Hewlett Packard Enterprise

Example architectures for data security and the GDPR

Use cases for application of pseudonymization and encryption to protect data

Contents

Technical white paper

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 from privacy and data breaches. The regulation comes into effect on 25 May 2018 and imposes heavy fines—up to 4% of annual revenue—on organizations for noncompliance.

While the GDPR mandates a number of measures to protect EU citizen data such as data portability, consent and revocation, age verification, and the right to be forgotten, achieving compliance in large measure comes down to good data security.

The GDPR recommends pseudonymization and encryption as two mechanisms that can be used to protect personally identifiable information (PII). Vast amounts of information exist on **what** data needs to be protected, though there is very 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 HPE SecureData core technologies and platform, and then describes architectures and strategies adopted by two of HPE's customers to secure PII data:

- A large European mobile operator that uses the HPE 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 PII 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 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. HPE Format-Preserving Encryption (HPE 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, HPE 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 HPE FPE protected data into these platforms enables organizations to perform analytics on de-identified data. This approach reduces the risk of data breaches and keeps the enterprise in 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 PII when these security controls are inevitably breached. Only selected applications and users that have been authenticated and authorized have access to decrypt data for use, in real time. Other applications operate with HPE FPE encrypted data to decrease the attack surface for retrieving sensitive PII 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.

Technology considerations for encryption and pseudonymization

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

HPE Format-Preserving Encryption

Organizations seeking GDPR compliance may have stored and processed sensitive PII data within various databases, applications, and systems for several years if not decades. Protecting this data with encryption using traditional techniques results in data 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 security overall and imposing additional costs for key management.

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

- Retain format and structure
- Retain logical data structure such as checksums, date validity
- Retain partial nonsensitive values in encrypted fields (partial fields)
- Retain relationships to 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 nonsensitive portions of fields available for applications while protecting the sensitive parts. HPE 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.

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

Scaling through HPE Stateless Key Management

As organizations protect multiple applications and sensitive PII 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 across multiple locations, traditional vault-based key management systems require continuous backup, synchronization, and protection that is burdensome and itself an increased security and compliance risk.

HPE 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. HPE 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 HPE Stateless Key Management is well matched with modern application architectures.

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

Broad platform support

Customers looking for GDPR compliance have sensitive PII 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, HPE 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 HPE 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 encryption for GDPR compliance considers solutions that provide native support for encryption and decryption on the widest number of platforms and systems.

The HPE SecureData solution is typically deployed in two layers:

- **Layer 1:** The HPE 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, HPE Vertica, z/OS, HPE NonStop, and Stratus VOS.

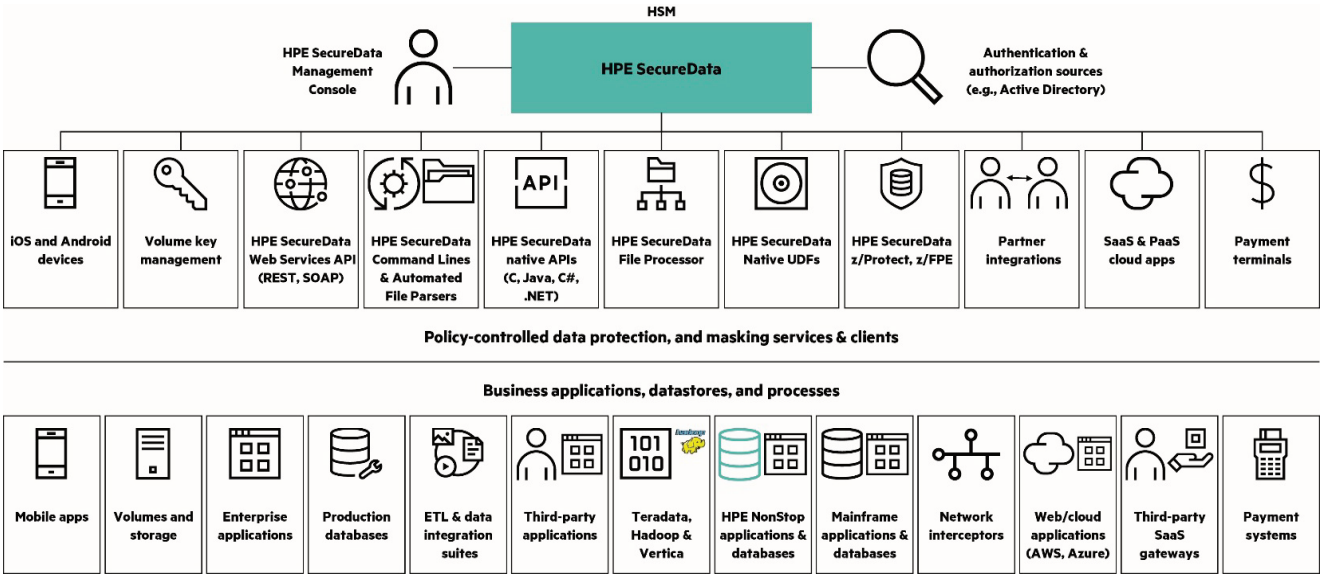

Figure 1. Layered HPE SecureData architecture

The extensive cross-platform support and wide number of integration options provided by HPE SecureData enable customers to perform encryption and decryption selectively as required by business process and applications. With HPE 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.

Architectural examples: HPE 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 PII data
- Comply with local data residency laws from multiple countries and GDPR
- Apply FPE to the PII 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 HPE FPE as a technology to pseudonymize PII data within call records before it is analyzed in Hadoop. The components deployed as part of this solution are shown here:

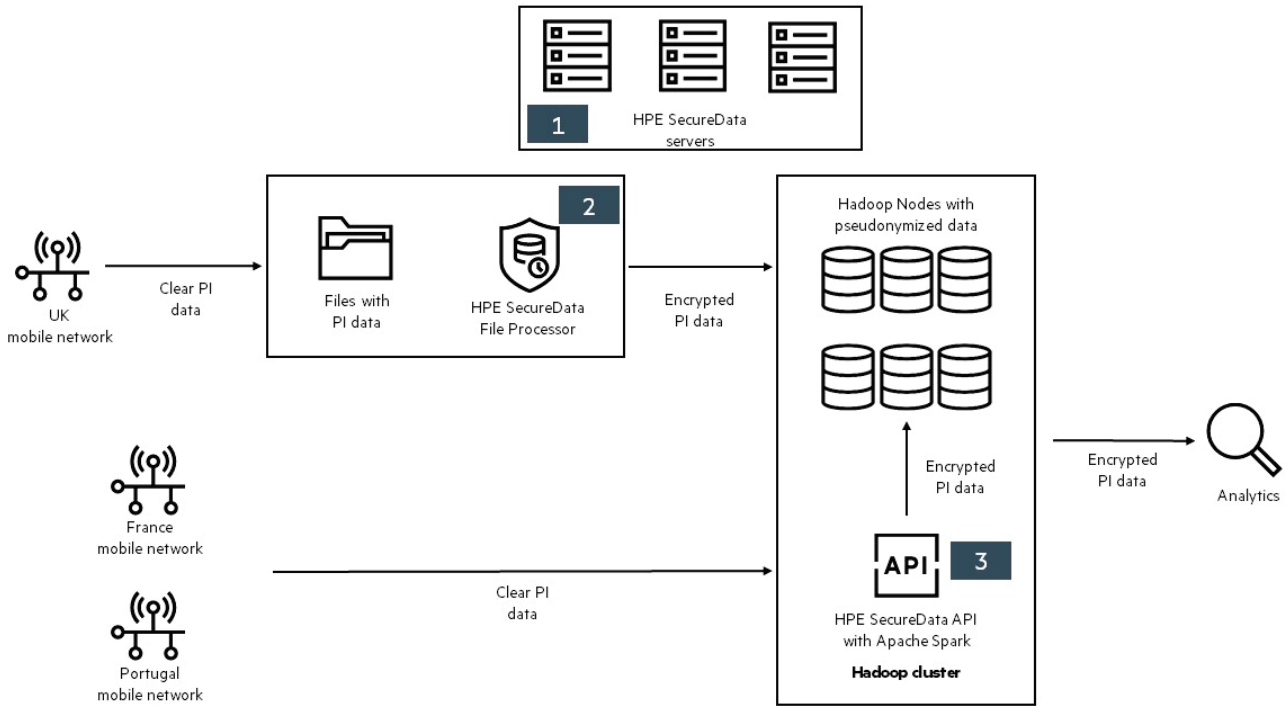

Figure 2. Data de-identification architecture in Hadoop

1. **HPE SecureData key servers:** These servers employ HPE Stateless Key Management technology from their data centers in Germany and Italy. The architecture of HPE 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. **HPE 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 HPE SecureData File Processor tool on servers in its landing zone to perform FPE on PII data within files before storage in Hadoop. The HPE SecureData File 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. **HPE 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 HPE SecureData Java APIs to encrypt sensitive PII data using HPE FPE as it is ingested into Hadoop.

The use of HPE 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 HPE 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 PII data and trigger notification requirements
- Compliance with several European data residency regulations including GDPR
- A single, enterprise-grade, scalable platform used to protect sensitive PII 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 PII 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 PII data across hundreds of applications collecting, storing, and processing PII data

[2] hortonworks.com/apache/hdfs/

Solution

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

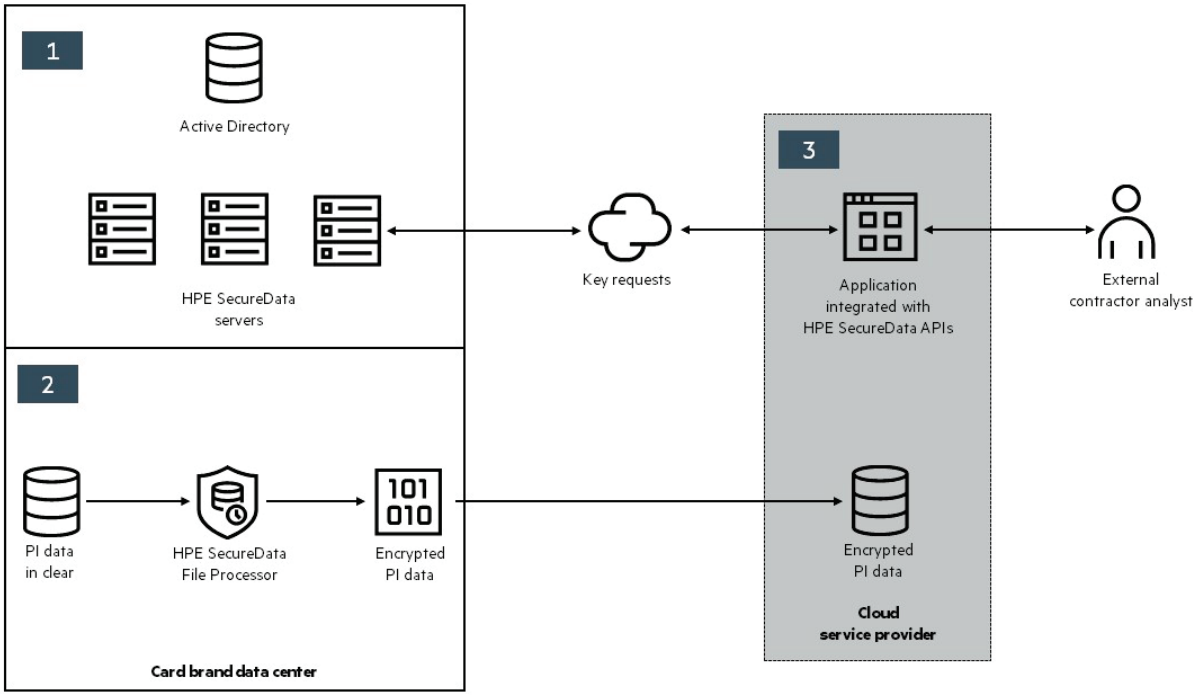

Figure 3. Data encryption architecture for the cloud

1. **HPE SecureData key servers:** A global infrastructure of load-balanced HPE SecureData key servers using HPE 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 compliance requirements.
2. **Migration of application data to the cloud:** The applications being moved to the public cloud had a significant amount of existing PII data within their databases. The HPE SecureData File Processor and command line tools were used to convert PII 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 HPE SecureData APIs:** The .NET-based card transaction analytics application was integrated with the HPE SecureData APIs to perform role-based decryption of PII data on an as-needed basis by analysts. The HPE 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 HPE 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 HPE 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 PII data. Recent advances in technology such as HPE Format-Preserving Encryption and HPE 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 PII 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.

Learn more at

voltage.com
hpe.com/software/datasecurity

Sign up for updates

© Copyright 2017 Hewlett Packard Enterprise Development LP. The information contained herein is subject to change without notice. The only warranties for Hewlett Packard Enterprise products and services are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. Hewlett Packard Enterprise shall not be liable for technical or editorial errors or omissions contained herein.

Microsoft and Windows are either registered trademarks or trademarks of Microsoft Corporation in the United States and/or other countries. Oracle and Java are registered trademarks of Oracle and/or its affiliates. Linux is the registered trademark of Linus Torvalds in the U.S. and other countries. All other third-party trademark(s) is/are property of their respective owner(s).

a00001230ENW, February 2017

Hewlett Packard Enterprise