KPMG

In data we trust, but how do we protect it?

Keeping pace with your data protection through proactive and automated data protection capabilities

kpmg.com

What does modern data protection look like?

What is the first thing you think of when you hear data protection? When most people think of data protection, they think of locking down their data assets and information, restricting access, and encrypting for starters. Data protection is much more than just keeping data safe. Data protection can and should be a continuous, programmatic approach to how the data that an organization creates, uses, and procures is protected, secured, monitored, privatized, and managed in an agile way, to adapt as the technology and regulatory landscape continues to evolve.

With traditional data protection, the key areas that are generally considered are data inventories, data loss prevention, data encryption, and data governance policies. This approach can be highly manual, perhaps understood and living out of spreadsheets. In the past, some may think of this approach as more rules based on a per-application basis and may notice this gets harder and harder to maintain as the number of applications and data volumes grows and the amount of data to protect increases. The bigger the spreadsheets get, the more policies you must worry about reviewing quarterly, annually, and sometimes manually and reactively try to keep pace with the proliferation of your data as technology advances and, in turn, evolves within your organization. When cloud is introduced into the data protection landscape, the traditional approach and governance challenges of data protection can compound and feel unsustainable to maintain.

Enter cloud technologies, and what will most do when migrating to the cloud? Lift and shift, is that what you and your organization have been doing? Or is it time to rethink data protection to make it work for you and your organization? The capabilities of cloud do present more opportunities for data protection, elevating the ability to do more faster, better, and more consistently when protecting your data assets. One common theme that many organizations first run into and observe, is that while the infrastructure and platforms are often managed by third parties and vendors, often the data management remains the customer's responsibility. The concept of data-democratization is wonderful and strategic for organizations leveraging cloud capabilities and mining their digital data assets for value. But in such an environment, the surface area for access to data assets can expand exponentially and needs additional consideration for evolving data protection during migration to cloud.

Digital transformation and cloud migration are some of the top priorities for organizations in the next few years and represents some of the most significant changes to underlying technology that most organizations will undertake, compared to the past 20 to 30 years. This presents an ideal opportunity to reimagine how organizations can better manage and protect increasingly important and valuable data assets. This means evaluating how one evolves practices around data protection for these cloud technologies and building in adaptability to "future proof" as best as possible. This means rethinking strategies around data management and data protection practices before moving workloads or even traces of data to cloud. (Think back to the old age of beta becoming production, get it right the first time if possible).

Whether prior, during or after a migration to cloud, new data protection capabilities are available and achievable. A driving principle to consider for the advancing data management and data protection capabilities is taking a proactive and automated approach to data protection. A foundation to drive data protection capabilities, is the use of metadata, such as through metadata data catalogs. But don't worry, it is not as technical as it may sound, and the past several years have brought to market plenty of user-friendly technologies built for business users and IT users to work together in easy-to-use applications.

While metadata underpins data protection and automation capabilities, it is just the magic behind the scenes. There are several approaches and advancing technologies that help make this easier for organizations to use such as data catalogs built into governance technologies, security applications and cloud platforms. Using a metadata-driven approach, organizations can embed policies and controls in human readable and machine-readable metadata that can be granularly applied to your sensitive data assets. To put that another way, think about a particularly sensitive data asset, let's call it "personal info id." And let's say you as an organization have a policy to restrict user, group, and service access to this data. The more traditional approach to data protection might simply try to protect this data by applying controls at the application level (e.g., via who has access to a report or a database table). But with a more metadata driven approach, we can consider applying that policy at the data asset and element level. So that "personal data" can have the proper controls and policies applied as a security attribute. Instead of setting controls in each application and solution, at multiple locations, what if you could set the policy once, tied to one's sensitive data asset, and let it travel with your data and enforce itself along the way? If "personal data" travels from on-premises database to one or more cloud environments, to multiple SaaS applications and reports, that is ok, the policies and controls can travel with the data (and

enforce along the way), and your data protection travels and becomes more consistent (and automated for your governance organization's sanity and peace of mind).

How traditional data protection differs from modern data protection

Traditional data protection is monolithic, meaning it is stored, managed, and consumed in one centralized location. A challenge for organizations moving to cloud is traditional or standard operating procedures for protection may not or not translate as well. As data scales exponentially and the boundary expands and blurs, data protection operating models will have to evolve. When undergoing a digital transformation to the cloud , let's take the example of how data becomes more decentralized.

The value of this more proactive approach is to meet the increasing speed and complexity of technology and data while systematizing adaptability. Regardless of where the data travels and the new applications an organization may introduce, the protection is embedded with the data asset(s). The key components to enabling metadata driven automation are data catalog technologies, curation by governance teams of scanned catalogs, sensitive data taxonomies and classification, and data lineage, along with rule-based and AI monitoring either through data governance technologies or other security and protection technologies.

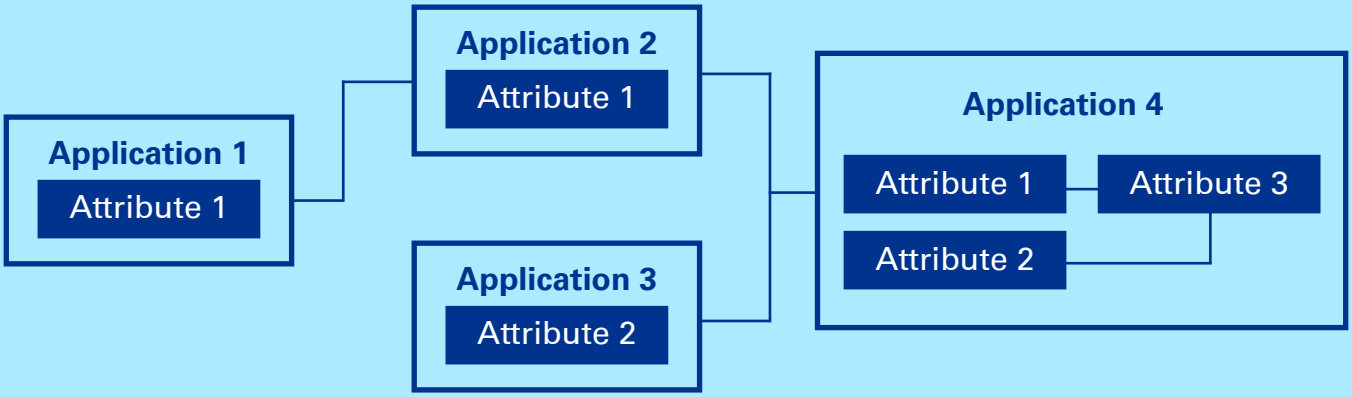
The future of data protection

The biggest challenge for cloud data protection today and tomorrow is how fast industries and the reliance on data are growing. The expanding perimeter and exponential increase in data volumes that need to be protected will keep professionals on their toes. The expanding volumes and distribution of data are hard to secure, while simultaneously pushing for open and accessible data. Data protection needs to evolve to be more agile to meet a changing technology landscape. There is a push and pull balance for democratizing data access for all, but how best to adapt this technology for data protection in a secure way? So, the question is, how do we insulate our approach to data protection to be more adaptable, flexible, and agile to meet continual technology changes? Since cloud is still a relatively new and growing space, data protection plays a crucial role in making or breaking effective migration to the cloud. That is, just on the technology front, external drivers such as regulatory requirements in Europe and the U.S. will challenge organizations to respond quickly and effectively when new rules are put in place, increasing the reliance around effective data management and data protection capabilities. How can a company meet evolving regulatory landscapes across North America, Europe, and Asia? And not simply meet but meet well and at speed. Organizations will face increasing overhead and demands to work fast and "prove it" when it comes to data protection. The ability to leverage metadata-driven technologies, with automation enabling the policy, enforcing and evidence, will save organizations and their employees many sleepless nights and lost weekends trying to respond to regulatory requests and actions.

For those operating across the globe and many regions, it must be taken into consideration when operating across jurisdictions and faced with mismatched regulatory standards. Not all standards are the same, but they must drive consistency wherever data travels, and can have the ability to meet the highest watermark needed for a particular standard, or the built-in flexibility to meet to be patchwork. There are country and industry specific standards that may meet similar and commonly accepted implementation approaches, or other standards may elevate requirements and may have vague or simply different requirements. If this sounds a bit confusing, it is because it is, and leaves organizations left to interpret standards, requirements and implementation approaches to protect and secure data across a number of regions, countries, and industries with similar, same or different patchwork standards and guidance. Then comes the question of how data protection will remain agile to meet existing and evolving patchwork regulatory guidance as well. Organizations will be forced to be nimble and be able to adapt quickly to guidance changes.

To make this a little more real and what your environment might look like, consider Google Cloud, which has many ways to protect customers data as well as their own, such as providing Identity and Access Management (IAM). Entitlements and access management, as well as metadata services like data catalog. Google Cloud also protects data by collaborating with other cloud technologies and third-party vendors for data protection platforms, applications, and services. Many Google Cloud resources are open source and offer a seamless collaborative experience with other providers. Google Cloud has data lineage capabilities as well which allows for traceability of data through its ecosystem for visibility to data flows, and application and consistency of policy in data flow. Data lineage is a data lifecycle that will tell you where data is originated from as well as any changes that are made to the data. Data lineage has many use cases in data protection and data governance. This is useful in identifying any data conflicts or enforcing data access control policies. It is a key variable in managing data access controls.

Example Data Lineage Lifecycle



In this graphic, it shows how data flows through its lifecycle. You will be able to see what application changes what attribute. This way, if there are any errors or an attribute is changed, you will be able to see everything. The lifecycle can also show if any applications are impacted by changing attributes.

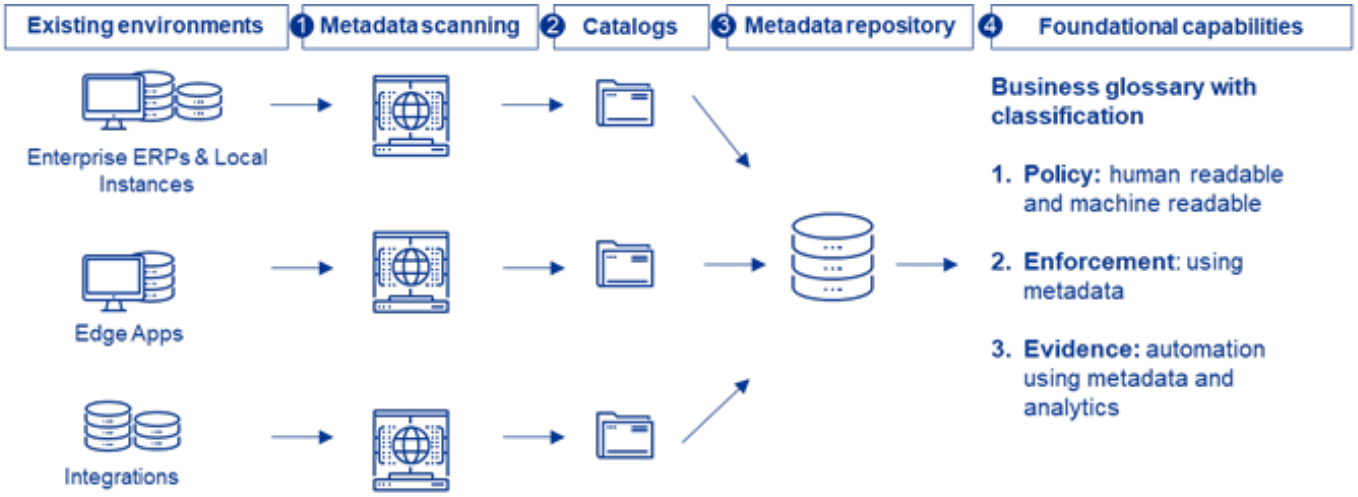
How metadata is used in data protection

Metadata is used to enhance data asset protection through automation using the metadata-driven attribute. Metadata can contain a number of descriptive attributes for a data asset and embed those attributes with the data asset. These attributes can capture specific policies, controls, sensitivity classifications, provenance, the owner, steward, creation date, or lifecycle rules among other things. With these attributes captured as metadata, this creates many attributes that can be used as tags, triggers, or functions to form the backbone of automated data protection. This metadata, much like an organization's data, can be operationalized to create dashboards, reporting, or analyzed to monitor, govern, and enhance how an organization approaches managing its data and protection. It is used to codify human readable policy and controls into machine readable policies and controls. The first step is to identify a data asset, then to associate the policy to this asset as a metadata attribute, which then enables further capability to automate the enforcement, and even further capability to automate the evidence and reporting of compliance.

Using metadata in the cloud, it enhances data asset protection through automation. When considering your data protection capabilities, it must follow regulations such as the GDPR, PCI-DSS, LGPD, BDSG, and China CSL depending on location. Just as companies have regulations on physical security such as needing a key card to enter a floor, the same applies to data assets. On top of these data regulations and standards, there are privacy standards that must be followed alongside them such as CCPA or CPRA.

There are many use cases for automated metadata-driven data protection. An example of this is metadata scanning operated via existing data management and/or catalog-specific scanning. Google Cloud also utilizes and protects metadata as well. By using data catalogs, cloud configuration, and data protection within data applications such as Big Query, Google Cloud employs metadata to personalize data protection for every customer. There is also an enforcement of data lifecycle to protect information across products and the Google Cloud ecosystem.

The flowchart below begins to conceptually show how catalogs, using metadata discovery, can create a single repository to enable policy, enforcement aka automation, and evidencing.



1. Metadata scanning operated via existing data management and/or catalog-specific scanning.

2. Catalogs represent technical metadata (scanned) and augmented by stewards for associated business glossary definitions.

3. Alignment of catalogs either via centralization or federation of catalogs provides a clear view of related metadata for lineage visualization.

4. Base capabilities enabled by a data catalog provides a business glossary containing single view of a data asset where policies can be defined in classifications for enforcement and evidencing.

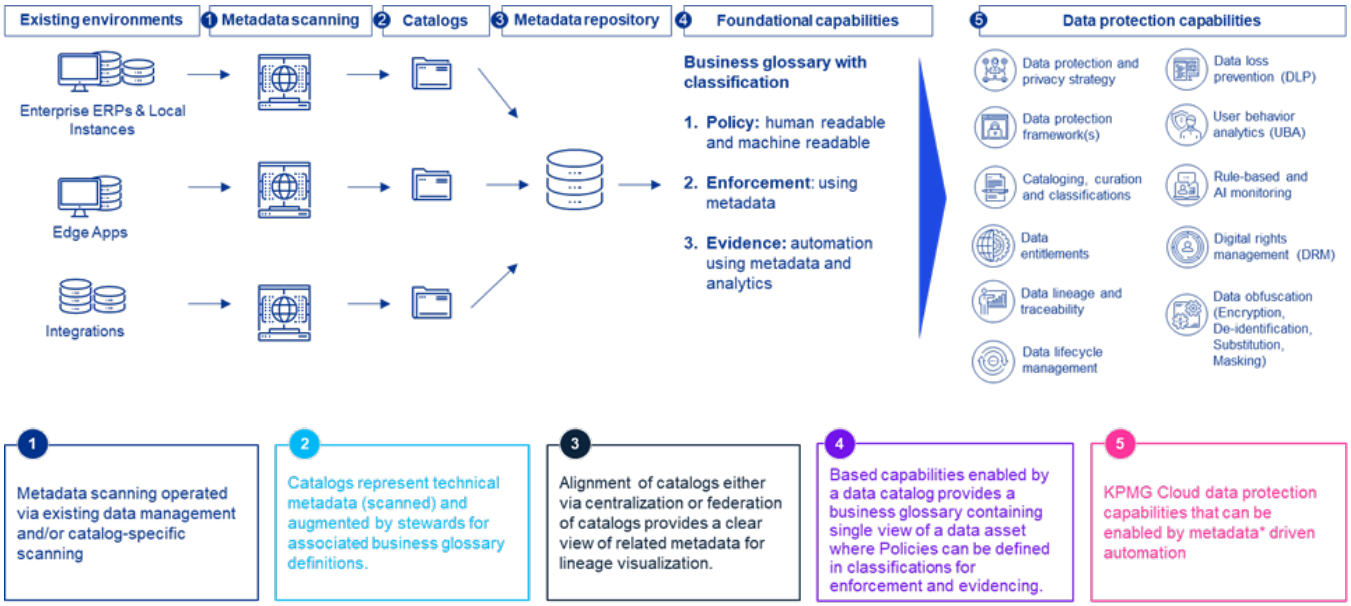
What is the art of the possible for automated data protection?

Using metadata scanning capabilities to discover technical metadata, an organization can then begin to curate the business glossaries as a part of a data catalog, that will allow for a more comprehensive analysis of the data that will later serve the basis for layering in classifications, and attributes to enable a number of data protection related capabilities. Active scanning also works with other applications such as mobile device management by storing data on the networks and preventing downloads to mobile devices. This also prevents the uploading of files to personal email addresses as well as printing sensitive information products and the Google Cloud ecosystem.

An example of data protection capabilities enabled with the suggested approach above is access and entitlements to data. Entitlements using User/Groups/Services are mapped to privileges and associated to data assets using a reference lookup table or service. Another example is data lifecycle management. This has several good data management and operational benefits, as well as protection benefits that limit the overall exposure and footprint of data available, automating availability based on rules instead of leaving data indefinitely or waiting for audits and batching of data for archival.

Entitlements and lifecycle management represent a couple of examples, but taking a bigger-picture view, see how the foundation laid using catalogs and metadata-driven attributes for automation can enable a number of other data protection related capabilities. A design that builds a strong foundation up-front, enables many downstream capabilities.

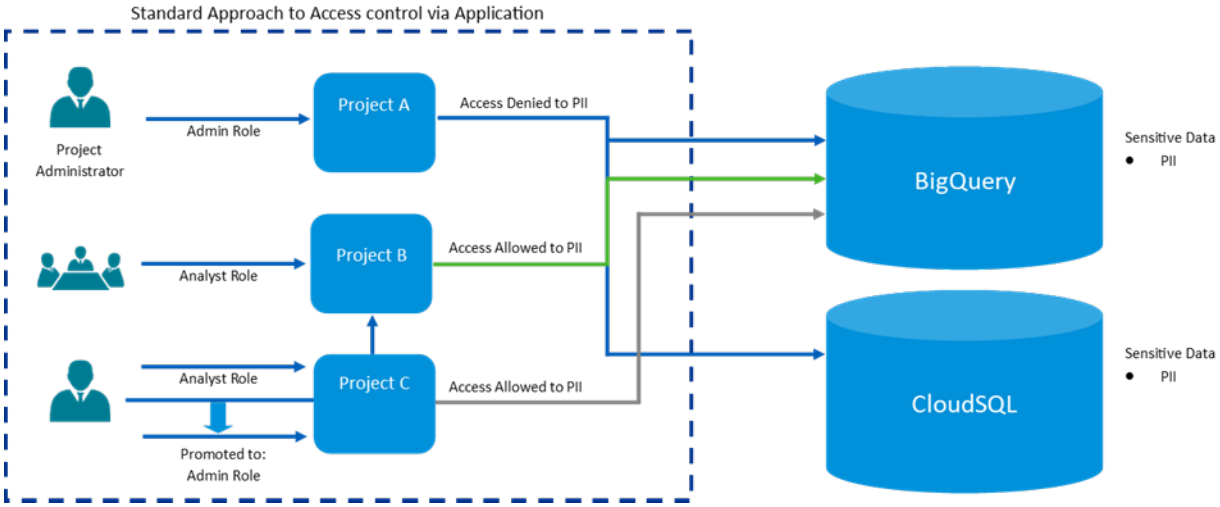
Use case for automated metadata for driven data protection



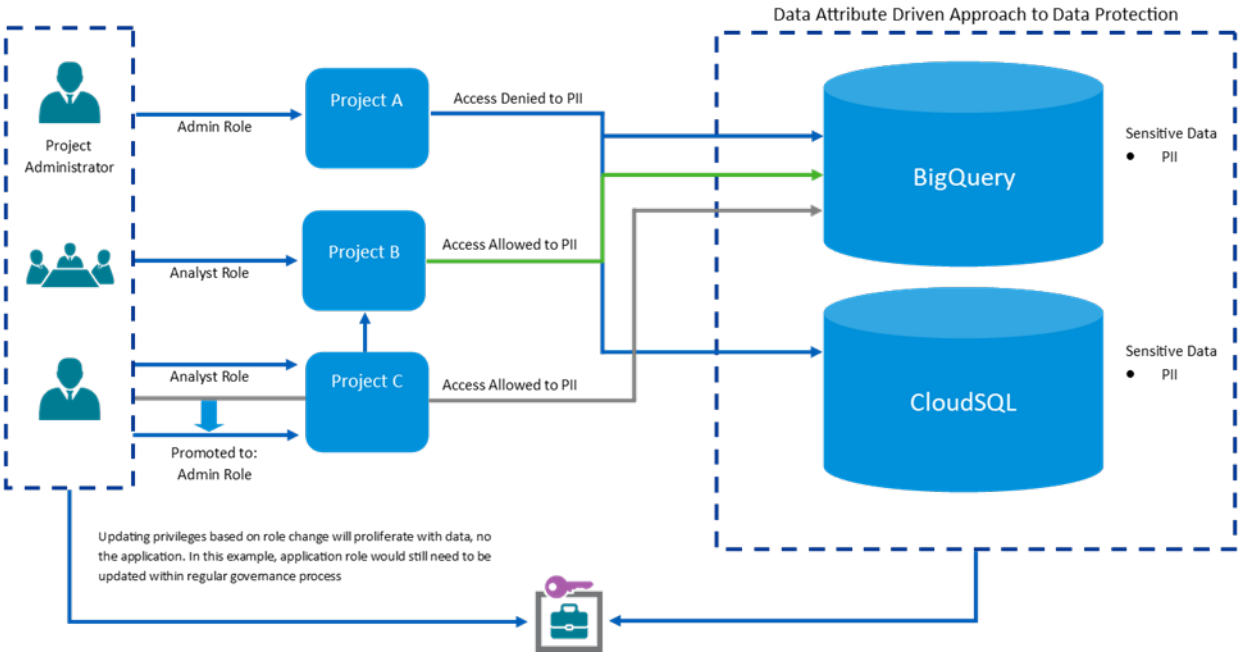
How data protection can be proactive

Proactive data protection is the act of preventing cyber attacks and incidents before they occur. It is achieved by moving away from a more traditional or legacy annual approach such as Excel-based tracking of policies as well as having role and attribute-based access controls. Traditionally, "wholesale" applications role-based access can be difficult to maintain in the long term since the volume of rules increases. Attribute-based access controls can be designed to travel with the data asset across ecosystems and applications for better consistency in policies and controls for data protection and privacy when associated to specific data assets.

Role-Based Access Controls



Attribute-Based Access Controls



Conclusion

Data protection will continue to be an ever-changing space, especially due to the technological advancements that are happening every day. As technology and data expand in volume and complexity, data protection will need to grow in sophistication and speed to mitigate risks in parallel. There will be an opportunity to learn new ways to protect data and make it more efficient in protecting customers. The future of data protection is metadata driven, automated, and proactive. As cloud adoption increases and emerging technologies become more prevalent in use, data protection will have to meet growing demand to protect data for customers and employees as well as meet an evolving technology and regulatory landscape. Companies must adapt and change now if they are to keep up with current trends and minimize any threats or vulnerabilities. Failure to do so could lead to tremendous issues in the future, and cost more than to initially transform.

Glossary

- CCPA: California Consumer Privacy Act
- CPRA: California Privacy Rights Act
- Data Democratization: This is the process of increasing accessibility of data so that the end user, no matter their technical background, can understand it.

Contact

Abhijeet Kulkarni
Managing Director, Advisory
Google Cloud Security Leader
KPMG LLP
T: 214-840-8889
E: abkulkarni@kpmg.com

Some or all of the services described herein may not be permissible for KPMG audit clients and their affiliates or related entities.

kpmg.com/socialmedia