comforte

DATA-CENTRIC SECURITY GUIDE

EVALUATING PROTECTION TECHNIQUES FOR ENTERPRISE DATA IN 2025

www.comforte.com

Contents

Purpose of this eBook

This eBook serves as a comprehensive, technically focused guide to data-centric security, designed for security leaders and technical teams. It provides insights into how organizations can better protect their sensitive data by leveraging both traditional and emerging security techniques. Key methods discussed include encryption, hashing, data masking, tokenization, and cutting-edge privacy technologies such as Homomorphic Encryption.

The primary objective is to equip security and data teams with the tools and knowledge needed to protect sensitive data proactively. By comparing established practices with innovative technologies, we highlight how data-centric security not only ensures regulatory compliance but also drives business value and mitigates risk.

This eBook aims to:

> **Educate**	Offer a clear understanding of the principles of data-centric security and their importance in today's threat landscape.
> **Evaluate**	Provide a comparative analysis of the related protection methods, helping organizations align techniques with their control framework and security needs.
> **Empower**	Equip decision-makers to safeguard data effectively while enabling secure business innovation.

01

UNDERSTANDING DATA-CENTRIC SECURITY

> **Data-centric security** is a cybersecurity approach focused on protecting the data itself at every stage of its lifecycle rather than only securing the networks, applications, or systems that hold or transfer it.

The Evolving Data Security Landscape

Organizations increasingly rely on data to fuel innovation, make informed decisions, and achieve operational excellence. However, this dependency presents intensifying challenges in safeguarding sensitive information while maintaining efficiency.

The modern security landscape is marked by rising complexity and sophisticated cyber threats. In 2024, we have seen the global average data breach cost rise by 10%, reaching USD 4.88 million—the most significant increase since the pandemic. Personally identifiable information (PII) has become a prime target for attackers, accounting for 46% of exposed data in breaches, further underscoring the critical need for robust enterprise-wide data protection strategies.[1]

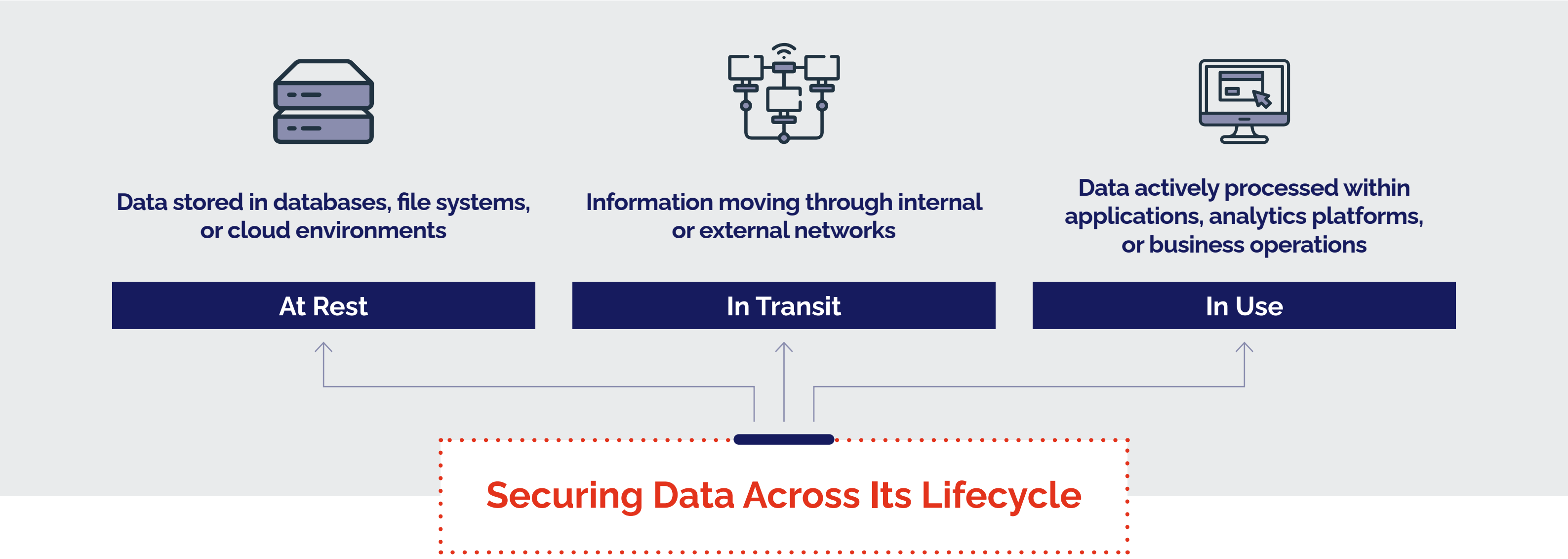
Security and business leaders face a difficult choice: impose stringent data controls that could limit innovation or choose less restrictive measures that increase security risks.

Data-centric security represents a paradigm shift, focusing protection on an organization's most valuable asset—its data. This strategy keeps sensitive data secure across environments, aligning protection with key business objectives like compliance, efficiency, and data-driven decisions while enabling innovation and safeguarding critical assets.

The data-centric security market is expected to reach USD 16.7 billion by 2030, with a compound annual growth rate (CAGR) of 16.8%.[2]

[1] *Cost of a Data Breach Report* 2024, IBM, 2024

[2] *Data-Centric Security - Global Strategic Business Report*, Global Industry Analysts, Inc. 2024


Data stored in databases, file systems, or cloud environments
At Rest
Information moving through internal or external networks
In Transit
Data actively processed within applications, analytics platforms, or business operations
In Use
Securing Data Across Its Lifecycle

Data-centric security moves beyond traditional perimeter defenses by prioritizing the protection of the information, rather than merely securing the underlying infrastructure. This approach ensures that data is secured at every stage of its lifecycle—whether at rest, in transit, or in use—offering consistent protection regardless of location or access method.

The four essential components of data-centric security include:

01 **Data Discovery and Classification** Identify and categorize sensitive data by location and importance to support effective risk analysis and prioritize protection.

02 **Data Access Control and Permissions** Define access controls based on roles and least privilege, ensuring only authorized individuals can access specific data.

03 **Data Protection** Apply safeguards such as encryption and tokenization to secure data at rest, in transit, and in use, rendering it unusable to unauthorized entities.

04 **Data Auditing and Threat Detection** Continuously monitor data usage to detect anomalies or unauthorized access in real-time.

Key Principles:

1. **Protect data as early as possible in its lifecycle** Secure data through all stages to prevent unauthorized access, while keeping it usable within systems.

2. **De-protect data only when absolutely necessary** Ensure security measures remain with the data, preserving its protection regardless of where it is stored or how it is transmitted.

Key Drivers of the Shift Towards Data-Centric Security

> **Exponential Data Growth**
Unprecedented volumes of data require focused efforts to protect sensitive information across its movement.

> **Data Decentralization**
Cloud services, IoT, and remote work have led to more distributed data environments, reducing the effectiveness of traditional perimeter defenses.

> **Business Imperatives**
Privacy-preserving security enables secure analytics, AI, and insights, driving innovation and competitive advantage.

> **Evolving Threat Landscape**
Sophisticated attackers bypass traditional defenses, emphasizing the need to secure the data itself.

> **Regulatory and Compliance Pressures**
Frameworks such as GDPR, CCPA, and PCI DSS require stringent data protection measures to maintain compliance and avoid penalties.

> **Data Privacy Concerns**
Increased public awareness of privacy necessitates robust data protection to maintain trust and prevent reputational harm.

Limitations of Traditional Perimeter-Based Security Models

Perimeter-focused security models, reliant on firewalls and endpoint protection, are increasingly vulnerable to advanced attacks, underscoring the need for a comprehensive control framework and attack surface reduction.

Traditional perimeter-based security models were designed for an era when data was centralized in on-premises data centers. These models relied heavily on securing the network's edge, assuming that strong perimeter defenses would adequately safeguard the data within. However, modern threats and technological shifts have exposed significant weaknesses in this approach.

Advanced attack techniques, such as lateral movement, supply chain compromises, and social engineering, easily exploit vulnerabilities in perimeter defenses. Once attackers breach the network, they often move undetected for extended periods, accessing sensitive data and systems. On average, organizations take over 200 days to detect a breach and an additional 73 days to contain it, a delay that compounds the overall impact.[3]

Data-centric security aligns with a Zero Trust strategy by considering and verifying every interaction with data. This approach reduces data exposure and mitigates risks, even when other security layers fail. Relying solely on perimeter defenses in today's landscape is like locking the front door while leaving the windows wide open—an invitation for attackers to exploit unchecked vulnerabilities.

Core Benefits of a Data-Centric Approach

Enhanced Security	Directly protects data, addressing legacy vulnerabilities and defending against advanced threats, strengthening overall security.
Support for Data-Driven Innovation	Enables the secure use of sensitive data for analytics, AI, and insights balancing innovation with privacy and security.
Operational Efficiency	Reduces administrative burden and optimizes workflows, lowering costs to focus resources on critical tasks.
Comprehensive Threat Mitigation	Protects against external attacks, insider threats, and accidental data leaks, maintaining security even after breaches.
Streamlined Regulatory Compliance	Aligns with regulatory requirements, reduces penalties, and enhances organizational security posture.

Talk to an Expert Today

Schedule a consultation to explore how data-centric security adds value.

[3] *Cost of a Data Breach Report* 2024, IBM, 2024

02

IMPORTANT ATTRIBUTES OF DATA-CENTRIC PROTECTION METHODS

When evaluating a data protection solution, organizations must balance security, regulatory compliance, operational efficiency, and usability.

Traditional data protection approaches, such as encryption and hashing, have long served as essential tools for safeguarding sensitive data. Recently, newer techniques such as tokenization have gained traction.

As organizations navigate a growing landscape of data protection options, the complexities surrounding their distinctions, overlaps, and broader system implications can be overwhelming. Misunderstandings about specific implementations—often exacerbated by inconsistent or misleading marketing—can cloud the decision-making process. These factors make it difficult for companies to accurately assess which solution best aligns with their unique requirements.

Each method offers distinct technical characteristics that influence its effectiveness and suitability for different use cases. Below are some critical attributes to consider, including reversibility, key management, determinism, format preservation, referential integrity, data utility, and collision resistance.

Reversibility: Balancing Privacy and Utility

What use cases require data restoration, and how frequently?

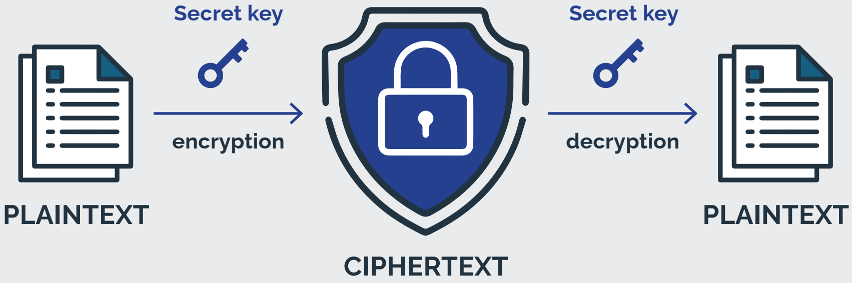
Reversibility is a key consideration in data protection and defines whether original data can be restored after being protected. Reversible techniques allow authorized parties to restore data to its original form, which is vital in use cases like financial transactions. In contrast, irreversible methods—such as hashing—alter the data permanently, making restoration impossible. These methods are suited for applications focused on privacy, such as anonymizing customer data for analysis, where data recovery isn't required.

Secret Symmetry: Key Management Strategies

How should key access be controlled to minimize the risk of unauthorized decryption?

01

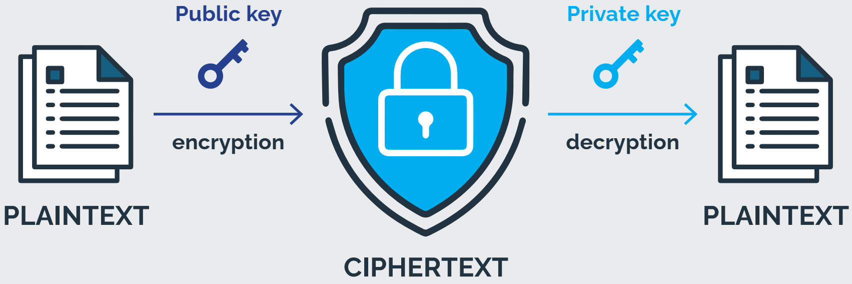
Symmetric encryption uses the same key for both encryption and decryption. As a result and by design, all involved parties with access to the protection secret can both protect and deprotect data.



Symmetric Encryption

02

Asymmetric encryption involves a pair of keys: a public key for encryption and a private key for decryption. This separation allows encryption without revealing the private key, limiting decryption capabilities to authorized entities only.



Asymmetric Encryption

Post-Quantum Cryptography: Preparing for a Quantum Future

What is post-quantum cryptography and how can it protect against future threats?

Post-quantum cryptography focuses on developing encryption methods that remain secure in the face of quantum computing capabilities. Quantum computers have the potential to quickly break current asymmetric algorithms like RSA and ECC using Shor's algorithm, which could undermine encryption protocols such as HTTPS, SSL, and TLS that are essential for securing online communications and transactions.

This risk gives rise to the "Harvest Now, Decrypt Later" threat, where attackers store encrypted data with the intention of decrypting it once quantum computing becomes fully operational. To address this, researchers are working on quantum-resistant algorithms designed to withstand quantum attacks. As these technologies develop, organizations must begin transitioning to post-quantum systems to protect their data in a future where quantum computing is a reality.

Determinism in Data Protection

Which business processes require deterministic outputs, and how would they be affected by non-deterministic methods?

Deterministic protection ensures that the same input produces the same output every time it is protected with the same key, critical in systems that require data consistency, such as databases. Non-deterministic methods generate different outputs for the same input, even when the same key is used. While this enhances security by making patterns harder to detect, it complicates processes where consistent data references are necessary.

Ensuring Consistent Data Relationships

Referential integrity is a resulting property of any deterministic protection technique. It ensures that relationships between data elements are preserved after protection mechanisms are applied. This is crucial in environments where different datasets or systems need to maintain consistent links between protected data elements, such as in banking, where broken data relationships can undermine important business operations such as fraud detection.

Format Determination

Which applications require data to maintain its original format, and why is this critical?

Format Determination maintains the desired format of the protected data elements during the protection operation. This includes preserving the length and the character set of the underlying sensitive data, essential where data needs to flow through systems that expect specific formats, such as transactional databases.

Format-preserving protection techniques maintain the usability of sensitive information by ensuring it remains in a format that allows for ongoing analysis, processing, and sharing. This approach minimizes the need for system adjustments, reducing operational disruption while maintaining security. It ensures that data protection does not compromise business functionality or the ability to leverage data for analytical purposes.

Collision Resistance: Safeguarding Data Integrity

How do collision vulnerabilities affect the security of your applications and databases?

Collision resistance ensures different inputs do not result in the same protected output. This is important for applications that rely on precise data mapping, such as relational databases. Strong collision resistance ensures that even with the same key, different data elements produce unique encrypted values. If using the same secret on the same data element results in different protected elements over time, this is a sign that a specific protection method is not collision-free/resistant. The same is true if the deprotection operation yields different results over time.

03

EXPLORING DATA-CENTRIC PROTECTION TECHNIQUES

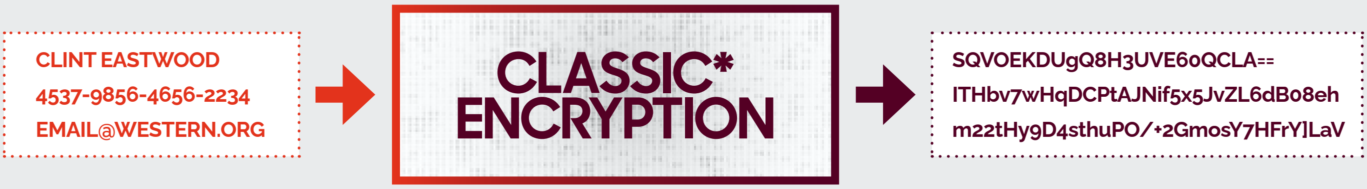
> When it comes to sensitive information, a data-centric approach is proven to have fewer security gaps and risks, as security travels with the data whether at rest, in use, or in transit, rather than relying on additional security methods when data moves beyond applications, databases, or storage systems.

When implementing a data-centric approach to safeguarding sensitive information, organizations must evaluate the security benefits and operational trade-offs of each data protection technique. This chapter explores common data protection methods, their impact on data properties, technical considerations, and relevant use cases, while also addressing performance and regulatory compliance to ensure a secure and compliant environment.

Protection Method Properties

	Format-Preserving Hashing	Data Masking	Classic Encryption	FPE / Tokenization	Synthetic Data	Fully Homomorphic Encryption
Reversibility Ability to retrieve original data.	non-reversible	non-reversible	reversible	reversible	non-reversible	reversible
Format Preservation Maintains original data format (e.g., numeric, date).	Yes	Yes	No	Yes	Yes (in some cases)	No
Collision Resistance Likelihood of different inputs yielding the same output.	Possible	Possible	Unlikely	Unlikely	Possible	Possible
Referential Integrity Preservation Keeps relationships between data.	Partial	No	Yes	Yes	Partial	Yes
Maturity Level Level of proven reliability and enterprise adoption.	High	High	Very High	High	Moderate	Emerging
Computation Capability on Protected Data Ability to compute on protected data.	No	No	No	No	Partial	Yes
Compliance Suitability for PII/PCI/PHI Meets regulatory requirements for sensitive data	Partial	Yes	Partial	Yes	Partial	No
Data States Protected If the data is stored, transmitted, or actively used	Data at Rest	Data at Rest	Data at Rest, in Transit	Data at Rest, in Transit, in Use	Data at Rest, in Transit, in Use	Data in Use
Suitability for High-Sensitivity Data Appropriate for highly sensitive data types (e.g., financial, medical).	Moderate	Moderate	High	High	Moderate	High

Classic Encryption



Algorithm: Des; Mode: CBC; Output encoded using Base64

Encryption is a fundamental data-centric protection method that converts plaintext (readable data) into ciphertext (encrypted data), ensuring that only authorized users with the decryption key can access the original information. This process relies on cryptographic algorithms and keys to secure data, allowing ciphertext to be reverted back into plaintext by those with the appropriate key.

Classic encryption provides both confidentiality and data integrity. It is deterministic and collision-free, meaning identical inputs consistently generate the same ciphertext, while different inputs produce unique outputs.

There are two primary forms of this protection:

01	**Symmetric encryption**	Which includes standards like Advanced Encryption Standard (AES) and Data Encryption Standard (DES).
02	**Asymmetric encryption**	Represented by algorithms such as RSA and Elliptic Curve Cryptography (ECC).

Use Cases for Classic Encryption

Classic encryption is primarily employed to secure data at rest (e.g., databases, devices) and data in transit (e.g. network transmissions).

Examples include full-disk encryption, file-level encryption, and database encryption to protect data at rest. For data in transit, protocols such as SSL/TLS establish secure communication channels, while end-to-end encryption safeguards information against unauthorized access during transmission.

A significant drawback of traditional encryption is that it changes the data structure, which can cause compatibility issues with systems requiring specific formats. For instance, systems designed to handle email addresses may struggle with encrypted data. To address these challenges, **format-preserving encryption (FPE)**—also known as encryption-based tokenization—allows for data protection while maintaining the original format. For example, a 16-digit credit card number remains a 16-digit number, ensuring compatibility with legacy systems. Further details on FPE are provided later in this report.

Key Considerations of Classic Encryption

> **Data Structure Changes**	Encrypted data typically has a different format and length compared to its original form, potentially causing system compatibility issues.
> **Workflow Disruption**	Applications must decrypt data before using it, which can temporarily expose sensitive information and interrupt workflows.
> **Key Rotation**	Key rotation periodically replaces encryption keys to reduce compromise risk but presents challenges, especially in high-volume or real-time systems. Without proper planning, it may lead to data unavailability and downtime.

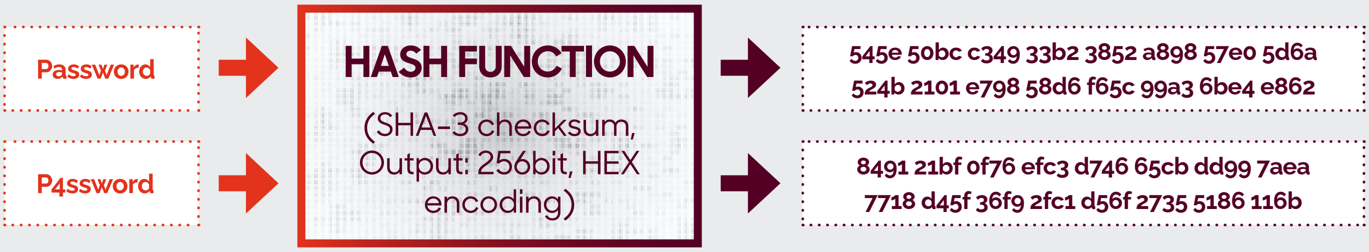
Synthetic Data

Synthetic data is artificially generated information designed to replicate the statistical properties and distributions of real-world datasets without containing any actual sensitive data. It can be generated through methods such as sampling, which draws values from statistical distributions of existing datasets, and simulation, which creates new datasets based on mathematical models or rules that mirror real-world processes. It is a valuable tool for organizations seeking to support testing, development, and analytics in controlled environments where privacy and compliance are critical.

Despite its advantages, synthetic data has notable limitations:

> **Complexity Gaps**	It may fail to fully capture the intricacies and edge cases present in real-world datasets, impacting the accuracy and reliability of applications trained on it.
> **Model Dependency**	The quality of synthetic data depends heavily on the strength of the models used for its generation. Weak models can introduce biases or fail to represent the diversity of real-world scenarios, limiting its effectiveness in certain contexts.

Hashing


Password
P4ssword
HASH FUNCTION
(SHA-3 checksum, Output: 256bit, HEX encoding)
545e 50bc c349 33b2 3852 a898 57e0 5d6a 524b 2101 e798 58d6 f65c 99a3 6be4 e862
8491 21bf 0f76 efc3 d746 65cb dd99 7aea 7718 d45f 36f9 2fc1 d56f 2735 5186 116b

Password Hashing

Hashing is a one-way, deterministic method that converts data into a fixed-size output called a hash or digest. Unlike encryption, it is irreversible, meaning data cannot be restored to its original form. It's commonly used for data integrity verification and securely storing sensitive information, such as passwords.

Hash functions are designed to be collision-resistant, meaning it is highly unlikely for two different inputs to produce the same output. Once data is converted, it becomes computationally infeasible to reverse-engineer the original input from the digest, making this approach particularly effective for protecting sensitive information.

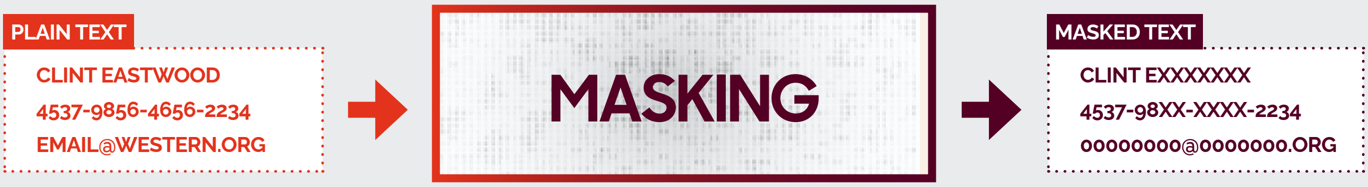
Secure Password Storage

To protect user credentials, systems store hashed versions of passwords instead of plaintext. When a user attempts to log in, the system processes the entered password to generate a corresponding hash and compares it with the stored value. This method relies on the deterministic nature of Hashing. As a result, password validation can be performed by comparing the hashes, eliminating the need to store or manage plaintext passwords.

Format-Preserving Hashing (FPH)

Traditional hash functions, such as SHA-256, generate fixed-size outputs that do not retain the original format of the data, making it difficult to integrate into systems that depend on specific data structures. Format-preserving hashing (FPH) overcomes this limitation by producing hashes that preserve the original format, enabling seamless integration into these systems while maintaining the security of standard hashing methods.

Data Masking



PLAIN TEXT

CLINT EASTWOOD
4537-9856-4656-2234
EMAIL@WESTERN.ORG

MASKING

MASKED TEXT

CLINT EXXXXXXX
4537-98XX-XXXX-2234
00000000@0000000.ORG

Data Masking

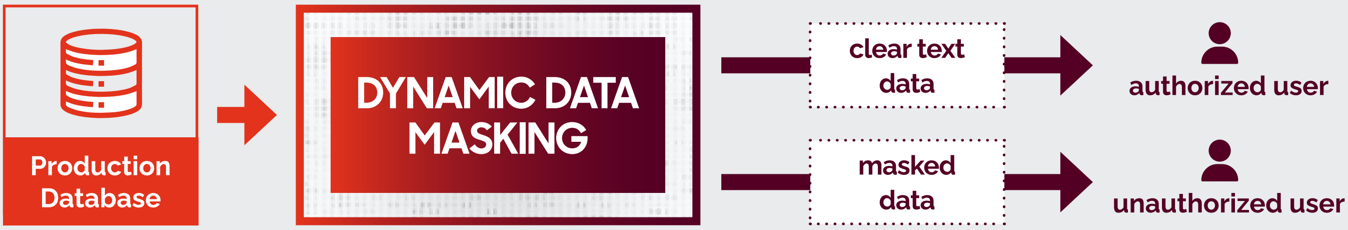
Data masking replaces sensitive data with fictional but realistic data, allowing for testing, development, or analysis without exposing the original data. Techniques such as substitution (replacing data with random but similar values), shuffling (rearranging data within a column), randomization (slightly altering numeric or date fields), and character masking (hiding portions of data) ensure that the data remains usable for various applications while keeping sensitive information hidden.

Given the irreversible nature of this protection method, masking typically does not involve any secrets. While the format of the masked value is highly configurable, there is a substantial risk of collision since there is no one-to-one correlation between masked values and their plaintext equivalents.

Data Masking Example

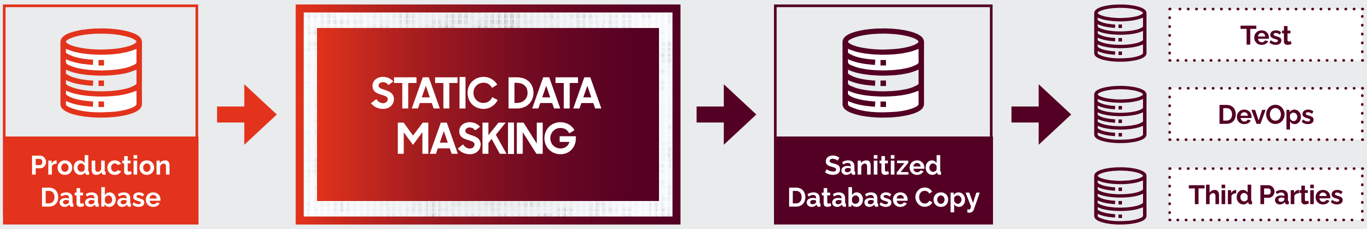
Customer Support at a Financial Institution

When a customer disputes a credit card transaction, the customer service agent needs to verify which card was used. However, it's unnecessary for the agent to have access to the full credit card number. By applying data masking, the system only reveals the last four digits of the card number, enabling the agent to identify the card in question while ensuring the full card details remain protected.



Dynamic Data Masking (DDM)

DDM provides real-time data protection by altering sensitive information based on user permissions and request types. Implemented via an application or agent, it intercepts data streams and replaces sensitive values according to predefined rules. While DDM ensures that sensitive data is protected during access, it does not modify the data stored in the underlying database. As a result, additional security measures, such as encryption or tokenization, are necessary to protect data at rest.

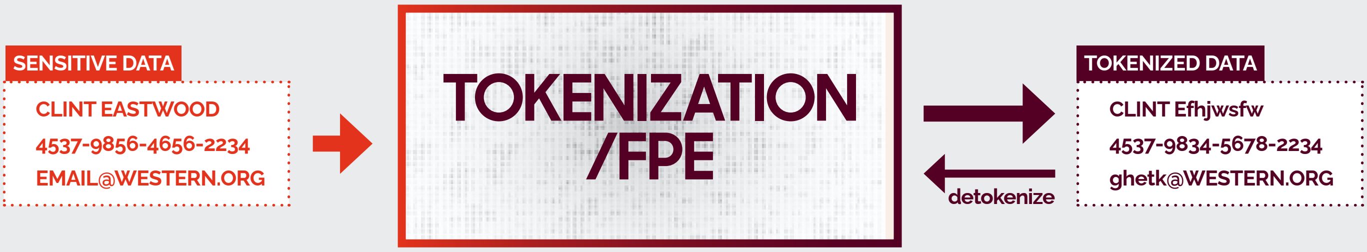


Static Data Masking (SDM)

SDM creates anonymized copies of sensitive data in non-production environments, such as for testing or analytics. It permanently transforms sensitive data while retaining its original structure, ensuring that it cannot be reverted to its original form. SDM is applied only to data at rest in non-production settings, leaving the original data intact in production environments. To ensure full protection, SDM should be complemented with other security methods to safeguard the underlying sensitive information.

Masking methods, including DDM and SDM, do not protect the original data at rest. It is critical to implement additional security measures, such as encryption or tokenization, to protect sensitive information effectively.

Tokenization



Data Tokenization

Tokenization replaces sensitive data elements with non-sensitive equivalents known as tokens, which retain the original format and structure. These tokens are meaningless without access to the secure tokenization system, making this method highly effective for protecting payment data, PII, and other sensitive datasets. When integrated into an organization's control framework, tokenization reduces the attack surface and prevents unauthorized access to sensitive information.

Tokenization relies on secure mappings maintained by a dedicated tokenization system. This system ensures that identical inputs always yield the same token, preserving referential integrity across datasets. Deterministic and collision-free, tokenization techniques allow enterprises to process or analyze data securely, without exposing sensitive values.

While tokenization is reversible, this process—often linked to pseudonymization in privacy contexts—restores the original data only when absolutely necessary. By replacing sensitive data with tokens during routine operations, tokenization enables secure business workflows without compromising usability.

Tokenization secures data throughout its entire lifecycle—whether the data is at rest, in transit, or in use.

Tokenization in Data-Centric Security

It's important to distinguish **data-centric tokenization** in enterprise environments from broader tokenization uses. In blockchain, tokenization converts physical assets into digital tokens on a distributed ledger. Digital wallets like Apple Pay, replace payment data with non-sensitive tokens for transaction security. For enterprises, data-centric tokenization protects sensitive data while enabling its use in business processes, analytics, and compliance. Recognizing these differences helps organizations select the right approach to meet data protection, business, and regulatory needs.

Poker Chip Analogy

A useful analogy for understanding tokenization is the use of poker chips in a casino. Imagine a bustling casino where players use chips instead of cash. By using chips as intermediaries, players can avoid handling large sums of cash, which can be easily lost or stolen.

While poker chips can be exchanged for cash, the actual money remains securely stored by the casino, out of sight and reach of the players. The chips themselves have no real value outside the casino; they are simply pieces of plastic. Similarly, tokens in data-centric security cannot be exploited for any intrinsic value on their own; they require a secure mechanism to be linked back to the sensitive information they represent. In this way, poker chips act as stand-ins, protecting the actual money while allowing the games to continue seamlessly, much like how tokens safeguard sensitive data while enabling its secure use in various applications.

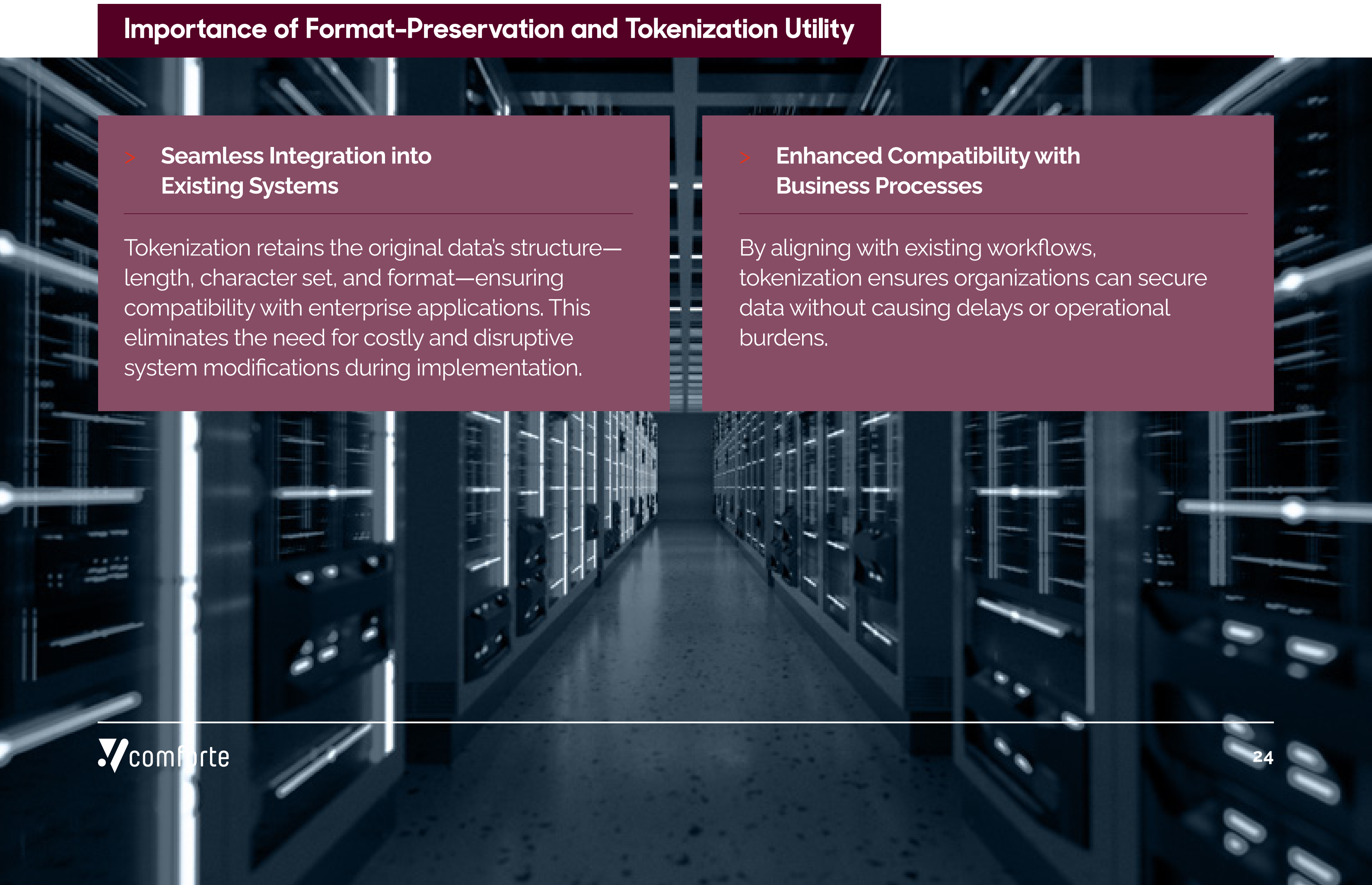
Tokenization is used to secure a wide range of sensitive data types, including:

- Payment card information
- Bank account details
- Social Security numbers (SSNs)
- Phone numbers
- Passport numbers
- Driver's license information
- Email addresses
- Health records and personal health information (PHI)
- Intellectual property (IP)

Common use cases for tokenization include:

Payment Processing	Reduces fraud risks by eliminating the need to store cardholder information.
PCI DSS Compliance	Limits storage of Primary Account Numbers (PANs) and sensitive financial data, reducing the scope of audits.
Financial Services	Secures customer data, such as bank account numbers, supporting safe transactions and compliance with industry regulations.
Business Intelligence and Data Analytics	Allows usage and analysis of sensitive data without exposing real values.
Cloud Strategy and Data Sharing	Secures data before migrating to public clouds or sharing with third parties.
Regulatory Compliance	De-identifies sensitive data to meet GDPR, CCPA, and HIPAA requirements.

Importance of Format-Preservation and Tokenization Utility



> Seamless Integration into Existing Systems

Tokenization retains the original data's structure—length, character set, and format—ensuring compatibility with enterprise applications. This eliminates the need for costly and disruptive system modifications during implementation.

> Enhanced Compatibility with Business Processes

By aligning with existing workflows, tokenization ensures organizations can secure data without causing delays or operational burdens.

Types of Tokenization: Vault-Based v Vaultless Tokenization

Tokenization methods can be classified into two primary types: **vault-based** (stateful) and **vaultless** (stateless), each with its distinct approach to securing sensitive data.

Vault-Based Tokenization

Vault-based systems store mappings between sensitive data and their corresponding tokens in a centralized database (token vault). While effective in protecting sensitive information, vault-based systems present a few notable challenges.

As the vault expands with new mappings, performance can degrade, particularly in high-transaction environments such as payment gateways or retail systems. The reliance on a centralized database also creates scalability concerns, as the vault must accommodate a growing volume of data. This can lead to synchronization and latency issues, increasing the complexity of system management, especially across distributed environments.

Vaultless (Stateless) Tokenization

In contrast, vaultless systems represent a newer generation of data protection methods. By eliminating the need for a centralized database, they utilize predefined algorithms—such as static table-based or encryption-based methods. This stateless approach effectively addresses the scalability and performance limitations of vault-based systems.

Stateless schemes, like static table-based tokenization and encryption-based tokenization (or format-preserving encryption), address the limitations and complexities of schemes using a token vault.

Static Table-Based Approach

This method employs pre-generated tables filled with random values, which serve as the protection mechanism. These tables, compact enough to be stored in memory, permit deterministic permutation of input data without requiring a central database for mapping. This method effectively resists collisions and preserves the deterministic nature required for certain business applications, ensuring that identical inputs consistently produce the same token.

Encryption-Based Method (Format-Preserving Encryption)

This technique extends traditional encryption algorithms, typically AES, to produce outputs that retain the original input structure. The key used in the encryption and decryption operations acts as the protection secret. While FPE enhances both security and usability, it is not without limitations. Certain FPE algorithms have been declared to be not secure by cryptanalysis as it is theoretically possible to perform successful attacks, making it important to carefully select secure encryption methods for sensitive data protection.

ORIGINAL SENSITIVE DATA	
Name	John Smith
SSN	024-76-8823
Email	Jsmith@corp.com
Pass	N8880925
DOB	10/27/1976

→ tokenization process →

100% TOKENIZED OUTPUT	
Name	Uhea Pmwun
SSN	976-02-1394
Email	Kowipy@glor.com
Pass	N2783904
DOB	05/15/1969

Enhance Your Security

Connect with us to develop
a data-centric strategy that aligns
with your business goals.

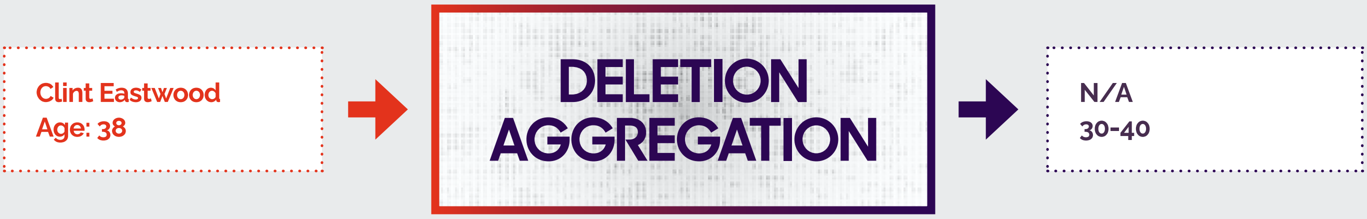
04

KEY CONSIDERATIONS FOR AN ENTERPRISE-WIDE DATA-CENTRIC APPROACH

Anonymization v Pseudonymization: Data De-identification Techniques

Organizations handling personally identifiable information (PII) face increasing pressure to manage privacy while maintaining data utility. Two primary data de-identification techniques—anonymization and pseudonymization—offer unique advantages and challenges depending on the application and regulatory landscape.

Anonymization: Irreversible data protection



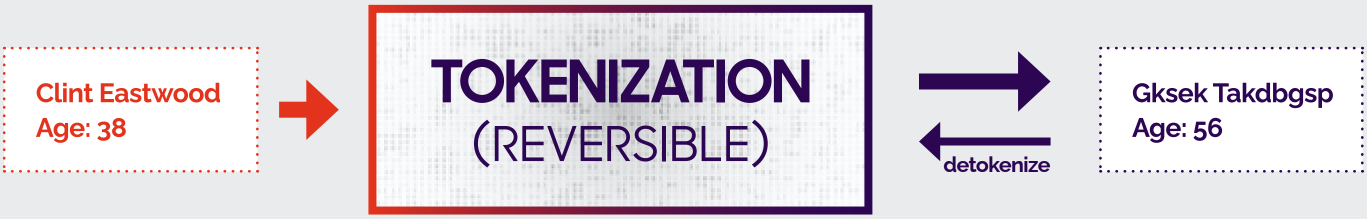
Anonymization using aggregation or deletion

Anonymization transforms data so it cannot be traced back to an individual, often using aggregation or masking identifiers. Once data has undergone this transformation, it no longer qualifies as PII under regulations such as GDPR. This compliance advantage exempts anonymized data from many privacy laws, making it particularly valuable for large-scale data sharing, machine learning, and trend analysis where individual identities are irrelevant.

There are some limitations to this method of de-identification: data that has been anonymized cannot be re-linked to individuals, making it unsuitable for processes that require traceability, like audits or customer support. Achieving effective anonymization is also challenging. Even a residual identifier can expose the data to re-identification attacks, potentially leading to regulatory penalties.

Ensuring anonymity requires rigorous processes to eliminate re-identification risks entirely.

Pseudonymization: Reversible de-identification



Pseudonymization using tokenization

Pseudonymization, such as from tokenization of sensitive information, by contrast, is a reversible method that replaces identifiable information with pseudonyms, retaining the ability to reconnect data to individuals if necessary. This flexibility is beneficial for business operations requiring re-linking, such as fraud detection or customer service.

Although pseudonymized data offers privacy advantages, it remains under regulatory scrutiny due to the risk of re-identification.

For instance, the security of this data depends on protecting "additional information" (e.g., encryption keys) that enable re-linking. Article 4, Section 5 of GDPR underscores that pseudonymization is effective only if these re-linking elements are safeguarded against unauthorized access. Breaches involving pseudonymized data can still lead to privacy violations; however, reporting requirements may be less stringent if the keys remain secure during an incident.

Under GDPR, pseudonymization is considered a strong privacy measure, provided that additional technical and organizational safeguards are in place to prevent unauthorized re-identification.

Emerging Privacy-Enhancing Technologies in Data Security

As privacy regulations tighten, organizations are adopting advanced **Privacy-Enhancing Technologies (PETs)** to strengthen data security. These tools add layers of protection beyond traditional de-identification techniques. Key PETs include Fully Homomorphic Encryption (FHE) and Differential Privacy (DP).

> **Fully Homomorphic Encryption (FHE)**

FHE enables computations on encrypted data without decryption, allowing secure processing without exposing raw data. Unlike standard encryption, FHE preserves computational accuracy, producing the same results as on unencrypted data.

While FHE offers robust security, it comes with high computational costs, especially for large datasets. This limits its practicality for real-time applications, as encrypted data processing significantly impacts performance and scalability.

> **Differential Privacy (DP)**

Differential Privacy protects individual data by adding controlled noise to datasets, preventing specific data points from being identified. It operates on a “privacy budget,” balancing data accuracy with privacy protection, allowing secure analytics without revealing sensitive information.

The challenge with DP lies in maintaining data utility. Excessive noise can obscure insights, while insufficient noise increases privacy risks. Additionally, DP is most effective when applied during data collection; if applied later, raw data may remain exposed.

Secret Isolation: Strengthening Core Data Security Practices

Isolating protection systems and their associated secrets—like cryptographic keys or tokens—is critical to prevent unauthorized access. Whether using tokenization, encryption, or one of the other data protection methods discussed thus far in this paper, strong isolation of secrets enhances overall security.

- **Central Access Model**

  Secrets and protection systems are centrally managed in a secure environment. This model allows for tighter access controls and simplified security protocols, minimizing the attack surface and making it harder for attackers to target sensitive data.

- **Shared Access Model**

  Secrets are distributed across multiple entities needing data access, complicating risk management. Each entity must be protected, increasing the risk, especially when key rotations or revocations occur due to access changes.

- **Risks of Sharing Secrets**

  Traditional encryption methods often rely on shared secrets, and managing these securely across entities is essential. For example, the shared access model requires careful key management to prevent unauthorized access, while tokenization better aligns with central access, reducing exposure. Generally, as the number of entities with access to shared secrets grows, so does the risk of compromise, highlighting the need for robust secret isolation.

- **Protection Secret Rotation**

  Rotating protection secrets regularly is a key security practice that limits exposure if keys are compromised. However, rotation introduces operational challenges, such as across data protected with different keys over time. While encryption often requires frequent key rotation (e.g., in AES-GCM to prevent vulnerabilities), an approach that uses tokenization, for example, can require fewer rotations due to its inherent isolation making it a more resilient option in certain scenarios.

The greater the isolation of a system, the more secure it tends to be. However, extreme isolation can lead to impracticality; a system that is completely secure may become entirely inaccessible. While such a system achieves perfect security, it ultimately undermines its intended purpose by failing to provide usable access.

To illustrate, consider securing access to a building. In a **central access model**, a single guard holds the key (protection secret) and allows authorized individuals access, effectively preventing unauthorized entry and facilitating auditing. In contrast, a **shared access model** distributes keys to all authorized personnel, complicating security and auditing efforts. While the central model is secure, it risks becoming a single point of failure if the guard is unavailable. Introducing a team of trained guards can enhance security without sacrificing accessibility. Similarly, in software systems, deploying a security-hardened central access system in a clustered configuration optimizes security, availability, scalability, and reliability.

Regulatory and Compliance Frameworks

Data-centric security simplifies global compliance by implementing consistent protections across data flows, enabling organizations to streamline processes and quickly adapt to new laws.

Since 2013, over 40 billion data records have been compromised, with fewer than 10% protected by effective measures like encryption or tokenization. Such incidents expose organizations to financial, reputational, and legal risks, underscoring the importance of compliance with global regulatory standards.

Non-compliance can result in severe penalties, as regulators increasingly require proactive measures and timely responses to breaches. For example, GDPR offers reduced reporting obligations if compromised data is protected through pseudonymization or encryption. Many other jurisdictions are adopting similar frameworks, increasing the focus on a data-centric security approach for safeguarding sensitive data.

Liability and obligation in case of a data breach

Regulations like GDPR, PCI DSS, and HIPAA emphasize transparency and accountability during breaches. Penalties are typically imposed not for the breach itself, but for inadequate preparedness. Organizations can avoid or mitigate fines if the breached data was encrypted or tokenized, highlighting the importance of strong data security practices to reduce legal risks.

PCI DSS: Securing payment card data

The Payment Card Industry Data Security Standard (PCI DSS) mandates that organizations protect cardholder data during transactions, ensuring it remains unreadable to unauthorized entities.

A key focus of the standard is minimizing the scope of cardholder data environments, which simplifies compliance efforts and reduces risk. It also enforces strict controls to protect sensitive payment information, particularly during transmission over public networks.

A key aspect of this framework is Requirement 3: Protect Stored Cardholder Data. Under PCI DSS v4.0, organizations must implement granular protection mechanisms to ensure that Primary Account Numbers (PANs) are unreadable wherever stored. Options for achieving this include tokenization, truncation, one-way hashing, or encryption with proper key management.

> The introduction of PCI DSS v4.0, the most significant update since the standard's inception in 2004, reflects the growing complexity of payment systems and the adoption of cloud platforms, emphasizing more advanced protection mechanisms.

Get the Assessment

Access an independent assessment of how comforte Data Protection supports PCI compliance and safeguards your data.

COALFIRE

WHITE PAPER

COMFORTE AG SECURDPS PLATFORM

PCI DSS v4.0 TECHNICAL ASSESSMENT

VIKRAM DHABAL DEB, SENIOR CONSULTANT | CISA, CISSP, QSA, PCI SSCLA

comforte

877-224-8077 | info@coalfire.com | Coalfire.com

GDPR: Privacy and accountability

GDPR emphasizes core principles such as data minimization, purpose limitation, and privacy by design and default. Article 25 requires integrating privacy protections into systems and processes from the outset, rather than adding them as an afterthought.

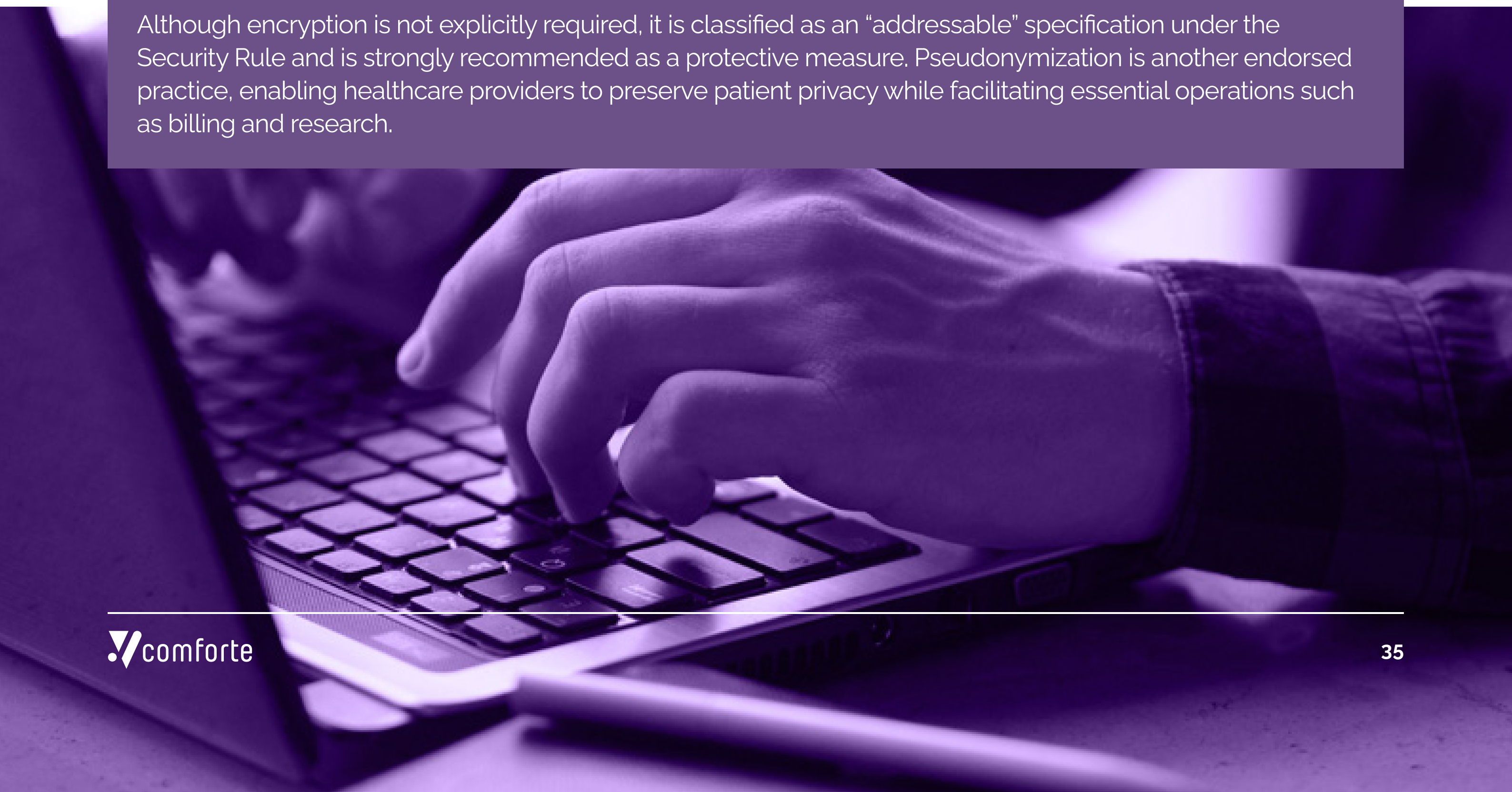
Pseudonymization plays a central role in protecting personal data within this regulatory framework. Defined in Article 4(5) as the substitution of identifying elements with alternatives like unique codes, pseudonymization ensures that data cannot be linked to an individual without additional information. This supplementary data must be stored separately and secured with robust technical and organizational safeguards. While pseudonymized data remains classified as personal data, it reduces exposure risks and strengthens compliance by limiting the potential for unauthorized re-identification.

Article 32 further reinforces the importance of pseudonymization by requiring organizations to implement appropriate technical and organizational measures to secure personal data. Pseudonymization, alongside encryption, is explicitly recognized as a technique for protecting data during processing, storage, and transmission. These measures align with GDPR’s overarching goal of mitigating risks associated with data breaches, unauthorized access, or misuse.

HIPAA: Safeguarding health information

In the U.S., the Health Insurance Portability and Accountability Act (HIPAA) mandates protection for Protected Health Information (PHI), focusing on privacy, security, and breach notification. The HIPAA Security Rule specifically mandates that healthcare organizations secure sensitive patient information both in transit and at rest.

Although encryption is not explicitly required, it is classified as an “addressable” specification under the Security Rule and is strongly recommended as a protective measure. Pseudonymization is another endorsed practice, enabling healthcare providers to preserve patient privacy while facilitating essential operations such as billing and research.



05

THE INDUSTRY SPECIFIC APPLICATIONS AND USE CASES

Protection Method Use Cases

	Format-Preserving Hashing	Data Masking	Classic Encryption	FPE / Tokenization	Synthetic Data	Fully Homomorphic Encryption
End-to-end data lifecycle protection	🟨	🟩	🟩	🟩	🟥	🟥
Integrates without changes to existing systems	🟨	🟨	🟥	🟩	🟩	🟥
Privacy Enabled Analytics	🟥	🟥	🟩	🟨	🟥	🟩
Secure Data Sharing	🟨	🟨	🟥	🟩	🟥	🟨
Secure Cloud Use	🟥	🟥	🟨	🟩	🟥	🟨
Database Joins	🟥	🟥	🟩	🟩	🟥	🟩
Customer Loyalty Programs	🟥	🟥	🟨	🟨	🟥	🟨
Payments Processing	🟥	🟥	🟨	🟩	🟥	🟨

Fully supported, optimal solution.	🟩
Supported with some limitations, moderate fit.	🟨
Not supported, inadequate solution.	🟥

> The control methods that form part of a data-centric security approach have broad applicability across industries. This section examines how different sectors can protect sensitive information for different use cases, ensure regulatory compliance, and unlock the full potential of their data to drive innovation and growth.

Value Creation Through Secure Data Utilization

Security concerns and compliance constraints often limit data use, holding back innovation. Data-centric security removes these barriers by enabling secure collaboration and integration of sensitive information across modern hybrid environments. With data protected throughout its lifecycle, organizations can:

- **Drive digital transformation** by securely accessing data across departments.
- **Improve operational efficiency** by reducing data sharing obstacles.
- **Accelerate revenue growth** through data-driven innovation and improved decision-making.

Data-centric security empowers data scientists to generate insights from datasets without compromising privacy, allows product development teams to create offerings using de-identified customer data, and promotes secure cross-department collaboration with high-quality, protected datasets for strategic initiatives.

Risk Mitigation and Threat Reduction

> **The impact of a breach is greatly reduced. If you lose credit card data, the banks will fine you based on the loss and card replacements value. It could be as much as $50 per card number you lose. If you think about how many cards we deal within a year, it could be devastating. You'll be paying all the banks to replace those cards; it could be very catastrophic in that sense.**
>
> CISO, Leading Retailer, TEI Forrester Report

Despite cybersecurity advances, breaches remain a challenge, with detection averaging over 200 days according to IBM's 2024 Cost of a Data Breach Report. Data-centric security mitigates these risks by limiting breach impact and reducing insider threats:

- **Perimeter breach minimization:** De-identified data reduces exposure even if network defenses fail.
- **Insider threat reduction:** Data obfuscation reduces damage from malicious or accidental leaks.
- **Accidental exposure protection:** Misconfigurations have reduced impact due to built-in protections.

Cloud and Digital Transformation Challenges

Hybrid cloud environments, which involve multi-cloud deployments, complicate data security. Traditional methods, like data-at-rest encryption, often fall short for cloud-native operations. Data-centric security secures sensitive information throughout its lifecycle and aligns with Zero Trust principles, ensuring that data remains protected even in complex, distributed systems.

Zero Trust security in cloud and hybrid environments

Zero Trust, where no internal or external entity is inherently trusted, integrates seamlessly with data-centric security. With granular access controls and protection that travels with the data, organizations reduce unauthorized access risks in cloud and hybrid environments, ensuring continuous verification and least-privilege access.

Secure Access Service Edge (SASE) architecture and data protection

SASE architecture plays a pivotal role in modern cloud environments by integrating identity-driven and cloud-based security systems. This integrated approach enables secure, reliable access to applications and data, regardless of the user's location. As organizations increasingly adopt cloud environments, Security Service Edge (SSE), a critical subset of SASE, becomes central to ensuring that sensitive data is protected throughout its lifecycle.

SSE integrates security solutions such as encryption, tokenization, and data masking to protect data from unauthorized access and cyber threats. These methods align with data-centric protection principles, ensuring sensitive information remains secure even during processing or transit. By unifying multiple security functions into a centralized framework, SSE enables consistent policy enforcement across diverse environments, strengthening an organization's overall security posture and its ability to address evolving threats

Banking Security and Payments

In banking, safeguarding sensitive customer data is critical for maintaining trust and ensuring compliance. As digital transformation accelerates and cloud adoption rises, new security challenges emerge—particularly regarding cross-border data flows and compliance with regulations such as PSD2 in the EU, which governs open banking and secure data sharing via APIs. Tokenization plays a vital role in securing payment data, which is a prime target for cybercriminals, improving transaction monitoring, fraud detection, chargebacks, and regulatory reporting.

With the shift to cloud-based payment systems and AI-powered fraud detection, protecting data before transmission to the cloud helps mitigate vulnerabilities and manage emerging risks.

Ensuring Data Protection in Analytics Platforms

The rise of cloud-native architectures and Big Data analytics presents new challenges in securing vast amounts of structured and unstructured data. Traditional encryption methods can impact performance, leading some organizations to hesitate in securing sensitive data.

Data-centric security offers a solution enabling the use of tokenized data for secure processing and analysis. For example, a retailer can anonymize PII in transaction data, allowing analysts to explore trends without exposing customer identities. Tokenized data can be stored in cloud data warehouses like Snowflake, where queries operate on tokens, with detokenization occurring only when necessary.

Data Sharing and Data Sovereignty

Sharing sensitive information with external partners or across departments introduces significant security risks, particularly in cross-border scenarios.

Organizations may find compliance with data localization laws would mean substantial investments in local data centres and infrastructure, which can be financially prohibitive.

Anonymization and pseudonymization techniques to protect sensitive information enable businesses to share only the necessary data with partners, ensuring privacy during collaborations. Tokenization effectively safeguards data during cross-border transfers, allowing sensitive information to remain within its country of origin while tokenized data is shared internationally.

Data-centric security addresses key data-sharing and sovereignty challenges by:

01 **Ensuring Compliance with Country-Specific Data Protection Laws**

Protect sensitive data while maintaining operational flexibility.

02 **Enabling Global Data Processing with Local Control**

Transfer data securely across jurisdictions without compromising control over sensitive information.

03 **Utilizing Tokenization for Data Localization**

Allow organizations to store and process tokens in foreign environments, while the original data stays within the required jurisdiction.

06

HOW TO IMPLEMENT DATA-CENTRIC SECURITY

5 Steps to Implement Data-Centric Security

Implementing a data-centric security strategy requires aligning security controls with business objectives, regulatory needs, and existing IT infrastructure. This strategy must address not only the technical architecture but also operational processes and governance practices that ensure data security across the organization.

01 Locate Sensitive Data

Identify and map sensitive data, tracking its movement and uncovering unauthorized sharing or shadow IT. Automated discovery tools streamline this process.

02 Minimize Data and Reduce Scope

Reducing the amount of sensitive data processed is key. The rationale is simple: less data means a smaller attack surface. Assess which data is essential and eliminate redundant or non-essential information.

03 Conduct Data Protection Risk and Impact Assessments

Regular risk assessments help understand evolving threats and preemptively address security gaps, particularly during changes like mergers or system upgrades.

04 Define Policies and Apply Protection Methods

Establish clear security policies based on data classification. Use role-based access controls to restrict data access and implement centralized access control for real-time enforcement and visibility into usage.

05 Maintain an Audit Trail

Detailed logging of data access is vital for compliance and security. A centralized access model ensures all interactions with sensitive data are tracked for detecting threats and supporting investigations.

Remember the Human Element

Even with advanced technology, the human factor remains a crucial element in data security. Regular training on data protection practices reduces the risk of insider threats, both accidental and malicious. Security awareness programs ensure employees understand their role in protecting sensitive data and align with organizational policies. By fostering a security-conscious culture, businesses empower their workforce to act as the first line of defense against emerging threats.

Data-Centric Security Solutions Quick Checklist

✓ **Centralized Management Platform**
A solution that offers a single interface for managing security across all environments, integrating with major platforms for consistent policies.

✓ **Automated Data Discovery**
Tools for discovering, classifying, and cataloging sensitive data for compliance with regulations like GDPR, PCI DSS, and HIPAA.

✓ **Diverse Protection Methods**
The ability to securely apply encryption, tokenization, or format-preserving encryption (FPE) throughout the data lifecycle.

✓ **Policy Management**
Facilitate the creation and maintenance of adaptable security policies to ensure consistent data protection.

✓ **Robust Access Controls**
Role-based access that enforces least privilege and separation of duties.

✓ **Audit and SIEM Integration**
Built-in auditing capabilities that integrate with existing security information and event management (SIEM) frameworks.

✓ **Scalability**
The ability to scale without performance degradation to handle growing workloads.

✓ **High Availability**
Built-in fault tolerance and automatic failover to maintain security during system failures.

✓ **Flexibility**
Support for various deployment models, including multicloud and hybrid, with minimal disruption.

✓ **Seamless Legacy Integration**
Integration with existing infrastructure—both on-premises and cloud—without significant re-engineering.

✓ **Trusted Vendor Support**
Choose a vendor with a strong track record and battle-tested technologies that demonstrate effectiveness in real-world production environments. Look for a partner who provides expert guidance and dedicated support throughout the deployment process.

> Implementing effective data-centric security requires more than awareness—it demands action.

Turning Insights into Action

Organizations must evaluate available protection techniques, assess their suitability for specific environments, and understand their broader implications. While this document provides foundational insights, consulting industry experts will offer valuable perspectives. Engage vendor solutions, conduct security audits, and prepare for future threats to strengthen your data protection strategies.

Unlock the Full Potential of Your Data
Empowering Growth, Mitigating Risk, Ensuring Compliance

Comforte Data Protection

Data-centric security offers game-changing strengths for securing sensitive information. While the choice of data protection method depends on specific use cases, system requirements, and the data involved, understanding the strengths and limitations of each method allows organizations to develop effective data protection strategies that meet security needs without compromising operational efficiency.

Overall, organizations should integrate data-centric controls into their core operations, viewing it not just as a compliance requirement but as a core element of their digital strategy. By doing so, they can create a resilient data ecosystem that supports decision-making, innovation, and customer value while enhancing their overall security posture.

At comforte AG, we understand that data is your organization's most valuable asset. With over 25 years of expertise in securing mission-critical systems, we are a trusted partner in data-centric security, helping businesses navigate complex security challenges while driving growth and ensuring compliance.

Our comprehensive approach to data protection offers a range of advanced solutions tailored to meet the diverse needs of modern enterprises. From encryption and tokenization to format-preserving hashing and data masking, we deliver robust methods to secure sensitive data—whether at rest, in transit, or in use.

Got 20 minutes?

Schedule a personalized session with our data security experts to explore how comforte Data Protection can drive enterprise-wide value through data-centric security.

comforte

Head office address
comforte AG
Abraham-Lincoln-Str. 22
65189 Wiesbaden
Germany

Visit our website: comforte.com
Phone: +49 611 93199 00
Sales@comforte.com

Disclaimer

The information in this paper is for educational and informational purposes only, focusing on data protection and cybersecurity strategies for organizations. Cybersecurity is an evolving field, and the strategies outlined may not be suitable for every organization or situation. Readers should conduct their due diligence and consult qualified professionals before implementing any recommendations or solutions presented here.

Comforte AG disclaims any liability for losses, damages, or other liabilities arising from the use of this paper or the application of its content.