GDPR: Impact to Your Data Management Landscape

How MongoDB Can Put You on a Fast Track to Compliance

Table of Contents

Introduction

Cyber-crime is forecast to cost the global economy $6 trillion by 2021, up from $3 trillion in 2016. Described by some as the "greatest threat to every company in the world", public concern for the safety of data is growing – not just in how criminals might use stolen data to commit fraud, but also in how personal data is used by the organizations we engage with. Many people are asking whether data provided in exchange for goods, services, and employment could be used to:

- Damage our reputations?
- Deny us access to the healthcare or financial services we might need?
- Discriminate against us based on our political views, religion, associations, or ethnicity?
- Reduce our autonomy, freedom, and individuality?

The European Union (EU) General Data Protection Regulation (GDPR) 2016/679 is designed to confront these concerns. Protection and privacy of individuals – "data subjects" in GDPR terminology – becomes not just a legal obligation placed on organizations collecting and processing our data, but also entrenches data privacy as a fundamental human right of all EU citizens.

The GDPR is a global regulation and applies to any organization handling EU citizen data, irrespective of physical presence within the EU. It was introduced May 24, 2016, and enforcement started on May 25, 2018. Fines for noncompliance are up to €20m or 4% of global revenues, whichever is greater, in addition to compensation payable to data subjects, and the suspension of any further collection and processing of EU citizen data.

A range of requirements and controls are defined by the GDPR to govern how organizations collect, store, process, retain, and share the personal data of EU citizens. However, Gartner predicts that more than 50% of companies affected by the GDPR will not be in full compliance with its requirements by the end of 2018 – 6+ months after the regulation came into force.

In this white paper, we explore the specific requirements mandated for data protection as part of the GDPR, and discuss how MongoDB can provide the core technology foundations to help organizations accelerate their path to

compliance.

Disclaimer

For a full description of the GDPR's regulations, roles, and responsibilities, it is recommended that readers refer to the text of the GDPR (Regulation (EU) 2016/679), available from the Official Journal of the European Union, and refer to legal counsel for the interpretation of how the regulations apply to their organization. Further, in order to effectively achieve the functionality described below, it is critical to ensure that the database is implemented according to the specifications and instructions detailed in the MongoDB security documentation. Readers should consider engaging MongoDB Global Consulting Services to assist with implementation.

GDPR Primer

The former EU data protection legislation (Data Protection Directive 95/46/EC) was introduced back in 1995, but was increasingly regarded as insufficient, both for today's privacy demands, and those envisaged in the future:

- Implementation varied across each member state, creating complexity, uncertainty, and cost. Inconsistencies affected both user trust in an emerging digital economy, and EU competitiveness in the global market.
- Technology enhancements over the past 20+ years now allow both private enterprises and public authorities to collect and make use of personal data on an unprecedented scale in order to pursue their activities. The emergence of social networking, cloud computing, eCommerce, web services, mobile devices and apps, Internet of Things, machine learning, and many more render the existing regulation inadequate.

The reforms introduced by the GDPR are designed to provide EU citizens with more control over their own personal data. In this context, the scope of personal data has been expanded – it includes anything that can uniquely identify an individual, such as a name, an identification number, location data, an online identifier, or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural, or social identity of that individual.

In EU research, nine out of ten Europeans had expressed concern about mobile apps collecting personal data without their consent, and seven out of ten worried about the potential use that companies may make of the data that they disclosed. The GDPR attempts to address these concerns through a range of new measures:

- Individuals must provide explicit consent to data collection – "consent by default" is no longer valid. The organization seeking consent must also provide clear information on how that data will be used, for how long it will be retained, and how it will be shared with third parties. Individuals can retract consent at any time, without prejudice. Additional permissions must be requested from the individual if the data is to be used for processing purposes beyond the original consent.
- A "right to be forgotten", also known as "right to erasure", requires deletion of data when owners ask for it to no longer be retained, and there is no legitimate reason for an organization to refuse the request.
- Organizations must provide easier access to an individual's data, enabling them to review what data is stored about them and how it is processed, who it is shared with, along with the ability to migrate that data between service providers without restriction.
- A right to review is required for how automated decisions computed against personal data have been made, for example, by machine learning algorithms declining transactions based on risk scores.
- Disclosure within 72 hours must be made to a member state's "supervisory body" (a member state's independent public authority overseeing GDPR implementation) when personal data has been breached, enabling individuals to be informed and take appropriate remedial action.
- Data protection has to be by design and by default, requiring data protection controls to be built into products and services from the earliest stage of development, and the adoption of privacy-friendly default settings in all applications collecting personal data.

- Punitive financial recourse (e.g., 4% of global revenue or €20m) will be made against any organization proven not to comply with the regulations.

The new regulations seek to provide clarity and consistency in how privacy rules are applied, not just across the EU, but also globally to every organization processing citizen data as part of offering products and services in the EU. Some organizations may regard the GDPR as being onerous, but for others, it will present an opportunity to build greater trust with their customers, and provide a source of competitive advantage. We explore this further in the Customer Experience Transformation section later in the paper.

The GDPR introduces specific terminology to define roles and responsibilities within organizations, including:

- **Data Protection Officer (DPO)**, an individual employed by the data controller or processor, with responsibility for advising on GDPR regulation, reporting to the highest management level. The DPO is ultimately answerable to the local supervisory authority.
- **Data controller**, typically the organization with whom the data subject (the individual) is sharing the data.
- **Data processor**, an organization and/or individual working on behalf of the controller, e.g., a direct employee such as a business analyst or a developer, or an external service provider, such as a credit rating agency or a payroll processor. A data processor is any entity or individual with access to personal data.

GDPR's Definition of a Data Breach

It is very important to understand what a data breach means in context of this new regulation. The GDPR applies a much broader definition than only loss of confidentiality or unauthorized processing of personal data, demonstrating that data protection methods extend beyond narrow concepts of access. It also encompasses availability and integrity. The GDPR text states:

> *"'personal data breach' means a breach of security leading to the accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to, personal data transmitted, stored or otherwise processed"*

The following section of the paper examines GDPR requirements, and maps them back to the required database capabilities. Please note that the list below is illustrative only, and is not designed to be exhaustive.

Mapping GDPR to Required Database Capabilities

Like other regulations designed to enforce data security and privacy standards (e.g., HIPAA, PCI DSS, SOX, FISMA, FERPA), GDPR compliance can be achieved only by applying a combination of controls that we can summarize as People, Processes, and Products:

- "People" defines specific roles, responsibilities, and accountability.
- "Processes" defines operating principles and business practices.
- "Products" defines technologies used for data storage and processing.

As with any data security regulation, enabling controls in a database storing personal data is just one step towards compliance – people and processes also are essential. There are, however, specific requirements stated in the GDPR text that define a set of controls organizations need to implement across their data management landscape. We can group these requirements into three areas:

- **Discover:** scope data subjects to the regulation.
- **Defend:** implement measures to protect discovered data.
- **Detect:** identify a breach against that data, and remediate security and process gaps.

Discover

Before implementing security controls, an organization first needs to identify personal data stored in its databases, and for how long the organization is permitted to retain that data. They also need to assess the potential impact to the individual, should the personal data be disclosed to an unauthorized party.

Identification of Impact to Personal Data

The GDPR requires organizations to undertake a Data Protection Impact Assessment, documented in Article 35 (clause 1) of the GDPR text, stating:

> *"Where a type of processing in particular using new technologies, and taking into account the nature, scope, context and purposes of the processing, is likely to result in a high risk to the rights and freedoms of natural persons, the controller shall, prior to the processing, carry out an assessment of the impact of the envisaged processing operations on the protection of personal data."*

It is therefore important to have access to tools that enable the data controller to quickly and conveniently review their database content, and as part of an ongoing discovery process, to inspect what additional data will be captured as new services are under development.

Retention of Personal Data

As noted in "Information to be Provided", Article 13 (clause 2a), the GDPR text specifies that at the time data is collected from an individual, the organization must state:

> *"the period for which the personal data will be stored, or if that is not possible, the criteria used to determine that period"*

Therefore, a required capability that the organization needs to implement is the ability to identify personal data, and securely erase it from the database once the expiration period has been reached, or an individual specifically requests erasure. As a result, storage, including backups, should have the ability to provably erase data as requested by the owner.

Defend

Once the organization has conducted its Discover phase, with an Impact Assessment and expiration policies defined, they need to implement the controls that will protect citizen data.

General Security Requirements of the GDPR

The "Security of Processing", Article 32 (clause 1) provides an overview of security controls an organization needs to enforce:

> *"….the controller and the processor shall implement appropriate technical and organisational measures to ensure a level of security appropriate to the risk, including inter alia as appropriate:*
>
> *(a) the pseudonymisation and encryption of personal data; (b) the ability to ensure the ongoing confidentiality, integrity, availability and resilience of processing systems and services; (c) the ability to restore the availability and access to personal data in a timely manner in the event of a physical or technical incident; (d) a process for regularly testing, assessing and evaluating the effectiveness of technical and organisational measures for ensuring the security of the processing."*

Each of the bulleted clauses is further expanded upon within the GDPR text, as follows.

Access Control

The GDPR emphasizes the importance of ensuring that only authorized users can access personal data. As stated in the text "Data Protection by Design and by Default", Article 25 (clause 2):

> *"The controller shall implement appropriate technical and organisational measures for ensuring that, by default, only personal data which are necessary for each specific purpose of the processing are processed"*

This requirement is further reinforced in Article 29, "Processing Under the Authority of the Controller or Processor", stating

> *"The processor and any person acting under the authority of the controller or of the processor, who has access to personal data, shall not process those data except on instructions from the controller…."*

Within the database, it should be possible to enforce authentication controls so that only clients (e.g., users,

applications, administrators) authorized by the data processor can access the data. The database should also allow data controllers to define the specific roles, responsibilities, and duties each client can perform against the data. For example, some clients may be permitted to read all of the source data collected on a data subject, while others may only have permissions to access aggregated data that contains no reference back to personal identifiers. This approach permits a fine-grained segregation of duties and privileges for each data processor.

Pseudonymisation & Encryption

In the event of a breach, the pseudonymisation and encryption of data is designed to prevent the identification of any specific individual from compromised data. In the recitals section of the GDPR text, pseudonymisation means:

> *"….the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information"*

Clause 28 of the general regulations states:

> *"The application of pseudonymisation to personal data can reduce the risks to the data subjects concerned and help controllers and processors to meet their data-protection obligations."*

One of the most effective and efficient means of pseudonymising data is based on the access control privileges defined in the previous step. The database redacts personal identifiers by filtering query results returned to applications.

Encryption is specifically referenced in Article 32 (clause 1) referenced above. The advantages of encryption are further expanded in the text for "Communication of a Personal Data Breach to the Data Subject", Article 34 (clause 3a), stating communication to the data subject **is not** required if:

> *"the controller has implemented appropriate technical and organisational protection measures, and those measures were applied to the personal data affected by the personal data breach, in particular those that render the personal data unintelligible to any person who is not authorised to access it, such as encryption;"*

The database should provide a means to encrypt both data "in-transit" using network connections, and data "at-rest" using storage and backups.

Resilience and Disaster Recovery

As stated in bullets B and C in the "The Security of Processing", Article 32 cited above, systems and service availability, along with a means to restore data in a timely fashion, are both core operational requirements of the GDPR.

As a result, the database needs to offer fault tolerance to systems failures, along with backup and recovery mechanisms to enable disaster recovery.

Data Sovereignty: Data Transfers Outside of the EU

Chapter 5 of the GDPR is dedicated to how the transfer of personal data outside of the EU should be handled – defining when such transfers are permissible and when they are not. Key to understanding data transfer is that EU citizen rights under the GDPR accompany the data to wherever it is moved globally, where the same safeguards must be applied. To summarize the chapter, Article 45 (clause 1) states:

> *"A transfer of personal data to a third country or an international organisation may take place where the Commission has decided that the third country, a territory or one or more specified sectors within that third country, or the international organisation in question ensures an adequate level of protection."*

To support globally distributed applications, organizations are increasingly distributing data to data centers and cloud facilities located in multiple countries across the globe. In context of the GDPR, it should be possible for the database to enforce data sovereignty policies by only distributing and storing EU citizen data to regions recognized as complying with the regulation.

Detect

In the event of a data breach, the organization must be able, in timely fashion, to detect and report on the issue, and also generate a record of what activities had been performed against the data.

Monitoring and Reporting

Monitoring is always critical to identifying potential exploits. The closer to real time, the better chance of limiting the impact of data breaches. For example, sudden peaks in database resource consumption can indicate an attack in progress at the very moment it happens.

In the GDPR text "Notification of a Personal Data Breach to the Supervisory Authority", Article 33 (clause 1), it is stated:

> *"In the case of a personal data breach, the controller shall without undue delay and, where feasible, not later than 72 hours after having become aware of it, notify the personal data breach to the supervisory authority…."*

As a result, the database should offer management tools that enable constant monitoring of database behavior to proactively mitigate threats, and that enable the organization to report on any breaches within the specified timeframes.

Auditing

"Data Protection by Design and by Default", Article 25 (clause 2) emphasizes the requirement to maintain a log of activities performed against the data:

> *"….Each controller and, where applicable, the controller's representative, shall maintain a record of processing activities under its responsibility"*

"Processor", Article 28 (clause 3H) further expands on the requirement for auditing, stating that the data processor:

> *"makes available to the controller all information necessary to demonstrate compliance with the obligations laid down in this Article and allow for and contribute to audits, including inspections, conducted by the controller or another auditor mandated by the controller."*

The database needs to offer a mechanism to record database activity, and present that activity for forensic analysis when requested by the controller.

How MongoDB Can Help Meet GDPR Requirements

While data protection regulations such as GDPR, HIPAA, PCI-DSS, and others stipulate requirements that are unique to specific regions, industries or applications, there are foundational requirements common across all of the directives, including:

- Restricting data access, enforced via predefined privileges and roles
- Measures to protect against the accidental or malicious disclosure, loss, destruction, or damage of personal data
- The separation of duties when accessing and processing data
- Recording user, administrative staff, and application activities with a database

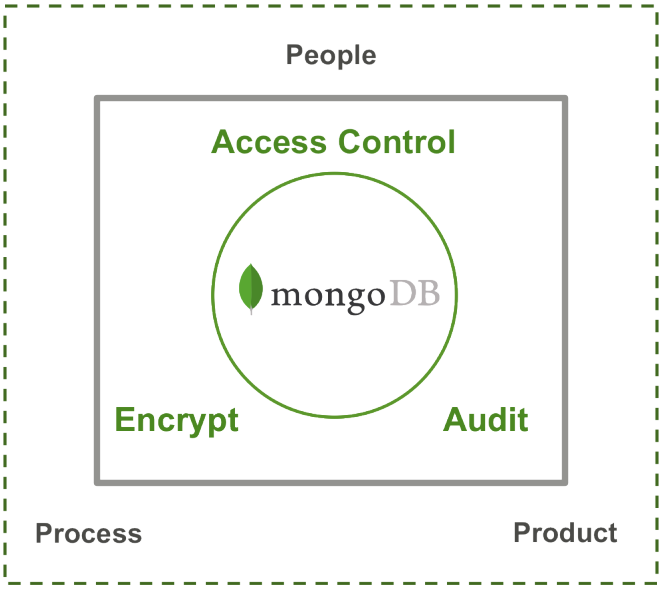

Figure 1: MongoDB End to End Security Architecture

These requirements inform the security architecture of MongoDB, with best practices for the implementation of a secure, compliant data management platform.

Using the advanced security features available in MongoDB Enterprise Advanced and the MongoDB Atlas cloud database service, organizations have extensive capabilities to implement the data discovery, defense, and detection requirements demanded by the GDPR.

Discover	Defend	Detect
Identify Personal Data • MongoDB Compass • Expressive Queries & Analytics • Schema Governance	**Access Control** • Authentication (i.e. LDAP, Kerberos) • Authorization (RBAC) • IP Whitelisting & VPC Peering	**Monitor & Report** • Real-Time Alerting
Personal Data Retention • TTL Indexes	**Pseudonymisation & Encryption** • Read-Only Views • Log Redaction • TLS/SSL Network Encryption • Encrypted Storage Engine	**Audit** • MongoDB Audit Log • MongoDB Change Streams
	Resilience & DR • Replica Sets • MongoDB PIT Backup & Recovery	
	Data Sovereignty • MongoDB Zones	
MongoDB Training & Global Consulting		

Table 1: Mapping GDPR requirements to MongoDB Enterprise Advanced capabilities

Discover

Identification of Personal Data

There are multiple ways to inspect database content. The most common method is to query the database and extract all records to identify the tables and rows (collections and documents, in MongoDB terminology) containing user data. However, this approach also requires significant manual analysis of the schema to track what data is stored, and where, while imposing processing overhead on the database itself.

MongoDB provides a much simpler approach with Compass, the GUI for MongoDB. Compass enables users to visually explore their data, providing a graphical view of their MongoDB schema by sampling a subset of documents from a collection, thereby minimizing database overhead and presenting results to the user almost instantly.

Schema visualization with MongoDB Compass enables the user to quickly explore their schema to understand the frequency, types, and ranges of fields in each data set. The user doesn't need to be conversant with the MongoDB query language – powerful ad-hoc queries can be constructed through a point and click interface, opening up the discovery and data loss prevention process beyond developers and DBAs to Data Protection Officers and other business users. Data governance controls enforced by MongoDB ensure that once a schema for storing customer data has been defined and documented, no new fields can be added without first updating schema validation rules.

Beyond Compass, the MongoDB query language and rich secondary indexes enable users to query and analyze the data in multiple ways. Data can be accessed by single keys, ranges, text search, graph, and geospatial queries through to complex aggregations, returning responses in milliseconds. Data can be dynamically enriched with elements such as user identity, location, and last access time to add context to Personally Identifiable Information (PII) , providing behavioral insights and actionable customer intelligence. Complex queries are executed natively in the database without having to use additional analytics frameworks or tools, and avoiding the latency that comes from ETL processes that are necessary to move data between operational and analytical systems in legacy enterprise architectures.

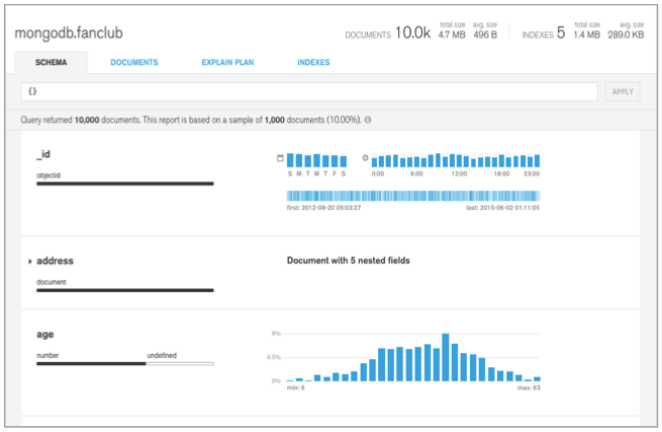

Figure 2: Data discovery with MongoDB Compass GUI-based schema exploration

Retention of Personal Data

Through the use of the special-purpose TTL (Time-To-Live) index, administrators can automate the expiration of EU citizen data from a database. By configuring the required retention period against a date field in the document (i.e. the date on which the user data was collected or last

accessed), MongoDB will delete the document once the period has been reached, using an automated background process that runs against the database every 60 seconds.

Compared to implementing expiration code at the application level, which must then regularly scan the database to find records that need to be deleted, the MongoDB TTL index dramatically simplifies the enforcement of data expiration policies. It also imposes significantly lower database overhead.

Defend

Access Control

Access control to a database can be separated into two distinct stages: 1. Authentication, designed to confirm the identity of clients accessing the database. 2. Authorization, governing what that client is entitled to do once they have access to the database, such as reading data, writing data, performing administrative and maintenance activities, and more.

MongoDB Authentication

MongoDB provides multiple authentication methods, allowing the approach best suited to meet the requirements of different environments. Authentication can be managed from the database itself, or through integration with external authentication mechanisms.

MongoDB Atlas enforces in-database authentication via the SHA2 standard and LDAP. As the MongoDB Atlas service runs on public cloud platforms, it also implements additional security controls to reduce the risk of unauthorized access. An Atlas cluster by default will disallow direct access from the internet. Each Atlas cluster is deployed within a virtual private environment (e.g., AWS or GCP Virtual Private Cloud, Azure Virtual Network), and that private environment is by default configured to allow no inbound access. Also IP whitelisting can be used to restrict network access to a database (i.e., application servers are prevented from accessing the database unless their IP address has been added to the whitelist for the appropriate MongoDB Atlas group) The Atlas AWS VPC peering option allows peering an organization's Atlas network to its own AWS VPC network, thereby ensuring network traffic never traverses the public internet, and instead uses the internal private network.

MongoDB Enterprise Advanced also allows SHA2 authentication, LDAP, and IP whitelisting, with additional integration options for Kerberos, or x.509 PKI certificates.

LDAP is widely used by many organizations to standardize and simplify the way large numbers of users are managed across internal systems and applications. In many cases, LDAP is also used as the centralized authority for user access control to ensure that internal security policies are compliant with corporate and regulatory guidelines. With LDAP integration, MongoDB Enterprise Advanced and MongoDB Atlas can authenticate and authorize users directly against existing LDAP infrastructure to leverage centralised access control architectures.

MongoDB Authorization

Over ten predefined roles supporting common user and administrator database privileges provide Role Based Access Control (RBAC) capabilities. With MongoDB, these can be further customised through User Defined Roles, enabling administrators to assign fine-grained privileges to clients, based on their respective data access and processing needs. To simplify account provisioning and maintenance, roles can be delegated across teams, ensuring the enforcement of consistent policies across specific data processing functions within the organization.

MongoDB also supports authorization via LDAP, in addition to authentication discussed above. This enables existing user privileges stored in a LDAP server to be mapped to MongoDB roles, without recreating users in MongoDB itself. This integration strengthens and simplifies access control by enforcing centralised processes.

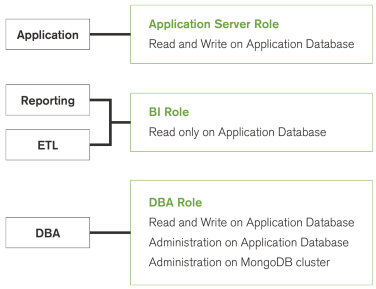

Figure 3: MongoDB user defined roles permit separation of duties across different classes of clients

Review the Authorization section of the documentation to learn more about role-based access control in MongoDB.

Pseudonymisation & Encryption

As discussed earlier, pseudonymisation and encryption of data is designed to prevent the identification of any specific individual in the event of data being accessed by an unauthorized party.

Pseudonymisation

MongoDB provides multiple levels of pseudonymisation. Through read-only views, MongoDB can automatically filter out specific fields, such as those containing PII of citizens when a database is queried. Rather than query collections directly, clients can be granted access only to specific, predefined views of the data. Permissions granted against the view are specified separately from permissions granted to the underlying collection, and so clients with different access privileges can be granted different views of the data.

Read-only views allow the inclusion or exclusion of fields, masking of field values, filtering, schema transformation, grouping, sorting, limiting, and joining of data across multiple collections. Read-only views are transparent to the application accessing the data, and do not modify the underlying raw data in any way.

MongoDB Enterprise Advanced can also be configured with log redaction to prevent potentially sensitive information, such as personal identifiers, from being written to the database's diagnostic log. Developers and DBAs who may need to access the logs for database performance optimization or maintenance tasks still get visibility to metadata, such as error or operation codes, line numbers, and source file names, but are unable to see any personal data associated with database events.

Encryption

Encryption can protect data in transit and at rest, enabling only authorized access. Should unauthorized users gain access to a network, server, filesystem or database the data still can be protected with encryption keys.

Support for Transport Layer Security (TLS) allows clients to connect to MongoDB over an encrypted network channel, protecting data in transit. In addition, MongoDB encrypts data at rest in persistent storage and in backups.

Using the MongoDB Atlas managed database service, TLS is the default and cannot be disabled. Traffic from clients to Atlas, and between Atlas cluster nodes, is authenticated and encrypted.

MongoDB Enterprise Advanced and MongoDB Altas also offer the Encrypted Storage Engine, making the protection of data at-rest an integral feature of the database. By natively encrypting database files on disk, administrators reduce both the management and performance overhead of external encryption options, while providing an additional level of defense. Only those staff with the appropriate database credentials can access encrypted personal data. Access to the database file on the server would not expose any stored personal information.

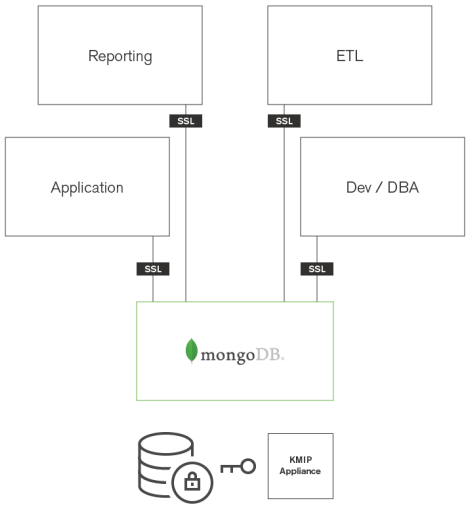

Figure 4: End to End Encryption – Data In-Flight and Data At-Rest

The storage engine encrypts each database with a separate key. MongoDB recommends encryption keys be rotated and replaced at regular intervals, and by performing rolling restarts of the replica set, keys can be rotated without database downtime. Database files themselves do not need to be re-encrypted when using a Key Management Interoperability Protocol (KMIP) service, thereby also avoiding the performance overhead incurred with key rotation.

Refer to the documentation to learn more about encryption in MongoDB.

Resilience and Disaster Recovery

To protect service availability and recover from events that cause data corruption or loss, MongoDB offers fault tolerance to systems failures, along with backup and recovery tools for disaster recovery.

Resilience

Using native replication, MongoDB maintains multiple copies of data in what are called replica sets. A replica set is a fully self-healing cluster distributed across multiple nodes to eliminate single points of failure. In the event a node fails, replica failover is fully automated, eliminating the need for administrators to intervene manually to restore database availability.

The number of replicas in a MongoDB replica set is configurable: a larger number of replicas will provide increased data availability and protection against database downtime (e.g., in case of multiple machine failures, rack failures, data center failures, or network partitions). Replica sets also provide operational flexibility by providing a way to upgrade hardware and software without requiring the database to be taken offline. Replica set members can be deployed both within and across physical data centers and cloud regions, providing resilience to regional failures.

Disaster Recovery

Data can be compromised by a number of unforeseen events: failure of the database or its underlying infrastructure, user error, malicious activity, or application bugs. With a backup and recovery strategy in place, administrators can restore business operations by quickly recovering their data, enabling the organization to meet regulatory and compliance obligations.

The operational tooling provided as part of MongoDB Enterprise Advanced and the MongoDB Atlas managed database service can continuously maintain database backups for you. If MongoDB experiences a failure, the most recent backup is only moments behind the operational system, minimizing exposure to data loss. The tooling offers point-in-time recovery of replica sets and cluster-wide snapshots of sharded clusters. These operations can be performed without any interruption to database service. Administrators can restore the database to precisely the moment needed, quickly and safely. Automation-driven restores allow a fully configured cluster to be re-deployed directly from the database snapshots in a just few clicks, speeding time to service recovery.

You can learn more about backup and restore in MongoDB Enterprise Advanced from the Ops Manager documentation, and from the documentation for MongoDB Atlas.

Data Sovereignty: Data Transfers Outside of the EU

To support data sovereignty requirements, MongoDB zones allow precise control over where personal data is physically

stored in a cluster. Zones can be configured to automatically "shard" (partition) the data based on the user's location – enabling administrators to isolate EU citizen data to physical facilities located only in those regions recognized as complying with the GDPR. If EU policies towards storing data in specific regions change, updating the shard key range can enable the database automatically to move personal data to alternative regions. Zoned sharding is available to MongoDB Atlas customers as part of the Global Sharding service.

Beyond geo-specific applications, zones can accommodate a range of deployment scenarios – for example supporting tiered storage deployment patterns for data lifecycle management, or segmenting data by application features or customers.

You can learn more about MongoDB zoned sharding from the documentation.

Detect

Monitoring

Proactive monitoring of all components within an application platform is always a best practice. System performance and availability depend on the timely detection and resolution of potential issues before they present problems to users. Sudden and unexpected peaks in memory and CPU utilization can, among other factors, be indicative of an attack, which can be mitigated if administrators are alerted in real time.

The operational tooling provided with MongoDB Enterprise Advanced and the MongoDB Atlas service provide deep operational visibility into database operations. Featuring charts, custom dashboards, and automated alerting, MongoDB's operational tooling tracks 100+ key database and systems health metrics including operations counters, memory, CPU, and storage consumption, replication and node status, open connections, queues, and many more. The metrics are securely reported to a management UI where they are processed, aggregated, alerted, and visualized in a browser, letting administrators easily track the health of MongoDB in real time. Metrics can also be pushed to Application Performance Management platforms such as AppDynamics and New Relic, supporting centralised visibility into the global IT estate.

Figure 5: Ops Manager Offers Charts, Custom Dashboards & Automated Alerting

Custom alerts can be generated when key metrics are out of range. These alerts can be sent via SMS and email, or integrated into existing incident management and collaboration systems such as PagerDuty, Slack, HipChat, and others to proactively warn of potential issues and help prevent outages or breaches.

The operational tooling also enables administrators to roll out upgrades and patches to the database without application downtime. Using the MongoDB Atlas database service, patches are automatically applied, removing the overhead of manual operator intervention.

Auditing

By maintaining audit trails, changes to personal data and database configuration can be captured for each client accessing the database, providing a log for compliance and forensic analysis by data controllers and supervisory authorities.

The MongoDB auditing framework logs all access and actions executed against the database, including:

- Administrative actions such as adding, modifying, and removing database users, schema operations, and backups.
- Authentication and authorization activities, including failed attempts at accessing personal data.
- Read and write operations to the database.

Administrators can construct and filter audit trails for any operation against MongoDB Enterprise Advanced. They

can capture all activities, or just a subset of actions, based on the requirements stipulated by the data controller and auditors. For example, it is possible to log and audit the identities of users who accessed specific documents, and any changes they made to the database during their session. Learn more from the MongoDB Enterprise Advanced auditing documentation.

Beyond auditing of database operations, MongoDB change streams allows controllers to be automatically notified of all changes to customer data in real time. A monitoring application can register for notifications whenever a collection containing customer data is modified, enabling the administrator to inspect what has changed, without constantly querying the entire collection to identify new or updated customer data.

Services to Help Your Teams Create a Secure Database Environment

The GDPR text explicitly states the requirement for training in the text "Binding Corporate Rules", Article 47 (clause 2n)

> *"the appropriate data protection training to personnel having permanent or regular access to personal data."*

MongoDB provides extensive training and consulting services to help customers apply best security practices:

- The MongoDB Security course is a no-cost, 3-week online training program delivered by MongoDB University.
- MongoDB University also offers a range of both public and private training for developers and operations teams, covering best practices in using and administering MongoDB.
- MongoDB Global Consulting Services offer a range of packages covering Health Checks, Production Readiness Assessments, and access to Dedicated Consulting Engineers. The MongoDB consulting engineers work directly with your teams to guide development and operations, ensuring skills transfer to your staff.

Case Studies

MongoDB has been downloaded over 35 million times and counts 50% of the Fortune 100 as commercial customers of MongoDB's products and services. Among the Fortune 500 and Global 500, MongoDB customers include:

- 40 of the top financial services institutions
- 15 of the top retailers
- 15 of the top telcos
- 15 of the top healthcare companies
- 10 of the top media and entertainment companies

MongoDB is used by enterprises of all sizes, and from all industries to build modern applications, often as part of digital transformation initiatives in the cloud. An example of such a company is Estates Gazette, the UK's leading commercial property data service.

Estates Gazette (EG)

The company's business was built on print media, with the Estates Gazette journal serving as the authoritative source on commercial property across the UK for well over a century. Back in the 1990s, the company was quick to identify the disruptive potential of the Internet, embracing it as a new channel for information distribution. Pairing its rich catalog of property data and market intelligence with new data sources from mobile and location services – and the ability to run sophisticated analytics across all of it in the cloud – the company is now accelerating its move into enriched market insights, complemented with decision support systems.

To power its digital transformation, Estates Gazette migrated from legacy relational databases to MongoDB, running an event-driven architecture and microservices, deployed to the Amazon Web Services (AWS) cloud. The company is also using MongoDB Enterprise Advanced with the Encrypted storage engine to extend its security profile, and prepare for the EU GDPR.

You can learn more by reading the Estates Gazette case study.

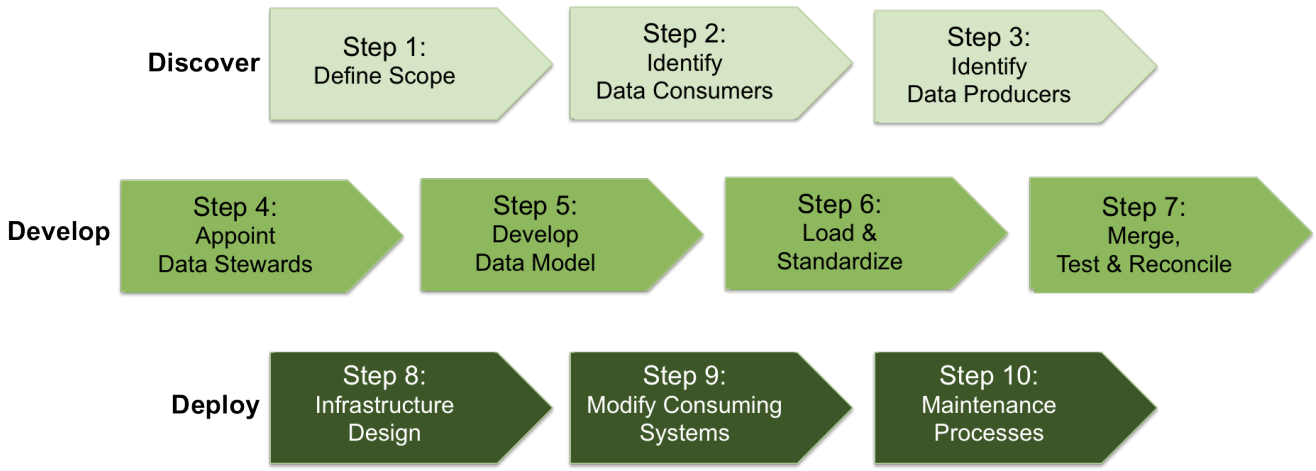

Figure 6: 10-step methodology to building a single customer view with MongoDB

Leading European Retailer

As part of its ongoing digital transformation that extends customer engagement beyond brick and mortar stores to mobile channels, the retailer with over 50,000 employees and €4.5bn in annual sales, was building a new mobile app offering opt-in marketing services to collect customer data, storing it in MongoDB.

As part of its GDPR readiness, the retailer employed MongoDB Global Consulting Services to advise on data protection best practices, taking advantage of the MongoDB Enterprise Advanced access controls, encryption, and auditing framework. By using MongoDB consultants in the design phase of the project, the retailer has been able to adopt a "security by design and by default" approach, while enhancing its security posture.

Using the GDPR for Customer Experience Transformation

To comply with the GDPR, organizations will need to identify all personal data within their systems. Forward-looking companies can leverage the regulations for personal data discovery processes to transform interactions with their customers.

Marketing and sales groups have long seen the value in aggregating data from multiple, disconnected systems into a single, holistic, real-time representation of their customer. This single view can help in enhancing customer insight and intelligence – with the ability to better understand and predict customer preferences, behaviors, and needs.

However, for many organizations, delivering a single view of the customer to the business has been elusive. Technology has been one limitation – for example, the rigid, tabular data model imposed by traditional relational databases inhibits the schema flexibility necessary to accommodate the diverse customer data sets contained in multiple source systems. But limitations extend beyond just the technology to include the business processes needed to deliver and maintain a single view.

MongoDB has been used in many single view projects across enterprises of all sizes and industries. Through the best practices observed and institutionalized over the years, MongoDB has developed a repeatable, 10-step methodology to successfully delivering a single view.

You can learn more by downloading the MongoDB single view whitepaper, covering:

- The 10-step methodology to delivering a single view
- The required technology capabilities and tools to accelerate project delivery
- Case studies from customers who have built transformational single view applications on MongoDB

Conclusion

It takes much more than security controls of a database to achieve GDPR compliance. However, MongoDB is offering a holistic vision of how database customers can accelerate

a path to meeting regulations scheduled for enforcement from May 2018.

Using the advanced security features available in MongoDB Enterprise Advanced and the MongoDB Atlas managed database service, organizations have extensive capabilities to implement the data discovery, defense, and detection requirements demanded by the GDPR. Methodologies used in successfully delivering customer single view projects can be used to support data discovery, and used to innovate in delivering a differentiated customer experience.

Further Reading

- MongoDB Security documentation
- MongoDB Security Architecture whitepaper
- MongoDB Atlas Security Controls whitepaper

We Can Help

We are the MongoDB experts. Over 5,700 organizations rely on our commercial products. We offer software and services to make your life easier:

MongoDB Enterprise Advanced is the best way to run MongoDB in your data center. It's a finely-tuned package of advanced software, support, certifications, and other services designed for the way you do business.

MongoDB Atlas is a database as a service for MongoDB, letting you focus on apps instead of ops. With MongoDB Atlas, you only pay for what you use with a convenient hourly billing model. With the click of a button, you can scale up and down when you need to, with no downtime, full security, and high performance.

MongoDB Stitch is a serverless platform which accelerates application development with simple, secure access to data and services from the client – getting your apps to market faster while reducing operational costs and effort.

MongoDB Mobile (Beta) MongoDB Mobile lets you store data where you need it, from IoT, iOS, and Android mobile devices to your backend – using a single database and query language.

MongoDB Cloud Manager is a cloud-based tool that helps you manage MongoDB on your own infrastructure. With automated provisioning, fine-grained monitoring, and continuous backups, you get a full management suite that reduces operational overhead, while maintaining full control over your databases.

MongoDB Consulting packages get you to production faster, help you tune performance in production, help you scale, and free you up to focus on your next release.

MongoDB Training helps you become a MongoDB expert, from design to operating mission-critical systems at scale. Whether you're a developer, DBA, or architect, we can make you better at MongoDB.

Resources

For more information, please visit mongodb.com or contact us at sales@mongodb.com.

Case Studies (mongodb.com/customers)
Presentations (mongodb.com/presentations)
Free Online Training (university.mongodb.com)
Webinars and Events (mongodb.com/events)
Documentation (docs.mongodb.com)
MongoDB Enterprise Download (mongodb.com/download)
MongoDB Atlas database as a service for MongoDB (mongodb.com/cloud)
MongoDB Stitch backend as a service (mongodb.com/cloud/stitch)

mongoDB

US 866-237-8815 ▪ INTL +1-650-440-4474 ▪ info@mongodb.com