Data Discovery: Key to Data Privacy and Cyber Resilience

As organizations shift, re-prioritize, and transform their businesses in response to COVID19, regulatory pressures, cloud acceleration, and IT modernization efforts, adapting to change and being resilient has quickly become a must-have corporate skill set. In today's climate, cyber resilience expands beyond disruptions caused by cyber-attacks, and extends to the challenges surrounding data resiliency. Data resiliency depends on how well an organization understands, manages and protects its data. Meeting these challenges head on ensures sensitive data, critical information infrastructures, and corporate reputations are protected.

As data volumes continue to grow in both structured and unstructured applications, many organizations have also seen their cloud object and file stores grow as they accelerate their cloud objectives. Data privacy regulations and data protection strategies compound these challenges and introduce complexity. In response, many organizations are simply looking for a place to begin.

Organizations are leveraging data discovery as a critical step to identifying risk. Some, embarking on manual discovery efforts, are coming up well short of expectations as the data quickly becomes stale; while others are limiting scope to looking for just personal data without a plan for what to do with it when they find it. The most successful data discovery approaches allow organizations to build an understanding of their data, what value and risk it represents, and what actions can be taken to contain costs, detect, and protect sensitive data, and comply with regulatory policies.

Voltage Data Discovery Solution

A key component for customers looking to tackle these challenges is the Voltage Data Discovery solution. Voltage Data Discovery enables organizations to gain a deep understanding of the data contained within structured and unstructured data repositories. This understanding helps detect value and risk, and protect sensitive and high-value data, while providing flexible approaches that evolve to serve your needs over the different use cases throughout the lifecycle of your data. These key use cases include:

Data Minimization

Organizations are under significant pressure to only store and keep information they need. Voltage Data Discovery solution can analyze and detect data based on its value or business purpose to the organization. Rich analytics-driven dashboards and drilldowns identify functional and organizational data based on characteristics such as age, department, application, and data type. This information can be used to identify unstructured data that can be deleted or migrated to cheaper storage locations, containing cost and reducing risk of over-retention, and structured data and/ or application data sets that can be archived or deleted, reducing cost and improving performance while ensuring relevant business data is preserved and accessible.

Questions to Ask Yourself About Sensitive Data Discovery:

- Do you know where your risky data is?
- Do you have an approach that translates to actionable next steps when you do find it?
- Can you respond to auditors when asked about how you manage sensitive data, official records, and PII?
- Does your data protection plan include data in use?
- Can your business users access protected data easily inside their line of business applications?

Data Privacy Readiness and Sensitive Data Detection

During data discovery phases like data privacy readiness, a critical activity is detecting the location of risky and sensitive data. Voltage Data Discovery solution provides support for detecting personal data types across 39+ countries and economic regions. Contextually aware grammars and custom grammar capabilities enable highly accurate PII detection producing fewer false positives and driving greater efficiency and automation around handling sensitive data. Voltage Data Discovery also provides continuous or periodic scans along with random sampling across structured and unstructured repositories allowing organizations to identify and prioritize sensitive data hot spots and take more surgical approaches to protect data.

As data ages over its lifecycle, its business use and usefulness to the organization changes. What doesn't change is the ongoing associated risk around that data in the event of a breach. The significant rate of change of unstructured data lends itself to three distinct modes—active, inactive, or dark.

- Active data is that data that is currently in use by the business. It tends to reside in lead applications or file locations while it is in use—this is the data that drives your business.
- Inactive data is data that hasn't been accessed or modified for months and can be archived or deleted—this is the data that takes up space, increases costs and increases your complexity and risk by it lying around dormant.
- Dark data is data that has been copied, saved, moved or migrated without a true understanding of what value it represents or what sensitive data is present. Dark data can be as innocuous as a copy of a product brochure or as risky as a spreadsheet with 30,000 customers' personal information inside it.

As a result of this volatility unstructured data discovery is well suited to continuous scanning and discovery.

DATA DISCOVERY SOURCES

Voltage Data Discovery support a broad range of repositories, and over 1,000 different data formats, to analyze and protect sensitive data across the enterprise—in structured data across databases and data applications and unstructured data in on-premises and cloud repositories. These repositories include:

Unstructured Data Repositories

- NT file shares
- SMB (Samba)
- Microsoft Exchange
- Microsoft SharePoint
- Micro Focus Content Manager
- Documentum
- Filenet
- SharePoint Online & Teams
- Office 365
- Box
- GoogleDrive
- Gmail
- Microsoft Azure Blob
- Microsoft Azure file and object stores
- Amazon S3 object stores (S3)
- Custom ingest connector support available*

Structured Applications and Data Sources

Database

- Oracle SQL
- IBM DB2
- Microsoft SQL
- IBM DB2 iSeries
- MySQL
- Teradata
- PostgresSQL
- Sybase
- SAP HANA
- IBM Informix
- Netezza*
- Presto*
- Snowflake*
- AWS Redshift*
- AWS Athena*

Big Data Sources

- Apache Hadoop HDFS
- Hadoop Hive
- Hadoop HBase
- Vertica
- Cassandra
- MongoDB
- Teradata
- Additional data sources available*

Business Applications

- Oracle eBusiness Suite
- PeopleSoft
- Siebel
- Salesforce.com
- SAP
- JDEdwards
- Any (like custom) that stores in a JDBC source

***Integration with Additional Data Sources and Business Applications via:**

- Swagger API
- JDBC type-4 (structured databases, applications, and big data source)
- IDOL connector framework (~150+ connectors)

Sampling, Tagging and Enrichment

Prioritizing your data discovery projects is critical. Voltage Data Discovery provides risk-based random sampling across both structured applications and unstructured repositories highlighting risk and helping guide future efforts to assess and mitigate risk. Categorization of sensitive data can be

automated, tagged, and metadata enriched based on pre-built sensitive data grammars and classifications. These grammars are highly precise, contextually-aware analysis that help support a broad range of regulatory guidelines including GDPR, CCPA, PIPEDA, POPI, KVKK, as well as, PCI, PHI and custom use cases.

As you develop your 3rd party vendor relationships these categories and weighted tags can be leveraged to highlight, monitor, and report on data shared with 3rd parties and its usage, thus helping ensure continuity with these sharing agreements.

> If data is the new currency, data protection is the central bank for an organization looking to protect its assets and preserve their worth.

Scale to Meet New and Evolving Workloads
As organizations look to meet the evolving workloads, wrangle enterprise data, comply with global regulations and contain the cost associated with managing and securing sensitive data—data discovery must scale along with those needs. Voltage Data Discovery is ideally suited to meet these challenges. It's built to scale up and out, drive operational efficiency through data minimization, and intuitively identify risk while reducing the total cost of compliance.

Actionable Outcomes: Data Protection
If data is the new currency, data protection is the central bank for an organization looking to protect its assets and preserve their worth. For Voltage Data Discovery, data protection is powered by our patented data security and encryption capabilities. This integration ensures that when sensitive data is detected, Voltage data security can protect and secure the data in place, in motion, and while in use. The flow of business continues securely as data is accessible for business users—and blocked or rendered harmless for unauthorized users.

Voltage data security integrated into the data discovery process flows protects data:

- **Inside structured business applications** where sensitive data is encrypted at field level using Voltage SecureData with format-preserving encryption (FPE), secure stateless tokenization (SST), or format-preserving hash (FPH).
- **Inside data lakes** where it enables secure data analytics for faster insights and reduced risk of a breach or proliferation of sensitive data.
- **Inside archived data repositories** where data is kept for data preservation and records compliance obligations. Sensitive data is protected inside the official record contained within a data archive ensuring data usage, permissions, and retention schedules are met.
- **At the endpoints** based on transparent and persistent policies beyond the edge of the network with Voltage SmartCipher file level encryption for unstructured data.

Voltage Data Discovery can mask or intelligently encrypt the database data in place by acting on production instances while ensuring data integrity. This allows you to manage not only archive databases, but the full relational database management system (RDBMS) lifecycle—protecting all data and adhering to the latest regulatory guidelines.

Actionable Outcomes: Data Preservation
One area of data discovery that is under-valued is the retention and preservation of sensitive and high-value data. Many global data privacy regulations require policies that track the business purpose for sensitive data. Voltage Data Discovery solutions help meet these compliance guidelines through the ability to control and map retention and preservation policies of sensitive data over its data lifecycle. Organizations who overlook this piece, or believe it is someone else's problem are at risk of non-compliance with records management obligations, and will be bound to time-consuming manual processes connecting the dots between data usage, and business purpose for Subject Rights Requests.

> A successful Data Discovery strategy not only requires the ability to connect and analyze disparate sources where sensitive data resides. It should support actionable steps that help drive and protect the business.

The Advantage of Voltage Data Discovery (Beyond Data Privacy) Is Delivering Actionable, Scalable Outcomes

Voltage Data Discovery solutions can help detect, protect, and evolve as your business transformation journeys unfold. Data Discovery is mission critical and a fundamental part of data minimization, data privacy readiness, data protection, and data preservation—and can act as a catalyst for building greater data resiliency and supporting your broader cyber resiliency programs. Voltage Data Discovery solutions include:

Voltage File Analysis Suite—Data Discovery
FAS Data Discovery SaaS file analysis solution enables organizations to quickly find, secure and protect sensitive and high-value data. FAS Data Discovery provides complete visibility and insight across unstructured data silos, helps contain data management costs, while delivering actionable analytics that improve efficiency, data quality and data privacy compliance. Contextually-aware, AI-driven grammars reduce false positives, and quickly identify high-value assets (e.g., contracts, intellectual

property, patents, etc.) personal and sensitive data types (e.g., PI/ PII, PCI, PHI, etc.).

Voltage Structured Data Manager—Data Discovery
Structured Data Manager (SDM) enables discovery, analysis, and classification of data and scanning for personal and sensitive data in any database accessible thru JDBC. SDM automates application lifecycle management and structured data optimization by relocating inactive data from expensive production systems and legacy databases, while preserving data integrity and access. With SDM, you can retire outdated applications through an automated process of extracting, validating, and deleting data. This unique solution significantly reduces capital expenses and administrative costs, improves end user productivity and overall IT staff efficiency, helps you to respond quickly to legal and compliance requests, and allows you to get more value from your data.

Voltage SmartCipher
Voltage SmartCipher simplifies unstructured data security, delivering control over the use and proliferation of sensitive files for secure collaboration and improved privacy compliance. It provides persistent file encryption, and complete control and visibility, over file usage and disposition across platforms. By combining critical technology features into a single solution for endpoint privacy and security, SmartCipher simplifies compliance and risk control with a single endpoint solution that transparently works with any datatype, on-premises or cloud solution, including transparent file encryption, usage controls, content inspection, and activity monitoring.

Voltage SecureMail
Voltage SecureMail is an essential part of any privacy compliance program or transition to Office 365 cloud. SecureMail is the best of breed end-to-end encrypted email solution available for desktop, cloud, and mobile that is scalable to millions of users while keeping Personally Identifiable Information (PII) and Personal Health Information (PHI) secure and private. SecureMail adds end-to-end encryption to Office 365, with flexible deployment options, additional compliance and collaboration features, ease of use, and privacy in the cloud.

Voltage SecureData
Micro Focus Voltage SecureData provides an end-to-end data-centric approach for enterprise data protection. By leveraging Voltage Format Preserving Encryption (FPE), Format-Preserving Hash (FPH), Secure Stateless Tokenization, and Stateless Key Management, SecureData protects sensitive structured data over its entire lifecycle—from the point at which it's captured and throughout its movement across the extended enterprise, without gaps in security. SecureData "de-identifies" data, rendering it useless to attackers, while maintaining its usability and referential integrity for data processes, applications, and services. SecureData enables the adoption of a continuous data protection model wherever data flows, in analytic platforms and applications in hybrid multi-cloud environments and native cloud-services.

Micro Focus Content Manager
Content Manager's data lifecycle management solution for both structured and unstructured data helps organizations meet data privacy, regulatory and productivity requirements. With more than 2,400 customers and nearly three million daily users, Content Manager provides a full range of content services, including document, retention, legal hold records and disposition management activities.
It can be deployed on-premises and from the cloud via a managed solutions provider. The solution provides high-value content management especially for government, life science, manufacturing, financial services, natural resource management, and telecommunications organizations.

Contact us at **CyberRes.com**
Like what you read? Share it.

CyberRes
A Micro Focus line of business