ESG Lab Review

Informatica Secure@Source

Date: April 2016 **Author:** Tony Palmer, Senior Lab Analyst

Abstract

This ESG Lab review documents hands-on testing of Informatica Secure@Source data security intelligence platform. Testing focused on the ability of Secure@Source to provide visibility into organizations' sensitive data and give those organizations insights into the risks and potential financial impact exposure of their sensitive data across the enterprise, to prioritize their data security investments.

The Challenges

The cybersecurity industry has been talking about the intersection of big data and cybersecurity analytics for years. Recently published ESG research confirms that data security intelligence is a key component of cybersecurity, and organizations need to pay attention.[1] While nearly half (48%) of surveyed organizations say they use sensitive data access and usage information as part of their threat management analysis activities and processes today, most organizations do not have deep insights or a clear, up-to-date understanding of the risks or potential financial impact of a breach of their sensitive data.

The data they do have is collected from many disparate parts of the organization and is typically months out of date because of the manual processes used to collect and analyze it. The people who are intimately familiar with the data—CIOs, architects, and data analysts—know this is a problem and know they need tools to fix it, but the rest of the organization—even the information security team—is rarely even aware of the issue.

There is plenty of confusion in the market around data, context, and security solutions. How is Informatica different from a traditional security, information, and event management (SIEM) solution? Most cybersecurity solutions focus on the network, endpoints, and access and do not understand or integrate the content or the context of the sensitive data. This context is critically important and correlation between data risk and activity on the data is key. Many data security solutions treat the problems in silos – focusing on databases, file systems, cloud and Big Data separately. When enterprises and large organizations are supporting thousands to tens of thousands of databases, Hadoop, and cloud applications in their environment, a tool that can provide visibility across all these disparate data stores, define hierarchies, and provide logical organization—uncovering, measuring, and assigning sensitive data risks—is not just a "nice to have"—it's essential.

The Solution: Informatica Secure@Source

Secure@Source, the Informatica data security intelligence platform, is designed to provide organizations with visibility of sensitive data and its risks to enable businesses to prioritize security investments, manage audits for privacy and industry legislation and detect anomalous activities in a timely manner. Secure@Source applies prebuilt and customizable definitions to specific domains of data to identify and locate sensitive data and classify them based on an organization's levels of sensitivity. Secure@Source also analyzes the volume of sensitive data being moved and/or copied from one data store to another because data proliferation increases the potential threat surface. It then collects and correlates user access i.e. who has access to which sensitive data, and user activity information. Based on these insights on sensitive data Secure@Source provides risk scores that can be used to quantify and continuously measure sensitive data risk and cost across the organization, over time, visually and in a consistent way.

[1] Source: ESG Research Report, *Threat Intelligence and Its Role Within Enterprise Cybersecurity Practices*, June 2015.

To accomplish this, Secure@Source provides key capabilities including:

- Automated Discovery & Classification: finds sensitive data via automated analysis of both data and metadata, along with appropriate context (e.g. user name and SSN appearing together) in targeted sources, which can be databases, mainframes, Hadoop, and SaaS applications. This allows organizations to define, discover, and analyze data domains and classifications with deep context.
- Visibility to Location and Proliferation: provides insight into where sensitive data resides, data movement, and replication to targets both within the organization and to public cloud applications.
- Collection & Correlation of User Access & Activity: incorporates user access information from LDAP and IAM systems, along with user activity logs from databases, Hadoop distributions, SaaS applications, and Database Activity Monitoring (DAM) feeds.
- Risk and Cost Analysis: determines the level of sensitive data risk via analysis and weighting of multiple factors including the protection status of the data, user access and activity, data location and proliferation, the cost of exfiltrated data due to a breach, and classification.
- Security Ecosystem Connectivity: integrates with a broad array of security products and solutions to enhance enterprise data privacy and security risk analysis.
- Visualization and Reporting: delivers active and clickable elements, enterprise level summaries for the C-level executives and drill downs by departments and data stores to offer more detailed information, providing organizations with enterprise-wide visibility of sensitive data risks and allows them to standardize measurement across departments and identify key areas of improvement.

Identification of High Risk Conditions and Anomalies: enables the security operations team to define rules to identify and detect high risk conditions as well as monitor for unusual user activities and data movement.

ESG Lab Tested

First, ESG Lab took a look at the Secure@Source main dashboard, shown in Figure 1. The dashboard shows key risk indicators at the top. These indicators are viewable organization-wide but can also be controlled by: privileges (role-based access), and the data that users have access to. It's important to note that the UI is completely active, and every object on the page is clickable, with links to context-sensitive details. The blue bar shows the percentage of the organization's data that has been scanned. The solid portion is indicates the percentage of data stores that have been fully scanned, while the dotted part indicates the percentage of data stores that are partially scanned.

Figure 1. The Secure@Source Dashboard

Under the blue bar on the left is ***Risk Score***, which shows the overall data risk level for the entire organization with month-to-month trending. Secure@Source determines the risk score based on multiple attributes of the data and users' access. To the right of ***Risk Score*** is protection status. When Informatica finds sensitive data via discovery or input from API integration, this area shows whether that data is protected or unprotected. Next is the sensitivity level of the data. Sensitivity level is stacked with the most sensitive data on top. The default categories of sensitivity are ***Restricted***, ***Confidential***, ***Internal***, and ***Public***, but these are completely customizable. Finally, ***Data Risk Cost*** indicates the monetary value of the organization's sensitive data to provide potential risk and liability guidance to executives.

Risk scoring is based on many attributes in addition to sensitivity level. Attributes like protection status, sensitive fields, level of proliferation, risk cost (the monetary value of the organization's data), user access, and activity. Activity count refers to how much the data is accessed. This is translated into events and impressions—as an example, in a healthcare database, the statement ***select * from patients*** would constitute a single event. Depending on the number of records actually exposed to the user, an event would translate into many more impressions.

ESG Lab also looked at how data is collected and analyzed. Secure@Source uses the concept of data domains to define how data fields are recognized as sensitive. Examples of data domains might include age, birth date, Social Security number (SSN), credit card number, or anything else an organization might want to identify as sensitive. Secure@Source ships with

hundreds of predefined data domains and content rules. Users can customize existing data domains or create new ones to meet their specific needs.

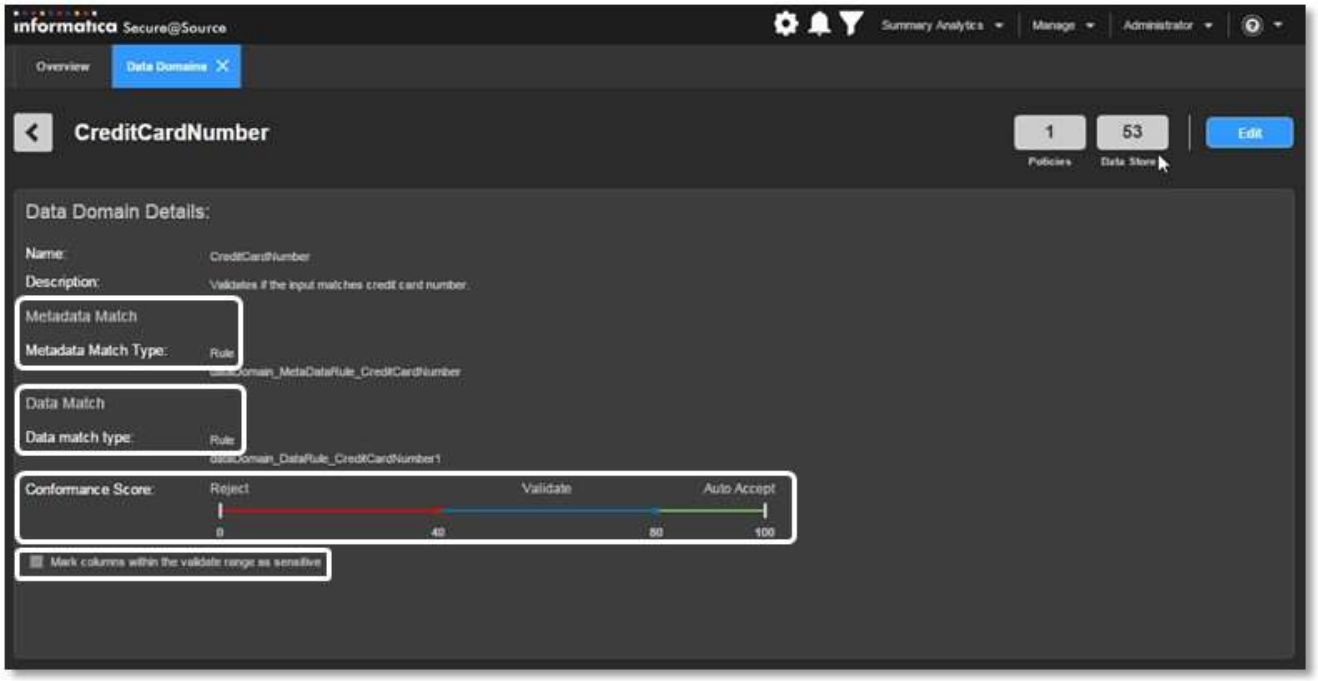
ESG Lab edited the ***CreditCardNumber*** Data Domain, shown in Figure 2. A data domain match can be confirmed via metadata like column headers, the actual data using pattern matching, reference tables (Informatica provides predefined tables), or rules. The default is to use rules.

The conformance score allows organizations to automate how sensitive data will be accepted or rejected for the domain and when curation (validation by a human) will be required. The action ***Mark columns within the validate range as sensitive*** allows for further tuning of the organization's risk posture.

Data Domain Rules

A rule is a complex workflow that allows for multiple steps in the validation process and enables Secure@Source to incorporate blacklists, whitelists, pattern matching, and algorithmic methods to increase accuracy and reduce false positives.

Figure 2. Defining a Data Domain in Secure@Source



Data domains are used in policies. Policies are often focused on business-oriented standards, like Payment Card Industry (PCI) standards, or those governing personally identifiable information (PII). Secure@Source ships with many predefined policies, and all policies are highly customizable. A user selects the data domains to scan, then sets the data match conditions using Boolean (and/or) logic. This enables Secure@Source to dramatically reduce false positives by checking multiple fields in combination. A single field by itself is not necessarily personally identifiable unless it is combined with other fields—SSN, age, or birth date, for example.

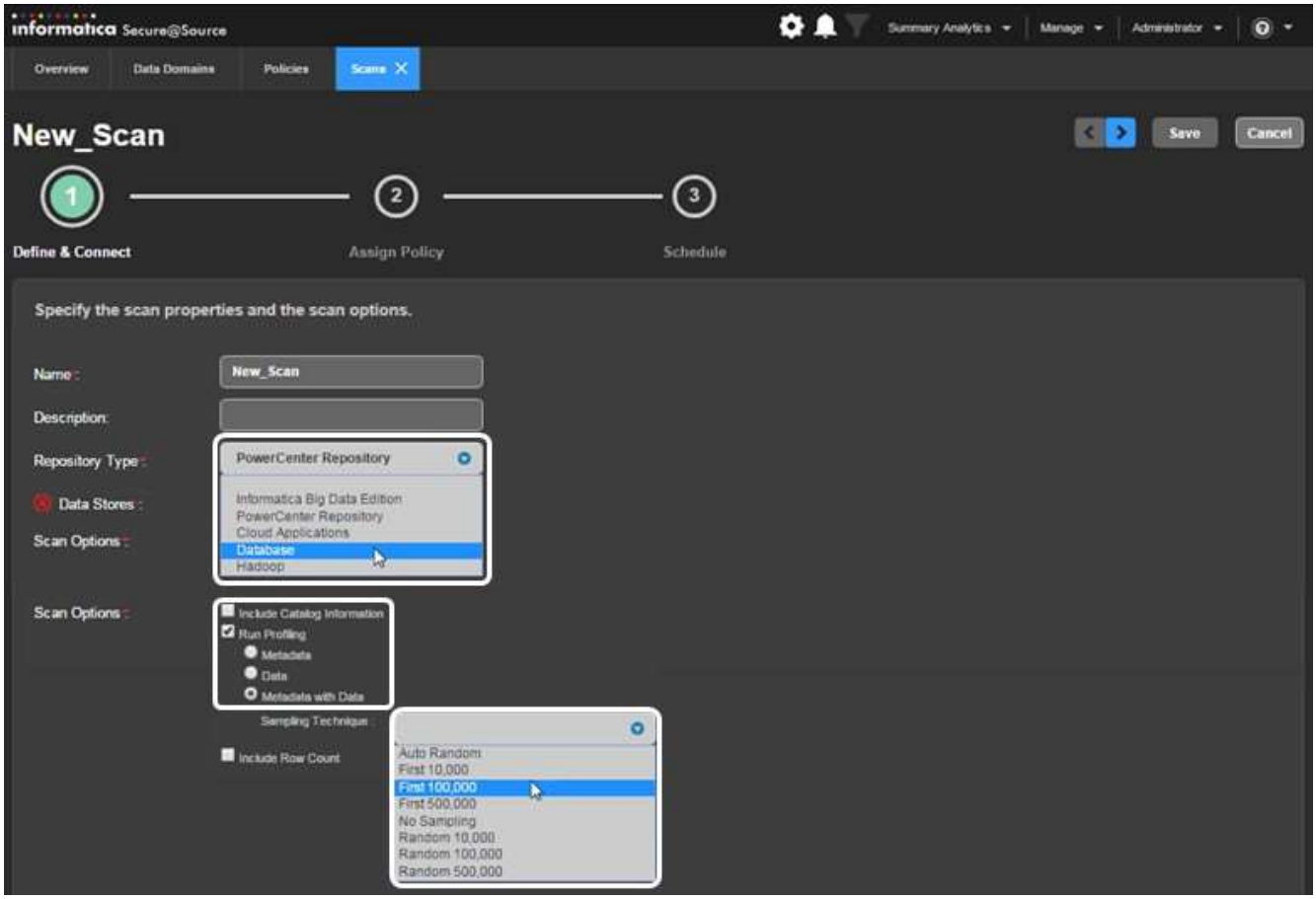
Policies also allow organizations to define the cost of each data loss or exfiltration occurrence. This can be quite subjective depending on the industry, type of data breached, and whether soft or indirect costs are included. Industry estimates vary from less than one dollar up to $200 per record. Informatica took a different approach, looking at what cyber insurance companies are actually paying organizations that suffer a breach, which averages from $75 to $100 per record. Informatica sets the default cost to slightly below this range to be conservative, but organizations can adjust that as they see fit.

Secure@Source uses the concept of a scan to get data into the system. Scans can be configured for different repository types as seen in Figure 3. Organizations can leverage Informatica's PowerCenter data integration platform or Informatica

Big Data Edition for quick time to value, or pull data directly from relational databases, cloud applications like SalesForce, or Hadoop.

Scans can be configured to look only at the database catalog (metadata) or set to perform deeper discovery of the data, or both and allow for customization of the sample size and technique.

Figure 3. Scanning for Sensitive Data



Secure@Source offers a variety of sampling options. By default, a random sample based on statistical analysis of the data in the repository is used, but the system can be configured to sample the data in a number of different ways, or simply scan all data. Once a scan has been completed, Secure@Source provides a scan report, which provides the risk score for the repository and breaks out details like the number of sensitive domains, fields, and records found.

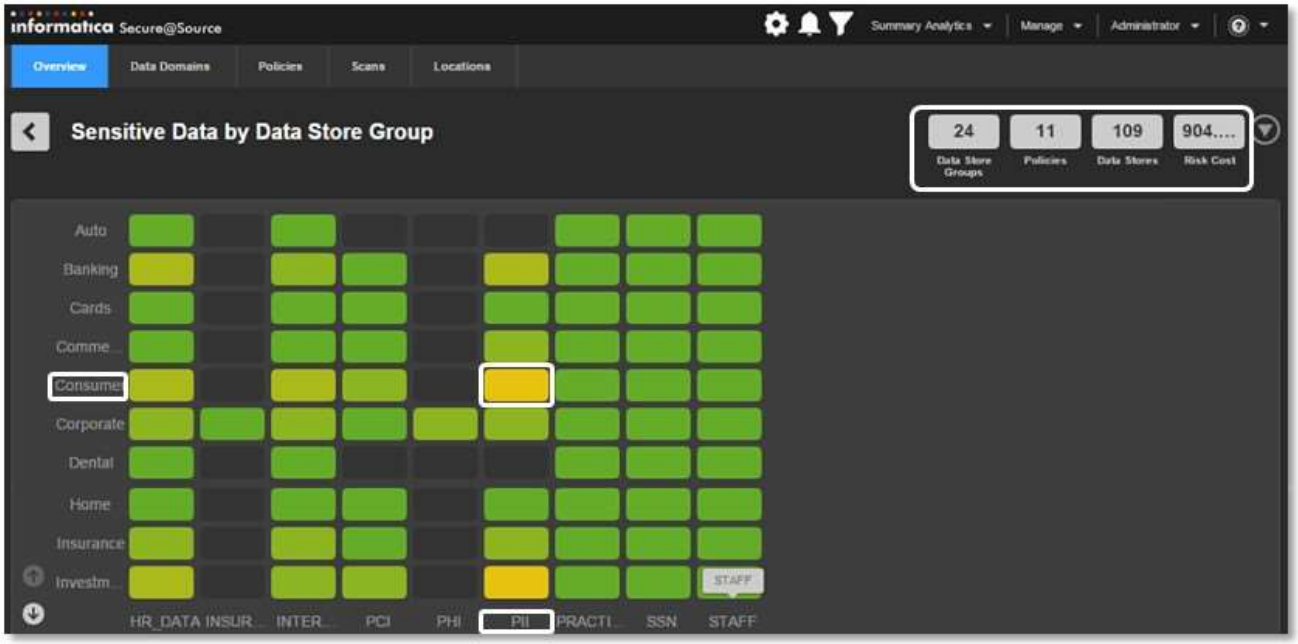
Secure@Source supports user synchronization over LDAP for Active Directory or Tivoli, or from homegrown solutions, as seen in finance and healthcare, which might use Active Directory in addition to other systems. Secure@Source integrates with numerous identity access management (IAM) solutions.

Finally, ESG Lab looked at the various visualizations, reports, and alerts of sensitive data that Secure@Source provides to organizations. ***Sensitive Data Locations***, as shown in Figure 4, provides organizations with a geographical heat map that shows the risk score, number of data stores, and sensitivity level of the data by location. This view, like all others in Secure@Source, is active; clicking on any element will enable users to drill down to more detail for a specific location, or pivot to another view.

Figure 4. Secure@Source Visualization—Location Detail

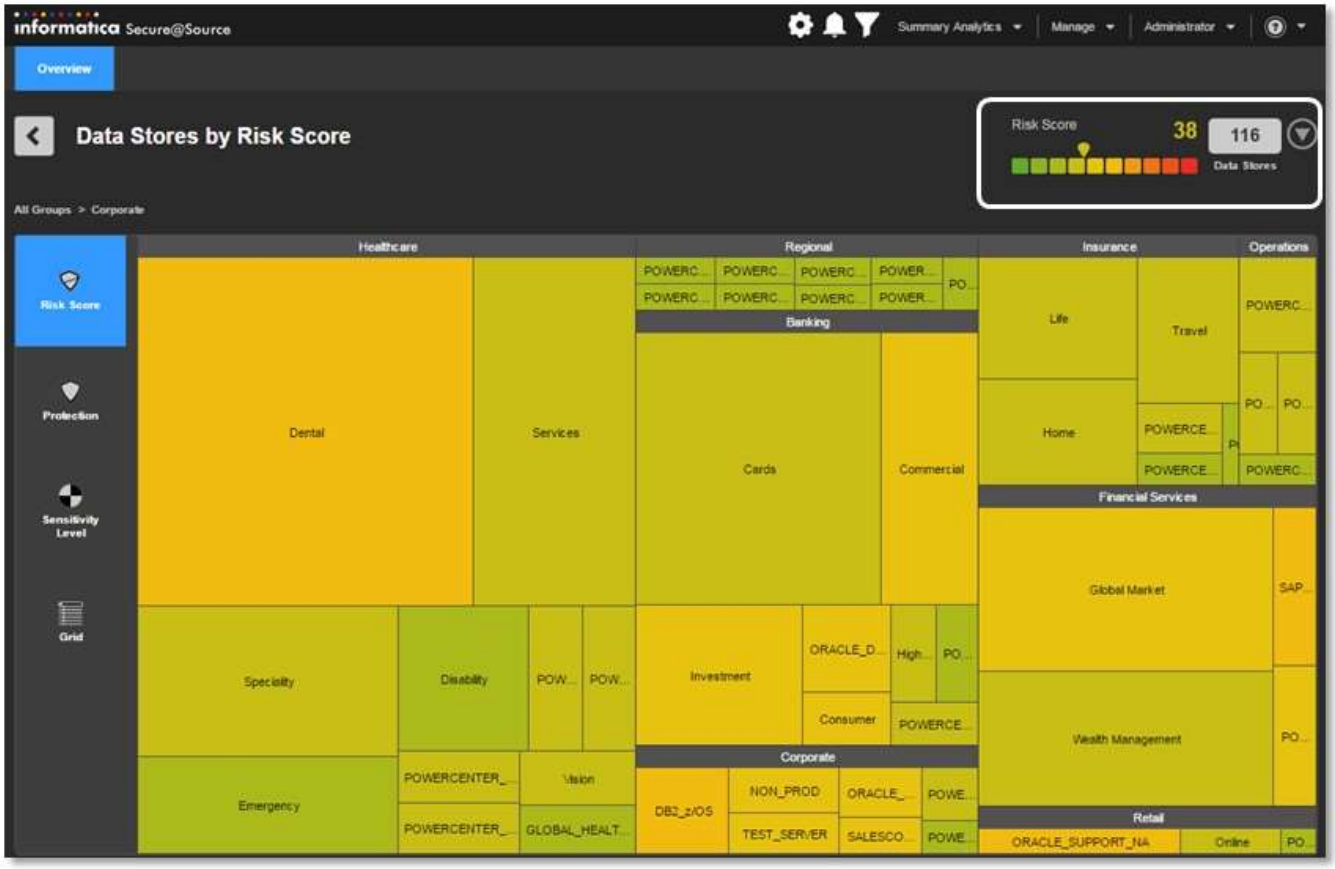
Next, ESG Lab looked at logical groupings of sensitive data. Secure@Source provides a wide variety of views of organizations' sensitive data, all accessible with a few clicks from the main dashboard or by pivoting from within a view. Figure 5 shows sensitive data defined by policies on the x-axis cross-correlated with data store groups on the y-axis.

Figure 5. Visualization—Sensitive Data Grouped by Data Store



In addition to the grid view, Secure@Source provides a heat map of the top data stores. The heat map can be viewed based on risk score, data protection status, or data sensitivity level within data store groups Figure 6 shows a view of data stores color-coded by risk score.

Figure 6. Visualization—Risk Scores for Data Store Groups

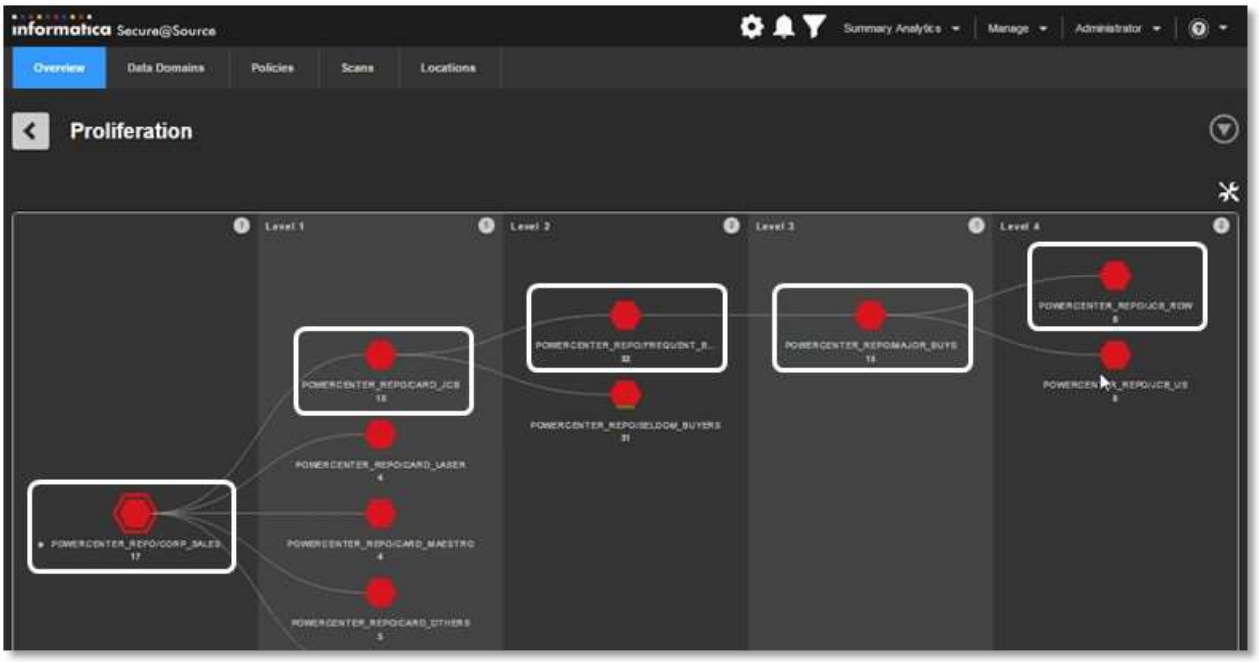


In any of these views, users get a summary view of the risk score and other details at the top, and a detailed view at the bottom, based on what element is selected in the chart. Clicking on ***Risk Score Details*** shows how the weighted factors contribute to the risk score.

Next, ESG Lab looked at data proliferation. Data proliferation can be examined from many angles and sorted by data store, by groups and policies, by department, and by user.

Figure 7 shows a detailed view of a source—a production data store —and all downstream targets. In this example, we can see that the sensitive data in this production database propagates to several downstream targets. The organization can use this tool to determine how sensitive data flows across data stores, whether it is protected, and whether the proliferation of that sensitive data is really necessary.

Figure 7. Visualization—Data Proliferation

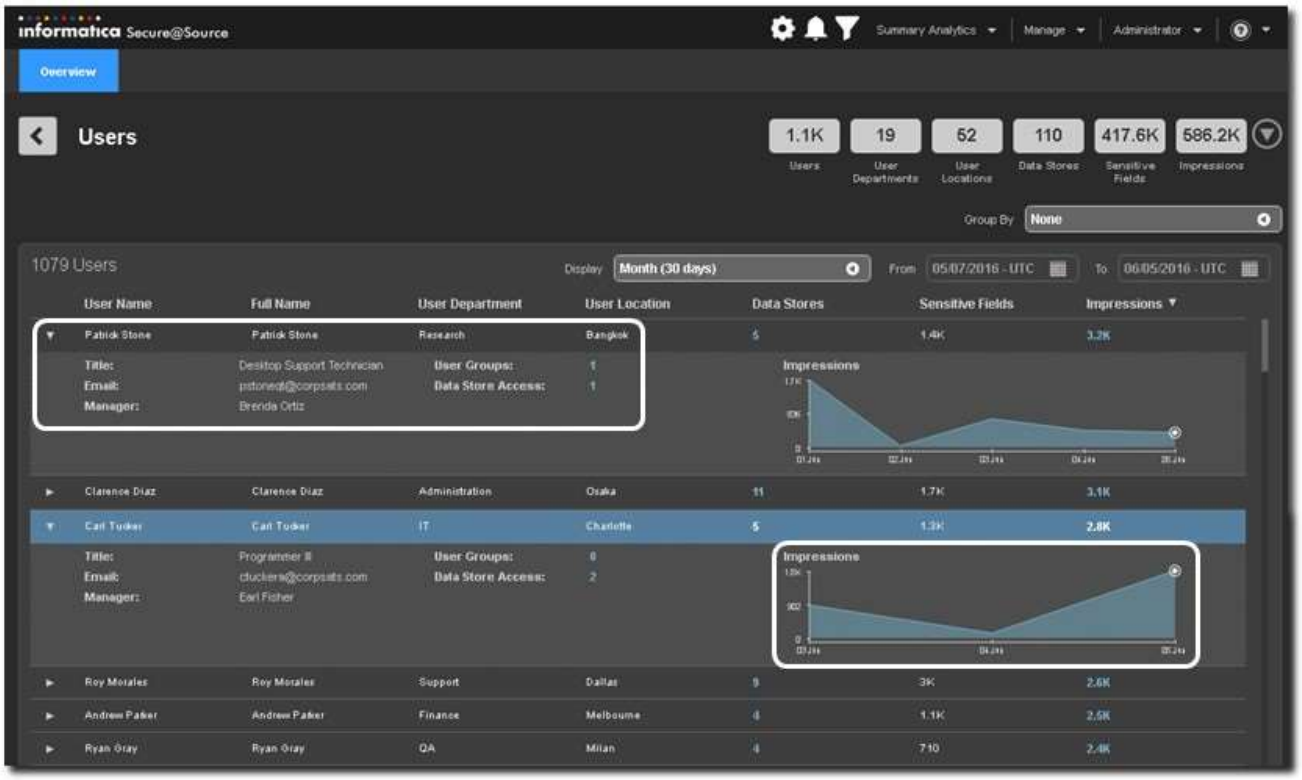


Secure@Source provides the same deep analytics for users. ***Top Users*** list shows users ranked by activity, summarizing how many data stores the users have touched and how many impressions they've requested as shown in Figure 8.

Figure 8. Visualization—User Access and Activity

Integration with LDAP provides details about each user. Secure@Source provides detailed information on user activity including the actual information viewed, and the number of impressions and sensitive fields, as shown in Figure 9.

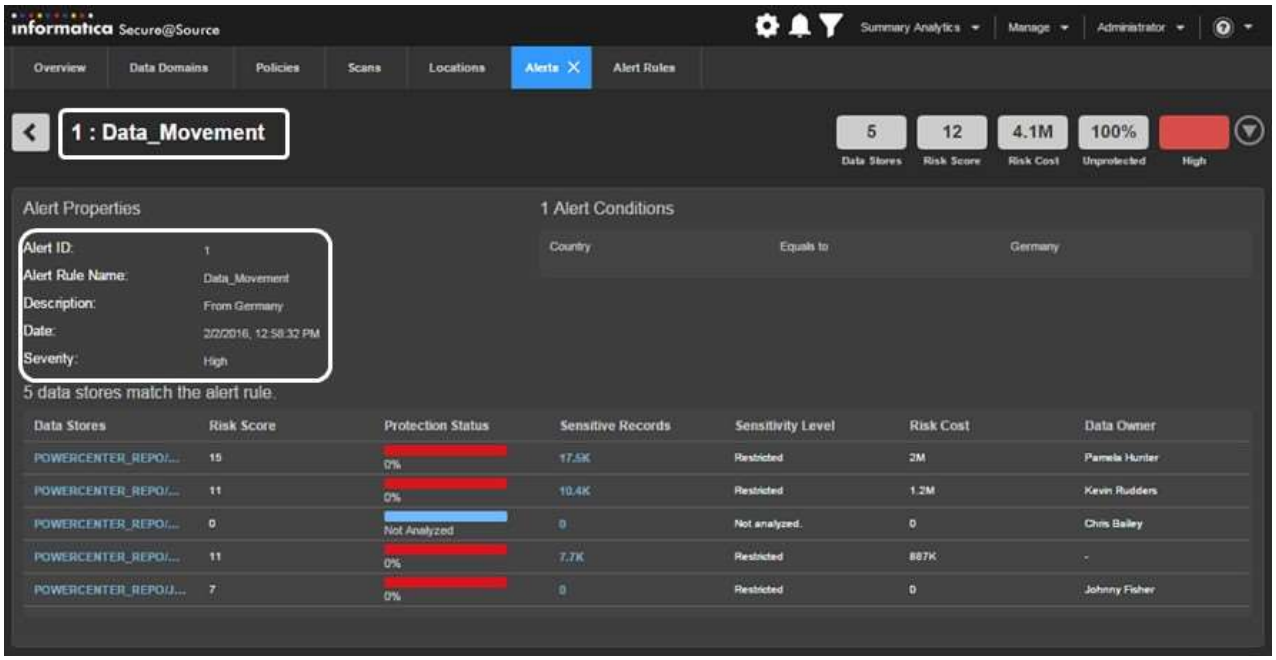
Figure 9. Visualization—User Activity



Administrators can click on a user to see the systems she has access to and pivot bi-directionally; clicking on a system that a user has access to shows all users and groups that have access to that system.

Summary Analytics provides predefined views of information grouped by data store, by department, and by users. Users can be ranked by overall activity, or by percent change to identify anomalies. Secure@Source also cross-correlates data store departments with user departments to highlight anomalies—for example, when sensitive data in one geography is moved to another geography. Administrators can also create alerts and schedule notifications to provide "management by exception," as seen in Figure 10.

Figure 10. Alerting



Why This Matters

Cybersecurity was the most-cited IT priority in ESG's most recent IT spending intentions survey,[2] while respondents to a separate survey reported that the threat landscape continues to grow more dangerous.[3] Nearly half (48%) of surveyed organizations say they collect and analyze sensitive data access and usage information as part of their threat management analysis activities and processes today.[4] Thanks to manual processes being used to collect and analyze the data, many of those organizations do not have a clear, up-to-date understanding of the risk or potential financial impact of their sensitive data.

A solution that combines intelligence and analytics on data security can provide organizations with visibility into what sensitive data they have and where it resides, along with insight into user activity such as information about the users who have access, what actions they are taking, and the status of the systems or data when they do so. This can enable organizations to understand and mitigate risk to their sensitive data across departments and lines of business.

ESG confirmed the ability of Secure@Source to provide meaningful metrics for CISOs to report up to the board. Rather than a report that simply says, "We did not have a data breach this month," A CISO can show metrics and trends like risk reduction, protection of sensitive data, control of access, and reduction of proliferation.

Secure@Source provides additional contextual information including location, groups, data movement, access, and activity, along with the value of that information. The user interface provides deep analytics with a multitude of representative views. Organizations can configure alerts to manage by exception or export information to report performance over time.

All of this functionality comes with a high level of automation, which simplifies the discovery, analysis, and mitigation of risks associated with sensitive data.

[2] Source: ESG Research Report, *2016 IT Spending Intentions Survey*, February, 2016.
[3] Source: ESG Research Report, *Cyber Supply Chain Security Revisited*, September 2015.
[4] Source: ESG Research Report, *Threat Intelligence and Its Role Within Enterprise Cybersecurity Practices*, June 2015.

The Bigger Truth

As stated in the introduction, there is plenty of confusion in the market surrounding data, context, and security solutions. Most cybersecurity solutions focus on the network, endpoints, and access and do not understand or integrate the context of an organizations' data. This context is critically important and correlation between data risk and data access is key to minimizing the financial risks of a data breach, which can be substantial.

When enterprises and large organizations are supporting thousands to tens of thousands of databases in their environments, a tool that can define hierarchies and provide a logical organization of data is not just a "nice to have"—it's essential. Cybersecurity has remained at the top of the IT priority list for the last four years, according to ESG research. When asked to consider their organizations' most important IT priorities for 2016, cybersecurity initiatives were the most-often cited (identified by 37% of respondents), and securing sensitive data is a vitally important element.[5]

Secure@Source is designed to provide organizations with clear, automated visibility into sensitive data. Prebuilt and customizable definitions of specific domains of data identify and locate restricted, private, and sensitive data, empowering businesses to understand activity and identify high-risk conditions or possible malicious or compromised users, and mitigate them. Secure@Source allows organizations to use the built-in discovery capability or import from data loss prevention (DLP) systems and behavior analytics platforms, map the data across their internal user data, and cross-correlate.

Secure@Source is able to provide meaningful metrics and trends like risk reduction, protection of sensitive data, control of access, and data proliferation. Additional contextual information helps organizations to precisely target sensitive data specific to their business while reducing false positives. ESG Lab found the user interface to be slick and powerful, providing deep analytics that enable administrators to rapidly pivot across departments or lines of business, with visualizations of risks, relationships, and potential costs.

ESG Lab validated that Informatica Secure@Source can help organizations to not only gain visibility into the sensitive data they have and the risks associated with it, but also leverage that intelligence to optimize the amount they are spending on protection according to risk. Any organization that is serious about gaining a clearer understanding of the sensitive data it has and reducing its exposure to data breaches would be smart to take a closer look at Secure@Source data security intelligence.

[5] Source: ESG Research Report, *2016 IT Spending Intentions Survey*, February, 2016