Cybersecurity
Action Team

Modern Data Security

A path to autonomic data security

Dr. Anton Chuvakin, John Stone

Google Cloud

Table of contents

Executive summary 3

Your data has fallen out of love with your security model 4

Challenges of the classic data security model 5

How cloud is changing data security 6

- Data governance 6
- Speed and scale 7
- Data loss prevention 7
- Segmentation 8
- Data encryption 8
- Data access 9

What should your next steps be? 10

Pillars for building modern data security 10

- Automated/embedded classification and encryption 10
- Integrated access to data over any channel 12
- Policy intelligence leads to autonomy 13
- Reduced friction and complexity 14
- Measurability vs. business outcomes 15
- Visibility of the data processing supply chain 16
- Data lifecycle transparency 18
- Data security as enabler 18

Ready to move to new-world security? 19

Executive summary

Your business sits at a critical juncture as you face adapting old-world security models to the new world of data in the cloud. If you don't change and adapt, you'll not only have to deal with increased security risks, but you'll also limit the value to be derived from your data, stall innovation, and compromise governance.

Here we'll examine in detail how data security in the cloud differs from more traditional, on-premise data security. We'll discuss how the cloud is changing every aspect of data security – from data loss prevention and data access to segmentation, encryption, and governance. We'll present the pillars essential to building modern data security. Note that this paper primarily focuses on data confidentiality and while it touches on integrity issues, it does not cover data availability.

Finally, we'll leave you with the concepts and tools you'll need to start implementing an autonomic data security model today. This table compares and contrasts the key differences.

Old-world data security	**New-world data security**
Manual, user-driven classification, with confusing layers of encryption	Automated/embedded classification and encryption
Data is accessed separately in each channel and access controls are separate	Integrated access to data over any channel
Policies are manual, and granular policies overwhelm security teams	Policy intelligence leads to autonomy
High and growing complexity of many data security safeguards, each with its own rules	Reduced friction and complexity
Compliance focus and no direct link to business outcomes	Measurability vs. business outcomes
Opaque data supply chains, no central visibility	Visibility of the data processing supply chain
Many data lifecycles run at the same time, distributed over data types	Data lifecycle transparency
Data security as friction or compliance burden	Data security as enabler

Your data has fallen out of love with your security model

"90% of all data today was created in the last two years – that's 2.5 quintillion bytes of data per day." – Domo, "Data Never Sleeps 5.0"

This stat from Domo would be mind-boggling if it weren't for the fact that it's already five years old.

Five years ago, none of us would have predicted a global pandemic and the effect it would have on all facets of our life and work. Even data did not escape its impact as, according to Statista:

> The total amount of data created, captured, copied, and consumed globally is forecast to increase rapidly, reaching 64.2 zettabytes in 2020. Over the next five years up to 2025, global data creation is projected to grow to more than 180 zettabytes. In 2020, the amount of data created and replicated reached a new high. The growth was higher than previously expected caused by the increased demand due to the COVID-19 pandemic, as more people worked and learned from home and used home entertainment options more often."

One outcome is that your business needs around using data and deriving value from it have also changed. You're relying more on the power of technologies like cloud computing and AI, which gives you greater accessibility to keener insights from your data. Your organization is no longer just crunching the same datasets. Data moves, shifts, and replicates as you mingle datasets and gain new value in the process. All the while, your data resides in – and is being created in – new places.

At the same time, data breaches have been on the rise, with threats such as ransomware presenting real risks to the availability of data. Large disruptions to business operations are putting already-strained data security models under further pressure. In 2021 alone, over five thousand confirmed data breaches were committed. According to other estimates, the average cost of a data breach in 2021 was the highest in 17 years – an estimated $4.24M.

Focusing more broadly, on all security incidents, our GCAT Threat Horizons intel report #1 came to the same conclusion. "The shortest amount of time between deploying a vulnerable cloud instance exposed to the internet and its compromise was determined to be as little as 30 minutes."

We've also seen the impact that ransomware operations can have on businesses, with numerous published cases of security threats, such as the Colonial Pipeline attack. Modern ransomware incidents involve not just malicious hackers encrypting data, but exfiltrating it and stealing it, too.

We should also consider new elements, such as third-party libraries and components in your software stack, that could lead to unintended consequences and, in some cases, a data breach. As mentioned in GCAT Threat Horizons intel report #2:

> "During the month following the vulnerability's disclosure, there was extensive scanning across the Internet. Google Cloud and other providers had a unique vantage point over this and used this to good effect to help customers identify vulnerabilities as well as watch for the evolution of attempted exploitation to rapidly assure mitigations were effective for cloud infrastructure and customers. Google Cloud is continuing to see scanning (400K times a day) and expects similar, if not more scanning levels against all providers, and so we recommend continued vigilance in ensuring patching is effective."

Fortunately, not all is doom and gloom. IBM's Cost of a Data Breach report observes: "Organizations further along in their cloud modernization strategy contained the breach on average 77 days faster than those in the early stage of their modernization journey." Without a doubt, a lot of change and disruption over the past few years has challenged the traditional data security model.

Challenges of the classic data security model

Before the cloud, your data resided on-premises, often inside many corporate data center servers. You used closed-sourced products to store and manage data. Flash forward to the present and you have data in the cloud – including multi-tenant software with distributed data, and different hardware and software components interacting continuously.

To emphasize the difference between securing the place where the data lives versus securing the data itself, consider how securing a container where data is housed is very different from securing the data itself, wherever it lives.

Dealing with implementing and integrating myriad security tools from different vendors can impede efforts to create a cohesive security strategy. A recent article from IDG lists some specific challenges, including lack of interoperability among security tools, broken functionality, limited network visibility, false alarms, and lack of skills. As we'll detail later in this paper, modern autonomic data security in the cloud eliminates this fractured approach.

How cloud is changing data security

Without question, the cloud is changing data security in significant ways. Here are some cloud computing challenges we are facing in the modern world.

Data governance

The topic of governance and data security in the cloud takes on increased importance for regulated companies (like those in the banking and financial services industry).

Responding to new and changing regulations can slow things down when it comes to managing data, and taking a long time to gain the insights needed to make decisions to stay ahead of the competition is never good for business. Speed is of the essence here, and it's often essential that access decisions be made within minutes, not months. Manual exception management also becomes impossible at cloud scale, without changes to both technology and processes.

Equally important is the need to govern the data lifecycle. Data retention policies dictate how companies must save and maintain data for regulatory purposes. Tension can occur between regulatory compliance and internal company policies regarding how quickly a company proactively deletes data for legal exposure and liability purposes. Implementing the right plan for breaking up data in this way – what you don't need vs. what you do and for how long – is also a data security concern.

This has been a huge challenge for many organizations, even before cloud became an option. Who created what data, where it is, and who has access to it have always been challenges – and are some that many companies still struggle with.

Cloud changes data governance and data lifecycle management due to its scale and speed of the processes. Just as with other aspects of data security, cloud speed and scale make existing approaches ineffective or, sometimes, impossible to implement.

Speed and scale

Cloud also sped up many of the IT processes, driving the need to accelerate many of the data security processes. For example, making decisions on who can access the data cannot take months to achieve.

Cloud also brings an incredible scale of computing. Where gigabytes once roamed, petabytes are now common. This means that many data security approaches, especially the manual ones, are no longer practical. The very nature of the public cloud speed and scale destroys some traditional practices and approaches. At the same time, new approaches become possible: encrypting all data by default; rotating keys across your entire environment within minutes; ensuring all connections between users and systems are encrypted by default and others.

Data loss prevention

When data primarily resided on-premises, the key question for IT administrators and security teams was often: “What’s crossing my boundary?” This placed the emphasis on network-based controls. This aligns essentially to the well-worn analogy of the “walled castle” security model: build high enough walls and a moat with hungry alligators to keep threats on the outside. In fact, some organizations are so focused on DLP as a border control that they consider DLP to be a magical solution to all data risk, while it is as important to reduce your risk footprint by knowing what/where/who/how and when the data is stored and used.

Today, the question has evolved to be: “Where is my data? What value does it hold? Who and what has access to it? Is it still in the right context?”

Data loss prevention, as it was practiced years ago, just doesn’t fit the realities of cloud computing today. However, the need for technologies that focus on detecting exfiltration, discovering sensitive data, or performing other data-aware security tasks is higher than ever. Data loss prevention in the age of cloud is not about blocking the flow of data. Instead, it’s about knowing where the data is, what it is, and who has access to it.

Segmentation

Let's return to the walled castle model. We know by now that walls should be relegated to history books. Cloud has dramatically changed the practice of network security, including network segmentation. Many of the traditional on-premises concepts that worked really well, such as a DMZ, along with many traditional network architectures, are either not applicable in the cloud or not optimal for cloud computing.

But that doesn't mean that DMZ and similar concepts should be completely left behind. Instead, its principles can be adjusted to the modern environment. For example, using microsegmentation with access governed by the identity in context is a more modern approach to DMZ. Making sure that the right identity in the right context has access to the correct resource gives you strong control. Even if you get it wrong, microsegmentation can limit the fallout to a much smaller scale.

Technologies such as containers already have these elements in place. Having a layered approach and not relying on a single control are key building blocks towards a 'Zero Trust approach'.

Some organizations practice network security in the cloud as if it were a rented data center, thus not utilizing any of the cloud-native data security controls and relying on traditional control that they can bring with them. If that's your case, you'll end up getting fewer benefits while suffering from many of the pitfalls. This signals that you should be taking a different look at cloud security and consider the examples just given.

Data encryption

Encryption is one component of a broader security strategy. It adds a layer of defense for protecting data. It ensures that if the data accidentally falls into attackers' hands, they cannot access the data without also having access to the encryption keys. In many cases, the old wisdom that "encryption is easy, encryption key management is hard" means that to be an effective modern data security safeguard, key management needs to be rapidly modernized.

Traditionally, encryption meant encrypting the storage media, or setting up some form of an encrypted tunnel between two endpoints. This still holds true today, however, certain security challenges that drove this encryption activity are no longer such a big issue in the cloud. For example, now you have less need to encrypt for physical threats, because your cloud providers are ultimately responsible for securing hardware, not

your individual enterprise. It's unlikely that someone will steal a hard drive from, say, a Google data center (to be clear, Google still encrypts all data at the hardware level).

The scale of encryption key management in the cloud changes from having a couple of hundred or few thousand on-premises endpoints to multiple thousands. That's why requiring encryption keys in the cloud challenges key management at scale. Couple that with short-lived resources such as containers that only require key material for a short period of time, and you have key lifecycle management that's often unchanged since the early 2000s.

In the cloud, encryption may exist for reasons other than security, such as government regulations and compliance. For example, you may have a requirement that a cloud user encrypts the data in a way that prevents access by anybody other than the client. That's a newer kind of risk that needs to be considered.

Data access

From one point of view, the layers of security used on-premises are logical and familiar to many security professionals, especially if you began your career before cloud. You have security controls in the database, in the servers, in the data center, with all of it behind your firewalls.

This model means that every time we needed to access data from the outside, every time we needed to poke holes into the perimeter, the castle walls went from impenetrable stone to Swiss cheese. And once inside the perimeter, traffic was typically more trusted – something that attackers loved. This has been a driving factor behind the Zero Trust concept, and even though Zero Trust has been around for a while, it's still not implemented in most organizations, whether for users or computing services.

What's more, remote access has been put under further pressure during the pandemic. While widespread remote access has worked from a technical point-of-view, data governance generally has not been updated to match the new paradigm. Now data lives in myriad locations and requires access from different networks, devices and systems, but much of the current security model is not geared toward this.

What should your next steps be?

Your data may have fallen out of love with your security model, but attackers haven't. It's time to shift focus and build a modern approach to data based on autonomic security.

Pillars for building modern data security

We've identified some issues around the classic approach to data security and the changes triggered by the ubiquity of the cloud. The case is compelling for adopting a modern approach to data security. We contend that the optimal way forward is with autonomic data security. Just like with Autonomic Security Operations, this approach can help transform data security and make it ready for the future.

Simply put, autonomic data security is security that's integrated throughout the data lifecycle. It makes things easier on users, freeing them from having to define and redefine myriad rules about who can do what, when, and with what data. It's an approach that keeps pace with constantly evolving cyberthreats and business changes. In this way, you can keep your IT assets more secure and your business decisions speedier. Sounds like magic, right? So what are the essential pillars for building this new approach? (Spoiler alert: It's not magic, but a constant willingness to evaluate, change, and adapt.)

Automated/embedded classification and encryption

Let's start with data classification - a process for attaching labels to data, typically based on sensitivity or other dimensions. When your data is located on-premises within your databases or other data stores, you most likely need to employ some form of tooling plus the skilled resources to do the task. The challenge here is that it's hard to ensure all data is classified correctly – and that the classification remains in line with the data throughout its lifecycle.

> **Consider this scenario:** A data scientist runs an experiment using some moderately sensitive data. This data is then transformed by combining it with different datasets and then deriving new insights from it. This data now plots a clear path of how to optimize your customer engagement experience in a way that would lead to a 15% increase in your customer base.

This means that this data is carrying much more value than it did before. As such, this leaves your data classification running behind by both running even more behind and by missing more and more data.

After classification, you still need to consider how encryption should be done. You have many options, from encryption algorithms to the storing and management of keys, along with the need to meet FIPS and other requirements for compliance purposes. Keep in mind that in many cases, a lot of data that's stored at rest in an on-premises environment remains unencrypted – adding to the challenge.

In contrast, when your data resides in the cloud, both classification and encryption can be determined, assigned and enabled automatically, by default. Consider default encryption at rest in Google Cloud, where stored data is always encrypted. You can choose which encryption methods apply, including the use of Google provided keys or encryption keys you manage, but the starting point is encryption by default.

One way to put classification,encryption, and data de-identification (another strategy for securing sensitive data) together could be by applying things such as a predefined template for securing a data warehouse for confidential data. This blueprint includes pipelines that de-identify and re-identify data in two ways:

- The first pipeline de-identifies confidential data using pseudonymization.
- The second pipeline re-identifies confidential data when authorized users require access.

These examples illustrate the automated and embedded principles for encryption and classification. To go a step further, you could create an ingestion pipeline that classifies the data as it enters the cloud. You could also set automated life-cycle policies around the data (for example, data older than 30 days and containing PII is automatically crypto-shredded).

Encryption also works for when the virtual machine is running in use via Google Cloud Confidential Computing. This means that the virtual machine processing the data can be encrypted including with keys that the cloud provider does not possess.

Factoring in automated classification and encryption makes data security easier. You don't have to do any retrofitting or add on various data components. This autonomic approach throughout the data lifecycle creates a frictionless experience for users, with faster, easier, automatic adaptation to managing assets, threats, and business needs.

Integrated access to data over any channel

Data doesn't sit still – it travels. It's processed. It's accessed at different points, at different times, and in different ways. Sometimes third parties, partners, and customers legitimately access it. Security needs to be part of all the technology stack layers – and not just data at rest but also data in transit and data being processed. This means that data should be protected at all times, and only approved access in the correct context by the appropriate authorized resources, users, and applications can be enforced at all times, no matter where the data resides.

Google Cloud provides more flexible approaches to this comprehensive data security within an on-premises data center, where the ability to automate and embed controls across the technology stack makes this a reality.

With Google Cloud, we provide more nuanced control of which device, which person, and which location can access data, which is more aligned with Zero Trust principles. By comparison, managing data access on-premises depends on more coarse-grained rules. As a result, rules don't change as often as the business demands. And overly broad rules are often set, which can increase data exposure and business risk.

Layering a combination of coarse-grained and fine-grained capabilities gives you a model that spans different channels.

A virtual private cloud (VPC) service control is an example of a coarse-grained approach. Using VPC service controls ensures that only resources that are part of the perimeter can interact with the data in question, providing a layer of protection when data is being used. Furthermore, the control prevents data at rest from being exfiltrated as the control confines that data to the perimeter only and, in this case, also secures it in transit by limiting where it can be moved.

A more nuanced control example is using Identity-Aware Proxy (IAP), as discussed in this Bank of Anthos use case. Access to a GKE control plane is enabled through a bastion host, with one host in each environment. Each bastion is protected by IAP, and context-aware policies can be applied to ensure that its access is only allowed from the appropriate endpoint under the right context with an authorized user.

Another example is crypto-isolation, which is when two datasets have different data encryption keys. These two datasets can co-mingle and – because they have different

encryption keys – they remain crypto-isolated. This concept is already in use on Google Cloud through our default encryption.

These examples show how controls can make up layers that can be integrated over every channel and embedded into deployment blueprints as part of your continuous integration and continuous delivery (CI/CD) pipeline to provide an automated rollout.

It should be noted that for the purpose of this paper we have not included all points of access channels. An important part in this whole chain is the strong endpoint control; they are critical as they are often the first or the last mile of the journey. Upleveling these to take a browser-based access approach in the form of Chrome gives you a big leap in establishing Zero Trust controls, and also allows for other benefits such as Safebrowsing and making sure that your corporate password is not entered into non corporate sites

Policy intelligence leads to autonomy

Compared to on-premises data, many cloud elements are API-driven and can be leveraged to create an increased level of automation, policy enforcement, and granular access to data. This makes data more secure and more usable by your business because less integration effort is required compared to an on-premises environment.

The cloud also offers great intelligence in identity systems, defining intent and policy at a higher level so that data is only accessible to whoever needs to use it for the business – and nobody else.

Expressing your security principles via policy as code is an example of policy enforcement. Your policies are then rolled out as code across your organization to establish built-in guardrails. From a security perspective, this means that your developers can be given more autonomy since you've already set certain guardrails and built templates of controls that they can use. From there, you can implement drift detection to ensure that these practices are being adhered to and to quickly spot deviations from the code and templates.

Policy Intelligence tools help you understand and manage policies to proactively improve your security configuration. Policy Intelligence in Google Cloud already employs this approach and reduces risk with automated policy controls.

A good security principle to implement at all times is “least privilege access.” In order to achieve this, a tool such as Recommender can be used to help remove unwanted access to Google Cloud resources, with machine learning making smart access control recommendations. With Recommender, your security teams can automatically detect overly permissive access and rightsize them based on similar users in the organization and their access patterns. As an example, let’s say that a set of permissions hasn’t been used in 90 days. The tool will then recommend that you revoke the role. You could also take that a step further and trigger an automated response to remove the permissions altogether. A greater level of autonomy is achieved by having the system figure out the right set of permissions based on the context of how it’s being used.

Risk and compliance as code (RCaC), is another example of security policy enforcement on Google Cloud. It gives you the ability to assert infrastructure and policies as code, while detecting drift and noncompliance.

Reduced friction and complexity

Securing data oftentimes focuses on minimizing risk, as in minimizing the risk of an unauthorized party being able to access data, or the risk of data becoming unavailable due to an internal or external event. An often-overlooked element is security control usability. Consideration needs to be given as to whether or not it’s still relevant to achieve the originally intended outcome by implementing the control in the first place. This can create friction, because at times security controls unduly make things hard on the user of that control. Complexity also arises when on-premises controls are retrofitted into cloud – with some of those controls predating the existence of the cloud.

A modern approach to data security involves understanding how security controls and their technical components achieve their purpose of securing the data – and how it all affects the user journey. This requires a mindset shift, as you now need to start thinking about security as a product. Taking this type of approach will also focus your efforts on reducing both friction and complexity. After all, if your product is not doing its job effectively, why would the user want to use it?

Here’s an example. By creating organization policy guardrails through organization policy constraints, you’re able to abstract some security layers away from users. This in turn reduces complexity, removing yet another thing users have to think about. Plus, these guardrails get applied across your organization. So, when you understand user

interaction in relation to an organizational policy, you can then set a policy that prevents public access to Google Cloud storage. This control prevents existing and future cloud storage resources from being accessed via the public internet.

From a risk mitigation point of view, this is an excellent point of security control. When you begin to understand the use cases of the teams using the platform, you may learn that certain use cases actually require this control. Having this knowledge helps you rightsize the control and its applicability. It reduces friction, versus a blanket deployment without full usage understanding.

Measurability vs. business outcomes

Hand in hand with reducing friction and complexity is the ability to measure the usage, uptake, and user experience of the controls to gain insight into how these align to business outcomes. As you begin to think of security as a product, you can then seek to measure how the product performs. When you launch a new product, it's expected that a key component will be measuring the product's performance. But this approach is very seldom taken with security controls. As a pillar of modern data security, measurement can provide data points to the effectiveness of your security control and usage.

Take the following simple example: Fictitious Company A takes six weeks to deploy version updates to an application. From a security perspective, a new control and process is introduced to lower the risk associated with that application. However, this means Company A will need 12 weeks to deploy a version update to users. Is this still a good control aligned with the business outcome? Was the benefit of lowering the risk worth the extra six weeks of deployment? Remember, this is where measurement can provide insights.

Taking a creative approach to measurement can inform innovations. For example, in many highly regulated industries, recertification is a key concept. The more automated this process can become, the better it is for the business. That's because the potential for human error is less and it's easier to demonstrate results to auditors.

This raises questions like: "How do I measure what data people are using?" More specifically: "How do I determine if and for how long people should still have access to that data?" Instead of doing recertification at the end of the year, the goal is to measure constantly throughout the year. This enables you to recertify in a much

shorter, more efficient time period, or even on an ongoing basis. This approach to recertification brings both security and business benefits.

How you can think of this challenge is best couched in the "joiners, movers, leavers" concept as it relates to providing secure, dynamic data access. This is easy enough to do when someone leaves the company. If it's done right, the system automatically kicks the user out. Likewise, it's easy when someone joins the company. You give them the right access and they can go in and do their job.

Access for a user who moves from department to department poses more of a challenge. The optimum solution does not require an immediate granular rule definition when the user starts a new job. This kind of quantitative recertification works this way: You just move. Access to where you've been is dropped and you're recertified with your new access. Here again, measurement is required to determine business outcomes.

Visibility of the data processing supply chain

Most companies face continual pressure to launch applications faster. To achieve this, shared libraries or components are often typically used instead of recreating from scratch. Open source is a great tool for this. But from a security perspective, you should always take your software bill of materials into consideration. In doing so, you can better understand how these components interact with your data and how you can best factor in optimal data security. Traditionally, most organizations have not considered this when employing open source software.

A prime example comes from the open source software community, as illustrated by the impact of the Apache Log4j vulnerability. As discussed in a recent Google blog, "More than 35,000 Java packages, amounting to over 8% of the Maven Central repository (the most significant Java package repository), have been impacted . . . with widespread fallout across the software industry."

Gaining visibility into your data processing supply chain is the starting point to understanding your risk and setting appropriate security controls that can ultimately be embedded and automated to help lower the risk. Just like with software, the data processing supply chain has processors and their suppliers with the need to gain visibility over the entire chain, and then control it.

This is also something that's very commonly overlooked in traditional security approaches, however, and could lead to various impacts, such as what we saw with the SolarWinds breach.

An approach to think of here is supply chain levels for software artifacts (SLSA). SLSA framework formalizes criteria around software supply chain integrity to help the industry and open source ecosystem secure the software development lifecycle. SLSA introduces this by providing levels with increasing integrity guarantees to give you confidence that software hasn't been tampered with and can be securely traced back to its source. Here's a summary of the SLSA levels.

Summary of SLSA levels

Level	Description	Example
1	Documentation of the build process	Unsigned provenance
2	Tamper resistance of the build service	Hosted source/build, signed provenance
3	Extra resistance to specific threats	Security controls on host, non-falsifiable provenance
4	Highest levels of confidence and trust	Two-party review and hermetic builds

Using such an approach would allow for insight into the supply chain process, the risk thereof, and the measure you can take to lower the risk to your data. Google Cloud Build already supports SLSA Level 1.

Not only is supply in terms of third party libraries important but also considerations of supply in the sense of your Cloud Service Provider. Where is my data located, which controls do I have to safeguard it and how do I monitor access to it from a CSP

perspective. From a Google Cloud perspective it has taken the utmost care to ensure that there a contractual safeguards in our Data Processing terms and also technical controls ranging from Assured Workloads, Access Transparency and Access Approval to Sovereign Cloud offerings

Data lifecycle transparency

Data lifecycle transparency includes every aspect of data lineage and every movement of data beyond who accessed it when and where. It involves who created the data, how it's used, its retention, and even its destruction, closely aligned with compliance requirements that specify how long data should be retained and stored.

This requires that you have a robust data lifecycle management approach in place, which can be a difficult challenge. Understanding what you have out there is a good first step. As discussed, automated classification is a pillar in the modern approach to data security, one that would answer key questions like: "What data do I have? And how is it classified?"

Tying those answers together, you could set automated policies that might say: If confidential data of type X is not used for 30 days, it should be set in cold storage through a retention policy. By measuring and understanding the use of the data, you could also reduce access permissions to only a group of archive-retention administrators. Another scenario: If data is classified as type Y and not used for 30 days, it gets scheduled for deletion.

Now you can see how the pillars start to work together.

On Google Cloud, Data Catalog is a technology that brings together key aspects to data lifecycle transparency. It provides a fully managed, highly scalable data discovery and metadata management service designed to aid in answering questions like: "Is my data fresh, clean, validated, and approved for use in production? Who is using my data and who is the owner? And who and what processes are transforming the data?"

Answering these questions can help you set automated policies and gain a better understanding of lifecycle transparency, all the way to the decommissioning of data.

Data security as enabler

Having a good data security model in place does not mean data needs to be confined to an island to be secure. Having an autonomic data security model in place means that the right parties have access to support business collaboration without having to grant unilateral access.

This can also lead to data security being an enabler. Many important research, business, and social questions can be answered by combining datasets from independent parties where each party holds their own information about a set of shared identifiers (such as email addresses), some of which are common.

An example of what is already an enabler today is Confidential Computing that has helped unlock computing scenarios that have previously not been seen as possible.

But when you're working with sensitive data, how can one party gain aggregated insights about the other party's data without either of them learning any information about individuals in the datasets? Although the promise of fully homomorphic encryption is still some time away from being more viable in day-to-day usage, Confidential computing already provides some applications of this, taking this a step further multi-party computation can additionally add benefits to the above question .

To enable secure data sharing, Google has already provided open source availability of Private Join and Compute, a new type of secure multi-party computation (MPC) that augments the core private set intersection (PSI) protocol to help organizations work together with confidential datasets while raising the bar for privacy.

Having the pillars of the autonomic data security model in place allows you to take advantage of forward-leaning concepts like MPC, giving you a good foundation to build upon. Not having this in place is like building the tenth story of a building without the supporting infrastructure – and we all know how that will end.

Ready to move to new-world security?

Taking the precepts, concepts, and forward-looking solutions presented here into consideration, we strongly believe that now is exactly the right time to assess where you and your business are when it comes to data security.

To prepare for the future, we recommend you challenge your current model and ask critical questions, evaluate where you are, and then start to put a plan in place of how

you could start incorporating the autonomic data security pillars into your data security model.

The path to new-world data security starts by asking the right questions.

Key data security questions	**New-world data security**
What data do I have?	Automated/Embedded classification and encryption
Who owns it?	Automated/Embedded classification and encryption, Policy Intelligence leads to autonomy
Is it sensitive?	Automated/Embedded classification and encryption
How is it used?	Measurability vs. business outcomes, data life-cycle transparency, visibility of the data processing supply chain
What is the value in storing the data?	Measurability vs. business outcomes

Please reach out to your Google Cybersecurity Action Team if you would like to engage in further discussions around what you can do to implement a modern approach to autonomic data security.