White Paper

Security

Example Architectures for Data Security and the GDPR

MICRO FOCUS

Table of Contents

Introduction

The European Union (EU) General Data Protection Regulation (GDPR) is the most significant development in data privacy in decades. Its aim is to protect EU citizens and residents from breaches of their personal data. The regulation comes into effect on 25 May 2018 and imposes heavy fines—up to 4% of annual revenue—on organizations for noncompliance.

GDPR mandates a number of specific measures to protect EU data subjects and their personal data including requiring certain cyber security measures, prompt notification in the event of a breach, and mandatory encryption of the most sensitive categories of personal data. In large measure, achieving compliance comes down to good data security.

The GDPR recommends pseudonymization and encryption as two mechanisms that can be used to protect personal data. Vast amounts of information exist on what data needs to be protected, though there is relatively little public knowledge about how an organization can deploy technologies and processes to secure this data.

This paper introduces typical business use cases for applying pseudonymization and encryption, provides an overview of the Voltage SecureData core technologies and platform from Micro Focus, and then describes architectures and strategies adopted by two of Micro Focus's customers to secure personal data with SecureData:

- A large European mobile operator that uses the SecureData data protection technologies to protect mobile subscriber information in a Hadoop data lake
- A global card brand and card issuer that leverages data protection to secure data as it is migrated to the cloud and uses the same architecture to protect sensitive customer personal data within its on-premises environment

Pseudonymization and Encryption: What's the Difference?

The GDPR specifically calls out the use of pseudonymization and encryption mechanisms as acceptable means for protecting personal data. Pseudonymization is often used as a general term that can apply to various techniques for data de-identification when the pseudonym or surrogate data can be used in business processes. Field-level encryption and tokenization are both examples of pseudonymization.

The GDPR is careful not to prescribe specific forms of encryption or pseudonymization. In the IDC white paper "Enabling GDPR Compliance Through Innovative Encryption and Key Management Approaches,"

reference is made to legacy encryption methods that render data unrecognizable and break business processes. However, GDPR calls out two important encryption features: the ability to decrypt the data when necessary and the ability to continue to run business processes on the encrypted data. Voltage Hyper Format-Preserving Encryption (FPE) exceeds these guidelines at enterprise scale.

Use Cases for Pseudonymization and Encryption

- **"Secure analytics" for data warehouses and Hadoop:** Big Data technologies including data warehouse platforms such as Teradata, Micro Focus® Vertica, and Hadoop hold seemingly unlimited promise to enable organizations to gain new analytic insights and operational efficiencies. Organizations are streaming, feeding, continuously analyzing, and storing sensitive data fields such as names, addresses, email addresses, geo locations, phone numbers, and card or bank account numbers within these platforms. Obtaining a return on investment from these platforms requires opening up the data to data scientists for analysis.

  However, expanding access to sensitive data exposes the organization to the risk of data breaches through insider theft, data mishandling, or the security of a third-party. Global organizations that collect data from points of presence in several European countries into a central repository or data lake are subject to additional GDPR and data residency issues. Passing Voltage FPE protected data into these platforms enables organizations to perform analytics on de-identified data. This approach reduces the risk of data breaches and enables the enterprise to maintain its compliance with regulations such as GDPR.
- **Migration to the cloud:** Organizations adopt cloud-computing strategies to gain significant market advantages and realize economic savings. For sensitive corporate and customer data such as medical or financial data, adopting new cloud capabilities imposes unique challenges, business risks, and compliance complications due to the nature of cloud architecture. Replacing identifiable data with an encrypted value narrows possible exposure of sensitive data and can greatly reduce audit scope and compliance costs.
- **Protecting data in live production systems:** Organizations store and process sensitive data in a number of production applications, databases, and systems. These systems are typically behind infrastructure and network-based security controls such as firewalls, access control lists, and database activity monitoring systems. Field-level data protection technologies ensure that attackers do not have access to real personal data when these security controls are inevitably breached. Only selected applications and authenticated, authorized users have the ability to decrypt data for real-time use. Other applications operate with Voltage FPE encrypted data to decrease the attack surface for retrieving sensitive personal data within an enterprise's infrastructure, lowering the organization's risk.
- **Development and test systems:** Generating data for development and test environments presents serious challenges for enterprise security and risk management. When data is copied from production databases and used directly, large volumes of private data accumulate on unprotected

servers and workstations. Control of the data is lost, exposing the enterprise to needless risk. Outsourced or offshore quality assurance and development services further exacerbate these risks. An alarming number of data breaches, along with regulatory compliance requirements such as GDPR, highlight the need to de-identify sensitive data when moving from production to test, development, and training environments. Passing encrypted data into these systems protects sensitive data against loss and theft while providing businesses with the agility required in their application development process.

Data Anonymization with Voltage Format-Preserving Hash

In specific use cases, such as Article 17—Right to erasure ('right to be forgotten') or in the creation of test data for example, the need to recover masked data may be an unnecessary risk—or further, may be explicitly undesired, as in the case of permanently enforcing right to be forgotten. Voltage Format-Preserving Hash (FPH) operates with the same benefits as FPE for structure, logic, partial field application and so forth, but with the added benefit of non-recovery of original data. This enables FPH to offer high-performance data usability—unlike traditional one-way transformation techniques, such as SHA-256—in a non-disruptive and more flexible approach toward data masking.

Technology Considerations for Encryption and Pseudonymization

There are a number of technical considerations for organizations looking to protect personal data using encryption and pseudonymization.

Voltage Format-Preserving Encryption

Organizations seeking GDPR compliance may have stored and processed sensitive personal data within various databases, applications, and systems for several years if not decades. Protecting this data with encryption using traditional techniques results in data that is incompatible with existing schemas, data structures, and processing requirements. Encrypting structured formatted fields such as customer names, national ID numbers, passport numbers, phone numbers, GPS locations, and dates of birth would require significant database schema and application changes to accommodate the protected data in its new format. Data decryption is then required for each analysis and use, decreasing overall security and imposing additional costs for key management.

Voltage FPE is a fundamental innovation enabling the Voltage SecureData data-centric platform to provide high-strength data encryption. Technical properties of data encrypted by Voltage Hyper FPE include:

- Retain format and structure
- Retain logical data structure such as checksums, date validity
- Retain partial non-sensitive values in encrypted fields (partial fields)
- Retain relationships with other fields and referential integrity where needed
- Retain the meaning in data and cross data relationships across records to preserve analytic meaning

These properties enable applications, analytic processes, and databases to use the protected data for the vast majority of use cases, even across distributed systems, platforms, and tools. Protection is applied at the field or partial-field level, leaving non-sensitive portions of fields available for applications while protecting the sensitive parts. Voltage FPE can, if required, preserve referential integrity across data sets so protected data can be consistently referenced and joined. This is especially critical where common identifiers such as phone numbers or IDs are used as references across disparate data sets.

Voltage FPE adheres to the AES-FF1 per the NIST SP-800-38G FPE standard[1] which Micro Focus helped pioneer. This provides enterprises with confidence in the security proofs and standards underpinning Voltage Hyper FPE.

Scaling Through Voltage Stateless Key Management

As organizations protect multiple applications and sensitive personal data types with encryption, they face increasing challenges with scaling their key management systems. Traditional encryption key management systems store encryption keys in a back-end database or vault, which causes scalability, backup, and disaster recovery issues. When dealing with field-level encryption across heterogeneous systems and multiple locations, traditional vault-based key management systems require continuous backup, synchronization, and protection which is burdensome and, itself, an increased security and compliance risk.

Voltage Stateless Key Management provides dynamically generated keys to be securely derived as needed with no storage or database management. Database synchronization and backups are not required, minimizing the risk of key loss. Voltage Stateless Key Management integrates with existing identity management infrastructure such as external LDAP directories. Permission to decrypt or detokenize can incorporate user roles and groups to simplify management based on identity management system policies.

Role-based, field-level data access empowers applications and users to view and use only the data they are authorized to access. The simplified implementation and high-performance, scalable, distributed processing of Voltage Stateless Key Management is well matched with modern application architectures.

Broad Platform Support

Customers looking for GDPR compliance have sensitive personal data in a number of different platforms and systems, including:

- Platforms such as Windows, Linux, HP-UX, Solaris, AIX, and others
- Databases such Oracle, DB2, Microsoft SQL Server, and others
- Mission-critical platforms such as z/OS, HPE NonStop, Stratus Virtual Operating System (VOS), and others
- Data warehouse platforms such as Teradata, Micro Focus Vertica, and leading Hadoop distributions
- Cloud platforms such as Amazon Web Services and Microsoft Azure
- Mobile devices that are based on iOS and Android

This data is also transported within disparate systems through extract, transform, and load (ETL) tools such as NiFi, Sqoop, Informatica, IBM DataStage, and Microsoft Server Integration Services (SSIS). Modern organizations also perform analysis on this data using a wide variety of business intelligence tools.

The main benefit of Voltage Hyper FPE as a field-level protection technology is that data can be protected using strong encryption as soon as it is captured and then the data stays protected at-rest, in-motion, and in-use as it is proliferated throughout an enterprise. It is critical that an organization planning to use

1 **nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-38G.pdf**

encryption for GDPR compliance considers solutions that provide native support for encryption and decryption on the widest number of platforms and systems.

The Voltage SecureData solution is typically deployed in two layers:

- Layer 1: The Voltage SecureData virtual appliances support authentication, authorization, key management, policy management, and integration with hardware security modules for root key hardware storage, used for key derivation. This layer provides secure and dual controlled web-based interfaces for managing, monitoring, auditing, and operating a deployed solution. It allows central management of data format policy for data encryption and tokenization, authentication and authorization controls, and central audit and monitoring of the modules in Layer 2.
- Layer 2: This layer includes a number of flexible and easy-to-use policy-controlled application programming interfaces (API), command line tools, file processor tools, database, and user-defined functions that can be used to encrypt or tokenize data. These tools are available on a number of platforms and are native on Windows, Linux, AIX, HP-UX, Solaris, various Hadoop distributions, Teradata, Micro Focus Vertica, z/OS, HPE NonStop, and Stratus VOS.

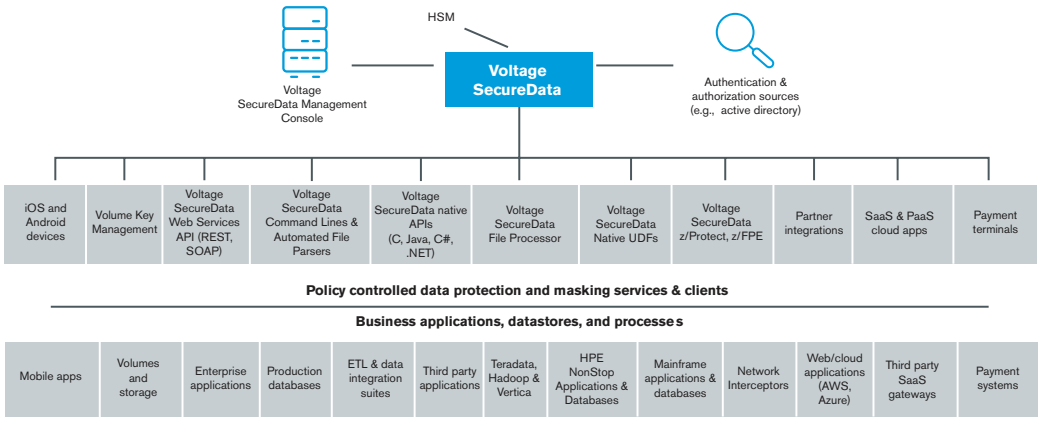

Figure 1. Layered Voltage SecureData architecture

The extensive cross-platform support and wide number of integration options provided by Voltage SecureData enable customers to perform encryption and decryption selectively as required by business process and applications. With Voltage FPE, by preserving the meaning and logic in the data, implementation is streamlined and simplified as most applications and processes can operate using encrypted data—so an implementation does not require application or process changes for the vast majority of cases. This dramatically simplifies deployments compared to traditional data encryption where integration and key management are invasive and complex.

SecureData Sentry Transparent Deployment to Accelerate Time to Value

With migration to hybrid IT and an increasing reliance on SaaS applications, organizations may not have the accessibility or development resources for API-level integration. Along with Layer 2 approaches, a data privacy broker solution, Voltage SecureData Sentry, enables a transparent encryption method by intercepting sensitive data flowing through the network to support on-premises and hybrid cloud-deployed applications. SecureData Sentry simplifies hybrid IT migration, accelerates time to value by quickly enabling security compliance, and offers consistency for end-to-end data protection, without having to break open applications and extensively re-qualify IT architectures.

Architectural Examples: Voltage Format-Preserving Encryption and GDPR

Large European Telco: Data De-Identification of Call Records in Hadoop for "Secure Analysis"

A large European mobile operator collects massive data sets from its mobile subscribers in a number of European countries. Data is moved to data centers in Germany and Italy for analysis in a 140-node Hadoop cluster. The operator expects to process over 11 billion records daily.

BUSINESS NEED

- Protect massive data sets including contact records, location, IMEI, IMSI, subscriber data, application, text, call data, and other personal data
- Comply with local data residency laws from multiple countries and GDPR
- Apply FPE to the personal data connected from various European points of presence to comply with data residency regulations and with GDPR while retaining the ability to analyze the data to detect access fraud, gain user pattern insights, and debug network fault scenarios

SOLUTION

The operator uses NIST-approved Voltage FPE as a technology to pseudonymize personal data within call records before it is analyzed in Hadoop. The components deployed as part of this solution are shown here:

1. **Voltage SecureData key servers:** These servers employ Voltage Stateless Key Management technology from their data centers in Germany and Italy. The architecture of Voltage SecureData enables these servers to be deployed in separate key jurisdictions. For example, this ensures that data processed in Germany is protected with keys generated in Italy, so a government authority seizing the data cannot seize the key servers required to identify that data.
2. **Voltage SecureData File Processor on a landing zone:** Many Hadoop architectures deploy a landing zone where incoming data is preprocessed, formatted, and normalized before ingestion into HDFS.[2] The operator deployed the Voltage SecureData File Processor tool on servers in its landing zone to perform FPE on personal data within files before storage in Hadoop. The Voltage SecureData File

2 **hortonworks.com/apache/hdfs/**

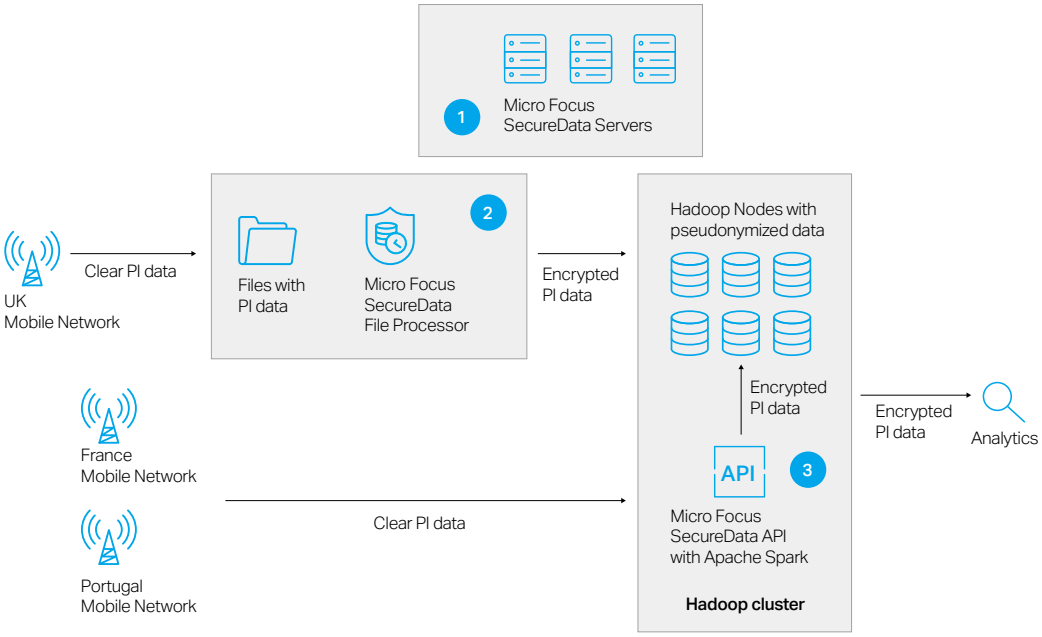

Figure 2. Data de-identification architecture in Hadoop

Processor tool encrypts sensitive fields within structured files of various formats including but not limited to comma separated, XML, JSON, record delimited, and positional.

3. **Voltage SecureData APIs integrated into Apache Spark:** The operator also uses Apache Spark for fast in-memory processing of data as it is ingested into Hadoop. They were easily able to integrate the Voltage SecureData Java APIs to encrypt sensitive personal data using Voltage FPE as it is ingested into Hadoop

The use of Voltage FPE guarantees referential integrity and enables pseudonymized data to retain its characteristics such as length and data type. The operator performs all their analysis on pseudonymized data, with no requirements to de-identify data to its original form.

BENEFITS

The deployment of Voltage SecureData provided the telco operator with a number of benefits, including

- Protection of their most valuable and vulnerable systems such as their Hadoop data lake
- Analytics system breaches no longer expose personal data and trigger notification requirements
- Compliance with several European data residency regulations including GDPR
- A single, enterprise-grade, scalable platform used to protect sensitive personal data within other platforms and systems

Global Card Brand: Encryption of Data for Moving to the Cloud

A global card brand and card issuer moves a number of applications to the Azure public cloud to reduce costs and capture a quicker time to market by enabling agile development strategies. Research showed one application storing personal data could realize over 50% savings in infrastructure by moving to the cloud. That transaction analysis application was written in the .NET language and was used by a number of contractors, posing security and cost concerns associated with allowing access to the card brand's network. Moving data in the clear to the cloud would introduce a number of risks including the possibility of a data breach, data jurisdiction challenges, and potential breach of compliance with regulations including GDPR.

Business Need

- Support for a large-scale hybrid infrastructure with a mix of legacy, enterprise, and cloud platforms. Support for operating systems such as z/OS, Windows, Linux, HPE NonStop, and database platforms such as Oracle, Microsoft SQL, and DB2. Support for data warehouse platforms such as Teradata and storage and processing platforms such as Hadoop
- Protect data immediately from specific applications as they are moved to the cloud
- Scale to protect billions of instances of personal data across hundreds of applications collecting, storing, and processing personal data

Solution

The card brand deployed Voltage Hyper FPE to protect data as it is moved to the cloud. The architecture deployed in the solution is shown here:

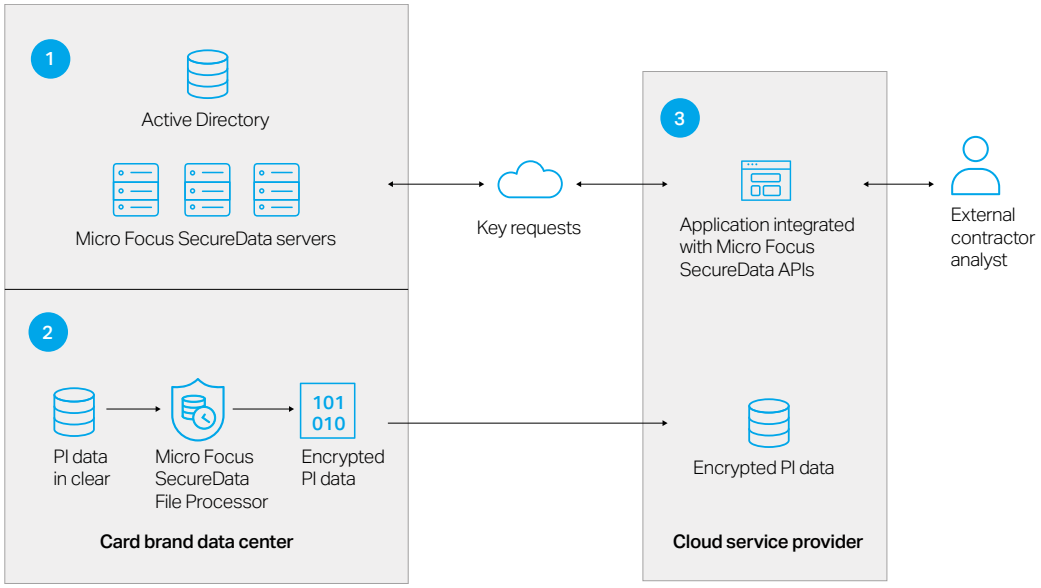

Figure 3. Data encryption architecture for the cloud

1. Voltage SecureData key servers: A global infrastructure of load-balanced Voltage SecureData key servers using Voltage Stateless Key Management was deployed within the card brand's network. Deploying key servers on-premises enabled full and complete control of encryption keys at all times, satisfying a number of internal security policies and external regulatory and compliance requirements.
2. Migration of application data to the cloud: The applications being moved to the public cloud had a significant amount of existing personal data within their databases. The Voltage SecureData File Processor and command line tools were used to convert personal data from its clear form to its format-preserved and encrypted form. This integration was performed using batch scripts with the ETL toolset.
3. Integration of cloud applications with Voltage SecureData APIs: The .NET-based card transaction analytics application was integrated with the Voltage SecureData APIs to perform role-based decryption of personal data on an as-needed basis by analysts. The Voltage SecureData APIs call back to the key servers deployed within the card brand's internal infrastructure to download keys for decryption. Each call is authenticated with the card brand's enterprise Active Directory infrastructure. All key requests and responses are centrally logged for alerting and reporting.

BENEFITS

The deployment of Voltage SecureData provided a number of benefits to the card brand:

- Immediately protected data within specific applications with the ability to scale
- Easily moved several dozens of applications to the cloud with significant cost savings
- Compliance with internal security standards and external regulations, such as GDPR

This deployment of Voltage SecureData has been extended to over 130 applications within the card brand's infrastructure. These include applications deployed on mainframe platforms, data warehouses, and Hadoop, as well as a number of distributed operating systems.

Conclusion

Complying with the GDPR regulations is driving organizations to adopt encryption and pseudonymization as techniques to protect customer personal data. Recent advances in technology such as Voltage Hyper Format-Preserving Encryption and Voltage Stateless Key Management are enabling customers to deploy highly scalable data protection technologies with minimal change to existing platforms and systems.

Organizations are taking a step-by-step approach to GDPR compliance, whereby the enterprise first protects sensitive personal information in their most vulnerable systems including Hadoop, data warehouses, and in applications deployed in the cloud. This provides an organization with a template to roll out data protection to other applications, platforms, and systems.

Additional contact information and office locations:
www.microfocus.com

www.microfocus.com