Google Cloud Whitepaper
January 2021

Designing and deploying a data security strategy with Google Cloud

Authors:

Andrew Lance, Founder & Principal, Sidechain
Dr. Anton Chuvakin, Head of Solutions Strategy, Google

Google Cloud

Table of contents

Executive summary

Many organizations are looking to the public cloud to solve a variety of business challenges, whether it's achieving greater scalability, global resilience, increased reliability and performance, or to bring applications to market faster. The journey to move applications and workloads to the cloud is complex though, with **data security** often being top of mind. Some organizations are hesitant to migrate sensitive data to the cloud, as they explore their security options and struggle to understand regulatory demands.

Simply applying a data security strategy designed for on-premise workloads isn't adequate. It lacks the ability to address cloud-specific requirements and doesn't take advantage of the great amount of security services and capabilities Google Cloud has to offer.

Successfully deploying data to Google Cloud Platform in a way that delivers adequate protections, meets compliance requirements, and reduces risk, requires reassessing legacy data security strategies and creating a *cloud-ready* program. This involves understanding how the cloud impacts your current strategy and pivoting the strategy to embrace the great capabilities available in Google Cloud.

It also uses 3 key pillars as the center of an effective cloud-based data security strategy:

- **Identity**: Understanding the identity of users, machines, and applications as they create, modify, store, use, share, and ultimately delete data is core to an effective cloud-centric data security strategy.
- **Access Boundaries**: Creating guardrails for how data is accessed, by whom, and under what circumstances is the second core pillar for protecting data in the cloud. Your cloud-focused data security program should use *identity* to control *access* through policy, and take advantage of the many tools and services available in GCP that make this effective and easy.
- **Visibility**: Once the data guardrails are in place, use powerful visibility services in Google Cloud to audit usage and provide compliance reporting that demonstrate how data is controlled and accessed, not only by your own cloud administrators but by Google personnel that assist with your cloud infrastructure. These visibility tools provide rapid detection of anomalies and threat detection, enabling targeted response activities.

Google Cloud not only offers the services and tools to adequately protect data with even immense security requirements, it does so without steep learning curves such that even resource-constrained security teams can achieve a powerful data security posture and accelerate the delivery of security to their organizations.

Disclaimer

The content contained herein is correct as of January 2021, and represents the status quo as of the time it was written. Google Cloud's security policies and systems may change going forward, as we continually improve protection for our customers.

Data security challenges in the cloud

Many security professionals who have managed data security programs for on-premise environments are now finding that their organization's shift to the cloud requires a re-thinking of how data is protected. Questions arise, such as:

- How does my data security strategy need to change to accommodate a shift to the cloud?
- What new security challenges for data protection do I need to be aware of in the cloud?
- What does my cloud provider offer that could streamline or replace my on-premise controls?

As cloud is adopted for workloads containing a variety of data sensitivity, this paper addresses how to shift your data protection program to be cloud-ready. In many instances, as we'll see, cloud-based data security strategies still leverage the same familiar security patterns that have protected data on-premise for ages, while offering a more modern approach to security infrastructure that embraces the efficiencies and power that cloud platforms offer.

On-premise data security strategies usually consist of at least three foundational parts:

1) Managing data as a lifecycle
2) Data classification efforts
3) Data protection policies based on (1) and (2) that inform appropriate security controls

Data security controls will frequently include data encryption requirements, both at rest and in transit, and an increasing ability to protect data in-use by leveraging secure compute technology. Other common security components for data protection include key management, data-loss prevention, activity monitoring, and a variety of protections governing data use at the endpoint.

When shifting to a cloud-ready data security strategy, many of these foundational elements will remain, yet must be adapted to the cloud platform you adopt. Some on-premise controls aren't needed anymore, or can be replaced by cloud-native security capabilities. You will also be adding controls to your data security strategy, as cloud platforms may introduce new data concerns that didn't exist in the data center.

Managing a data security strategy for cloud is different than for on-prem environments. There's an entirely new set of technology and controls that

must be in place to adequately protect data in the cloud. There may be data sovereignty issues, for instance, with storing data on international servers that wouldn't be the case if such data were only stored in country-resident data centers. There are shared responsibility matrices that must be understood, and compliance requirements that require new kinds of reporting and auditing. As we will see, much of a cloud-centric data security program is less about managing full-stack threats, now that you are sharing security responsibility with your cloud provider, and more about creating data guardrails through access permissions and policies.

Whether your organization is lifting and shifting workloads to the cloud, or building cloud-native applications (or a combination of both), your data security strategy needs to be cloud-ready. The goal of this paper, whether you are starting a new data security program that is cloud-native from the beginning, or transitioning an existing on-premise data security program to the cloud (or are somewhere in between), is to give you guidance that will enable you to protect data in the cloud with a robust data security program that uses *cloud-native thinking* and a set of solutions that accelerate your ability to achieve high standards for data protection.

The 3 pillars of effective cloud security

While there are numerous security services, solutions, products, and technical controls at our disposal when creating a data security program for the cloud, there are *three pillars* of security controls for protecting data in the cloud: **Identity, Access,** and **Visibility**. These pillars represent the core of a security program, around which additional controls can be added to create whatever kind of comprehensive program is appropriate for your organization. These three pillars are considered core because every program will have these as a necessary component, and missing any one of them dramatically compromises the effectiveness of the program. Not all data, for example, needs to be encrypted at rest, but *all* data, regardless of its classification or stage in the lifecycle, must have deliberate **access** guardrails in place that grants the right entitlements for data usage (even if such access is ultimately "public" access).

Identity

Every data access policy rests on the integrity of *identity*. Operational data workflows depend on the identity of both humans, machines, and processes. Governance is a function of identity, and data sensitivity is often defined in terms of who gets access. In general, identity is a fundamental pillar supporting an effective data security strategy. Get this even partially right, and you've solved a big part of your data security challenges. Get this wrong, and no amount of other controls will offer anywhere close to adequate protection.

Start by identifying *who* needs access to data. This will likely be a combination of humans, services, and machines. Users, of course, are a diverse way that data is accessed: customers inputting data through an application, system administrators, analysts, internal users (customer service reps, for example), developers, and DevOps engineers, to name a few. GCP enables users to be managed directly within the Identity and Access Management console, added as members to a project, or through Cloud Identity, a service acting as a centralized hub for defining, setting up, and managing users and groups.

A distinct advantage of Google Cloud Platform over traditional data center infrastructure is that the many services available to you are already designed to work with each other, and this "service collaboration" is achieved by authorizing a special kind of account called a *service account*. For example, a Compute Engine VM may run as a service account, and that account can be given the permissions to access the data resources it needs. All services in GCP will use a service account when accessing data stores and other digital assets, which makes it easy for you to establish identity for them. Services accounts are managed through Cloud Identity and Access Management.

Finally, machine-to-machine identity is increasingly important in DevOps orchestration, automated code-to-production pipelines, IoT implementations, and other architectures. Certificates have long been one of the most common ways to identify and authenticate devices over networks, and this is no different in cloud infrastructures. To manage device identity through certificates, Google offers Certificate Authority Service (CAS), a highly scalable and available service that simplifies and automates the management and deployment of private CAs.

Boundary and access

Protecting data that never needs to be accessed is simple - encrypt it, store the key in an inaccessible way, and put the data where you like. Voila - protected. It's also unusable, and there *are* use cases for this, we'll discuss them later, but the vast majority of data security use cases involve securing data that *is* in regular use. Understanding **who** needs access is necessary, but ultimately you've got to start granting access, and this is where the real work begins.

Creating access rules to sensitive data is by far the hardest part of a data security strategy. This is why establishing data boundaries with strict access policies is so core to your strategy, a pillar that not only prevents access, but grants access, usually through myriad security controls and solutions. You will enable these access boundaries through multiple services and controls, providing a layering effect that protects your data according to its sensitivity.

Access boundaries to data can be managed at three levels:

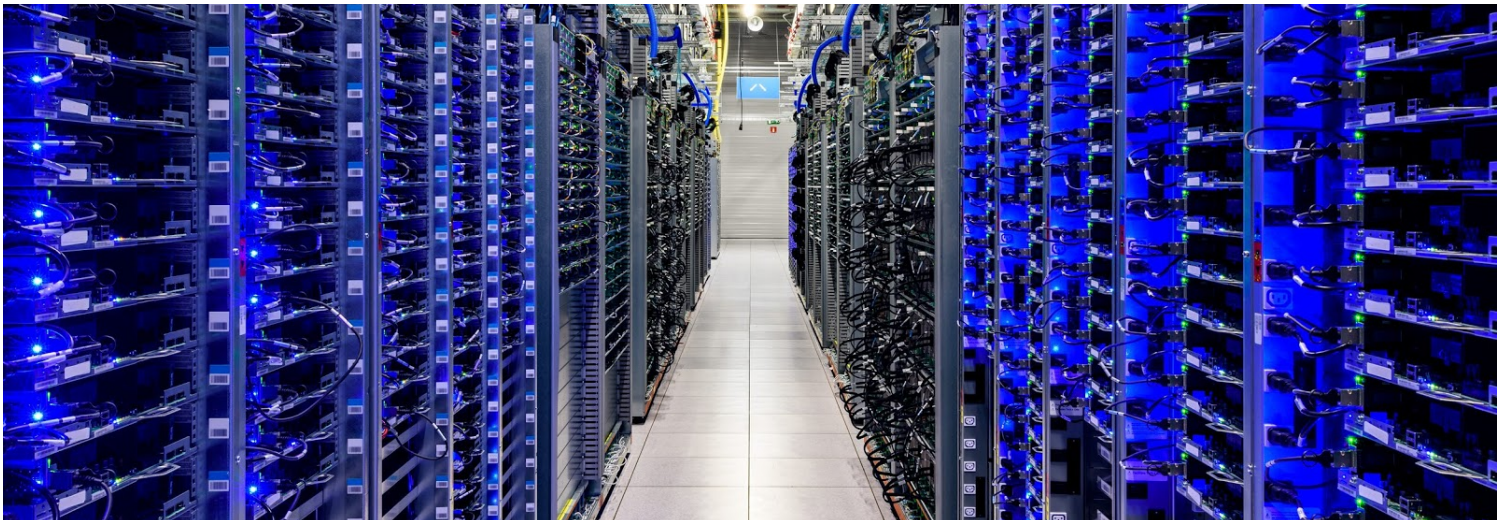
1) Network layer access
2) IAM access controls
3) Provider/Customer segregation

Network layer access

In on-premise security models, access controls often started at "layer 3", using network-based controls and attempting to "control the perimeter." While this approach by itself isn't effective for security protections, it is still a necessary part of any access control strategy. Within GCP, there are many capabilities that enable you to *segment* the access across your entire cloud infrastructure, starting with Google Cloud Firewalls. Use firewall best practices to segment access at the network level to sensitive data and sensitive workloads. Use Firewall insights and other capabilities within the Network Intelligence Center to fine-tune least-privilege access controls at the network.

Virtual Private Cloud (VPC) Service Controls also give you the ability to segment access to data resources. With VPC Service Controls, you create perimeters that protect the resources and data of services you explicitly specify, for example Cloud Storage buckets, Bigtable instances, and BigQuery datasets. This ensures that data resources within a perimeter are accessed *only* from clients within the authorized VPC networks. It's also important to note that these controls are a security defense independent of Identity and Access Management (IAM) (see below). VPC Service Controls prevent data from being exfiltrated or otherwise moved out of the control boundary.

There are other cloud segmentation tools available to you within GCP. Infrastructure is managed at the top level by "Projects" and these are strictly-enforced boundaries for Google Cloud systems. You may, for example, have non-production projects, and Production projects, each are separately managed, and access to each is governed through policy. You should use projects to segregate environments, applications, and other use cases.



IAM access controls

The vast majority of access rules, however, will come through Cloud Identity and Access Management (IAM). This is the master control center for authorizing who can take action on any particular resource within the GCP environment you manage. Several aspects of IAM help to manage access from a security standpoint.

IAM policies propagate down the hierarchy structure of your GCP environment:

- **Organization level**: This resource represents your company. IAM roles granted at this level are inherited down to all resources within the organization.
- **Project level**: Projects are a way of creating a boundary of services and resources.
- **Resource level**: Individual services and objects managed within a project

Google Cloud found that most permissions granted to cloud users aren't actually used within 90 days. This means that most users are simply over-provisioned when it comes to data access. And while convenient for managing large organizations with many administrators and projects, inheritance can sometimes be at odds with the security best practice of "least privilege access," granting broader permissions than necessary through inheritance. Take care when assigning permissions that they are done deliberately and by design, according to which identities need access to resources and for how long they need that access. See Designing Resource Hierarchies to help with how to manage IAM rules within your GCP environment.

Your data security strategy must account for every access type to data resources: users accessing data directly, service accounts accessing data, API access, other programmatic access, and how other applications or services will be accessing data. As discussed earlier, assuming there is an identity tied to each and every access to data, you'll be able to control it through Identity and Access Management policies.

Provider/customer segregation

IAM controls are vital for managing how *your* users are accessing data resources. But what about preventing unauthorized access to your data by the cloud provider? After all, moving sensitive data to the cloud requires a significant amount of *trust*, no matter how well designed the security is. It's not just trust in the cloud provider you need to be conscious of. Perhaps you trust your cloud provider, but not the country where they operate, and under whose laws they are governed. Geopolitical realities may require you to be extra cautious when storing highly sensitive data with your cloud provider. You may also have compliance requirements to encrypt your data in the cloud with keys ultimately in your custody.

To that end, Google Cloud Platform offers unique capabilities for ensuring your data is safe even from the cloud provider itself. Google Cloud External Key Manager enables you to use keys that you manage within a supported external key management partner to protect data within Google Cloud. With Cloud EKM, you control the location and distribution of your own managed encryption keys. Those keys are never stored or cached within Google Cloud. By leveraging Key Access Justifications, you also manage access to your keys, granting permission by Google to access your keys only according to your policy, with complete transparency to how Google is using your keys.

While encryption of data-at-rest and in-motion are a common requirement of data security strategies (and happen automatically within the Google Cloud), data typically must be decrypted for processing. Confidential Computing is a breakthrough technology that encrypts data while in use, all in real-time. By using Confidential VMs, you can run your workloads in private, encrypted services where your data is not exposed to the cloud provider.

Maintaining access controls

Many organizations have processes for enabling access to systems, applications, and data. After all, people need to get their work done. Very few, however, have processes for either reviewing that access periodically after-the-fact, or deprovisioning usage over time. The attack surface of accounts that have access to data that don't need it anymore is enormous, and rife for abuse. Traditionally, this can be an enormously difficult task because there is no unified tracking of account usage across the disparate systems and data platforms users may have access to within a data center. GCP makes this easier by providing IAM Recommender, a tool that looks at the permissions a user has versus which permissions they've used over the past 90 days. This gives you a good sense for how "over-permissioned" a user may be, and whether they're even using their access anymore.

For high sensitivity data classifications, access should never be granted indefinitely. Time-binding should always accompany access to this class of data, even if it requires a manual policy to review and revoke.

Visibility

One of the most pervasive problems with on-premise security strategies is that security is really a byproduct of stitching together disparate technologies and solutions that were never designed to integrate. When multiple tools, vendors, and technologies need to be used to secure the environment, managing across those solutions for consistency and visibility is a colossal task for resource-stretched security teams.

This is where the cloud offers a game-changing experience that is unparalleled with on-premise security efforts. Because native cloud platform services are all integrated seamlessly, it offers an opportunity to leverage consistent monitoring, detection, logging, and auditing, all combined with powerful analytics and discovery tools enabled by machine learning to provide insights and control. Beyond monitoring and logging, centralized management of security controls provides unified single-pane-of-glass visibility into how security for data is applied across your organization. There is simply no way on-prem security programs can achieve the same level of consistency without massive resources.

First off, if you haven't had the right tooling in place to perform data discovery on your legacy data, you can leverage Google Cloud Data Loss Prevention (DLP) to perform security scanning of your cloud data. Cloud DLP has native support for scanning and classifying sensitive data in Cloud Storage, BigQuery, and Datastore, as well as a streaming API to support additional data sources and custom workloads. Not only will Cloud DLP find sensitive data, identifying over 130 built-in data element types, it can automatically classify, mask, tokenize, and transform sensitive elements to enable you to better manage the risk of collecting, storing, and using data. In other words, it can integrate with your data lifecycle processes to make sure that data in every stage is protected.

In addition to understanding your data through DLP security scanning, you also need to know what's happening with your data as it moves along the lifecycle. How is it being accessed? How is it being moved and shared? Are permissions changing on it? Another significant benefit of GCP is that logging and auditing is natively aggregated by Google Cloud Logging. The Cloud Logging service is not only exhaustive in terms of events captured on the cloud platform itself, but it's extensible, enabling you to add additional sources if necessary, for example if you bring third-party security products to the cloud to help manage data. You can segment your logs by region, can store them in buckets, and integrate custom code for processing logs. You can also export logs to BigQuery to perform security and access analytics to help identify unauthorized changes and inappropriate access to your organization's data.

Security Command Center also helps identify and resolve insecure access problems to sensitive organizational data stored in the cloud. Through one management interface, many data security features can be activated to scan for a wide variety of security vulnerabilities and risks to your cloud infrastructure. A host of security assessment capabilities exist for your data, including the ability to monitor for data exfiltration, scan storage systems for sensitive data (Cloud DLP is integrated into the console), and detect which Cloud Storage buckets are open to the internet.

Controls that enable the pillars

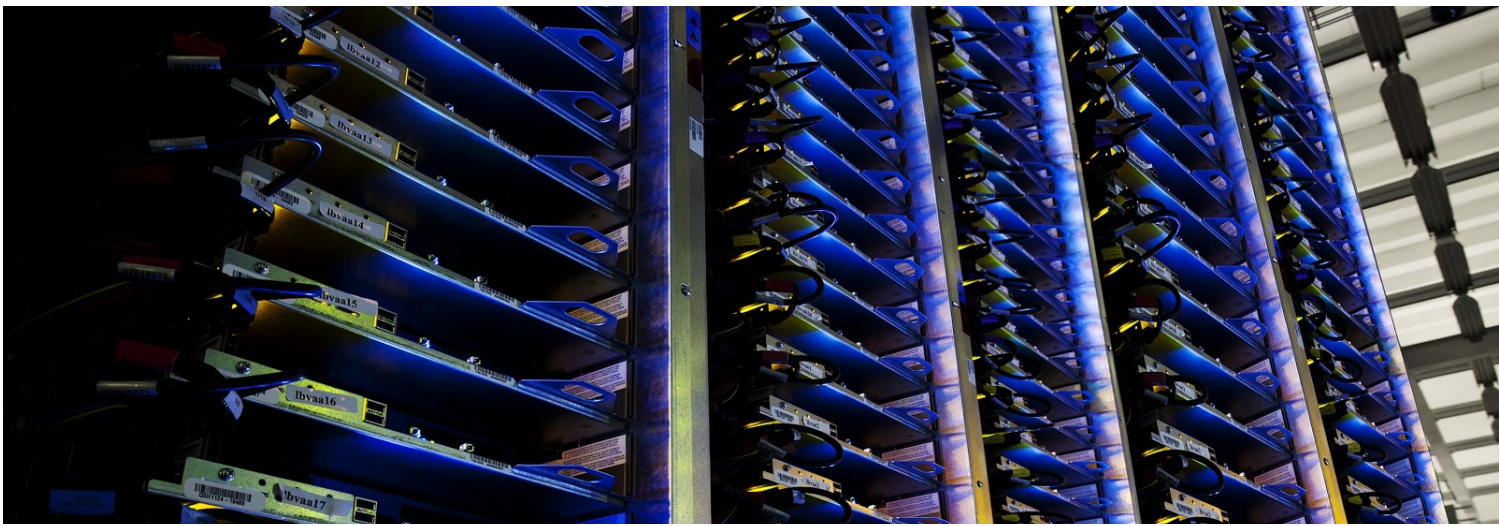
Using encryption to protect data

Data encryption has become a new norm for data of all sensitivities. Whether for structured data stored in databases, unstructured data like file shares and bucket storage, or other use cases like tokenizing data fields, encrypting data has become a commonly prescribed method for securing data in many regulations and compliance mandates. Encrypting data is also a strong way of segmenting and blocking unauthorized access, if you exert enough control over how encryption is implemented. Let's take a look at a few ways encryption can impact your data security strategy.

Unique among cloud providers, data is automatically encrypted at rest within GCP. This is provided completely transparently, and requires no intervention or setup on your part. In fact, no matter what other security controls you exert on data, it will *always* be encrypted at rest within GCP. Depending on your encryption requirements regarding key management, this may be sufficient to meet your needs.

GCP gives you a continuum of encryption key management options, depending on your particular needs. For example, if you have certain key management requirements such as the need to rotate your encryption keys, customer-managed encryption keys (CMEK) use Cloud KMS, GCP's world-class key management platform, to give you that control. Cloud KMS also enables you to encrypt your data with either software-backed encryption keys or FIPS 140-2 Level 3 validated HSM's. It also takes care of scaling out your key management needs across targeted regions, providing redundancy and global availability of keys. And if your key management needs require you to generate your own keys using your on-prem key management system, Cloud KMS enables you to bring-your-own-key (BYOK) to the cloud.

For even more control, however, GCP has implemented a powerful separation of duties between your most sensitive data and the cloud provider. Cloud External Key Manager (EKM) protects your data at rest in BigQuery and Compute Engine by using encryption keys that are stored and managed in a third-party key management system that you control outside Google infrastructure. Because your data within GCP is encrypted by keys you store outside of the cloud, you have high assurance that your data cannot be accessed whatsoever, and you truly achieve a secure Hold-Your-Own-Key (HYOK) model for key management. Additionally, Key Access Justifications work with Cloud EKM to give you complete visibility and transparency into every request made for an encryption key. Policies can be designed around these key access requests, approving or denying them as necessary.



Additional data security controls

There are other, more narrow, use cases for various sensitive data that will be used in the cloud.

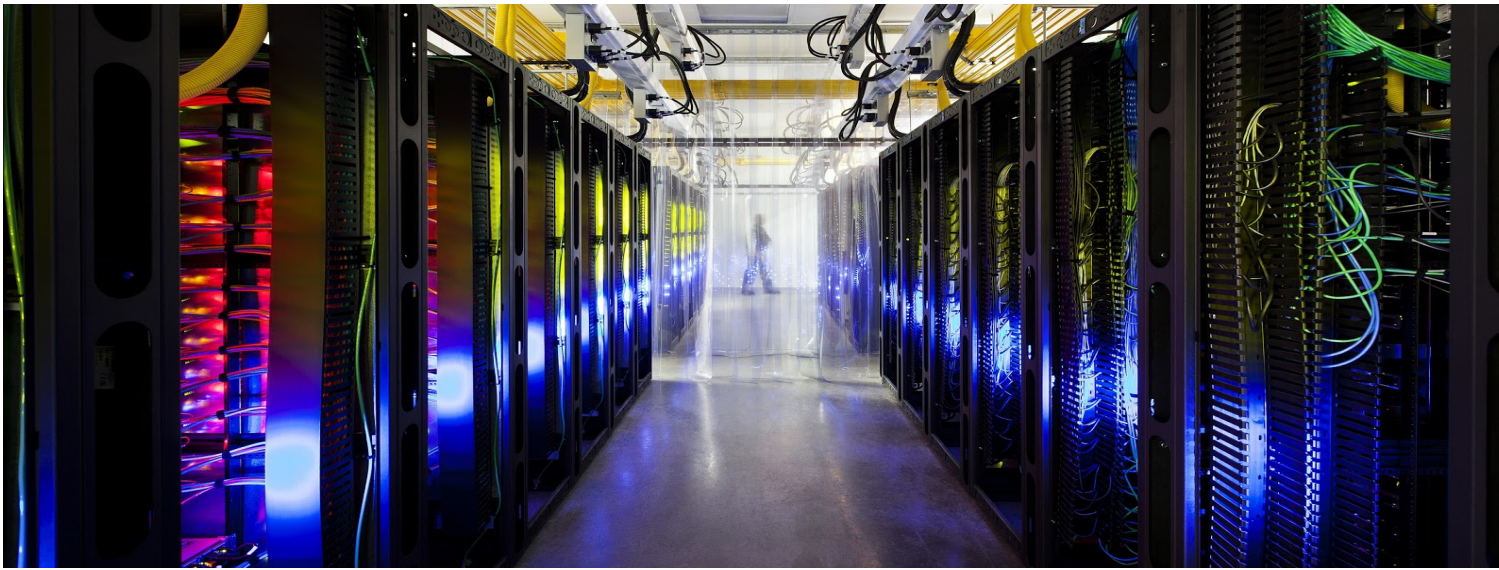
Secrets Management

Passwords, credentials, connection strings, API keys, and various other sensitive data fields or strings of data are often buried in configuration or properties files in traditional on-prem tools. This all changes with the cloud, because managing these *secrets* is fully integrated into the API and most services directly. These sensitive fields of data can now be stored in a powerful, yet simple, Secret Manager that secures and centralizes management and access of these secrets. Secrets can be automatically rotated, and applications can be configured to automatically use the latest version of a secret. Every interaction with Secret Manager generates an audit log, so you'll always have full transparency to every access to every secret. Cloud DLP also has a category of detectors to help you identify credentials and secrets that could be protected with Secrets Management.

Data Catalog

Getting control of all the data assets you manage in the cloud, particularly at scale, is a monumental task. As data is spread out across services and data stores, storage buckets and databases, and managed by different teams across your organization, finding and tagging data, classifying it, and then automatically enforcing data security policies based on those classifications and tags, is almost impossible to achieve with most legacy on-premise infrastructures, if for no other reason than, again, the fact that on-premise environments are managed by a collection of disparate products and solutions that simply can't integrate, coordinate, and orchestrate the way fully integrated cloud services can. Data Catalog solves this problem.

By using a powerful user interface, even non-technical individuals on your team can manage data at scale. Using Data Catalog, team members can find, curate, and use metadata to describe your data assets managed in the cloud. With Data Catalog, users primarily search for data assets, then tag the assets with metadata. By integrating with Cloud DLP to automatically identify sensitive data, Data Catalog can help accelerate your data classification efforts. Once data is tagged, users querying or using that data through Data Catalog views are restricted on what they can access using Cloud IAM access controls.



Rethinking your data security strategy for cloud-readiness

If you currently manage data protection on-premise, your strategy will change when applied to the cloud. Cloud-ready data security strategies may involve all of the new areas of concern that will need to be addressed as you pivot your strategy for the cloud. Let's discuss some strategies and considerations when adapting your current data protection program to the cloud.

When we move data to Google Cloud Platform, some decisions need to be made in order to inform how data protection should be cloud-enabled:

1) What are the regional/geopolitical implications with moving data to a cloud provider?
2) What workloads should be moved?
3) How will those workloads be moved?
4) Use the data security cloud-readiness checklist

What are the regional/geopolitical implications with moving data?

Depending on the nature of the data you manage, regulatory implications may be something to closely look at when designing your cloud strategy. Data protection laws across the globe are forecasted to increase by 65% over the next three years (Gartner). This means that as data is managed in a global cloud, strict attention must be given to how data is managed, stored, and utilized. Data residency laws may require that data gathered in one country remains in that country, which will obviously impact how data is deployed within your cloud architecture. Transborder data transfers are another area of uncertainty, and guidance is only beginning to emerge by regulators, particularly within the EU. A comprehensive data security strategy for the cloud will need to accommodate the latest developments in this area, particularly if personal and consumer data is managed.

Another growing area of concern particularly among European cloud customers is the ability to not only apply strong security controls to meet security policy and regulatory requirements, but retain greater security and autonomy for how data is managed, stored, and processed within the cloud. This is commonly discussed under the umbrella term of data sovereignty. Strong trends in the EU are pointing towards further regulation governing data residency and maintaining strict control over how European data is transferred not only within the EU but outside of the bloc.

Your strategy will not only need to adapt to these emerging trends, but should consider proactively implementing mechanisms to address these concerns within your target cloud platform. For example, you may require visibility and recorded audit when the cloud provider personnel needs to access data, such as fixing an outage, or when Support is attending to a request. Indeed, you may consider controls that prevent the provider from accessing sensitive data altogether, or an ability to implement controls that detect and prevent unauthorized access to data by various services running in the cloud platform. Google Cloud Platform offers several unique and powerful capabilities in this area, as will be discussed below.

What data is going to be moved?

Choosing which data will be moved to the cloud is usually derived by fundamental business drivers. Perhaps an application can scale better in the cloud, generating more revenue opportunities for the business. Perhaps the business is downsizing its reliance on data centers and colocation facilities, moving infrastructure to the cloud as a strategic initiative or part of a digital transformation effort. Regardless of the business driver, once a workload or application is identified as a candidate to move to the cloud, your data security strategy should kick in and inform the decision.

A cloud-ready strategy should be able to contribute to understanding:

- The *classification* and *sensitivity* of data
- Whether or not the data is regulated
- The compliance requirements for the data
- Can adequate security controls for the data can be met by the cloud platform

How will workloads be moved?

Workloads, that is, all system components that comprise an application, including data stores, tend to migrate to the cloud in one of three ways:

1) Lift and shift
2) Application migration
3) Cloud-native rebuild

The Migrating to Google Cloud solutions document explains that in a lift and shift migration, “you move workloads from a source environment to the target cloud environment with minor or no modifications or refactoring. The modifications you apply to the workloads to migrate are only the minimum change you need to make in order for the workloads to operate in the target environment.” Lift and shift strategies are beneficial operationally because they minimize application changes, limit the cloud skills resourcing required to support the cloud, and are usually a faster way to migrate workloads to the cloud. Likewise, since there are minimal changes being made, the goal from a data security perspective is to maintain as much of the on-prem strategy as possible.

Application migration, sometimes called an *improve and move* effort, moves components of the application to the cloud over a number of releases, architecting the application to take advantage of new cloud capabilities as components are migrated. Migration efforts enable your application to take advantage of cloud-native capabilities without wholesale changes to the entire workload.

Cloud-native rebuilds are a complete rewrite of the application into a cloud-native version. Usually taking the longest of the three approaches, a cloud-native rebuild is a challenging undertaking, requiring an understanding of the capabilities your chosen cloud platform offers and requiring the engineering skillsets to utilize them. While the effort to build an application specifically for the cloud can be daunting, the benefits can be well worth it, achieving maintainability, scalability, security, and the ability to rapidly enhance the application. Rebuilds often require new workflows and management processes, and will require a rethinking of data security strategy when bringing them online in the cloud platform.

There are other ways that workloads end up in the cloud. “Developer-led” projects happen when a development team build something in the cloud, and IT is forced to support it. This happens frequently with various lines of business, all who may end up making cloud-based decisions without consulting IT or Security. An analytics team may decide to use Google BigQuery without ever thinking about consulting the security team (who may be perceived as the team that will just slow down their initiative).

The service model of the cloud environment is also a factor when shifting your data security strategy. The shared security model between IaaS, PaaS, and SaaS are different, requiring you to maintain varying degrees of responsibility towards security in each. Understanding how your data will be used in the cloud, and under which model, will influence the security controls you can bring to bear when designing and deploying your data protection strategy.

How can I best prepare my data security strategy for the cloud?

The following is a checklist of cloud preparedness items to consider as you shift your data security program to be cloud-ready.

- ❑ **Is your data classification changing?**

  As you move data to the cloud, will it require changing the classification of any data? Chances are, it won't, but audit your data classification to make sure.

- ❑ **Is your data lifecycle changing?**

  How does cloud fit into your data lifecycle? Is there a cloud migration process that needs to take place? Is new data being generated in the cloud?

- ❑ **Are you adopting a single cloud or multi-cloud strategy?**

  Will data be centralized onto a single cloud platform or will it be spread across multiple cloud providers? Will data be replicated onto multiple clouds? What security controls do each platform offer to adequately secure data?

- ❑ **Prepare to leverage security controls in the cloud**

  The cloud offers powerful native controls to protect data that you'll want to leverage. Understand what your options are for leveraging them.

- ❑ **Can you de-identify or tokenize data moving to the cloud?**

  By using de-identification to reduce the sensitivity of data, it may facilitate moving to the cloud.

- ❑ **Create a data migration plan that includes data inventory**

  It may be necessary to create a data migration plan that includes a data inventory of data assets moving to the cloud. This plan should describe *how* data is migrated, including necessary security controls during migration.

- ❑ **Establish where data assets will land in the cloud**

  Data assets moving to the cloud may go into storage buckets, structured databases, a secret manager, or other data stores. Understand where data will land and be managed.

- ❑ **Are there data sovereignty needs that must be addressed?**

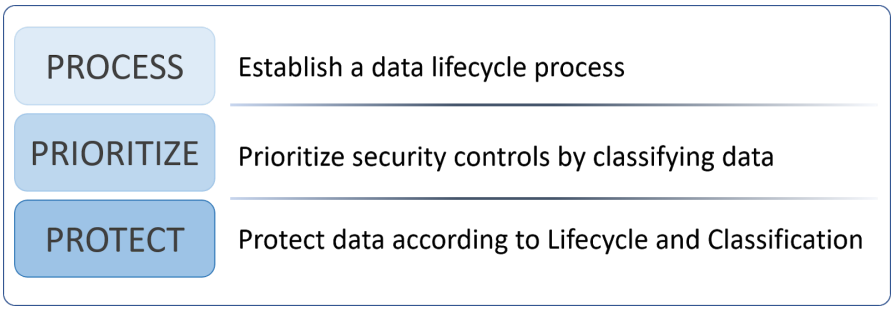
  As data moves to the cloud, understand whether localization, sovereignty, and regional attestation will play a role in governing the data.

- ❑ **Understand any skills-gap with leveraging cloud-native services**

  Does your team have the necessary skills and certifications to implement cloud-native security capabilities.

Data security foundations

The following foundations are data protection best practices, and should be considered in every data protection strategy. Whether you are designing a cloud-native data protection program, or pivoting an existing program to support your organization's migration to the cloud, these fundamentals should play a role.



PROCESS: Data should be managed in a lifecycle

Data doesn't exist statically. It is created, stored, accessed, manipulated, shared, backed-up, archived, and sometimes, destroyed.

When developing a data security program, managing data within the context of a lifecycle enables teams to apply the correct controls to the data as appropriate to its phase. Initially, think broader than security, and define the context for the data.

For each phase, consider the following:

- Purpose and value
    - Why is data collected, stored, and used? What business function does it serve? How does it fit into our business operations and how does it deliver value to our organization or our customers?
    - Is it necessary to collect and retain this data according to data minimisation guidelines? And if so, can it be tokenized/masked/redacted to reduce risk and still provide value?

- Privacy
    - Are there privacy concerns with the data being collected? Is there authorization to use the data and how is the privacy and protection of the data governed? What controls do end-users have over the data collected and stored?
- Data ownership
    - For each phase, who is ultimately responsible for the data? Who is accountable for the data?
- Liability
    - What are the implications if the privacy or protection of the data is compromised? What are the legal and regulatory requirements associated with this data in terms of notification and incident response?
- Security
    - How is the data secured in each phase? Is there unified security across the lifecycle or are there hand-off points? What security is necessary for the data, and what are the audit requirements of the security controls for the data in each phase?
- Standards and Data Quality
    - How is the data changed, manipulated, or otherwise modified? Are there legal or compliance requirements around change management? Is data quality a factor when collecting, storing, and using data?
- Regulations
    - How is the data regulated and what are the process and technology implications of that? Are there technical controls, governance controls, or privacy controls to account for? Is data residency a factor in how data is stored?

By considering the above attributes of the data, it informs the next characteristic of a strong data security program: classification.

PRIORITIZE: Data should be classified based on its sensitivity

Many organizations have wasted effort applying stringent security to data that doesn't need it. It's imperative to *classify* data based on its sensitivity so that *appropriate* security controls can be applied. By segmenting data based on its classification, time and energy (and budget) can be applied to the data that needs it most.

Data classification should be performed as early in the data management lifecycle as possible, ideally in the *Create* stage. Data classification efforts usually require only a handful of different categories of priority. For example, consider the following 4 category scheme:

- **Public**: Data has been approved for public access
- **Internal**: Non-sensitive data that is not released to the public
- **Confidential**: Sensitive data, general distribution
- **Restricted**: Highly-sensitive, restricted distribution, regulated data

Use all the knowledge gained by evaluating the *context* of data above (in the Process effort) to really understand your data, and why it's important to secure. Use big brush strokes for data; you're not evaluating every file in a file share, for example. Try and bucket data together, for example:

- All production databases supporting customer-use applications are classified *Confidential*
- All *restricted* data created by a certain business group (Executive Staff, HR, etc.) will be stored in dedicated repositories

One of the difficult tasks of classifying data is really understanding the criteria for each dataset. What makes a certain dataset *confidential* vs. *restricted*? Why is a certain PDF *internal* and not *confidential*? Creating as much clarity as you can on what the criteria is for each category will help when determining how data should be classified.

Data classification isn't just an exercise in understanding data or merely to fulfill a compliance requirement. In the next phase, **Protect**, these data classifications are used to implement appropriate data protection controls.

Cloud-based services from Google greatly help manage data discovery and classification by enabling you to catalog your data through both automated tools and an intuitive user interface. We shall see how these tools can be extremely effective at helping you as you move data to the cloud.

PROTECT: Data should be appropriately protected according to its phase and classification

Only after data has a) been defined within the context of its lifecycle, and b) classified based on its sensitivity and risk, can you begin to start assigning the right security controls to protect it. Class

Using the framework above, controls can then be applied to protect data in an appropriate way. This is usually accomplished by finding the right tools and technologies that address the specific ways data must be protected per its classification and phase of lifecycle. This is no trivial task. Finding the *right* tools varies, depending on your definition of *right.* How does the solution integrate into existing products and processes? Are the right skillsets in place to maintain the solution? Is it within an acceptable budget? Does it align with future roadmap? There is no shortage of vendors, consultants, and partners to help with this decision-making process.

It's important to note that the above process – Process, Prioritize, and Protect – can and should be applied to data regardless of whether it's managed on-prem, in the cloud, or with any other service provider. The implementation will be different, but the strategy remains the same: data must be managed in a lifecycle, data must be classified by sensitivity, and it must be protected according to both.

Putting it all together

Operationalizing a data security program in the cloud isn't about implementing every control a provider offers, but neither is it about dragging every control you had in the data center to the cloud. In real life, security is as much about budgets, working within the skill sets of your team, and how many resources you have, as it is about risk management. Implementation is often about knowing which controls to choose and, perhaps more difficult, which controls to hold on.

Imagine we're a financial services organization with a strategic, board-level mandate to shift application infrastructure to the cloud. We know we're having various teams use the Google Cloud environment: development teams from various lines of business, security, operations, a small DevOps team working on Kubernetes proof-of-concepts, and cloud administrators.

We have lots of unstructured data, file shares, file servers, and many of these have been in use for over a decade. This is organizational data as well as customer data. These file repositories have been in Production use for well over a decade. Over that time, applications have come and gone, regulations have emerged that govern our data, and we didn't originally architect for customer data segregation.

Per our data security strategy, before we jump into securing our data, we need to get a handle on identities managed in the cloud. It's absolutely critical that we deliberately provision our users in the cloud and not over-provision-by-default as we work quickly to migrate users. As engineering teams are provisioned in IAM, how do we know which GCP services they will need access to? It would be tempting to give them broad access to anticipate their needs - this is not recommended. In fact, by default they have access only to the specific assets they are working with: Compute Engine VM's, individual Storage Buckets, etc. This creates a flurry of activity and access requests early on in the project - again, the temptation to over-provision creeps in - but it's vital that users and service accounts are restricted.

Several newer applications can be modified for the cloud to take advantage of cloud-native architectures, but it's the DevOps team working on integrating a few pilot applications into Google Kubernetes Engine. For this, they have secured various authentication needs with certificates and, having not run a Certificate Authority before, are leveraging the Google Cloud Certificate Authority Service due to its ease of integration.

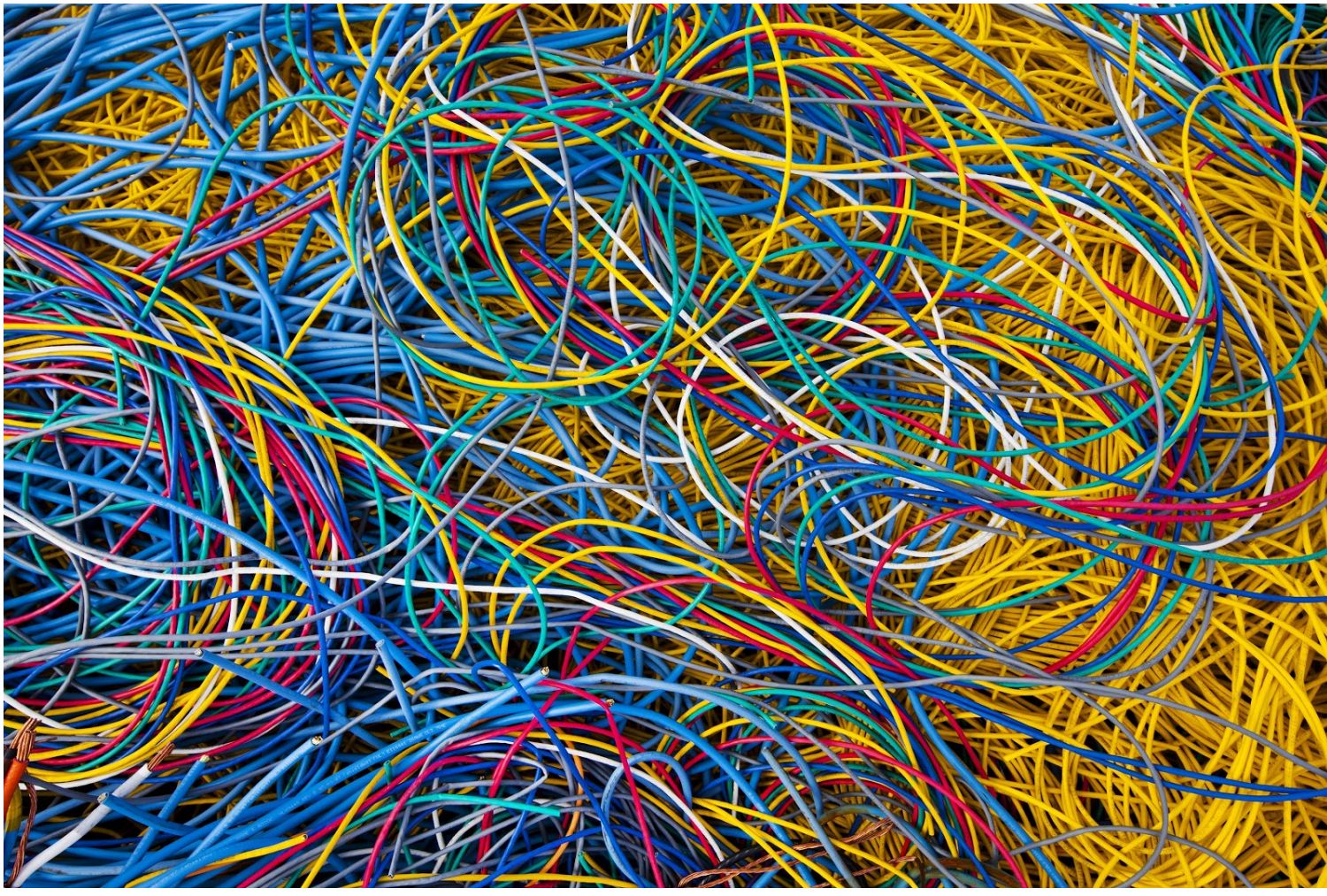
It's been determined, however, that due to the massive volumes of legacy unstructured data associated with older applications, these need to be moved using a lift and shift approach, minimizing the amount of changes to the applications, and doing a wholesale migration of the data. There were concerns about this, as several larger customers would raise eyebrows knowing their data was stored in the cloud as these applications were migrated. The data migration, therefore, would go through a process by which enhanced security would be applied to sensitive customer data.

First, the unstructured data would be staged in locked down Cloud Storage Buckets, encrypted by default, and only accessible to the data team working on this phase of the project. The team used Cloud DLP to scan for sensitive customer data, including custom fields such as account numbers and various other customer identifiers. Once the sensitive data was identified, it was migrated to dedicated Storage

Buckets that were encrypted by Customer Managed Encryption Keys (CMEK) to enable periodic key rotation. This data was also restricted to a minimal number of human users, and only the required service accounts from the Compute servers and databases. The rest of the data was segmented into other Storage Buckets where default encryption and more relaxed IAM controls would be applied.

Migrating databases was a different process. Engineering already knew exactly which databases contained sensitive and critical customer information, so no discovery process was required to know where additional security controls would be needed. As applications and databases were migrated, the Cloud SQL instances were encrypted using CMEK. For a few crown jewel databases, the team is exploring using Cloud DLP to tokenize certain columns as a near-term project.

Rather than bringing on-premise, third-party vendor products to the cloud for security, the Security team is piloting the vulnerability scanning and threat detection in Security Command Center (SCC), while still conducting regular pen tests of the applications (trust, but verify). Likewise, they are building security compliance automation to analyze IAM permissions and leverage IAM Recommender to ensure that users aren't overprovisioned, and using Logging and Audit data to aggregate security events like key access, sensitive data access, and permission change events. Much of this information they could never get from third party teams managing on-prem infrastructure, due to complications of technology integration, and internal politics.



Accelerating security

The organizations that will utilize every security capability GCP has to offer are few and far between. There's just too much. Most organizations don't need that level of security based on the risks and compliance requirements they face. Deploying a successful data security strategy is about choosing *appropriate* solutions based on the level of security risk carried by the data. Some data, given its nature, will be highly regulated, highly sensitive, and/or mission critical, and thus, introduces significant risk that must be mitigated to preserve business continuity and growth. This data, however, is on the extreme end of necessary protections, and most data will fall somewhere along a spectrum of risk.

Part of running a successful data security program in the cloud is recognizing that all of this *is* hard to implement, maintain, and operationalize over time, especially as teams change and resource-taxed security teams are required to manage more with less. But this is the whole objective of a cloud-native data security strategy that prioritizes **identity, access boundaries,** and **visibility**. These pillars, if focused on, give security teams the biggest return on technology investment for securing data. Laying in additional controls as necessary, and using cloud-native, integrated services, *accelerates security*, enabling even modest security teams to create robust security postures using GCP services and solutions.

When you design your data security strategy to be cloud-native using Google Cloud Platform security services, you take advantage of capabilities in the cloud that simply *cannot be replicated on-premise*, whether that's aggregating security visibility across the entire infrastructure into a seamless management console, having integrated security event logging at your fingertips, leveraging recommendation tools that use machine learning to find over-provisioned accounts and alert you proactively, or encrypting every byte of data by default with no additional management or overhead. These are data protection capabilities very few teams running on-premise will ever achieve, and yet are fully within reach by security teams using the cloud.

Your journey to the cloud

You will invariably need to confront data security requirements in your journey to the cloud, and performing a “lift and shift” for your data security program won’t work to address the unique opportunities and challenges the cloud offers. What worked to secure data on-premise isn’t entirely adequate when data and workloads move to the cloud, so you need to assess your data security strategy and adapt it to be cloud-ready, addressing new concerns and needs of the business operating in the cloud.

As you first assess, then adapt, your data security needs for the cloud, think cloud-native and consider all of the controls and services Google Cloud Platform offers with your data migration efforts. Inventory the security controls required on-premise, and map them to the powerful capabilities GCP delivers, focusing on the three pillars of cloud data security: identity, access boundaries, and visibility. Understand the *impact* the cloud will have on your data security program that didn’t exist on-premise, for instance whether there are data sovereignty concerns for data stored in the cloud, or what it means to have adequate accountability from your provider about how they are accessing your infrastructure. GCP accels at offering not only comprehensive controls to address all aspects of a cloud-ready data security strategy, but does so with intuitive onboarding and adoption so even resource-constrained teams can achieve their data protection goals.

As your organization moves its infrastructure and operations to the cloud, shift your data protection strategies to *cloud-native thinking*. Leveraging Google Cloud Platform to rapidly deploy a data security strategy for your data and workloads, you will greatly accelerate your ability to deliver the data protection standards your business deserves.