSQL Data Platform

Real-Time Security for Today's FinTech

Aerospike for Financial Fraud prevention gives you:

Industry Leading Reliability

Our self healing clusters deliver Five-9's of uptime

Reduced Risk

We enable 10-times the data to be used for fraud predictions

Lowest TCO

Our Hybrid Memory Architecture™ allows 5x reduction in server footprint

Want to learn more?

See our customer case studies at www.aerospike.com/customers/

CYBERSECURITY Sourcebook 2020

From the publishers of database TRENDS AND APPLICATIONS BDQ BIG DATA QUARTERLY

PUBLISHED BY Unisphere Media—a Division of Information Today, Inc.

EDITORIAL & SALES OFFICE 121 Chanlon Road, New Providence, NJ 07974

CORPORATE HEADQUARTERS 143 Old Marlton Pike, Medford, NJ 08055

CYBERSECURITY SOURCEBOOK (ISBN: 2376-7383) is published annually by Information Today, Inc., 143 Old Marlton Pike, Medford, NJ 08055

POSTMASTER
Send all address changes to:
Cybersecurity Sourcebook, 143 Old Marlton Pike, Medford, NJ 08055

PRINTED IN THE UNITED STATES OF AMERICA

Cybersecurity Sourcebook is a resource for IT managers and professionals providing information on the enterprise and technology issues surrounding cybersecurity and the key challenges, opportunities, and technologies, as well as the approaches being evaluated, adopted, and bringing success. The *Cybersecurity Sourcebook* provides in-depth articles on the expanding range of cybersecurity technologies and best practices. Articles cover encryption and data masking, database auditing, database administration, IoT and connected devices, the business of data security, and regulatory compliance.

Information Today, Inc.

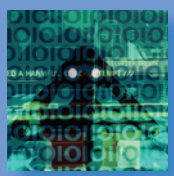
CYBERSECURITY SOURCEBOOK 2020

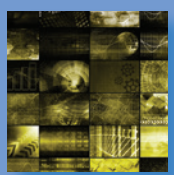
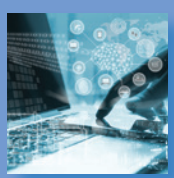
CONTENTS

introduction

cybersecurity updates






It's 2020—Do You Know Where Your Cybersecurity Gaps Are?

By Joyce Wells

IT HAS ALWAYS BEEN IMPORTANT TO CAREFULLY MANAGE DATA and protect it as a valuable asset. But with an increasing array of laws that address data privacy, such as GDPR, CCPA, and new ones expected to follow, it is more critical than ever to put safeguards in place and also know exactly where data is across the enterprise. The embarrassment to companies that fail to address heightened data protection mandates, in addition to the risk of fines and loss of customer trust, is just too great.

According to Dell Technologies' recently introduced "Global Data Protection Index 2020 Snapshot," even with increased investment in data protection measures, disruptive events from cyberattacks, data loss, and downtime pose threats to high-value data. Dell Technologies surveyed 1,000 IT decision makers across 15 countries and 14 industries and found that, in 2019 alone, the average cost of data loss event was $1 million. According to the report, organizations are now managing 13.53 petabytes of data, 40% more than 2018. In addition, 82% have suffered a disruptive event in the last 12 months—up from 76% in 2018—and more than half of respondents are also struggling to find data protection solutions for emerging technologies such as 5G and edge infrastructure and AI/machine learning platforms.

Additionally, research from Ponemon Institute found that 56% of IT security practitioners know their organization's security infrastructure has gaps in coverage that allow attackers to penetrate its defenses. Moreover, 63% of IT security leaders don't share information with their boards on a regular basis. This research closely followed an earlier Ponemon report that found that enterprises are spending $18.4 million on average every year on cybersecurity investments, but 53% are unsure about whether the tools they're using are actually effective. Furthermore, only 41% of companies can accurately identify their own cybersecurity gaps and fix them.

Clearly, more needs to be done. With the 2020 *Cybersecurity Sourcebook*, our goal is shine a light on the pitfalls to avoid and the key approaches and best practices to embrace when addressing data security, governance, and regulatory compliance. The thought-provoking articles on the many interlocking aspects of cybersecurity serve to provide a composite view of the steps to take to safeguard data now and in the future. ■

In Fraud Prevention, the Database Matters (a Case Study)

AEROSPIKE

BARCLAYS IS A VENERABLE, world-leading financial institution that processes 30 million-plus payment transactions a day for its 20 million-plus customers. They know that successful prevention of transaction fraud detection requires a database with ultra low latency. The original data architecture was overly complex as they had to maintain multiple bespoke engineering solutions which struggled to leverage their large-scale user profile datasets across use cases across their business units.

In 2019, it became clear to Barclays that their custom-engineered solutions could no longer meet their objectives. For example, as more time is taken during an end-to-end fraud detection process, more risk is introduced. These risks include stand-in processing (offline account authorization) and the rise of data consistency issues, which in turn can lead to increased false positives and false negatives for subsequent transactions.

A comprehensive analysis of the problem revealed that most of these issues could be traced to a non-optimal database architecture. To that end, Aerospike worked with their Vice President, Enterprise Fraud Architect, Dheeraj Mudgil to put in place the right database for fraud prevention.

Below we summarize the situation and factors in his database choice. (*This case study was presented at the Big Data London Conference, 2019. The video of the presentation can be watched here.*)

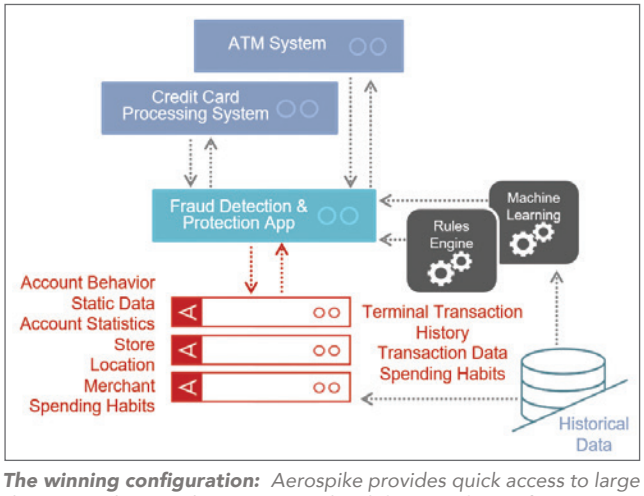
TRENDS THAT MATTER IN FRAUD PREVENTION

Data growth

- 5B internet users (Oct'19) represents an 83% increase in the last 5 years.
- 463 exabytes of data produced daily by 2023

Customer behavior

- Decreasing patience
- Increasing expectations regarding user experience
- Expectation of flawless protection from loss
- Decreasing loyalty due to availability of choice



The winning configuration: *Aerospike provides quick access to large datasets without cache misses. Predictability, simplicity of operations, scalability, high performance, and strong consistency meet SLA security window requirements.*

Fraudster behavior

The sophistication of techniques and tools in the fraud supply chain are growing rapidly. Any user inability to keep pace will lead to a new vulnerability. There are so many steps in credit card authorization (i.e. opportunities for fraud) and so much data to be evaluated— all this needs to be done essentially instantly and with perfect protection from fraud.

Net-net: Exponential growth in complexity of the software is needed to manage this data.

THE KEY NEEDS OF A DATABASE TO PREVENT CREDIT CARD FRAUD

In the new system designed for BarclayCard with the Aerospike data platform, these were the key needs:

Performance — Quick access to large data sets

- One hop to the data from client
- Fast disk access [This indeed is the secret sauce – Patented Technology]
- No cache misses
- Supports IMDB
- Parallel Fetch

Predictability — Helps utilizing set 'Time Budget' effectively

- Known path of data retrieval
- No cache misses
- Reduced Jitter [written in C]
- Simple Architecture – Help TCO & Extensibility
- No caching layer to setup and manage
- Reduced RAM footprint & Cluster size

Supported — The use-cases

- Scaling needs
- Strong consistency & durability
- Standard security features

SUCCESSFULLY BRIDGING THE INNOVATION GAP WITH REAL-TIME DECISIONING

When implementing a new database technology, the payment fraud team at Barclays ended up with a fraud-detection system which solved its problems. The resulting solution scaled the Barclays dataset from 3TB to 30TB-plus over the course of three years, shared fraud rules across platforms, and facilitated machine learning consistently with an aim to achieve a maximum of two digit (<100) millisecond response time for the 99.99 percentile of transactions.

If your role includes more than credit card-fraud prevention, in this Solution Brief you can see these examples in the use cases of Identity Resolution, Settlement and Clearing, Digital Identity Tracking, Frictionless Digital Wallet and more.

CALL TO ACTION

If you have already sorted this issue out (but read this article to see just how your peer did it) may we challenge you to consider the ideas put forward in our recent webinar AI, Machine Learning and Beyond: Changing the Future of Finance.

10 Essentials for Protecting Data

The severity of today's cybersecurity threats, combined with continued stunning growth in data volumes, underscores the urgent need to protect cloud data through a comprehensive certification such as FedRAMP.

By Joe Arthur

Data security is one of the most persistent issues facing federal, state, and local governments as well as commercial enterprises today, and often is one of the most alarming.

A data hack can seemingly come out of nowhere. When it does, it immediately becomes the one, all-encompassing priority that overtakes whatever else organizations expected to be doing that day, week, or month. From ransomware and malware to denial-of-service (DoS) attacks and the notoriously annoying phishing attempts that keep popping up in inboxes, more than 3,800 publicly disclosed data breaches compromised 4.1 billion records in the first 6 months of 2019. This is according to a recent *Forbes* article, which noted, "even more remarkable is the fact that 3.2 billion of those records were exposed by just eight breaches."

The severity of today's cybersecurity threats, combined with continued stunning growth in data volumes, underscores the urgent need to protect cloud data through a comprehensive certification such as the Federal Risk and Authorization Management Program, or FedRAMP. The government-wide program was initiated to establish a risk management, authorization, and continuous monitoring process for the use of cloud computing services.

Cybersecurity is a functional need organizations can never lose sight of, a poster child for the old, overused cliche, "it's a journey, not a destination."

Here are 10 issues that agencies and companies need to consider to ensure they are on the right track.

1. A Big Risk Hits Small Organizations

The first step to building a cybersecure organization is realizing all companies and agencies are at risk. For several years now, it's been a rule of thumb among data security specialists that if organizations think they are impervious to attack, they're actually the most vulnerable.

Here are three takeaways from the "2019 MidYear QuickView Data Breach Report" by RiskBased Security, which found that 2019 was on track to be the worst year on record for breach activity:

- The number of reported breaches grew 54% between mid-year 2018 and 2019.
- The number of exposed records jumped 52%.
- The overwhelming majority of breaches were small, exposing 10,000 or fewer records.

One reason the risk is so rampant is that smaller organizations have less time and resources to optimize their cybersecurity, making them prime pickings for data criminals.

*For several years now, it's been a **rule of thumb** among data security specialists that if organizations think they are **impervious to attack**, they're actually the most vulnerable.*

2. Realize You're Not Alone

The list of the top 10 data breaches hitting U.S. state and local governments shows that hackers aren't the only problem organizations face.

"Some of the biggest and most significant government data breaches come down to human error: from lost hard drives, misconfigured databases, and physical device theft to simple mistakes that lead to millions upon millions of leaked Social Security numbers, names, addresses, voting affiliations, and other sensitive data," *Digital Guardian* reported in a 2018 article. "Adding insult to injury," it stated, "U.S. taxpayers usually end up footing the bill for the aftermath, including years of free identity theft and credit monitoring for the victims."

Here is the Top 10 list:

- The U.S. voter database, 191 million records, December 2015
- The National Archives and Records Administration, 76 million records, October 2009
- The Department of Veterans Affairs, 26.5 million records, May 2006
- The U.S. Office of Personnel Management, 21.5 million records, June 2015
- The Virginia Department of Health Professions, 8.3 million records, May 2019
- The Office of the Texas Attorney General, 6.5 million records, April 2012
- The Georgia Secretary of State office, 6.2 million records, November 2015
- Tricare, 4.9 million records, September 2011
- The South Carolina Department of Revenue, 3.6 million records, October 2012
- The State of Texas, 3.5 million records, April 2011

The list leaves out commercial data breaches such as the well-publicized Target hack (2013) that helped bring cybersecurity to the attention of managers everywhere. It also reveals two critical points: Data loss has been going on for a long time and it affects respected, established organizations that were credible before and after their breaches. The objective is not to single them out, but to address an issue that affects every company and every government agency at every level.

3. Upgrades Deliver Better Cybersecurity

It's a mistake to postpone an organization's cybersecurity planning, but it's easy to understand how it happens.

Planning costs money and soaks up valuable time to change something as fundamental as the way an organization protects its IT systems. However, it costs far more to leave legacy systems in place and vulnerable. There may be many reasons for an agency or company to upgrade its ERP and associated systems—and as those reasons accumulate into an irresistible need, enhanced security is one of the most important gains expected from modernizing software.

In the end, enhanced data security is one of the best reasons to break away from old, obsolete systems that have probably been in place for far too long, and that can't begin to meet the latest compliance standards. At a time when companies and public agencies are looking to maximize efficiency and reduce costs, financial and program management software is stretching to track operations across multiple platforms, and supply chains are becoming more complex, cybersecurity is just one of many compelling reasons to upgrade.

4. Security Across Your Supply Chain

Today's cybersecurity challenges extend beyond in-house systems, which are as strong as their weakest link.

One of the enduring lessons of the Target data breach was that it originated with a vendor so small that it almost certainly wasn't on the security team's radar—until that small company's vulnerability became an entry point to the retail giant's business. In an era of unprecedented complexity, supply chains likely originate 80% of the data that organizations rely on to deliver on their missions. A company or agency securing its own system is just the essential first step. The next challenge is to extend that protective umbrella to every piece of external data that enters its system.

5. A New Wave of Opportunity Awaits

The good news is that it isn't all about threats and potential loss. Enhanced cybersecurity is just one of the advantages organizations tap into when they move their operations into the cloud.

The emergence of smart city strategies is opening the door to wider collaboration, coordination, and optimization across service areas, agencies, and levels of government.

Internet of Things (IoT) technology offers a wealth of sensor data to optimize operations and capture the most granular updates on equipment performance and material flows.

Cloud-based asset management systems help maximize the performance and extend the operating life of expensive and often-specialized capital equipment and property.

Across every aspect of a business, cloud computing offers greater access

*Enhanced data security is **one of the best reasons** to break away from old, obsolete systems that have probably been in place for far too long and can't begin to meet the latest **compliance standards.***

and efficiency with routine, seamless updates that keep operations more current than any on-premise system. But it's only safe to make the move if businesses and agencies have a reliable, secure pathway for bringing all of that data to the cloud.

6. Older IT Isn't Up to the Challenge

The benefits of modern IT infrastructure are just one upgrade away and the need is acute. We constantly hear from CFOs and CIOs whose legacy systems fall short of organizational objectives, are often out-of-date, and frequently hamper efficient operations.

Those issues reflect an ongoing risk to operations posed by legacy systems that are familiar to agencies and companies. Every single day, those systems eat away at an organization's effectiveness, blocking performance improvement, limiting access to best practices, isolating it from emerging technologies, and failing to deliver the ease of use that the next generation of millennial employees expects on the job.

7. FedRAMP Delivers Data Safety and Security

FedRAMP is a one-stop resource for governments at all levels as well as regulated companies that are intent on keeping their data safe and secure. Its primary mission is to keep federal data and U.S. citizens safe in an environment of ever-escalating threats. The program is also open to state and local governments and commercial enterprises that are prepared to leverage its stringent authorization process to increase security, confidence, and innovation in their own cloud strategies.

AI, machine learning, and IoT have the potential to transform organizations' missions and drive business success—but cloud migration is a necessary first move. FedRAMP authorization ensures that every layer of an organization's IT structure, from the operating system to industry-specific applications and data analytics, is continuously monitored and assessed, and that new innovations are quickly integrated into a secure architecture.

8. New Expectations for Contractors

The Department of Defense is working to protect controlled unclassified information within the supply chain and contractor networks. Expected to begin appearing as a requirement in 2020, the Cybersecurity Maturity Model Certification establishes five levels of progressively rigorous security controls that operate across 14 different control families based on standards such as NIST SP 800-171, NIST SP 800-53, and ISO 27001. According to *Government Computer News*, a FedRAMP authorization may satisfy many of the CMMC requirements. Both programs have similar control families—including access control to awareness and training, security assessment, and system and information integrity. Building a deliberate, integrated framework will ensure that an organization's vendors and partners are onboard with the plan as it embarks on its cybersecurity journey.

9. New Expectations from Users

Another reason to embrace a more cybersecure architecture is that an organization's clients, customers, and stakeholders are demanding it.

In 2018, a survey of 374 Infor customers across multiple industries listed innovation, security and compliance, performance and scalability, user experience and adoption, and total cost of ownership as the five top reasons to move to the cloud. Most of the arguments against the transition had to do with system security—which is precisely where FedRAMP comes in. The certification is so comprehensive that organizations' data is probably more at risk in an internal, on-premise system than in a state-of-the-art cloud environment. The longer a company delays the transition, the more serious that risk becomes.

10. Getting the Transition Done

If an agency or company is thinking of FedRAMP authorization for its own operations, the first thing to understand is that it won't be out there alone.

Experienced, third-party cybersecurity advisors are available to guide the process. Once the system is in place, a third-party assessment organization (3PAO) conducts an independent audit to ensure that the organization's security controls meet FedRAMP requirements, while assisting with document development and providing ad hoc engineering support as needed. Both of these highly trained professionals are paid by the cloud services provider the company selects to house its data. ■

Joe Arthur is a vice president, regulatory industry SaaS, at Infor and has been with the company for 6 years. He is responsible for leading the Infor Regulated Industries SaaS (IRIS) business, as well as driving growth for the full IRIS solution. Arthur is an expert in transforming organizations through his extensive aerospace, defense, government contracting, and commercial IT experience. He also has technical acumen specialized in areas including enterprise services, cloud solutions, technology infrastructure, cybersecurity solutions, FedRAMP, and more. Prior to Infor, Arthur worked at Arthur Consulting Services Inc., NJVC-LLC, and Lentech, Inc. He holds a BSBA in finance and computer science from Northeastern University and is currently based in Ashburn, Va.

The Changing Role of the DBA in a Data Protection–First World

DBAs have always been responsible for monitoring and improving performance, but cybersecurity regulations are elevating their roles to the next level.

By Matt Hilbert

DBAS HAVE ALWAYS BEEN CENTRAL TO how organizations manage, store, and use data. First, they were gatekeepers, limiting access to production environments and carefully shepherding database changes through to avoid the risk of data loss or system downtime.

The rise of DevOps, however, changed the game and encouraged them to be more open data enablers. With changes to front-end applications often requiring the database at the back end to be updated more frequently as well, there is a growing demand to share copies of databases with developers to test their changes against. Limiting access to production environments and excluding the database from DevOps hinder the faster pace of development that can otherwise be achieved.

Now, with the global rise of tougher regulations such as GDPR and CCPA, DBAs need to embrace a new role as guardians, protecting data, yet still ensuring it is available in secure, anonymous ways to enable faster development without the risk of breaches.

With more than 62% of the world's population being protected by more stringent data privacy laws moving forward, according to Redgate analysis, there are new challenges for DBAs. Becoming a true data guardian now involves taking 10 steps within four areas:

1. Identifying and cataloging your data

- Where is your data?
- What is your data?
- Where are the risks to your data?

Data spreads across organizations, so DBAs need to create a record of every database, every instance of it, and who has access to it. This can be a bigger task than it first appears because data is used in so many ways. A remote office may have a copy of a customer database open to every employee—for example, a large number of people might have historic access to your production database, and production data is often used in business analysis, sales, and marketing.

The next task is to identify what that data is. It might be standard personal data, such as names and addresses or telephone numbers, or it could be more sensitive, such as a person's ethnic origin or details about the individual's health.

With that knowledge, DBAs will be able to spot any risks that exist and categorize

the data with a taxonomy that allows them to differentiate between personal and sensitive data. Columns can then be tagged to identify what kind of data they contain, and therefore, which need to be protected.

2. Protecting your data

- Reduce the attack surface area.
- Mask data outside production.

Once you have a picture of your data, you can take steps to protect it. As a default, aim to consolidate the storage of data into as few locations as possible, and only let individuals view, modify, or delete personal data that is relevant to their job roles.

Focus on reducing the attack surface area and masking data outside production. Bear in mind that most breaches are not caused by outside hackers, but instead are due to unauthorized access by contractors, third parties, and users without the appropriate permission. That's why companies need to move to a default methodology of "least access," whereby people are only allowed to access the data they need in order to perform their jobs.

This, of course, raises another thorny issue because developers have become accustomed to having access to production data to test their proposed changes again—yet production databases invariably contain the kind of personal data that needs to be protected.

This is where data masking measures such as pseudonymization, encryption, anonymization, and aggregation should be adopted, preferably using a third-party tool to ease the process. These protect data while providing a realistic, accurate set of information that matches the size, distribution characteristics, and referential integrity of the original.

3. Bringing DevOps to your data

- Standardize team-based development.
- Version-control database code.
- Automate where possible.

The rise of DevOps has seen developers being expected to develop the database alongside applications, switching from coding in Java one moment to using languages such as T-SQL the next. Because T-SQL is a looser declarative language, as opposed to an imperative language such as .NET, there are many different styles in use. This can lead to confusion, especially over time, when multiple people have worked on the same code base.

To overcome this, development needs to be standardized—not by forcing developers to change how they work, which would be unpopular and counterproductive, but by adopting tools that can automatically change code to a team's standard style in seconds and perform static code analysis as code is written. This makes the overall code base easier to understand and also flags errors earlier in the development pipeline.

Similarly, version control is becoming standard in DevOps, with developers checking their changes into a common repository so that one source of truth is maintained. The same approach can be used in database development, preferably using tools that integrate with those used for application version control.

Once you have version control in place, you can then look to automate parts of the development process to make it more reliable. Every time a change is committed to version control, for example, a continuous integration process can be triggered to test the change and flag any errors in the code. Errors can be fixed immediately and tested again, before the change is then passed along to a release management tool, where it can be reviewed before being deployed to production.

Going back to the guardian role, this approach aligns with data privacy requirements and helps with compliance as it enables database updates to be delivered in a consistent, repeatable, and reliable way and provides an audit trail of those changes.

4. Monitoring your data

- Back up every change.
- Monitor for compliance.

Every DBA understands the importance of backups, but new data privacy and protection requirements add extra considerations. For example, businesses are expected to be able to restore availability and access to personal data, should any issues occur. Backup schedules will also need to accommodate additional requirements such as data being held for no longer than is necessary. Once the processing for which the data was collected is complete, it will need to be deleted from the backup along with the original database it was stored on.

Businesses therefore need to standardize backup regimes, centralize the management of backups, encrypt them, and have the ability to restore and validate backups when required.

DBAs have always been responsible for monitoring and improving performance, but cybersecurity regulations move this to the next level. For example, companies must now monitor and manage access and ensure data is available and identifiable. If a data breach does occur, they must report it, describing the nature of the breach, the categories and number of individuals concerned, the likely consequences, and the measures taken to address it.

This all makes having an advanced monitoring solution a necessity, enabling DBAs to keep track of the availability of servers and databases containing personal data, and be alerted to issues that could lead to a data breach before it happens.

The DBA's New Role

In a security-led world, personal data has moved from being a business asset to a business risk. DBAs therefore need to embrace their data guardian role to safeguard privacy and focus on security, while still ensuring faster development by adopting the right processes, tools, and mindsets. ■

Matt Hilbert is a technology writer at Redgate Software with 20 years' experience working for many of the world's biggest tech companies—and many of the smallest.

Beyond Cybersecurity Technology: Tips for Preventing and Dealing With Data Breaches

Organizations must ensure that developers are fully educated on the security features of the entire development stack and the risks to the organization of improper configuration and use.

By David Busby

According to Accenture's ninth annual Cost of Cybercrime study, the number of cyberattacks continues to rise and take more time to resolve. Organizations participating in the study saw an average of 145 attacks in 2018, up from 130 in 2017. The good news in the report was that prioritizing technologies to improve cybersecurity protection can reduce the consequences of attacks and "unlock future economic value as higher levels of trust encourage more business from customers."

However, technology can take you only so far. "Whether by accident or intent, many employees are often the root cause of successful cyberattacks," the report noted. As demonstrated by several recently publicized data breaches, a key area of concern is how developers deploy their databases, creating completely unnecessary vulnerabilities that let adversaries stroll into their networks with little effort. To mitigate this threat, companies must reduce the potential for human error, react quickly and appropriately to breach notifications, and embrace the security research community.

***Businesses must create** a culture of security. They need to stop thinking that **'move fast and break things'** will lead to successful innovation.*

Human Error and Data Breaches

Examples of common breaches likely caused by developer errors include those at Verifications.io and Lumin PDF.

In the case of Verifications.io, security researcher Bob Diachenko discovered a non-password-protected 150GB MongoDB instance that included four databases containing plain text entries. He confirmed that the database was owned by Verifications.io, an email verification company that helped clients remove inactive addresses from their bulk email lists. The breach exposed between 800 million and 1 billion records (depending on how they are counted), including verified emails, phone numbers, addresses, dates of birth, social media account details, and financial information.

Lumin PDF, a cloud-based service that lets users view, edit, and share PDF files, recently confirmed that a third party gained access to its system and stole user information, including email addresses and encrypted passwords. The adversary then published a download link to the company's entire user database, a 2.25GB ZIP file that holds a 4.06GB CSV file containing the user records of 24,386,039 Lumin PDF users. Writing on a forum, the adversary claimed to have obtained the data from a MongoDB database that was accessible online without a password.

While it would be reckless to assume that we know exactly how and why these companies left themselves so vulnerable, there are some common reasons that such breaches are occurring. One is that developers, under pressure from above to rapidly bring a product to market, quickly spin up a database to meet their needs and then put it into production without adequate forethought regarding how the database should be configured or isolated.

Another reason is the popularity of MongoDB and Elasticsearch. While these are certainly good products and easy to use, certain options in their default configurations, if not modified, could lead to the exposure of data. MongoDB, for example, has no default authentication set, so if a developer spins up a database and never changes this setting, the database will be accessible without authentication. In versions 2.6 or later, MongoDB does not bind to all interfaces by default, meaning users must change the configuration to bind to external IP addresses in order to function. Users should consider adding authentication at this stage as a matter of course. Elasticsearch now has X-Pack, which, at the appropriate paid-for tier, provides an authentication layer and other security-related features, built in. However, businesses must pay to use the authentication and other security features of X-Pack, so many companies either don't implement this, or implement it for the trial period and let it expire, again leaving the database open with no authentication requirements.

In addition to the lack of authentication, organizations are failing to properly configure network isolation. Take the Capital One breach. More than 100 million consumer applications for credit, including Social Security numbers, were exposed to the public. The problem seems to have stemmed at least in part from a misconfigured open source Web Application Firewall (WAF) being used for an Amazon Web Services (AWS) deployment. The WAF should provide protection against several vulnerabilities that attackers commonly use. However, the WAF was apparently left in the default configuration, which allowed the adversary to manipulate the firewall using the well-known Server Side Request Forgery (SSRF) attack. In addition, the WAF was configured to reach an IP address range that was too broad, allowing access to the AWS metadata service. Finally, and this is speculation, AWS identity access management may have been configured with too many permissions. Also, the AWS CloudTrail service may not have been properly configured based on the facts that the WAF allowed access to the AWS S3 bucket data, and the scope was not completely clear for what was accessed.

Once again, simply focusing on security and ensuring proper configuration of the deployment could have prevented this breach.

What Can Companies Do?

Organizations must ensure that developers are fully educated on the security features of the entire development stack and the risks to the organization of improper configuration and use.

It does no good to build the most secure bank vault in the world if the doors to the vault and the bank are simply left open to the public. And if bank personnel don't understand how to properly lock the doors, they can never make the bank safe. Likewise, developers must understand how to configure their tools for security. In part, this means they should never assume default configurations are secure. In the case of Capital One, developers also needed to understand how to properly configure the cloud to provide appropriate identity and access management (IAM) permissions, as well as the WAF, including the use of whitelists which, if correctly configured, would have blocked 90% of typical attack attempts.

Further, while a firewall is essential for security, by itself, it is never enough. Other tools, whether MongoDB, Elasticsearch, AWS, or any other tool in the stack, must also be properly configured, including authentication requiring valid credentials.

Beyond the development team, vulnerability scanners can be deployed to help spot security issues before applications

are put into production. As organizations begin to grow, they can also create security teams whose role is to oversee the security of production environments. These could include, for example, AWS Certified Architects, who have a deep understanding of AWS security issues. Expert third-party consultants are also available to review the security of application stacks.

Businesses must also create a culture of security. They need to stop thinking that "move fast and break things" will lead to successful innovation. They must stop putting pressure on developers to bring products to market as fast as possible without regard for security. This is a culture change that must start at the top.

Security Researchers and Breach Notifications

How a company responds to a breach notification is just as important as how it tries to prevent breaches in the first place. A screaming headline about millions of records being exposed is almost never how a company learns about its problem. Often, notification comes from a well-intentioned security researcher, or "ethical hacker," who has discovered the vulnerability and is working in a lawful manner to inform the company in the hopes the vulnerability will be quickly closed.

For example, Google's Project Zero, a team of security analysts focused on finding zero-day vulnerabilities, has a 90-day disclosure policy. Once Project Zero notifies a company of a vulnerability, that organization has up to 90 days to fix the problem before Project Zero goes public. The idea is that it is far better for the company to be able to say the problem has already been fixed when the problem is disclosed, than for people to see a headline that it has ignored the problem for 90 days and it still exists.

When a notification comes in from a security researcher or other source, the company may have the opportunity to act quickly before any real damage is done. The company should accept the report, ask the researcher for any additional details or other evidence of the vulnerability, and work with the researcher to bring the breach under control and limit the exposure. If there is evidence that sensitive information was accessed, the company should also consult with its legal or compliance department to follow proper disclosure guidelines.

Unfortunately, many companies respond to this type of breach notification by ignoring it or getting angry. Sometimes a company requests an NDA in an effort to prevent the research from going public and turns to company lawyers in hopes of avoiding any public airing of the breach.

> ***Companies that want to successfully*** *defend themselves against data breaches must go beyond cybersecurity technologies. They must* ***reduce the potential for human error*** *and know how to react when something goes wrong.*

This is usually the worst reaction a company can have. Breaches typically become public anyway—sometimes thanks to a frustrated researcher going public out of desperation—resulting in more damage to the brand than if the company simply owned up to the breach in the first place.

Also consider that it is likely that security researchers who reach out to a company are acting in good faith. Otherwise, they would have taken the exposed data and used it for a nefarious and profitable purpose. It is far better to appreciate the efforts of these good-faith security researchers—sometimes all they want is a modest "thank you" in the form of swag or other simple acknowledgement—and you can work with them to resolve the issue and follow any necessary disclosure guidelines.

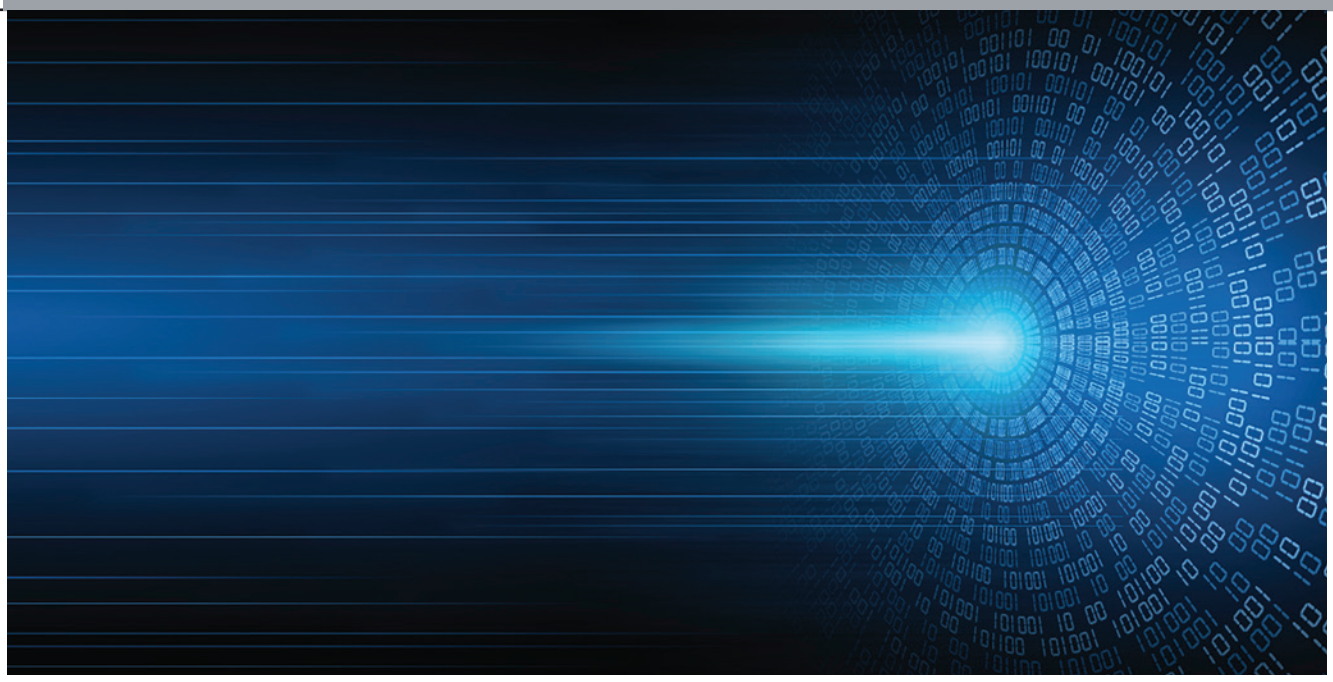
And rather than sitting back and hoping they never receive a breach notification—or ignoring the problem altogether—companies have the option to proactively mitigate the risk of data breaches utilizing a bug-bounty program, such as Bugcrowd, HackerOne or Open Bug Bounty, or hiring a security consultant, such as Rapid7, The Phobos Group, or others.

Companies that want to successfully defend themselves against data breaches should go beyond cybersecurity technologies. They must reduce the potential for human error and know how to react when something goes wrong. They must also embrace the security research community and engage in cybersecurity training across the organization to ensure the following:

- Developers know how to secure the technology stack they are using.
- Employees understand how to safely handle data.
- Executives and the legal team are familiar with deadlines related to breach notification, as well as any regional, state, federal, or international regulatory requirements, including evolving privacy regulations such as GDPR and CCPA.

The organization must also have an incident response plan in place and regularly run tabletop exercises to test the response process, evaluate its effectiveness, and continually improve it. A security expert should also be added to the staff or at least be easily available for consulting. While no security strategy can ever promise to be 100% effective, given the ever-increasing cybersecurity threats, not taking these steps to protect the organization will almost certainly lead to disaster. ■

David Busby is an information security architect at Percona, a company that delivers enterprise-class MySQL, MariaDB and MongoDB, PostgreSQL, and other open source database solutions and services. He has more than 20 years of experience in DevOps, databases, and security and is CISSP-qualified.



Bringing Dark Data to Light: How to Handle the Next Great Business Resource

Dark data collected during business operations that otherwise goes unused is difficult to monitor, meaning it's hard to notice when information has been replicated, leaked, tampered with, lost, or stolen.

By George Kobakhidze

Dark data is a hot topic in the field of data management. Many perceive it as scary and aren't sure where to start in making it something of value. To back up, dark data is defined as data collected during business operations that otherwise goes unused. This unmanaged content is difficult to monitor, meaning it's hard to notice when information has been replicated, leaked, tampered with, lost, or stolen. It's easy to understand the ominous nature of the discussion around it.

"Dark" sounds foreboding, but only serves to highlight the fact that it's not understood. Similar to dark matter in astronomy, it isn't that dark data is to be feared, but rather that we haven't fully realized its potential.

That said, much of the discussion around dark data treats it as little more than an indeterminable nuisance. While there are difficulties in dealing with dark data, it would be foolish to continue ignoring it, as it is both possible to "tame" dark data and to utilize it as an asset, provided a thoughtful, concise, coherent plan is put in place. How can an organization tame something that, by definition, requires a lack of understanding?

Identify the Monster Under the Bed

Dark data is data that isn't currently understood. "Currently" is the operative word. The first step is identifying what exists within the dark data. This is easier said than done and, in fact, is very difficult to do. As such, it's best to separate dark data into smaller pieces, and then break it down.

The goal is to know the repositories of data possessed by an organization, and then to understand the footprint left by that data. Hypothetically, let's say that an organization has a petabyte of data. Upon breaking it down, it's revealed that one-third of this

is comprised of file shares, one-quarter is SharePoint sites, and the rest is email. That's about 80% of the way to understanding dark data, because now, at least, there's an idea of what's out there. This allows further segmentation, enabling things to be further broken down into an examination of the distribution of data.

If one-third of that dark data is comprised of file shares, activity is a good metric to measure. If, of the 200 file shares present, 100 of them are actively managed and modified by users, while the other half are dormant, then the organization has discovered the dangerous part of its dark data and can begin to tame it. If data is sitting unassessed, unattended, or otherwise unoccupied, it's a ticking time bomb—both wasting storage space and containing potentially compromising information. Repeating this procedure allows the taming process to be completed.

Classifying, separating, and codifying dark data into something understandable gives an organization the ability to say, "I don't know everything about this data, but it's being actively used." The danger comes from unknown data that isn't actively used.

After an organization has finished codifying its dark data, it can become an invaluable asset. At the end of the day, dark data is data. It's unstructured data, meaning that instead of traditional 1s and 0s, it's all other content; but it's data, nonetheless. Rather than being processed by simple computers though, dark data is produced by the most sophisticated computers to exist: humans. In the dialogue around data, it's described as "the new oil" or "the next gold rush." That data though is incredibly processed and created by simpler computers. It's easy to place value on it, and therefore clear to call it the next hot commodity. Unstructured data is created by humans, which makes it more difficult to value but potentially worth more than oil and gold combined.

Train the Beast and Make It Work for You

Whereas regular, structured data is only useful after processing, dark data is useful in its raw state. It was created by a human for a reason. It can tell a story about a human, the team the human was working with, the state of the team, and the work that was being performed. The inherent value is much greater than structured data. Yes, it's a huge task to analyze it, but there's an incredible amount of value in a raw state. To understand this value though, there must be a plan in place: What is the organization trying to identify?

Ironically, the easiest way to do that involves examining the most structured parts of unstructured data—the metadata. Metadata, such as dates created, accessed, and modified, can help further break things down. It allows context to be determined. Say that out of a 100TB hard drive, most of the data was accessed or otherwise modified around 2007 or 2008.

If data is sitting unassessed, *unattended, or otherwise unoccupied,* ***it's a ticking time bomb****—both wasting storage space and containing potentially* ***compromising information.***

Dark data could be pertinent to understanding historical events such as the crash of 2008. A financial institute could examine that dark data and answer important questions such as these: How have we evolved, if at all? How responsible were we in this issue? What can we do to ensure this never happens again? Dark data can certainly have value extracted from it, but there must be an understanding of what is being sought. Otherwise, there's a giant content index sitting unused.

Move Past the Monsters Guarding the Castle

Understanding dark data can also remove wasteful duplicative effort and increase productivity. The best way to achieve this is to culturally and technologically educate employees to prevent them from doing the same type of work over and over again. Employees often end up recreating the same documents when presented with a specific question that has been answered before.

A better working knowledge of what has been created before would make a difference in allowing employees to remove redundant workflows. Structured workflows allow employees to know the sources of old information, which is where dark data comes in. Presumably, if dark data is "tamed," an organization has some sort of search index. If a mailbox goes back 10 years, a comprehensive understanding of dark data makes searching that backlog much easier than approaching searches with native tools.

Bringing the Beasts to the 'Good Side'

Understanding dark data allows users to cover ground quickly across incredibly wide spans of time. To know this, though, requires proactive, proper education of employees.

This also allows organizations to prevent insider threats. When discussing cybersecurity, it's obviously important to address firewalls and spam filters, but insider threats can be even more damaging than an outside attack. The best way to mitigate this beyond a working knowledge of dark data stores is access management.

Prioritizing privileged users and only allowing access to data within certain windows of time prevent accidental or purposeful leaks. Limiting data with even more rulesets over time is a proactive way to prevent inside threats and can stop things slipping through the cracks. For example, perhaps an employee—and not even a disgruntled one—is accessing data storage from more than 10 years ago. Any personal information created years ago would be capable of being compromised. This may not necessarily be information about the company, but personal, individual privacy

While there are difficulties *in dealing with dark data, it would be foolish to continue ignoring it, as it is both possible to 'tame' dark data and to utilize it as an asset, provided a* ***thoughtful, concise, coherent plan*** *is put in place.*

that could be violated. The company could and should be held accountable for that data, but all of it could be mitigated with proactive access management and control of dark data.

A Call for Legislation

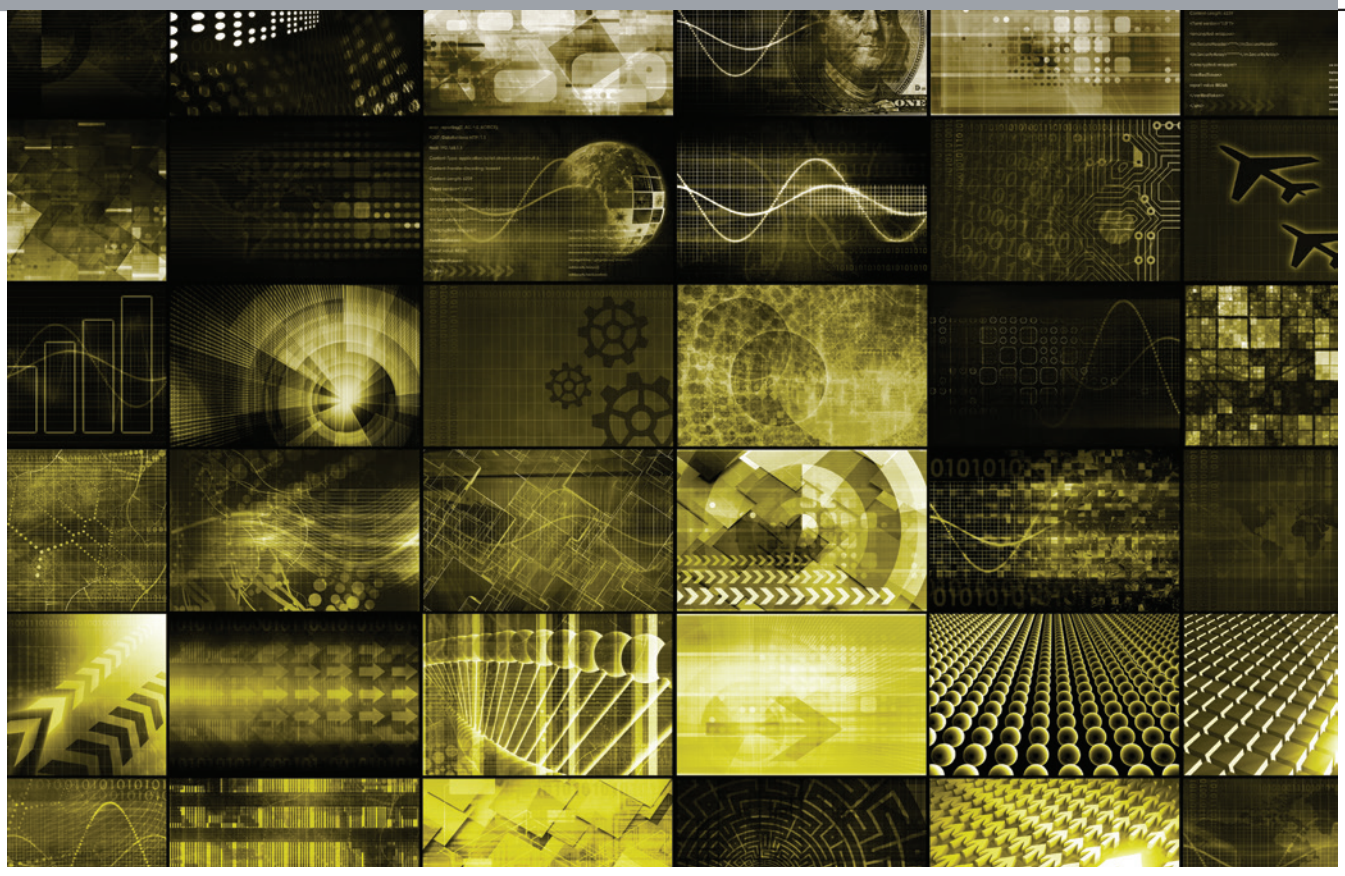
To hold those companies accountable, there should be legislation in the U.S., such as Europe's GDPR. Taking it a step further, U.S. legislation needs to be more technologically thoughtful.

GDPR was intended to be vague, as it gives more power to the people, which is important. That vagueness though can take some of the teeth out of legislation by making it impractical. Enterprise-level organizations likely have robust storage for dark data, but a small website could have different struggles. Smaller sites likely have schemas that don't allow data deletion.

If, in the event a customer recognizes her "right to be forgotten," the whole site could be broken down because of a small site schema error. Maybe those websites shouldn't be allowed in theory, but in practice, that kind of legislation is most damaging to small business owners and entrepreneurs. With the vast amount of dark data created, there should be legislation, but the feasibility of execution of that legislation should be deeply considered. The court of public opinion is a great way to hold data mis-managers accountable, but technologically thoughtful legislation would ultimately do a better job.

Dark data is just data at the end of the day and is therefore an asset that can be leveraged by businesses. This will require a deep understanding of dark data that will involve complicated thinking, but that doesn't mean it shouldn't be dealt with at all. It won't be easy, and it won't be quick, but there are numerous advantages in bringing dark data to light. ■

George Kobakhidze is a solutions engineer for large-scale data governance software solutions with billions of documents in scope. His primary focus at ZL Technologies is data privacy, ediscovery, and data retention management. He earned his bachelor's degree in applied mathematics from the University of California–Los Angeles.



Why Organizations Are Ignoring Vulnerability Reports—And How to Fix the Problem

Instead of being heeded and acted upon, vulnerability reports are often minimized, challenged, or simply ignored.

By Martin Lemay

The message on the business landscape is clear: Vulnerabilities are everywhere. They're in software, in hardware, in processes, and even in people. It only takes a small, unguarded threat vector for hackers to attack, invade, steal data, and wreck reputations.

Yet, despite this awareness, many organizations exhibit a strong resistance—if not outright hostility—toward vulnerability reports. Instead of being heeded and acted upon, these reports are minimized, challenged, or simply ignored. Unfortunately, this pushback is the daily reality for many "bug hunters," security researchers, penetration testers, and internal security teams.

At the same time, it is also true that many vulnerability reports are unintelligible, out of context, or riddled with technical inaccuracies. However, this does not prevent some reporters from aggressively requesting (i.e., virtually demanding) fees for reports that, ultimately, have little or no value. Failure to comply can lead to public shaming, escalation to the CEO, or publishing reports that claim to "expose" a business.

Organizations need to have a clear understanding of what must happen to

Organizations need to have *a clear understanding of what must happen to bridge the gap between* ***vulnerability reporters and actionable threat*** *intelligence to keep their data, customers, and* ***reputations safe.***

bridge the gap between vulnerability reporters and actionable threat intelligence to keep their data, customers, and reputations safe. At a high level, the solution involves the following paradigm shifts:

- Organizations must establish a culture of strong information security, which includes implementing clearly defined processes and procedures to manage vulnerability reports and enable remediation.
- Reporters must understand business-related risks and documentation detail requirements, and they must also ensure that their reports are well-organized, robust, relevant, and accurate.

Let's dive deeper into these shifts and consider how to create common ground between reporters and organizations (including specific business units), so they can dramatically reduce—if not eliminate—frustrations, threats, and scenarios where both sides lose versus both sides win.

Organization Failure

We all know the main goal of business shareholders is to create value. While benefit realization and resource optimization are often at the core of business concerns, risk optimization is usually left aside or visited when absolutely necessary. This kind of situation leaves the organization in a state of unpreparedness—with poor visibility and the inability to react and respond to information security threats.

Not surprisingly, such organizations are shocked by vulnerability reports. They can't anticipate those situations and treat them as incidents. Without a proactive approach, reporters have a hard time finding the email of the responsible function for vulnerability management and might end up emailing the sales function, the CEO, an abandoned mailbox, reaching out via social media or, even worse, just give up trying. Also, some assumptions and expectations might not be shared with them. The report might end up with too little information or contain undesirable and unnecessary information.

Another common inconvenience is feeling harassed by frequent emails from the reporter on a status update. This is most likely because communication delays and frequency are not shared. The worst of all problems is trying to agree on a degree of impact (severity) for each vulnerability. Why does the reporter insist so much on the urgency of remediating an issue the business does not consider impactful? Well, does the business clearly define what is considered impactful? All these issues will spark some disagreement and frustrations between involved parties, which justifies the need of a formal process that is defined, documented, and communicated to appropriate parties. This process is called "responsible disclosure."

What Can Organizations Do?

A formal responsible disclosure process must be designed, developed, implemented, and published on the business's main website. It must be easy to find and always available. This process leverages proactivity in handling vulnerability reports by defining authorized reporting channels, assumptions, and expectations, as well as communication delays and frequency and legal terms and conditions. The more transparent the process, the better it is for everyone. This proactive approach will eliminate frictions and surprises that lead to the blame game and state of crisis management.

The responsible disclosure process must have clear degrees of severity to which the reporter can agree. The information security industry relies heavily on the Common Vulnerability Scoring System (CVSS), an industry standard to evaluate a vulnerability's severity according to its characteristics. It should actively be used when communicating with external reporters. However, this system may not fit well internally. In most cases, it is better to derive the CVSS score with business-approved metrics to produce a risk or a priority score that your internal teams can agree on. As an example, a high-severity vulnerability according to CVSS could derive to a medium or low priority issue because the affected system sits in a lab that is not connected to the internet.

Reporting Failure

It is sad to say, but not all reports are equal. In fact, it even gets trickier when considering that one report that suits one company's needs might not fit a different company's needs. The reality is that the reporter has a direct impact on the success of having his report accepted or not. Common factors and actions that contribute to this failure include (but are not limited to) the following:

- A violation of responsible disclosure process
- Poor writing and communication skills
- An unethical approach, such as asking for a "reward" (ransom) for the full report

The responsible disclosure process should be publicly available from the main website of the target organization. It should include all that is needed to report to the appropriate personnel and establish assumptions and expectations, as well as terms and conditions. Not complying with those rules is definitely a hostile way to report a vulnerability. No matter how severe the vulnerability, such behavior might just close any chance to obtain a follow-up or a reward.

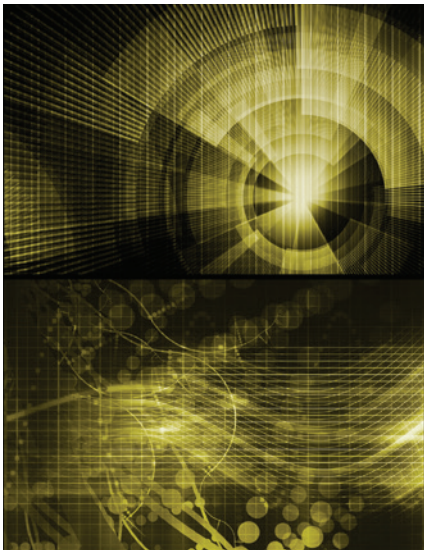
Writing and communication skills are also very important and are often neglected by technical people. It is not unusual to obtain reports written with poor and unintelligible sentences. Sometimes, the vulnerability is well-described, but its impacts on the business are misinterpreted due to poor explanation. Reporters need to focus on how it can hurt the business—for example, saying the risk involves "breaching the confidentiality of the customer PII data from your production database."

Not all bounty programs provide financial compensation. In the instances where responsible disclosure does *not* mention any reward, asking for money in exchange for a full report is not acceptable and could be considered extortion. Security researchers looking for compensation should focus on bounty programs that explicitly mention financial rewards. Any type of harassment, aggressivity, or threatening will not be tolerated.

What Can Vulnerability Reporters Do?

Reporters should first look for the presence of a responsible disclosure process and follow it carefully. For organizations that do not publish information about such processes, they should first be contacted by their support department. This is typically where all external requests come from. Reporters must not send the report until they reach a contact that is responsible for handling the issue. Reporters must also keep in mind that reporting vulnerabilities, especially to a business that is not ready, is a joint operation—in which the reporter must communicate and behave with extra diligence in regard to business representatives. If the business feels threatened, the cooperation will fail. This is where good communication and writing skills come in handy.

The reporter must also consider the business threat model and make sure the vulnerability is impactful. Each vulnerability should be reported with at least a description, a severity score (using CVSS), an analysis explaining factors affecting the probability and impacts of the vulnerability exploitation, and remediation options. References, screenshots, and any other information that may help understand risk and technical aspects of the vulnerability should also be included. For reporters that are internal to the organization, it is not recommended to rely solely on CVSS to evaluate a vulnerability severity. Internal teams have access to more information and should speed up the process by deriving the CVSS severity to a risk or priority score. This extra step will provide a better alignment with business needs in terms of risk optimization goals and risk appetite.



A formal responsible disclosure process *must be designed, developed, implemented, and published* ***on the business's main website.***

The terms of any agreement for a reporter to be paid or rewarded for extra work requested by the target business should be communicated early and should not restrict the business access to vulnerability information. The same applies if the reporter wants to publish the report publicly. This discussion should be held early to avoid cooperation failure. Currently, in the industry, the rule is to allow the business to fix the vulnerability before going public. For non-cooperative businesses, a deadline of 90 days is usually the norm before going public. However, reporters should be aware of risks of legal actions from a business that does not consent to go public with a vulnerability report.

Is It That Easy?

While designing, implementing, and maintaining a responsible disclosure process will leverage collaboration between reporters and business representatives, there still need to be other processes in place. Such additional processes are needed to remediate and manage risks related to vulnerability reports to avoid unnecessary delays, emergency change errors, and sensitive information leakage.

A good place to start is with the NIST draft on Secure Software Development Framework (SSDF). The practice group called Respond to Vulnerabilities (RV) further speeds up and secures the remediation of vulnerability reports. Organizations can also rely on third-party assistance, such as bug bounty platforms, assessors, and auditors, to help them put in place and refine their program.

On the reporter side, staying professional, focused, technically accurate, and communicative is an art that should be continuously improved upon to help keep businesses, their partners, and customers safe. ■

Martin Lemay is the chief security officer at Devolutions, where he leads the information security program from cyber-risk management to application and infrastructure security. Over the past 10 years, he has acquired a solid technical background as a security professional with a specialization in penetration testing and has operated in most industry sectors, including—but not limited to—banking, financial, energy, healthcare, transportation, and telecommunications.

Securing Your Cloud Data Lake With an In-Depth Defense Approach

By nature, cloud data lakes are the first place where data lands. Because of this, they have become the most attractive target for cybercrime.

By Jacques Nadeau

Cybersecurity attacks have been increasing at an exponential rate. In 2018 alone, more than 2,000 data breaches were reported. The impact of these attacks has been calculated at more than $6 trillion. Given these statistics, the security of data lakes is of paramount importance. We all understand the value of a cloud data lake.

Modern cloud technologies make cloud data lakes easy to set up and maintain, and in addition to being virtually limitless, they provide separation of compute and storage, allowing users to run any engine on top of their data. By nature, cloud data lakes are the first place where data lands. Because of this, they have become the most attractive target for cybercrime. For these reasons, organizations need to adopt especially stringent security controls.

Data Lake Security—Understanding the Requirements

Industries have developed standards and regulations to better protect data. Examples of this include CCPA, to enhance privacy rights and consumer protection for residents of the state of California; FISMA, to ensure the security of data in the federal government; GDPR, for the protection of EU citizen data privacy; and HIPAA standards, for managing healthcare information. While all of them are different, and each one treats a different symptom, these ever-evolving regulations have several requirements in common: access control, auditing, and encryption.

Cloud vendors such as Azure and AWS offer several features that help industries implement security best practices on their cloud data lakes to meet these requirements. These built-in controls go all the way from identity security to security management.

Locking Down the Data Layers

Cloud data lakes are the place where all data lands. However, the more granular view lets us see that the basic structure of a cloud data lake is comprised of network interfaces and data in file formats such as Parquet and JSON, as well as technologies that group all these files into tables such as Hive Metastore and AWS Glue. There is also a semantic layer as part of the architecture—technologies such as Dremio, Spark, and Hive enable data analysis directly from the data lake and, most importantly, protocols and client interfaces that allow users to consume the data.

Understanding the difference between the layers of a cloud data lake is important, because each layer will require a different kind of security depending on the accessibility needed by the user dealing with it. The best way to implement security at this level is to decide who is going to be allowed to have access to each of the layers.

Following the least-permissions-required approach to each layer is a fundamental principle to ensure that each user has just the right amount of permissions to complete his tasks without compromising the integrity of the data. Examples of permissions around the data layers include storage buckets being accessible only to compute engines and data engineers, security permissions that are configured using resource-based identity and access management policies, data tables being accessible to data engineers and

data scientists with permissions configured through the implementation of users and roles, and semantic layers accessible to business analysts with the help of access policies defined on Active Directory or other authentication systems.

The complexity of a data pipeline directly affects the security of a cloud data lake; there have been situations where policies are not implemented correctly and millions of rows of sensitive data (i.e., voting records, medical records, and credit card information) have been left in unsecured public storage buckets.

The following are some guidelines that will help avoid these kinds of mishaps:

Design your semantic layer around secure zones: Identify who needs access to what assets. For example, administrators and data engineers will need access to physical data sources, while access to virtual datasets and curated data will be sufficient for analysts and data scientists.

Apply column- and row-level permissions in the semantic layer: By doing this, you eliminate complexity by not having to make changes at the application level or create multiple protected versions of the same dataset. Data consumers simply receive the data they need with implemented security that will allow them to see just what they need.

Apply permissions based on the capabilities of the service: Cloud vendors such as AWS and Azure provide a variety of mechanisms, including IAM policies, role-based access, encryption at rest and transit, and key management, just to name a few.

Secure and govern user access: This will always ensure that the enterprise's data is not open to the public. It also provides the opportunity to identify who can access the data, as well as what actions they can take.

Secure and govern users' rights: This allows companies to control what privileges an authenticated entity can have within the system. It is imperative to have a security plan laid out before the data lake is created, as this will provide an opportunity to define roles and privileges.

Leverage metadata governance: Securing your data is only part of the story—securing metadata is just as important. Armed with metadata, an attacker can target users as well as applications within your organization and gain access to data. Metadata controlling systems such as AWS's Glue can help alleviate that issue through IAM-based policies. Similarly, Azure Data Catalog allows you to specify who can access the data catalog and what operations they can perform.

Avoid Data Copies

The majority of data breaches that we have witnessed are the result of human errors. However, most of the errors that humans make are due to the complexity of the infrastructures that they are dealing with. Not having a basic understanding of the impact that changing a security setting may have on an entire cloud environment is the reason for most of the issues and breaches seen today.

The worst enemy of data security and governance is the lack of a self-service environment. By allowing users to utilize a self-service environment on the data lake, enterprises eliminate the need for users to create multiple copies of the same data every time they need to make a small change to it. Data lake engines allow users to query data directly from the data lake, thus eliminating the need for data copies, which are hard to secure.

Keep It Simple

There are two types of security risks: exogenous—driven by external attackers—and endogenous—driven by employees exposing data. Unfortunately, many of the systems put in place to protect against the former are increasing the risk of endogenous attacks. Security systems sometimes can be so complex that users try to work around them. In order to manage this type of risk, enterprises need to focus on simplicity and give users enough tools so they don't attempt to work outside the system that is in place for them.

Enterprises should provide a governed mechanism for data sharing that avoids disconnected copies and also avoids restricting access to data unnecessarily; this will only stifle self-service and drive users to less-governed alternatives.

Enterprises should also enable coarse-grained ownership when possible. The scalability and elasticity of the cloud make it easier to create separate resources for different teams. Full resource isolation is emerging as a common model for data lakes and data warehouses, allowing data teams to use their resources without sharing them with other organizational units. Additionally, access control is easier to set up and maintain.

Self-Service Is the Foundation of Governance

Implementing security measures to keep attackers in check and avoid data from leaking out can be a daunting task if not implemented correctly. Security policies can do more harm than good if they are perceived by users as roadblocks. The good news is that self-service is emerging as the fundamental element of data security and governance; it allows users to have access to properly secured and curated data, thus avoiding the need to work around security hurdles to complete their jobs. Additionally, it allows for admins to keep track of what actions are being taken against what assets through features such as data lineage and activity monitoring.

When implementing a security model on your cloud data lake, always start simply, and only add complexity as needed while keeping the user experience in mind. This way, the number of security mishaps caused by endogenous reasons can be reduced to a minimum. ■

Jacques Nadeau is the co-founder and CTO of Dremio. Previously, he ran MapR's distributed systems team; was CTO and co-founder of YapMap, an enterprise search startup; and held engineering leadership roles at Quigo, Offermatica, and aQuantive. Nadeau is co-creator and PMC chair of Apache Arrow, a PMC member of Apache Calcite, a mentor for Apache Heron, and the founding PMC chair of the open source Apache Drill project.

Three Steps to Manage Shadow IT in the Enterprise

The primary concern with bringing unauthorized apps onto the network is that enterprise IT teams may be unaware of the dangers that these tools could introduce.

By Sean Armstrong

IT TEAMS FACE AN UPHILL BATTLE IN managing the deluge of apps that now populate enterprise networks. A big reason for this is that employees have access to a wide array of SaaS and cloud-delivered platforms—both business-critical tools and otherwise—that they can deploy in just a few clicks. This puts the onus on network teams to lay out specific policies about what kind of tools they'll allow on the network, as apps that could pose a threat to network performance are now increasingly common and easier than ever to deploy.

Holding employees to task and enforcing these policies, however, are easier said than done.

This is especially apparent when considering Gartner's estimates that anywhere between 20% and 50% of enterprise app spending take place without IT's knowledge or consent—aka shadow IT.

The rise of shadow IT mirrors the rise of SaaS, as tools delivered "as a service" are, by design, easier to deploy (often requiring simply a web browser and registration) and more cost-effective than legacy solutions, thereby giving individual employees more agency to seek out and start using new apps at their discretion.

The primary concern with bringing unauthorized apps onto the network is that it can make enterprise IT teams unaware of potential dangers that these tools might introduce—in particular, data leakage and falling out of compliance with privacy regulations such as SOC 2, GDPR, and CCPA. But it's not always (or even usually) the case that shadow IT is conducted with negative intent. In reality, it often comes down to a matter of

teams preferring certain platforms to complete day-to-day tasks over those approved by IT (e.g., preferring Zoom over Microsoft Teams for unified communications, or UC).

When this practice is on the rise at an organization, it's usually in response to dissatisfaction with the policies and tools approved by IT. Moreover, users may be to blame for their own dissatisfaction, such as when non-approved apps sap up network capacity that was originally allocated for approved apps, thereby impacting the performance of both "shadow" tools and approved ones.

So where can IT teams get started in keeping a handle on shadow IT? Here are some tips:

Get to Know Your Network

As the name implies, shadow IT occurs when network teams don't have an understanding of all the tools leveraging their network. It's more than just an issue of malware hiding in the shadows: Teams also need to have insight into every application living on the network to evaluate how non-critical tools (or alternative apps) are impacting the performance of approved business-critical solutions. Having an understanding of employee habits and preferences versus what's prescribed by company policy can help inform IT teams on how to better allocate network capacity in the future.

Give Your Network a Sanity Check

Once IT has gotten a sense of all the applications populating the network, teams can then explore what existing policies (and approved apps) are helping the business, and identify where things could improve. Perhaps a team that abandoned Skype for Business in favor of Zoom was onto something, for instance, and the whole company would benefit from a new default UC tool.

IT can also explore whether the reason approved tools are being abandoned is a deeper performance issue that IT might not have been aware of. The network team could then take steps to remedy this chronic issue and get all users back on the same page.

***The rise of shadow IT mirrors the rise of SaaS**, as tools delivered 'as-a-service' are, by design, easier to deploy **(often requiring simply a web browser and registration)** and more cost-effective than legacy solutions.*

Put Your Learnings to Work

All of these steps are driving toward the goal of giving enterprise IT the visibility it needs to successfully monitor and manage the network and all the apps within it. With this visibility, IT can more easily spot signs of hazardous shadow IT and better enforce (and inform) its network policies. This doesn't necessarily mean dedicating members of IT staff to policing end users, however. Rather, teams need to employ lightweight, low-overhead solutions that can deliver active insights on app performance in near real time without overcomplicating the task of network monitoring. Combined with passive traffic analysis to identify what apps are running at each enterprise location, IT can gain a full picture of the app landscape.

Once armed with active and passive visibility across the enterprise network, IT teams can start building bridges between themselves and the users who turned to shadow IT in the first place. That way, end users and network teams can approach the company's tech stack collaboratively, recommending new tools or taking a proactive approach to remedying performance issues. ■

Sean Armstrong has been in product management at AppNeta for more than 11 years. As vice president of products, he revels in taking deep network information and making it interesting and digestible to businesses. Prior to his time at AppNeta, Armstrong worked for RSA, the security division of EMC, as a senior product manager. He received his bachelor's degree in management science and computer information systems from Virginia Polytechnic Institute, and currently resides in the greater Boston area.

Safeguarding Your Data in an Online World

GDPR, the recently enacted CCPA, and similar requirements around the world are forcing organizations to take steps to protect data and manage it properly. But when it comes to individuals, whether consumers or employees, more needs to be done to educate them on the seriousness of a compromise or not having backups in place.

By Rick Vanover

In a world ruled by data, it is frightening to think how few people take protecting their information seriously. While it has become easy to share everything from an email address and mobile number to more sensitive personal information, what are the consequences of not safeguarding data?

The likes of GDPR and the recent CCPA, as well as similar requirements around the world, are forcing organizations to take the necessary steps to protect data and manage it properly. But when it comes to individuals, whether consumers or employees, more needs to be done to educate them on the seriousness of a compromise or not having backups in place. This is a much more relevant concern for any organization that is in the midst of a digital transformation. If you are correctly transforming digitally, the need for resilient backup and recovery techniques requires strong safeguarding of critical data. This becomes a more essential matter as organizations are pressed to retain data longer and more organizations are subject to data being under the purview of regulatory or compliance guidelines.

Although this heightened data regulation helps keep organizations in line with legal requirements, consumers do still carry a responsibility for protecting their personal data. This data can take the form of photographs, documents, and other important records that are increasingly stored on cloud-based services in addition to identification numbers, bank details, and so on. One easy place to start is the shared responsibility model for most cloud services; it puts the responsibility of data solely in the hands of the cloud subscribers.

Question Everything

It all starts by questioning everything in the digital world. People need to understand what they are agreeing to and the trade-offs involved, especially when it comes to mobile apps and freely available services. The majority simply accept the terms and conditions without reading them through. While not accepting them may mean being unable to use a specific app or service, that may be better than the alternative of having personal data spread across the web. This is especially true in the case of anything free. Ask yourself, "Why is it free?" and then you may think twice before unknowingly handing an organization all of the peripheral data that your smart device can provide them.

And being cautious does not necessarily mean focusing only on the cloud. A person must consider the implications of older technology. Just think of all those flash drives people have lost over the years. Some might contain innocuous bits of data, while others could provide a malicious user with a treasure trove of information that can be used to compromise a person.

The popularity of FaceApp put the spotlight on the rights of the individuals and what companies can do with their data, in this instance, their photographs. Some argued that this was a form of spyware that could store people's personal photos on their servers for perpetuity. Cynics counter-argued that if a site such as Facebook already has that information, what difference does it make if others have it as well?

These examples are not limited to public social media sites. Consider businesses that partner with other organizations in various ways—is that data tended to in the same way?

Regardless, people should bear in mind that even if they are using trusted platforms such as Apple, Android, and Microsoft, it does not mean that every app they are using is secure. The app stores simply cannot check all security aspects of any given app. In the

case of FaceApp, if a person is unsure about the merits of sharing photographs, then it is advisable to simply not install the app.

These examples are mobile-first, but organizations need to consider that these types of scenarios can also apply to cloud and data center workloads and data.

Public Cloud = Natural Choice

Every organization should be in some stage of a journey to one or more cloud technologies. The natural subsequent consideration with the platform selection is to examine the changes associated with all data management, data protection, and disaster recovery (DR) contained in the public cloud. Specifically, organizations should consider the complete platform selection as a comprehensive pairing of cloud-native services and protection techniques to meet the same objectives that are on-premise: Reduce down time, avoid data loss, and recover quickly from any type of operational outage.

Compared to on-premise IT deployment, a recommendation is to have the backup and DR practices implemented and standardized before services are deployed in the public cloud.

Native Cloud Platform Capabilities and ISVs for Backup

It is a natural choice to seek to address backup, recovery, and disaster recovery with native cloud capabilities. ISVs, however, can offer services for multiple clouds as well as connecting on-premise experiences.

Azure Backup for Virtual Machines, for example, has the concept of a Recovery Services Vault. This will protect virtual machines and put them in the same location that they are running. Caution should be noted here about keeping backup data in the same location as the production source. ISVs in the space have taken an approach of supporting multiple regions for cross-region backup and DR. Each individual ISV will be at a different stage in its journey in this space, and organizations should press for public cloud road maps per cloud for backup and DR offerings.

The native backup offering for AWS, AWS Backup, is available to protect many services in AWS. AWS Backup leverages a backup vault, which is tied to the AWS account, and lets different resources be assigned to a backup plan.

In both native situations, one fundamental risk is in place: The same cloud account is used. ISVs have taken note of this and support multi-account backup within the same cloud. A critical recommendation is to have the backup of public cloud workloads leveraging a different account than where cloud resources are consumed. Naturally, in the cloud, there is often buildup and removal of services. The risk of deleting one or more components or data (such an Azure VM, EC2 instance, or S3 bucket) associated with backup data can be high.

Hybrid Cloud for Enterprises

For the foreseeable future, organizations will balance some form of a hybrid approach to the public cloud. This means some key services will reside on-premise, some will reside in the public cloud, and some key services may be specifically spread between on-premise and the public cloud. Hybrid organizations consistently prefer to have similar capabilities as their on-premise backup and DR strategies. It is wise to demand from ISVs a public cloud road map to see if their offering is in alignment with your public cloud strategy.

In this hybrid cloud approach, organizations will have a preference to maintain some of the capabilities that are familiar with on-premise backup and DR processes. This gives ISVs that are in use on-premise an opportunity to deliver the aforementioned road map or focus on standardization for consistent capabilities on-premise and in the public cloud for backup and DR. Switching or consolidating vendors for this reason is worth the effort now for the long-term cloud strategy.

New Options for Organizations in the Public Cloud to Safeguard Data

Recent advances from both AWS and Azure have introduced compelling new benefits for IT organizations, including the following:

- **Minimal retraining.** By leveraging the same management software and technologies as on-premise, organizations can quickly extend to the cloud in these platforms.
- **Adjacent services.** If other AWS and Azure solutions are to be used (such as AWS S3), these are very close to the source of the VMware options in the public cloud from a transfer and latency perspective. Backing up data to S3 or BLOB (binary large object) storage from these services can be a natural tiering of data for abstraction as well. (Continue with the advice to use separate AWS or Azure accounts.)
- **Replication options.** Many ISVs support replication options from on-premise infrastructures to VMware Cloud on AWS. This enables the flexibility of moving DR from on-premise to VMware Cloud on AWS, DR from one VMware Cloud on AWS availability zone to another, or DR from VMware Cloud on AWS to on-premise.

Education, Education, Education

Simply put, users need to be better educated on how best to safeguard their data. It is important for employees to understand that there are serious repercussions for a company that does not comply with legislation.

The importance of data availability throughout all this is essential. People expect companies to have data always available and accessible. This can take the form of products and services but also secure access to their own data (photos, documents, etc.). To this end, consumers must make three copies of their data, store two of those on different storage media, with one off-site (such as the cloud). This becomes essential to protect data in the always-on digital world of today. ■

Rick Vanover (Cisco Champion, VMware vExpert) is senior director of product strategy for Veeam Software. Vanover's experience includes system administration and IT management, with virtualization, cloud, and storage technologies being the central theme of his career recently. Follow him on Twitter @RickVanover or @Veeam.

CCPA Forces Modern Approaches to Customer Information Governance

Wrangling vast volumes of customer information to understand where to begin with compliance is arguably the single biggest challenge organizations face.

By Tara Combs

The CCPA deadline has come and gone. And, while the California attorney general won't enforce the act until July 1, just 5 months before its effective date, a recent survey revealed that only slightly more than one in 10 business owners and executives were aware of whether the law even applied to their business. Alarmingly, almost half had never heard of the regulation.

CCPA isn't the first of its kind. The act closely follows GDPR, which went into effect in 2018 to protect data privacy and security for consumers in the European Economic Area. We can expect to see a similar enforcement path—with the real fallout for non-compliant organizations coming once a data breach has occurred. Notably, Marriott and British Airways were fined £99 million and £183 million, respectively, for their failure to comply with GDPR, which was discovered as the result of data breaches. While the CCPA fines are not nearly this high, at only $7,500, the reputational damage can still be substantial.

Furthermore, we can expect that CCPA is only the beginning in the U.S. There is similar proposed legislation in Massachusetts, New Mexico, New York, and Washington state—much of which closely aligns with CCPA. And, consumers are increasingly seeking more accountability from businesses as they become aware of the role organizations play in their data privacy and security. With this in mind, organizations not immediately impacted by CCPA should still take note and act fast to clean up their information act.

The Path to Information Chaos

Wrangling vast volumes of customer information to understand where to begin with compliance is arguably the single biggest challenge organizations face. In fact, a survey by AIIM revealed that 75% of organizations see information chaos as a major problem. Factors contributing to confusion include the following:

- **The simplification of the customer journey**—Consumer information doesn't simply reside in "name" and "address" fields within structured

databases any longer. It lives in photos, scanned documents, PDFs of resumes, emails, and myriad other forms.

Businesses have tried to make customer experiences easier whether the customer is an end user or internal user, such as the relationship between an employee and human resources. However, as experiences are streamlined, information management is made more complex.

- **Digital transformation**—As enterprises have become more digital, they've adopted disparate operational tools that house important information. AIIM reports that 52% of enterprises have at least three enterprise content management systems, and 22% have more than five.

Unfortunately, many of those systems don't "talk" to each other, so there isn't a simple way to query an organization's systems to gather all information about a single customer, for instance. Instead, information may live in dozens of systems that must be individually parsed through.

- **Operational inconsistencies**—Many organizations continue to manually manage records. With up to 2,000 systems in some enterprises, it's not surprising that business users, even those with the best of intentions, don't consistently file information when and where they should. Business users also often don't know which information must be retained. And, that's without considering employees who might view information management as a low-priority task.

To combat these inconsistencies, many organizations have hired records managers to oversee documents and ensure retention schedules are followed for various records. However, with each new standard or regulation, these managers are fighting an increasingly uphill battle to stay on top of demands.

To undergo the required transformation *to meet new information governance challenges, businesses* ***must appoint*** *a data security and compliance officer who can lead the charge in* ***mapping out all of the organization's systems.***

Looking at the current state of the information, security, and regulatory landscape, it's clear that we've reached a breaking point. Traditional methods of records management and information security are not only siloed, but also leave room for human error. The way it's always been done simply won't suffice any longer.

Next-Generation Information Management

To overcome the information governance challenges and ensure compliance with new and upcoming regulatory demands, businesses must employ modern approaches that not only take advantage of the latest technology, but also consider information at an enterprise level.

Building a Strong Foundation

As with any strategic initiative, the foundation of a next-generation information management program first and foremost requires planning and a substantial investment of time and financial resources. In an economic impact assessment released in August 2019, the California Department of Justice forecast that compliance with CCPA would cost $467 million–$16.5 billion between 2020 and 2030. However, the regulation is expected to protect the $12 billion worth of personal information used in advertising annually in California alone.

To undergo the required transformation, businesses must appoint a data security and compliance officer. This individual will lead the charge in mapping out all of the organization's systems and identifying what types of information reside in them. While the initial activity of creating a data map is a huge undertaking—particularly for enterprises that have 2,000-plus systems—it is essential. The map will provide a complete picture of what information the business has and where it originates. This will allow an assessment of personally identifiable information (PII) risks and, if the organization sells consumer information, help to identify where the business must provide consumers a "right to opt out."

Building a data map is frequently a rushed task or skipped altogether, particularly as CCPA does not mandate the exercise. It is, however, worth noting that other regulations such as GDPR require organizations to implement this best practice. When businesses forgo the map, they often decide to only tackle a subset of their systems—perhaps the most obvious enterprisewide repositories. As a result, PII stored in peripheral tools is often not brought under an organization's management programs, leaving the business vulnerable.

Technology Streamlines Compliance and Security

With a firm grasp of their information assets, organizations can begin leveraging technology in a smart manner. A few of the technology-based practices businesses will begin to use as they grapple with information security and privacy include the following:

- **Consolidating Information in Place**—Previously, businesses that wanted a single view of their customers, for instance, needed to pull and combine data from multiple systems for each inquiry. The solution to the manual headaches these efforts created was to migrate data from multiple systems into a single repository. But this can be a costly and time-consuming activity that disrupts business

users' workflows as they are forced to learn new tools.

Federated governance hubs use prebuilt connectors and APIs to bring together information from common business platforms and create a single interface for applying information management rules. Hubs can even be extended beyond enterprise systems to local drives so that documents stored on desktops are managed and properly secured. As a result, existing business user applications are not disrupted. Not only does the federated hub streamline information management, it also provides organizations with a single interface for responding to ediscovery requests. One study found document review and analysis made up nearly three-quarters of ediscovery costs, resulting in an average cost of $18,000 per gigabyte. Automating this process can significantly reduce those litigation and regulatory expenses.

- **Tapping AI**—IDC has predicted that there will be 175ZB of data by 2025. AI and machine learning are helping businesses quickly cut through this staggering amount of information to provide greater control of document classification, retention, and security.

For instance, machine learning algorithms can cull through 40 years of scanned insurance documents to quickly identify and obfuscate social security numbers and tag documents with the appropriate security level. The same exercise can be applied to other PII—addresses, phone numbers, etc.

- **Automating Compliance and Security**—Automation is one of the most effective strategies for reducing the burden and minimizing the risks of information management. With an effective automation program in place, compliance happens seamlessly in the background with very little or no intervention. Instead of requiring business users to understand records management, business rules and metadata ensure records are created, managed, and archived or destroyed on schedule.

We can expect that CCPA** is only the beginning in the U.S. There is similar proposed legislation in Massachusetts, New Mexico, New York, and Washington state—**much of which closely aligns with CCPA.

In the previous example of documents with Social Security numbers, the AI-generated security mark can be used to automatically limit who has access to the document internally. Automation settings can also provide varying levels of information access—allowing some individuals to view the full Social Security number while limiting others to the last four digits.

In addition to the time-saving benefits, organizations that automate information management throughout the entire lifecycle reduce their vulnerability and liability. AIIM has estimated that up to 70% of the data in unmanaged servers is redundant, obsolete, and trivial (ROT)—in other words, information clutter. Holding onto it impairs businesses' abilities to demonstrate compliance with regulations and slows down fulfillment of discovery requests. ROT is also typically unmanaged and unknown, contributing to a greater likelihood of theft or breach.

With CCPA at businesses' doorsteps and other regulations quickly following behind, the time for companies to take action and clean up their information act is now. While it is a challenging process, the risks are too great for businesses to stand idly by. Thankfully, modern approaches supported by technological advances from the last 10 years can help to streamline information governance and, ultimately, compliance. ■

Tara Combs is the senior information governance specialist for Alfresco. In this role, she helps organizations understand how to meet their information governance and records management requirements and mandates, as well as use records management modernization as a catalyst for their larger digital transformation needs. Previously, Combs worked extensively in the enterprise content management/records management market space as a solutions consultant for government and corporate organizations.

With GDPR in Full Swing, CCPA Takes Off. Here's How Organizations Can Prepare—And Cope With—SRRs

Complying with subject rights requests, or SRRs, requires that organizations establish a privacy management program well in advance of receiving them so they can "hit the ground running" and avoid becoming deluged by the flood of incoming demands—especially in the early days of CCPA.

By Sovan Bin

CCPA took effect on Jan. 1, 2020, following the May 25, 2018, launch of the landmark global compliance regulation GDPR. When California begins enforcing CCPA on July 1, 2020, any for-profit entity doing business in California that collects, shares, or sells California consumers' personal data will be governed by CCPA if it:

- Has annual gross revenues in excess of $25 million; or
- Possesses the personal information of 50,000 or more consumers, households, or devices; or
- Earns more than half its annual revenue from selling consumers' personal information.

Data privacy regulations have focused on holding organizations accountable for breaches of their systems and the personally identifiable information (PII) they hold. In fact, a foundational premise of CCPA is that consumers "own" their privacy information. However, while CCPA acknowledges the primacy of consumers' rights regarding the information that organizations hold, much less attention has been paid to how consumers can take action on their own.

A tenet of CCPA is that consumers should feel free to exercise their rights to safeguard their personal data—and hence the incorporation of what CCPA refers to as subject rights requests, or SRRs. (A data subject, or simply "subject," is defined as an identifiable individual about whom personal data is held.) What's more, consumers should demand that organizations remain transparent about the usage of their personal data so they understand what information the organization holds, how it is being used, and who it is being shared with.

That said, complying with SRRs requires that organizations establish a privacy management program well in advance of receiving requests. The goal is to "hit the ground running" and avoid becoming deluged by the flood of incoming requests—especially in the early days of CCPA. And then comes the hard work: drawing up a data inventory of all the organization's IT environments, establishing what information is classified as personal data under the CCPA, and mapping the flow of data through your applications that use it.

SRRs: A 'Foundational Requirement' of CCPA

That is why SRRs have become central to consumers' data privacy rights under CCPA. They cover a defined set of rights where individuals have the power to make requests regarding their data, and where organizations handling this data must address these requests in a defined time frame—which, for CCPA, is 45 days. Gartner cautions that SRRs will play an overarching role in enforcement of CCPA.

What's more, CCPA differs from GDPR in its definition of an "entity" (the data subject). "The GDPR is specifically focused on all data related to the EU consumer/citizen whereas the CCPA considers both the consumer and household as identifiable entities."

Given the primacy of consumer data, organizations that are subject to CCPA need to turn their focus to protecting the consumer data they hold, which should be their highest ideal. Still, Gartner cautions that subject rights requests left unmanaged have the potential of becoming "death by a thousand cuts."

SRRs come in three categories:

- *Right to know*: These rights focus on providing individuals with access to their data. This class of requests includes the most commonly sought SRRs, typically known as subject access requests (SARs) or data SARs (DSARs), where individuals seek to view what data the organization holds on them.
- *Right to correct:* These rights focus on allowing individuals to manipulate their data or their preferences. At the extreme, corrective rights allows individuals to delete their records.
- *Right to object:* These rights focus on allowing individuals to control how their data is processed. Under CCPA, individuals have the capacity to object to the sale of their data to a third party.

Flow mapping can be a massively complex and tedious undertaking, of course. But it can become painful in highly distributed infrastructures, according to Gartner. "The question is, why is ensuring GDPR [or CCPA] compliance so difficult? The answer lies in the complexity of a given organization's technology infrastructure, which is laden with dozens if not hundreds of systems. Any one of those systems, which seldom talk to each other, can hold various customer records."

Organizations that bring a high level of transparency to SSRs inevitably increase customer intimacy while strengthening their brand image. And, in doing so, they meet the highest of ideals: protecting consumer rights.

Keep in mind that businesses must meet every SRR within 45 days. Here is a six-step process that sets the stage for success:

1. Establish a privacy risk register, where the organization can log and validate repositories of personal data, calculate the risk of each entry, and use it to prioritize remediation tasks.
2. Divide the discovery exercise into two parts: one dealing with information currently held, and the other focused on new information that the organization is generating or appropriating.
3. Ensure that new information introduced into the system has the metadata that would allow it to be tracked and managed properly.
4 Capture, catalog, and prioritize large repositories of personal data—such as HR data, CRM records, and customer care logs—as they represent risk to a large number of individuals.
5. Enable your employees and partners to introduce new personal data repositories they discover into the existing privacy risk register. Doing so creates an iterative, crowdsourced process that maximizes the amount of personal data you can manage for any individual.
6. Define consumer rights workflows and steps in detail. Automate consumer rights management with a data privacy compliance automation platform.

And remember that enforcing compliance can be a notoriously complex challenge. "A CCPA-covered business is required to respond to at least two requests from any individual consumer in a 12-month period, provide a toll-free number for consumer information requests, and prominently link to an opt-out page from the company's homepage or any other page where personal information is collected," according to the law firm Gunderson Dettmer.

> A premise of CCPA is that individuals have the right to request what data an organization is holding about them, why the organization is holding that data, and who else their information is disclosed to. Individuals exercise that right—which is essentially a consumer right—via a formal mechanism called an SRR. Individuals can quickly track down more information on SRRs by searching on that term in the CCPA.

Perhaps the most crucial aim of every organization subject to data privacy regulations is to prepare for the likelihood of an audit. But isn't complying with the "letter of the law or the regulation" sufficient preparation? Unfortunately, no. Enforcing compliance is not the same as documenting compliance. To cope with the documentation efforts, companies can automate the stewardship of personal data in software and eliminate weeks or months of tedious, error-prone manual processes, while producing proof of compliance for auditors.

Automation isn't always the best solution to complex problems. But, in the case of CCPA, it may be the only solution that allows organizations to cope with the immense scope of data privacy regulations, which, above all, exist to protect consumer rights. ■

Sovan Bin is CEO of Odaseva, a company he founded in 2012 to answer the need for better data protection and governance in cloud services. Bin also spent 6 years at Salesforce leading the architect team in Paris, where he was 1st CTA (certified technical architect) in EMEA.

From the Mainframe Era to the Internet of Things—And What Lies Ahead With Edge Computing

For risk management and cybersecurity professionals tasked with integrating and protecting the systems and data of the future, it is valuable to understand the evolution of computing and security in order to address tomorrow's risks.

By Martin J. Frappolli

Today, with a quick screen tap on a personal handheld processor, people seeking a ride can open a cloud-based app and connect with a drivers who can pinpoint an exact pick-up location via GPS, arrange a dropoff, and make a payment without the need to physically exchange cash or card. Who would have imagined this reality when just 40 years ago, before the internet was mainstream, the only computing interface was conducted on a green-screen "dumb terminal" limited to display and data entry? Nevertheless, this current moment is the reality for only a fleeting period of time before a new, transformative—and disruptive—technological era begins.

The migration from the mainframe era to the Internet of Things (IoT) has led to countless life- and industry-changing innovations. And, it is only a signal of what is to come with edge computing and beyond. However, just as with all advancements, it also brings risks and threats. Given how rapidly technology is being developed and adopted in critical mass, it is vitally important to global, corporate, and personal security to ensure the integrity of each development, device, and network, especially now when so much sensitive data is aggregated and exchanged on these platforms.

For risk management and cybersecurity professionals tasked with integrating and protecting the systems and data of the future, it is valuable to understand the evolution of computing and security in order to address tomorrow's risks.

The Path to IoT

The journey from mainframe to IoT was driven by pioneering engineers, technicians, and risk and security experts who broke boundaries to create solutions and maximize efficiencies. Mainframe computing began revolutionizing the way work was done and how information was stored. Little by little, computers were taking on tasks that had always required human labor. Though the typical access was through a simple data

entry and display terminal, the mainframe still allowed for smarter work. The "one brain" system was relatively secure with minimal exposure points—which meant that there was little concern about network safety and cybersecurity (a term yet to be coined). Data processing professionals could create and maintain a secure system for any company, regardless of size.

Enter the PC. The need to allow more applications for diverse user communities within a firm, combined with the emerging ability to create them, helped usher in the distributed computing era: an empowering time when individuals now had access to independent processing power with desktop PCs instead of mere dumb terminals connected to the mainframe. However, by enabling more and more people to manipulate data and information on a particular network, the distributed era landed at a significant and consequential new intersection: increased capabilities and complexity and the need for more security.

On the heels of the distributed era, the internet entered the mainstream and was integrated into homes and workplaces. This moment in time forever changed every aspect of how the world connected and worked. What started as a tool for government use quickly evolved to public and commercial use from dial-up to wireless connections and from desktops to smart phones. Every new point of connection brought with it new capabilities and new vulnerabilities. Cyber-risks, introduced during the distributed era, now escalated in seriousness with malware and denial-of-service attacks. New legislation was enacted to address threats, while security vendors designed products available to protect data.

And then came IoT. While the concept of smart devices had long been top-of-mind, they did not become widely accessible until around 2008. Today, just about *anything* with a plug can connect to a network and be used for precise purposes. IoT has brought us smart houses, turned mobile devices into personal medical monitoring tools, and streamlined countless aspects of every industry. These capabilities rely on the sharing of sensitive data, which increases the need to secure the many points of exposure.

Learning from breaches and hacks *allows cybersecurity professionals to better understand how to identify vulnerabilities and determine* ***how best to mitigate the risk.***

What's Next

IoT stands to further increase efficiencies as edge computing is integrated. Edge computing effectively moves computing power as close to the IoT device as possible. Not only will it create better performance and lower latency, it can also greatly increase capabilities. The number and impact of the revolutionary outputs we saw with 4G—services such as Uber, higher-definition television, and more—will likely be greatly exceeded with 5G and edge computing.

Ideas once thought futuristic, such as autonomous vehicles or vehicle-to-vehicle communication, will enter the mainstream with the intention of making roadways safer and reducing accidents. Edge computing will likely bring advancements in telemedicine that can break down geographic barriers to care and provide life-saving services to people, regardless of location.

Edge computing will also drive continued movement from AI and machine learning to deep machine learning—all while continuing to lower latency. This will enable even smarter use of technology to bring applications such as facial recognition to airports and to better secure infrastructure, including power grids, fuel lines, and roadways. But as we invest more and more in the development of these technologies—and trust their capabilities to keep us safe—it again becomes even more vital to protect them.

Securing the Future

Computing has made seismic shifts in every industry and across the globe. And now, on the precipice of even more significant change, cybersecurity and risk management professionals are tasked again with securing a complex landscape. Lessons from the past make clear that no industry is immune from the risk of data exploitation. In the first half of 2019, 32 million medical records were breached. In the same year, more than 106 million American and Canadian Capital One customers had their information exposed following a breach, and a WhatsApp hack targeted highly sensitive information from military and government sources, as well as members of the media. These breaches not only caused significant financial losses and damage to reputations, they also compromised personal and national security. Learning from breaches and hacks allows cybersecurity professionals to better understand how to identify vulnerabilities and determine how best to mitigate the risk.

The advancements brought forth as IoT and edge computing evolve will lead to better medical care; safer roadways, skies, and railways; the birth of new industries; and, ultimately, new eras of technology, again shattering boundaries that once stood in the way. Staying up-to-date with technology, the data it generates, and the vulnerabilities it exposes along the entire risk continuum is critically important to ensuring the security of the entity a risk manager or cybersecurity professional is charged to protect.

Now is the time to look beyond present computing capabilities risks and provide counsel on how to address them. The future of computing is immensely promising, but so are the stakes when it comes to cybersecurity. ■

Martin J. Frappolli, CPCU, FIDM, AIC-M, is senior director of knowledge resources, The Institutes Risk and Insurance Knowledge Group.

IRI
Total Data Management

MASK PII EVERYWHERE

Founded 1978, IRI — The CoSort Company — delivers uniquely fast, affordable and consolidated software for big data discovery, integration, migration, governance, and analytics.

To protect data at risk in structured, semi- and unstructured data sources, DBAs, CISOs, DevOps, and compliance officers use Gartner-advised IRI 'shield' tools or services to:

- Classify, find, and audit PII on-premise or in the cloud
- Statically or dynamically mask PII to stem data breaches
- Encrypt, hash, and tokenize PANs per PCI-DSS
- Score re-ID risk and anonymize for HIPAA/FERPA
- Comply with GDPR/CCPA Erasure, Portability & Rectification provisions
- Subset or random-synthesize referentially correct test data

Learn more at: **www.iri.com/products/iri-data-protector**

IRI, The CoSort Company
www.iri.com/products/iri-data-protector

AEROSPIKE

Aerospike is the global leader in next-generation, real-time NoSQL data solutions for any scale.

Aerospike enterprises overcome seemingly impossible data bottlenecks to compete and win with a fraction of the infrastructure complexity and cost of legacy NoSQL databases.

Aerospike's patented Hybrid Memory Architecture™ delivers an unbreakable competitive advantage by unlocking previously unimaginable value from vast amounts of data at the edge, to the core and in the cloud.

Aerospike empowers customers to
- instantly fight fraud;
- dramatically increase shopping cart size;
- deploy global digital payment networks; and
- deliver instant, one-to-one personalization for millions of customers.

Learn more at **www.aerospike.com**.

Aerospike
www.aerospike.com