Crittografia, pseudonimizzazione, tokenizzazione e anonimizzazione: una overview delle principali tecniche per elaborare i dati in modo sicuro

17 Marzo 2023 - 13 min. read

Data Security and Governance | *GDPR*

La data protection (o impropriamente “privacy”) e la data security sono oggi sempre più importanti, soprattutto con l’aumentare dei processi basati sulla raccolta di grandi quantità di dati (big data) e il generale aumento della digitalizzazione delle informazioni in ogni settore.

La **data protection** ruota attorno al diritto delle persone di controllare le proprie informazioni personali e di impedire che vengano divulgate ad altri illegalmente. Implica la **protezione delle informazioni personali** dall'accesso, l'utilizzo o la divulgazione da parte di entità non autorizzate e la garanzia che gli individui abbiano il controllo su come i loro dati personali vengono raccolti, utilizzati, condivisi e archiviati.

Per **data security**, invece, si intende la protezione dei dati da accessi non autorizzati, furto, danneggiamento o distruzione. Implica la **salvaguardia di sistemi informativi, reti e database da violazioni della sicurezza** e la garanzia che le informazioni sensibili siano protette da minacce sia interne che esterne.

In altre parole, data security è un concetto più ampio che si occupa genericamente di garantire la riservatezza e la protezione delle informazioni dal rischio di compromissione.

Con data protection si intende più nello specifico la protezione dei dati personali degli utenti, specialmente riguardo alle normative vigenti ed applicabili.

Per garantire la privacy e la sicurezza dei dati, le organizzazioni devono **implementare politiche e procedure appropriate**, come controlli degli accessi, crittografia, firewall e

monitoraggio della sicurezza, per impedire l'accesso non autorizzato ai dati personali e/o ai dati sensibili. Sia la privacy che la sicurezza dei dati sono necessarie per proteggere gli individui e le organizzazioni dai eventuali danni, e sono essenziali nell'era digitale odierna, in cui grandi quantità di dati personali e sensibili vengono raccolti e archiviati online.

In questo articolo, ci concentreremo sulla **protezione dei dati in termini di riservatezza, in particolare in relazione ai requisiti GDPR**, e spiegheremo i principali concetti di security che possono essere utilizzati per garantire agli utenti la privacy richiesta dalla legge.

Data security

La sicurezza dei dati è essenziale per diversi motivi: in primo luogo, aiuta a **proteggere la privacy delle persone impedendo l'accesso non autorizzato alle loro informazioni personali**. La protezione della privacy di un individuo è particolarmente importante in settori come sanità, finanza e public sector, dove i dati sensibili come le cartelle cliniche, le informazioni finanziarie e l'identificazione personale devono essere protetti dall'accesso non autorizzato.

È anche fondamentale per **mantenere l'integrità dei dati**; implementando misure di sicurezza come crittografia e backup dei dati, gli sviluppatori possono garantire che i dati non vengano manomessi o persi a causa di guasti hardware o software.

Infine, la sicurezza dei dati è essenziale per **mantenere la reputazione e la credibilità delle imprese**. Le violazioni dei dati e gli attacchi informatici possono avere gravi conseguenze per le aziende, tra cui perdite finanziarie, responsabilità legali e danni alla reputazione. Implementando solide misure di sicurezza dei dati, le aziende possono dimostrare il proprio impegno a proteggere i dati sensibili e mantenere la fiducia dei propri clienti.

L'aspetto centrale della data security è la capacità di elaborare e archiviare i dati per ridurre al minimo il rischio di esporre informazioni sensibili, consentendo il normale funzionamento dei sistemi e lo svolgimento delle attività di manutenzione e gestione richieste; limitando al contempo la disponibilità di dati sensibili al minor numero di sistemi e occhi consentiti dal caso d'uso.

Mantenere il giusto **equilibrio tra sicurezza e agilità degli operatori** è fondamentale perché entrambi sono essenziali per il buon funzionamento di un'organizzazione. Le misure di sicurezza sono necessarie per garantire la privacy richiesta dalla normativa; queste misure aiutano a garantire che un'organizzazione possa operare senza interruzioni e che le informazioni riservate siano mantenute al sicuro.

È essenziale garantire che gli operatori autorizzati possano svolgere le proprie mansioni in modo efficace ed efficiente. I dipendenti devono essere liberi di accedere alle informazioni e alle risorse di cui hanno bisogno per svolgere il proprio lavoro senza inutili ostacoli o limitazioni. Se gli operatori sono troppo limitati, potrebbero non essere in grado di svolgere i propri compiti in modo efficace, con conseguente diminuzione della produttività e del morale.

Pertanto, è essenziale mantenere un equilibrio tra la sicurezza e la capacità degli operatori di accedere alle informazioni di cui hanno bisogno per garantire che l'organizzazione possa operare in modo efficiente ed efficace proteggendo le proprie risorse e le informazioni sensibili.

Ci sono quattro tecniche fondamentali che possono essere utilizzate per produrre dataset che gli operatori possono manipolare in sicurezza in base al livello di informazioni di cui hanno bisogno: **crittografia**, **pseudonimizzazione**, **tokenizzazione** e **anonimizzazione**. Questo articolo esaminerà le tecniche sopra menzionate, con considerazioni generiche per la loro implementazione nelle architetture cloud.

Senza ulteriori indugi, Iniziamo la panoramica delle tecniche.

Crittografia

La crittografia è una tecnica ampiamente utilizzata per proteggere i dati e garantire la riservatezza, l'integrità e l'autenticità che si basa sulla trasformazione dei dati in un formato illeggibile da attori non autorizzati. Questa trasformazione si ottiene utilizzando vari algoritmi di crittografia come **Advanced Encryption Standard (AES)** e **Rivest–Shamir–Adleman (RSA)**. I dati possono essere decifrati solo utilizzando una chiave specifica nota solo alle parti autorizzate. La crittografia può essere utilizzata anche per garantire l'integrità e l'autenticità dei dati utilizzando firme digitali e funzioni hash.

Ai sensi del GDPR, le aziende devono proteggere i dati personali da accesso, modifica e divulgazione non autorizzati. La crittografia può essere utilizzata per raggiungere questo obiettivo crittografando i dati personali e garantendo che solo le parti autorizzate abbiano accesso alle chiavi di decrittazione. Ciò significa che anche se un'entità non autorizzata ottiene l'accesso ai dati, non potrà leggerli senza la chiave di decrittazione.

Un altro aspetto critico del **GDPR** è il diritto alla cancellazione, noto anche come diritto all'oblio. **Le aziende devono garantire che i dati personali vengano cancellati su richiesta dell'interessato o alla scadenza del periodo di conservazione**. La crittografia può essere utilizzata per raggiungere questo obiettivo cancellando in modo sicuro le chiavi di

decrittazione, rendendo di fatto illeggibili i dati crittografati. Operare cancellando la chiave può essere più veloce e portare a meno errori rispetto alla ricerca di tutte le occorrenze dei dati utente e alla loro eliminazione.

Pseudonimizzazione

La pseudonimizzazione è il processo mediante il quale si **impedisce a un individuo di essere identificato attraverso i propri dati**. Il GDPR è particolarmente severo per quanto riguarda la pseudonimizzazione: bisogna garantire l'assoluta impossibilità di risalire all'identità del titolare dei dati da parte di soggetti diversi dal titolare del trattamento.

Questa tecnica protegge i dati personali rendendo impossibile il collegamento all'identità individuale originale (senza possedere l'algoritmo o la tabella di pseudonimizzazione) pur consentendo l'utilizzo dei dati per scopi specifici. La pseudonimizzazione viene spesso utilizzata quando vi è la necessità di condividere dati ma dove è essenziale proteggere la privacy delle persone.

Esempi di casi in cui è possibile utilizzare la pseudonimizzazione includono la ricerca medica, le sperimentazioni cliniche, marketing e analisi dei social media, in particolare durante la creazione di set di dati per machine learning, report o statistiche.

Un buon algoritmo di pseudonimizzazione sostituisce le informazioni identificative di una persona, come il nome, l'indirizzo o la data di nascita, con uno pseudonimo o un altro identificatore artificiale.

Lo pseudonimo risultante è unico per l'individuo ma non rivela alcuna informazione di identificazione personale. I dati originali, l'algoritmo e/o la matrice di transcodifica vengono memorizzati separatamente, consentendo al sistema di funzionare normalmente.

Uno dei principali vantaggi della pseudonimizzazione è che consente la condivisione dei dati proteggendo la piena riservatezza dell'identità dell'interessato. Questa tecnica riduce anche i rischi associati alle violazioni dei dati, poiché anche se i dati pseudonimizzati vengono rubati, è impossibile ricollegarli a individui specifici.

Tuttavia, è essenziale notare che **la pseudonimizzazione non garantisce l'anonimato assoluto**. È ancora possibile identificare nuovamente i dati se qualcuno può accedere ai dati pseudonimizzati e originali.

Tokenizzazione

La tokenizzazione è un'altra tecnica utilizzata per elaborare i dati in modo sicuro e **comporta la sostituzione dei dati sensibili con un identificatore o un token univoco**. I dati

originali vengono archiviati in modo sicuro in un database indipendente, mentre il token rappresenta i dati in altri sistemi. La tokenizzazione è comunemente utilizzata nell'elaborazione dei pagamenti, dove è essenziale proteggere i dati finanziari sensibili, come i numeri di carta di credito, pur consentendo l'elaborazione delle transazioni.

La tokenizzazione in genere comporta l'utilizzo di un **algoritmo di tokenizzazione**, che genera un token univoco per ogni dato sensibile. Il token è in genere una stringa di caratteri generata casualmente o un valore hash. I dati originali vengono quindi crittografati e archiviati in modo sicuro in un sistema indipendente e ad accesso controllato.

Uno dei principali vantaggi della tokenizzazione è che fornisce un elevato livello di sicurezza per i dati sensibili, poiché i dati originali non vengono mai trasmessi o archiviati in forma non crittografata. Questa tecnica **semplifica anche la conformità alle normative sulla protezione dei dati**.

Tuttavia, è importante notare che **la tokenizzazione non fornisce l'anonimato assoluto**, poiché è ancora possibile ricollegare il token ai dati originali se qualcuno ha accesso sia al token che al data vault.

Anonimizzazione

L'anonimizzazione è una forma più estrema di elaborazione dei dati che comporta la **rimozione di tutte le informazioni identificative dai dati**. L'anonimizzazione viene in genere utilizzata quando non vi è alcuna necessità legittima di conservare i dati identificativi e dove è essenziale proteggere la privacy delle persone. Esempi di casi in cui è possibile utilizzare l'anonimizzazione includono l'analisi di dati sanitari in ambito di ricerca, l'analisi demografica e i sondaggi sull'opinione pubblica.

L'anonimizzazione in genere comporta la rimozione di qualsiasi informazione identificativa dai dati, come nomi, indirizzi o altre informazioni personali.
Secondo la maggior parte delle leggi e dei regolamenti, l'anonimizzazione produce gli stessi effetti dell'eliminazione dei dati o dell'eliminazione della chiave di crittografia utilizzata per crittografarli perché l'aspetto cruciale è rendere impossibile l'ottenimento delle informazioni identificative.

I dati risultanti vengono aggregati o riepilogati per fornire approfondimenti senza rivelare informazioni personali. Esistono varie tecniche per ottenere l'anonimizzazione dei dati, ad esempio il data masking, generalizzazione o la semplice eliminazione dei dati interessati.
Introduzione al GDPR su AWS
Amazon Web Services (AWS) offre vari servizi e infrastrutture che possono essere utilizzati

per implementare la gestione delle richieste GDPR; quello che segue è un elenco dei tipici servizi interessati:

- **Amazon S3**: Amazon S3 è un servizio di archiviazione di oggetti altamente scalabile e sicuro in grado di archiviare e gestire i dati personali. Con S3, puoi configurare facilmente le lifecycle policy, per eliminare o archiviare automaticamente i dati in base ai periodi di conservazione.
- **Amazon EC2/Fargate/Lambda**: Forniscono capacità di elaborazione scalabile nel cloud, consentendo di creare ed eseguire applicazioni che gestiscono le richieste GDPR. La tua applicazione può essere configurata per essere eseguita in un ambiente VPC (Virtual Private Cloud) sicuro, garantendo che i dati siano protetti e che il perimetro sia ben delimitato e controllato.
- **Amazon RDS**: Amazon RDS è un servizio di database relazionale gestito che fornisce un modo semplice per configurare, gestire e ridimensionare un database relazionale. RDS supporta una varietà di motori di database, tra cui MySQL, PostgreSQL e Oracle, semplificando l'archiviazione e la gestione dei dati personali. Supporta anche la crittografia at rest completamente gestita da AWS e facoltativamente la crittografia in transito se supportata dal DBMS.
- **Amazon Kinesis**: Amazon Kinesis è un servizio di streaming di dati in tempo reale che può essere utilizzato per raccogliere, elaborare e analizzare dati da varie fonti. Con Kinesis, puoi elaborare in modo rapido ed efficiente le richieste GDPR non appena arrivano, assicurandoti che vengano gestite in modo tempestivo ed efficiente.

Oltre a questi servizi, AWS fornisce anche l'infrastruttura per supportare la conformità al GDPR, tra cui:

- **AWS Identity and Access Management (IAM)**, che puoi utilizzare per controllare in modo granulare chi ha accesso alle risorse cloud, agli oggetti S3 e alle chiavi crittografiche gestite da AWS.
- **AWS Key Management Service (KMS)**: KMS è un servizio gestito che semplifica la creazione e il controllo delle chiavi di crittografia utilizzate per proteggere i dati personali archiviati in AWS. Con KMS, puoi creare e gestire chiavi, definire criteri e controllare l'accesso alle chiavi.
- **AWS CloudTrail**: CloudTrail è un servizio che consente di registrare, monitorare continuamente e conservare gli eventi relativi all'audit che si verificano nel tuo account AWS.

Nel complesso, AWS fornisce un solido insieme di servizi e infrastrutture che possono essere utilizzati per implementare la gestione delle richieste GDPR. Sfruttando questi servizi, puoi garantire che i dati personali vengano archiviati, elaborati e trasmessi in modo sicuro e conformi ai requisiti del GDPR.

Per implementare le tecniche precedentemente descritte in un'applicazione cloud, di solito dovresti:

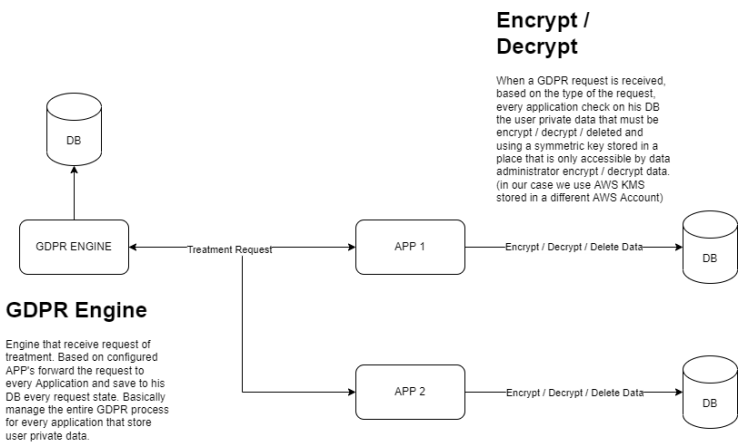
- Identificare i dati sensibili che devono essere protetti, comprese le informazioni di identificazione personale (PII), le informazioni finanziarie, le informazioni sulla salute o qualsiasi altro dato sensibile.
- Scegl una tecnica appropriata per il tipo di dati con cui stai lavorando e il livello di protezione richiesto.
- A seconda della tecnica scelta, potrebbe essere necessario aggiungere funzionalità specifiche al codice. Ad esempio, se si utilizza la crittografia, è necessario utilizzare una libreria di crittografia per crittografare i dati sensibili prima di archiviarli nel database. Allo stesso modo, se utilizzi la tokenizzazione, devi generare un token per ogni dato sensibile e archiviare i token nel tuo database invece dei dati originali (quando possibile)

Esiste anche un altro approccio: se stai creando un'applicazione basata su microservizi, puoi creare un servizio specifico per gestire la protezione dei dati e tutte le funzionalità specifiche della normativa.

Un progetto d'esempio: GDPR system centralizzato

Di seguito è riportato un esempio di come abbiamo costruito **un servizio per gestire la crittografia dei dati e richieste specifiche relative alla conformità GDPR**.

Questo sistema è progettato per **servire più applicazioni con più origini dati**. Il requisito principale è centralizzare tutto il lavoro relativo al GDPR e la gestione delle richieste degli utenti, come per esempio il "diritto all'oblio".



Servizio per gestire la crittografia dei dati e richieste specifiche relative alla conformità GDPR

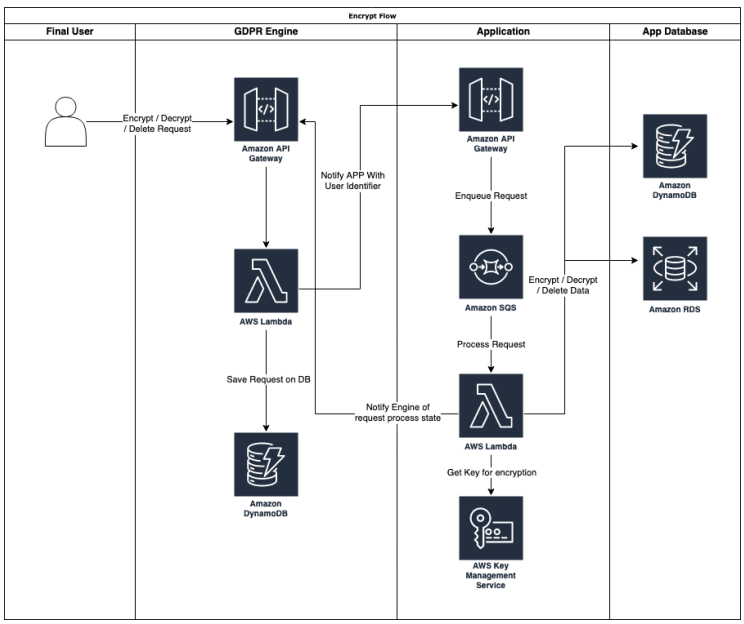
Il Core GDPR System centralizza, gestisce e inoltra ogni richiesta dell'utente alle APP configurate. Deve implementare un callback API che riceve la risposta della richiesta elaborata dall'applicazione. Questo perché il tempo per elaborare una richiesta GDPR potrebbe essere elevato a causa del numero di record e dati sensibili che devono essere elaborati.

Ogni APP deve implementare un'API che riceva la Richiesta GDPR e, in base al tipo di richiesta, cifrare/decifrare/eliminare i dati sensibili dal proprio DB. Questo lavoro deve sempre rispondere al sistema GDPR con il dettaglio e lo stato della richiesta elaborata. Ogni richiesta ha l'identificativo dell'utente che vuole che i suoi dati sensibili siano cancellati dal sistema.

La crittografia/decrittografia utilizza una chiave simmetrica che deve essere archiviata in un luogo sicuro a cui solo l'amministratore dei dati può accedere. (in un ambiente AWS, in genere utilizziamo una chiave AWS KMS archiviata in un account AWS diverso, accessibile solo dal codice dell'applicazione back-end e dall'amministratore dei dati).

Flusso di cifratura / decifratura / eliminazione

Vediamo come funziona il flusso di crittografia:



Come possiamo vedere dal diagramma di flusso sopra, un utente può **generare una richiesta di crittografia, decrittografia o eliminazione al motore GDPR.** Nel nostro caso, abbiamo deciso di creare questo motore sfruttando i servizi AWS Serverless Managed come:

- AWS API Gateway, che ci consente di esporre l'API
- AWS Lambda, che eseguirà il codice che gestisce tutte le richieste.
- AWS DynamoDB come archivio dati per il motore GDPR.

Quando viene ricevuta una richiesta di trattamento, il motore GDPR salva la richiesta nella tabella DynamoDB e controlla le APP configurate che devono essere notificate e, per ognuna, chiama l'API fornita.

Quando l'applicazione riceve la richiesta, la invia a una coda AWS SQS. Questo perché, nelle applicazioni che archiviano ed elaborano una grande quantità di dati con informazioni relative agli utenti, possiamo ricevere molte richieste relative al GDPR. A seconda della struttura e del volume dei dati, l'elaborazione di una singola richiesta potrebbe richiedere del tempo.

Utilizzando una coda, possiamo **disaccoppiare il servizio di elaborazione dall'applicazione principale**, consentendo loro di scalare in modo indipendente. Il disaccoppiamento con SQS può anche essere utile per creare solidi sistemi di retry e garantire che ogni richiesta venga conservata fino a quando non viene soddisfatta.

È inoltre possibile **rilevare i guasti e reindirizzare le richieste non riuscite a una coda dedicata** (DLQ) per modificare la potenza di calcolo del consumatore, ad esempio passando da una funzione Lambda a un servizio Fargate. Inoltre, possiamo rilevare fallimenti di secondo livello e allertare un operatore.

Per ulteriori informazioni sul disaccoppiamento dei servizi tramite SQS potete fare riferimento ad un mio precedente articolo dove viene illustrato nel dettaglio il meccanismo di trigger.

Ogni richiesta viene quindi elaborata in base al suo tipo:

- **Crittografare**. Utilizzerà una chiave KMS gestita dal cliente archiviata in un account diverso accessibile solo dal nostro back-end dell'applicazione e dall'amministratore dei dati per crittografare le informazioni sensibili dell'utente.
- **Decrittare**. utilizzerà la stessa chiave KMS per decrittografare le informazioni sensibili dell'utente.
- **Eliminare**. Rimuoverà o sostituirà con una stringa casuale i dati sensibili dell'utente.

Quando la richiesta viene elaborata, indipendentemente dallo stato positivo o negativo dell'elaborazione della richiesta, notificherà il motore GDPR, che salverà sulla tabella

DynamoDB il risultato.

Conclusioni

Abbiamo fatto una panoramica di alto livello delle principali tecniche disponibili per proteggere ed elaborare in modo sicuro i dati nelle moderne applicazioni cloud-native. Ma molte tecnologie, servizi e modelli di progettazione specifici possono essere utilizzati per creare applicazioni conformi al GDPR.
Se sei interessato a questo argomento, inviaci un messaggio, lasciaci un commento e seguici per altri articoli su GDPR e AWS!

About Proud2beCloud

Proud2beCloud è il blog di beSharp, APN Premier Consulting Partner italiano esperto nella progettazione, implementazione e gestione di infrastrutture Cloud complesse e servizi AWS avanzati. Prima di essere scrittori, siamo Solutions Architect che, dal 2007, lavorano quotidianamente con i servizi AWS. Siamo innovatori alla costante ricerca della soluzione più all'avanguardia per noi e per i nostri clienti. Su Proud2beCloud condividiamo regolarmente i nostri migliori spunti con chi come noi, per lavoro o per passione, lavora con il Cloud di AWS. Partecipa alla discussione!

Alessio Gandini

Cloud-native Development Line Manager @ beSharp, DevOps Engineer e AWS expert.Computer geek da quando avevo 6 anni, appassionato di informatica ed elettronica a tutto tondo. Ultimamente sto esplorando l'esperienza utente vocale e il mondo dell'IoT.Appassionato di cinema e grande consumatore di serie TV, videogiocatore della domenica.
