White Paper

Detect and Protect:
A Data-Centric Approach to Security

This edition published April 2017

Table of Contents

Executive Summary

As sensitive data proliferates in cloud and mobile environments, data lakes, and other big data repositories, it poses a growing challenge to data security. Indeed, Gartner says data security governance and the orchestration of data security policies across disparate data silos and platforms will be critical challenges for organizations over the next decade.[1]

Protecting expanding data environments calls for a "Detect and Protect" approach to data security that uses next-generation tools to monitor data movement and access, track user access and behavior that may indicate data misuse, and automate and orchestrate the dynamic application of security controls. To deliver more proactive and intelligent sensitive data risk monitoring, investment prioritization, and protection, these tools must include all the following capabilities:

- Automated and integrated sensitive data discovery, proliferation analysis, detection of anomalous user activities, multi-factor data risk analytics, and automated orchestration of remediation in a single platform
- Broad coverage across cloud, Hadoop, and legacy environments
- Support for diverse and complex data classifications, proliferation and risk analytics, user access and activity correlation, user behavior analytics (UBA) and orchestration of data security controls

This paper explores how these capabilities reduce the risk of data breaches and misuse while improving compliance with data privacy regulations such as GDPR, HIPAA, and FINRA, especially when used on a comprehensive single platform that also lowers total cost of ownership. In addition, the paper describes what organizations should consider when evaluating functionality in these six critical areas:

- Sensitive data elements (data domains)
- Policy management
- Discovery and classification
- Protection and orchestration
- User access and user activities
- Alerting and actions

Finally, the paper addresses what it takes to implement a "Detect and Protect" solution and highlights the benefits of a properly deployed data security solution.

Controlling the Data Explosion

The days of slow-growing, fixed data infrastructures are over. Today, data doubles every 12 to 18 months, bringing with it a host of challenges:

- Diversity in the form of structured, semi-structured, and unstructured data
- Proliferation as data moves around the globe from data stores including Hadoop nodes, cloud instances, file servers, and relational databases
- More users from more locations, functions, and geographies
- Exponential growth in on-site, remote, and mobile access points

Given this shift, organizations can no longer rely on segmented and often disconnected approaches to finding, analyzing, monitoring and protecting sensitive data. They urgently need to shift to a new data security paradigm that takes a holistic approach:

- A complete, enterprise-wide view of sensitive data
- Risk analysis that makes it easier to prioritize security programs and investments
- Continuous monitoring of sensitive data and its use to provide early warnings of misuse or breach
- Integrated protection that connects discovery, risk, and monitoring to automated remediation

CISOs, enterprise architects, and privacy and compliance officers know they need continuous, contextual, and responsive protection for enterprise data assets to spot new threats as they arise. However, to achieve this goal at enterprise scale, they must tackle several challenges:

Standardization, Consistency, and Auditing

To protect thousands of data stores, organizations need to standardize data definitions and associated protection policies. Without central policy management, security practitioners and architects cannot consistently manage policy changes or audit and track compliance.

Accuracy and Effectiveness

To maintain a high and reliable level of accuracy in a constantly changing environment, a data protection solution requires data context and user behavior analytics. Without them, the solution will generate a high number of false positives, thus rendering itself ineffective at continuous “detection and protection.”

Scalability and Time to Value

Achieving risk reduction within tight deadlines calls for a data security solution that can scale to protect many data stores in far less time than it would take to protect each data store individually. This is a pressing concern for security and compliance officials who must, for example, meet the GDPR deadline and enforcement date of May 25, 2018 or risk fines of up to 4 percent of annual revenues. Without intuitive usability, orchestration of actions, workflows, machine learning, and automation, a data security solution can neither scale nor deliver speedy time to value.

Continuous Protection Without Interruption

A data protection solution must be flexible enough to maintain data protection and compliance in an environment in which data, usage, users, and regulations are in constant flux. If it cannot support policy customization and regular reassessment of risk, it cannot provide truly continuous detection and protection.

Out-of-the-Box Readiness for Deployment in Non-Greenfield Environments

CISOs today need not just a data-centric approach to security with new capabilities driven by machine learning, but the ability to derive more value from current security controls wherever possible. Any addition to the security ecosystem should be API-driven for easy integration with already implemented security controls such as SSO, password management, ranger, sentry, shield, encryption, and tokenization solutions.

Support for Hybrid Environments

Organizations in every industry, from financial services and healthcare to transportation and entertainment, are at some stage of digital transformation. At this stage of their journey to the cloud, their world is a hybrid one. Any data security solution they choose must cover all their data sources, whether traditional, big data, or cloud-based.

Evaluating “Detect and Protect” Solutions

To secure sensitive data, organizations must start with the basics:

- Understanding where sensitive data resides
- Systematically and consistently identifying and classifying sensitive data
- Setting policies for handling it properly
- Implementing technical controls that enforce data handling rules
- Educating users about their role in keeping sensitive data safe

To support these basic requirements, a platform that manages sensitive data must include all of the following functional areas:

Sensitive Data Elements (Data Domains)

This capability involves the ability to create and modify search definitions for data domains such as credit card numbers, addresses, names, etc.

Ideally, it should come with out-of-the-box data domains that include most domains defined in major regulations such as PII, PCI, and PHI, as well as the ability to create custom domains to cover additional requirements. In addition, it should include the following features:

- Ability to create search definitions for both metadata and data
- User interface to create advanced search rules such as the ability to do logical checks against data (e.g., check sums)
- Ability to create search definitions using pattern matching (including regular expressions), reference tables, and rules
    - Ability to minimize false positives through conformance scoring (only accept if > xx% of selected rows match), blacklists, and whitelists

In addition, consider looking for these features:

- Ability to modify the search criteria for out-of-the-box data domains
- Ability to create advanced rules beyond regular expressions
- When searching both data and metadata, ability to create a rule prioritizing one or the other when results conflict

Policy Management

This capability defines sensitive data classification policy by sensitive data elements and protects these elements in compliance with all applicable regulations and laws. It should easily create as many custom policies as necessary, each one containing, at a minimum, the following information:

- One or more sensitive data elements, with the ability to define complex combination rules. (e.g., only match if Name + Address + SSN are present)
- Classification level (sensitivity level)
- Breach cost for each row that matches the policy

In addition, consider looking for these features:

- Policy management that is part of the platform rather than a separate application
- Ability to assign classification levels at the policy level
- Ability to build sophisticated, multi-element matching rules within the policy
- Availability of all data domains without needing to import them
- Ability to associate a cost to sensitive data

Discovery and Classification

This capability automatically locates and classifies sensitive data, then calculates risk/breach cost based on defined policies.

Traditional discovery and classification tools identify databases containing sensitive data by listing the table/column names where that data resides. While this approach offers some insight into where protection may be needed, it does not provide any guidance on where to start or indicate whether the data has already been protected. The better alternative is a common risk framework that compares data stores to each other to generate a risk score that drives the protection effort.

The capability should include multiple adjustable factors in its risk framework to calculate the risk for each data store:

- Classification/sensitivity level of the data
- Amount of sensitive data
- Protection status of the data
- Number of sensitive fields
- Breach cost of the data
- Sensitive data movement; how many targets the data moves to
- Number of users with access to the data
- User activity count against the sensitive data

The solution should meet an organization's specific needs around scalability, the ability to scan across multiple vendor platforms, and different data repositories such as databases, Hadoop, cloud, etc.

In addition, consider looking for these features:

- Agentless scanning
- The ability to calculate a risk score for each individual data store, with details about how the risk is calculated
- Ability to schedule scan jobs
- False positive detection and handling
- Ability to configure scans to use metadata, data, or both
- Automated email and other notifications when a job is complete

Protection and Orchestration

This capability automatically applies defined protection policies to encrypt, mask, or otherwise protect sensitive data. It should be launchable directly within the target data store or by initiating a protection workflow.

The capability should support persistent data masking, dynamic data masking, encryption, access controls, and other data protection methods. Protection policy creation should be bi-directional: protection rules can be built within the platform and pushed to protection tools, or vice versa.

In addition, consider looking for these features:

- Protection tools that operate across all supported data store types
- Integration with third-party protection tools and vendors
- A common policy framework for discovery and protection

User Access and User Activities

This capability makes the platform more dynamic and improves risk scoring by capturing who has access to sensitive data and how it is used. To correlate user activities against sensitive data, this capability must be able to ingest raw user activity log files in real time, then integrate them with LDAP and Active Directory data such as users, user group memberships, user aliases, and user access to data stores, tables, and columns. It should be able to report on users and groups with access to sensitive data by user(s), data store, time of occurrence, and other criteria, and incorporate this user activity into each data store's overall risk scoring. It should also leverage machine learning techniques to capture and generate automatic alerts for user anomalies.

In addition, consider looking for these features:

- Ability to ingest user activities across all supported data store types
- Ability of Syslog server to ingest industry standard activity logs
- Ability to import user access rights
- Availability of user behavior analytics

Alerting and Actions

This capability creates online and email alerts to inform users of exceptions and other defined conditions that require their attention. It should include the ability to create scripts that trigger automatic actions in response to an alert (e.g., moving users from one LDAP group to another).

Implementing a Solution

An organization cannot achieve a comprehensive view of the risks associated with sensitive data until it can identify data location, volume, proliferation (where the sensitive data is created and how it flows through the organization), and protection status (how the data is currently protected by data security controls). The security team needs to confirm what it already knows, but it also needs to understand what it may be overlooking. Regular aggregated risk scoring can deliver a quantitative measure of the security team's blind spots, helping it set priorities and focus 80 percent of its efforts on the 20 percent of data at highest risk.

Understanding User Activity, Behavior, and Anomalies

Data does not compromise itself; a user is always involved somehow. To determine why and how, the security team needs to capture the type, volume, location, and other characteristics of data each end user accesses, then apply User Behavioral Analytics (UBA) to create profiles and baselines of user activity. In this way, the organization can identify normal user behavior so that abnormal behavior stands out sharply. The additional context makes detection of suspicious events more accurate so IT can defend against or remediate potential threats faster.[2]

Trigger Alerts For Policy or Behavior Exceptions

Assigning specific protection techniques and tools for each type of policy, noncompliance or suspicious activity makes the security team more efficient and effective.

In some cases, it might work best to take a fully integrated, automated approach to detection and protection. For example, when a user is accessing more SSN records than normal, a script could automatically activate to dynamically mask the data or move the user to a high-risk user group in LDAP. In other cases, the security team might simply want to generate an alert that requests or requires manual intervention by the data owner or application owner.

As the security team configures and rolls out the data protection solution, it will need to take the following steps:

- Define what protection to use for what asset type and user
- Establish an appropriate level of integration with the protection method
- Monitor the effectiveness of the solution continuously and adapt as required
- Expand scope to more users and more information asset types

Focus on the Data

Knowing where data resides, how it's being used, and who is using it is critical to determining how to protect it. Data servers in a secure location may not require encryption at the disk or file level, but databases hosted by a partner in another location or even another country will need stronger protection to prevent loss of control. Data lakes accessed by multiple users and applications will need to take location, time, and necessary levels of access into account in setting up access rules. Data controlled by GDPR will need dynamic access rights that adapt to users' time and location.

Test environments pose their own challenges. Environments being used for functional testing need realistic data to continue operating smoothly. Protection applied to columns of sensitive data needs to ensure the relationships between tables remain intact.

In cross-system business processes, protecting sensitive columns must not break the business process. Persistent masking can be deployed in reporting, analytical, and test environments that have little or no need to restore the protected data to its original value.

In production systems, protecting data at rest may be less important than protecting data appropriately for different groups of users. In these cases, data needs to be completely protected from some users and only partially protected from or entirely available to others. A banking call center provides an obvious example of this: DBAs don't need to see a caller's SSN at all; customer service reps need to see the last four digits; and back office users who validate the callers' credit history need to see the SSN in Clear. This fine-grained access control needs to take place dynamically.

Protection is a Team Sport

Application owners and security analysts must cooperate closely with the DBAs who maintain data operationally. A business process orchestration tool can automate and measure the handoff process among these three groups, ensuring the right level of data security at all times, updating internal systems to reflect that data is protected, and recalculating the risk associated with the data so the security analyst who initiated the protection request can be confident the protection job is complete.

Automation is the Way Forward

Policy-based data policing ensures that even in a fast-changing environment, data remains up-to-date and secure. The security team can implement rules that continuously scan new data and changes to metadata. If an anomaly emerges, the data protection solution can create an alert for a policy violation and suggest corrective action—for example, alerting the owner of an application to take corrective steps to protect data if a new column that contains account numbers appears in an existing system due to a metadata change. Alternatively, the solution can take action automatically—for example, identifying someone downloading a report containing sensitive data as a DBA, who would not ordinarily be allowed to download that information, and automatically shifting that user into a group without access to the report.

While policy-driven automation is a tremendous help to the security analyst, it can be difficult to get right on the first try. It is generally best to test the policies first in a small group of users or applications, learn from errors, and roll them out gradually.

Business Benefits

A properly deployed "Detect and Protect" data security solution offers business value beyond simply reducing risk. By providing a comprehensive and continuous view of risk for all data assets and users, it facilitates governance of security policy and control, ensuring that data owners don't sacrifice flexibility for compliance while keeping data security policies consistent across systems and data sets. It also creates a defensible legal position in response to a data breach or audit challenge.

By supporting an approach to data protection that focuses on applying the right method at the right time to the right type of data, the solution also encourages collaboration between data owners and application owners and improves their buy-in to security measures.

Finally, the solution delivers faster time to value and lower operational costs. By automating data protection based on risk detection associated with users and/or data stores, it frees IT staff and budget for more strategic tasks.

Intelligent Detection and Protection with Informatica

Informatica Secure@Source is a comprehensive detection and protection solution with the strongest set of capabilities currently available to power the entire data security cycle. Using Secure@Source, organizations can manage data security across the enterprise:

- **Know where sensitive data resides** – Global visibility, classification, discovery and reporting of the organization's sensitive data and where it is proliferating.
- **Recognize what is at risk** – Continuous multi-factor risk scoring based on data sensitivity, protection status, risk cost, and more using Informatica's innovative interactive risk analytics technologies.
- **Understand who is using the data** – Visibility into users with access to sensitive data, with usage monitoring and highly accurate anomalous behavior detection.
- **Orchestrate effective remediation actions** – Risk-based remediation recommendations and orchestration of data security measures including dynamic or persistent data masking, encryption, blocking and other solutions from Informatica and third-party providers.

According to Gartner, "Security spending on detection and response to attacks on systems and services must be refocused to monitor and protect applications and data. Security and risk management leaders should use data security governance to prepare risk-based security strategies and infonomics to reset budgets."[3] With Informatica, your organization can refocus its spending wisely and patch security holes that weren't previously visible.

[1] Gartner, "Application and Data Security Primer for 2017," Brian Lowans, Ayal Tirosh and John Girard, 24 January 2017

[2] Ibid.

[3] Gartner, "Predicts 2017: Application and Data Security." Brian Lowans, Neil MacDonald, Marc-Antoine Meunier and Brian Reed, 22 November 2016

About Informatica

Digital transformation is changing our world. As the leader in enterprise cloud data management, we're prepared to help you intelligently lead the way. To provide you with the foresight to become more agile, realize new growth opportunities or even invent new things. We invite you to explore all that Informatica has to offer—and unleash the power of data to drive your next intelligent disruption. Not just once, but again and again.

Informatica

Worldwide Headquarters, 2100 Seaport Blvd, Redwood City, CA 94063, USA Phone: 650.385.5000 Fax: 650.385.5500 Toll-free in the US: 1.800.653.3871 informatica.com linkedin.com/company/informatica twitter.com/Informatica

IN09_0617_3332