Couchbase

START A REVOLUTION

Best Practices for Data Protection and Security in the Couchbase Data Platform

Best Practices for Data Protection and Security in the Couchbase Data Platform

Introduction

The emergence of European Union's General Data Protection Regulation (GDPR) which comes into effect in May 2018 highlights the regulatory role in ensuring a strong and up-to-date framework for the management and processing of personal data. A number of the principles in GDPR can be found in other regulations such as the Health Insurance Portability and Accountability Act (HIPAA) and the Payment Card Industry Data Security Standard (PCI DSS), which ensure data is processed and secured appropriately.

Most sensitive data covered by different regulations, be it personal, financial, or otherwise, will invariably reside in a database. This document presents the Couchbase Data Platform's best practices for regulation and security. The core focus is on the role of the database as a central store of data within an organization for personally identifiable information (PII) and the tools and processes that can be applied to the data and database to protect and secure it.

While this paper focuses on the Couchbase Data Platform, regulation and security best practices for an organization must be holistic beyond just the database and include:

1. Assessments of people and organizations who process or access data and the policies around the data itself.

2. All components of the security landscape including the physical facilities, hardware, networks, operating systems, and applications.

Security, privacy, and personal information matter

Data security threats are continuously rising, with the attack frequency, severity, and sophistication **increasing every year**. In order to secure data against these increasing threats, organizations must make security a top priority.

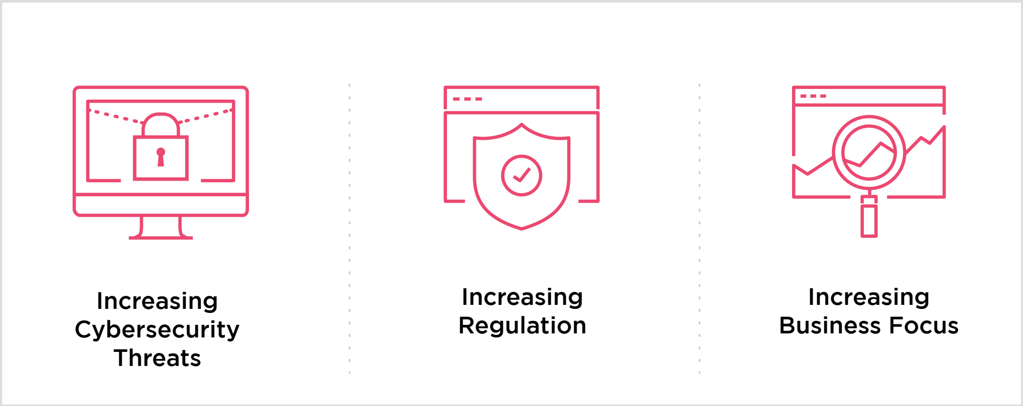
> **“Cybersecurity exercises are going to be absolutely expected of all companies by regulators.”**
>
> – Michael Vatis,
> founding Director
> of the FBI’s National
> Infrastructure
> Protection Center

Additionally, there is a trend of increasingly strict privacy and protection regulation for citizens around the globe. For example, GDPR creates a consistent baseline for EU citizens’ personal data online, including introduction of some significant new rights for individuals and corresponding obligations for global organizations. GDPR is just one example of regulations currently in effect or being developed around the world. Not meeting these regulations for an organization’s particular geography, industry, or data types can mean significant fines and penalties.

The cyberthreat and regulatory response are two strong arguments for driving your security program and protection of PII.

Increasingly, however, management of PII and its supporting security programs are also driven by commercial imperatives.

Management of an individual’s data is part of the wider customer experience and interactions. As more activity transfers into digital markets where customer data is central to interactions and transactions, **businesses that do not manage their customer data well and build trust will be at a competitive disadvantage**. Customers will shift to services that are trusted and offer better security and privacy capabilities. At the extreme, individuals will use their legal rights to bar or remove their data from organizations that do not. Regulators can also prohibit an organization from accessing or processing personal data. In any market, especially digital, that company would, in effect, be out of business.



Increasing Cybersecurity Threats

Increasing Regulation

Increasing Business Focus

Businesses spend inordinate amounts of time trying to optimize customer experience. PII is the customer’s actual data footprint and should be treated as an extension of them. **As businesses treat their customers as golden, they should treat their data footprint the same.**

This responsibility extends throughout the whole organization. Security and compliance teams must be aware of what data is being captured and stored in order to advise how it must be secured, R&D and operations teams must understand the **benefit of securing PII** in order to do so proactively and successfully.

See Appendix B for more about personally identifiable information (PII).

Privacy by Design

Privacy by Design is an approach that promotes **privacy and data protection compliance from the start**. Privacy considerations have often, at best, been bolted on as an afterthought and at the worst, been ignored altogether. As more regulation has been put into effect, Privacy by Design ensures adequate privacy and security. Businesses who adhere to this process reduce the risk of cybersecurity breaches. And, as regulators increasingly look for strong upfront development processes supporting privacy, businesses that secure data can be trusted and will be in a stronger competitive position in the digital marketplace.

80% of consumers globally say trust is a key driver of brand loyalty; 45% of consumers globally switched providers in the last year because they lost trust in a company.

– A New Slice of PII With a Side of Digital Trust, Accenture 2018

Beyond the best practice recommendations is this document, assessing the value of a Privacy by Design program may be an additional consideration.

See Appendix C for more about Privacy by Design.

Securing PII in the Couchbase Data Platform

The Couchbase Data Platform provides built-in security across the entire platform and is designed to integrate with any enterprise environment. While there are many ways to meet a particular security requirement, implementing the strictest of security controls to all data across all parts of an organization may not be practical.

The Couchbase Data Platform is capable of ensuring a high level of security **without significant operational or compliance overhead**. As some of the processes described in this paper do have an impact on performance and operations (i.e., encryption and auditing require extra CPU cycles), it is advisable to make use of the Couchbase Professional Services team to determine the best configuration and architecture. In many cases, Couchbase performs benchmarks in order to advise organizations based upon their specific workloads and requirements. As these benchmarks show, the Couchbase Data Platform **maintains performance with easy operational processes**.

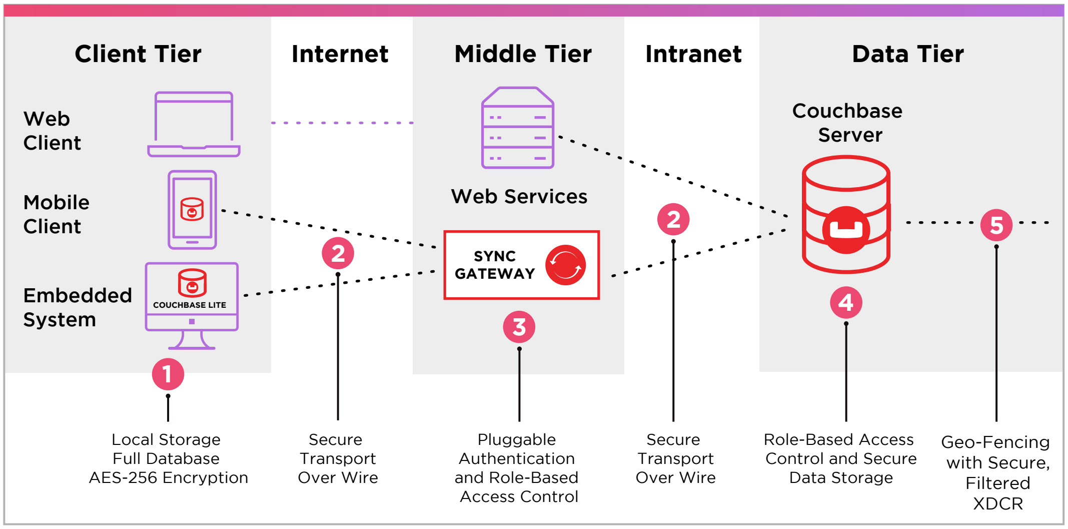
"There's no silver bullet solution with cyber security, a layered defense is the only viable defense."

– James Scott, Senior Fellow, Institute for Critical Infrastructure Technology

The main focus of this paper is on live production environments storing sensitive data, including PII. Organizations should think carefully about securing this sensitive data **in all environments, from development to test to support to storage scenarios**. Simply minimizing the existence of sensitive data outside of production is the cleanest approach to ensure it cannot be exposed or misused.

Security from core to edge

Most discussions about data and database security are limited to what can be done within the bounds of an organization's firewall/infrastructure. Anything outside of that is often left up to developers, or even the end user. However, with built-in synchronization of data from the "core" (datacenter or cloud) to the "edge" (mobile device or embedded system), the Couchbase Data Platform makes it possible to provide end-to-end security controls.


Client Tier
Internet
Middle Tier
Intranet
Data Tier
Web Client
Mobile Client
Embedded System
COUCHBASE LITE
Web Services
SYNC GATEWAY
Couchbase Server
1
2
3
2
4
5
Local Storage Full Database AES-256 Encryption
Secure Transport Over Wire
Pluggable Authentication and Role-Based Access Control
Secure Transport Over Wire
Role-Based Access Control and Secure Data Storage
Geo-Fencing with Secure, Filtered XDCR

Full-Stack Security Controls for Enterprise Security Compliance

In the above diagram, you can see the three products that make up the Couchbase Data Platform:

- **The core:** Couchbase Server is a clustered, distributed database designed to run within a datacenter or cloud LAN environment. It may also replicate to one or more Couchbase Server clusters, likely in other datacenters or clouds.
- **The edge:** Couchbase Lite is an embedded database for managing data on a mobile device or embedded system.
- **In the middle** is Couchbase Sync Gateway, which manages the connectivity and data routing between one or more Couchbase Server clusters and one or more instances of Couchbase Lite. Sync Gateway will typically reside in a DMZ or internet-facing zone of an organization's network.

While the requirement to secure data remains unchanged, those requirements and features associated with them will differ at each layer. For example, authentication and role-based access must be granular down to the individual user at the edge, while it can be coarser grained, application-level access at the core. While the edge is inherently untrustworthy from an access perspective, firewalls and business processes can be used to control access at the core. The Couchbase Data Platform is designed with all of this in mind.

More than 65% of organizations allow unrestricted, unmonitored, and shared use of privileged accounts, which severely limits auditability and personal accountability.

– Best Practices for Privileged Access Management, Gartner 2017

Throughout this document, references to the Couchbase "Data Platform" imply that a feature or discussion applies at all levels, whereas Couchbase "Server," "Sync Gateway," or "Lite" will be used to reference capabilities or requirements of those products specifically.

Regardless of where data is captured, transferred, or stored, there are four fundamental security concerns:

1. Access control
2. Encryption/Masking/Redaction
3. Discovery/Sovereignty/Retention
4. Auditing/Reporting

See Appendix A for a Couchbase security checklist.

Access control

From both a regulatory and business standpoint, **access control is the first core building block for securing data**. At its simplest, access control is about identifying and verifying that a user is who they say they are (*authentication*), and then what they are allowed to see or do (*authorization*). Whether a user is an individual or an application makes no difference.

Security best practices lay out two concepts: separation of duties and least privileged access. From a practical perspective, together these call for separate credentials for every user and that those users are only allowed to perform the minimal level of activity required.

At the "core," users are either database administrators, individual developers, or applications/services. At this level, Couchbase Server provides coarse-grained access control that integrates with an enterprise's internal security controls. These users typically number from one to the low hundreds and need access to all or large portions of the dataset.

At the “edge,” a user directly accesses an application. These users can number into the millions, many times sitting outside of a trusted network and granted access to their own individual slice of a larger dataset. To meet these requirements, Couchbase Sync Gateway implements fine-grained access control that scales to millions of users, authenticates based on both internal and external authentication mechanisms, and handles the dynamic linkage between a user and what individual pieces of data they can read and/or write. Couchbase Sync Gateway also has a built-in user for administration.

An embedded system or a device is typically accessed by only one user at a time. Therefore, Couchbase Lite is a single-user database. Any credentials or access to data are verified by Couchbase Sync Gateway.

Authentication

The first step of access control is to determine *who* is trying to access the data – **users must be clearly and strongly authenticated**. While this will depend on an organization’s own standards and capabilities, certificate-based authentication is currently the strongest form that Couchbase provides.

Couchbase Server supports various password, certificate, and third-party based authentication models to fit the environment:

- Password-based: The Couchbase Data Platform supports built-in, password-based authentication for both administrators and applications. For additional security, password strength policies should be set which are operationalized and enforced (complexity of the password, lifecycle, updating of the password, etc.). The transmission of credentials for both administrators and application users can be encrypted with Transport Level Security (TLS) and/or hashed.
- Certificate-based: Couchbase also supports the use of X.509 certificates to authenticate users. X.509 certificates provide an additional layer of security where the certificate authority (CA) validates identities and issues certificates. The Couchbase Data Platform supports both self-signed as well as CA-signed certificates across all TLS-enabled services. While there is a bit more operational overhead in the setup of this method, it provides much stronger security as well as management capabilities at scale.
- Third-party/external authentication:
    - LDAP/AD: Login and connection attempts to Couchbase can be directed to authenticate against an LDAP or Active Directory layer.
    - PAM: Pluggable Authentication Modules (PAM) provide an authentication framework that allows multiple, low-level authentication schemes to be used by a single API. The Couchbase Data Platform primarily supports local Linux users but also leverages PAM for third-party tools such as Callsign or Kerberos.

Couchbase Sync Gateway also supports password-based, built-in provider, and pluggable authentication methods:

- Password-based: Usernames and passwords can be defined within Couchbase Sync Gateway and used to authenticate users and/or devices as they connect.
- Built-in providers: Couchbase Sync Gateway has built-in support for Facebook, Google+, and OpenID Connect authentication providers.
- Custom authentication: A pluggable authentication service allows any external application to handle authentication on behalf of Couchbase Sync Gateway.

As it is a single-user database embedded into and accessed exclusively by the application, Couchbase Lite requires no authentication.

Authorization

Once a user has been authenticated, *authorization* determines what that user is allowed to do. The Couchbase Data Platform employs **Role-Based Access Control**: users are mapped to roles which determine the actions they are authorized to perform.

At the "core," Couchbase Server separates its roles between administrator/operations and application/data access. Each user may have zero or more roles which can be as broad or as restrictive as required, from a Full Administrator having access to all administrative functions and data, a Bucket Admin only having permission to control the settings of a particular dataset, to read-only, query-only, search-only, etc. To protect against escalation of privileges, administrators are limited to modifying permissions levels *below themselves* if at all.

At the "edge," there is often a requirement for more dynamic, programmatic assignment of roles down to the individual document or field level. Couchbase Sync Gateway allows for both static and dynamic role assignment for individual users. Read access can be controlled down to an individual document, while write access can be controlled down to one or more fields within a document. This goes hand-in-hand with the routing of data to individual devices that Couchbase Sync Gateway manages: documents that can be read by an individual user are synchronized to their device and changes to that data are either accepted or rejected when synchronized back.

As an embedded, single-user database, Couchbase Lite has no need for roles or authorization.

Storing credentials outside of the Couchbase Data Platform

Later in this document we will discuss the secure transmission of credentials (username/password/certificates) **into** the Couchbase Data Platform as well as how to secure them **within** the platform. However, it is also important to be aware of the security of these credentials **outside** of the Couchbase Data Platform.

Across all of the Couchbase Server SDKs, there is support for securely transmitting usernames and passwords as well as X.509 certificates. In some cases this is done via environment variables, and in other cases this is done by native keystores such as the JVM keystore for Java.

At the edge, it is a best practice not to embed any sensitive usernames or passwords within a mobile application. Rather, secure on-device keystores should be leveraged where available as well as server-side authentication/authorization (supported through Couchbase Sync Gateway). Couchbase Lite supports securing credentials in a local database using AES 256-bit encryption and developers need to manage the encryption keys accordingly.

Securing access to the systems

This document is not intended to be a comprehensive guide to facilities, network, or system-level security, but it is important to take these topics into consideration as well. These may be offloaded by your infrastructure provider but it is everyone's responsibility to be aware of the entire security landscape. These can include:

- Physical – Buildings, datacenters, cages, servers
- Network – Firewalls, iptables, WAN encryption
- Operating system – User management, security patches and updates
- Application – Credentials
- Key management – Rotation, revocation, remediation

PII access control cheat sheet

1. **Security and regulatory processes for PII and other sensitive data is a significant strategic decision and teams should get the right level of buy-in and support, including:**
    a. Senior sponsorship and support for privacy good practices
    b. Security is a two-way street:
        i. Security and compliance teams need support to understand the data landscape of their customer.
        ii. Development, operations, and architecture teams must understand the need for security of that data and why PII is so significant to an organization.

2. Clear and enforced doctrines of **"separation of duties"** and **"least privilege"**: no entity should be allowed to access PII without a clear business need, and nothing is granted by default. For example, top-level administrator access should be scoped and narrowed for what is necessary for their role (e.g., setting up other roles, configuration of the product, etc.), and every task within the application performed with lower-level roles.

3. Applications should be treated like any other user with regards to PII: provided access to only the **dataset and tasks needed for the running of the application**.

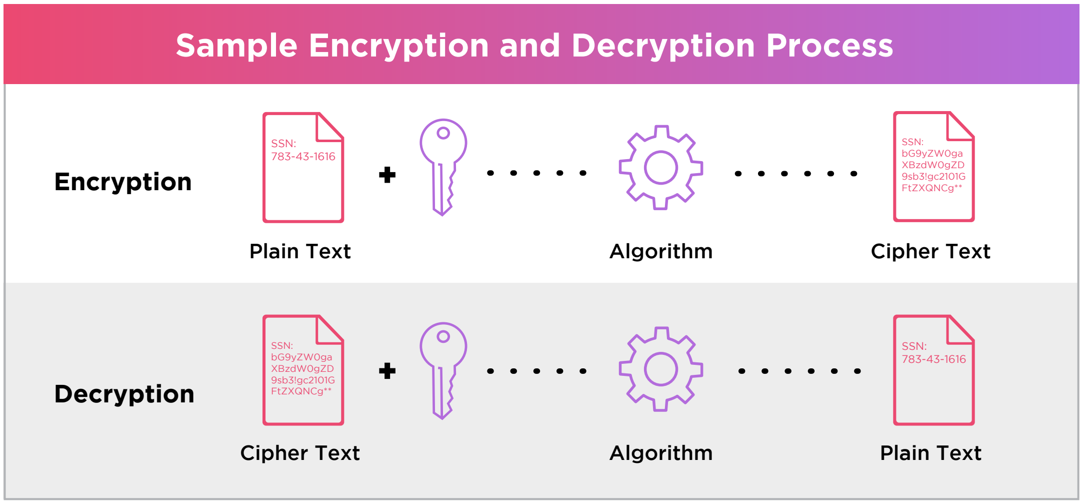
4. Organizations need to enforce such processes as every entity accessing the database is **identifiable with unique credentials, and that access is based on strong authentication**.

5. Access to facilities, networks, and systems must be secured.

6. The access controls policy and environment **should be regularly reassessed**.

Encryption

Encryption is another workhorse of privacy regulation. The core objective of encryption is to ensure that sensitive data is not accessible in the event of unauthorized access (a breach).



Encryption Process

Unencrypted information, often referred to as *plaintext*, is encrypted using an encryption algorithm and an encryption key. This process generates *ciphertext* that can only be viewed in its original form if decrypted with the correct key.

Information transmitted over a network is vulnerable to eavesdropping by unauthorized parties. Information stored on disk is vulnerable to compromises at the physical or operating system (OS) layers. Either could lead to the exposure of sensitive user data and/or the capture of credentials that could be used for wider, unauthorized access.

In the context of a database, encryption is used to protect the user/application data as well as credentials and other metadata used to connect and gain authorized access. The Couchbase Data Platform and its partner network support the strong encryption of this data and metadata both in transit over the network and/or while stored on disk.

Encryption on the wire

Traditional database systems are deployed and accessed from within a corporate firewall. This level of protection against outside attackers used to be enough. In the modern world, however, databases are expected to replicate and synchronize data over the public internet. The benefit of doing so requires an added layer of protection, which the Couchbase Data Platform addresses.

In the core, Couchbase Server uses TLS to encrypt data and credentials passed between clients and servers, and between clusters (typically across datacenters). Currently, Couchbase Server does not support native encryption of data between the nodes of an individual cluster. A Couchbase Server cluster is confined to a single LAN environment and therefore all nodes are secure by a single firewall domain, greatly limiting the scope of vulnerability. If required, the intra-cluster traffic can be secured further with iptables or ipsec configurations and a future release will provide native encryption to make this easier to manage. Keep in mind that an individual Couchbase Server cluster is confined to a LAN environment. Cross datacenter replication (XDCR) *does* provide native encryption and is suitable for transmitting data between clusters over the public internet.

Couchbase Server automatically generates a self-signed certificate which is propagated throughout all the nodes of a cluster. In many cases, a self-signed certificate does not conform to enterprise security standards so Couchbase Server also supports CA-signed (X.509) certificates across all TLS-enabled interfaces.

Couchbase Sync Gateway and Couchbase Lite are intended to face the public internet for all data transfer and support HTTPS through X.509 certificates provided by the administrator.

Encryption at rest

Encryption at rest refers to the encryption of data residing on physical media. This level of encryption is designed to protect against unauthorized access to the database files either from within the operating system or to the physical disks themselves.

Couchbase Server relies upon and supports on-disk encryption solutions provided by third-party software vendors which deny data access to anyone who does not possess an appropriate encryption key or is otherwise noncompliant with the configured security policy. Couchbase partners with Vormetric, Gemalto, and Protegrity, and also supports LUKS, Windows file-system encryption, and Amazon’s encrypted EBS. A native on-disk encryption capability is currently being evaluated for development.

Couchbase Sync Gateway does not store any data and therefore does not need to encrypt it at rest.

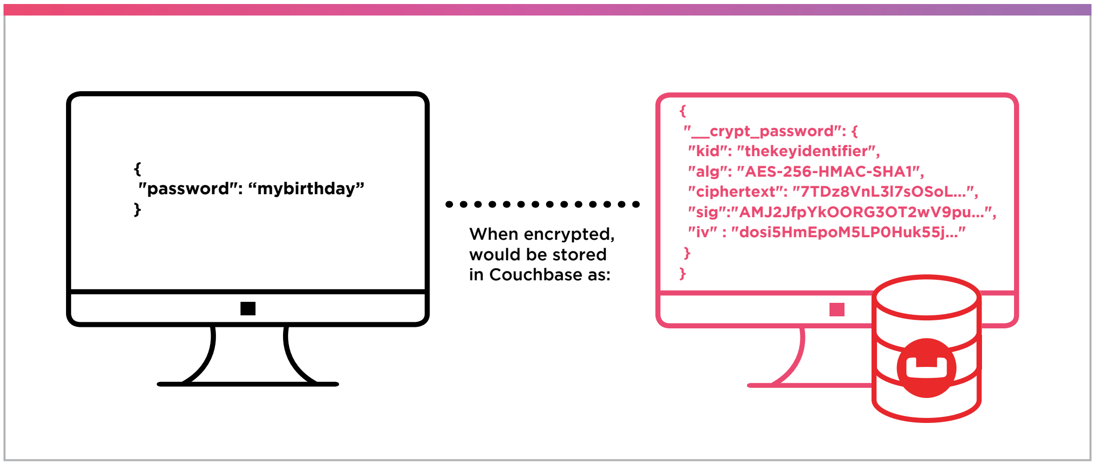
Couchbase Lite supports securing data “at rest” in a local database using AES 256-bit encryption. The encryption key is applied when the embedded database is created and the same key is needed to access the embedded database.

Data-level encryption

While encryption of data in transit and at rest provides protection against unauthorized **external** access, the data is still accessible by any authorized user and the database software itself.

Data-level encryption provides an extra layer of protection by encrypting user data within the database itself. Not only is it encrypted over the network and on disk, but requires a separate key from the application to decrypt.

The Couchbase Data Platform supports native data-level encryption. This is also available through some of the third-party technologies that offer encryption of data at rest such as Vormetric, Gemalto, and Protegrity. Both options support encrypting an entire document or specific fields within (such as credit card or Social Security numbers). Data encrypted in this manner is stored as an opaque "blob" within Couchbase which limits the indexing and querying capabilities.



```
{
 "password": "mybirthday"
}
```

```
{
 "__crypt_password": {
  "kid": "thekeyidentifier",
  "alg": "AES-256-HMAC-SHA1",
  "ciphertext": "7TDz8VnL3l7sOSoL...",
  "sig":"AMJ2JfpYkOORG3OT2wV9pu...",
  "iv" : "dosi5HmEpoM5LP0Huk55j..."
 }
}
```

Couchbase Server Secret-Management

To support its normal operation, Couchbase Server must store certain usernames, passwords, certificates, and internal tokens on disk. To protect this information against OS and physical breaches, Couchbase Server supports the encryption of these credentials on disk with an AES 256-bit algorithm in GCM mode and is protected by a master password that can be rotated as needed.

Key management

Managing encryption keys is a critical part of the security framework. Typically, the key management process for the Couchbase Data Platform would fit with the wider key management processes of the organization including rotation, revocation, and remediation if necessary.

Data masking and redaction

When it is necessary to access or process sensitive data in an unencrypted form, data masking and redaction are common approaches. Data masking is an umbrella term for approaches like pseudonymization and anonymization that protect confidential information by decoupling it from an individual's identity. **Pseudonymization** refers to the substitution of PII or other sensitive data with tokenized values so that linkage to an identity is not possible without additional information and security controls. **Anonymization** is the complete obfuscation of an individual's PII with no link back to the original. **Redaction** is simply the removal of sensitive data while allowing the non-sensitive data to be accessed. These practices eliminate the ability to identify an individual based upon their characteristics while still allowing the resulting/remaining data to provide benefit. For example, production support teams may need access to data about an individual but not their credit card or Social Security numbers. There may also be development or test processes where datasets need to closely mirror that of production while not exposing the actual sensitive data or PII.

An often overlooked, yet important, aspect of data masking is the discovery of PII and other sensitive information in a dataset. The Couchbase Data Platform supports a variety of flexible search, query, and index functions to identify this sensitive data for redaction or masking.

Depending on the individual requirements, there are a few approaches to masking PII or sensitive data within the Couchbase Data Platform:

- Couchbase Server's MapReduce views and the Eventing service both allow for the programmatic creation of "materialized views." These representations of the main dataset are augmented by either redacting sensitive values and/or replacing them with tokens/random values. Separate security controls can be applied to the original dataset, the creation of these views, and access to them.
- Couchbase's query language, N1QL, has deep support for JOIN semantics. This allows sensitive data to be separated from non-sensitive data within the database and combined only when authorized.
- When querying data through N1QL, the result set can be manipulated to redact or mask specific values, enforced through an organization's coding practices and APIs.
- At the edge, Couchbase Sync Gateway allows for fine-grained control over which data is synchronized to which devices.
- Finally, Couchbase works with best-of-breed vendors such as Gemalto and Vormetric for pseudonymization and anonymization at the application level.

Log redaction

The Couchbase Data Platform produces a rich set of logs for usage tracking and troubleshooting. While actual user data is never intentionally written to logs, a variety of other data can be deemed potentially sensitive such as document keys, usernames, index definitions, query strings, etc. These are automatically tagged in the logs and can be redacted upon collection.

PII encryption and masking best practices cheat sheet

1. **Only store sensitive data that you need for your business.** The concept of both data minimization and legitimacy of processing is increasingly highlighted in regulation.
2. **Use widely accepted algorithms and widely accepted implementations** (e.g., GCM, CCM). For personal information, aim to use an implementation that is FIPS 140-2 certified, including in an Authenticated Encryption mode.
3. **Encryption keys need to be separately stored and subject to strong protection including:** a specific key lifecycle (creation, rotation, etc.), physical and logical separation of the keys from the encrypted data, and key generation lifecycle process. Role-Based Access Controls also apply to key management including separation of duties and dual control for critical key management tasks.
4. **All connections to the database should be encrypted** as well as internal connections across the "database instances."
5. **Data at rest must be encrypted** to mitigate threats targeting the operating system or physical environments in addition to the database itself.
6. Where PII needs to be legitimately processed, use pseudonymization and other masking techniques to **maintain individual privacy and reduce the risk of breach**.
7. The Couchbase Data Platform supports search and management of PII (e.g., delete). Ensure there are clearly agreed-upon processes for:
   a. **Searching and detecting the data/PII**
   b. **Accessing, transferring, and deleting data/PII**
8. **Regulation links**
   a. **These best practices need to be cognizant of the actual regulation(s) being applied.** For example, whereas GDPR takes a principled position that data should be encrypted, it does not specifically indicate the standards or technologies required whereas U.S. federal regulations have specific encryption standards and levels that must be met.
   b. Financial services make regular use of pseudonymization. PCI DSS requires the **masking of credit card information in various scenarios**.

Data discovery, sovereignty, and retention

Determining what data is retained, where, and for how long is becoming one of the most critical functions from a regulatory perspective. Regulations such as GDPR are increasing the rights of individuals to be made aware of and optionally remove their data held by an organization. Additionally, these regulations provide for the control of the **physical storage location** of data. In many cases, security controls must be tightened, or conversely can be relaxed, if sensitive data is retained for longer or shorter periods of time.

The Couchbase Data Platform supports the discovery of what data is stored within the platform, controlling its geographic storage location, and deleting that data according to various policies.

> **"... we base our online user experience around what consumers want. We shape our products and services around what consumers want. We need to shape our data protection approach around what consumers expect."**
>
> – Elizabeth Denham,
> UK Information
> Commissioner, 2017

Discovery

- The Couchbase Data Platform provides rich search, query, analytics, batch, and streaming capabilities to discover sensitive and PII data either on demand or through regular processing.
- There are a variety of important aspects to the discovery of data: timeliness when data changes, performance impact on the application, and operational overhead. The Couchbase Data Platform is uniquely suited to address all of these aspects due to its architecture and consolidation of access patterns.
- Partnerships and integrations with technologies like Informatica, Pentaho, Kafka, Spark, Storm, Hadoop, and more further support this requirement.

Sovereignty

- The cross datacenter replication (XDCR) feature within Couchbase Server provides a filtering mechanism to control which documents are replicated from source to destination clusters.
- At the edge, Couchbase Sync Gateway implements granular security policies down to an individual user and role level, making it easy to control data replication between Couchbase Server and a large number of Couchbase Lite instances.

Retention

- Using the same tools as for discovery, the Couchbase Data Platform allows for the periodic or batch deletion of data. The Couchbase Data Platform is capable of handling the deletion of large volumes of data without a performance impact or operational overhead.
- The Couchbase Data Platform supports time to live (TTL) applied either on individual documents or across an entire bucket. Once the TTL has passed, that data will no longer be accessible and is eventually purged from the system.
- At the edge, Couchbase Sync Gateway will automatically manage the deletion of data on devices and embedded systems when it is deleted from the core. In addition, Couchbase Lite supports deletion of data locally.

Auditing and reporting

To complete any security framework, a comprehensive auditing and reporting capability is necessary. There is no prescriptive set of events to audit as it depends on the organization, service, and dataset in operation, and the breadth of PII. The Open Web Application Security Project suggests certain areas that **should and should not** be logged for auditing purposes. Regulatory requirements will generally set some audit requirements for data and PII:

- All activities that impact PII are formally tracked.
- There is integrity of data and audit events, and logs cannot be overwritten or tampered with in any way.
- There is some notification to impacted users and the regulatory authorities if there is a breach.

"… there are only two different types of companies in the world: those that have been breached and know it and those that have been breached and don't know it."

– Ted Schlein,
Kleiner Perkins
Caufield & Byers

The Couchbase Data Platform provides comprehensive auditing facilities. The audit records created by Couchbase capture information on who has performed what action, when, and whether successful or not:

1. **Who:** The administrator, user, or application performing an action.
2. **What:** The action performed. Couchbase Server supports auditing of administrative and applicative actions including (but not limited to) connection/login attempts, cluster topology changes, dataset creation/deletion, N1QL queries, individual document reads and writes, etc. Couchbase Sync Gateway maintains an access log of connection and synchronization attempts. As mentioned earlier, Couchbase Lite is an embedded single-user database and accessed exclusively by the application so there is no need for auditing at that layer.
3. **When:** The time stamp that corresponds to each recorded action.
4. **How:** The success or failure of the action.

Auditing within Couchbase Server comes with additional security and performance controls:

- Auditing is off by default and can be enabled dynamically at runtime. When enabled, all administrative events are always audited, but application events (query statements, document read/write access) can be enabled at a more granular level.
- A user whitelist can be provided to further filter audit events.
- Events are audited asynchronously.
- Audit records are stored as JSON in a separate log file that should be tightly secured on the filesystem.
- Changes to the audit settings are restricted to specific administrative roles and are themselves audited.

These audit logs also need to be integrated into a reporting environment and monitored constantly. **The regulatory requirements on reporting are growing.** For example, GDPR introduces a requirement that all organizations must report certain types of personal data breach to the relevant supervisory authority **within 72 hours of becoming aware of the breach**. If the breach is likely to result in a high risk of adversely affecting individuals' rights and freedoms, **organizations must also inform those individuals without undue delay**. As mentioned previously, the Couchbase Data Platform is able to identify users and their data which can be used to assess the scale of a breach and/or to notify the relevant parties.

PII auditing and reporting best practice cheat sheet

Organizations have undertaken the following preparatory work in order to act in the event of data loss:

1. **Recognize a personal data breach** and understand that a personal data breach isn't only about loss or theft of personal data.
2. **Prepare a response plan** for addressing any personal data breaches that occur.
3. **Allocate responsibility for managing breaches** to a dedicated person or team.
4. **Know how to escalate a security incident** to the appropriate person or team in an organization.
5. Have processes in place to **assess the likely risk to individuals** as a result of a breach.
6. Know the relevant supervisory authority and have a process in place for notification within the specified timeframe.
7. **Have a process to inform affected individuals** about a breach when it is likely to result in a high risk to their rights and freedoms.
8. Know how to **inform affected individuals without undue delay** and know what information about a breach should be provided to individuals.
9. **Document all breaches**, even if they don't all need to be reported.

Conclusion and references

This document introduced the general regulatory and security environment organizations face and the drivers for stronger security and safeguards for sensitive data, especially personally identifiable information (PII). The three core drivers include **increased cybersecurity threats, tougher regulatory response, and security's commercial role in building trust with customers.**

With a focus on the Couchbase Data Platform, this document introduced the key functions for securing data as well as best practices for organizations to employ to create secure data environments. In addition, we recommend working closely with the Couchbase Professional Services team to ensure environments are configured not only for security requirements but also for performance and operations.

Finally, while the Couchbase Data Platform is central to securing data and PII, it does need to fit into a wider security model for the whole organization. A true layered defense includes the database and links and integrates with all the other parts of the organization.

Next steps:

- Contact us to discuss your security and GDPR requirements.
- Couchbase Security Health Check

Additional resources:

- Couchbase security documentation
- Couchbase corporate security
- Couchbase and GDPR
- Webinar: "GDPR: It's All About Digital and Digital Is All About Trust"
- Ebook: "Don't Waste Your GDPR Effort on Narrow Compliance"
- Datasheet: GDPR and the Couchbase Data Platform

Disclaimer

The information in this document may not be construed or used as legal advice about the content, interpretation, or application of any law, regulation, or regulatory guideline. Customers and prospective customers must seek their own legal counsel to understand the applicability of any law or regulation on their processing of personal data, including through the use of any vendor's products or services.

Appendix A: Couchbase security checklist

1. **Access control:**
   - ❑ Implement Role-Based Access Control
   - ❑ Create unique user accounts for each individual and application that accesses the platform, and assign roles following a principle of least privilege
   - ❑ Leverage strongest available authentication mechanisms
   - ❑ Ensure secure storage and transfer of credentials or certificates

2. **Secure network communication:**
   - ❑ Enable encryption for all network connections and block insecure ports
   - ❑ Replace self-signed certificates with external CA certificates

3. **Secure data storage:**
   - ❑ Encryption at rest
   - ❑ Data-level encryption

4. **Limit data exposure:**
   - ❑ Control geographic distribution of data
   - ❑ Regularly scan for sensitive data
   - ❑ Mask/redact sensitive and PII data
   - ❑ Leverage log redaction when collecting and transferring logs

5. **Auditing:**
   - ❑ Enable auditing
   - ❑ Review audit logs manually and programmatically for anomalies

6. **Regular review:**
   - ❑ Assess security from core to edge and perform regular security health checks
   - ❑ Review enhancements provided with each new version of the Couchbase Data Platform
   - ❑ Apply upgrades to the Couchbase Data Platform and SDK software as well as application frameworks, operating systems, networking infrastructure, etc.

Appendix B: More about PII

Personally identifiable information **(PII)** is any information that can be used to uniquely **identify, contact, or locate an individual, or can be used with other sources to uniquely identify a person.**

PII is expanding with the increasing scale and breath of digitization and digital activity and is growing to include biometric, visual, genomic, and device information.

Regulations such as the European Union's General Data Protection Regulation (GDPR) are expanding the definition of PII to reflect digital activity as well as giving individuals much stronger rights over PII. It is important to note that different regulation or regulatory bodies **will have varying definitions of exactly what constitutes PII.**

Typical PII:

- First or last name (if common)
- Date of birth
- Country, state, or city of residence
- Credit card numbers
- Immunization history/medical records
- Age
- Telephone numbers
- Email addresses
- Gender
- Race
- Criminal record

Potential PII in digital:

- IP address
- Cookies
- Device identifiers (computing, mobile, other connected devices)
- Comments (on Facebook, blogs, etc.)
- Photos of people
- Video/Audio (including live cams)
- Geolocation or mapping data
- Mobile app user data web tracking
- And more ...

Appendix C: The 7 Privacy by Design principles

1. **Proactive not reactive** – the Privacy by Design approach is characterized by proactive rather than reactive measures. Taking a clear commitment, at the highest levels, to set and enforce high standards of privacy, generally higher than the standards set out by global laws and regulation.

2. **Privacy as the default** – seeks to deliver the maximum degree of privacy by ensuring that personal data are automatically protected in any given IT system or business practice.

3. **Privacy by Design** – is embedded into the design and architecture of IT systems and business practices. The result is that privacy becomes an essential component of the core functionality being delivered.

4. **Full functionality** – Privacy by Design seeks to accommodate all legitimate interests and objectives in a positive-sum "win-win" manner. Privacy by Design avoids the pretense of false dichotomies, such as privacy vs. security, etc.

5. **End-to-end security** – having been embedded into the system prior to the first element of information being collected, extends securely throughout the entire lifecycle of the data involved.

6. **Visibility and transparency** – Privacy by Design seeks to assure all stakeholders that whatever the business practice or technology involved, it is in fact, operating according to the stated promises and objectives, subject to independent verification.

7. **Respect for user privacy** – Privacy by Design requires the development process keep the interests of the individual uppermost by offering such measures as strong privacy defaults, appropriate notice, and empowering user-friendly options.

2440 West El Camino Real | Ste 600
Mountain View, California 94040

1-650-417-7500

www.couchbase.com

About Couchbase