



UNIVERSIDAD POLITÉCNICA SALESIANA

SEDE QUITO

CARRERA DE COMPUTACIÓN

**ESTADO DEL ARTE ACERCA DE PATRONES DE ALMACENAMIENTO
DE BIG DATA EN EL ÁREA DE SALUD**

Trabajo de titulación previo a la obtención del
Título de Ingeniero en Ciencias de la Computación

**AUTORES: JONATHAN ANDRES BARRIONUEVO PORTERO
JOSUE VLADIMIR MORALES RIVERA
TUTOR: GUSTAVO ERNESTO NAVAS RUILOVA**

Quito - Ecuador
2023

CERTIFICADO DE RESPONSABILIDAD Y AUTORÍA DEL TRABAJO DE TITULACIÓN

Nosotros, Jonathan Andrés Barrionuevo Portero con documento de identificación N.º 1726020322 y Josue Vladimir Morales Rivera con documento de identificación N.º 1752702074; manifestamos que:

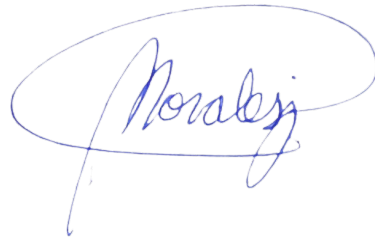
Somos los autores y responsables del presente trabajo; y, autorizamos a que sin fines de lucro la Universidad Politécnica Salesiana pueda usar, difundir, reproducir o publicar de manera total o parcial el presente trabajo de titulación.

Quito, 14 de agosto de 2023

Atentamente,



Jonathan Andrés Barrionuevo Portero
1726020322



Josue Vladimir Morales Rivera
1752702074

CERTIFICADO DE CESIÓN DE DERECHOS DE AUTOR DEL TRABAJO DE TITULACIÓN A LA UNIVERSIDAD POLITÉCNICA SALESIANA

Nosotros, Jonathan Barrionuevo con documento de identificación No. 1726020322 y Josue Vladimir Morales Rivera con documento de identificación No. 1752702074, expresamos nuestra voluntad y por medio del presente documento cedemos a la Universidad Politécnica Salesiana la titularidad sobre los derechos patrimoniales en virtud de que somos autores del Artículo Académico: “Estado del Arte Acerca de Patrones de Almacenamiento de Big Data en el Área de Salud”, el cual ha sido desarrollado para optar por el título de: Ingeniero en Ciencias de la Computación, en la Universidad Politécnica Salesiana, quedando la Universidad facultada para ejercer plenamente los derechos cedidos anteriormente.

En concordancia con lo manifestado, suscribo este documento en el momento que hago la entrega del trabajo final en formato digital a la Biblioteca de la Universidad Politécnica Salesiana.

Quito, 14 de agosto de 2023

Atentamente,

Jonathan Andrés Barrionuevo Portero
1726020322

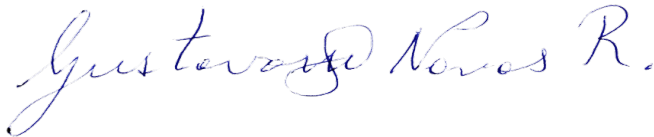
Josue Vladimir Morales Rivera
1752702074

CERTIFICADO DE DIRECCIÓN DEL TRABAJO DE TITULACIÓN

Yo, Gustavo Ernesto Navas Ruilova con documento de identificación N° 1705675625, docente de la Universidad Politécnica Salesiana, declaro que bajo mi tutoría fue desarrollado el trabajo de titulación: ESTADO DEL ARTE ACERCA DE PATRONES DE ALMACENAMIENTO DE BIG DATA EN EL ÁREA DE SALUD, realizado por Jonathan Andrés Barrionuevo Portero con documento de identificación No. 1726020322 y Josue Vladimir Morales Rivera con documento de identificación No. 1752702074, obteniendo como resultado final el trabajo de titulación bajo la opción Artículo Académico que cumple con todos los requisitos determinados por la Universidad Politécnica Salesiana.

Quito, 14 de agosto de 2023

Atentamente,



Ing. Gustavo Ernesto Navas Ruilova, MSc
1726834730

ESTADO DEL ARTE ACERCA DE PATRONES DE ALMACENAMIENTO DE BIG DATA EN EL ÁREA DE SALUD

1st Jonathan Andres Barrionuevo Portero 2nd Josue Vladimir Morales Rivera 3rd Gustavo Ernesto Navas Ruilova
Universidad Politécnica Salesiana Universidad Politécnica Salesiana Universidad Politécnica Salesiana
Quito-Ecuador Quito-Ecuador Quito-Ecuador
jbarrionuevop@est.ups.edu.ec jmoralesr1@est.ups.edu.ec gnavas@ups.edu.ec

Resumen—En el presente documento presentamos un estado de arte acerca de patrones de almacenamiento de Big Data en el área de Salud.

Para llevar a cabo este proceso, se tiene como objetivo identificar los principales patrones de almacenamiento utilizados en área de la salud, dando a conocer las tendencias de la industria, los desafíos y las principales preocupaciones respecto a como la implementación de estos patrones de almacenamiento pueden afectar a la información que se aloja, el acceso a estas y los retos que supone para su análisis y preservación.

La metodología SMS (Mapeo Sistemático de la Literatura) propuesta por Paternoster et al. [1] y Petersen et al [2] implementada en este artículo ayudó a recopilar un inicial de 530 artículos alojados en bases de datos académicas como IEEE, Web of Science y Scopus, los cuales fueron evaluados mediante un proceso de 7 etapas para finalmente obtener un final de 70 artículos, los cuales se presentan como los más relevantes para el presente trabajo.

Gracias a los resultados obtenidos, se puede concluir que los patrones de almacenamiento no estructurados y que pueden ser alojados en la nube son aquellos que se encuentran en tendencia ya que la industria de la salud dado al constante crecimiento de la información que se generan requieren soluciones que puedan escalar de forma flexible y ágil, asegurando la privacidad y la seguridad de la información almacenada en los diferentes servicios de la Nube.

Palabras Clave—Big Data, Patrones de almacenamiento, Privacidad de datos, Seguridad de datos, Salud, Arquitecturas

Abstract—In this paper we present a state of the art about Big Data storage patterns in the Healthcare area.

In order to carry out this process, we aim to identify the main storage patterns used in the healthcare area, providing industry trends, challenges and main concerns regarding how the implementation of these storage patterns may affect the information that is hosted, the access to it and the challenges it poses for its analysis and preservation.

The SMS (Systematic Literature Mapping) methodology proposed by Paternoster et al. and Petersen et al. implemented in this article helped to collect an initial 530 articles hosted in academic databases such as IEEE, Web of Science and Scopus, which were evaluated through a 7-stage process to finally obtain a final 70 articles, which are presented as the most relevant for the present work.

Thanks to the results obtained, it can be concluded that the unstructured storage patterns that can be hosted in the cloud are those that are in trend since the health industry given the constant growth of the information generated require solutions

that can scale in a flexible and agile way, ensuring the privacy and security of the information stored in the different cloud services.

Keywords—Big Data, Storage patterns, Data Privacy, Data Security, Health, Architectures

I. INTRODUCCIÓN

El crecimiento de la información generada por el área de la salud ha experimentado un crecimiento exponencial, esta información es generada u obtenida a través de diferentes dispositivos médicos, historias clínicas y dispositivos de IoT [3], lo que ha resultado como consecuencia el nacimiento del concepto de Big Data en el área de la salud (BDS). Este término hace referencia a la enorme cantidad de datos que se generan en el ámbito mencionado y que mediante su almacenamiento y análisis, tiene el potencial de proporcionar información valiosa sobre los pacientes, prevención de enfermedades y la gestión de la salud de la población [4].

Sin embargo, gestionar y almacenar esta enorme cantidad de información plantea un reto significativo para los sistemas de salud ya que organizar y analizar grandes cantidades de datos de tipo estructurados y no estructurados de manera eficiente y efectiva requieren un enfoque diferente al almacenamiento tradicional y estos son conocidos como patrones de almacenamiento de Big Data en el contexto de la salud [5]. Estos patrones proporcionan la capacidad de almacenar grandes cantidades de datos de manera escalable de forma local (on-premise) o de forma remota en la nube (cloud) y permiten el análisis rápido y eficaz de la información médica lo cual permite tomar mejores decisiones clínicas y mejorar la atención médica [6].

Las soluciones de almacenamiento on-premise hacen referencia a la implementación de la infraestructura de almacenamiento y procesamiento de datos dentro de las instalaciones físicas de la entidad de salud [3]. En este enfoque, la entidad tiene un control sobre sus sistemas y datos. Por otra parte las implementaciones en la nube, también conocidas como servicios cloud, son ofrecidos por terceros a las entidades de salud, con el mismo principio de funcionamiento que una implementación on-premise sin embargo toda la información y el procesamiento se hace de forma remota [?].

Con el fin de tener una visión más clara respecto a la uso de los patrones de almacenamiento de BDS, se aplicó la metodología del Mapeo Sistemático de la Literatura (SMS - Systematic Mapping System). A través de 7 etapas [7] las cuales permitieron recopilar un total de 70 artículos relacionados al tema del presente estado del arte permitiéndonos generar una clasificación de dichos artículos estructurando la misma para la posterior investigación de los patrones de Almacenamiento de BDS, identificando sus principales usos, las tecnologías mayormente utilizadas para la implementación de patrones de almacenamiento así también los retos de seguridad que afronta la implementación de esta tecnología en el área de la salud.

Este documento está compuesto por 4 secciones: II) Metodología: en esta sección abordaremos la explicación teórica del proceso realizado para la extracción de artículos relevantes para el presente estado del arte, III) Caso de estudio: se abordará de forma detallada el proceso que se ha seguido con base a la metodología presentada, IV) Resultados y Síntesis: se mostrarán los resultados de la investigación realizada y se dará respuesta a las preguntas de investigación propuestas y finalmente la última sección de Conclusiones.

II. METODOLOGÍA

Para el desarrollo de la presente estado de arte se utilizó la metodología de mapeo sistemático (SMS), este proceso riguroso de revisión de literatura científica tiene un total de 7 pasos distribuidos en 3 fases como se observa en la Figura 1.

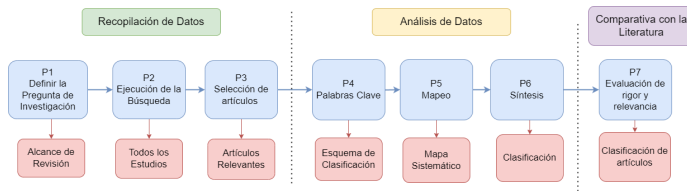


Fig. 1: Diagrama de Proceso de Estudio de Mapeo Sistemático de acuerdo con Navas et al. [7]

La metodología de SMS planteada se agrupa en 3 secciones o fases de acuerdo con lo propuesto por Navas et al. [7] en la metodología GSMS:

- *Recopilación de datos:* se definen las preguntas de investigación y se realiza una búsqueda de artículos relacionados con la pregunta de investigación tras ello se realiza la selección de artículos relevantes para el estado del Arte.
- *Análisis de datos:* en esta etapa deberemos definir las palabras claves mediante un esquema de clasificación, el Mapeo y una síntesis de los artículos seleccionados
- *Comparativa con la literatura:* se evaluará los artículos por rigor científico y relevancia industrial clasificando los mismos.

Dentro de la primera etapa consta de 3 pasos [7], los cuales podemos definir como :

- 1) *Definir la pregunta de investigación:* se plantean que interrogantes podemos esclarecer y bajo este concepto se recopilaran los artículos en pasos posteriores.
- 2) *Ejecución de la búsqueda:* el resultado esperado en esta etapa es obtener una lista de artículos encontrados en fuentes de carácter científico que sea relevante para el tema a estudiar utilizando una cadena de búsqueda.
- 3) *Selección de artículos:* aplicando criterios de exclusión e inclusión se seleccionan los artículos que bajo estos preceptos fueran más relevantes permitiendo realizar un proceso de clasificación de los mismos.

Durante la segunda etapa, se realiza un proceso de conceptualización y análisis de la información, el proceso de clasificación se realiza de forma sistemática donde las palabras claves se utilizan para clasificar los artículos relevantes reduciendo el tiempo necesario para llevar a cabo este proceso; esto permite obtener un set de palabras claves de los diferentes artículos, lo cual, permite tener una comprensión de alto nivel sobre la naturaleza y contribución del trabajo investigativo [2].

Posterior a ello, el mapeo sistemático nos permite crear un esquema de clasificación donde se colocan los artículos en categorías, de forma que se presenta la frecuencia en la que se presenta cierta publicación en cada categoría permitiendo visualizar en que categorías se ha hecho hincapié en investigaciones anteriores y, por tanto, identificar temas que no se han tratado previamente y que pudieran ser abordados en futuros trabajos.

A continuación, en esta etapa en la síntesis se pueden identificar los conceptos en cada estudio clasificado utilizando el nombre original del autor estos conceptos son organizados de forma tabulada con el fin de establecer comparativas entre los estudios más relevantes [1].

Finalmente tenemos la evaluación de rigor y pertinencia que tiene como objetivo presentar la validez de los resultados y sus ventajas concretas aplicando una sencilla función de clasificación, para evaluar el rigor científico y la relevancia industrial de cada estudio primario. El modelo sugerido por Paternoster et al. [1] proporciona un conjunto de rúbricas para medir el rigor científica y la relevancia industrial, estos aspectos deberán ser evaluados en cada uno de los artículos recopilados, en cuanto al rigor es tomado en consideración aspectos como: contexto, diseño del estudio y validez y cuanto a la relevancia hace referencia al realismo del ambiente en el que el estudio fue realizado y el grado en que el método de investigación elegido contribuye a la posibilidad de transferir los resultados a los profesionales teniendo en consideración aspectos como: sujetos, contexto y escala.

Todo este proceso nos da como resultado una clasificación sistemática de todos los artículos recopilados permitiendo presentar un estado del Arte que ayude a futuras investigaciones sobre el uso de patrones de Big Data en el campo de la salud.

III. CASO DE ESTUDIO

Para la elaboración del presente estado del arte se aplicó la metodología SMS, siguiendo el siguiente proceso dividido en 3 etapas (Figura I).

A. Etapa 1: Recopilación de Datos

1) *Definir la Pregunta de Investigación:* En esta etapa se utilizó el método PICO para la obtención de las preguntas de investigación que ayudarán en definir y delimitar el área de estudio del presente estado del arte con el fin de obtener artículos referentes patrones de almacenamiento de BDS, como recurso de investigación bibliográfica [8], los criterios de búsqueda definidos pueden verse en la Tabla I.

Tabla I: Método PICO

PICO	Consideraciones
Population(P): ¿Quién?	Área de Salud
Intervention(I): ¿Qué?, ¿Cómo?	Técnicas Utilizadas para almacenar y organizar grandes cantidades de datos (Big Data)
Comparison (C): ¿Con qué comparar?	Efectos de los patrones de almacenamiento de Big Data en el área de salud
Outcomes (O): ¿Qué se busca conseguir/mejorar?	La eficiencia en la toma de decisiones médicas, la mejora de investigación científica, privacidad y seguridad de los datos

Preguntas de Investigación:

- *SMSP1: "¿Cuáles son los factores clave que influyen en la elección entre soluciones de almacenamiento de big data locales (On-premises) o basadas en la nube (Cloud) en el sector sanitario?"*.
- *SMSP2: "¿Cómo afectan los patrones de almacenamiento de big data a los resultados de la atención sanitaria, y cuáles son las implicaciones para los proveedores de atención sanitaria y los responsables políticos?"*.
- *SMSP3: "¿Cuáles son los patrones de almacenamiento de big data más comunes utilizados en el sector sanitario, y cuáles son las ventajas y desventajas de las implementaciones on-premise y en la nube?"*.

Se determinaron las palabras clave con base a las preguntas de investigación obtenidas con el método PICO estos términos son: Data Privacy, Data Security, Data Governance, big data storage patterns, health area, almacenamiento en la nube, distributed system, scalable system, bases de datos distribuida, clousters storage, block storage, sistemas de archivos distribuidos, architectures; de estos se estructuró una cadena de búsqueda la cual será utilizada en los repositorios seleccionados haciendo uso de expresiones booleanas (AND, OR, NOR) las cuales ayudarán a estructurar la cadena de búsqueda presentada a continuación: ("**BigData**" OR "**Big Data**") AND "**Storage**" AND "**HealthCare**".

Tabla II: Resultados de la cadena de búsqueda

Repositorio	Tipo de Artículo	Resultados
IEEE	Artículos y Conferencias	150
Web of Science	Artículos	200
Scopus	Artículos	180
	TOTAL	530

Criterios de Elegibilidad: Con el fin de seleccionar los artículos más relevantes se utilizaron criterios de inclusión y exclusión de la metodología SMS especificados en la Tabla III teniendo en cuenta los siguientes conceptos:

- **Criterios de Inclusión:** La literatura define al criterio de inclusión como las características de tipo temporal, geográfica o idiomática que ayudan a seleccionar una muestra de una población de estudio [9].
- **Criterios de Exclusión:** Se conoce a los criterios de exclusión como las características no deseadas para la extracción de una muestra de una población de estudio [9].

2) *Ejecución de la búsqueda:* Durante el desarrollo de esta etapa se hace uso de la cadena de búsqueda que se estableció en la etapa anterior y se aplica en cada uno de los 3 repositorios seleccionados, que en el caso del presente estado del arte son: IEEE, Scopus y Web of Science, en estos se obtuvo en formato de archivo CSV el listado de los artículos que tuvieran relación con la cadena de búsqueda teniendo en consideración que se tomarían en cuenta los criterios incluidos en la tabla II, esto con la finalidad de garantizar que fueren artículos relevantes para el presente estudio.

3) *Selección de Artículos:* Una vez obtenidos la totalidad de 530 artículos se procede a filtrar los mismos, con el fin de reducir la cantidad de artículos a un mínimo de 60 hasta un máximo de 100 artículos, a efecto de lo cual se tomaron los criterios de inclusión y exclusión previamente mencionados en la Tabla IV. Aplicando una escala del 1 al 5 donde se mantienen los siguientes criterios:

- 1) No existe Relación
- 2) Existe poca Relación
- 3) Medianamente Relacionado
- 4) Existe Relación
- 5) Altamente Relacionado

Filtro 1. Haciendo uso de los criterios de inclusión y exclusión mencionados en la Tabla IV, de ellos se eligieron el rango de fecha entre 2019 hasta 2023, el idioma inglés y los artículos que tuvieran libre acceso, una vez aplicados estos filtros, quedando con un total de 373 Artículos.

Filtro 2. Se evaluó los artículos restantes calificando cada uno con base a una escala de 1-5 que indica el nivel de correlación que los títulos de los artículos presentaban con el presente estado del arte, adicionalmente con ayuda de la herramienta de OpenAI ChatGPT para contrastar con la calificación que se dio de forma manual, obteniendo un total

de 133 artículos y seleccionando aquellos con una calificación de 4 y 5.

Filtro 3. Tomando consideración las palabras claves contenidas en cada uno de los artículos con ayuda de ChatGPT de OpenAI como herramienta de apoyo se revisaron de las palabras claves de cada uno de los artículos, comparando estas con las palabras claves generadas mediante el método PICO, de igual forma se realizó la calificación de los artículos con la escala obteniendo finalmente un total de 114 artículos.

Filtro 4. Finalmente se realizó una lectura del abstract de cada artículo y se evaluó la relación de los mismos con el tema del presente estado del arte calificando con la escala (1-5) obteniendo finalmente un total de 70 artículos.

Tabla III: Resultados de Selección de Artículos

Filtro	Art. Iniciales	Criterio	Art. Finales
1	530	Criterios Inclusión y Exclusión	373
2	373	Correlación de Títulos	133
3	133	Correlación de Keywords	114
4	114	Lectura de Abstract	70
		TOTAL	70

4) *Palabras Clave (Keywording):* Habiendo aplicado los correspondientes filtros a los artículos seleccionados se procede a realizar una lectura de cada uno de los 70 artículos finales, esto con el propósito de identificar conceptos y/o palabras clave que sean relevantes para el presente trabajo, este proceso se realizó mediante una lectura rigurosa de cada uno de los artículos extraídos.

Durante esta etapa se utilizó Microsoft Excel como herramienta que nos permitió documentar y categorizar las palabras clave identificadas, conforme se llevaba a cabo el proceso de lectura y adicionalmente se utilizó la herramienta Atlas.ti (Figura 3) con el fin de obtener el número de veces que cada una de las palabras seleccionadas se repite en cada uno de los documentos revisados; una vez finalizado este proceso se obtuvo un total de 163 términos los cuales podemos observar en la Figura 2. Una vez obtenidos los términos estos fueron clasificados de tal forma que se obtuvieron Algoritmos, Modelos, Formatos de Codificación, Técnicas y Tipos de almacenamiento esto dando como resultado una clasificación más completa respecto a cuales son los aspectos principales de los patrones de almacenamiento de Big Data en el área de salud, de acuerdo con los artículos revisados.

5) *Mapeo:* Durante el desarrollo de la etapa de keywording se identificó de forma recurrente tres aspectos principales: seguridad, privacidad y almacenamiento; a su vez, dentro de los artículos que abordaban estos aspectos se identificaron subcategorías que abarcaban de mejor forma diferentes clasificaciones que exploraremos a continuación.

Dentro de los artículos que abordaban el tema de seguridad se identificaron sub-categorías adicionales como algoritmos de encriptación, modelos de control de accesos, formatos de

codificación. Estas sub-categorías permiten una clasificación más detallada de los artículos y facilita el análisis de los mismos.

Se observó que, dentro de los artículos revisados que tratan el tema de privacidad, se identificaron sub-categorías que abordan técnicas de preservación de la privacidad y modelos para la protección de datos sensibles. Además, se encontró una amplia discusión sobre el almacenamiento en los trabajos revisados, el cual se subdividió en temas como el almacenamiento distribuido en la nube, la utilización de fog computing y el empleo de sistemas de archivos distribuidos, entre otros.

Este proceso dio como resultado el diagrama de relación jerárquica que puede ser apreciado en la Figura 4. A continuación se detalla cada una de las categorías identificadas y su ocurrencia dentro de los artículos seleccionados:

a) *Patrones de Almacenamiento de Big Data en el área de salud:* De acuerdo con Hong et al. [3] en el contexto de la salud Big Data se presenta con características propias además de las que podemos observar en otros campos, pudiendo ser: incompleta, heterogénea, puntual, longeva y privada; esto conlleva a plantear una serie de retos a la hora de almacenar, extraer y compartir estos datos.

Siendo los patrones de almacenamiento la forma en cómo esta información es almacenada, se ha podido identificar gracias al proceso de Keywording las 3 características predominantes en el área de la salud (Figura 5)

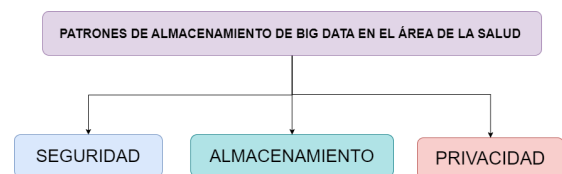


Fig. 5: Características de patrones de almacenamiento de Big Data en el área de Salud

Los patrones de almacenamiento de Big Data tienen como prioridad garantizar la seguridad, privacidad y almacenamiento adecuado de la información médica, de acuerdo con los artículos revisados, estas son fundamentales para cualquier patrón de almacenamiento, en caso de que una de estas falle, se generarían amenazas significativas que afectarían a la toma de decisiones por parte de los profesionales de la salud poniendo en peligro la vida de sus pacientes.

Seguridad

Preservar la integridad y confidencialidad sobre la información sensible de los pacientes es fundamental, la exposición no autorizada de esta información podría resultar en violaciones a la privacidad de los pacientes, robo de identidad, discriminación y otros riesgos para su seguridad y bienestar, a lo cual establecer mecanismos robustos con el fin de prevenir accesos no autorizados.

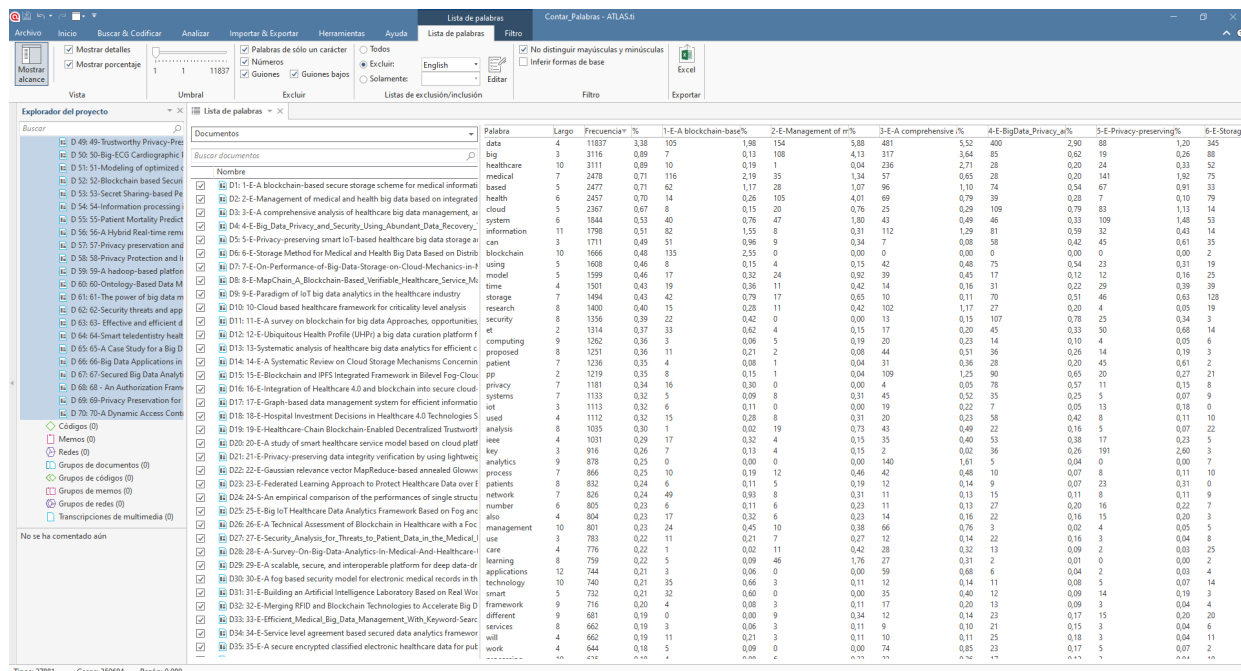


Fig. 3: Conteo de Frecuencia de palabras clave en Atlas.ti

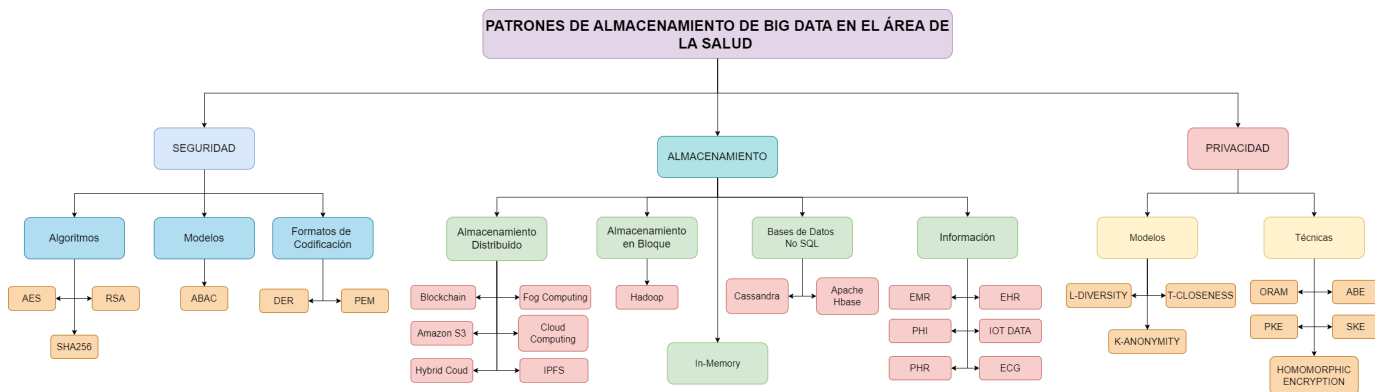


Fig. 4: Diagrama de relación Jerárquica

esta forma se garantiza que aunque la información almacenada en la estructura de Big Data fuere extraída de forma ilegal, esta no pueda ser leída por terceros que no poseen las llaves de acceso para la misma. En la Tabla VI podemos ver de forma etiquetada los algoritmos de seguridad que han sido recopilados de los artículos revisados en el presente estado del arte.

N°	Algoritmo de Seguridad	Referencia
AS1	SHA256	[12], [23], [24], [14], [25]
AS2	AES	[26], [27], [28], [23], [29], [30], [31], [32]
AS3	RSA	[26], [16], [27], [33], [23], [34], [35], [32], [36], [14]

Tabla VI: Ejemplos de Algoritmos de Seguridad

Formatos de Codificación para la Seguridad

En [37] se menciona que estos formatos se utilizan para proteger la información almacenada en los patrones de almacenamiento de acceso no autorizado, modificación o destrucción de los mismos, recordemos que estos formatos son utilizados a menudo para transportar la información en sistemas de información para la salud [38], esto va de la mano junto con los algoritmos de seguridad que ayudan a garantizar la confidencialidad de la información durante su envío.

Los formatos de codificación más destacados que han sido identificados en el presente trabajo son Distinguished Encoding Rules (DER) y Privacy Enhanced Mail (PEM).

DER se presenta como un formato de codificación determinista lo que implica que solo existe una forma de codificar una estructura de datos siendo esta la notación de sintaxis abstracta 1 (ASN.1 - Abstract Syntax Notation One) [39]. Es uno de los dos formatos de codificación más comunes para certificados

X.509 y claves privadas, siendo el otro PEM [40].

Por otra parte, PEM es una versión codificada en base64 del formato DER, es uno de los formatos más comunes para certificados X.509 y claves privadas [41], este es especialmente utilizado en contextos sanitarios donde es importante poder leer y verificar fácilmente el contenido de un archivo [42]; este formato de codificación se utiliza en el área de la salud para el almacenamiento de claves de cifrado o la transmisión de historiales médicos electrónicos (HCE).

En la Tabla VII podemos ver de forma etiquetada los Formatos de Codificación que han sido recopilados de los artículos revisados en el presente estado del arte.

N°	Formato de Codificación	Referencia
FC1	DER	[27], [17], [43], [33], [44], [35], [45], [14]
FC2	PEM	[27], [30], [31], [45]

Tabla VII: Ejemplos de Algoritmos de Seguridad

Almacenamiento

De acuerdo con Sarkar [46] Podemos observar que el almacenamiento de Big Data en el área de salud se ha vuelto un reto importante ya que con el fin de poder satisfacer la demanda de almacenamiento que tienen los servicios médicos han debido crearse nuevas estructuras de almacenamiento que permitan soportar esta cantidad creciente de información, solo en el almacenamiento de imágenes DICOM que ayudan al diagnóstico de cáncer se han requerido peta-bytes en espacio de almacenamiento [46], con el fin de poder mantener estudiar y preservar esta información que es de alto valor tanto como para los pacientes como para profesionales de la salud.

Hay que recordar también que [10] la información almacenada en los diferentes patrones de almacenamiento de Big Data no se limita únicamente a registros médicos de historias clínicas si no también se habla de registros Financieros, Genoma, Físicos, Perfiles psicológicos etc, por lo cual las estructuras de almacenamiento deben ser capaces de adaptarse a todo tipo de información con o sin estructura y garantizar que estas no sufran pérdidas en su integridad debido a compresión y otros procesos.

Tipos de Información Almacenada

Dentro del área de la salud, en los diferentes patrones de almacenamiento de Big Data se puede encontrar historias clínicas, imágenes médicas, información relacionada con el cuidado de la salud, toda esta información puede ser: estructurada, no estructurada y semi estructurada. En [47] se hace mención a la información almacenada y cómo el conocer la estructura de la información almacenada ayuda a un procesamiento más eficiente, para lo cual en la Tabla VIII podemos ver de forma etiquetada los tipos de información almacenada que han sido recopilados de los artículos revisados en el presente estado del arte.

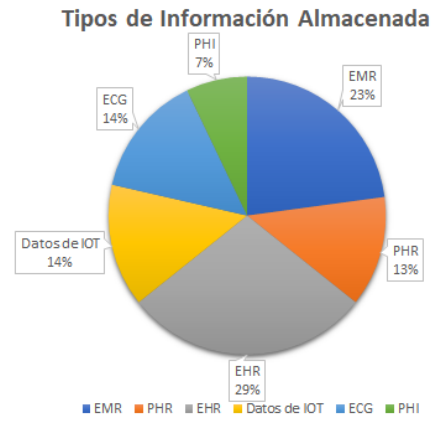


Fig. 6: Tipos de Información Almacenada

N°	Tipos de Información	Referencia
IA1	EMR	[48], [49], [16], [50], [51], [43], [23], [52], [53], [54], [55], [18], [30], [56], [45], [36]
IA2	PHR	[16], [50], [43], [57], [58], [31], [59], [32], [60]
IA3	EHR	[12], [15], [61], [62], [48], [49], [63], [16], [64], [65], [43], [44], [29], [57], [55], [18], [20], [21], [32], [66], [67], [60], [36], [68], [21]
IA4	IOT DATA	[12], [26], [13], [69], [70], [71], [17], [65], [53], [72], [19], [25], [35]
IA5	ECG	[26], [48], [49], [63], [73], [33], [52], [72], [74], [54], [30], [21], [25], [59], [32], [45]
IA6	PHI	[33], [52], [55], [34], [75]

Tabla VIII: Ejemplos de Información Almacenada

Almacenamiento en Bloque (Block-Storage)

Este almacenamiento es un método que ayuda a organizar la información en bloques de tamaño predeterminado, cada bloque es asignado con un identificador único, el tamaño de estos bloques típicamente es de 512 bytes a 4 kilo-bytes estos bloques son almacenados de forma lineal [76], acceder a estos es realmente eficiente ya que es posible acceder a ellos de forma independiente y ayudan a mantener un alto rendimiento en las aplicaciones en donde se utiliza, en el caso de su uso en Patrones de almacenamiento de BDS en la Tabla IX podemos ver de forma etiquetada un ejemplo de información almacenado en bloque que ha sido recopilado de los artículos revisados en el presente estado del arte.

N°	Almacenamiento en Bloque	Referencia
ABI	Hadoop	[12], [61], [26], [69], [77], [62], [70], [48], [71], [49], [63], [78], [16], [17], [64], [65], [33], [54], [21], [25], [35], [79], [67], [60], [68]

Tabla IX: Ejemplo de Almacenamiento en Bloque

Almacenamiento In-Memory

En [80] se habla este tipo de almacenamiento, el cual guarda la información en la memoria de acceso aleatorio (RAM) de la computadora, permitiendo acceder de forma sencilla a la información, esto usualmente con el propósito únicamente de agilizar procesamiento de información previo a ser enviado a una estructura de almacenamiento a largo plazo, de igual forma se ha encontrado referencias que han sido identificadas en la Tabla X estas han sido recopilado de los artículos revisados en el presente estado del arte.

N°	Almacenamiento In-Memory	Referencia
AIM	In-Memory	[69], [71], [49], [63], [78], [81], [17], [64], [82], [21], [25], [67], [60]

Tabla X: Ejemplo de Almacenamiento In-Memory

Almacenamiento en Bases de Datos no-SQL

El almacenamiento de Big Data supone un gran reto ya que gran parte de la información que esta maneja se presenta como información estructurada, no estructurada o semi estructurada [47], en respuesta a este reto se han implementado bases de datos no SQL que permiten almacenar información que cumple con lo antes mencionado, al ser mucho más flexibles permite transportar esta información en archivos de tipo JSON garantizando que no existirán pérdidas o si las hubiere estas serán mínimas. En la Tabla XI podemos ver de forma etiquetada los ejemplos de Base de Datos no-SQL que han sido recopilados de los artículos revisados en el presente estado del arte.

N°	Bases de Datos No-SQL	Referencia
NS1	Cassandra	[62], [65], [54], [32],
NS2	Apache Hbase	[61], [69], [62], [63], [30], [19],

Tabla XI: Ejemplo de Almacenamiento en bases No-SQL

Almacenamiento Distribuido

Como se menciona en [83] el almacenamiento distribuido ayuda a permitir que la información sea más accesible y escalable al estar distribuida en múltiples nodos dentro de una red, así también esta forma de almacenar información permite particionar la información, replicar la información en diferentes nodos con el fin de mantener una redundancia, permitiendo de igual forma acceder a la información de forma eficiente obteniendo únicamente lo necesario sin afectar al rendimiento tanto de la red como de la aplicación que se utiliza. En la Tabla XII podemos ver de forma etiquetada los ejemplos de Almacenamiento Distribuido que han sido recopilados de los artículos revisados en el presente estado del arte.

N°	Almacenamiento Distribuido	Referencia
AD1	BlockChain	[12], [69], [71], [73], [27], [51], [52], [72], [55], [84]
AD2	Hybrid cCloud	[26], [13], [77], [70], [48], [78], [43], [29], [85], [30]
AD3	Amazon S3	[62], [78], [16], [54], [30]
AD4	Fog Computing	[61], [26], [13], [69], [77], [62], [70], [71], [63], [78], [73], [27], [28], [65], [51], [62], [23], [52], [53], [85], [18], [30], [34], [25], [59], [35], [60], [14]
AD5	Cloud Computing	[12], [15], [61], [26], [13], [69], [77], [62], [70], [48], [71], [49], [63], [78], [73], [63], [86], [50], [28], [17], [64], [65], [87], [43], [33], [23], [52], [53], [72], [29], [85], [57], [74], [54], [18], [19], [20], [24], [34], [21], [58], [31], [25], [59], [88], [32], [56], [45], [75], [36], [21]
AD7	IPFS	[12], [71], [73], [27], [51], [87], [36]

Tabla XII: Ejemplos de Almacenamiento Distribuido

Privacidad

La Protección de información sensible y personal contenida en los data-sets que son recopilados, almacenados y en algunos casos analizados en el área de salud son de alto valor [37], ya que en el área de salud esta información brinda información vital para toma de decisiones respecto a la salud de sus pacientes. La privacidad dentro de los patrones de almacenamiento de Big Data en el área de salud se comprenden como medidas que ayudan a asegurar la información personal de los individuos solo sea accesible para el personal autorizado, minimizando el riesgo de atentar contra las regulaciones que se tienen respecto a la protección de esta información.

Modelos para proteger la Privacidad

Los modelos de privacidad relacionados con los patrones de almacenamiento de Big Data en el área de salud hacen referencia a marcos conceptuales o enfoques diseñados para proteger la privacidad de información de carácter sensible de los pacientes dentro de los data-sets [89], estos modelos proporcionan directrices, principios y técnicas que evaluaremos posteriormente para que se garantice un almacenamiento y análisis en el que se respete la privacidad del paciente cuya información se está utilizando con el fin de obtener cierta información con base los datos recopilados del mismo. En la Tabla XIII podemos ver de forma etiquetada los ejemplos de Modelos de Privacidad que han sido recopilados de los artículos revisados en el presente estado del arte.

N°	Modelos de Privacidad	Referencia
MP1	K-anonymity	[26], [73], [17], [64], [88], [68]
MP2	L-Diversity	[17], [64], [18], [24], [88], [68]
MP3	T-Closeness	[49], [17], [64], [88], [68]

Tabla XIII: Ejemplo de Modelos de Privacidad

Técnicas para proteger la privacidad

En [89] se menciona que estas técnicas son mecanismos o aproximaciones que ayudan a salvaguardar información sensible que se encuentra almacenada en los data-sets, estas técnicas buscan asegurar que la información personal del paciente como por ejemplo PHR o PHI solo pueda ser accesible para individuos autorizados o sistemas y que esta información sea almacenada de forma segura en concordancia con las regulaciones de cada país. En la Tabla XIV podemos ver de forma etiquetada los ejemplos de técnicas para proteger la privacidad que han sido recopilados de los artículos revisados en el presente estado del arte.

N°	Técnicas de Privacidad	Referencia
TP1	Homomorphic Encryption	[26], [71], [16], [28], [17], [64], [43], [30], [34], [25], [88]
TP2	ORAM	[26], [49], [64], [18], [24], [54]
TP3	ABE	[26], [13], [27], [43], [53], [29], [14]
TP4	PKE	[26], [53], [66]
TP5	SKE	[26], [13], [49], [72]

Tabla XIV: Ejemplo de Técnicas para proteger la Privacidad

IV. RESULTADOS Y SÍNTESIS

1) SMSP1: ¿Cuáles son los factores clave que influyen en la elección entre soluciones de almacenamiento de big data locales (On-Premises) o basadas en la nube (Cloud) en el sector sanitario??

Tomando en consideración los estudios revisados durante el presente estado del arte, se pudo definir una taxonomía que se centra en los patrones de almacenamiento de Big Data en el área de salud, en consecuencia se obtuvo una jerarquía que consta de 3 niveles que corresponden a:

- a) Seguridad
- b) Privacidad
- c) Almacenamiento

Estos 3 factores responden a las necesidades clave que tiene el área de salud al momento de decidir cuál solución de almacenamiento de Big Data se acopla mejor a ellos, dado que el sector de salud maneja información sensible de pacientes, requieren soluciones que presten una seguridad robusta y presenten medidas de seguridad que garanticen el anonimato de sus pacientes [64]; las soluciones locales ayudan a tener un control directo sobre la información permitiendo a las organizaciones implementar protocolos a la medida de sus necesidades [90], por el contrario las soluciones basadas en la nube que si bien ofrecen funciones de seguridad avanzadas, es necesario que las organizaciones examinen las prácticas en el tratamiento de datos de los proveedores de servicios en la nube y que estas no violen la seguridad o la privacidad de los pacientes [91].

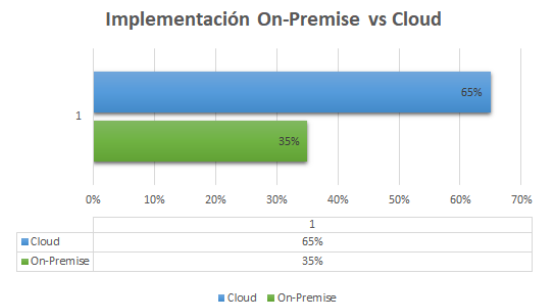


Fig. 7: Soluciones On Premise vs Soluciones Cloud

De un total de 70 artículos revisados hemos podido determinar que en un 65% de ellos se ha mostrado que existe una tendencia creciente por adoptar soluciones de almacenamiento de Big Data basadas en la Nube como se puede observar en la Figura 5, dado que proveen a las organizaciones dedicadas a la salud una capacidad de almacenamiento ilimitado, tomando en cuenta que el problema principal de almacenamiento en el área de salud viene dado de un crecimiento exponencial de la información generada, esto hace que sean una opción más viable, adicionalmente, proveen ventajas como copias de seguridad automáticas y redundancia de datos, sumado a esto el alto costo de una implementación de infraestructura y personal que de mantenimiento a la misma hace que las soluciones On-premise no sean tan atractivas, sin embargo aún existe un 35% de instituciones que aún utilizan este tipo de infraestructura o a su vez optan por un sistema híbrido entre estas dos soluciones para el almacenamiento de Big Data.

2) SMSP2: ¿Cómo afectan los patrones de almacenamiento de big data a los resultados de la atención sanitaria, y cuáles son las implicaciones para los proveedores de atención sanitaria y los responsables políticos??

Los patrones de almacenamiento tienen un impacto muy significativo en los resultados de la atención sanitaria, debido a que un almacenamiento eficiente permite a quienes brindan los servicios en el área de salud poder acceder a la información de sus pacientes, analizarla mediante modelos predictivos, modelos de Machine Learning que ayudan a identificar patrones, anomalías y pudiendo predecir potenciales riesgos para la salud del paciente [92].

Con una estructura apropiada para el almacenamiento de Big Data los responsables políticos y proveedores al analizar la información referente a la salud de la población a gran escala pueden identificar tendencias, patrones y factores de riesgo de la salud pública [93]. Al conocer las necesidades sanitarias de poblaciones específicas, los responsables políticos pueden diseñar planes para intervenir de forma específica en ciertos

sectores sociales, asignar recursos de forma eficiente y aplicar estrategias de prevención [94]. Este enfoque puede ayudar a mejorar los resultados en los servicios de salud brindados en una comunidad específica, reducir costes sanitarios y mejorar la calidad de vida de los pacientes.

También es importante recordar que a medida que los patrones de almacenamiento de Big Data permiten recopilar la información de los pacientes, es importante que los proveedores de servicios así también como los responsables políticos están en la obligación de establecer marcos sólidos de gobernanza de datos, garantizar cumplimiento de la normativa de cada país y aplicar medidas de seguridad robustas para proteger la privacidad de los pacientes y evitar accesos no autorizados o filtraciones de datos [95].

3) **SMSP3: “¿Cuáles son los patrones de almacenamiento de Big Data más comunes utilizados en el sector sanitario, y cuáles son las ventajas y desventajas de las implementaciones on-premise y en la nube?”**

De acuerdo con [96], existen 2 grandes tipos de patrones de almacenamiento, los patrones de almacenamiento de datos estructurados los cuales utilizan bases de datos tradicionales como SQL y los patrones de almacenamiento de datos no estructurados, este último se presenta de forma recurrente en la revisión de los artículos recopilados en el presente estado del arte.

Tomando en consideración la información recabada el 100% de los artículos revisados hacen mención a patrones de almacenamiento de datos no estructurados, tomando en consideración esto se encontró que los patrones de almacenamiento más comunes son: Blockchain, Amazon S3, IPFS, Cloud Computing, Hybrid Cloud y Fog Computing.

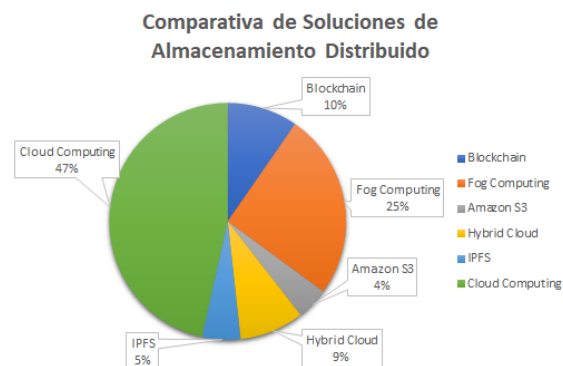


Fig. 8: Soluciones de Almacenamiento Distribuido

Como se observa en la Figura 6 de acuerdo al análisis

realizado con base a los artículos revisados, el patrón de almacenamiento de Big Data más común en el área de salud es Cloud Computing con un 45%, esto en concordancia a los datos obtenidos que indican que en el área de salud mantiene la tendencia a utilizar servicios en la nube.

a) **Ventajas Implementación On-Premise**

La implementación de una solución On-Premise tiene como ventaja el control total de las organizaciones, brindando un mayor control sobre la infraestructura de datos [97], adicional a ello permite al sector de la salud cumplir con mayor facilidad la normativa que existe en cada país.

b) **Desventajas Implementación On-Premise**

La principal desventaja que presenta esta solución radica en los costos iniciales elevados, debido a que se requiere una inversión significativa en infraestructura y personal para mantener los sistemas [98], adicional a ello posee una escalabilidad que se ve limitada por la dificultad en la gestión de nuevos equipos o equipos adicionales para el sistema ya implementado [99].

c) **Ventajas Implementación en la nube**

Las soluciones en la nube ofrecen un precio menor de implementación inicial dado que no se requiere una gran infraestructura para su implementación, ofrece una escalabilidad y flexibilidad que se adapta a las necesidades cambiantes del cliente [100], además las actualizaciones y mantenimiento de la infraestructura son responsabilidad del proveedor de servicios en la nube [101].

d) **Desventajas Implementación en la nube**

Estas soluciones dependen íntegramente del acceso a internet por lo que se puede afectar el acceso y procesamiento de datos en momentos en que la conexión sea inestable o no esté disponible [11]. Además uno de los problemas más graves radica en garantizar que quién fuere el proveedor de servicios en la nube garantice la privacidad y seguridad de la información que la normativa de cada país requiera.

La implementación y uso de los patrones de almacenamiento de Big Data en el área de salud presentan una serie de desafíos y oportunidades para mejorar la calidad y la eficiencia de los servicios sanitarios, estos deben ser implementados como soluciones de almacenamiento escalables, distribuidos y tolerantes a fallos, sobre todo tomando en consideración que no existe un único patrón que se adapte a todas las necesidades si no que es necesario elegir el más adecuado según las características y los requisitos que tenga cada escenario.

1) *Evaluación de rigor y relevancia:* En la presente sección se han evaluado los criterios presentados en la Tabla XV con la finalidad clasificar rigor científico y relevancia industrial de los 70 artículos finales que se obtuvieron mediante la metodología

de mapeo sistemático de la literatura propuesto por Navas et al. [7].

Tabla XV: Criterios de Rigor Científico y Relevancia Industrial

Rigor Científico	Relevancia Industrial
Métodos y diseño de investigación claramente descritos	Describe o presenta desafíos reales en la industria
Aplica metodologías de investigación	El artículo fue realizado con base a casos de estudios provenientes de la industria
Los resultados se acoplan a la aplicación de la metodología	El artículo potencialmente puede generar beneficios o mejoras en los procesos de la industria

De acuerdo con los criterios presentados anteriormente se han evaluado los artículos obtenidos, estos han sido clasificados de forma sistemática haciendo uso de la herramienta Microsoft Excel para documentarlos y la herramienta Atlas.ti con el fin de realizar la búsqueda de palabras clave que ayuden a corroborar los criterios presentados den la Tabla XV se ha obtenido la siguiente gráfica.

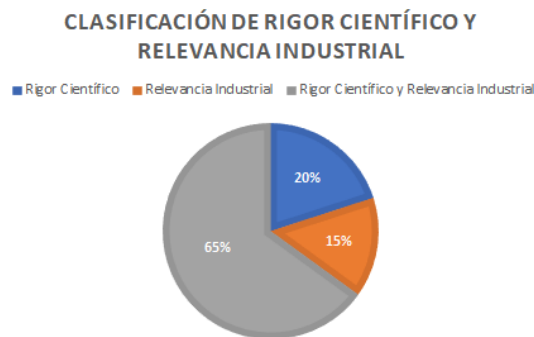


Fig. 9: Clasificación de Rigor Científico y Relevancia industrial

Podemos denotar que el 20% de los artículos que han sido clasificados presentan únicamente rigor científico y el 15% de los artículos cumplen únicamente con los criterios de relevancia industrial, sin embargo tenemos un %65 de artículos que presentan ambas clasificaciones, esto nos indica que los artículos seleccionados gracias la metodología de mapeo sistemático según Navas et al. [7], ayuda a filtrar artículos que presentan problemas reales y están acorde a los estándares industriales actuales en el momento de su publicación y pudiendo utilizar el presente estado de arte como un punto de partida para futuros trabajos que busquen ampliar la información respecto a patrones de almacenamiento de Big Data en el área de Salud.

V. CONCLUSIONES

Una vez finalizada la investigación aplicando la metodología de mapeo sistemático de la literatura propuesto por Petersen et al. [2] y Paternoster et. al. [1] con el agrupamiento propuesto Navas et al. [7] para la metodología GSMS, se ha podido

llegar a la siguientes conclusiones:

El uso de la metodología de mapeo sistemático (SMS) fue de gran ayuda para poder filtrar los artículos que contuvieran información relevante sobre patrones de almacenamiento de Big data, llevando a cabo 7 etapas las cuales permitieron reducir de un total inicial de 530 artículos académicos referentes al tema del presente estado del arte filtrar los mismos hasta llegar a un total de 70 artículos obtenidos de las plataformas científicas Web of Science, Scopus e IEEE, que brindaron información relevante para este trabajo.

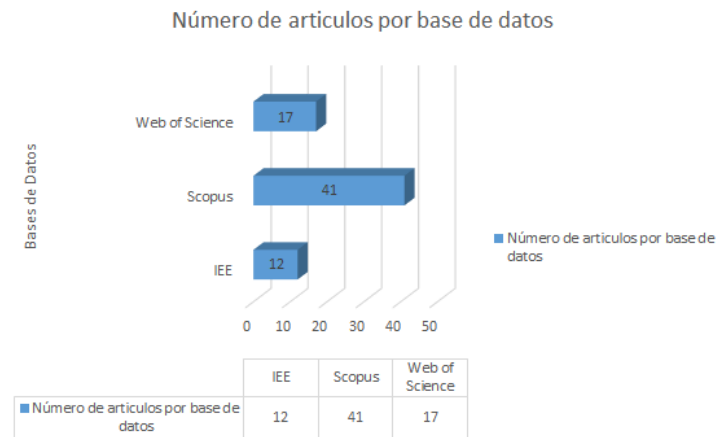


Fig. 10: Número de Artículos por Plataforma

Los patrones de almacenamiento de Big Data son estructuras diseñadas para asegurar el almacenamiento seguro y privado de información personal y médica de pacientes en un sistema de salud. Esta información es muy variada y en su gran parte no estructurada, lo que representa un reto frente a las estructuras de almacenamiento que pueden ser implementadas o las que deberán ser diseñadas a la medida dependiendo de las circunstancias de cada centro médico y área de la salud específica donde se requiera su implementación.

En este estudio se ha observado una tendencia creciente a hacer uso de soluciones basadas en la nube frente a soluciones on-premises esto debido a su facilidad de implementación y bajo costo inicial, el cual se presenta como una solución escalable y asequible para los problemas de almacenamiento que a diferencia de una implementación local requiere de una planificación anticipada y no responde a la creciente necesidad de almacenamiento de información médica.

La clasificación de los patrones de almacenamiento de BDS se puede dividir en dos grandes grupos: almacenamiento estructurado y almacenamiento no estructurado.

Dentro de los patrones de almacenamiento estructurado, se destacan el almacenamiento en bases de datos relacionales y Hadoop. Sin embargo, se han vuelto rápidamente una solución poco eficiente debido a la gran cantidad de datos que se generan. Por lo cual las soluciones de almacenamiento de datos no estructurado están ocupando su lugar. Como se mencionó anteriormente las soluciones en la nube son las más utilizadas

en la actualidad. Dentro de los patrones de almacenamiento no estructurado podemos encontrar las soluciones de almacenamiento distribuido como: Cloud Computing, BlockChain e IPFS, estas se adaptan de mejor forma a las necesidades actuales del área de la salud.

También se ha observado que existen soluciones de almacenamiento híbridas las cuales combinan el almacenamiento en la nube y on-premise con el fin de aprovechar ambas soluciones. Esto permite lograr una escalabilidad y acceso remoto a los datos gracias a la tecnología de la nube y al mismo tiempo se protegen datos sensibles o críticos en entornos on-premise para garantizar un mayor control y seguridad de la información.

Según los artículos revisados, se ha identificado una creciente preocupación por el tratamiento de los datos que compañías proveedoras de servicios en la nube puedan dar ya que es crucial mantener la privacidad de los pacientes y garantizar que no existirán filtraciones de datos, para abordar esta problemática se ha identificado la implementación de diferentes modelos y enfocados en la seguridad y la privacidad, formatos de codificación y técnicas para asegurar la seguridad de la información almacenada como se observa en la Figura 4.

La evolución y el constante crecimiento de la información planean desafíos para futuras generaciones en cuanto al tratamiento de la información médica, ya que esta es invaluable para identificar patrones, prevenir enfermedades e inclusive mejorar el estilo de vida de los pacientes basados datos reales. La información es el futuro por ello es importante garantizar que su preservación, almacenamiento y tratamiento serán adecuados para su futuro uso.

REFERENCES

- [1] N. Paternoster, C. Giardino, M. Unterkalmsteiner, T. Gorschek, and P. Abrahamsson, "Software development in startup companies: A systematic mapping study," *Information and Software Technology*, vol. 56, no. 10, pp. 1200–1218, 2014. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0950584914000950>
- [2] K. Petersen, S. Vakkalanka, and L. Kuzniarz, "Guidelines for conducting systematic mapping studies in software engineering: An update," *Information and Software Technology*, vol. 64, pp. 1–18, 2015. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0950584915000646>
- [3] L. Hong, M. Luo, R. Wang, P. Lu, W. Lu, and L. Lu, "Big data in health care: Applications and challenges," *Data and Information Management*, vol. 2, no. 3, pp. 175–197, 2018. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S25433925122000791>
- [4] S. Khanra, A. Dhir, A. K. M. N. Islam, and M. Mäntymäki, "Big data analytics in healthcare: a systematic literature review," *Enterprise Information Systems*, vol. 14, no. 7, pp. 878–912, 2020. [Online]. Available: <https://doi.org/10.1080/17517575.2020.1812005>
- [5] A. Shaikat, R. Rakesh, M. Indrajit, and S. B. Kanti, "A systematic review of healthcare big data," *Hindawi*, vol. 2020, no. 7, pp. 878–912, 2020. [Online]. Available: <https://doi.org/10.1155/2020/5471849>
- [6] S. Pouyanfar, Y. Yang, S.-C. Chen, M.-L. Shyu, and S. S. Iyengar, "Multimedia big data analytics: A survey," *ACM Comput. Surv.*, vol. 51, no. 1, jan 2018. [Online]. Available: <https://doi.org/10.1145/3150226>
- [7] G. Navas and A. Yagüe, "Glaserian systematic mapping study: An integrating methodology," in *Proceedings of the 17th International Conference on Evaluation of Novel Approaches to Software Engineering - Volume 1: ENASE*, INSTICC. SciTePress, 2022, pp. 519–527.

- [8] T. F. Frandsen, M. F. Bruun Nielsen, C. L. Lindhardt, and M. B. Eriksen, "Using the full pico model as a search tool for systematic reviews resulted in lower recall for some pico elements," *Journal of Clinical Epidemiology*, vol. 127, pp. 69–75, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0895435620305692>
- [9] L. M. Connelly, "Inclusion and exclusion criteria," *Medsurg Nursing*, vol. 29, no. 2, p. 125, Mar 2020, copyright - Copyright Anthony J. Jannetti, Inc. Mar/Apr 2020; Última actualización - 2021-06-02. [Online]. Available: <https://www.proquest.com/scholarly-journals/inclusion-exclusion-criteria/docview/2388933304/se-2>
- [10] H. Kupwade Patil and R. Seshadri, "Big data security and privacy issues in healthcare," pp. 762–765, 2014.
- [11] S. Nazir, S. Khan, H. U. Khan, S. Ali, I. García-Magariño, R. B. Atan, and M. Nawaz, "A comprehensive analysis of healthcare big data management, analytics and scientific programming," *IEEE Access*, vol. 8, pp. 95 714–95 733, 2020.
- [12] Z. Sun, D. Han, D. Li, X. Wang, C. C. Chang, and Z. Wu, "A blockchain-based secure storage scheme for medical information," *Eurasip Journal on Wireless Communications and Networking*, vol. 2022, 12 2022.
- [13] Y. Yang, X. Zheng, W. Guo, X. Liu, and V. Chang, "Privacy-preserving smart iot-based healthcare big data storage and self-adaptive access control system," *Information Sciences*, vol. 479, pp. 567–592, 4 2019.
- [14] A. S. Shahraki, C. Rudolph, and M. Grobler, "A dynamic access control policy model for sharing of healthcare data in multiple domains," *Proceedings - 2019 18th IEEE International Conference on Trust, Security and Privacy in Computing and Communications/13th IEEE International Conference on Big Data Science and Engineering, TrustCom/BigDataSE 2019*, pp. 618–625, 8 2019.
- [15] Y. Ye, J. Shi, D. Zhu, L. Su, J. Huang, and Y. Huang, "Management of medical and health big data based on integrated learning-based health care system: A review and comparative analysis," *Computer Methods and Programs in Biomedicine*, vol. 209, 9 2021.
- [16] H. B. Mahajan, A. S. Rashid, A. A. Junnarkar, N. Uke, S. D. Deshpande, P. R. Futane, A. Alkhayat, and B. Alhayani, "Integration of healthcare 4.0 and blockchain into secure cloud-based electronic health records systems," *Applied Nanoscience (Switzerland)*, vol. 13, pp. 2329–2342, 3 2023.
- [17] R. Patan, S. Kallam, A. H. Gandomi, T. Hanne, and M. Ramachandran, "Gaussian relevance vector mapreduce-based annealed glowworm optimization for big medical data scheduling," *Journal of the Operational Research Society*, vol. 73, pp. 2204–2215, 2022.
- [18] S. Dhote, S. Baskar, P. M. Shakeel, and T. Dhote, "Cloud computing assisted mobile healthcare systems using distributed data analytic model," *IEEE Transactions on Big Data*, 2023.
- [19] W. Bovenizer and P. Chetthamrongchai, "A comprehensive systematic and bibliometric review of the iot-based healthcare systems," *Cluster Computing*, 2023.
- [20] P. Zhang and M. N. K. Boulos, "Privacy-by-design environments for large-scale health research and federated learning from data," *International Journal of Environmental Research and Public Health*, vol. 19, 10 2022.
- [21] I. Hussain and S. J. Park, "Big-ecg: Cardiographic predictive cyber-physical system for stroke management," *IEEE Access*, vol. 9, pp. 123 146–123 164, 2021.
- [22] P. Kaur, M. Sharma, and M. Mittal, "Big data and machine learning based secure healthcare framework," *Procedia Computer Science*, vol. 132, pp. 1049–1059, 2018, international Conference on Computational Intelligence and Data Science. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S187705091830752X>
- [23] E. MJ, J. Jobin, and J. Dona, "A fog based security model for electronic medical records in the cloud database 2553," 2021.
- [24] C. Xiao, F. Wang, T. Jia, L. Pan, and Z. Wang, "48-construction system and preliminary application of big data platform for liver disease and liver cancer," 2022.
- [25] C. Li, "Information processing in internet of things using big data analytics," *Computer Communications*, vol. 160, pp. 718–729, 7 2020.
- [26] S. Funde and G. Swain, "Big data privacy and security using abundant data recovery techniques and data obliviousness methodologies," *IEEE Access*, vol. 10, pp. 105 458–105 484, 2022.
- [27] M. S. Islam, M. A. B. Ameen, M. A. Rahman, H. Ajra, and Z. B. Ismail, "Healthcare-chain: Blockchain-enabled decentralized trustworthy system in healthcare management industry 4.0 with cyber safeguard," *Computers*, vol. 12, 2 2023.

- [28] J. Xu, L. Wei, W. Wu, A. Wang, Y. Zhang, and F. Zhou, "Privacy-preserving data integrity verification by using lightweight streaming authenticated data structures for healthcare cyber-physical system," *Future Generation Computer Systems*, vol. 108, pp. 1287–1296, 7 2020.
- [29] K. S. K. Maathavan and S. Venkatraman, "A secure encrypted classified electronic healthcare data for public cloud environment," *Intelligent Automation and Soft Computing*, vol. 32, pp. 765–779, 2022.
- [30] V. Kashyap, A. Kumar, A. Kumar, and Y. C. Hu, "A systematic survey on fog and iot driven healthcare: Open challenges and research issues," *Electronics (Switzerland)*, vol. 11, 9 2022.
- [31] P. Sarosh, S. A. Parah, G. M. Bhat, A. A. Heidari, and K. Muhammad, "Secret sharing-based personal health records management for the internet of health things," *Sustainable Cities and Society*, vol. 74, 11 2021.
- [32] P. S. Sen and N. Mukherjee, "An ontology-based approach to designing a nosql database for semi-structured and unstructured health data," *Cluster Computing*, 2023.
- [33] A. Bahmani, A. Alavi, T. Buergel, S. Upadhyayula, Q. Wang, S. K. Ananthkrishnan, A. Alavi, D. Celis, D. Gillespie, G. Young, Z. Xing, M. H. H. Nguyen, A. Haque, A. Mathur, J. Payne, G. Mazaheri, J. K. Li, P. Kotipalli, L. Liao, R. Bhasin, K. Cha, B. Rolnik, A. Celli, O. Dagan-Rosenfeld, E. Higgs, W. Zhou, C. L. Berry, K. G. V. Winkle, K. Contrepolis, U. Ray, K. Bettinger, S. Datta, X. Li, and M. P. Snyder, "A scalable, secure, and interoperable platform for deep data-driven health management," *Nature Communications*, vol. 12, 12 2021.
- [34] V. Stephanie, I. Khalil, M. Atiqzaman, and X. Yi, "Trustworthy privacy-preserving hierarchical ensemble and federated learning in healthcare 4.0 with blockchain," *IEEE Transactions on Industrial Informatics*, 2022.
- [35] M. K. Hassan, A. I. E. Desouky, S. M. Elghamrawy, and A. M. Sarhan, "A hybrid real-time remote monitoring framework with nb-woa algorithm for patients with chronic diseases," *Future Generation Computer Systems*, vol. 93, pp. 77–95, 4 2019.
- [36] H. N. K. P and A. Professor, "An authorization framework for preserving privacy of big medical data via blockchain in cloud server," *IJACSA International Journal of Advanced Computer Science and Applications*, vol. 13, p. 133, 2022. [Online]. Available: www.ijacsa.thesai.org
- [37] R. D. Thantilage, N.-A. Le-Khac, and M.-T. Kechadi, "Healthcare data security and privacy in data warehouse architectures," *Informatics in Medicine Unlocked*, vol. 39, p. 101270, 2023. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2352914823001144>
- [38] K. Abouelmehdi, A. Beni-Hssane, H. Khaloufi, and M. Saadi, "Big data security and privacy in healthcare: A review," *Procedia Computer Science*, vol. 113, pp. 73–80, 2017, the 8th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN 2017) / The 7th International Conference on Current and Future Trends of Information and Communication Technologies in Healthcare (ICTH-2017) / Affiliated Workshops. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1877050917317015>
- [39] International Organization for Standardization (ISO), International Electrotechnical Commission (IEC), *ISO/IEC 8825-1:2015 - Information technology — ASN.1 encoding rules: Specification of Basic Encoding Rules (BER), Canonical Encoding Rules (CER) and Distinguished Encoding Rules (DER)*, Std., 2015.
- [40] R. Housley, "Evolution of email security standards," *IEEE Communications Magazine*, vol. 60, pp. 6–9, 11 2022.
- [41] G. Schryen, "A quantitative and qualitative analysis of the PEM format for internet electronic mail security," *Journal of the Association for Information Systems*, vol. 19, no. 5, pp. 393–421, 2018.
- [42] V. K. Jha, G. K. Gupta, and M. Chhabra, "PEM: Privacy enhanced mail security analysis and evaluation," in *2019 International Conference on Communication and Signal Processing (ICCSP)*. IEEE, 2019, pp. 0452–0456.
- [43] S. Kundella and R. Gobinath, "A survey on big data analytics in medical and healthcare using cloud computing," *INTERNATIONAL JOURNAL OF SCIENTIFIC TECHNOLOGY RESEARCH*, vol. 8, 2019. [Online]. Available: www.ijstr.org
- [44] A. Damiani, C. Masciocchi, J. Lenkiewicz, N. D. Capocchiano, L. Boldrini, L. Tagliaferri, A. Cesario, P. Sergi, A. Marchetti, A. Luraschi, S. Patarnello, and V. Valentini, "Building an artificial intelligence laboratory based on real world data: The experience of gemelli generator," *Frontiers in Computer Science*, vol. 3, 12 2021.
- [45] S. Nazir, M. Nawaz, A. Adnan, S. Shahzad, and S. Asadi, "Big data features, applications, and analytics in cardiology - a systematic literature review," *IEEE Access*, vol. 7, pp. 143 742–143 771, 2019.
- [46] B. K. Sarkar, "Big data for secure healthcare system: A conceptual design - complex amp; intelligent systems," *SpringerLink*, Mar 2017. [Online]. Available: <https://link.springer.com/article/10.1007/s40747-017-0040-1>
- [47] B. J. Awrahman, C. Aziz Fatah, and M. Y. Hamaamin, "A review of the role and challenges of big data in healthcare informatics and analytics," *Computational Intelligence and Neuroscience*, vol. 2022, p. 1–10, 2022.
- [48] D. B. P, "Cloud based healthcare framework for criticality level analysis," *International Journal of Engineering and Advanced Technology (IJEAT)*, pp. 2249–8958, 2019. [Online]. Available: www.ijeat.org
- [49] F. A. Satti, T. Ali, J. Hussain, W. A. Khan, A. M. Khattak, and S. Lee, "Ubiquitous health profile (uhpr): a big data curation platform for supporting health data interoperability," *Computing*, vol. 102, pp. 2409–2444, 11 2020.
- [50] J. W. Oh, J. K. Kang, and J. Y. Park, "A study of smart healthcare service model based on cloud platform: Focus on small and medium sized hospitals," *Medico-Legal Update*, vol. 19, pp. 435–440, 7 2019.
- [51] G. Al-Sumaidae, A. Alexandridis, R. Alkhudary, and Z. Zilic, "A technical assessment of blockchain in healthcare with a focus on big data," *Proceedings - 2022 IEEE International Conference on Big Data, Big Data 2022*, pp. 2467–2472, 2022.
- [52] X. Chen, H. Zhu, D. Geng, W. Liu, R. Yang, and S. Li, "Merging rfid and blockchain technologies to accelerate big data medical research based on physiological signals," *Journal of Healthcare Engineering*, vol. 2020, 2020.
- [53] C. Li, M. Dong, J. Li, G. Xu, X. B. Chen, W. Liu, and K. Ota, "Efficient medical big data management with keyword-searchable encryption in healthchain," *IEEE Systems Journal*, vol. 16, pp. 5521–5532, 12 2022.
- [54] N. C. Taher, I. Mallat, N. Agoulmine, and N. El-Mawass, "An iot-cloud based solution for real-time and batch processing of big data: Application in healthcare," 2020.
- [55] R. Daraghmeah and R. Brown, "A big data maturity model for electronic health records in hospitals," *2021 International Conference on Information Technology, ICIT 2021 - Proceedings*, pp. 826–833, 7 2021.
- [56] K. E. Alhajaj and I. A. Moonesar, "The power of big data mining to improve the health care system in the united arab emirates," *Journal of Big Data*, vol. 10, 2 2023.
- [57] L. Nayak and V. Jayalakshmi, "A study of securing healthcare big data using dna encoding based ecc," *Proceedings of the 6th International Conference on Inventive Computation Technologies, ICICT 2021*, pp. 348–352, 1 2021.
- [58] A. Karunamurthy and M. Aramudhan, "Modeling of optimized data processing framework for potential knowledge discovery and recommendation based on healthcare big data," pp. 134–140, 2022. [Online]. Available: www.ijitee.org
- [59] A. Alourani, K. Tariq, M. Tahir, and M. Sardaraz, "Patient mortality prediction and analysis of health cloud data using a deep neural network," *Applied Sciences (Switzerland)*, vol. 13, 2 2023.
- [60] F. López-Martínez, E. R. Núñez-Valdez, V. García-Díaz, and Z. Bursac, "A case study for a big data and machine learning platform to improve medical decision support in population health management," *Algorithms*, vol. 13, 4 2020.
- [61] S. Nazir, S. Khan, H. U. Khan, S. Ali, I. Garcia-Magarino, R. B. Atan, and M. Nawaz, "A comprehensive analysis of healthcare big data management, analytics and scientific programming," *IEEE Access*, vol. 8, pp. 95 714–95 733, 2020.
- [62] U. Demirbaga and G. S. Auja, "Mapchain: A blockchain-based verifiable healthcare service management in iot-based big data ecosystem," *IEEE Transactions on Network and Service Management*, vol. 19, pp. 3896–3907, 12 2022.
- [63] S. Khan, H. U. Khan, and S. Nazir, "Systematic analysis of healthcare big data analytics for efficient care and disease diagnosing," *Scientific articles*, vol. 12, 12 2022.
- [64] G. Dhiman, S. Juneja, H. Mohafez, I. El-Bayoumy, L. K. Sharma, M. Hadizadeh, M. A. Islam, W. Viriyasitavat, and M. U. Khandaker, "Federated learning approach to protect healthcare data over big data scenario," *Sustainability (Switzerland)*, vol. 14, 3 2022.

- [65] H. Alshammari, S. A. El-Ghany, and A. Shehab, "Big iot healthcare data analytics framework based on fog and cloud computing," *Journal of Information Processing Systems*, vol. 16, pp. 1238–1249, 2020.
- [66] G. Dhanalakshmi, S. Mary, R. Asha, R. Dharani, V. Rajeswari, and M. Gobinath, "Securing an e-health care information systems on cloud environments with big data approach design engineering securing an e-health care information systems on cloud environments with big data approach," 2020. [Online]. Available: <https://www.researchgate.net/publication/354892101>
- [67] A. Cuzzocrea, G. M. Grasso, and M. Nolich, "Effective and efficient distributed management of big clinical data: a framework," *Int. J. Data Mining, Modelling and Management*, vol. 11, pp. 284–313, 2019.
- [68] M. Patel, V. K. Prasad, P. Bhattacharya, M. Bhavsar, and M. Zuhair, "Privacy preservation for big data healthcare management," *Proceedings of 3rd International Conference on Intelligent Engineering and Management, ICIEM 2022*, pp. 211–216, 2022.
- [69] H. Chen, Z. Song, and F. Yang, "Storage method for medical and health big data based on distributed sensor network," *Journal of Sensors*, vol. 2023, 2023.
- [70] A. Bahmani, A. Alavi, T. Buerger, S. Upadhyayula, Q. Wang, S. K. Ananthkrishnan, A. Alavi, D. Celis, D. Gillespie, G. Young, and et al., "A scalable, secure, and interoperable platform for deep data-driven health management," *Nature Communications*, vol. 12, no. 1, 2021.
- [71] S. Haque, Z. Eberhart, A. Bansal, and C. McMillan, "Semantic similarity metrics for evaluating source code summarization," *IEEE International Conference on Program Comprehension*, vol. 2022-March, pp. 36–47, 2022.
- [72] S. Benila and N. U. Bhanu, "Service level agreement based secured data analytics framework for healthcare systems," *Intelligent Automation and Soft Computing*, vol. 32, pp. 1277–1291, 2022.
- [73] A. Mehbodniya, R. Neware, S. Vyas, M. R. Kumar, P. Ngulube, and S. Ray, "Blockchain and ipfs integrated framework in bilevel fog-cloud network for security and privacy of iomt devices," *Computational and Mathematical Methods in Medicine*, vol. 2021, 2021.
- [74] I. of Electrical, E. Engineers, M. R. I. I. of Research, and Studies, "Proceedings of the international conference on machine learning, big data, cloud and parallel computing : trends, prespectives and prospects : Comitcon-2019 : 14th-16th february, 2019," 2019.
- [75] A. Rehman, K. Haseeb, T. Saba, J. Lloret, and U. Tariq, "Secured big data analytics for decision-oriented medical system using internet of things," *Electronics (Switzerland)*, vol. 10, 6 2021.
- [76] Y. Wang, S. Cheng, X. Zhang, J. Leng, and J. Liu, "Block storage optimization and parallel data processing and analysis of product big data based on the hadoop platform," *Mathematical Problems in Engineering*, vol. 2021, p. 1–14, 2021.
- [77] A. S. Rajawat, S. Goyal, P. Bedi, A. Shrivastava, N. B. Constantin, M. S. Raboaca, and C. Verma, "Security analysis for threats to patient data in the medical internet of things," pp. 248–253, 2022.
- [78] A. Tahir, F. Chen, H. U. Khan, Z. Ming, A. Ahmad, S. Nazir, and M. Shafiq, "A systematic review on cloud storage mechanisms concerning e-healthcare systems," *Sensors (Switzerland)*, vol. 20, pp. 1–32, 9 2020.
- [79] H. Harb, H. Mroue, A. Mansour, A. Nasser, and E. M. Cruz, "A hadoop-based platform for patient classification and disease diagnosis in healthcare applications," *Sensors (Switzerland)*, vol. 20, 4 2020.
- [80] R. F. Famutimi, M. O. Oyelami, A. O. Ibitoye, and O. M. Awoniran, "An empirical comparison of the performances of single structure columnar in-memory and disk-resident data storage techniques using healthcare big data," *Journal of Big Data*, vol. 10, no. 1, 2023.
- [81] M. Aldwairi, M. Jarrah, N. Mahasneh, and B. Al-khateeb, "Graph-based data management system for efficient information storage, retrieval and processing," *Information Processing and Management*, vol. 60, 3 2023.
- [82] R. Famutimi, M. Oyelami, A. Ibitoye, and O. Awoniran, "An empirical comparison of the performances of single structure columnar in-memory and disk resident data storage techniques using healthcare big data," 2023. [Online]. Available: <https://doi.org/10.1186/s40537-023-00691-x>
- [83] O. Jinadu, O. Johnson, and M. Ganiyu, "Distributed database system optimization for improved service delivery in mobile and cloud bigdata applications," *International Journal of Computer Science and Mobile Computing*, vol. 10, no. 9, p. 38–45, 2021.
- [84] M. Babar, M. U. Tariq, M. D. Alshehri, U. Fasee, and M. I. Uddin, "Smart teledentistry healthcare architecture for medical big data analysis using iot-enabled environment," 2022.
- [85] T. Javid, M. Faris, H. Beenish, and M. Fahad, "Cybersecurity and data privacy in the cloudlet for preliminary healthcare big data analytics," pp. 7–10. [Online]. Available: <http://ushistoryscene.com/article/second-industrial-revolution>
- [86] —, "Cybersecurity and data privacy in the cloudlet for preliminary healthcare big data analytics," pp. 1–4, 2020.
- [87] A. S. Rajawat, S. B. Goyal, P. Bedi, A. Shrivastava, N. B. Constantin, M. S. Raboaca, and C. Verma, "Security analysis for threats to patient data in the medical internet of things," *Proceedings of the 2022 11th International Conference on System Modeling and Advancement in Research Trends, SMART 2022*, pp. 248–253, 2022.
- [88] M. A. Chamikara, P. Bertok, D. Liu, S. Camtepe, and I. Khalil, "Efficient privacy preservation of big data for accurate data mining," *Information Sciences*, vol. 527, pp. 420–443, 7 2020.
- [89] H.-Y. Tran and J. Hu, "Privacy-preserving big data analytics a comprehensive survey," *Journal of Parallel and Distributed Computing*, vol. 134, pp. 207–218, 2019. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0743731519300589>
- [90] N. A. Azeez and C. V. der Vyver, "Security and privacy issues in e-health cloud-based system: A comprehensive content analysis," *Egyptian Informatics Journal*, vol. 20, no. 2, pp. 97–108, 2019. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1110866517302797>
- [91] S. Tanwar, J. Vora, S. Tyagi, N. Kumar, and M. Obaidat, "Ensuring privacy and security in e- health records," 08 2018.
- [92] P. Kaur, "Big data analytics in healthcare: A review," *International Journal of Engineering and Technical Research*, vol. 10, 07 2021.
- [93] K. Batko and A. Ślezak, "The use of big data analytics in healthcare," *Journal of Big Data*, vol. 9, no. 1, 2022.
- [94] S. Galea and S. Abdalla, "Covid-19 pandemic, unemployment, and civil unrest: Underlying deep racial and socioeconomic divides," *JAMA*, vol. 324, 06 2020.
- [95] N. Ngesimani, E. Ruhode, and P.-A. Harpur, "Data governance in healthcare information systems: A systematic literature review," *SA Journal of Information Management*, vol. 24, 08 2022.
- [96] Y. Aytas, *Big Data Storage*, 07 2021, pp. 41–61.
- [97] A. Q. Khan, N. Nikolov, M. Matskin, R. Prodan, H. Song, D. Roman, and A. Soylu, "Smart data placement for big data pipelines: An approach based on the storage-as-a-service model," in *2022 IEEE/ACM 15th International Conference on Utility and Cloud Computing (UCC)*, 2022, pp. 317–320.
- [98] H. Zaydi and Z. Bakkoury, "Towards a global and abstract end-to-end architecture for data analysis and transformation with ml/dl application cases: Medical iot and ioht," in *2021 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*, 2021, pp. 1–6.
- [99] H. Huang, P. Zhu, F. Xiao, X. Sun, and Q. Huang, "A blockchain-based scheme for privacy-preserving and secure sharing of medical data," *Computers Security*, vol. 99, p. 102010, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0167404820302832>
- [100] A. Pesqueira, M. Sousa, and S. Bolog, *Implementation of Big Data and Blockchain for Health Data Management in Patient Health Records*, 05 2023, pp. 17–26.
- [101] R. Raja, I. Mukherjee, and B. K. Sarkar, "A systematic review of healthcare big data," *Scientific Programming*, vol. 2020, p. 1–15, 2020.