# Security in NoSQL Databases for Big Data in the Cloud: Systematic Mapping and LGPD Compliance Guidelines

Vitor Mayorca Camargo [0009-0008-7741-7468] and
Carlos José Maria Olguín [0000-0003-4030-9195]

Colegiado de Ciência da Computação
Universidade Estadual do Oeste do Paraná – UNIOESTE
Cascavel, PR – Brasil
vitor.camargo2@unioeste.br; carlos.olguin@unioeste.br

**Abstract.** The convergence of Big Data, NoSQL databases, and cloud computing introduces significant security vulnerabilities that conflict with the requirements of data protection legislation such as the Brazilian General Data Protection Law (LGPD). This work investigates the main security challenges of NoSQL databases used in Big Data contexts within cloud environments through a Systematic Mapping of the Literature (SM). The search strategy, applied to the ACM Digital Library and Scopus databases, led to the selection and analysis of 159 primary studies. Results show that the absence of native encryption, authentication failures, and injection vulnerabilities are the most common flaws, with research predominantly focused on MongoDB and a significant gap regarding audit and logging mechanisms. As a main contribution, based on the qualitative synthesis of the 21 most relevant works, a technology-independent Best Practices Guide was developed. This guide consolidates governance guidelines, recommending the implementation of fine-grained access controls, security middleware, and secure disposal techniques such as crypto-shredding, to support compliance of these environments with Brazilian legislation.

**Keywords:** data security, NoSQL databases, Big Data, systematic mapping, cloud computing.

# Seguridad en Bancos NoSQL para Big Data en la Nube: Mapeo Sistemático y Directrices de Cumplimiento con la Ley de Protección de Datos Personales (LGPD)

**Resumen.** La convergencia entre Big Data, bases de datos NoSQL y la computación en la nube introduce vulnerabilidades significativas de seguridad que entran en conflicto con las exigencias de la Ley General de Protección de Datos (LGPD). Este trabajo investiga los principales desafíos de seguridad en bases de datos NoSQL utilizadas en el contexto de Big Data en entornos de nube mediante un Mapeo Sistemático de la Literatura (MS). La estrategia de búsqueda, aplicada en las bases ACM Digital Library y Scopus, resultó en la selección y análisis de 159 estudios primarios. Los resultados indican que la ausencia de

cifrado nativo, fallas de autenticación y vulnerabilidades de inyección son las deficiencias más recurrentes, con una predominancia de investigaciones centradas en MongoDB y una marcada carencia en relación con mecanismos de auditoría y registros. Como contribución principal, y basándose en la síntesis cualitativa de los 21 trabajos más relevantes, se elaboró una Guía de Buenas Prácticas independiente de la tecnología. Esta propuesta consolida directrices de gobernanza, sugiriendo la implementación de controles de acceso de granularidad fina, middleware de seguridad y técnicas de eliminación segura (Crypto-shredding), con el objetivo de orientar la adecuación de estos entornos a la legislación brasileña.

**Palabras clave:** seguridad de datos, bases de datos NoSQL, Big Data, mapeo sistemático, computación en la nube.

# 1 Introducción

El fenómeno del Big Data, comúnmente definido por un conjunto de características conocidas como las 5 "V" (Volumen, Velocidad, Variedad, Veracidad y Valor), representa una nueva generación de tecnologías y arquitecturas diseñadas para extraer valor económico de conjuntos de datos masivos, complejos y, muchas veces, no estructurados (Ishwarappa y Anuradha, 2015). Este paradigma digital, alimentado por diversas fuentes como redes sociales, transacciones en línea y dispositivos conectados al Internet de las Cosas (IoT), impone un nuevo desafío para el sector tecnológico: la gestión y el procesamiento eficiente de datos a gran escala (Khan et al., 2019).

Frente a este escenario, dos tecnologías se han vuelto esenciales para este ecosistema. La computación en la nube ofrece escalabilidad bajo demanda y un modelo de negocio basado en el pago por uso, eliminando la necesidad de realizar grandes inversiones en infraestructura propia (Surbiryala y Rong, 2019). Por su parte, las bases de datos NoSQL proporcionan flexibilidad de esquema y rendimiento optimizado en entornos distribuidos, siendo consideradas una respuesta directa a las limitaciones de los SGBD relacionales tradicionales ante las demandas del Big Data (Oussous et al., 2017). La sinergia entre estas tecnologías dio origen al modelo Database as a Service (DBaaS), en el cual un proveedor de nube ofrece una base de datos flexible y escalable bajo demanda (Abourezq y Idrissi, 2016).

No obstante, la convergencia entre Big Data, bases de datos NoSQL y computación en la nube introduce nuevos tipos de vulnerabilidades. La relativa inmadurez de las bases NoSQL, diseñadas con enfoque en rendimiento y escalabilidad, hace que a menudo descuiden mecanismos de seguridad que son estándar en los SGBD relacionales (Fahd et al., 2019). Esta inmadurez se traduce en fallas críticas, como la prevalencia de configuraciones predeterminadas inseguras: muchas implementaciones, como MongoDB, vienen con la autenticación deshabilitada por default, permitiendo acceso irrestricto si están expuestas a Internet (Ferrari et al., 2020; Rubira, 2025). La ausencia de cifrado nativo y la flexibilidad de formatos de datos como JSON abren la puerta a nuevas formas de ataque, como la inyección específica para NoSQL (Awad et al., 2024; Ron et al., 2015).

Este escenario adquiere mayor relevancia debido a las exigencias legales establecidas por la Ley General de Protección de Datos (LGPD) de Brasil, Ley n.º 13.709/2018 (Brasil, 2018), que impone requisitos estrictos de seguridad y gobernanza

para el tratamiento de datos personales. La importancia del tema fue reforzada por la Enmienda Constitucional n.º 115/2022, que elevó la protección de datos personales a la categoría de derecho fundamental (Rank y Berberi, 2022). Por lo tanto, la falta de una comprensión sistemática de los riesgos combinados de este ecosistema tecnológico y la ausencia de un conjunto consolidado de buenas prácticas de seguridad representan una brecha crítica.

En este sentido, el presente trabajo tiene como objetivo general analizar los desafíos de seguridad que enfrentan las bases de datos NoSQL en el contexto de Big Data en entornos de nube. Para lograr este propósito, se definieron los siguientes objetivos específicos: (a) realizar un mapeo sistemático sobre seguridad en bases de datos NoSQL, Big Data y computación en la nube; (b) identificar los principales riesgos, desafíos y vulnerabilidades; (c) evaluar las principales formas de mitigación propuestas en la literatura; (d) identificar el estado del arte en el área; y (e) elaborar una guía de buenas prácticas de seguridad orientada al cumplimiento de la LGPD.

En cuanto a los trabajos relacionados, Raposo et al. (2019) y Raposo (2024) realizaron revisiones sobre la LGPD y la computación en la nube, respectivamente, sin abordar los desafíos específicos del paradigma NoSQL. Belarmino (2024) analizó soluciones tecnológicas para la LGPD desde una perspectiva generalista, sin considerar las problemáticas de Big Data. No se identificó ningún mapeo sistemático que integre simultáneamente los dominios de bases de datos NoSQL, Big Data, computación en la nube y LGPD, brecha que el presente trabajo busca cubrir.

# 2 Marco Teórico

## 2.1 Big Data y Computación en la Nube

El término Big Data se refiere a conjuntos de datos cuyo tamaño y complejidad superan la capacidad de los sistemas tradicionales para capturarlos, almacenarlos y analizarlos, y se caracteriza por las "5 V": Volumen, Velocidad, Variedad, Veracidad y Valor (Khan et al., 2019). La computación en la nube permite el acceso bajo demanda a recursos computacionales configurables con un esfuerzo mínimo de gestión (Magoulès et al., 2013). La sinergia entre ambos paradigmas es especialmente relevante: la nube proporciona la infraestructura necesaria para manejar grandes volúmenes de datos con escalabilidad elástica y reducción de costos (Surbiryala y Rong, 2019).

## 2.2 Bases de Datos NoSQL

Las bases de datos NoSQL surgieron como respuesta a las limitaciones de los SGBD relacionales ante las demandas de aplicaciones web a gran escala. Diseñadas para operar en clústeres con escalabilidad horizontal, adoptan el modelo de consistencia eventual (BASE) en lugar de la consistencia estricta (ACID), priorizando la disponibilidad y la tolerancia a particiones según el Teorema CAP (Zaki, 2014). Se clasifican en cuatro modelos principales: clave-valor (Redis), orientadas a documentos (MongoDB), orientadas a columnas (Cassandra, HBase) y orientadas a grafos (Neo4j). Su diferencia más relevante respecto al modelo relacional es la adopción de un esquema dinámico (schema-on-read), especialmente ventajoso en escenarios de Big Data con datos semiestructurados o no estructurados (Oussous et al., 2017).

### 2.3 Seguridad y Privacidad de los Datos

La seguridad de la información se basa en la tríada CID: Confidencialidad, Integridad y Disponibilidad. En entornos NoSQL, las vulnerabilidades más frecuentes son la configuración incorrecta, ya que muchos sistemas, como MongoDB, se distribuyen con la autenticación deshabilitada por defecto (Ferrari et al., 2020), y la inyección de código específica para NoSQL, que aprovecha formatos como JSON para manipular consultas (Fahd et al., 2019; Ron et al., 2015). Para mitigarlas, la criptografía es fundamental tanto en tránsito (SSL/TLS) como en reposo. En cuanto al control de acceso, modelos como RBAC son ampliamente utilizados, aunque diversos autores destacan la necesidad de enfoques más granulares como ABAC, que evalúan el contenido de los datos y el contexto del usuario (Mohamed et al., 2024). La implementación de middleware de seguridad en la capa de aplicación surge como alternativa para compensar las limitaciones nativas de los sistemas NoSQL.

### 2.4 El Contexto Legal: LGPD

La Ley General de Protección de Datos (LGPD), Ley n.º 13.709/2018, se inspira en el Reglamento General de Protección de Datos (GDPR) de la Unión Europea y tiene como objetivo regular el tratamiento de datos personales en Brasil. La normativa establece principios, derechos y obligaciones, exigiendo que el tratamiento de datos se realice con finalidades legítimas y mediante la adopción de medidas de seguridad adecuadas (Fernandes et al., 2021). El cumplimiento resulta especialmente complejo en entornos de Big Data en la nube, ya que el principio de minimización de datos entra en conflicto con la práctica de acumular grandes volúmenes de información para análisis futuros (Santos et al., 2023). En este contexto, la intersección entre la LGPD y el Big Data constituye un campo de tensión significativo, que exige una reestructuración profunda de la gobernanza de datos a gran escala (Rank y Berberi, 2022).

## 3 Metodología

La investigación empleó la metodología de Mapeo Sistemático de la Literatura (MS), siguiendo las directrices de Felizardo et al. (2017), desarrollada en las fases de planificación, ejecución y presentación de resultados. La estructuración del estudio se realizó conforme a las directrices de la declaración PRISMA 2020 (Page et al., 2022).

### 3.1 Preguntas de Investigación (PI)

Se formularon cuatro Preguntas de Investigación: PI1 ¿Cuáles son las principales vulnerabilidades y riesgos de seguridad discutidos en la literatura para bases de datos NoSQL en entornos de nube y Big Data? PI2 ¿Qué tipos de soluciones y buenas prácticas se proponen para mitigar las vulnerabilidades identificadas? PI3 ¿Qué herramientas, técnicas o criterios son utilizados por los principales sistemas de gestión de bases de datos y proveedores de nube para garantizar la seguridad? PI4 ¿Cuál es el perfil de la investigación en el área, incluyendo la distribución temporal, los tipos de estudio y los canales de publicación?

### 3.2 Estrategia de Búsqueda

Los términos clave se dividieron en dos grupos: población/contexto (NoSQL, Cloud, Big Data) e intervención/enfoque (Security, Privacy, Encrypt*, Access Control, LGPD, GDPR). La cadena de búsqueda final fue: ("NoSQL" OR "Not only SQL") AND ("Cloud Computing" OR "Cloud" OR "Big Data" OR "Distributed Environment*") AND (Security OR Privacy OR Vulnerab* OR "Data Protection" OR Encrypt* OR Crypt* OR "Access Control" OR LGPD OR GDPR). La búsqueda se realizó exclusivamente con términos en inglés, ya que en las bases seleccionadas los títulos y resúmenes están obligatoriamente indexados en este idioma.

Se seleccionaron como fuentes de datos ACM Digital Library y Scopus. La elección de ACM se justifica por ser uno de los principales recursos para revisiones sistemáticas en el área de computación. Scopus se incorporó por ofrecer una amplia cobertura de trabajos de alta calidad en tecnología y ciencias de la computación (Burnham, 2006).

### 3.3 Criterios de Seleçción

Los criterios de inclusión establecieron que los estudios debían: (CI1) abordar directamente la seguridad en bases de datos NoSQL; (CI2) situarse en el contexto de Big Data o computación en la nube; (CI3) ser artículos completos; y (CI4) presentar una contribución explícita relacionada con la seguridad. Por su parte, los criterios de exclusión eliminaron: (CE1) estudios no escritos en inglés; (CE2) duplicados; (CE3) literatura gris; y (CE4) trabajos sin acceso al texto completo.

### 3.4 Conducción y Extracción de Datos

Las búsquedas se realizaron en septiembre de 2025, lo que resultó en un total de 350 artículos iniciales. La clasificación se efectuó mediante el software Zotero, utilizando un sistema de etiquetas para categorizar cada estudio según las preguntas de investigación. Para la PI1 (Vulnerabilidades) y la PI2 (Soluciones), se adoptó un enfoque inductivo, en el que las categorías surgieron a partir de la lectura de títulos y resúmenes. El análisis cuantitativo se automatizó mediante scripts en Python, empleando las bibliotecas Matplotlib y Seaborn. Además, se realizó una síntesis cualitativa de 21 artículos seleccionados por presentar arquitecturas de seguridad replicables, validaciones experimentales o propuestas de adecuación regulatoria, los cuales sirvieron como base para la elaboración de la Guía de Buenas Prácticas.

## 4 Resultados

### 4.1 Selección de los Estudios

La etapa de identificación incluyó un total de 350 publicaciones (328 de Scopus y 22 de ACM Digital Library). La eliminación de duplicados excluyó 16 registros, y 8 estudios fueron descartados por idioma, quedando 264 artículos para la fase de cribado. En la etapa de elegibilidad se aplicaron los criterios de inclusión y exclusión, clasificándose una proporción significativa (105 artículos) como "Fuera de Alcance",

ya que, aunque contenían las palabras clave, abordaban la seguridad en la nube de manera general sin centrarse en bases de datos, o trataban el rendimiento de bases NoSQL sin considerar aspectos de seguridad. El proceso completo de selección, conforme a las directrices PRISMA (Page et al., 2022), se muestra en la Figura 1. Al finalizar el filtrado, se seleccionaron 159 estudios para la extracción de datos y la síntesis cualitativa.

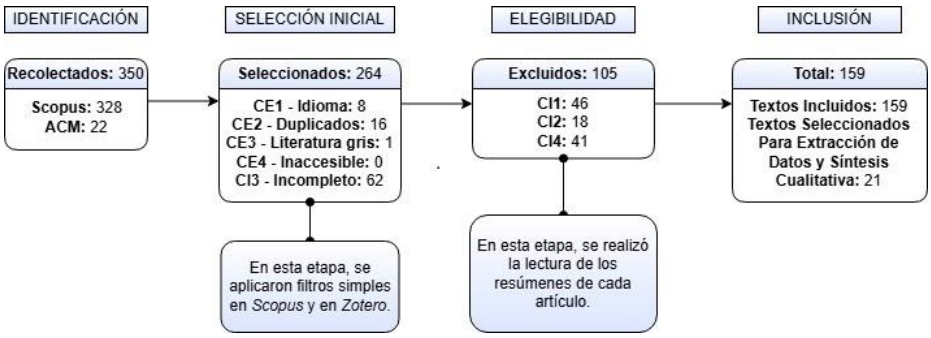


**Fig. 1.** Proceso de selección de los estudios.

### 4.2 Principales Vulnerabilidades y Riesgos (PI1)

La categorización de los 159 estudios permitió identificar las fallas de seguridad más frecuentes. La "Criptografía" (por su ausencia o implementación deficiente) es el aspecto más citado, presente en 71 estudios. También destacan los estudios de naturaleza "Genérica/Review" (48 artículos), vulnerabilidades de "Autenticación" (26 estudios), "Inyección NoSQL" (13 estudios), "Configuración Inadecuada" (13 estudios), "Auditoría/Logs" (9 estudios), "Amenazas Internas" (6 estudios), "Privacidad de Datos Sensibles" (5 estudios) y "Denegación de Servicio" (3 estudios).

Esta predominancia de la criptografía como vulnerabilidad no debe interpretarse solo como una falla de desarrollo, sino como un trade-off arquitectónico consciente. Para cumplir con los requisitos de alta disponibilidad y escalabilidad establecidos por el Teorema CAP, muchas soluciones NoSQL han sacrificado históricamente garantías de seguridad en favor del rendimiento (Goel y Hofstede, 2021; Zahid et al., 2014).

### 4.3 Mecanismos de Protección y Soluciones (PI2)

La criptografía encabeza las propuestas de mitigación, siendo el tema principal en 62 estudios. Las soluciones orientadas a la privacidad (anonimización, privacidad diferencial) aparecen en 44 estudios, seguidas por el control de acceso (42 estudios), middleware/proxy (23 estudios), inteligencia artificial/machine learning (10 estudios), auditoría/logs (9 estudios), sistemas IDS/IPS (7 estudios) y blockchain (4 estudios).

Es interesante observar que el número de estudios que señalan la criptografía como solución supera al de aquellos que la identifican como vulnerabilidad cuando está ausente, lo que sugiere que la comunidad científica la considera un mecanismo principal para garantizar la confidencialidad en entornos de nube. Otra tendencia relevante es la adopción de middleware o proxy como mecanismo de protección (23

estudios), una estrategia que permite incorporar capas de seguridad a bases de datos legadas o inmaduras sin modificar el código fuente del SGBD. A pesar de esta diversidad de enfoques preventivos, se identifica una brecha significativa en la literatura respecto a los mecanismos de detección y respuesta a incidentes: soluciones fundamentales como auditoría y logs (9 estudios) y sistemas de detección de intrusiones (7 estudios) siguen siendo escasamente exploradas. Esta disparidad resulta preocupante, dado que el cumplimiento de la LGPD también exige la capacidad de demostrar que los datos no han sido accedidos de manera indebida.

### 4.4 Tecnologías y Bases de Datos Investigadas (PI3)

La categoría "Genérico" (independiente de la tecnología) es predominante, abarcando 101 artículos. Entre las tecnologías específicas, MongoDB es la más estudiada (38 artículos), seguida de Cassandra (16), HBase (6), Redis (5) y otros sistemas como DynamoDB, Neo4j y Couchbase (6 en la categoría "Otros"). La relevancia académica de MongoDB refleja tanto su posicionamiento en el mercado como el historial de incidentes de seguridad asociados a sus configuraciones por defecto (Rathore y Bagui, 2024; Singh et al., 2024).

Esta concentración evidencia una brecha significativa respecto a modelos como los de clave-valor (Redis) y los orientados a grafos (Neo4j). Dado que las arquitecturas de Big Data suelen integrar distintos tipos de bases NoSQL en una misma aplicación, esta carencia representa una oportunidad relevante para futuras investigaciones.

### 4.5 Panorama de la Investigación y Análisis Cruzado (PI4)

El análisis temporal muestra una baja producción entre 2013 y 2015, seguida de un crecimiento a partir de 2016, con un pico en el período 2017–2018 (23 y 21 estudios, respectivamente), que coincide con la implementación del GDPR y la creación de la LGPD. La mayoría de los trabajos (63%) se publicó en revistas científicas. El cruce con las tecnologías analizadas (PI3) revela una predominancia de estudios de carácter genérico y una concentración de propuestas basadas en inteligencia artificial y blockchain dentro de esta categoría. MongoDB y Cassandra presentan un enfoque centrado en criptografía y control de acceso, lo que configura un perfil predominantemente correctivo. Las intersecciones vacías evidencian importantes brechas en la literatura, especialmente en la ausencia de mecanismos de auditoría y detección de intrusiones en tecnologías como HBase, Redis y Neo4j.

## 5 Discusión

### 5.1 Amenazas a la Validez

La validez de constructo se garantizó mediante la construcción iterativa de la cadena de búsqueda, calibrada a partir de búsquedas piloto. Dado que el estudio fue realizado por un único investigador, existe riesgo de subjetividad; sin embargo, este se mitigó mediante un protocolo con criterios objetivos y la automatización del proceso de extracción con Python. La validez externa está limitada por la restricción a dos bases de datos (ACM y Scopus), decisión adoptada conscientemente para viabilizar la

ejecución del estudio. En cuanto a la validez de conclusión, se emplearon categorías estandarizadas de OWASP y, cuando fue necesario, la técnica de keywording (Petersen et al., 2015; Zhou et al., 2016).

### 5.2 Síntesis de los Hallazgos

El análisis de los datos revela que la criptografía y el control de acceso se han consolidado como los mecanismos de protección más investigados por la comunidad científica. La recurrencia de vulnerabilidades asociadas a la ausencia de criptografía refleja un trade-off arquitectónico consciente impuesto por el Teorema CAP, donde la priorización de la disponibilidad ha implicado históricamente el sacrificio de controles de seguridad nativos. La adopción de middleware/proxy resulta prometedora, ya que permite incorporar capas de seguridad en bases de datos legadas sin modificar el código fuente del SGBD, aunque tecnologías emergentes como la Inteligencia Artificial y Blockchain permanecen mayoritariamente en etapa teórica, sin validaciones en SGBDs específicos. En contraste, los mecanismos de auditoría, detección de intrusiones y la seguridad en bases orientadas a grafos y clave-valor constituyen las principales brechas identificadas, con implicaciones directas sobre la capacidad de respuesta ante incidentes y el análisis forense exigidos por la LGPD.

## 6 Guía de Buenas Prácticas para el Cumplimiento de la LGPD

A partir de la síntesis cualitativa de los 21 estudios primarios seleccionados, esta sección consolida los resultados en un conjunto de directrices prácticas, cuyo mapeo a los requisitos normativos de la LGPD se resume en la Tabla 1. Uno de los principales objetivos al elaborar esta guía fue garantizar la independencia tecnológica, es decir, asegurar que las buenas prácticas propuestas sean aplicables a cualquier SGBD NoSQL. La estructuración se basa en la premisa de que la seguridad en bases de datos NoSQL en entornos de nube no puede depender exclusivamente de los mecanismos de seguridad nativos de los SGBDs, ya que la literatura evidencia que estos, en muchos casos, resultan insuficientes para cumplir con las exigencias normativas.

### 6.1 Gobernanza y Preparación de los Datos

En consonancia con el Art. 50 de la LGPD, que promueve la adopción de programas de gobernanza y buenas prácticas, se recomienda la clasificación automatizada de la sensibilidad de los datos durante su recolección, separándolos en categorías: Públicos, Internos, Confidenciales y Sensibles. El enfoque propuesto por Heni y Gargouri (2015) sugiere un flujo de homogeneización seguido de la detección de sensibilidad, con el objetivo de aislar atributos identificadores. Además, el uso de lenguajes de políticas como XACML permite que el acceso sea concedido o denegado dinámicamente en función de metadatos (Ghazi et al., 2016).

### 6.2 Control de Acceso y Gestión de Identidades

El control de acceso en entornos NoSQL distribuidos requiere una granularidad mayor que la ofrecida por los modelos basados únicamente en roles (RBAC), conforme al Art.

46 de la LGPD, que exige la adopción de medidas adecuadas para proteger los datos personales frente a accesos no autorizados. El sistema debe restringir el acceso a nivel de campo o atributo específico dentro de un documento JSON. Huang et al. (2019) demuestran que el uso de mecanismos de Control de Acceso Basado en Atributos (ABAC), integrados en el núcleo de la base de datos, permite validar dinámicamente el contexto de la solicitud antes de conceder el acceso a los datos.

**Tabla 1.** Mapeo de Requisitos de la LGPD e Soluciones Técnicas.

| Requisito LGPD | Medida Técnica | Referencia |
|---|---|---|
| Gobernanza (Art. 50) | Clasificación y Etiquetado | Heni y Gargouri (2015); Ghazi et al. (2016) |
| Seguridad (Art. 46) | ABAC, MFA, Contexto | Huang et al. (2019); Sicari et al. (2022) |
| Privacy by Design (Art. 46) | Middleware, KMS, Multi-Cloud | Rafique et al. (2021); Ahmadian et al. (2017) |
| Minimización (Art. 6°) | Priv. Diferencial, Enmascaramiento | Ye et al. (2023); Cuzzocrea y Shahriar (2017) |
| Prestación de Cuentas (Art. 6°) | Auditoria, ML | Hauger y Olivier (2018); Praveen et al. (2022) |
| Eliminación (Art. 18) | Crypto-shredding | Yoon y Lee (2018) |
| Prevención (Arts. 6° y 46) | Multi-Cloud, Secret Sharing | Pontes et al. (2019); Kohler et al. (2019) |

### 6.3 Arquitectura de Protección y Privacidad

La arquitectura debe considerar que la infraestructura de nube es un entorno no confiable, implementando las medidas técnicas de protección exigidas por los artículos 46 y 47 de la LGPD. La criptografía nativa de las bases de datos generalmente solo protege contra el robo físico del disco. Para garantizar la privacidad frente al proveedor de nube, los datos deben cifrarse en la capa de aplicación o mediante un middleware de seguridad antes de ser enviados a la base de datos, conforme a la arquitectura propuesta por Rafique et al. (2021).

Las claves criptográficas no deben almacenarse en el mismo servidor que los datos cifrados. Se recomienda utilizar un servicio externo de gestión de claves (KMS) o módulos de seguridad hardware (HSM), asegurando que el administrador de la base de datos no tenga la capacidad técnica para descifrar la información (Ahmadian et al., 2017). Para datos de alta criticidad, la fragmentación y distribución de secretos entre múltiples nubes garantiza que una eventual filtración en un proveedor exponga solo fragmentos ininteligibles de información, lo que incrementa la resiliencia y la confidencialidad del sistema (Pontes et al., 2019; Kohler et al., 2019; Santos y Masala, 2019).

### 6.4 Procesamiento Seguro y Minimización

La utilidad de los datos para el análisis no debe comprometer la privacidad de los titulares, en cumplimiento del principio de necesidad previsto en el Art. 6°, inciso III de la LGPD, que limita el tratamiento al mínimo indispensable. Para permitir que la base de datos realice operaciones de búsqueda y ordenamiento sin exponer los datos en texto plano, pueden emplearse esquemas de cifrado determinístico (DET) o cifrado que preserva el orden (OPE) en columnas indexadas, según lo detallado por Waage y Wiese (2018).

En escenarios de Big Data Analytics donde el objetivo es extraer tendencias estadísticas y no información individual, la aplicación de técnicas de Privacidad Diferencial Local permite recolectar estadísticas en bases de datos clave-valor garantizando matemáticamente el anonimato de los individuos (Ye et al., 2023). Además, los datos reales nunca deben utilizarse en entornos de desarrollo o prueba; en su lugar, deben aplicarse técnicas de enmascaramiento irreversible que preserven la integridad referencial sin exponer información real (Cuzzocrea y Shahriar, 2017).

### 6.5 Auditoría, Análisis Forense y Defensa Activa

La capacidad de respuesta ante incidentes es un requisito obligatorio para la notificación de violaciones, conforme a lo establecido en el Art. 48 de la LGPD. Los registros estándar (logs) de los sistemas NoSQL suelen limitarse al registro de errores u operaciones del sistema, lo que dificulta la atribución forense (Hauger y Olivier, 2018). Se recomienda configurar mecanismos de auditoría con un nivel de detalle granular, validar rigurosamente las entradas mediante el uso de statements parametrizados como medida de defensa frente a inyecciones NoSQL (Sachdeva y Gupta, 2018), así como utilizar modelos de machine learning supervisado para la detección de patrones complejos de inyección (Praveen et al., 2022).

### 6.6 Gestión del Ciclo de Vida y Eliminación Segura

Para garantizar el derecho a la eliminación previsto en el Art. 18 de la LGPD, en sistemas distribuidos donde asegurar la sobrescritura física de todos los fragmentos es técnicamente inviable, se recomienda aplicar la técnica de crypto-shredding, que consiste en destruir la clave criptográfica asociada a los datos. Yoon y Lee (2018) demostraron experimentalmente que las operaciones de eliminación estándar en bases de datos NoSQL suelen dejar artefactos recuperables en el disco, lo que convierte la destrucción de claves en una de las pocas estrategias capaces de garantizar la irreversibilidad exigida por la normativa.

## 7 Conclusiones

El presente estudio logró su objetivo general de analizar los desafíos de seguridad que enfrentan las bases de datos NoSQL en el contexto de Big Data en la nube. El Mapeo Sistemático de 159 estudios permitió organizar el conocimiento disperso sobre el tema, revelando que la criptografía y el control de acceso son los mecanismos más investigados. Sin embargo, se observó que la literatura sobre auditoría y logs sigue

siendo muy limitada, lo que afecta tanto el análisis forense como la respuesta ante incidentes. La mayoría de las investigaciones se enfocó en MongoDB o en soluciones genéricas, lo que ha provocado una escasez de estudios sobre bases de datos clave-valor (Redis), orientadas a columnas (HBase) y orientadas a grafos (Neo4j).

Otra conclusión relevante se refiere a la naturaleza arquitectónica de las vulnerabilidades identificadas. La ausencia de mecanismos de seguridad robustos habilitados por defecto en la mayoría de los sistemas NoSQL no es simplemente un descuido, sino un compromiso consciente impuesto por el Teorema CAP, que históricamente ha priorizado el rendimiento sobre la seguridad.

La Guía de Buenas Prácticas propuesta demuestra que la seguridad en entornos NoSQL en la nube no puede depender únicamente de las funcionalidades nativas de los SGBD, sino que requiere la incorporación de capas externas de gobernanza y middleware. Las limitaciones de esta investigación se centran en la restricción de las fuentes de búsqueda a dos bases de datos y en la selección subjetiva de los 21 artículos utilizados como base para la guía..

## Referencias


Abourezq, M. y Idrissi, A. (2016). Database-as-a-service for big data: An overview. *International Journal of Advanced Computer Science and Applications*, *7*(1), 157-177.

Ahmadian, M., Pliatsios, D., Shulman, A. y Pechenizkiy, M. (2017). SecureNoSQL: An approach for secure search of encrypted NoSQL databases in the public cloud. *International Journal of Information Management*, *37*(2), 63–74.

Awad, K. U., Cardoso, L. S. y Bussador, A. (2024). NoSQL e segurança: Um estudo de análise para prevenção de injeção em bancos de dados NoSQL. *Actas del XXI Congresso Latino-Americano de Software Livre e Tecnologias Abertas (Latinoware 2024)*, p. 464–467.

Belarmino, G. de S. (2024). Revisão Sistemática de Soluções Tecnológicas e Organizacionais para Conformidade com a LGPD [Tesis de Maestria, Universidade Federal da Paraíba]. https://repositorio.ufpb.br/jspui/handle/123456789/33072

Brasil. (2018). Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD).

Burnham, J.F. (2006). Scopus database: a review. *Biomedical Digital Libraries*, *3*(1). https://doi.org/10.1186/1742-5581-3-1

Colombo, P. y Ferrari, E. F. (2019). Access control technologies for Big Data management systems. *Cybersecurity*, *2*(1).

Cuzzocrea, A. y Shahriar, H. (2017). Data masking techniques for NoSQL database security: A systematic review. IEEE International Conference on Big Data.

Fahd, K., Venkatraman, S. y Hammeed, F. K. (2019). A comparative study of NoSQL system vulnerabilities with big data. *International Journal of Managing Information Technology*, *11*(4), 1–19.

Felizardo, K. R., Nakagawa, E. Y., Fabbri, S. C. P. F., y Ferrari, F. C. (2017). Revisão Sistemática da Literatura em Engenharia de Software – Teoria e Prática. Elsevier.

Fernandes, M. A. de S., Canedo, E. D., Costa, P. H. T. y Masson, E. T. S. (2021). Impactos da lei de proteção de dados (LGPD) brasileira no uso da computação em nuvem. *Revista Ibérica de Sistemas e Tecnologias de Informação*, (E42), 374–385.

Ferrari, D., Carminati, B., Ferrari, E. y Tran, N. H. (2020). NoSQL breakdown: A large-scale analysis of misconfigured NoSQL services. Annual Computer Security Applications Conference, 567–581.

Ghazi, Y., Masood, R., Rauf, A. y Shibli, M. A. (2016). DB-SECaaS: A cloud-based protection system for document-oriented NoSQL databases. *EURASIP Journal on Information Security*.
Goel, K. y Hofstede, A. H. T. (2021). Privacy-breaching patterns in NoSQL databases. *IEEE Access*, 9, 35229–35239.
Hauger, W. K. y Olivier, M. S. (2018). NoSQL databases: Forensic attribution implications. IFIP Advances in Information and Communication Technology.
Heni, H. y Gargouri, F. (2015). A methodological approach for big data security: Application for NoSQL data stores. East European Conference on Advances in Databases and Information Systems.
Huang, L., Zhu, Q. y Wen, Y. (2019). An attribute-based fine-grained access control mechanism for HBase. Computers & Security.
Ishwarappa e Anuradha, J. (2015). A brief introduction on big data 5Vs characteristics and Hadoop technology. Procedia Computer Science, 48, 319–324.
Khan, N., Alsaqer, M., Shah, H., Badsha, G., Abbasi, A. A. y Salehian, S. (2019). The 51 v's of big data. International Conference on Omni-Layer Intelligent Systems, 19–24.
Kohler, J., Jünemann, K., Bristow, R. y Bayer, N. (2019). A security-by-distribution approach to manage big data in a federation of untrustworthy clouds. IEEE Consumer Communications & Networking Conference.
Magoulès, F., Pan, J. y Teng, F. (2013). Cloud Computing: Data-Intensive Computing and Scheduling. CRC Press.
Mohamed, A. K. Y. S., Ahmed, T., Benyahia, A. y Zulkernine, M. (2024). A systematic literature review of authorization and access control for different database models. *International Journal of Web Information Systems*, 20(1), 1–23.
Oussous, A., Benjelloun, F., Lahcen, A. A. y Belfkih, S. (2017). NoSQL databases for big data. *International Journal of Big Data Intelligence*, *4*(3), 171.
Page, M. J. et al. (2022). A declaração PRISMA 2020. *Revista Panamericana de Salud Pública*, 46, e112.
Petersen, K., Vakkalanka, S. y Kuzniarz, L. (2015). Guidelines for conducting systematic mapping studies in SE: An update. Information and Software Technology, 64, 1–18.
Pontes, R., Maia, F., Paulo, J., Villari, M. y Oliveira, R. (2019). D'Artagnan: A trusted NoSQL database on untrusted clouds. ACM/IFIP International Middleware Conference.
Praveen, S. P., Dcouth, A. y Mahesh, A. S. (2022). NoSQL injection detection using supervised text classification. IEEE International Conference on Electronics, Computing and Communication Technologies.
Rafique, A., Van Landuyt, D., Reniers, V., Lagaisse, B. y Joosen, W. (2021). CryptDICE: Distributed data protection system. Information Systems.
Rank, A. T. y Berberi, M. A. L. (2022). Big data e direitos fundamentais sob o enfoque da LGPD. *International Journal of Digital Law*, *3*(2), 9–26.
Raposo, C. F. L. (2024). Desafios e conformidade de cloud computing com a LGPD: Uma revisão sistemática. M Impacto Científico.
Raposo, C. F. L., Gomes, R. A. B. L., Lima, L. R. y Lima, J. C. L. (2019). LGPD: Revisão sistemática. *RACE - Revista de Administração do Cesmac*, 4.
Rathore, M. y Bagui, S. S. (2024). MongoDB: Meeting the dynamic needs of modern applications. Encyclopedia, *4*(4), 1433–1453.
Ron, A., Shulman-Peleg, A. y Bronshtein, E. (2015). No SQL, no injection? Examining NoSQL security. arXiv:1506.04082.
Rubira, L. A. (2025). Bancos de dados NoSQL: Vulnerabilidades de segurança encontrados no MongoDB [TCC]. UTFPR.
Sachdeva, V. y Gupta, S. K. (2018). Basic NoSQL injection analysis and detection on MongoDB. IEEE International Conference on Advances in Computing, Communication and Automation.

Santos, N. y Masala, G. L. (2019). Big data security on cloud servers using data fragmentation technique and NoSQL database. KES International Symposium on Intelligent Interactive Multimedia Systems, 5–13.
Santos, R. A., Scandolara, D. H. y Mohr, E. T. B. (2023). Desafios da LGPD na governança de dados pessoais em cenários de Big Data. 2ª Mostra Científica de Proteção de Dados.
Sicari, S., Rizzardi, A. y CoenPorisini, A. (2022). Security & privacy issues and challenges in NoSQL databases. Computer Networks.
Singh, P. K., Sharma, S., Chauhan, S. y Kumar, A. (2024). An analysis and overview of MongoDB security. SSRN.
Surbiryala, J. y Rong, C. (2019). Cloud computing: History and overview. IEEE Cloud Summit, 1–7.
Waage, T. y Wiese, L. (2018). CloudDBGuard: Sorting and searching on encrypted data in NoSQL cloud databases. Future Generation Computer Systems.
Ye, Q. et al. (2023). PrivKVM*: Key-value statistics with local differential privacy. IEEE TDSC.
Yoon, J. y Lee, S. (2018). A method and tool to recover data deleted from a MongoDB. Digital Investigation.
Zahid, A., Masood, R. y Shibli, M. A. (2014). Security of sharded NoSQL databases: A comparative analysis. Conference on Information Assurance and Cyber Security, 1–8.
Zaki, A. K. (2014). NoSQL databases: New millennium database for big data. *International Journal of Research in Engineering and Technology*, *3*(3), 403–409.
Zhou, X., Jin, Y., Zhang, H., Li, S. y Huang, X. (2016). A map of threats to validity of systematic literature reviews in SE. Asia-Pacific Software Engineering Conference, 153–160.