

Cyera's Next Generation Data Security Platform

Data Security Posture Management
is Built for the Cloud Era

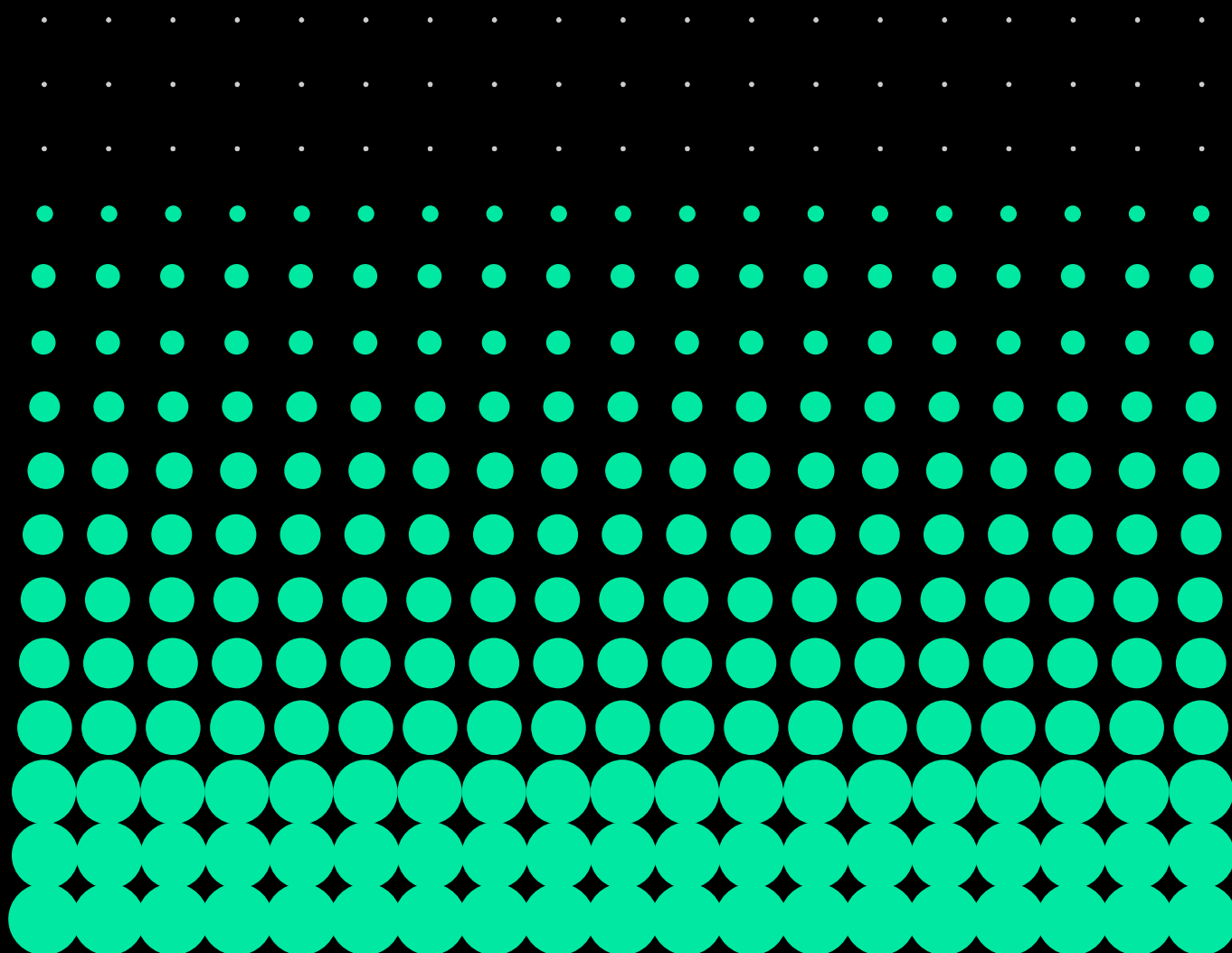


Table of Contents

Executive Summary	03
Modernizing data security for the cloud era	04
What makes data security so challenging?	06-08
Discovery and classification challenges	07
Data Loss Prevention (DLP) and Data Access Governance (DAG) challenges	07
Holistic Cloud Data Security with Cyera	09-19
Secure Deployment Architecture	10
Simple Agentless Connection Using a Read-only IAM Role	11
Dynamic Data Store Inventory Using Native APIs	12
Continuous Structured and Unstructured Data Discovery Across IaaS, PaaS, and SaaS	12
Automatic Sensitive Data Classification	14
Auto-Calibration and Contextual Data Enrichment	15
Multidimensional Exposure Assessment	16-19
Public Exposure	17
Encryption	17
Access	17
Data Sprawl	18
Logging	18
Backups	18
Automated Remediation Workflows	19
Summary	20-21
Data Discovery and Classification for the Cloud Era	20
Data Security Posture Management (DSPM)	20
Cloud DLP and Data Access Governance	21
About Cyera	21



Executive Summary

The world is experiencing a data revolution. In fact, 90% of the world's data was created in the past two years. This presents a very real challenge that legacy data security processes and tools were not designed to address. Manual processes, hardware- or software-centric deployments, and reactive responses to real-time exposures cannot keep pace with the dynamic, highly permissive nature of cloud environments. This leaves businesses exposed to increased amounts of risk as they embrace the cloud to innovate faster.

Security teams are struggling to keep pace with this data proliferation. For evidence of this, look no further than your news feed, where a new data breach, data leak, or ransomware attack has become a daily occurrence. Legacy data loss prevention, data access governance, and other siloed, highly manual solutions must be replaced with modern technology purpose-built to address these challenges so that security teams can create new processes around them to defend their data in the cloud era.

Cyera has introduced a revolutionary approach to data security, designed to manage data across the highly permissive, widely distributed, and massively scaled cloud landscape. We have taken a cloud-first approach to data security, but the novel approach to data discovery applies to on-prem data centers as well, which is important since nearly every business will maintain on-prem data for decades to come. Cyera's data security platform implements a non-invasive, fully automated data discovery scheme that maps an inventory of your enterprises' sensitive data and helps to put your data security objectives within reach.

Securing your organization's data is the foundation of any successful information security program. Cyera's multidimensional [data risk assessment](#) makes that process fast and easy for security teams. This paper will describe how the deep data context that Cyera develops identifies sensitive data exposure, helping you to prioritize only the most relevant and pressing issues stemming from real risks. Integrations with your existing tools and workflow processes enable you to streamline remediations with actionable guidance that significantly decreases mean time to resolution for critical security exposures.

Your security team needs a rapidly deployable, innovative solution that approaches data holistically, scales with your sprawling data landscape, and delivers value quickly and easily. Cyera is that solution.



Modernizing Data Protection For the Cloud Era

Businesses have moved boldly to adopt cloud services in an effort to better engage customers and improve the quality of service, as well as to create new revenue streams and opportunities for the business. According to Gartner, worldwide end-user spending on public cloud services is expected to reach nearly \$600 billion by 2023.¹

Today's business leaders don't care whether an application runs on premises or in the cloud. Investment decisions are defined by value delivery - and delivering that value quickly and sustainably is something that the cloud is uniquely positioned to enable. Cloud-first strategies are driving changes for businesses across every vertical, where leaders are driving a mindset shift away from the infrastructure teams manage to the services they provide.

"2022 Technology Spending Intentions Survey" highlights that more than 60% of organizations plan to increase their spending on both public cloud applications and infrastructure services in the coming year, and 44% of the survey respondents said their organization has a cloud-first policy on deploying new applications, unless there's a compelling reason to run them on premises.

Enterprise Strategy Group's (ESG)²

Despite this shift, a recent PWC report³ found that fewer than 40 percent of executives surveyed indicated they have fully mitigated the risks their bold moves incurred. Cybersecurity exposures and risks associated with cloud workloads are inherently different from those of legacy, on-prem infrastructure. Many of today's security teams are ill-equipped to quantify, understand, or mitigate the exposure that sprawling cloud environments represent.

A significant challenge stems from decades of siloed approaches to data, security policies, and the technologies designed to inform and enforce them. In the modern enterprise, data ownership and data stewardship responsibilities are fragmented. Different business units, departments, and disciplines ingest, create, and manipulate data to achieve their business goals. Each of these groups also procure solutions to help them manage their data. This distributed ownership and fragmented approach exposes businesses to significant risk — especially in light of growing regulatory pressures stemming from a seemingly unending stream of data breaches and ransomware attacks.

1. <https://www.gartner.com/en/newsroom/press-releases/2022-04-19-gartner-forecasts-worldwide-public-cloud-end-user-spending-to-reach-nearly-500-billion-in-2022>

2. <https://www.esg-global.com/2022-technology-spending>

3. <https://www.pwc.com/us/en/services/consulting/cybersecurity-risk-regulatory/library/global-digital-trust-insights.html>

In order to adapt, organizations must first identify and align on how the business defines “data” and the value that data represents to the business. For example, data can be source code, photos, algorithms, IoT sensor data, customer interactions, and more. Next, it is important to understand different stakeholders’ goals and motivations. This requires security practitioners to embrace their business counterparts’ need to use data, understand why and how they use it, and where that data comes from. Ultimately data has value when it is used. Understanding the journey data takes in your business environment will help you better identify the points in the journey, from access to use to the end of its useful lifecycle, where you can and should apply controls.

Modern security teams must recognize that they are operating within a system. This requires that they approach and think about controls and risk mitigation across the entire system. Data-centric security controls are only one part of that whole, and not all exposures represent the same level of risk or can have the same level of impact. To best partner with the business while mitigating risk, your security teams must organize their efforts to encompass:

- **Privacy strategy and data minimization efforts** on the front end where and when data collection happens
- **Security controls** that are applied to cloud infrastructure environments where data is processed and stored
- **Access controls** that you implement for your workforce, partners, and customers to maintain privacy, compliance, and the competitive advantage that comes from your intellectual property

The understanding of what data is sensitive, where it is located, and how employees need to access and use that data is critical to determine how you will prioritize your efforts. This includes identifying the appropriate controls to focus on based on which will mitigate the most significant risks, and then prioritizing which are the most pressing to implement based on the business’s goals.

The good news is that more than 70% of senior executives recognized improvements in cybersecurity last year — thanks to increased investments and greater collaboration from the C-suite.⁴ The introduction of cloud-native posture management, vulnerability mitigation, and user access has improved visibility and awareness of infrastructure, and some software-related threats. These outcomes can improve with a data-aware security architecture and approach. Data represents the most valuable commodity for a business, just as it represents the most significant insider risk, and the most valuable target for external threat actors. For there to be a real and lasting improvement in security outcomes, data security must be modernized for the cloud era. That will truly unlock the power of data for the enterprise.

That is Cyera’s vision, for every business to realize the full potential of data — collaboration, connection with customers, insight that fuels innovation — to power a new era of development, growth, and productivity.

4. <https://www.pwc.com/us/en/services/consulting/cybersecurity-risk-regulatory/library/global-digital-trust-insights.html>



What Makes Data Protection So Challenging?

The world is experiencing a data revolution. 90% of the world's data was created in the past two years. Modern businesses are creating and consuming data at an incredible pace, leveraging cloud technology to take advantage of the speed and agility it offers their teams to create new business opportunities and unlock the potential of customer engagements. This promised acceleration is why over 94% of enterprises are using cloud technologies today,⁵ with most leveraging multiple providers to enable their teams to maximize productivity, efficiency, and impact.



94% of enterprises are using cloud technologies today

The challenge is that legacy processes and tools were not designed for the permissiveness, and unrelenting pace of change that cloud technologies have introduced. Manual processes, hardware- or software-centric deployments, and reactive responses to real-time exposures cannot keep pace. This leaves businesses exposed to increased risk as they embrace the cloud to innovate faster and outpace their competitors.

Data security as a discipline needs to evolve to overcome the challenges that the cloud era has introduced. For evidence of this, look no further than your daily news feed, where a new data breach, data leak, or ransomware attack has become a daily occurrence. According to IBM, the average total cost of a data breach reached 4.35 million USD in 2022 (9.44 million USD in the US, where breaches were most costly), a nearly 13% increase from 2020. Cloud-based breaches accounted for 45% , and 83% of organizations studied had more than one data breach. Despite the attention paid to ransomware attacks in the media, the most common cause of a breach remains lost or stolen credentials.

Any security incident where one party gains unauthorized access to another party's information is a data breach. Recognizing that data security is just as relevant for managing insider risk as it is for mitigating external threats is an important foundation for addressing data security completely and correctly. Also foundational is the awareness of what data you manage, where it is located, and who has access to it. In fact, it is a precursor for most data security technologies to function correctly. However, for many tools humans have to do the discovery; you have to know where your data is, and then let the tool “discover” it by making the connection, often by deploying an agent and/or configuring a direct connection to the data store.

Proactive measures that aim to prevent data breaches are the most effective means of defense. They are also less expensive than responding after the fact. Aligning with business stakeholders on how you can remain in control of the data you process, and letting go of data you no longer need is not just important to remain in compliance, it helps to overcome many of the shortcomings of legacy data security tools and processes.

5.[chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://resources.flexera.com/web/media/documents/rightscale-2019-state-of-the-cloud-report-from-flexera.pdf](https://resources.flexera.com/web/media/documents/rightscale-2019-state-of-the-cloud-report-from-flexera.pdf)

Discovery and Classification Challenges

Legacy discovery and classification approaches were not designed for today's fast-moving, permissive multi-cloud environments. Rather, they were designed and built to manage the slow-changing legacy environments of the past. These solutions rely on antiquated deployment modes that include host-based agents, manual connection strings, and a need for every data owner to manually track what data exists, where it is stored, and how it should be classified in order for security teams to apply the appropriate controls.

Many tools claim to offer data discovery and classification capabilities, but the definition of those terms vary wildly. Legacy security technologies that claim to discover your data actually require you to do the discovery yourself. This can take the form of attestation, polling, or surveys, but in the end you are responsible for knowing where the data resides, and then "pointing" the technology at the data store (sometimes by deploying an agent, nearly always by maintaining connection strings). Classification is similar. Nearly every solution requires manual tagging, cumbersome data definitions, and regular expression tuning to function properly. With distributed ownership of the data, aligning business stakeholders who understand what your data represents with technical experts who can codify this with a regular expression is incredibly difficult. It is also extremely expensive. In Cyera's customer engagements, one financial services organization estimated the cost of their manual discovery and classification efforts to be \$200,000 USD per week.

Even newer cloud security posture management (CSPM) and SaaS security posture management (SSPM) tools are challenged to understand data, since they employ an inventory-centric approach to discovering data stores, and follow the same outdated manual for regular expression-based classification processes. Data requires a specific type and level of security based on the value it represents to the business, and that starts with a deep understanding of the data itself.

Data Loss Prevention (DLP) and Data Access Governance (DAG) challenges

Enforcing policy for data loss prevention and data access in the cloud is incredibly challenging. This is due to several factors, from the proliferation of environments and tools to distributed ownership and governance, to an explosion of tools that assess risk and apply controls differently. But the foundational problem remains a manual effort from multiple stakeholders that have very different objectives and goals. The business units that know why they manage the data that their applications, products, and partnerships rely on have different goals than the governance, risk and compliance teams who aim to manage risk for the business, and their goals are different from the security teams tasked with implementing controls.

Consider a simple and very common example: collecting customer data on a website. Modern websites are built on microservice architectures that implement an array of third-party components. Multiple stakeholders in a business influence the components that are added to a site - chat, personalization, social elements, various types of analytics, and more - and developers are constantly adding new functionality to attract and retain customers.

All of these efforts can create sensitive data exposures, whether from compromised supply chain components, over-sharing with third parties, managing data across regional boundaries, or storing customer data in a new data store that is not as secure as it should be. Legacy security solutions do not provide any utility to monitor and alert about these issues because they require you to inform them of all of this complexity and define policies before they can take action.

This reality led one former CISO to ruefully comment:

**“As security practitioners, we wanted data loss prevention,
but all the vendors gave us was DLP.”**

Industry Analyst and Former CISO

Legacy data loss prevention (DLP) solutions are broadly recognized as necessary evils — hugely expensive, complex products that take months, sometimes years, to architect, deploy, configure, and train. And for all of the investment, they ultimately produce too many false positives to ignore, resulting in noisy alerting mechanisms that create friction within the business.

Data access governance solutions face similar challenges. Provisioning and governing access has become a distributed discipline. Today, developers, marketing teams, and data analysts all have the ability to create users and provision access to sensitive data. Developers do it to deliver value faster via the applications they develop. Marketing teams do it to improve the user experience and maximize conversions and retention. And data analysts do it to mine the data a business manages in order to uncover and share insights that will help the business grow. None of these activities are nefarious or ill-intentioned, but all of them introduce risk. The [SolarWinds](#) hack is an example of supply chain compromise. [Sephora's recent CCPA fine](#) is an example of inadvertent misuse by a marketing team. And one Cyera customer highlighted that one of the hardest platforms for them to govern access to is Snowflake due to business stakeholders clamoring for the valuable insights the data analysts are exposing through their analyses.



Holistic Cloud Data Protection with Cyera

Cyera has introduced a revolutionary approach to data security, designed to manage data across the highly permissive, widely distributed, and massively scaled cloud landscape. At Cyera, we have taken a cloud-first approach to data security, but the novel approach to data discovery applies to on-prem data centers as well, which is important since nearly every business will maintain on-prem data for decades to come. Cyera has pioneered a non-invasive, fully automated data discovery scheme that maps an inventory of your enterprises' sensitive data and helps to put your data security objectives within reach.

The agile nature of the cloud makes it at best challenging, and at worst impossible for any business to maintain an up-to-date inventory of their cloud assets, including the data stores deployed across their data landscape. For this reason, Cyera has architected a fully automated process for continuously discovering data stores and data. The process focuses on leveraging native APIs to create and maintain a dynamic data store inventory in order to eliminate the effort and process overhead inherent in manual IT service catalog creation, and the reliance on agents being deployed to infrastructure environments. No agents, network footprint, or hardware are required. Our cloud-first approach allows us to leverage native APIs to discover every data store in the organization or account while remaining completely out-of-band. This means no performance overhead, no impact to data processing, and no ongoing maintenance.

Many discovery and classification tools take a very narrow view of environments, structured vs unstructured data, and the data stores that they support. Therefore, most businesses rely on multiple tools to discover specific data store types, each of which uses a different approach and manages the data separately. In addition, the typical means of discovering data stores include manual attestation, surveys, or agent-based scanning, all of which are very time-consuming, costly, and fraught with errors. They also represent a point in time, which is at odds with the rate of change introduced by cloud technology. Cyera addresses these issues with a holistic approach to discovery that minimizes human involvement, works across the cloud data landscape, and dynamically discovers new, changed, or eliminated data stores.



Note: Simplified Scenarios Represented

To simplify this document, all scenarios and explanations use AWS cloud terminology throughout. Please note that similar concepts apply to both Microsoft Azure and GCP cloud environments.

Secure Deployment Architecture

Cyera's cloud-native platform delivers highly scalable and highly available services, with security built in as a first principle. The platform consists of two services, the Data Analysis Service and Data Insights Service.

Data Analysis Service - this security-hardened Kubernetes cluster is Cyera's primary data discovery and classification service. It connects to the customer environment using cloud-native APIs and captures metadata and any identities that can be read or written to each data store. When required, the Data Analysis Service reads data store snapshots to perform data classification. Cyera will use existing snapshots, or create new snapshots if needed. If Cyera creates a snapshot, it is deleted once the analysis is complete. Cyera will use a direct read API for data stores that do not have a snapshot mechanism. The service can be deployed in a full-SaaS model inside of Cyera's cloud tenant or, if the customer wants to limit the information transferred outside of their cloud environment, as an 'outpost'. In both cases, customer data never crosses regions; all data analysis occurs in the same region where the data was originally discovered. All customer data analysis is performed inside the Data Analysis Service, then deleted immediately. For outpost deployments, communication with the data insights service uses a secure private link to Cyera's environment.

Data Insights Service - this service is hosted in Amazon AWS and is managed and operated by Cyera's cloud engineering team. This service hosts the web application that customers use to interact with the results of Cyera's discovery, classification, and risk assessment processes. Data insights are accessible via a web browser or programmatically via secure API endpoints.

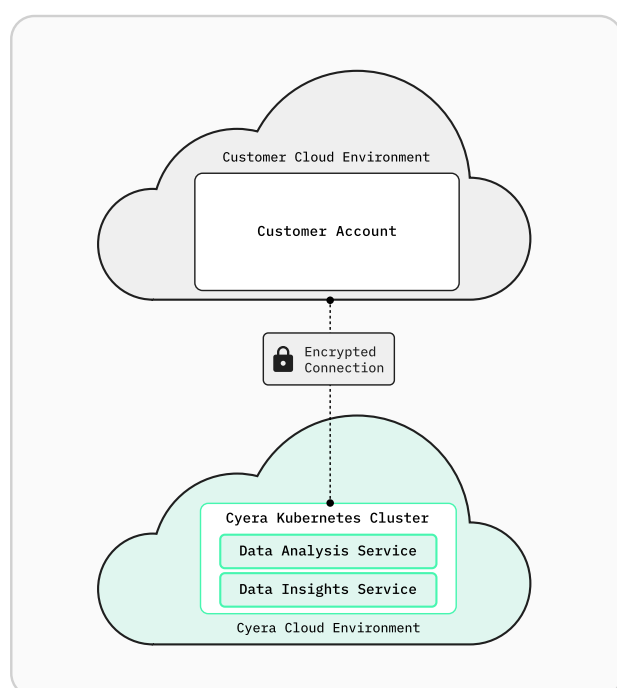


Figure 1: Full SaaS Deployment

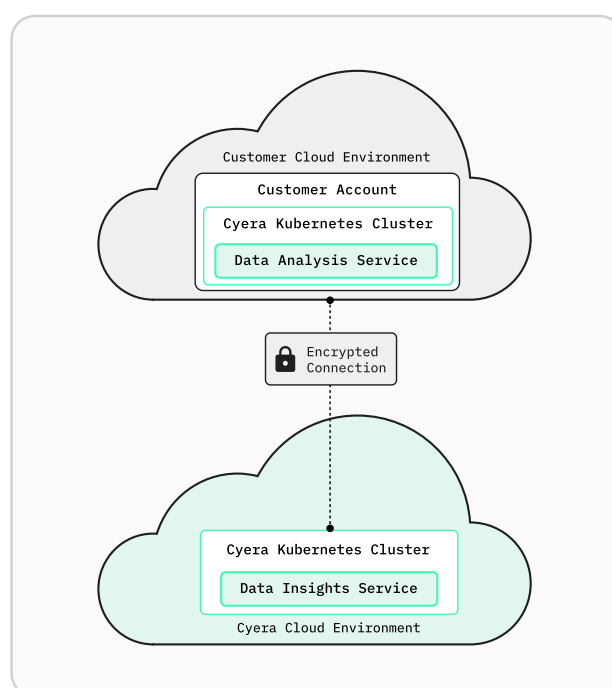


Figure 2: : 'Outpost' Deployment Model

Simple Agentless Connection Using a Read-only IAM Role

Cyera has architected our platform to maximize time to value, starting with the way we connect to an environment. Security teams connect Cyera to a cloud organization or account using a single identity and access management (IAM) role with a “cross-account assume role” trust policy. This is achieved by creating a CloudFormation stack in the relevant accounts. This stack grants Cyera’s platform temporary read-only access to data stores, environment logs and monitoring infrastructure, and the posture analysis services needed to discover and analyze a company’s data.

The role deployment can be scripted and executed by infrastructure as code tools like Terraform connect Cyera to new cloud accounts when they are created. This further accelerates time to value for organizations and ensures that a comprehensive data store inventory can be maintained. Because Cyera uses read-only access credentials, the platform cannot make any changes to your workloads or data stores, and won’t make any changes to or impact your runtime environment.

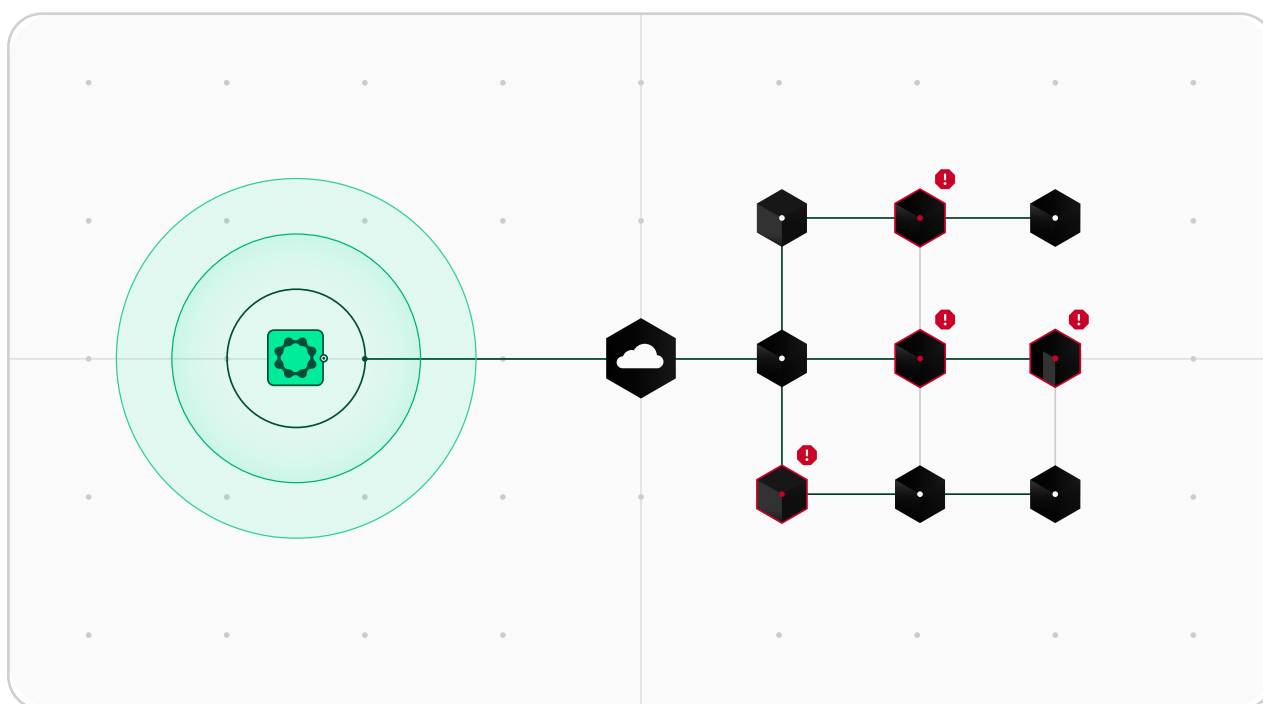


Figure 3: Cyera's Agentless Model

Dynamic Data Store Inventory Using Native APIs

Once Cyera's IAM role is deployed, native APIs are leveraged to ensure that we can discover and classify data across your entire cloud estate without impacting workload performance. The platform optimizes the scans of both structured and unstructured data to ensure maximum performance without compromising classification fidelity. This allows Cyera to dynamically generate a full data store inventory in minutes - including structured and unstructured data in buckets, blob storage, unmanaged, semi-managed, or DBaaS data stores, datalakes, and other data platforms.

Cyera's discovery process takes into account the environment that is being scanned and will optimize the discovery process to limit the time and compute power required to develop a data store inventory. This method affords Cyera the added benefit of detecting inactive (what we call "ghost") data stores. Ghost data stores represent data that is present in one or more snapshots for which no live data stores remain. This type of data can represent ransomware resilience and privacy risks, since the lack of a live data store suggests that there is no longer a business need to manage the data, and therefore security teams are likely no longer prioritizing security controls for the snapshots in the way that a live data store would warrant.

Continuous Structured and Unstructured Data Discovery Across IaaS, PaaS, and SaaS

Cyera has taken a holistic and cloud-first approach to data security. We support structured and unstructured data across buckets, object stores, unmanaged, semi-managed, and DBaaS structured databases, as well as data lakes and data management platforms. Because Cyera's cloud-native approach uses existing APIs, we can easily identify data stores installed inside virtual machines (i.e. EC2 instances). In fact, Cyera supports identifying data in EBS volumes and EBS snapshots, which are attached to an EC2 instance. This allows the platform to find unattached volumes and orphaned snapshots, which we identify as ghost data stores.

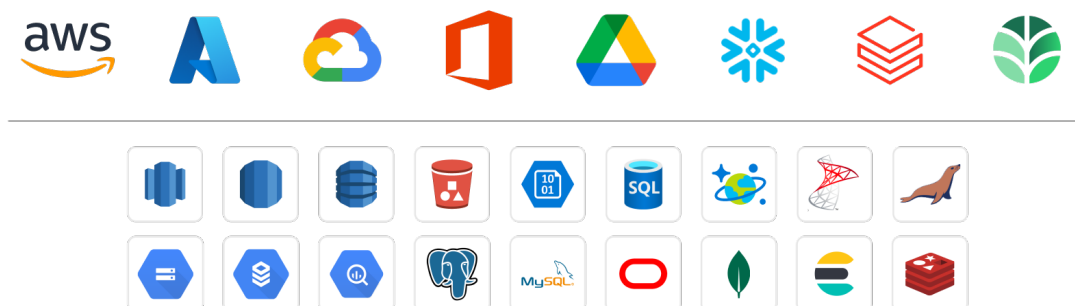


Figure 4: A sample of Cyera's growing list of supported data stores

Cyera's discovery process scans a wide array of unstructured files and formats and structured data stores. These include:

Unstructured Data:

- **Objects in Bucket Storage** - S3 Buckets are scanned using the native cloud provider API and are clustered based on metadata. After identifying potentially sensitive files, either these files, or parts of the files, will be copied for further analysis.
- **File types in bucket storage, or in SaaS environments:**
 - **Office Documents** - Documents, spreadsheets, presentations, including files with the extensions: DOCX, DOC, ODP, POTX, PPTX, XLSX, XLSB, XLS, XLSM, XLTX, XLT, OTS, ODS
 - **PDF files**
 - **Image files (OCR)** - various common formats e.g., JPEG, PNG, PDF.
 - **Plain-text files** - log and txt (and textual data in any file format)
 - **Security keys and certificates** - PEM, PGP, PPK, SSH, PW

Structured Data:

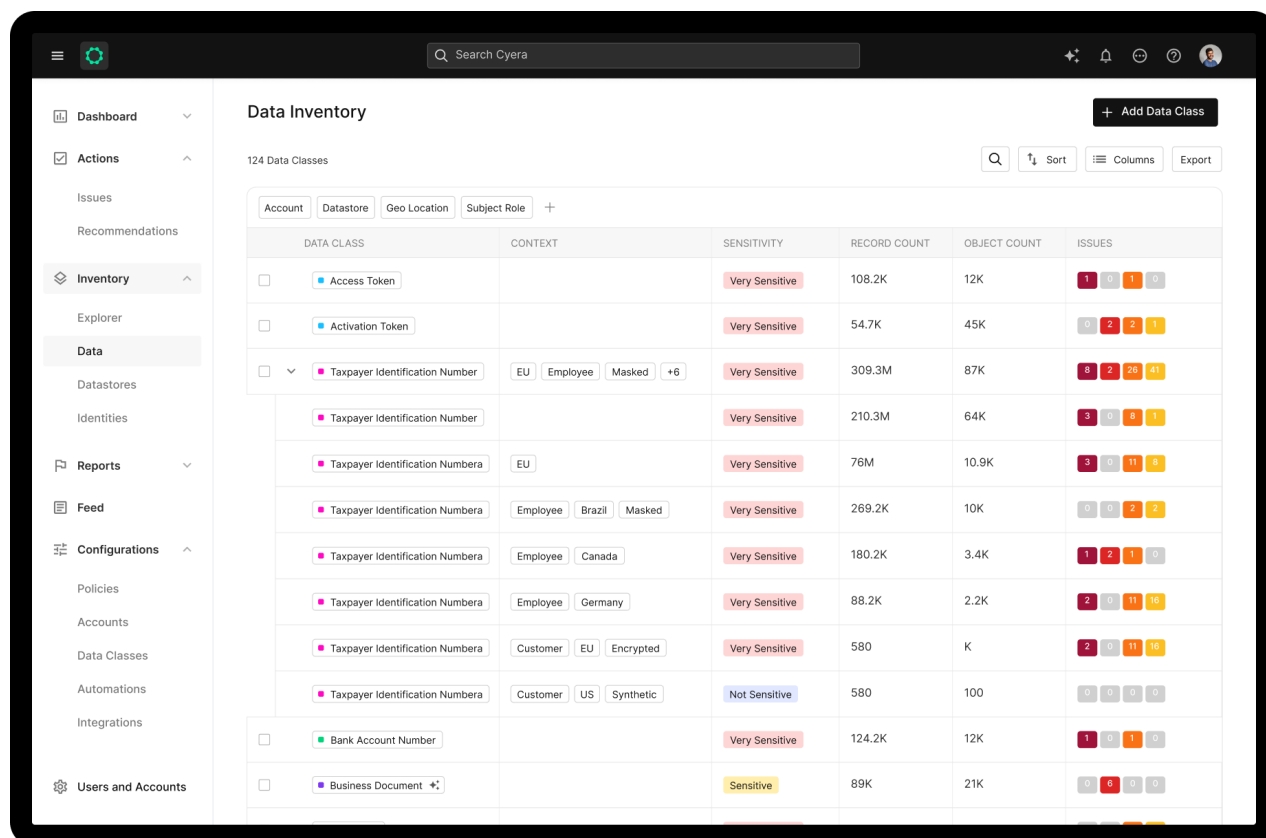
- **Objects in Bucket Storage** - S3 Buckets are scanned using the native cloud provider API and are clustered based on metadata. After identifying potentially sensitive files, either these files, or parts of the files, will be copied for further analysis.
- **Objects in Managed Databases** - when analyzing semi-managed or managed structured data stores (like Amazon RDS), or unmanaged ones (like database engines installed over EBS volumes - EC2's virtual disks), the system uses a backup of the DB instance to analyze its data. Although automatic backups are common in production databases, if a backup doesn't exist for a certain data store, one is created, analyzed, and deleted immediately post-analysis.
- **To detect and analyze potential data stores on EBS volumes without replicating the data**, Cyera leverages the AWS EBS direct API to access only a few sectors of the disk, then determines whether it contains database files and reads only the relevant portions of the disk.
- **Structured text files** - CSV, TSV, JSON
- **Big-data files** - Parquet, ORC, Avro, DTO
- Custom parsers can be added to support proprietary file formats

Automatic Sensitive Data Classification

Once Cyera's discovery process establishes a data store inventory, the next step is to identify the type of data each data store contains and its sensitivity. Cyera's discovery process reads the data present in each snapshot, which is stored in the proprietary format used by each data store and its specific version. Doing so requires that each snapshot be unencrypted, restored in an isolated environment, and attached to an adequate virtual machine used for scanning volumes (also known as a scanner virtual machine). Cyera mounts the volume to the virtual machine, and provisions the specific type and version of the data store engine from the initially scanned volume or block to read the data.

The data classification process reads the file or data store metadata or schema and takes random samples of the data. For unstructured data, Cyera will sample data from each file in the store.

For structured data, the database list, tables, and columns will be identified, and each table's data will be sampled to classify and detect sensitive data classes. Data is categorized as Personal, Financial, Health-related, Business & IP, and IT & Security. Data classes and a sensitivity level are assigned to each data element discovered. Cyera comes out of the box with hundreds of built-in data classifications, and the platform automatically discovers customer and environment-specific data types using a novel approach that draws context from the environment, data store, and the data itself.



The screenshot displays the 'Data Inventory' page in the Cyera interface. The page title is 'Data Inventory' with a subtitle '124 Data Classes'. A search bar and 'Add Data Class' button are at the top right. A sidebar on the left contains navigation links: Dashboard, Actions, Issues, Recommendations, Inventory (selected), Explorer, Data, Datastores, Identities, Reports, Feed, Configurations, Policies, Accounts, Data Classes, Automations, Integrations, and Users and Accounts. The main table has columns: DATA CLASS, CONTEXT, SENSITIVITY, RECORD COUNT, OBJECT COUNT, and ISSUES. The table lists various data classes like 'Access Token', 'Activation Token', and multiple instances of 'Taxpayer Identification Number' with different contexts (e.g., EU, Employee, Masked, Brazil, Canada, Germany, Customer, US, Synthetic). Sensitivity levels are marked as 'Very Sensitive' or 'Not Sensitive'. The 'ISSUES' column shows counts for different issue types.

DATA CLASS	CONTEXT	SENSITIVITY	RECORD COUNT	OBJECT COUNT	ISSUES
Access Token		Very Sensitive	108.2K	12K	1 0 1 0
Activation Token		Very Sensitive	54.7K	45K	0 2 2 11
Taxpayer Identification Number	EU Employee Masked +6	Very Sensitive	309.3M	87K	8 2 28 41
Taxpayer Identification Number		Very Sensitive	210.3M	64K	3 0 8 11
Taxpayer Identification Number	EU	Very Sensitive	76M	10.9K	3 0 11 6
Taxpayer Identification Number	Employee Brazil Masked	Very Sensitive	269.2K	10K	0 0 2 2
Taxpayer Identification Number	Employee Canada	Very Sensitive	180.2K	3.4K	1 2 1 0
Taxpayer Identification Number	Employee Germany	Very Sensitive	88.2K	2.2K	2 0 11 15
Taxpayer Identification Number	Customer EU Encrypted	Very Sensitive	580	K	2 0 11 16
Taxpayer Identification Number	Customer US Synthetic	Not Sensitive	580	100	0 0 0 0
Bank Account Number		Very Sensitive	124.2K	12K	1 0 1 0
Business Document		Sensitive	89K	21K	0 6 0 0

Figure 5: Cyera's Data Inventory Page

To reduce the volume of data we analyze and apply classification to data faster, Cyera uses multiple innovative techniques that accelerate the process without compromising the fidelity of our results. Since the platform has visibility into the entire data plane — across IaaS, PaaS, and SaaS — Cyera can cross-reference information among environments, recognizing previously analyzed and classified data with no impact on accuracy, but significantly accelerating time to value. In addition, the platform uses ML and NLP models to cluster similar data objects, which allows Cyera to analyze huge amounts of data efficiently without scanning all data or every object.

Auto-Calibration and Contextual Data Enrichment

Every environment that Cyera analyzes is unique. Businesses have unique data classes and proprietary data formats. Typically some form of data tagging has been done to categorize data as sensitive, confidential, or similar. To reduce false positives, the data classification process leverages patent-pending technology that uses multidimensional correlation. Cyera's platform combines pre-defined data classes (that were trained using traditional mechanisms including regular expressions and pattern-matching algorithms), with environment-specific analyses conducted by novel ML and NLP technologies, to reach a very high degree of accuracy. **The platform learns a customer's unique data and improves its accuracy with each additional account and subsequent scan, due to the increasing volume and variety of data available to the correlation engines.** The ultimate result is a capability that is similar to Exact Data Matching, but automatic.

Most businesses follow best practices and implement data security schemes that include encryption, hashing, and tokenization. Cyera identifies various types of encrypted, masked, and tokenized data out of the box. Our ML algorithms include treatment for names, addresses, email addresses, financial data, passwords and other secrets, and more. While some data classes cannot be identified as masked or tokenized without prior knowledge, such as client IDs, Cyera can integrate with data tokenization platforms to find objects that should have been tokenized and verify them.

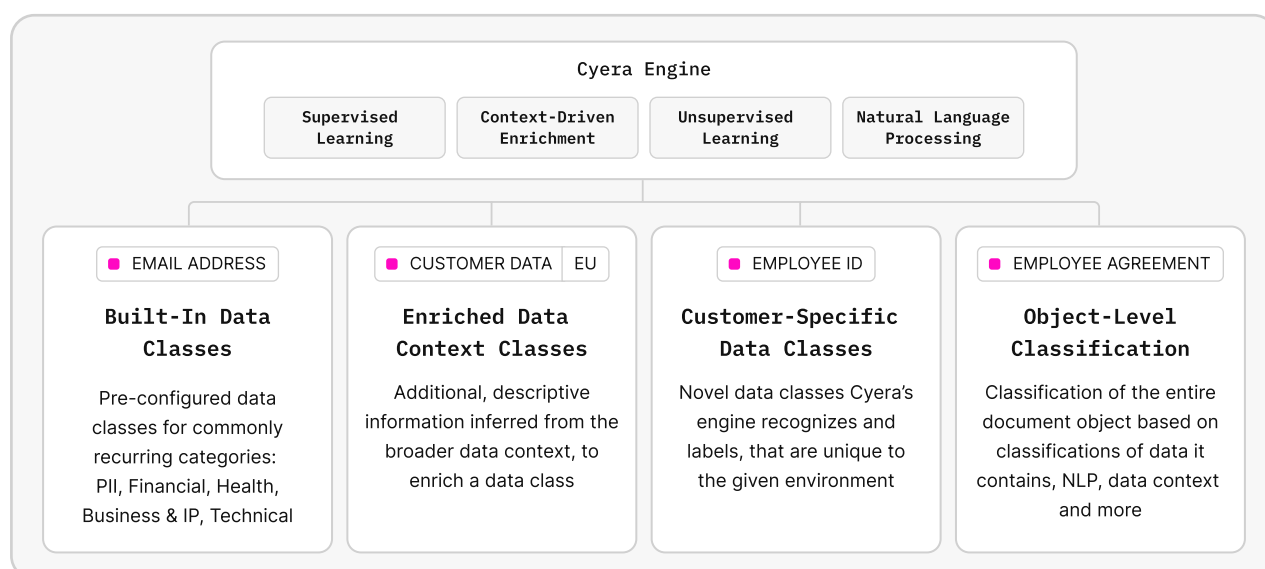


Figure 6: Cyera's Classification Engine

To provide the best insight into and ability to secure sensitive data, Cyera has developed a data classification scheme that merges the data type and data context. Cyera creates data classes as follows:

- **Data Subject Role** - for example, customer, employee, patient, etc.
- **Data Subject Residency** - the geo-location of the data subject, for example; "US", "EU", "California", etc.
- **Data-level Encryption** - is the data "plain", hashed, masked, tokenized, truncated, etc.
- **Identifiable Data** - can the data identify an individual
- **Synthetic Data** - is the data real or synthetic

This contextualized data classification is important in order to apply security policies appropriately. It also helps security teams better partner with their business counterparts because Cyera's data classes help to eliminate false positives that lead to noisy alerts. For example, if customer data is identified in a non-production environment, it is important to understand if that data is real or synthetic before creating a ticket regarding sensitive data in a non-production environment.

Multidimensional Exposure Assessment

Once data has been classified and contextualized, Cyera's Data Insights Service will assess exposure and risk. This is done by conducting a multidimensional data risk assessment and reviewing the sensitive data that is stored, accessed, and managed in the environments to which Cyera has been connected. The platform takes into account the environment, data store, sensitive data, and user access, and evaluates the exposure present against a set of security exposure, compliance, and risk assessment policy frameworks.

Cyera's exposure assessment identifies risks and validates security posture gaps against an established set of security and regulatory frameworks. When the policy engine identifies an exposure, it generates a prioritized issue based on the risk it poses to your business, and includes the context of the exposure and potential blast radius in terms of data volume, sensitive data type, user and role access, and the associated framework.

To make the exposures Cyera highlights as actionable as possible, the platform includes remediation guidance to resolve the exposure, and integrations with common toolchain technologies to automate the remediation workflow. This ensures that security teams can take action quickly and easily, and helps them to scale by eliminating the typical product training and industry expertise requirements that legacy tools require to take action.

To accommodate the unique needs of a customer's environment or business, and to make use of any existing policies they may have defined for an existing DLP or DAG initiative, **Cyera allows businesses to extend the built-in policies we deliver with user-defined custom policies.**

This includes incorporating specific guidance for business processes, accounting for a business unit or unique environment or product requirements and allowing for automated workflow remediation to resolve issues fast.

Cyera comes with built-in policies that focus on the following categories:



Public Exposure



Access



Logging



Encryption



Data Sprawl



Backup

| Public Exposure

Misconfigurations and relaxed permission settings routinely lead to publicly exposed sensitive data. While cloud solution providers and software vendors put tools in place to apply default controls that can mitigate this type of exposure, it is an all too common occurrence in cloud environments. Cyera's public exposure policies detect when sensitive, personal, financial, or health-related data is exposed to the internet, including when weak or default passwords that are easily compromised are in place.

- Show you who has access to your sensitive data
- Public exposure - public data, data shared with third parties
- Overly permissive data - salary document shared with the entire company



| Encryption

Encryption is a foundational means of safeguarding against unauthorized or unlawful processing of personal data and is one way that a business can demonstrate compliance by securing sensitive data. Cyera's policy engine, coupled with the contextualized data classes that the Data Analysis Services produces, highlights when critical vulnerabilities, including credentials stored in plain text or when PCI, HIPAA, or GLBA data are stored unencrypted in an environment. Because Cyera's data classes recognize hashed, tokenized, or synthetic data, noisy alerts for unencrypted data that does not constitute a sensitive data exposure are not raised, which allows security teams to stay focused on protecting the most sensitive data at risk.



| Access

Data security policies cannot prevent data loss if users are granted unrestricted access to data by default. It is critical that users are granted access to data using the least privilege model. However this can be challenging when data is not properly classified or tagged, or the business does not have a means of understanding what data is stored in a given folder, bucket, or database before granting access to it. Cyera's access policies are designed to enable DAG programs. DAG ensures users only have access to data when there is a legitimate business need. Data classification can help shed light on what data is where and who has access to it. This allows the IAM teams to quickly identify and clean up permissions that were not being strictly enforced.

| Data Sprawl

The proliferation in the number and different kinds of data that a business creates, collects, stores, shares, and analyzes has led to improved customer engagement, new business opportunities, and deeper insights into business operations and performance. But it also creates a significant challenge for security teams tasked with cyber resilience, ransomware resilience, and privacy.

Cyera's data sprawl policies highlight when data exists in a location that violates policies for security, risk, or compliance, especially when data security or privacy controls have not been applied correctly. These policies can address issues like:

- **Privacy violations** when data is stored in an unapproved country, or sensitive data exists in a non-production environment
- **Compliance violations** when financial, PCI (transaction), or PHI (health) data exists outside of a business's defined and protected zone, or when a specific type of data violates a compliance statute
- **Security & risk violations** when sensitive data has drifted to an environment with overly permissive access permissions that would constitute a compliance violation

| Logging

Logging is a critical component in establishing a proactive approach to security. Without effective logging, security teams cannot write and execute detections that are critical in hardening their environments, detecting threats, and complying with regulations. Security and IT Risk teams are forced to balance detailed logging - to enable monitoring, alerting, and forensic reviews - with cost - because robust logging can quickly become expensive in cloud environments. Cyera's policy engine helps teams maximize their data security posture while optimizing cloud costs by highlighting where their most sensitive data exists and when logging configurations expose them to undue risk. For example, the policy engine will highlight when administrative actions are not being logged, or access and changes to sensitive data will not be captured given the current logging configuration. The deep data classification and contextualization Cyera provides also enables security teams to be highly targeted in where they enable object-level logging for detection and response use cases where near-real-time to changes is required.

| Backups

Protecting sensitive data across multi-cloud environments includes ensuring that sensitive data store backup policies are managed correctly. This is an integral part of preparing for ransomware attacks, business continuity planning, and compliance with privacy, risk, and other compliance frameworks. Backup policies focus on whether sensitive data and audit logs have backups configured and whether sensitive data is configured with delete protection. When a policy produces an alert in a given environment, additional context is included to help security teams take action, including activity and change logs, context on the data store owner and environment, and visibility into the sensitive data classes in a given data store as well as samples of the data (based on the visibility rules configured for the environment and a user's level of access).

Automated Remediation Workflows

Taking action to remediate sensitive data exposure can be a time-consuming and noisy undertaking. Cyera was designed to streamline that process with prioritized alerts that benefit from the deep data context we apply when classifying data and by including environment, data store, and access context, along with remediation guidance. By combining this information with common security, operations, and DevOps workflow and toolchain components, Cyera can automate remediation workflows for common security exposures.

Cyera has prioritized integrations with tools that span:

- Automation
- SIEM
- CSPM and CWPP
- Vulnerability
- Data catalogs
- Workflow
- Identity management

Cyera's platform can easily integrate with additional solutions via a RESTful API.



Summary

Cyera is revolutionizing the way businesses secure their data. We have taken a cloud-first approach toward creating a data security platform that brings together sensitive data discovery and classification, data security posture management, data loss prevention, and data access governance. Cyera instantly provides companies visibility over all of their sensitive data, context over the risk it represents and their security exposure, and automated remediation to reduce the attack surface and ensure operational resilience.

Data Discovery and Classification for the Cloud Era

Data is growing at an incredible pace, due in large part to the critical role it plays in connecting people, creating new business opportunities, and deriving insights that fuel innovation. Cyera was purpose built as a cloud-native platform to empower businesses to harness the power of their data and to elevate security teams from gatekeepers to critical partners in enabling collaboration and sharing securely.

Effective and accurate data discovery and classification are foundational for any data privacy, governance, compliance, or data loss prevention effort. The ability to dynamically and continuously discover data across a company's landscape is critical given the rate of creation and change that the cloud and remote work have introduced. Cyera's agentless technology and novel technical innovations to streamline and accelerate data store inventory creation and data classification ensure that businesses will have the confidence that they always know what data they have, where it is located, and what it represents.

Data Security Posture Management (DSPM)

Cyera is a pioneer in the DSPM space. Our platform produces multidimensional data risk assessments, reviewing the sensitive data that is stored, accessed and managed in customer environments.

Cyera takes into account the environment, data store, sensitive data, and user access, and evaluates the exposure present against a set of risk assessment policies. Because the process is automatic and continuous, security teams are empowered with up-to-date context and understanding of sensitive data exposure, as well as insight into how to best partner with the business to address and remediate risk. Cyera's DSPM capability complements our discovery and classification insight by providing the answers to the questions, **"what sensitive data is exposed and how can I take action to remediate that risk?"**

Cloud DLP and Data Access Governance

Cyera was conceived and developed by operational cybersecurity experts with real-world experience with both defensive and adversarial security strategy and tactics. In the cloud era, attackers and insiders can behave in similar ways, presenting similar challenges to protecting sensitive data. Cyera combines critical discovery, classification, contextualization, exposure assessment, and remediation tools into a single platform because the aggregate context is necessary to secure data in sprawling cloud environments.

Cyera's data security platform focuses on pragmatic security controls. It highlights and prioritizes critical exposures, data posture issues, and risks associated with overly permissive access - whether from public exposure or insiders - to empower security teams in their quest to protect the business while enabling it to collaborate and grow. Cyera allows security teams to stay ahead of data regulations and risks. This frees up the rest of the organization to harness data and stay on the forefront of innovation.

About Cyera

Cyera is reinventing data security. Companies choose Cyera to improve their data security and cyber-resilience, maintain privacy and regulatory compliance, and gain control over their most valuable asset: data. Cyera instantly provides companies with a holistic view of their sensitive data and their security exposure, and delivers automated remediation to reduce their attack surface. Learn more at www.cyera.io, or follow Cyera on LinkedIn.

Trusted by:



info@cyera.io | cyera.io