UNIVERSIDAD CÉSAR VALLEJO

ESCUELA DE POSGRADO

PROGRAMA ACADÉMICO DE MAESTRÍA EN INGENIERÍA DE SISTEMAS CON MENCIÓN EN TECNOLOGÍAS DE LA INFORMACIÓN

Aplicación de Algoritmos Criptográficos en la Gestión de Almacenamiento de Datos en una Empresa de Seguridad Electrónica, Lima 2023

TESIS PARA OBTENER EL GRADO ACADÉMICO DE:
MAESTRO EN INGENIERÍA DE SISTEMAS CON MENCIÓN EN TECNOLOGÍAS DE LA INFORMACIÓN

AUTOR:

Torres Paredes, Carlos Martin (orcid.org/0000-0002-7464-5392)

ASESOR:

Dr. Acuña Benites, Marlon Frank (orcid.org/0000- 0001-5207-9353)

CO-ASESOR:

Dr. Pereyra Acosta, Manuel Antonio (orcid.org/0000-0002-2593-5772)

LÍNEA DE INVESTIGACIÓN:

Sistemas de Información y Comunicaciones

LÍNEA DE RESPONSABILIDAD SOCIAL UNIVERSITARIA:

Desarrollo económico, empleo y emprendimiento

LIMA – PERÚ

2023

Dedicatoria

A mis padres, Olga y Carlos, que con su ejemplo de vida y continuas enseñanzas me inculcaron el valor de la educación y el trabajo, como único camino para superarnos y ser hombres de bien y de provecho para la sociedad.

Agradecimiento

A los docentes del programa de Maestría en Ingeniería de Sistemas de la Universidad César Vallejo, por la generosa entrega de sus conocimientos y experiencia para mejorar nuestra formación académica.

A nuestro asesor, el Doctor Marlon Acuña, por su infinita paciencia e indesmayable tesón para animarnos, orientarnos y guiarnos en la conclusión de nuestros proyectos de investigación.

Índice de contenido

Pág.

Índice de tablas

Pág.

Índice de figuras

Pág.

Diccionario de palabras

Algoritmo criptográfico

Secuencia de acciones aplicada a un grupo de datos a fin de transformar su presentación. Pueden ser orientados a la encriptación, donde los datos se convierten en un patrón distinto, difícil de asociar con los datos originales, u orientados al resumen (hash), donde de los datos se obtiene una huella digital característica.

Ancho de banda

Capacidad de transferencia de datos por unidad de tiempo de un sistema informático. Dependiendo de la velocidad a la que opere el sistema, los datos se transferirán en mayor o menor tiempo.

Deduplicación de datos

Técnica de identificación y remoción de datos duplicados, los cuales son reemplazados por un enlace lógico a una copia única, que se aplica en un sistema de almacenamiento con la finalidad de mejorar las características de gestión y operación del sistema.

Huella digital (fingerprint)

Secuencia de bits de longitud finita que se asocia a un conjunto específico de datos, independientemente de su tamaño. Se considera que es única y permite identificar la presencia del conjunto de datos independientemente de que se encuentren almacenados bajo otro nombre o mezclados con otros datos. Se le calcula utilizando un algoritmo criptográfico tipo Hash (Por ejemplo, el algoritmo MD5 genera secuencias de 128 bits).

Sistema de Almacenamiento de datos

Tipo particular de sistema informático orientado a guardar de forma de forma estructurada y confiable los datos que emplean los sistemas de información de una organización.

Resumen

La presente investigación tuvo el objeto de determinar el impacto de la aplicación de algoritmos criptográficos en la gestión del almacenamiento de datos de una empresa de seguridad electrónica de Lima. Metodológicamente la investigación tuvo un enfoque cuantitativo, con diseño pre-experimental, temporalmente transversal, de carácter aplicado y tipo correlacional. El instrumento empleado fue la ficha de registro, que se aplicó a treinta objetos de almacenamiento típicos empleados por la empresa donde se desarrolló el estudio, registrándose los valores antes y después de la aplicación de los algoritmos criptográficos. El análisis preliminar de los datos mostró, mediante la prueba de Shapiro-Wilk, que no se ajustaban al supuesto de normalidad, por lo que se aplicó la prueba de rangos con signo de Wilcoxon para muestras relacionadas. En todos los casos, vale decir hipótesis formuladas, se encontró diferencias estadísticamente significativas con valor de 0.000, menores al 0.05 tomado como criterio de aceptación, lo que nos llevó a rechazar la hipótesis nula y aceptar la hipótesis alterna, demostrándose que la aplicación de los algoritmos criptográficos tiene impacto positivo en la gestión del almacenamiento de datos.

Palabras claves: algoritmos criptográficos, almacenamiento de datos, funciones hash, deduplicación de datos.

Abstract

This research had the objective to determine the impact of the application of cryptographic algorithms in the data storage management of an electronic security company in Lima. Methodologically, the research had a quantitative approach, with a pre-experimental design, temporally was transversal, with applied nature and correlational type. The instrument used was the registration form, which was applied to thirty typical storage objects used by the company where the study was carried out, recording the values before and after the application of the cryptographic algorithms. Preliminary analysis of the data showed, by means of the Shapiro Wilk test, that they did not conform to the normality assumption, so the Wilcoxon signed-rank test for related samples was applied. In all cases, that is, formulated hypotheses, statistically significant differences were found with a value of 0.000, less than the 0.05 taken as acceptance criterion, which led us to reject the null hypothesis and accept the alternative hypothesis, demonstrating that the application of cryptographic algorithms has a positive impact on data storage management.

Keywords: cryptographic algorithms, data storage, hash, deduplication.

I. INTRODUCCIÓN

Vivimos un crecimiento cada vez mayor de datos digitales en los sistemas de información de las organizaciones, que se alojan en sistemas de almacenamiento de datos con variedad de funcionalidades y que poseen características operativas distintivas como son la velocidad de respuesta y la capacidad de almacenamiento.

Estos sistemas de almacenamiento guardan datos de varios usuarios, y muchas veces múltiples copias de los mismos datos, lo que reduce su eficiencia e incrementa los costos de operación. Más aun, las empresas almacenan datos de forma redundante, al menos dos copias, para asegurar la disponibilidad en caso de incidentes. Esta duplicación múltiple de datos deviene en una carga financiera pesada, que debe aliviarse de alguna forma. Para ello se emplean técnicas de deduplicación de datos que buscan reducir los costos de almacenamiento y mejorar su eficiencia, Ur Rehman et al. (2016).

Es posible aplicar técnicas de deduplicación de datos para identificar datos repetidos y reducirlos a una sola copia, lo que eleva la eficiencia de operación del sistema de almacenamiento. Un caso importante de aplicación son los ambientes virtualizados, donde se trabaja con instancias de máquinas virtuales, que comparten muchos datos, por ejemplo, a nivel de sistema operativo, de aplicaciones, etc. En esta situación se puede aplicar las técnicas de deduplicación para reducir el consumo de recursos de almacenamiento por parte de las máquinas virtuales que hayan sido aprovisionadas, Saharan (2020).

Se puede llevar los beneficios de la deduplicación de datos a múltiples campos de actividad humana. En el caso de la forensia digital, donde se analizan imágenes de dispositivos de almacenamiento, el volumen de casos, el volumen de datos, la falta de personal, etc., dificultan el proceso. Se ha obtenido reducciones importantes en el consumo de recursos mediante la aplicación de la deduplicación de datos de forma previa al proceso de forensia digital, Scanlon (2016).

Otro caso de uso es el almacenamiento en servicios de nube, donde abundan los datos de carácter multimedia, con alto grado de duplicidad, por lo que

es mandatorio aplicar tecnologías de deduplicación para alcanzar niveles adecuados de costo de almacenamiento y eficiencia de uso, He et al. (2020).

En la tesis doctoral de Ni (2019), se estableció la necesidad de contar con mecanismos eficientes de reducción de datos, y se exploró la deduplicación, identificando como factores claves la velocidad de procesamiento para calcular las huellas digitales de los datos, así como las velocidades de transferencia de los dispositivos de entrada y salida de datos, es decir unidades de almacenamiento. Para superar estas limitantes propuso un algoritmo paralelo de cálculo de huellas, con lo que logró mejoras significativas en la eficiencia de entrada/salida de datos.

En su tesis de maestría Liao (2022), repasó la importancia que los datasets han alcanzado en el desarrollo de modelos de aprendizaje profundo, pero resaltó que es sumamente importante velar por la calidad de estos datasets. Para tal fin propuso llevar a cabo la deduplicación de los datos que componen los datasets para evitar que datos duplicados impacten negativamente en los procesos de entrenamiento de modelos de aprendizaje profundo.

También en su tesis de maestría Qiu (2021), extendió la aplicación de la deduplicación de datos a la optimización de la computación sin servidores, donde los modelos son ejecutados en contenedores ligeros, pero de forma masivamente paralela, esta idea no es explotada por los proveedores de servicios de nube. Propuso un modelo de compartición de memoria para todos estos contenedores, donde, a fin de optimizar el empleo de recursos, se aplica la deduplicación a los objetos de datos residentes en memoria. Los resultados típicos de la propuesta muestran reducciones del orden del 20% del uso de memoria.

De la realidad problemática presentada se desprende el problema general que aborda la investigación: ¿Cómo impacta el uso de algoritmos criptográficos en la gestión de almacenamiento de datos de una empresa de seguridad electrónica? A su vez, podemos identificar los siguientes problemas específicos. Primer problema específico: ¿Cómo impacta el uso de algoritmos criptográficos en datos con alta similitud en la gestión de almacenamiento de datos de una empresa de seguridad electrónica? El segundo problema específico es: ¿Cómo impacta el uso algoritmos criptográficos a nivel de bloques en la gestión de la capacidad de

almacenamiento de datos en una empresa de seguridad electrónica? Como tercer problema específico tenemos: ¿Cómo impacta el uso de algoritmos criptográficos a nivel de bloques en la gestión de ancho de banda de almacenamiento de datos en una empresa de seguridad electrónica? El cuarto problema específico es: ¿Cómo impacta el uso de algoritmos criptográficos a nivel de bloques en la gestión del costo de almacenamiento de datos de una empresa de seguridad electrónica?

La realización del proyecto de investigación se basa en las siguientes justificaciones, desde el punto de las implicaciones prácticas, la tesis se justifica ya que permite abordar y resolver el problema de la ineficiente gestión de almacenamiento de datos debido a la duplicidad que se presenta en los datos almacenados. De forma específica, la tesis emplea como caso de estudio la gestión de almacenamiento de datos compuestos mayoritariamente por imágenes de cámaras de vigilancia electrónica, pero la solución propuesta puede ser empleada en otros escenarios donde se gestione el almacenamiento de datos de otras fuentes.

En cuanto a las implicaciones teóricas, la tesis encuentra justificación pues el tema de gestión eficiente del almacenamiento de datos, mediante la reducción de la duplicidad, ha sido mínimamente estudiado, y prácticamente no se ha aplicado en el caso de almacenamiento de imágenes. La propuesta que presenta la tesis, basada en algoritmos criptográficos, aplicados mediante técnicas de deduplicación, está orientada a incrementar la eficiencia, lo que no ha sido investigado adecuadamente, ya que en el enfoque convencional, estos algoritmos se emplean para protección de los datos almacenados, es decir estrictamente para su encriptación (ocultamiento), mientras que en el enfoque que propone la tesis, los algoritmos criptográficos que se emplean corresponden a funciones Hash, es decir se trabaja con la identificación de bloques de datos duplicados y su remoción. Los aportes mencionados posibilitarán trabajos futuros que profundicen en el tema.

Por el lado de las implicaciones metodológicas encontramos justificación adicional de la tesis ya que, para evaluar la eficiencia de la gestión del almacenamiento de datos, se definirán indicadores relevantes y se propondrá instrumentos particulares para realizar y registrar las mediciones correspondientes, como son, por ejemplo, las fichas de registro del nivel de reducción de los datos

almacenados. Los elementos mencionados permitirán el mejor estudio de una población compuesta por repositorios de almacenamiento de datos, donde el objeto de almacenamiento sean imágenes provenientes de cámaras de seguridad.

Una justificación adicional es la económica, que contempla que las organizaciones asignan cada vez más recursos para emplear tecnologías de la información (TI) para el soporte de sus operaciones de negocio, siendo el procesamiento, la conectividad y el almacenamiento de datos aspectos claves y, como parte importante de las TI, el almacenamiento de datos representa un reto constante por los requerimientos que impone (volumen de almacenamiento, tiempo de transferencia, costo del almacenamiento), los cuales escalan diariamente, por lo que cualquier iniciativa conducente a racionalizar su empleo y gestionar eficientemente los volúmenes de almacenamiento, es de alto impacto en la organización

Con respecto al objetivo general de la investigación, se ha planteado lo siguiente: “Determinar el impacto que tiene el uso de algoritmos criptográficos en la gestión del almacenamiento de datos en una empresa de seguridad electrónica”. Para alcanzar el objetivo general se han considerado los siguientes objetivos específicos. Primer objetivo específico: “Determinar el impacto que tiene el uso de algoritmos criptográficos en datos con alta similitud en la gestión de almacenamiento de datos en una empresa de seguridad electrónica”. El segundo objetivo específico es “Determinar el impacto que tiene el uso de algoritmos criptográficos a nivel de bloques en la gestión de la capacidad del almacenamiento de datos en una empresa de seguridad electrónica”. Como tercer objetivo específico tenemos: “Determinar el impacto que tiene el uso de algoritmos criptográficos a nivel de bloques en la gestión de ancho de banda de almacenamiento de datos de una empresa de seguridad electrónica”. El cuarto objetivo específico planteado es: “Determinar el impacto que tiene el uso de algoritmos criptográficos a nivel de bloques en la gestión de costo de almacenamiento de datos en una empresa de seguridad electrónica”.

En cuanto a la hipótesis general, esta se ha redactado en los siguientes términos: “La utilización de algoritmos de criptográficos impacta positivamente en la gestión de almacenamiento de datos de una empresa de seguridad electrónica”.

En base a la hipótesis general hemos podido plantar las siguientes hipótesis específicas, La primera hipótesis específica es: “La utilización de los algoritmos criptográficos en datos con alta similitud impacta positivamente en la gestión de almacenamiento de datos en una empresa de seguridad electrónica”. La segunda hipótesis específica es: “La utilización de los algoritmos criptográficos a nivel de bloques impacta positivamente en la gestión de la capacidad de almacenamiento de datos en una empresa de seguridad electrónica”. La tercera hipótesis específica es: “La utilización de los algoritmos criptográficos a nivel de bloques impacta positivamente en la gestión del ancho de banda de almacenamiento de datos de una empresa de seguridad electrónica”. La cuarta hipótesis específica es: La utilización de los algoritmos criptográficos a nivel de bloques impacta positivamente en la gestión de costo de almacenamiento de datos de una empresa de seguridad electrónica”.

II. MARCO TEÓRICO

Con la finalidad de presentar un contexto adecuado de antecedentes de la presente investigación, se ha revisado la literatura especializada y se ha seleccionado los antecedentes de mayor relevancia. Comenzaremos presentando los antecedentes nacionales para luego pasar a los internacionales.

En el trabajo de Morey et al. (2016), se desarrolló una investigación de tipo no experimental para evaluar el impacto de aplicas técnicas de deduplicación de datos. La población de estudio estuvo compuesta por máquinas virtuales que corrían sistemas operativos open source, Linux, y que estaban alojadas en servidores físicos. La única diferencia entre las máquinas virtuales era el sistema de archivos, habiéndose empleado SDFS y ZFS. Se evaluó el uso de CPU, obteniéndose 24.67% de ocupación para SDFS y 23.59% para ZFS que sería la mejor alternativa bajo ese criterio. También se evaluó el empleo de memoria RAM como consecuencia de los procesos de deduplicación, obteniéndose 96.85% para SDFS y 94.99% para ZFS que nuevamente se presenta como una mejor opción. Otro indicador importante que se midió en esta investigación fue el tiempo de descarga de un archivo que para SDFS fue de 8 minutos mientras que para ZFS solo fue de 5 minutos nuevamente a favor de ZFS. Finalmente se midió el uso de la capacidad de transferencia de datos, uso de red, obteniéndose 48.37% para SDFS y 29.45% para ZFS.

En la tesis doctoral de Larico (2020), se llevó a cabo una investigación aplicada de enfoque cuantitativo. La población estuvo constituida por el personal que labora en la dirección universitaria de sistemas, y se buscó evaluar el impacto de emplear arquitecturas hiperconvergentes para mejorar los parámetros de gestión de los recursos de TI. Los resultados muestran que el empleo de la arquitectura hiperconvergente tiene alto impacto en mejorar la gestión de los recursos de TI, y que el almacenamiento centralizado mejora la eficiencia y los parámetros de operación de dicho sistema de almacenamiento. Esta hipótesis fue aceptada con el valor de significancia (valor critico observado) de $0.003 < 0.05$.

En su tesis doctoral, Villa (2018), lleva a cabo una investigación de tipo no experimental, donde propone compara dos modelos de procesamiento y

almacenamiento de datos masivos en entornos de salud, uno basado en un servidor centralizado y otro en un clúster. La población fueron los datasets originales del Hospital de Texas Health Care Information Council, tomados para dos años consecutivos. Las pruebas incluyeron diversos tipos de acceso, encontrándose diferencias significativas en los indicadores. El tiempo de acceso promedio para las 5 consultas más frecuentes a la base de datos, bajo condiciones de carga intensa, fue de 6.29 segundos para el clúster y 14.489 segundos para el servidor central. Cuando se evaluó el impacto de estas diferencias, se obtuvo que el clúster permitía atender mayor cantidad de pacientes en comparación al servidor central.

En la tesis de maestría de Roldán et al. (2020), se lleva a cabo una investigación de carácter cualitativo, cuyo propósito fue evaluar el Sistema Informático de Denuncias Policiales para identificar los puntos de dolor del sistema y proponer un modelo mejorado para que la Policía Nacional del Perú pueda contar con datos confiables y seguros para trabajar estadísticas, efectuar seguimiento y tomar decisiones estratégicas. Dentro de las conclusiones, de carácter cualitativo por el tipo de investigación, se identificó el sistema de procesamiento y almacenamiento como inadecuado, con múltiples carencias y falencias, recomendándose la implementación de un sistema de procesamiento y almacenamiento de datos basado en clústers, con niveles de redundancia, con software de gestión de tanto del procesamiento como del almacenamiento de datos, así como el monitoreo y continuo para detectar oportunidades de mejora y optimización de la plataforma a lo largo de su vida útil.

En su tesis doctoral Ocsa (2018), desarrolla una investigación de carácter cualitativo con el propósito de proponer métodos de acceso eficientes y escalables a grandes volúmenes de datos. Como parte de sus conclusiones, se establece que la identificación rápida de un dato puede efectuarse con base en los algoritmos criptográficos Hash. La búsqueda puede ser por identificación exacta o por similitud. En la investigación se propone el algoritmo Deep frActal based Hashing (DAsH), que utiliza técnicas fractales para detectar contenidos específicos de interés con rapidez y precisión.

En el paper de Venish et al. (2016), se describe una investigación de tipo no experimental, donde se presenta la deduplicación de datos como una tecnología

emergente que reduce la utilización de los sistemas de almacenamiento y que es muy eficiente frente a la replicación de datos en backups. En esta línea, otorgan mucha importancia al algoritmo de fragmentación que permite dividir los objetos de datos en fragmentos que pueden ser identificados de forma única con una función criptográfica hash, con la finalidad de identificar y eliminar fragmentos redundantes. Por ello, llevan a cabo un ejercicio de rendimiento donde la población son cuatro algoritmos de deduplicación y se efectuó el análisis comparativo entre los algoritmos de fragmentación a nivel de archivo, a nivel de bloque, basados en contenido, de ventana deslizante y de dos niveles con dos denominadores (TTTD). Para evaluar los algoritmos, emplean la relación entre el objeto de datos comprimido y el objeto de datos original, obteniéndose los siguientes resultados: para bloques de 64 KB el algoritmo a nivel de archivo alcanza el 50% y a nivel de bloque llega al 35%, para bloques de 32 KB se alcanza 49% y 33% aproximadamente, para bloques de 16 KB se logra 48% y 32%, y para bloques de 8 KB se obtiene 47% y 31%. En conclusión, el resultado de sus pruebas muestra que la fragmentación a nivel de bloque es superior a otros esquemas y que, si bien se obtiene diferentes niveles de compresión dependiendo del tamaño de bloque elegido, los resultados son consistentes.

De acuerdo con Aronovich et al. (2016), es posible ampliar los beneficios que ofrece la gestión optimizada de almacenamiento de datos aplicando deduplicación a nivel de bloques de archivos, para lo que propone emplear una función de similaridad en contraste a una función de identidad donde emplea la función criptográfica SHA1. Para evaluar los resultados define una relación de compresión, resultante de dividir el tamaño original del archivo entre el archivo comprimido. Como caso de estudio trabaja con máquinas virtuales donde se ha instalado distribuciones de sistemas operativos Linux, obteniéndose los resultados mostrados en la tabla 1.

Tabla 1

Comparación de las relaciones de compresión logradas por deduplicación basada en identidad y basadas en similaridad

Nombre de archivo	Tamaño (MB)	Relación de Compresión Identidad	Relación de Compresión Similaridad	Ganancia por Similaridad
Centos-5.3-i386-server	2816	6.58	7.10	7.3%
Freebsd-6.4-i386	949	3.88	4.02	3.5%
Fedora-fc6-i386	265	5.84	6.20	5.8%

Nota: tomado de Aronovich et al. (2016)

Si bien los resultados muestran una ventaja en la aplicación de la deduplicación basada en similaridad, se precisa que los resultados son dependientes del tipo de datos y que la complejidad del software aumenta.

En la tesis de maestría de Pérez (2018), se analiza la problemática de la gestión de almacenamiento de datos distribuida, concluyéndose que este tipo de almacenamiento conduce a duplicidad pues los servidores de almacenamiento no cuentan con medios de comunicación que permitan adoptar medidas. En este sentido, se propone una solución centralizada para almacenamiento de datos, basada en una modelo de almacenamiento en red, NAS, y donde también se integran diversas funcionalidades. A fin de evaluar la idoneidad del modelo se utiliza una encuesta para definir el grado de satisfacción de los administradores de sistemas con respecto a la solución planteada. La pregunta específica que se formuló a un grupo de diez expertos fue “La solución de almacenamiento de datos informáticos como mecanismo para disminuir la duplicidad de la información la valoro como...”, obteniéndose los resultados recogidos en la tabla 2.

Tabla 2

Valoración de expertos sobre la conveniencia de que la solución de gestión de almacenamiento reduzca la duplicidad

Valoración	Número de opiniones de expertos
Muy Adecuado (MA)	8
Bastante Adecuado (BA)	1
Adecuado (A)	1
Poco Adecuado (PA)	0
Inadecuado (I)	0

Nota: tomado de Pérez (2018)

De acuerdo a Zhou et al. (2018), los sistemas de archivos utilizados en los sistemas operativos comerciales se pueden beneficiar notablemente con la aplicación de algoritmos de deduplicación. Esta investigación de tipo experimental se llevó a cabo con el Ubuntu kernel 3.16.0, un disco de 500 GB, 10 GB de memoria RAM y un procesador Core i3-2100. El tamaño de bloque de fragmentación se tomó en 4 KB, efectuándose pruebas con diferentes cargas de trabajo (10,000 , 50,000 , 100,000 y 500,000 archivos). Se tomó el sistema de archivos nativo de Ubuntu, el Ext4, como referente, aplicándose los algoritmos de deduplicación LessFS y LDFS, alcanzándose una relación de compresión de 40.9 y 40.8 respectivamente. Un resultado colateral obtenido es que, aunque se aumenta metadata correspondiente a los algoritmos, este incremento alcanza niveles bajos, de 71.8 MB y 72 MB para los algoritmos mencionados.

Según Diao et al. (2016), en una investigación de corte experimental, donde se tomó como población un dataset conformado por captura de descargas de Youtube, estableció que es posible reducir la cantidad de información redundante que circula por las redes. Esta reducción de tráfico no causaría pérdida de datos y más bien incrementaría la capacidad de transmisión disponibles (ancho debanda). A finde poder hacer los cálculos en tiempo real y hacer frente al tráfico masivo de los enlaces, la implementación se desarrolló empleando tarjetas gráficas, GPUs,

para acelerar los cálculos. Los resultados obtenidos muestran que la implementación permitió incrementar la tasa de transferencia en 14 veces comparativamente con la implementación de un algoritmo de deduplicación a nivel de software.

En la investigación de Xu et al. (2017), se abordó la gestión del almacenamiento de datos considerando no solo las unidades de almacenamiento, sino que se incluyó el tiempo de respuesta en acceder a los datos almacenados y el costo de llevar a cabo dicho almacenamiento. La investigación fue de tipo no experimental, pues se orientó a medir los indicadores asociados a los criterios mencionados. La población sobre la cual se trabajó fue un grupo de 1000 datasets, con tamaños entre 10 y 100 GB, un costo de procesamiento de $0.1, y de almacenamiento de $0.15 por GB por mes. Los resultados alcanzados muestran que la mejor estrategia de almacenamiento, conducente a mejores indicadores es la de almacenar parcialmente los datasets en un repositorio de acceso remoto y mantener, de forma local los restantes datasets. Esta estrategia puede generalizarse al almacenamiento de otros tipos de datos.

La investigación de Fehér et al. (2020), explora la aplicación de las técnicas de deduplicación estándares para lograr la compresión de los datos que eran leídos en medidores inteligentes de energía eléctrica que contaban con conexión a Internet y realizaban lecturas periódicas. En este contexto, el término compresión hace referencia a la reducción del volumen de datos transmitidos, mediante el empleo de la deduplicación. El autor reporta haber alcanzado un ratio máximo de reducción de la capacidad de valor 10, lo cual es un valor bastante alto de deduplicación que parece deberse a la alta similitud existente entre grupos de lecturas de datos de los medidores.

En el trabajo de Cheng et al. (2021), se abordó la problemática que presenta el empleo de Big Data y el retardo que supone el almacenamiento de los datos en servicios de nube. Para hacer frente a este problema se recurre al almacenamiento de datos en el borde de la red, y para mejorar la gestión de dicho almacenamiento, se aplican técnicas de deduplicación de datos, proponiendo una estrategia de almacenamiento de archivos en línea. Para evaluar el mérito de su propuesta, los investigadores efectúan baterías de pruebas obteniéndose reducciones en los

requerimientos de capacidad de almacenamiento del 44% (mínimo), 54.5% (promedio) y 35% (máximo).

Para Bharde et al. (2018), el problema de la latencia en el procesamiento y transferencia de datos en entornos de Internet de las Cosas (IoT) puede ser mitigado mediante el desarrollo de infraestructura versátil donde se apliquen tanto técnicas de compresión como de deduplicación de datos en repositorios ubicados en el borde la red, que luego son replicados a los servicios centrales de almacenamiento en nube. Este modelo optimización de la capacidad de almacenamiento se aplicó sobre datasets típicos de un entorno de sensores IoT, habiéndose obtenido los siguientes ratios entre la capacidad inicial y la capacidad optimizada: 1.61 (mínimo) y 2.94 (máximo).

En el paper de Almrezeq et al. (2021), exploran la utilización de la deduplicación a nivel de proveedores de servicios de almacenamiento en nube donde se enfrenta el problema de que los usuarios suelen subir sus datos encriptados, lo que impide una eficiente gestión del almacenamiento. Para superar esta barrera, los autores proponen un mecanismo de encriptación y deduplicación basado en los algoritmos criptográficos AES-CBC (encriptación) y funciones hash (identidad para deduplicación), el cual logra resultados interesantes en cuanto a rendimiento y seguridad. En este caso se alcanzó un ratio mínimo y máximo de Capacidad de almacenamiento con valores 1.34 y 2.66 respectivamente.

La investigación de Park et al. (2016), evaluó la aplicación de técnicas de deduplicación de datos en entornos de almacenamiento fuera de línea y para uso general, a diferencia del enfoque difundido de analizar entornos de almacenamiento que contengan datos provenientes de backups. A fin de mejorar el rendimiento de los sistemas de almacenamiento, se propuso aplicar selectivamente la deduplicación solo a los datos que cumplieran patrones de operación, como por ejemplo haber tenido accesos múltiples. Esta aplicación selectiva permitió alcanzar un ratio máximo de capacidad de almacenamiento de 7.14.

El artículo de Tyj et al. (2019), también abordó un área de estudio no convencional de las técnicas de deduplicación, tal como es la optimización del almacenamiento de datos compuestos por imágenes de máquinas virtuales en

ejecución (migración en vivo), lo que lleva a que el proceso de migración hacia y desde los servicios de almacenamiento en nube se acelere por cuanto se reduce la cantidad de datos que se transfiere. En la referida investigación se reportó un ratio mínimo de Capacidad de 5.88 y un ratio máximo de Capacidad de 9.76.

En Prathima et al. (2022), el giro de la investigación se orientó a servicios de almacenamiento en nube para usuarios que hayan desplegado escenarios de Internet de las Cosas (IoT). En estos servicios, independientemente de que se trate de múltiples usuarios independientes, las lecturas de las redes de sensores tienden a registrar muchos valores similares, vale decir suplicados, por lo que l agestión del almacenamiento se beneficia enormemente de la aplicación de algoritmos criptográficos a través de las técnicas de deduplicación de datos. En la investigación que se comenta, se mostró que la existencia de numerosos flujos de almacenamiento provenientes de estas redes de sensores también favorece el nivel de deduplicación. Entre los resultados reportados, se menciona el trabajo con más de 10,000 instancias de flujos de datos de sensores, habiéndose obtenido ratios de transferencia que oscilaban en el rango de 2.51 a 4.94.

Para Viji et al. (2019), la oportunidad de mejorar la gestión de los sistemas de almacenamiento se centró en la aplicación de técnicas de deduplicación sobre almacenamiento primario compuesto por discos duros convencionales (internos a los servidores), discos USB (externos) e incluso las unidades de cinta (Tape backup). En la aplicación de las técnicas sobre datos de carácter general se alcanzó una reducción del 68% para en la capacidad de almacenamiento empleada, identificándose que cuando se aplicaban las técnicas sobre datos compuestos por el respaldo de información (backups), el porcentaje de reducción subía hasta el 83%.

De acuerdo con la investigación de Daniel et al. (2021), donde se analizó el almacenamiento de datos externo a las organizaciones, en servicios comerciales de computación en nube, se propuso la modificación de las técnicas de deduplicación para que el almacenamiento fuera a la vez que eficiente también seguro. Los resultados reportados indican que se obtuvo una reducción del costo de almacenamiento del 20%.

En cuanto a las teorías relacionadas con la investigación podemos mencionar a Palacios et al. (2006), quienes hacen un recuento de los tipos de algoritmos criptográficos. Los algoritmos criptográficos son algoritmos matemáticos que actúan sobre un grupo de datos principal para transformarlos y/o obtener un grupo de datos procesados que está relacionado con el grupo principal. Esta transformación puede ser una encriptación o una función de resumen tipo Hash. Se distinguen tres tipos de algoritmos: los de clave secreta, donde la clave de encriptación usada para generar el criptograma en el lado de origen del mensaje es la misma que se emplea para desencriptar el criptograma en el lado destino. EL segundo tipo son los algoritmos de clave pública, en los cuales cada extremo de la comunicación maneja dos claves, una privada y otra pública, siendo estas complementarias, vale decir que lo que se encripta con una clave se desencripta con su clave complementaria. El tercer tipo de algoritmo criptográfico que mencionan es el algoritmo Hash o de Resumen, donde se emplean funciones criptográficas para procesar un objeto digital de información, como puede ser un archivo, y se obtiene una huella hash o huella digital (fingerprint), característica de los datos que la han generado. Este tipo de algoritmo es empleado para poder identificar grupos de datos similares: si las huellas hash de cada grupo de datos son iguales, los conjuntos de datos también son iguales. Justamente este último tipo de algoritmo es el que se empleará en la investigación.

Sobti et al. (2012), hacen una revisión de las funciones criptográficas Hash, presentando los servicios de seguridad que brindan, tales como la Integridad y Autenticación, la Firma Digital, Autenticación de usuarios, etc. Con respecto a la integridad indican que las funciones Hash pueden ser vistas como funciones de una sola vía (One Way Hash Functions), que reciben como argumento un objeto de datos y generan como salida una cadena de bits de longitud fija. Por ejemplo, la función MD5 genera cadenas binarias de 128 bits. En general, estas secuencias binarias de salida se consideran únicas ya que no se conoce método alguno que permita manipular y/o revertir la operación. Esta funcionalidad puede emplearse para verificar que, si dos grupos datos son iguales a nivel de contenido.

Shin et al. (2022), desarrollaron investigaron la deduplicación de datos segura y eficiente en ambientes de Edge Computing. En estos escenarios se

capturan datos masivos provenientes de sensores, siendo gran parte de ellos similares pues las magnitudes monitoreadas no varían continuamente ni de forma aleatoria. Esta escalaridad de los datos requiere una eficiente gestión del almacenamiento de datos, entendiéndose esta como la adopción de medidas que aseguren que el almacenamiento de los datos se hace de forma segura y eficiente. La eficiencia del almacenamiento de datos se refleja en la reducción del consumo de espacio de almacenamiento y los tiempos de transferencia de estos datos. La deduplicación de datos permite que un servidor de almacenamiento pueda maximizar la eficiencia del almacenamiento al almacenar una sola copia de los datos redundantes y remover otros duplicados.

En Adhab et al. (2022), se presenta una revisión de técnicas de deduplicación de datos, donde se emplean algoritmos criptográficos del tercer tipo, vale decir Hash o Resumen, lo que constituye la aplicación de algoritmos criptográficos para la gestión del almacenamiento de datos. La deduplicación de datos es una tecnología que ahorra espacio de almacenamiento. Puede identificar conjuntos de datos redundantes al compararlos mediante sus huellas hash. De esa forma puede mantener solo una copia de los datos y crear punteros lógicos que reemplacen las otras copias duplicadas. El volumen de datos es reducido mediante la deduplicación de datos, reduciéndose también el espacio de almacenamiento, el ancho debanda empleado para transferir los datos y el costo de almacenamiento de los datos. La deduplicación de datos puede ser empleada en cualquier sistema donde se almacene o transfiera datos.

En SNIA (2022), se define un sistema de almacenamiento de datos como una colección de elementos que controlan el almacenamiento que se efectúa en un conjunto de dispositivos de almacenamiento y que son gestionados por software para poder brindar servicios de almacenamiento a una o más computadoras.

Para Chen (2015), los sistemas de almacenamiento de datos pueden escalar solo si cuentan con vastos recursos computacionales, con gran capacidad de transferencia de datos y con muchos dispositivos de almacenamientos organizados en pools. El almacenamiento a gran escala requiere más que solo discos donde guardar los datos. Se requiere que los recursos de cómputo y de conectividad crezcan también para soportar las operaciones diarias propias del sistema de

almacenamiento como son la deduplicación, la compresión, la encriptación/desencriptación y otras.

De acuerdo con SNIA (2022), la eficiencia de un sistema de almacenamiento está dada por la razón entre la capacidad efectiva del sistema y su capacidad nominal. Para un sistema que inicia operación el indicador de eficiencia es bajo, pero un sistema con optimización de capacidad generalmente optimizará su eficiencia conforme se vaya cargando datos. No se podría estimar con exactitud, a priori, el nivel de eficiencia que se alcanzará.

Con respecto al acceso a los datos de un sistema de almacenamiento, Quiña et al. (2019), señala que la unidad de datos [donde se aplicará el algoritmo criptográfico propio de la deduplicación de datos] es la unidad de medida bajo la cual se identifica a un conjunto de datos, se le accesa y se le procesa. En el caso de la deduplicación este acceso puede darse a nivel de archivo (files) o a nivel de fragmentos o bloques (chunks). En el primer caso, los aplicativos los datos almacenados cuentan con un identificativo global que es el nombre del archivo, mientras que, en el caso del acceso a bloques, es posible acceder a subgrupos de datos definidos dentro de un archivo y a los cuales se les ha otorgado un identificativo particular. El acceso a nivel de archivo simplifica la operación de los sistemas de almacenamiento y el trabajo de los usuarios, pero el acceso a nivel de bloques abre la posibilidad de que se pueden efectuar otro tipo de operaciones con los datos más allá del acceso simple que es la lectura de datos, por ejemplo se puede trabajar en la recuperación de datos en archivos que hayan sido dañados o también se puede implementar procedimientos de optimización para reducir el espacio de almacenamiento o mejorar los tiempos de acceso a los bloques de datos.

En la tesis de maestría de Marín (2022), se concluye que es muy frecuente que haya datos duplicados debido a que las organizaciones almacenan masivamente los datos con que trabajan, siendo de gran importancia determinar qué datos están duplicados y luego aplicar algún mecanismo de optimización para unificar estos datos duplicados. Se han establecido múltiples métodos para este fin, destacando record linkage en el campo de la estadística, merge-purge, data

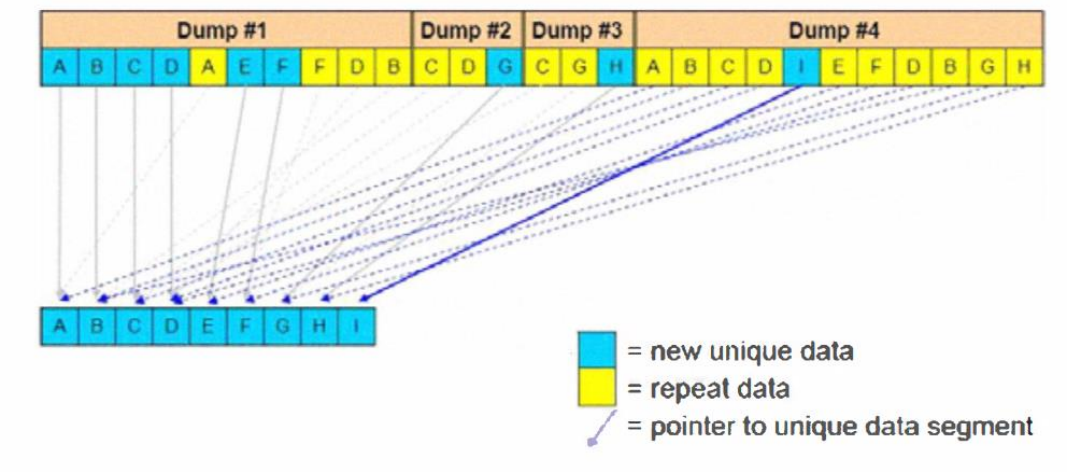
deduplication e instance identification en el campo de bases de datos, también se emplean algoritmos de probabilidad y de machine learning.

De acuerdo con Al Azad et al. (2022), cuando se generan datos en diversas fuentes, y estás son atendidas por varios servidores, se debe asegurar que los datos de interés se almacenen en el mismo servidor a fin de poder aplicar alguna técnica de detección de datos duplicados. Podemos considerar el caso de una cámara CCTV que genera vistas (fotogramas) consecutivos, los cuales son descargados en diferentes servidores, con lo que no sería posible deduplicar estos fotogramas. Se concluye que los modelos de almacenamiento de datos, para conjuntos de datos donde hay duplicidad, deben considerar el almacenamiento centralizado para poder aplicar optimizaciones en el uso de espacio de almacenamiento, concretamente la deduplicación.

Según Store et al. (2008), los grandes sistemas de almacenamiento de datos requieren ser gestionados a fin de mejorar sus parámetros de operación, tales como son la capacidad de almacenamiento, la capacidad de transferencia de datos desde y/o hacia el sistema, y el costo de operación. El punto más crítico es la utilización de la capacidad de almacenamiento. Con ese propósito se ensayan diversas técnicas como la compresión de datos, sin embargo, la deduplicación de datos es preferida sobre la compresión ya que la compresión graba todos los grupos de datos, así sean repetidos, pero la deduplicación identifica los grupos de datos repetidos y los reemplaza por una referencia lógica, reduciéndose la cantidad de datos a grabar. En la figura 1 se muestra un esquema de operación de la deduplicación de datos.

Figura 1

Esquema de deduplicación de datos



Nota: Gráfico tomado de Store et al (2008)

Mohamed et al. (2021), detallan los pasos que se deben seguir para efectuar la deduplicación de datos, los cuales pueden ser incorporados en un algoritmo:

1) Separar los datos de entrada en archivos o fragmentos (bloques)
2) Calcular la huella digital de cada archivo o fragmento (bloque)
3) Mediante la huella digital verificar si hay archivos o fragmentos (bloques) ya existentes en los datos almacenados
4) Si se detecta un archivo o fragmento (bloque) se reemplaza por una referencia al registro
5) Los datos replicados se descartan, dejándose solo un fragmento (bloque) representativo

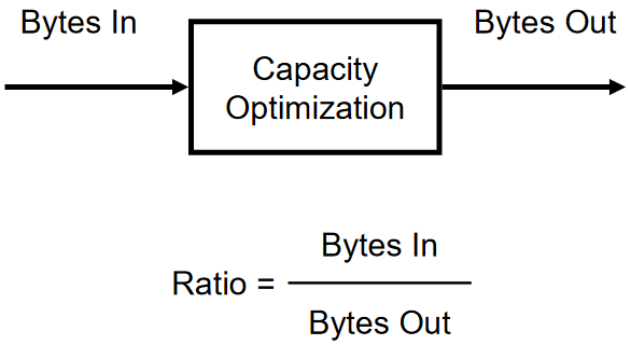
En Salazar (2018), se presenta un detallado recuento de las tecnologías de almacenamiento en centro de datos, definiéndose como características avanzadas de un sistema de almacenamiento: la replicación, los volúmenes delgados, la deduplicación, la destrucción segura de datos y el auto-nivelamiento. Con respecto a la deduplicación, se resalta que es un método que ahorra recursos, especialmente espacio de almacenamiento, y que debe incluir escaneos de los datos para detectar grupos de bytes redundantes que se descartarán, quedando solo una copia a la cual hacen referencias las instancias repetidas.

Para Kaur et al. (2022), la deduplicación de datos es efectiva para mejorar la eficiencia de los sistemas de almacenamiento de datos al reducir el costo del almacenamiento en base a la reducción de espacio de almacenamiento y del uso efectivo de ancho de banda. El ancho de banda o capacidad de transferencia de datos es un indicador numérico que relaciona la cantidad de datos transmitidos en un intervalo de tiempo.

En Dutch (2008), es posible revisar los indicadores y ratios de las dimensiones de la variable dependiente. Por ejemplo, la capacidad de almacenamiento se define como es espacio de disco empleado para almacenar un grupo de datos. Es posible definir un indicador que relacione el volumen de datos de entrada con el volumen de datos de salida. En la figura 2 se muestra un ejemplo de este indicador.

Figura 2

Indicador de reducción del uso de la capacidad de almacenamiento



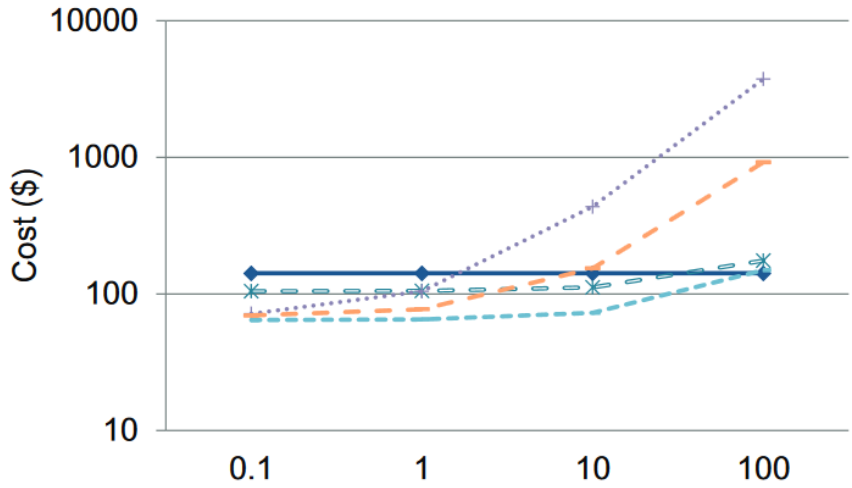
Nota: tomado de Dutch (2008)

Según Agarwala et al. (2011), el costo de operación de los sistemas de almacenamiento o simplemente costo de almacenamiento, se incrementa por el almacenamiento de datos duplicados múltiplemente, por lo que es necesario aplicar técnicas de reducción de datos tales como compresión y deduplicación, siendo la deduplicación la que presentara mayor impacto en el costo de operación del sistema de almacenamiento. El costo de almacenamiento es un indicador numérico

que resume la valoración, usualmente económica, de los recursos utilizados para llevar a cabo el almacenamiento de datos, y comprende diversas magnitudes, siendo las más importantes la capacidad de almacenamiento y el ancho de banda de transferencia de datos En la figura 3 se muestra el incremento del costo de operación para diversos niveles de compresión (curvas con variación creciente) y para la deduplicación (que prácticamente se mantiene en un costo constante, plano. Un indicador que se puede trabajar para recoger estos datos de costo es la razón entre el costo de almacenamiento de los datos originales y el costo de almacenamiento de los datos deduplicados.

Figura 3

Incremento del costo de almacenamiento según el volumen de datos y su nivel de compresión



Nota: tomado de Agarwala et al. (2011)

También en Kaur et al. (2022), se señala que, debido a la proliferación de redes sociales, los datos como las imágenes suelen estar duplicados ampliamente. En estos casos es posible abordar la deduplicación de imágenes que coincidan de forma exacta, sin embargo, en estas mismas redes sociales, se suele modificar parcialmente las imágenes, por lo que se esperaría mejores resultados si se aplica la deduplicación a imágenes cercanamente parecidas en lugar de imágenes exactamente iguales. Para determinar si dos grupos de datos, imágenes en este caso, son cercanos, se emplea el criterio de similitud o similaridad. El nivel de

similitud de datos [a que se aplica el algoritmo criptográfico] es un indicador numérico que expresa qué tanto ha variado un dato modificado con respecto a un dato original. Cuando el dato original no ha variado significativamente, mantiene la mayor parte de sus características.

Según Chen et al. (2021), existe una gran cantidad de datos redundantes en sistemas de almacenamiento centralizado de datos, como los basados en nubes de servicio. Por razones de privacidad los usuarios suelen encriptar los archivos, con lo que la detección de duplicidad en base a igualdad de archivos se dificulta, por lo que deben considerarse otros métodos de detección de duplicidad como, por ejemplo, el Hash perceptual, donde se trabajaría con imágenes que guardan similitud.

III. METODOLOGÍA

3.1 Tipo y diseño de investigación

Según Torres (2016), la elección del tipo y diseño de la investigación no está en función de lo que el investigador prefiera, sino más bien cuando se decide por cierto método de investigación se hace buscando que esta se adapte lo mejor posible al objeto de estudio.

El enfoque que se ha empleado en la investigación ha sido el cuantitativo, ya que se ha buscado medir la relación entre la variable independiente, “Aplicación de algoritmos criptográficos”, y la variable dependiente, “Gestión del almacenamiento de datos”. Según Otero (2018), el enfoque cuantitativo apunta a medir datos numéricos, se observan los procesos a través de la recopilación de datos que posteriormente son analizados para dar respuesta a las preguntas de investigación. En este enfoque juega un papel clave la estadística y sus diversos tipos de análisis.

En lo que al diseño metodológico se refiere, inicialmente se consideró trabajar con un diseño no experimental. Según Álvarez-Risco (2020), estas investigaciones se caracterizan porque en ellas no se manipula las variables de investigación. Por su parte, Reio (2016), señala que las investigaciones que siguen este diseño constituyen un gran aporte si se presentan los resultados correctamente porque pueden usarse cuando no es posible o no es deseable la experimentación. Sin embargo, al desarrollar la investigación se notó que sería necesario comparar mediciones antes y después de aplicar los algoritmos criptográficos en la gestión del almacenamiento de datos. En este sentido, Masid (2017), clasifica su investigación como pre-experimental ya que cuenta con tres grupos de estudio en los cuales analizó los valores obtenidos antes y después del tratamiento experimental. Por su parte, Ramos (2021), sostiene que en un diseño pre-experimental el investigador cuenta con un grupo de experimentación al cual se le aplica la acción de intervención. Al comparar estos dos diseños, nos pareció más adecuado clasificar el diseño como pre-experimental.

Si consideramos una visión temporal de la investigación, la podemos clasificar como transversal ya que la investigación se centra en recabar datos y evaluarlos en un tiempo determinado. Esta apreciación se valida con lo indicado

por Rodríguez et al. (2018), que señala que un estudio transversal es a la vez que analítico y descriptivo, siendo su objetivo establecer qué tanto se observa una característica en la población objeto de la investigación.

Por otro lado, la presente investigación es de carácter aplicado. Nicaragua (2018), puntualiza que estas investigaciones son desarrolladas para abordar problemáticas prácticas que requieren ser solucionadas mediante el desarrollo de soluciones tecnológicas que serán recogidas por instancias productivas. En nuestro caso, tenemos que la aplicación de algoritmos criptográficos en la gestión del almacenamiento de datos en una empresa de seguridad electrónica les permitirá reducir las capacidades de almacenamiento y transferencia de datos, así como el costo de dicho almacenamiento de datos, al remover los datos duplicados.

Desde el punto de vista del tipo de estudio, podemos clasificar la investigación como correlacional. En efecto, Ramos (2020), afirma que en este tipo de estudio hay la necesidad de presentar una hipótesis que vincule a dos o más variables. Si el trabajo es cuantitativo, las técnicas estadísticas harán posible extender a toda la población los hallazgos realizados.

.

3.2 Variables y Operacionalización

Variable Independiente

Aplicación de algoritmos criptográficos

Definición conceptual

La deduplicación de datos (DD) es una tecnología que ahorra espacio de almacenamiento. Puede identificar conjuntos de datos redundantes al compararlos mediante sus huellas hash. De esa forma puede mantener solo una copia de los datos y crear punteros lógicos que reemplacen las otras copias duplicadas. El volumen de datos es reducido mediante la DD, reduciéndose también el espacio de almacenamiento, el ancho debanda empleado para transferir los datos y el costo de almacenamiento de los datos. La DD puede ser empleada en cualquier sistema donde se almacene o transfiera datos, Adhab et al. (2022).

Variable Dependiente

Gestión del almacenamiento de datos

Definición conceptual

La eficiencia del almacenamiento de datos se refleja en la reducción del consumo de espacio de almacenamiento y los tiempos de transferencia de estos datos. La deduplicación de datos (DD) permite que un servidor de almacenamiento pueda maximizar la eficiencia del almacenamiento al almacenar una sola copia de los datos redundantes y remover otros duplicados, Shin et al. (2022).

Operacionalización de las variables

Definición Operacional de la Variable Independiente

A los objetos de almacenamiento (archivo o bloques) se les calcula la huella digital "hash", que se compara con las huellas hash de otros objetos de almacenamiento. Si las huellas coinciden entonces se acepta que los objetos de almacenamiento son iguales, reemplazándose las copias del objeto de almacenamiento por un enlace lógico a la primera copia. De esta forma se ahorra espacio de almacenamiento, Xia et al. (2016).

Definición Operacional de la Variable Dependiente

La eficiencia del sistema de almacenamiento se puede medir estableciendo la relación (ratio) entre el uso de recursos al almacenar datos de forma estándar y el uso de los mismos recursos al almacenar datos de forma optimizada utilizando deduplicación de datos. Los recursos relevantes a ser medidos son el volumen de datos almacenados, al ancho de banda empleado para transferencias de los datos y el costo de almacenamiento de datos, ZHou et al. (2013).

Las dimensiones de la variable dependiente son las siguientes:

Capacidad de almacenamiento, que se refiere al espacio de almacenamiento utilizado para alojar un grupo de datos, ZHou et al. (2013).

Capacidad de transferencia de datos, entendida como el ancho de banda que se emplea para transmitir los datos hacia y desde repositorios centralizados, también pudiendo expresarse como el tiempo que toma la transmisión, ZHou et al. (2013).

Costo del almacenamiento, que hace referencia al costo económico de operar el sistema de almacenamiento de datos, ZHou et al. (2013).

En el Anexo 1 se presenta la Matriz de operacionalización de variables.

3.3 Población, muestra y muestreo

Población

Según Ventura-León (2017), la población está constituida por el grupo de elementos que tiene las características que se desea estudiar. Sim embargo también se hace la distinción entre la población diana, que suele ser muy grande y a la cual no se puede acceder, y la población accesible, compuesta por menor número de elementos, que han sido seleccionados con criterios de inclusión y/o exclusión.

En la presente investigación la población de estudio estuvo compuesta por treinta objetos de almacenamiento representativos de los datos almacenados, y en su mayoría correspondían a videos demostrativos típicos producidos por el monitoreo y vigilancia electrónica mediante cámaras seguridad, donde se graban fotogramas correspondientes a la(s) zona(s) de interés.

Muestra

Según López-Roldán (2017), la muestra estadística es un subconjunto de elementos que poseen las características relevantes de una población.

En la investigación se seleccionó un total de treinta objetos de almacenamiento que el investigador considera son característicos y adecuados para llevar a cabo el estudio.

En este sentido, cabe aclarar que no se aplicó fórmula alguna para estimar el tamaño de la muestra, sino más bien criterios sugeridos por docentes expertos en métodos de investigación.

3.4 Técnicas e instrumentos de recolección de datos

De acuerdo con Arias (2020), la forma en que recabamos datos sobre el objeto de estudio es denominada “Técnica de recolección de datos”, y estos datos deben guardarse, al menos temporalmente, para que sean analizados y procesados. Justamente el elemento que nos ayuda a registrar los datos sea de forma manual o automática, en físico o digital, es conocido instrumento de recolección de datos. El instrumento que se empleó fue la Ficha de registro, la que nos permitió recopilar los datos relevantes de cada elemento de la población.

En el Anexo 2 se presenta el formato de la ficha de registro empleada en la investigación.

3.5 Procedimientos

En la presente investigación se ha llevado a cabo la revisión de la literatura especializada para tener una base teórica que permita establecer el alcance de la investigación, acotar el problema, presentar las definiciones de las variables de investigación y sus dimensiones, así como establecer la población objeto de estudio.

El siguiente paso fue solicitar la carta de presentación de la Unidad de Posgrado de la UCV (1074-2022-UCV-VA-EPG-F01/J) para poder pedir, formalmente, la autorización de la empresa EV Seguridad SAC para trabajar con sus registros de datos.

Posteriormente, al tratarse de un trabajo cuantitativo de tipo pre-experimental, se procedió a diseñar una ficha de registro que permita recolectar datos de las condiciones de operación del sistema de almacenamiento de datos antes y después de aplicar los algoritmos criptográficos tipo hash para identificar y remover los datos duplicados de los objetos de datos almacenados. Estos

algoritmos fueron aplicados habilitándose la deduplicación de volúmenes de almacenamiento bajo el sistema operativo Windows Server 2022.

Los objetos de datos analizados son, principalmente, los fotogramas generados en cámaras de seguridad que registran, de forma continua, la actividad desarrollada en puntos de interés.

El análisis se aplicó tanto a nivel de archivo como a nivel de fragmentos o bloques, ya que la organización bajo estudio no cuenta con repositorios centralizados o tipo nube, y tampoco se brindó acceso a volúmenes dedicados de almacenamiento

El procedimiento seguido consistió en analizar los tipos de objeto de almacenamiento con que trabaja la empresa para seleccionar, a priori, una muestra representativa, que estuvo compuesta principalmente por videos generados por diferentes tipos de cámaras de seguridad, bajo diferentes condiciones de calidad e iluminación, así como instancias de máquinas virtuales usadas para experimentación y prueba de drivers. En el caso de los videos, se trabajó con versiones demostrativas no comprimidas, extrayéndose de estos videos la secuencia de fotogramas que los compone, siendo estos datos masivos, aunque variables por la resolución y velocidad de muestreo de cada cámara. De esta forma, un objeto de almacenamiento estaba compuesto realmente por muchos archivos menores contenidos en un directorio. Los directorios correspondientes a los objetos de almacenamiento fueron copiados a volúmenes de almacenamiento de datos bajo el sistema operativo Windows Server 2022, tanto en unidades con almacenamiento estándar como en unidades con almacenamiento que tenía habilitada la deduplicación de datos.

Posteriormente, para cada objeto de almacenamiento se registró en la ficha correspondiente los valores de las dimensiones de la variable Independiente bajo las cuales se efectuó la medición, así como los valores resultantes de las dimensiones de la variable Dependiente, calculándose luego los valores correspondientes a los indicadores de la variable Dependiente (ratios).

Finalmente, los datos medidos y los indicadores calculados para los treinta objetos de almacenamiento, que constituían la población de estudio, fueron cargados y analizados estadísticamente con el aplicativo SPSS para determinar si

existía relación entre la aplicación de los algoritmos criptográficos y la gestión del almacenamiento de datos.

3.6 Métodos de análisis de datos

Debido al enfoque cuantitativo de la investigación, se recogió gran cantidad de datos a través del instrumento Ficha de registro. Este instrumento ayudó a verificar que la recopilación fuese completa para cada grupo de análisis (registro de datos y tratamiento posterior, hasta la generación de los indicadores). Esta información se ha presentado de forma descriptiva empleando técnicas estadísticas.

Por otra parte, el hecho de que la población fuese relativamente pequeña (30 elementos), nos llevó a escoger técnicas adecuadas. Por ejemplo, para validar la normalidad se ha empleado la prueba de Shapiro-Wilk debido a la cantidad de fichas de registro empleadas (muestras), y para contrastar las hipótesis se ha utilizado la prueba de Wilcoxon ya que las pruebas de normalidad señalaron que los datos no seguían una distribución normal y se requería una prueba no paramétrica.

3.7 Aspectos éticos

Según Espinoza (2019), una de las principales faltas éticas en la investigación científica es el empleo de material académico bajo la forma de textos, imágenes y tablas, sin llevar a cabo la referenciación adecuada, incluso aun cuando el investigador esté empleando sus publicaciones previas. De allí que sea importante aplicar un esquema riguroso de manejo de referencias para evitar cuestionamientos posteriores a la investigación que se realiza.

El maestrista que suscribe el presente documento deja constancia de que la presente investigación es original en cuanto a su concepción, teniéndose la propiedad intelectual de la misma.

En lo referente a la línea de investigación, se ha trabajado dentro de la línea de investigación específica “Sistemas de Información y Comunicaciones”, la cual está contemplada en la resolución de consejo universitario Nº0200-2018/UCV.

Con respecto a la resolución Nº107-2022-VI-UCV, la presente investigación se enmarcó en la línea de acción 3 de Responsabilidad Social Universitaria (Línea de acción RSU 3: Desarrollo económico, empleo y emprendimiento), así como el Objetivo de Desarrollo Sostenible 8 (ODS 8: Trabajo decente y crecimiento económico).

El presente reporte, correspondiente a la investigación desarrollada, se ha elaborado siguiendo las reglas APA (American Psychological Association) en su versión 7.

También cabe señalar que la investigación se ha desarrollado bajo el marco especificado en la resolución 110-2022-VI-UCV del Vicerrectorado de Investigación de la Universidad César Vallejo, que constituye la guía para la elaboración de productos de investigación de fin de programa.

Asimismo, se deja constancia que el presente reporte se ha evaluado con el aplicativo Turnitin para identificar el porcentaje de similitud de la presente investigación frente a otras publicaciones científicas, habiéndose tomado las medidas pertinentes para mantener un bajo nivel de similitud.

IV. RESULTADOS

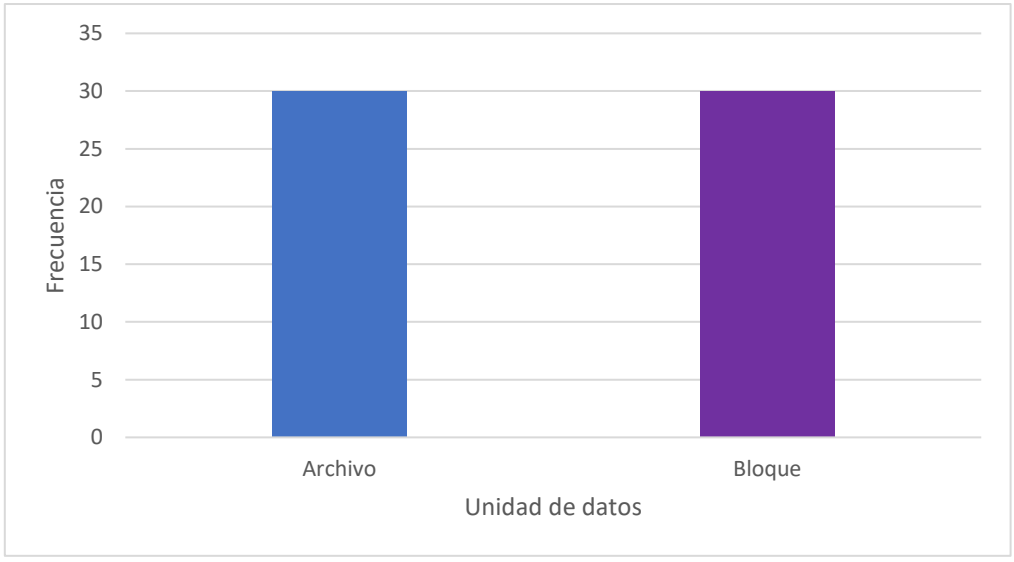
4.1 Análisis descriptivo

La muestra analizada en la presente investigación estuvo compuesta por treinta objetos de almacenamiento representativos de los tipos de datos utilizados regularmente en la empresa caso de estudio.

A estos objetos de almacenamiento se les aplicaron los algoritmos de encriptación tipo Hash que permiten deduplicar los datos almacenados e impactar en la gestión del almacenamiento. En la figura 4 se observa la cantidad de prueba efectuadas distribuidas por la unidad de datos a la cual se aplicó, pudiendo esta ser Archivo o Bloque.

Figura 4

Cantidad de pruebas analizadas por Unidad de datos



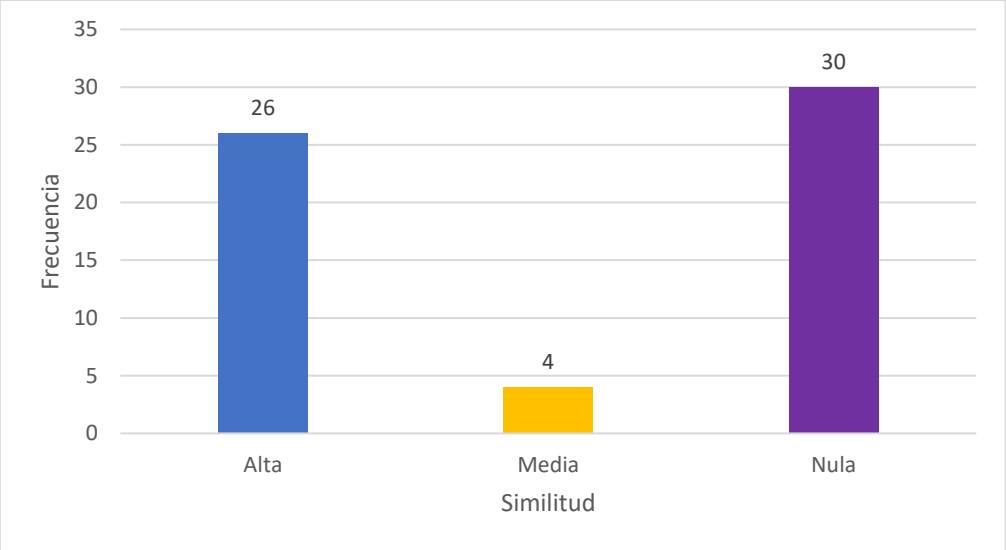
Nota: Elaboración propia, empleando SPSS versión 25

Un primer resultado que se obtuvo al analizar los objetos de almacenamiento en función al tipo de unidad de datos fue el grado de similitud que presentaban frente al tipo de deduplicación aplicada. La figura 5 muestra la distribución del grado de similitud de los objetos de almacenamiento analizados. Se aprecia que al aplicar los algoritmos criptográficos conducentes a la deduplicación de datos a nivel de archivo el grado de similitud encontrado fue nulo para los treinta casos, lo cual se

explica ya que cada tipo de objeto de almacenamiento era único, no existiendo, en general, archivos duplicados. Por el contrario, al aplicar las técnicas de deduplicación de datos a nivel de bloque se encontró que veintiséis objetos de almacenamiento, los archivos de video, presentaban similitud alta y cuatro objetos de almacenamiento, las instancias virtualizadas, presentaban similitud media.

Figura 5

Grado de similitud encontrado en el análisis de los objetos de almacenamiento



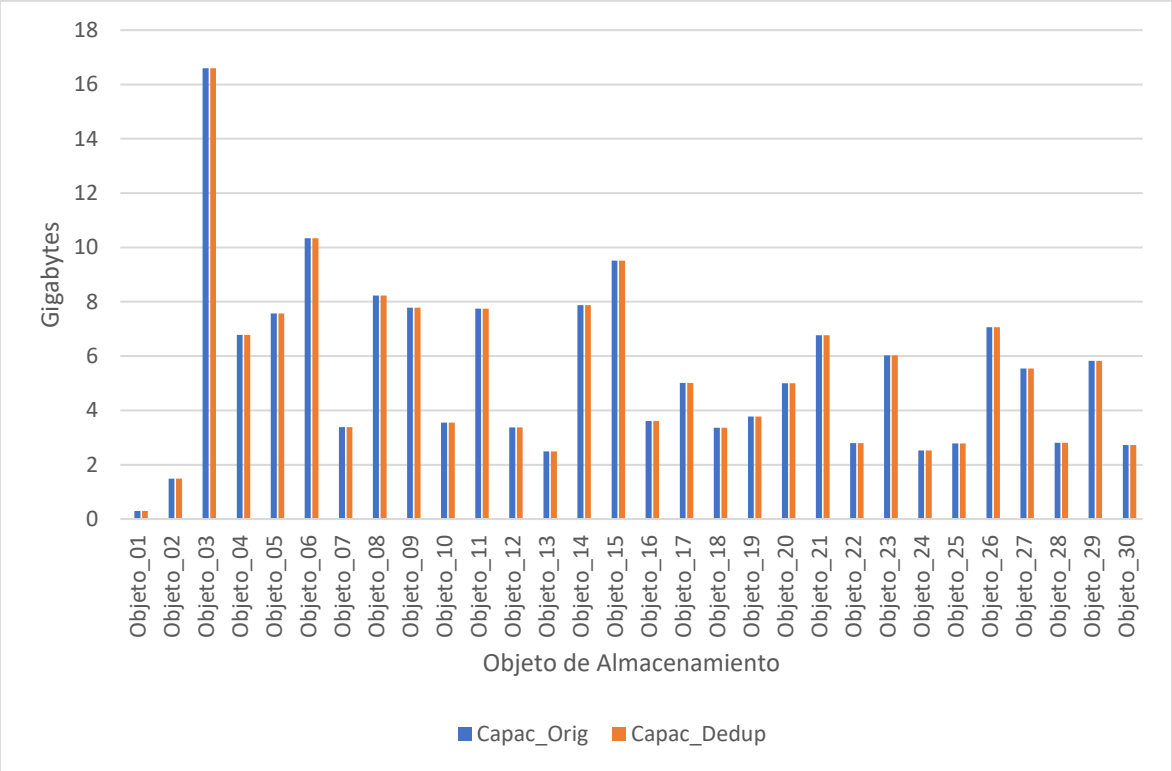
Nota: Elaboración propia, empleando SPSS versión 25

Una consecuencia directa de que los objetos de almacenamiento presentasen similitud nula al aplicarse la técnica de deduplicación de datos a nivel de archivo fue que estas técnicas no tuvieron impacto en la gestión del almacenamiento.

En efecto, en la figura 6 podemos ver que la capacidad de almacenamiento empleada originalmente y la capacidad de almacenamiento deduplicada fueron iguales.

Figura 6

Capacidad de Almacenamiento Original y Deduplicada para Unidad de datos "Archivo" (GB)



Nota: Elaboración propia, empleando SPSS versión 25

Una forma alterna de ilustrar este resultado es presentar el ratio de la capacidad de almacenamiento (relación entre la capacidad de almacenamiento empleada originalmente y la capacidad de almacenamiento empleada después de la deduplicación), el cual se ha mantenido en el valor de uno para todos los objetos de almacenamiento analizados. La figura 7 muestra lo que se acaba de indicar.

Figura 7

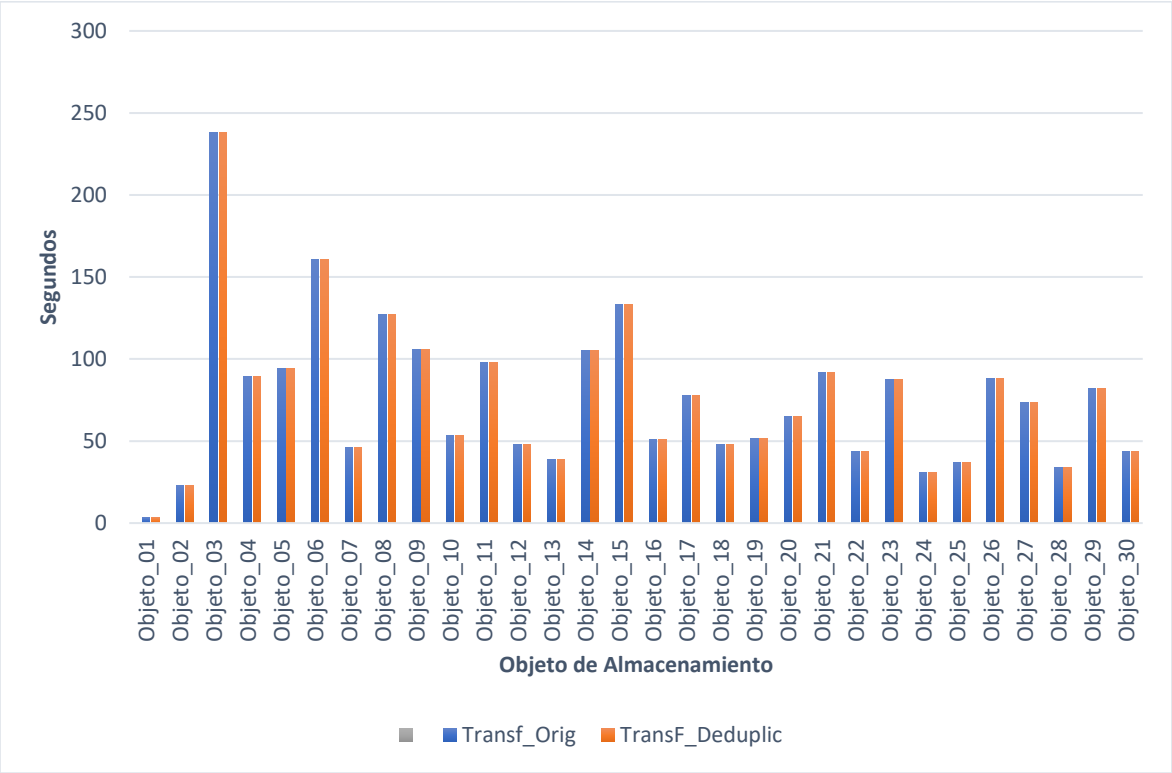
Ratio de Capacidad de Almacenamiento – Unidad de datos "Archivo"

Nota: Elaboración propia, empleando SPSS versión 25

Un resultado estadísticamente similar se ha obtenido al analizar los tiempos de transferencia de los datos originales y de los datos deduplicados. En ambos casos los tiempos se mantienen iguales debido a que no tuvo efecto la deduplicación a nivel de archivo. La figura 8 representa lo indicado.

Figura 8

Transferencia Original y Deduplicada para Unidad de datos tipo "Archivo"



Nota: Elaboración propia, empleando SPSS versión 25

Estos resultados también se pueden presentar mostrando el indicador ratio de transferencia que, al ser iguales los datos originalmente almacenados y los deduplicados, tomó el valor de uno. La figura 9 nos muestra el valor de dicho indicador para los objetos de almacenamiento analizados.

Figura 9

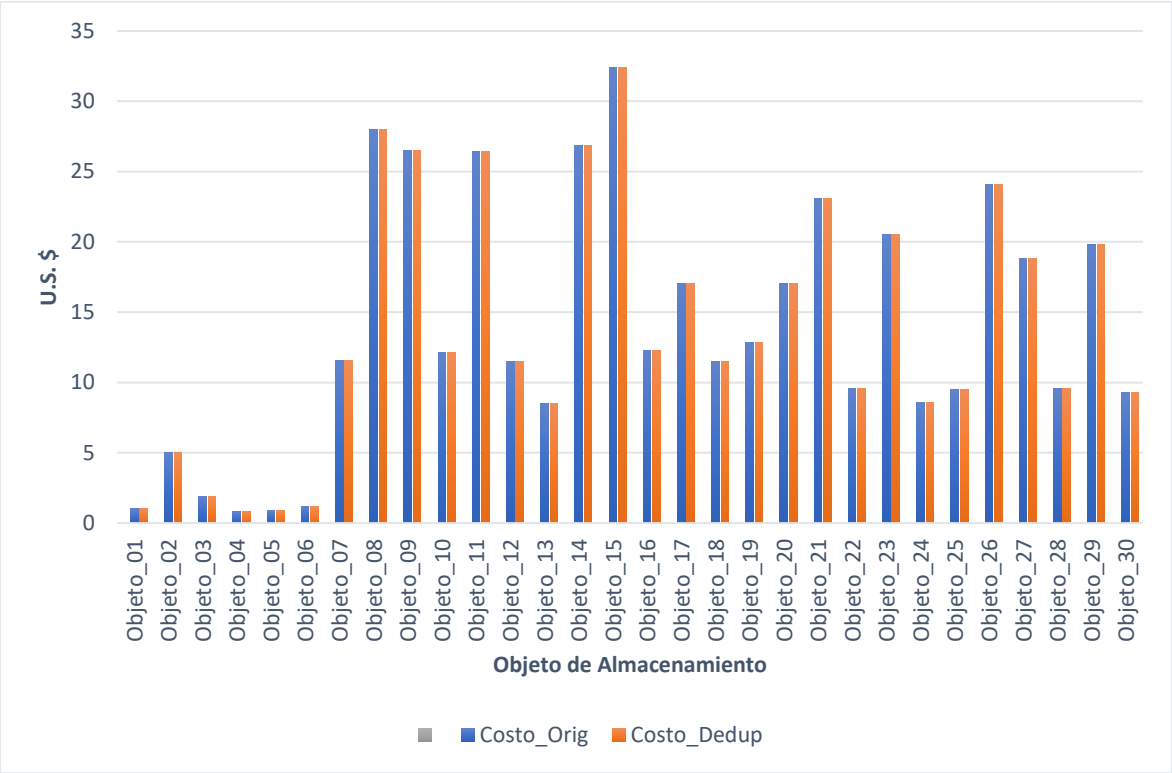
Ratio de Transferencia por objeto para Unidad de datos tipo "Archivo"

Nota: Elaboración propia, empleando SPSS versión 25

En cuanto al costo de almacenamiento de los datos, dado que no variaron las características de volumen de almacenamiento ni tiempo de transferencia, los cálculos de costo se mantuvieron iguales para el volumen de datos almacenado originalmente y para el volumen de datos almacenados después de la deduplicación a nivel de archivo. La figura 10 nos muestra estos resultados.

Figura 10

Costo Original y Deduplicado para Unidad de datos tipo "Archivo"



Nota: Elaboración propia, empleando SPSS versión 25

Tal como se ha hecho para las otras dimensiones, en este caso también podemos presentar el indicador ratio de costo de almacenamiento por objeto, que, en base a las consideraciones presentadas, toma el valor de uno para todos los casos. En la figura 11 podemos observar el valor de este indicador.

Figura 11

Ratio de Costo por objeto de almacenamiento para Unidad de datos tipo "Archivo"

Nota: Elaboración propia, empleando SPSS versión 25

El segundo grupo de pruebas fue la aplicación de los algoritmos criptográficos Hash, elemento central de las técnicas de deduplicación, pero esta vez a nivel de bloques o fragmentos. En este escenario, la deduplicación reconoce, al interior de los objetos de almacenamiento, la existencia de datos duplicados, procediendo a removerlos y producir ahorro en el uso de recursos y mejora en la gestión del almacenamiento.

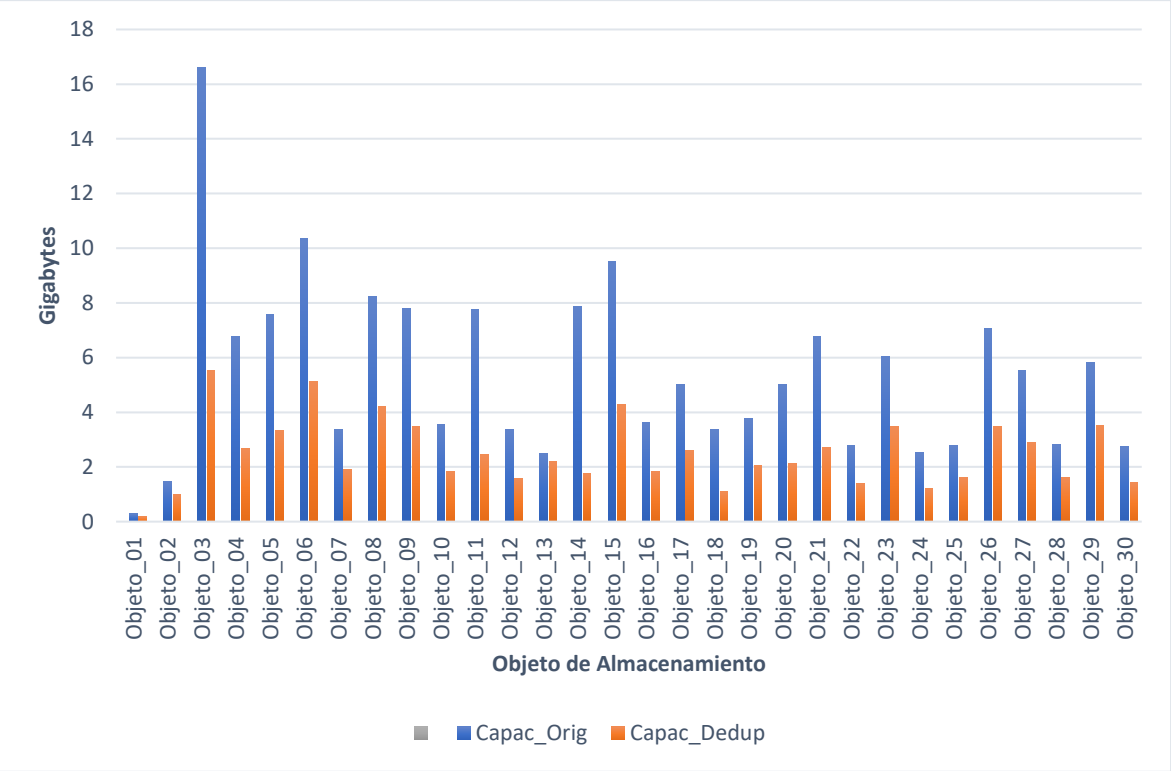
Para el caso de la empresa donde se efectuó la investigación, los objetos de almacenamiento que empelan son videos de seguridad que registran continuamente la misma zona de interés, de allí que se espere que gran parte de los datos digitales que componen el video sean similares y se logren niveles importantes de deduplicación.

En la figura 12 podemos observar un gráfico de barras paralelas de la capacidad original de almacenamiento que se ha empleado, comparada con la capacidad de almacenamiento empleada después de la deduplicación. Se aprecia que para todos los objetos de almacenamiento se ha logrado reducir la capacidad

empleada para almacenar los objetos, aunque los niveles alcanzados en cada objeto son distintos.

Figura 12

Capacidad de Almacenamiento Original y deduplicada para Unidad de datos tipo "Bloque"

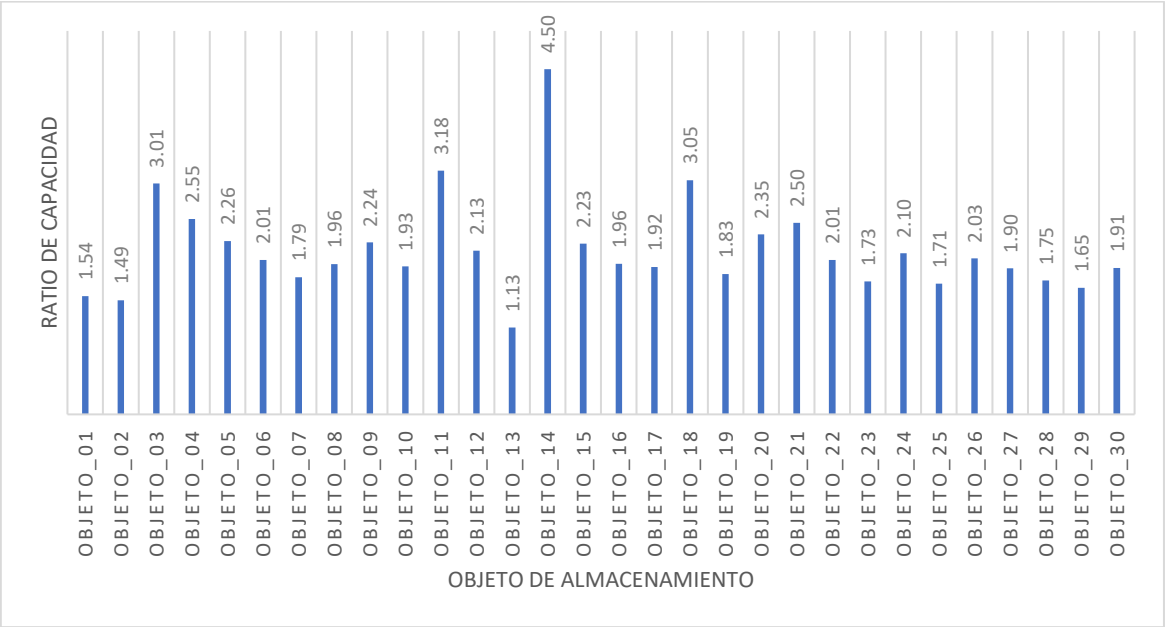


Nota: Elaboración propia, empleando SPSS versión 25

Esta última afirmación se puede observar con mayor facilidad en la figura 13, donde se muestra el indicador ratio de capacidad de almacenamiento, el mismo que resulta de efectuar el cociente entre la capacidad empleada originalmente y la capacidad de almacenamiento empleada después de efectuar la deduplicación a nivel de bloque. En todos los casos los indicadores son mayores a la unidad.

Figura 13

Ratio de Capacidad de Almacenamiento para Unidad de Datos tipo "Bloque"

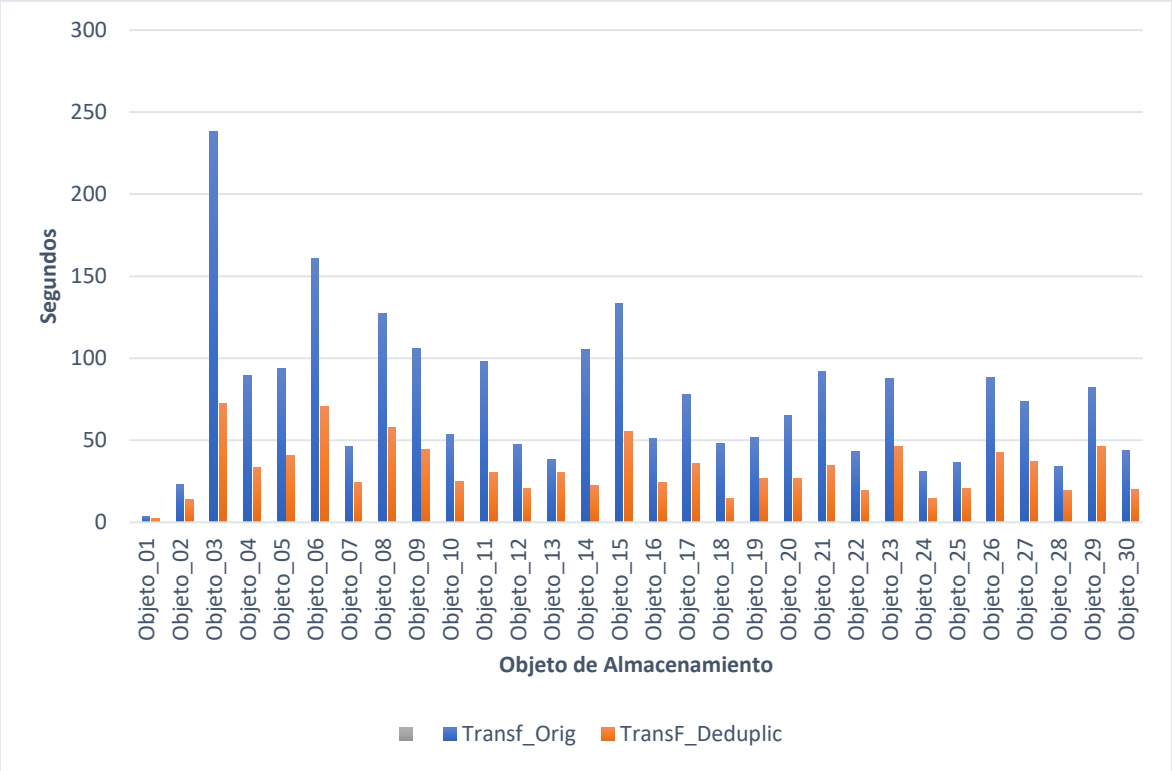


Nota: Elaboración propia, empleando SPSS versión 25

En cuanto al tiempo de transferencia original de los objetos de almacenamiento y el mismo tiempo medido luego de la aplicación de la técnica de deduplicación a nivel de bloque, vemos que se redujo según se muestra en la figura 14.

Figura 14

Transferencia Original y Deduplicada por objeto para Unidad de datos tipo "Bloque"

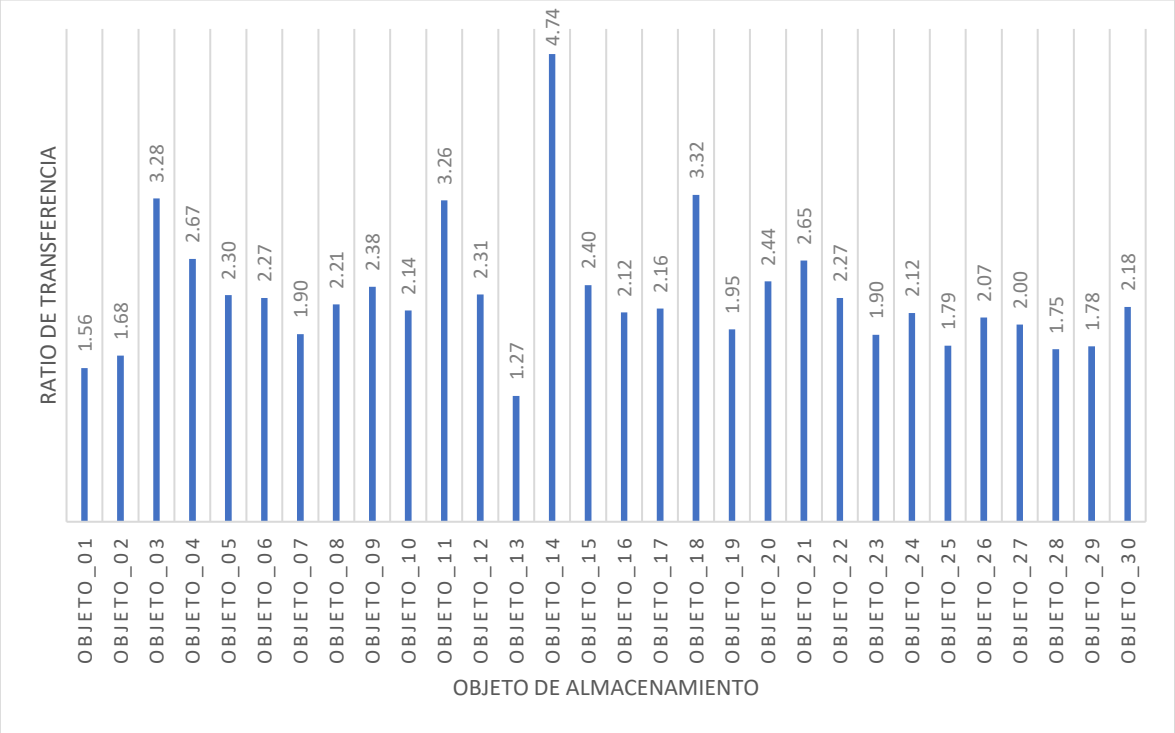


Nota: Elaboración propia, empleando SPSS versión 25

Estos resultados se pueden apreciar de mejor forma en la figura 15 donde se ha graficado el ratio de transferencia, que es la razón entre el tiempo empleado para transferir originalmente los objetos de almacenamiento y el tiempo empleado para transferirlos después de haber aplicado las técnicas de deduplicación. Cabe señalar que todos los ratios son mayores a uno, lo cual representa resultados positivos para la investigación.

Figura 15

Ratio de Transferencia por objeto para Unidad de datos tipo "Bloque"

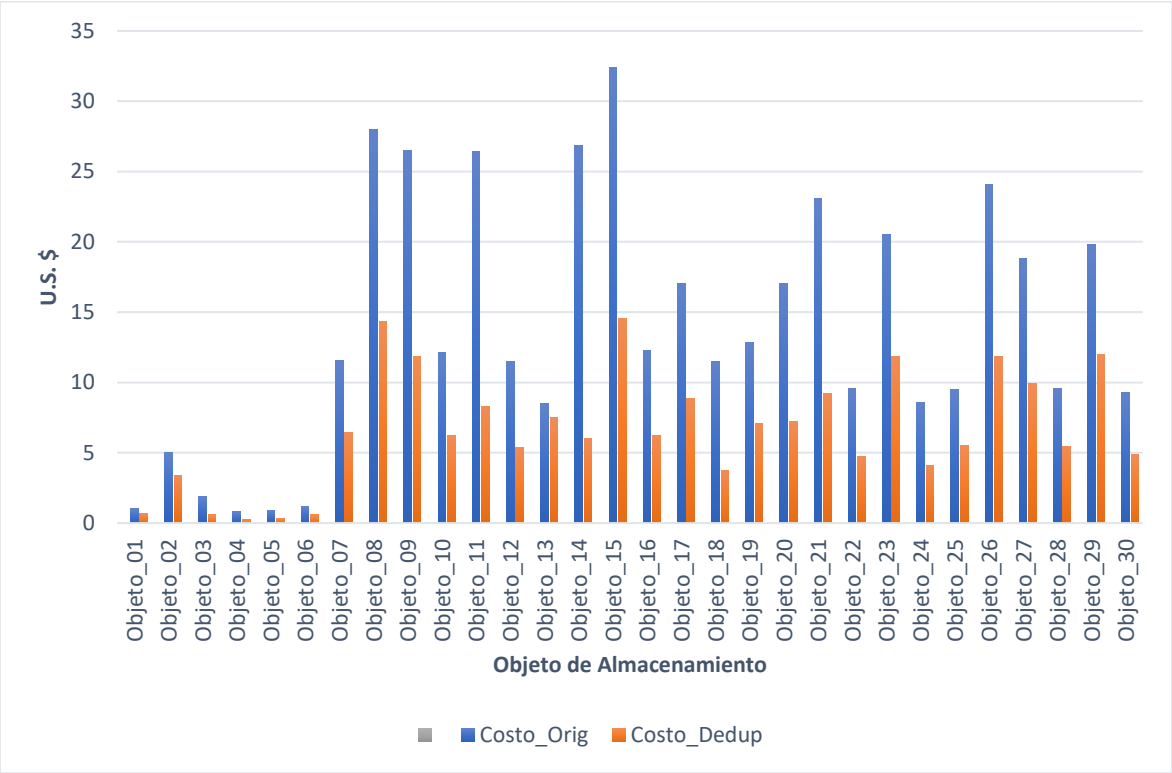


Nota: Elaboración propia, empleando SPSS versión 25

En el aspecto del costo de almacenamiento, este también se vio favorablemente impactado ya que el costo original se redujo en todos los casos tal como se muestra en la figura 16

Figura 16

Costo Original y Deduplicado para Unidad de datos tipo "Bloque"

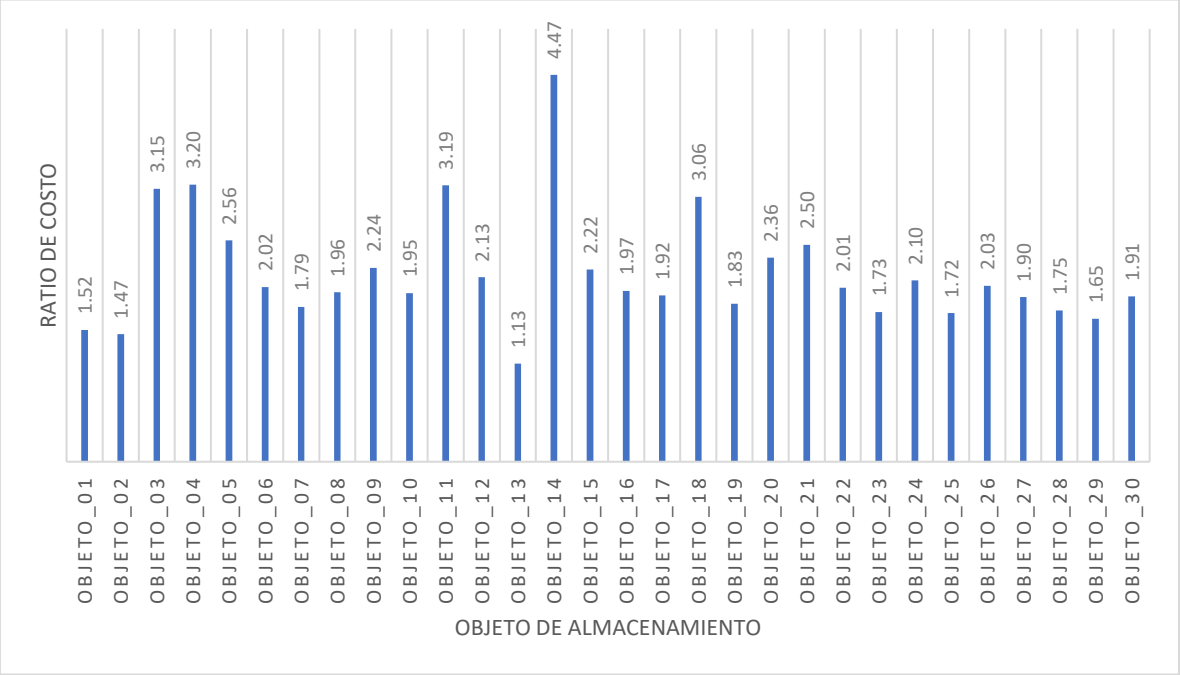


Nota: Elaboración propia, empleando SPSS versión 25

Igual que como se ha mostrado en las otras dimensiones, aquí también podemos ver en la figura 17 el ratio de costo, resultante de dividir el costo original del almacenamiento de un objeto entre el costo de almacenar el mismo objeto luego de aplicar la técnica de deduplicación a nivel de bloque. En este caso vemos también que los ratios alcanzaron valores superiores a uno en todos los casos.

Figura 17

Ratio de Costo de almacenamiento por objeto para Unidad de datos tipo "Bloque"



Nota: Elaboración propia, empleando SPSS versión 25

4.2 Análisis Inferencial

Tras presentar las estadísticas descriptivas, tenemos que llevar a cabo el análisis inferencial para poder establecer cuantitativamente la influencia del uso de algoritmos criptográficos en la gestión del almacenamiento de datos en una empresa de seguridad electrónica.

Para lograr este objetivo, se han efectuado mediciones antes de aplicar los algoritmos y después de aplicarlos.

Lo primero que debemos hacer es realizar la prueba de normalidad a fin de determinar si vamos a emplear pruebas paramétricas o no paramétricas, y para tal fin planteamos las siguientes hipótesis:

H_0 (Hipótesis nula): Los datos que se han analizado siguen una distribución normal.

H_a (Hipótesis alterna): Los datos que se han analizado no siguen una distribución normal.

El criterio de aceptación de la hipótesis nula H_0 es que el valor de la significancia (p-valor) sea mayor a 0.05, pero si el valor de la significancia es menor que 0.05 se procede a rechazar la hipótesis nula H_0 y se acepta la hipótesis alterna H_a.

Las pruebas se han efectuado utilizando el programa SPSS, versión 25, y como en la investigación se ha trabajado con 30 muestras, la prueba que corresponde aplicar es la de Shapiro-Wilk.

Prueba de normalidad: Ratios de la Gestión del Almacenamiento de Datos (Hipótesis general)

Tabla 3

Prueba de normalidad de Ratios, Hipótesis general

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
Ratio_Capac_Dedup	,195	30	,005	,830	30	,000
Ratio_Transf_Dedup	,215	30	,001	,831	30	,000
Ratio_Costo_Dedup	,200	30	,004	,851	30	,001

a. Corrección de significación de Lilliefors

Nota: Elaboración propia, empleando SPSS versión 25

En la tabla 3, se presenta la prueba de normalidad para los Ratios de Capacidad de almacenamiento, Transferencia de datos y Costo de almacenamiento, pudiéndose observar que, en los tres ratios mencionados, se ha obtenido una significancia con valor menor a 0.05, por lo que se rechaza la hipótesis nula (que los datos siguen una distribución normal) y se acepta la hipótesis alternativa de que los datos no siguen una distribución normal, por lo que se debe aplicar una prueba no paramétrica en su contrastación.

Prueba de normalidad: Alta Similitud de los datos (objetos de almacenamiento) (Hipótesis específica 1)

Tabla 4

Prueba de normalidad de Alta Similitud, Hipótesis específica 1

Alta Similitud de los datos	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
Ratio_Capac	,220	26	,002	,785	26	,000
Ratio_Transf	,234	26	,001	,793	26	,000
Ratio_Costo	,222	26	,002	,791	26	,000

a. Corrección de significación de Lilliefors

Nota: Elaboración propia, empleando SPSS versión 25

En la tabla 4, se observa el nivel de Alta Similitud en los objetos de Almacenamiento tanto para el Ratio de capacidad, Ratio de transferencia y Ratio de Costo la capacidad, presenta una significancia con un valor menor a 0.05, por lo que se rechaza la hipótesis nula (los datos siguen una distribución normal) y se acepta la hipótesis alternativa de que los datos no siguen una distribución normal, por lo que se debe aplicar una prueba no paramétrica. Es oportuno mencionar que, de los 30 objetos de almacenamiento analizados, solo veintiséis mostraron un nivel de similitud alto, de allí que la prueba considere solo estos 26 elementos.

Prueba de normalidad: Capacidad de Almacenamiento (Hipótesis específica 2)

Tabla 5

Prueba de normalidad para la Capacidad de Almacenamiento, Hipótesis específica 2

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
Capac_Orig	,159	30	,050	,897	30	,007
Capac_Dedup	,124	30	,200*	,955	30	,225

*. Esto es un límite inferior de la significación verdadera.

a. Corrección de significación de Lilliefors

Nota: Elaboración propia, empleando SPSS versión 25

En la tabla 5, se muestra la significancia obtenida para la capacidad original de almacenamiento (Pre-test) y la capacidad deduplicada (Post-Test), obteniéndose un valor menor a 0.05 en el caso de Pre-test, y mayor a 0.05 en el caso de Post-test. Al ser diferentes las significancias consideramos que los datos no siguen una distribución normal y corresponde aplicar una prueba no paramétrica.

Prueba de normalidad: Transferencia de datos (Objetos de almacenamiento) (Hipótesis específica 3)

Tabla 6

Prueba de normalidad para la Transferencia de datos (objetos de almacenamiento), Hipótesis específica 3

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
Transf_Orig	,149	30	,086	,885	30	,004
Transf_Deduplic	,136	30	,168	,941	30	,098

a. Corrección de significación de Lilliefors

Nota: Elaboración propia, empleando SPSS versión 25

En la tabla 6, se muestra la significancia obtenida para la transferencia original de los objetos de almacenamiento (Tiempo, Pre-test) y la transferencia

deduplicada de los mismos objetos (Tiempo, Post-Test), obteniéndose un valor menor a 0.05 en el caso de Pre-test, y mayor a 0.05 en el caso de Post-test. Al ser diferentes las significancias consideramos que los datos no siguen una distribución normal y también corresponde aplicar una prueba no paramétrica.

Prueba de normalidad: Indicador Costo de almacenamiento (Hipótesis específica 4)

Tabla 7

Prueba de normalidad para el Costo de Almacenamiento, Hipótesis específica 4

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
Costo_Orig	,147	30	,097	,945	30	,123
Costo_Dedup	,099	30	,200*	,955	30	,224

*. Esto es un límite inferior de la significación verdadera.

a. Corrección de significación de Lilliefors

Nota: Elaboración propia, empleando SPSS versión 25

En la tabla 7, se muestra la significancia obtenida para la transferencia original de los objetos de almacenamiento (Tiempo, Pre-test) y la transferencia deduplicada de los mismos objetos (Tiempo, Post-Test), obteniéndose un valor mayor a 0.05 en el caso de Pre-test, y también mayor a 0.05 en el caso de Post-test. Al ser ambas significancias mayores a 0.05 consideramos que los datos siguen una distribución normal y correspondería aplicar una prueba paramétrica, sin embargo, dado que el conjunto mayoritario de datos trabajados en la presente investigación no siguen una distribución normal, y gracias a una pauta brindada por el asesor, en este caso también se aplicará una prueba no paramétrica.

Contrastación de Hipótesis

A fin de alcanzar resultados con base analítica se procedió a llevar a cabo la contrastación de las hipótesis.

En las siguientes pruebas trabajaremos con un nivel de significancia o confianza igual a 0,05:

- Si (p-valor) es mayor o igual que 0,05 se acepta la hipótesis nula y se rechaza la hipótesis alterna
- Si (p-valor) es menor que 0,05 se rechaza la hipótesis nula y se acepta la hipótesis alterna

Hipótesis General

H_0: La utilización de algoritmos de criptográficos no impacta positivamente en la gestión del almacenamiento de datos de una empresa de seguridad electrónica

H_a: La utilización de algoritmos de criptográficos impacta positivamente en la gestión del almacenamiento de datos de una empresa de seguridad electrónica

Para poder probar la hipótesis general hemos definido las siguientes hipótesis específicas:

Tabla 8

Prueba de Wilcoxon para los Ratios, Hipótesis general

Estadísticos de prueba[a]

	Ratio_Capac_Dedup - Ratio_Capac_Orig	Ratio_Transf_Dedup - Ratio_Transf_Orig	Ratio_Costo_Dedup - Ratio_Costo_Orig
Z	-4,782[b]	-4,782[b]	-4,782[b]
Sig. asintótica(bilateral)	,000	,000	,000

a. Prueba de rangos con signo de Wilcoxon

b. Se basa en rangos negativos.

Nota: Elaboración propia, empleando SPSS versión 25

De la revisión de la tabla 8, encontramos que la significancia de los Ratios de la Gestión del almacenamiento de datos luego de aplicar los algoritmos criptográficos toma valores de 0.000 para los tres ratios, por lo que rechazamos la hipótesis nula H_0 y aceptamos la hipótesis alterna H_a: La utilización de algoritmos de criptográficos impacta positivamente en la gestión del almacenamiento de datos de una empresa de seguridad electrónica.

Hipótesis específica 1

Se efectuó la contrastación empleando la prueba de Wilcoxon para muestras relacionadas ya que la prueba de Shapiro-Wilk arrojó resultados que correspondían a una prueba no paramétrica

H_0: La utilización de los algoritmos criptográficos en datos con alta similitud no impacta positivamente en la gestión del almacenamiento de datos en una empresa de seguridad electrónica

H_a: La utilización de los algoritmos criptográficos en datos con alta similitud impacta positivamente en la gestión del almacenamiento de datos en una empresa de seguridad electrónica

Tabla 9

Prueba de Wilcoxon para el nivel de Alta similitud, Hipótesis específica 1

Estadísticos de prueba[a]

Impacto del nivel de Alta Similitud	Capac_Dedup - Capac_Orig	TransF_Deduplic - Transf_Orig	Costo_Dedup - Costo_Orig
Z	-4,458[b]	-4,457[b]	-4,457[b]
Sig. asintótica(bilateral)	,000	,000	,000

a. Prueba de rangos con signo de Wilcoxon

b. Se basa en rangos positivos.

Nota: Elaboración propia, empleando SPSS versión 25

En base a la tabla 9, y tomando nota de que el impacto del nivel de Alta Similitud da como resultado un valor de significancia de 0.000 para las tres dimensiones de la variable dependiente, se rechaza la hipótesis nula H_0 y aceptamos la hipótesis alterna H_a: La utilización de los algoritmos criptográficos en datos con alta similitud impacta positivamente en la gestión del almacenamiento de datos en una empresa de seguridad electrónica.

Hipótesis específica 2

Para llevar a cabo la contrastación se empleó la prueba de Wilcoxon para muestras relacionadas, esto debido a los resultados obtenidos en la prueba de Shapiro-Wilk que señalaban una prueba no paramétrica

H_0: La utilización de los algoritmos criptográficos a nivel de bloques no impacta positivamente en la gestión de la capacidad del almacenamiento de datos en una empresa de seguridad electrónica

H_a: La utilización de los algoritmos criptográficos a nivel de bloques impacta positivamente en la gestión de la capacidad del almacenamiento de datos en una empresa de seguridad electrónica

Tabla 10

Prueba de Wilcoxon para Capacidad de Almacenamiento, Hipótesis específica 2

	Estadísticos de prueba[a]
	Capac_Dedup - Capac_Orig
Z	-4,782[b]
Sig. asintótica(bilateral)	,000

a. Prueba de rangos con signo de Wilcoxon

b. Se basa en rangos positivos.

Nota: Elaboración propia, empleando SPSS versión 25

Considerando la tabla 10 y dado que el nivel de significancia es 0.000, se rechaza la hipótesis nula y aceptamos la hipótesis alterna H_a: La utilización de los algoritmos criptográficos a nivel de bloques impacta positivamente en la gestión de la capacidad del almacenamiento de datos en una empresa de seguridad electrónica.

Hipótesis específica 3

Para llevar a cabo la contrastación también se empleó la prueba de Wilcoxon para muestras relacionadas, esto debido a los resultados obtenidos en la prueba de Shapiro-Wilk que señalaban una prueba no paramétrica

H_0: La utilización de los algoritmos criptográficos a nivel de bloques no impacta positivamente en la gestión del ancho de banda del almacenamiento de datos de una empresa de seguridad electrónica

H_a: La utilización de los algoritmos criptográficos a nivel de bloques impacta positivamente en la gestión del ancho de banda del almacenamiento de datos de una empresa de seguridad electrónica

Tabla 11

Prueba de Wilcoxon para la Transferencia de datos (objetos de almacenamiento), Hipótesis específica 3

	Estadísticos de prueba[a]
	Transf_Deduplic - Transf_Orig
Z	-4,782[b]
Sig. asintótica(bilateral)	,000

a. Prueba de rangos con signo de Wilcoxon

b. Se basa en rangos positivos.

Nota: Elaboración propia, empleando SPSS versión 25

En referencia a la tabla 11 y dado que el nivel de significancia es 0.000, se rechaza la hipótesis nula y aceptamos la hipótesis alterna H_a: “La utilización de los

algoritmos criptográficos a nivel de bloques impacta positivamente en la gestión del ancho de banda del almacenamiento de datos de una empresa de seguridad electrónica".

Hipótesis específica 4

Para llevar a cabo la contrastación también se empleó la prueba de Wilcoxon para muestras relacionadas. Aunque para esta dimensión la prueba de normalidad arrojó que los datos si pueden considerarse como parte de una distribución normal, la selección de la prueba de Wilcoxon se hizo por cuanto las otras dimensiones de la variable dependiente no siguen distribuciones normales por lo que les corresponde aplicar pruebas no paramétricas.

H_0: La utilización de los algoritmos criptográficos a nivel de bloques no impacta positivamente en la gestión del costo de almacenamiento de datos de una empresa de seguridad electrónica

H_a: La utilización de los algoritmos criptográficos a nivel de bloques impacta positivamente en la gestión del costo de almacenamiento de datos de una empresa de seguridad electrónica

Tabla 12

Prueba de Wilcoxon para el Costo de Almacenamiento, Hipótesis específica 4

	Estadísticos de prueba[a]
	Costo_Dedup - Costo_Orig
Z	-4,782[b]
Sig. asintótica(bilateral)	,000

a. Prueba de rangos con signo de Wilcoxon

b. Se basa en rangos positivos.

Nota: Elaboración propia, empleando SPSS versión 25

Teniendo como marco la tabla 12 y dado que el nivel de significancia es 0.000, se rechaza la hipótesis nula y aceptamos la hipótesis alterna H_a: “La utilización de los algoritmos criptográficos a nivel de bloques impacta positivamente en la gestión del costo de almacenamiento de datos de una empresa de seguridad electrónica”.

Ya que hemos contrastado exitosamente las hipótesis específicas, podemos aceptar la hipótesis general alternativa H_a: “La utilización de algoritmos de criptográficos impacta positivamente en la gestión del almacenamiento de datos de una empresa de seguridad electrónica”

V. DISCUSIÓN

La investigación tuvo la finalidad de determinar el impacto de la aplicación de algoritmos criptográficos en la gestión del almacenamiento de datos de una empresa de seguridad electrónica en Lima. El diseño pre-experimental nos permitió medir y registrar las dimensiones de las variables identificadas en el estudio, en base a las cuales se calcularon los indicadores correspondientes encontrándose una significancia estadística, o p-value, de 0.000, menor que el valor 0.05 que se estableció como criterio para la aceptación o rechazo de la hipótesis nula. El valor encontrado, de 0.000, permitió rechazar la hipótesis nula y aceptar la hipótesis alternativa de que la aplicación de los algoritmos criptográficos impacta positivamente en la gestión del almacenamiento de datos de la empresa que se tomó como caso de estudio.

Los resultados mostrados, son similares a los obtenidos en la tesis doctoral de Larico (2020), donde su hipótesis de trabajo fue aceptada con el valor de significancia (valor critico observado) de 0.003, que fue bastante menor que el criterio de aceptación establecido de 0.05, pero muy cercano al presente trabajo, donde la hipótesis de trabajo se aceptó con una significancia de 0.000. En ambos estudios se concluyó que los cambios e innovaciones en la gestión de recursos informáticos alcanza un alto nivel de impacto, positivo, en los sistemas donde se aplican. La diferencia entre el alcance de las investigaciones radica en que en la tesis citada se evaluó el impacto de emplear arquitecturas hiperconvergentes para mejorar los parámetros de gestión de los recursos de TI mientras que en nuestra investigación se trabajó con una arquitectura estándar de servidores.

Uno de los resultados específicos obtenidos es el cuadro de ratios de Capacidad de almacenamiento para los treinta objetos de almacenamiento analizados, que se ha mostrado en la figura 13. Este ratio resulta de dividir la capacidad de almacenamiento empleada antes de la deduplicación entre la capacidad empleada luego de aplicar la deduplicación y, para nuestra investigación se obtuvo un rango de variación de los ratios desde 1.13 hasta 4.50, y un ratio

promedio de 2.1454. Pasaremos a comparar estos valores con los obtenidos en otras investigaciones

En Fehér et al. (2020), se han obtenido resultados mejores que los nuestros, con un ratio de Capacidad máximo de 10 frente a nuestro ratio de 4.50. En esta investigación se trabajó aplicando algoritmos criptográficos mediante técnicas de deduplicación a un solo tipo de objeto de almacenamiento, las lecturas efectuadas en medidores inteligentes de energía que luego eran transmitidas como mensajes y almacenadas en servicios comerciales de computación en la nube. A priori podemos teorizar que el alto grado de similitud de los datos, es decir las lecturas, es lo que permite alcanzar el nivel de ratio reportado.

Comparativamente, con la investigación de Cheng et al. (2021), en nuestra investigación hemos obtenido un ratio mínimo de Capacidad que es inferior, de 1.13 frente a 1.78 que es mejor para el estudio citado, pero también hemos alcanzado un ratio máximo superior con un valor de 4.50 versus 2.86. A nivel de ratios promedio, en el paper citado se obtuvo 2.1978 frente a 2.1454 en nuestro caso, que muestra una ligera ventaja para Cheng. Vemos que los valores reportados en el mencionado artículo son comparables con los obtenidos en nuestra investigación. Cabe señalar que para fines de comparación hemos convertido a ratios los porcentajes reportados en el paper: 44% (1.78), 54.5% (2.1978) y 65% (2.86).

Otro contraste con la literatura especializada lo tenemos en el paper de Venish et al. (2016), donde obtuvieron un mejor valor para el ratio mínimo de Capacidad, con un valor de 2 frente a 1.13 para nuestro caso, mientras que su ratio máximo de Capacidad de 3.23 fue inferior a nuestro resultado de 4.50. En la citada investigación se aplicó técnicas de deduplicación a nivel de la unidad de datos bloque o fragmento para reducir la utilización de capacidad de almacenamiento empleada, siendo en ese sentido investigaciones similares, pero aplicaron un enfoque más experimental pues variaron los tamaños de bloque empleados para generar variedad de resultados: 64 KB (ratio 2, 35%), 32 KB (ratio 3.03, 33%), 16 KB (3.125, 32%) y 8 KB (3.23, 31%). Está variación experimental a nivel de bloques no estuvo considerada dentro del alcance de la presente investigación, pero representa una interesante oportunidad para trabajos futuros.

Un caso de comparación de mucho interés para nosotros es el de Bharde et al. (2018), cuyos resultados son ligeramente mejores que los que hemos obtenido. En la mencionada investigación se reporta un ratio mínimo de Capacidad de 1.61 que es mejor que el nuestro de 1.13, asimismo también se reporta un ratio máximo de Capacidad de 2.94 pero este valor es inferior al ratio máximo de Capacidad que hemos alcanzado de 4.50. En el caso que se menciona se aplicó algoritmos criptográficos contenidos en métodos de deduplicación para identificar y remover similitudes en los datos que eran capturados en el borde de la red (Netwok Edge) donde se despliegan aplicaciones tipo Internet de las Cosas (IoT). Lo interesante del caso es que se trabajó con objetos de almacenamiento tipo datasets de video, a nivel de tramas o frames, lo cual es muy similar al tipo de objeto de almacenamiento predominante en nuestro estudio.

En el caso de la investigación de Aronovich et al. (2016), se han obtenido mejores ratios que en nuestra investigación, reportándose un mejor ratio mínimo de Capacidad de 3.88 frente a un valor de 1.13 para nosotros, también un mejor ratio máximo de Capacidad de 6.58 contra nuestro valor de 4.50. En la investigación que se compara aplicaron algoritmos criptográficos tipo Hash (o de identidad, como lo denomina Aronovich) a objetos de almacenamiento constituidos por sistemas operativos variados (Centos-5.3-i386-server, Freebsd-6.4-i386 y Fedora-fc6-i386), lo cual explicaría los mejores resultados obtenidos ya que en nuestro caso la población de estudio estuvo compuesta por otro tipo de objetos de almacenamiento, mayoritariamente secuencias de video.

El estudio de Almrezeq et al. (2021), obtuvo resultados mixtos. El ratio mínimo de Capacidad fue de 1.34 que es mejor que el 1.13 que encontramos en nuestras pruebas. Para el caso del ratio máximo de Capacidad de la citada investigación se obtuvo 2.66 contra 4.50 en nuestro caso, lo cual nos favorece. El escenario de aplicación citado por el paper fue de almacenamiento de datos en servicios de computación en la nube, donde se dio preferencia a la seguridad y al rendimiento, factores que, según lo revisado en otras fuentes bibliográficas, impacta en el volumen de almacenamiento de datos de forma directa.

Si comparamos nuestros resultados con los de Park et al. (2016), vemos que en dicha investigación se obtuvo mejores resultados, con un ratio máximo de

Capacidad de 7.14 frente a un valor de 4.50 en esta tesis. La diferencia encontrada es importante, por lo que es oportuno indicar que en el trabajo citado se aplicó un algoritmo selectivo de deduplicación de datos, donde solo se trabajaba con datos que cumplían con un patrón de acceso específico, principalmente que eran frecuentemente consultados, con lo que lograron un mayor rendimiento en tiempo de lectura y también ratio de Capacidad pues no se calculaba el indicador sobre todos los datos almacenados sino solo sobre los seleccionados por el algoritmo.

Otro de los resultados obtenidos en nuestra investigación es el concerniente a los ratios de transferencia, resultantes de dividir el tiempo empleado para transferir un objeto de almacenamiento original entre el tiempo empleado para transferir el mismo objeto tras haber aplicado las técnicas de deduplicación. Estos ratios reflejan la reducción de los tiempos de transferencia de los datos deduplicados, que constituyen una mejora en la gestión del ancho de banda disponible para el sistema de almacenamiento, por ejemplo, en casos de respaldo (backup). Nuestros ratios de tiempos de transferencia de los objetos de almacenamiento tienen un rango de variación que va desde 1.27 hasta 4.74, siendo el valor promedio 2.4309. A fin de valorar la bondad de estos resultados, los compararemos con los obtenidos en la literatura.

En la investigación de Morey et al. (2016), se obtuvo un ratio promedio inferior a nuestro caso en un primer ensayo, de 1.42 frente a 2.4309, mejorándose ligeramente en un segundo ensayo pero aún por debajo de nuestro valor promedio, 1.93 contra 2.4309. En efecto, en el paper referido, el primer ensayo se desarrolló aplicando técnicas de deduplicación sobre objetos de almacenamiento en el sistema de archivos ZFS, y el segundo ensayo se llevó a cabo almacenando objetos en el sistema de archivos SDFS, mientras que nuestra investigación aplicó las técnicas de deduplicación a objetos almacenados en el sistema de archivos NTFS. Es oportuno mencionar que el paper reporta sus resultados de forma porcentual, por lo que han sido convertidos a ratios para fines de comparación: 29.45% (1.42) y 48.37% (1.93) respectivamente para ZFS y SDFS.

La investigación de Tyj et al. (2019), obtuvo mejores resultados que los del presente estudio, con un ratio mínimo de Capacidad de 5.88 frente a un valor de 4.50 en nuestro caso, y un ratio máximo de Transferencia de 9.76 contra el valor 4.74 para nosotros. Ambos resultados superan nuestros ratios pero debe hacerse notar que esa investigación aplicaron un algoritmo adaptativo para la deduplicación y que la población de estudio estuvo constituida por instancias de máquinas virtuales en ejecución, lo cual supone, al menos conceptualmente, una ventaja frente a nuestro estudio donde se utilizó un algoritmo estándar para deduplicación y la población estuvo compuesta por objetos de almacenamiento provenientes de cámaras de seguridad.

En Prathima et al. (2022), se alcanzaron ratios de Transferencia ligeramente mejores que los que hemos obtenido. El intervalo de variación de los ratios estuvo entre 2.51 y 4.94 para la referida investigación frente a un intervalo de 1.24 a 4.74 en nuestro caso. El escenario de aplicación de la investigación fue el almacenamiento de datos provenientes de redes de sensores de Internet de las Cosas (IoT), donde podemos postular, a priori, que muchas lecturas de los sensores se repetirán favoreciendo la aplicación de técnicas de deduplicación. Otra diferencia encontrada entre ambos estudios es que en la investigación citada los ratios se obtuvieron cuando la cantidad de objetos transferidos fue realmente muy grande, arriba de las 10,000 instancias.

En el trabajo de Viji et al. (2019), los resultados reportados son mejores que los de nuestro estudio, alcanzando un valor representativo de 3.125 para el ratio de Capacidad, que podemos comparar con nuestro ratio de Capacidad promedio que es de 2.1454. El escenario de estudio del trabajo estuvo compuesto principalmente por lo que el autor denomina almacenamiento primario, vale decir discos duros convencionales, discos USB y unidades de cinta, comparativamente nuestro caso abordó el almacenamiento en discos convencionales. Un resultado adicional que menciona Viji es el ratio de Capacidad de los sistemas de respaldo de datos (Backups), para los cuales reporta un valor representativo de 5.88, que es mejor que nuestro valor promedio de 2.1454. Consideramos conveniente anotar que los resultados se reportaron de forma porcentual, por lo que se convirtieron a ratios para compararlos: 68% (3.125) y 83% (5.88).

También tenemos un grupo de resultados obtenidos en nuestra investigación que son los ratios de Costos de almacenamiento, resultantes de dividir el costo de almacenamiento del objeto original entre el costo de almacenamiento del objeto deduplicado. Como ya hemos venido comentando, aquí nuevamente vemos que nuestros resultados varían en el rango de 1.13 a 4.47, siendo el valor promedio 2.1812. Igualmente compararemos nuestros resultados con los ratios que reporta la literatura especializada.

En el caso de la investigación de Daniel et al. (2021), se obtiene un ratio de Costo de almacenamiento muy moderado de 1.25 frente a 2.1812 que arroja nuestro estudio. En la citada investigación se analizó el almacenamiento externo de datos, en servicios comerciales de computación en la nube, y se buscó proponer técnicas que permitan almacenar de forma segura los datos, pero también que sean eficientes, aplicándose tanto técnicas de deduplicación como de aseguramiento de los datos, Aquí es preciso señalar que el trabajo de Daniel aplicó consideraciones de seguridad para el almacenamiento, las cuales no han sido consideradas en la presente investigación. Podemos suponer, con buena base, que la adición de medidas de seguridad impacta incrementando el tamaño de los datos almacenados y, por ende, en el ratio de Costo, sin embargo, el valor presentado por Daniel sirve como referente y muestra una vez más que los valores alcanzados en este trabajo son compatibles con los de la literatura y también competitivos. Como en otros casos, el paper reporta porcentualmente su nivel de reducción de costos, que se transformó a ratio para fines comparativos: 20% (1.25).

VI. CONCLUSIONES

Primera:

Habiéndose alcanzado el Objetivo general, se concluyó que la utilización de algoritmos criptográficos impacta positivamente en la gestión del almacenamiento de datos de una empresa de seguridad electrónica, conclusión que se deriva de haber validado la Hipótesis general con la prueba de Wilcoxon con una significancia de 0.000.

Este impacto positivo se logra ya que los algoritmos referidos, específicamente los de tipo Hash, permiten identificar y remover datos duplicados con el consiguiente ahorro de capacidad de almacenamiento, tiempo de transferencia de datos y costo de almacenamiento.

Segunda:

Luego de lograrse el Objetivo específico 1, se determinó que la aplicación de algoritmos criptográficos en datos con alta similitud impacta positivamente en la gestión del almacenamiento de datos en una empresa de seguridad electrónica, conclusión a la que llegamos tras haber validado la primera hipótesis específica aplicando la prueba de Wilcoxon para muestras relacionadas (no paramétricas) con una significancia de 0.000.

El impacto positivo se logra al aplicar los algoritmos criptográficos a la unidad de datos adecuada, sea archivo o bloque, para lograr la deduplicación, es decir la identificación y remoción de datos duplicados.

Tercera:

Al alcanzarse el Objetivo específico 2 se ha establecido que la utilización de algoritmos criptográficos a nivel de bloques impacta positivamente en la gestión de la capacidad del almacenamiento de datos en una empresa de seguridad electrónica., pues se ha verificado la segunda hipótesis específica mediante el

empleo de la prueba de Wilcoxon para muestras relacionadas obteniéndose una significancia de 0.000

En efecto, debido a que los algoritmos criptográficos permiten deduplicar los datos almacenados, se reduce la capacidad de almacenamiento empleada.

Cuarta:

Mediante el cumplimiento del Objetivo específico 3 se logró validar que la utilización de algoritmos criptográficos a nivel de bloques impacta positivamente en la gestión del ancho de banda del almacenamiento de datos en una empresa de seguridad electrónica. Esta conclusión se desprende de los resultados obtenidos al aplicar la prueba de muestras relacionadas de Wilcoxon para validar la tercera hipótesis específica, donde la referida prueba arrojó una significancia de 0.000.

Esto se explica al considerar que el volumen de datos almacenado se reduce, y su transferencia se hace en menos tiempo, permitiendo transferir mayor cantidad de datos con los mismos recursos de ancho de banda, mejorándose la gestión.

Quinta:

Al alcanzar el Objetivo específico 4 se ha concluido que la utilización de algoritmos criptográficos a nivel de bloques impacta positivamente en la gestión del costo de almacenamiento de datos en una empresa de seguridad electrónica. puesto que al validarse la cuarta hipótesis específica con la prueba de Wilcoxon para muestras relacionadas se obtiene una significancia de 0.000

Esta conclusión final recoge el hecho de que la deduplicación reduce la cantidad de datos almacenados, lo cual no solo reduce la capacidad de almacenamiento empleada, los tiempos de transferencia utilizados sino, en general, las tareas afines que lleva a cabo un sistema de almacenamiento de datos como, por ejemplo, un escaneo de seguridad para verificar si los datos tienen contenido malicioso tales como virus, gusanos u otro tipo de malware, y todas actividades redundan en la reducción de costo de almacenamiento.

VII. RECOMENDACIONES

Primera:

Se recomienda a la subgerente el empleo de algoritmos criptográficos, a través de la aplicación de técnicas de deduplicación de datos, para mejorar positivamente la gestión del almacenamiento de datos de sus sistemas de información.

Segunda:

Con relación al tipo de datos u objetos de almacenamiento, se recomienda a la subgerente la aplicación de algoritmos criptográficos, a través de las técnicas de deduplicación, en aquellos datos que presenten alta similitud, tales como son los videos generados por el monitoreo de las cámaras de seguridad. Posteriormente pueden evaluarse nuevos objetos de almacenamiento, candidatos para la aplicación de estos algoritmos y técnicas, llevando a cabo un análisis preliminar de la similitud.

Tercera:

En lo referente a la gestión de la capacidad de almacenamiento de datos, se recomienda a la subgerente la aplicación de algoritmos criptográficos para mejorar las características operativas, ya que las técnicas de deduplicación, donde se emplean estos algoritmos, reducirán la capacidad de almacenamiento que se aprovisione.

Cuarta:

En base a los resultados obtenidos para tiempos de transferencia de datos y el empleo de ancho de banda, recomendamos a la subgerente la aplicación de algoritmos de encriptación, mediante la deduplicación, para mejorar positivamente la gestión del ancho de banda del almacenamiento de datos pues estos tiempos y consumo de ancho de banda se reducirán.

Quinta:

Refiriéndonos a la gestión del costo de almacenamiento de datos, recomendamos a la subgerente que adopte los algoritmos de encriptación, mediante la aplicación de la deduplicación de datos, a fin de mejorar sus indicadores de costo, los cuales se reducirán ya que, al trabajar con menor volumen de datos, todas las actividades relacionadas con el manejo de estos datos también reducen su consumo de recursos y por ende el costo de almacenamiento.

REFERENCIAS

Adhab, A. H., Hussien, N. A. (2022). Techniques of Data Deduplication for Cloud Storage: A Review.

Agarwala, S., Jadav, D., & Bathen, L. A. (2011, July). iCostale: adaptive cost optimization for storage clouds. In 2011 IEEE 4th International Conference on Cloud Computing (pp. 436-443). IEEE.

Al Azad, M. W., Mastorakis, S. (2022). The promise and challenges of computation deduplication and reuse at the network edge. IEEE Wireless Communications.

Almrezeq, N., Humayun, M., El-Aziz, A., Jhanjhi, N. (2021). An Enhanced Approach to Improve the Security and Performance for Deduplication. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 12(6), 2866-2882.

Álvarez-Risco, A. (2020). Clasificación de las investigaciones.

Arias Gonzáles, J. L. (2020). Técnicas e instrumentos de investigación científica.

Aronovich, L., Asher, R., Harnik, D., Hirsch, M., Klein, S., & Toaff, Y. (2016). Similarity based deduplication with small data chunks. *Discrete Applied Mathematics*, *212*, 10-22. https://doi.org/10.1016/j.dam.2015.09.018

Bharde, M., Bhattacharya, S., Shree, D. D. (2018). {Store-Edge}{RippleStream}: Versatile Infrastructure for {IoT} Data Transfer. In USENIX Workshop on Hot Topics in Edge Computing (HotEdge 18).

Chen, Y. F. R. (2015). The growing pains of cloud storage. *IEEE Internet Computing*, *19*(1), 4-7.

Chen, L., Xiang, F., Sun, Z. (2021). Image Deduplication Based on Hashing and Clustering in Cloud Storage. KSII Transactions on Internet and Information Systems (TIIS), 15(4), 1448-1463.

Cheng, G., Guo, D., Luo, L., Xia, J., Gu, S. (2021). LOFS: A Lightweight Online File Storage Strategy for Effective Data Deduplication at Network Edge. IEEE Transactions on Parallel and Distributed Systems, 33(10), 2263-2276.

Daniel, E., Susila, N., Durga, S. (2021). DSDA: A desirable and secure deduplication approach for outsourced data in cloud environment. Journal of Contemporary Issues in Business and Government Vol, 27(3).

Diao, K., Papapanagiotou, I., Hacker, T. J. (2016, December). HARENS: Hardware accelerated redundancy elimination in network systems. In 2016 IEEE International Conference on Cloud Computing Technology and Science (CloudCom) (pp. 237-244). IEEE.

Dutch, M. (2008, June). Understanding data deduplication ratios. In *SNIA Data Management Forum* (Vol. 7).

Espinoza, D. M. (2019). Consideraciones éticas en el proceso de una publicación científica. *Revista Médica Clínica Las Condes*, *30*(3), 226-230.

Fehér, M., Yazdani, N., Hansen, M. T., Vester, F. E., Lucani, D. E. (2020, December). Smart meter data compression using generalized deduplication. In GLOBECOM 2020-2020 IEEE Global Communications Conference (pp. 1-6). IEEE.

He, Q., Shao, B., Zhang, W. (2020). Data deduplication technology for cloud storage. Tehnički vjesnik, 27(5), 1444-1451.

Kaur, R., Bhattacharya, J., Chana, I. (2022). Deep CNN based online image deduplication technique for cloud storage system. Multimedia Tools and Applications, 1-34.

Larico Uchamaco, G. R. (2020). Sistemas hiperconvergentes para mejorar la gestión tecnológica en centros de datos de la Universidad Nacional Amazónica de Madre De Dios. Tesis Doctoral. Universidad Federico Villarreal.

Liao, Y. “Dataset deduplication with Datamodels” (2022). Master Thesis. Massachusetts Institute of Technology.

López-Roldán, P., Fachelli, S. (2017). El diseño de la muestra. Metodología de la investigación social cuantitativa.

Marín Y. “Metodología para la gestión de la calidad de los datos empleando un enfoque data driven para implementar procesos de evaluación y mejoramiento de los datos en iniciativas de gestión de datos maestros” (2022). Tesis de Maestría. Universidad Nacional de Colombia.

Masid Blanco, O. (2017). La metáfora lingüística en español como lengua extranjera (ELE). Estudio pre-experimental en tres niveles de competencia.

Mohamed, S., Wang, Y. (2021). A survey on novel classification of deduplication storage systems. Distributed and Parallel Databases, 39(1), 201-230.

Morey, D. G., Pérez, E. (2016) Análisis de métodos de deduplicación de datos aplicados a repositorios de Software Libre.

Ni, F. “Designing highly-efficient deduplication systems with optimized computation and I/O operations” (2019). Doctoral dissertation. The University of Texas at Arlington.

Nicaragua, Estelí. "Metodología de la investigación e investigación aplicada para Ciencias Económicas y Administrativas." *Revista de La Universidad Autónoma* (2018): 1-89.

Ocsa Mamani, A. V. (2018). Métodos semánticos para la recuperación de información en grandes volúmenes de datos: Una Arquitectura Escalable y Eficiente. Tesis Doctoral. Universidad Nacional de San Agustín.

Otero, A. (2018). Enfoques de investigación. *Métodos para el diseño urbano–Arquitectónico*.

Palacios, R., Delgado, V. (2006, January). Introducción a la Criptografía: tipos de algoritmos. En *anales de mecánica y electricidad*.

Park, S., Park, C. (2016). Offline selective data deduplication for primary storage systems. IEICE TRANSACTIONS on Information and Systems, 99(2), 370-382.

Pérez, V. “Solución para la gestión de almacenamiento de datos en las instituciones cubanas” (2018). Tesis de Maestría. Universidad de las Ciencias Informáticas.

Prathima, C., Muppalaneni, N. B., Kharade, K. G. (2022). Deduplication of IoT Data in Cloud Storage. In Machine Learning and Internet of Things for Societal Issues (pp. 147-157). Springer, Singapore.

Qiu, W. “Memory deduplication on serverless systems” (2021). Master Thesis. ETZ Zurich (Swiss Federal Institute of Technology Zurich)

Quiña, G. N., Yoo, S. G., Guarda, T. (2019). Recuperación de Datos en Dispositivos de Almacenamiento SSD Utilizando File Carving. Revista Ibérica de Sistemas e Tecnologias de Informação, (E17), 490-498.

Ramos, C. A. (2020). Los alcances de una investigación. *CienciAmérica: Revista de divulgación científica de la Universidad Tecnológica Indoamérica*, *9*(3), 1-6.

Ramos, C. A. (2021). Diseños de investigación experimental. CienciAmérica: Revista de divulgación científica de la Universidad Tecnológica Indoamérica, 10(1), 1-7.

Reio, T. G. (2016). Nonexperimental research: Strengths, weaknesses and issues of precision. *European Journal of Training and Development*.

Rodríguez, M., Mendivelso, F. (2018). Diseño de investigación de corte transversal. *Revista médica sanitas*, 21(3), 141-146.

Roldán, F., Sánchez, J. “Repotenciación del Sistema Informático de Denuncias Policiales para producir de manera eficiente información estadística en la Policía Nacional del Perú” (2020). Tesis de Maestría. PUCP. http://hdl.handle.net/20.500.12404/21735

Saharan, S., Somani, G., Gupta, G., Verma, R., Gaur, M. S., Buyya, R. (2020). QuickDedup: Efficient VM deduplication in cloud computing environments. Journal of Parallel and Distributed Computing, 139, 18-31.

Salazar, L. D. (2018). Tecnologías de Almacenamiento en Centro de Datos. *Tecnología Vital*, *2*(4).

Scanlon, M. (2016, August). Battling the digital forensic backlog through data deduplication. In 2016 sixth international conference on innovative computing technology (INTECH) (pp. 10-14). IEEE.

Shin, H., Koo, D., Hur, J. (2022). Secure and Efficient Hybrid Data Deduplication in Edge Computing. ACM Transactions on Internet Technology (TOIT).

Sobti, R., Geetha, G. (2012). Cryptographic hash functions: a review. *International Journal of Computer Science Issues (IJCSI)*, *9*(2), 461.

Storage Networking Industry Association [SNIA]. (22 de octubre de 2022). *SNIA Dictionary*. https://www.snia.org/education/online-dictionary/term/storage-efficiency

Storer, M. W., Greenan, K., Long, D. D., Miller, E. L. (2008, October). Secure data deduplication. In Proceedings of the 4th ACM international workshop on Storage security and survivability (pp. 1-10).

Torres, P. A. (2016). Acerca de los enfoques cuantitativo y cualitativo en la investigación educativa cubana actual. *Atenas*, *2*(34), 1-15.

Tyj, N. M. (2019). Adaptive deduplication of virtual machine images using AKKA stream to accelerate live migration process in cloud environment. Journal of Cloud Computing, 8(1), 1-12.

Ur Rehman, M. H., Chang, V., Batool, A., Wah, T. Y. (2016). Big data reduction framework for value creation in sustainable enterprises. *International journal of information management*, *36*(6), 917-928.

Venish, A., Siva Sankar, K. (2016). Study of chunking algorithm in data deduplication. In *Proceedings of the International Conference on Soft Computing Systems* (pp. 13-20). Springer, New Delhi.

Ventura-León, J. L. (2017). ¿Población o muestra?: Una diferencia necesaria. *Revista cubana de salud pública*, *43*(4), 0-0

Viji, D., Revathy, S. (2019, July). Various data deduplication techniques of primary storage. In 2019 International conference on communication and electronics systems (ICCES) (pp. 322-327). IEEE.

Villa, E. A. (2018). Modelo de evaluación de plataformas tecnológicas para el almacenamiento y procesamiento de grandes datos: Caso en salud asistencial (Doctoral dissertation, Universidad EAFIT).

Wu, H., Wang, C., Fu, Y., Sakr, S., Zhu, L., Lu, K. (2017). Hpdedup: A hybrid prioritized data deduplication mechanism for primary storage in the cloud. *arXiv preprint arXiv:1702.08153*.

Xia, W., Jiang, H., Feng, D., Douglis, F., Shilane, P., Hua, Y., Zhou, Y. (2016). A comprehensive study of the past, present, and future of data deduplication. *Proceedings of the IEEE*, 104(9), 1681-1710.

Xu, R., Zhao, K., Zhang, P., Yuan, D., Xie, Y., & Yang, Y. (2017, June). A novel data set importance based cost-effective and computation-efficient storage strategy in the cloud. In 2017 IEEE International Conference on Web Services (ICWS) (pp. 122-129). IEEE.

Zhou, R., Liu, M., Li, T. (2013, September). Characterizing the efficiency of data deduplication for big data storage management. In *2013 IEEE international symposium on workload characterization (IISWC)* (pp. 98-108). IEEE.

Zhou, Y., Deng, Y., Yang, L. T., Yang, R., Si, L. (2018). LDFS: A low latency in-line data deduplication file system. IEEE Access, 6, 15743-15753.

ANEXOS

Anexo 1: Matriz de Operacionalización de variables

Aplicación de algoritmos criptográficos en la gestión de almacenamiento de datos en una empresa de seguridad electrónica, Lima 2023

Variable	Definición conceptual	Definición operacional	Dimensión de la variable	Indicadores	instrumento	Fórmula	Escala de medición
Aplicación de algoritmos criptográficos	La deduplicación de datos (DD) es una tecnología que ahorra espacio de almacenamiento. Puede identificar conjuntos de datos redundantes al compararlos mediante sus huellas hash. De esa forma puede mantener solo una copia de los datos y crear punteros lógicos que reemplacen las otras copias duplicadas. El volumen de datos es reducido mediante la DD, reduciéndose también el espacio de almacenamiento, el ancho debanda empleado para transferir los datos y el costo de almacenamiento de los datos. La DD puede ser empleada en cualquier sistema donde se almacene o transfiera datos. (Adhab, A. H., & Hussien, N. A, 2022)	A los objetos de almacenamiento (archivo o bloques) se les calcula la huella digital "hash", que se compara con las huellas hash de otros objetos de almacenamiento. Si las huellas coinciden entonces se acepta que los objetos de almacenamiento son iguales, reemplazándose las copias del objeto de almacenamiento por un enlace lógico a la primera copia. De esta forma se ahorra espacio de almacenamiento. (Xia et al, 2016)	Unidad de datos para aplicación del algoritmo criptográfico	Unidad de datos	Ficha de Registro	Selección de unidad de datos	Archivo Fragmento o Bloque
			Nivel de similitud de los datos a que se aplica el algoritmo criptográfico	Similitud		Alta > 30% 30 > =Media > 10 10 >= Baja	Alta Media Baja
Variable	**Definición conceptual**	**Definición operacional**	**Dimensión de la variable**	**Indicadores**	**Instrumento**	**Fórmula**	**Escala de medición**
Gestión del almacenamiento de datos	La eficiencia del almacenamiento de datos se refleja en la reducción del consumo de espacio de almacenamiento y los tiempos de transferencia de estos datos. La deduplicación de datos (DD) permite que un servidor de almacenamiento pueda maximizar la eficiencia del almacenamiento al almacenar una sola copia de los datos redundantes y remover otros duplicados. (Shin, H., Koo, D., & Hur, J., 2022)	La eficiencia del sistema de almacenamiento se puede medir estableciendo la relación (ratio) entre el uso de recursos al almacenar datos de forma estándar y el uso de los mismos recursos al almacenar datos de forma optimizada utilizando deduplicación de datos. Los recursos relevantes a ser medidos son el volumen de datos almacenados, al ancho de banda o tiempo empleado para transferencias de los datos y el costo de almacenamiento de datos. (ZHou et al., 2013)	Capacidad de almacenamiento	Ratio de capacidad (Original/deduplicado)	Ficha de registro	RCA = (Capac_Orig)/(Capac_dedup)	Razón
			Capacidad de transferencia de datos	Ratio de transferencia (Original/deduplicado)		RBW = (BW_Orig)/(BW_dedup)	
			Costo de almacenamiento	Ratio de costo (Original/deduplicado)		RCO = (Costo_Orig)/(Costo_Dedup)	

Anexo 2: Matriz de Consistencia

Matriz de Consistencia				
Aplicación de Algoritmos Criptográficos en la Gestión de Almacenamiento de Datos en una Empresa de Seguridad Electrónica, Lima 2023				
Problema general	**Objetivo general**	**Hipótesis general**	**Variable Independiente**	**Metodología**
¿Cómo impacta el uso de algoritmos criptográficos en la gestión de almacenamiento de datos de una empresa de seguridad electrónica?	Determinar el impacto que tiene el uso de algoritmos criptográficos en la gestión del almacenamiento de datos en una empresa de seguridad electrónica.	La utilización de algoritmos de criptográficos impacta positivamente en la gestión del almacenamiento de datos de una empresa de seguridad electrónica	Aplicación de algoritmos criptográficos	**Tipo de investigación:** Aplicada **Diseño de Investigación:** Pre - Experimental **Enfoque:** Cuantitativo **Nivel:** Explicativo **Población:** 30 Objetos de almacenamiento, representativos de los datos que se almacena en la empresa de seguridad. **Muestra**: La muestra que se utilizará será igual al tamaño de la población. **Técnicas:** Observación directa. **Instrumentos:** **F**ichas de registro.
Problemas específicos	**Objetivos específicos**	**Hipótesis específicas**	**Variable dependiente**	
¿Cómo impacta el uso de algoritmos criptográficos en datos con alta similitud en la gestión de almacenamiento de datos de una empresa de seguridad electrónica?	Determinar el impacto que tiene el uso de algoritmos criptográficos en datos con alta similitud en la gestión del almacenamiento de datos en una empresa de seguridad electrónica.	La utilización de los algoritmos criptográficos en datos con alta similitud impacta positivamente en la gestión del almacenamiento de datos en una empresa de seguridad electrónica.	Gestión del almacenamiento de datos	
¿Cómo impacta el uso de algoritmos criptográficos a nivel de bloques en la gestión de la capacidad de almacenamiento de datos en una empresa de seguridad electrónica?	Determinar el impacto que tiene el uso de algoritmos criptográficos a nivel de bloques en la gestión de la capacidad de almacenamiento de datos en una empresa de seguridad electrónica.	La utilización de los algoritmos criptográficos a nivel de bloques impacta positivamente en la gestión de la capacidad del almacenamiento de datos en una empresa de seguridad electrónica.		
¿Cómo impacta el uso de algoritmos criptográficos a nivel de bloques en la gestión de ancho de banda de almacenamiento de datos en una empresa de seguridad electrónica?	Determinar el impacto que tiene el uso de algoritmos criptográficos a nivel de bloques en la gestión del ancho de banda del almacenamiento de datos de una empresa de seguridad electrónica.	La utilización de los algoritmos criptográficos a nivel de bloques impacta positivamente en la gestión del ancho de banda del almacenamiento de datos de una empresa de seguridad electrónica		
¿Cómo impacta el uso de algoritmos criptográficos a nivel de bloques en la gestión del costo de almacenamiento de datos de una empresa de seguridad electrónica?	Determinar el impacto que tiene el uso de algoritmos criptográficos a nivel de bloques en la gestión del costo del almacenamiento de datos en una empresa de seguridad electrónica	La utilización de los algoritmos criptográficos a nivel de bloques impacta positivamente en la gestión del costo de almacenamiento de datos de una empresa de seguridad electrónica		

Anexo 3 – Ficha de Registro (3.1 Ficha de Registro para Unidad de datos)

UCV Universidad César Vallejo	**INSTRUMENTO:** **Ficha de Registro**	**CÓDIGO:** UCV- ING SISTEMAS-TI-001.1	
	Recolección de datos	**Página:**	1 de 1
Investigador	Torres Paredes, Carlos Martin	**Tipo de Prueba**	Observación
Fecha Inicio	07/11/2022	**Fecha Final**	21/11/2022
Variable	**Indicador**	**Medida**	**Fórmula**
Aplicación de algoritmos criptográficos	Unidad de datos	Tipo: - Archivo - Bloque	Selección de unidad de datos

O

ITEM	FECHA	TIPO (PRE_TEST)	FECHA	TIPO (POST_TEST)
1				

Dr. Marlon Acuña Benites
DNI: 42097456
Ing. de Sistemas / Investigador

Anexo 3 – Ficha de Registro (3.2 Ficha de Registro para Similitud)

UCV Universidad César Vallejo	**INSTRUMENTO:** **Ficha de Registro**	**CÓDIGO:** UCV- ING SISTEMAS-TI-001.1	
	Recolección de datos	**Página:**	1 de 1
Investigador	Torres Paredes, Carlos Martin	**Tipo de Prueba**	Observación
Fecha Inicio	07/11/2022	**Fecha Final**	21/11/2022
Variable	**Indicador**	**Medida**	**Fórmula**
Aplicación de algoritmos criptográficos	Similitud	Nivel: - Alta - Media - Baja	> 30% > Alta >= 20% >= 20% > Media > 10% Baja >= 10%)

ITEM	FECHA	NIVEL (PRE_TEST)	FECHA	NIVEL (POST_TEST)
1				

Dr. Marlon Acuña Benites
DNI: 42097456
Ing. de Sistemas / Investigador

Anexo 3 – Ficha de Registro (3.3 Ficha de Registro para Capacidad de Almacenamiento)

UCV UNIVERSIDAD CÉSAR VALLEJO	**INSTRUMENTO:** **Ficha de Registro**	**CÓDIGO:** UCV- ING SISTEMAS-TI-001.1	
	Recolección de datos	**Página:**	1 de 1
Investigador	Torres Paredes, Carlos Martin	**Tipo de Prueba**	Observación
Fecha Inicio	07/11/2022	**Fecha Final**	21/11/2022
Variable	**Indicador**	**Medida**	**Fórmula**
Gestión de almacenamiento de datos	Capacidad de Almacenamiento	Gigabytes (GB)	$RCA = \frac{Capac\_Orig}{Capacid\_Dedup}$

ITEM	FECHA	CAPACIDAD ALMACENAMIENTO (PRE_TEST)	FECHA	CAPACIDAD ALMACENAMIENTO (POST_TEST)
1				

Dr. Marlon Acuña Benites
DNI: 42097456
Ing. de Sistemas / Investigador

Anexo 3 – Ficha de Registro (3.4 Ficha de Registro para Capacidad de Transferencia)

UCV UNIVERSIDAD CÉSAR VALLEJO	**INSTRUMENTO:** **Ficha de Registro**	**CÓDIGO:** UCV- ING SISTEMAS-TI-001.1	
	Recolección de datos	**Página:**	1 de 1
Investigador	Torres Paredes, Carlos Martin	**Tipo de Prueba**	Observación
Fecha Inicio	07/11/2022	**Fecha Final**	21/11/2022
Variable	**Indicador**	**Medida**	**Fórmula**
Gestión de almacenamiento de datos	Capacidad de Transferencia	segundos	$RT = \frac{Transf\_Orig}{Transf\_Dedup}$

ITEM	FECHA	TIEMPO DE TRANSFERENCIA (PRE_TEST)	FECHA	TIEMPO DE TRANSFERENCIA (POST_TEST)
1				

Dr. Marlon Acuña Benites
DNI: 42097456
Ing. de Sistemas / Investigador

Anexo 3 – Ficha de Registro (3.5 Ficha de Registro para Costo de Almacenamiento)

UCV UNIVERSIDAD CÉSAR VALLEJO	**INSTRUMENTO:** **Ficha de Registro**	**CÓDIGO:** UCV- ING SISTEMAS-TI-001.1	
	Recolección de datos	**Página:**	1 de 1
Investigador	Torres Paredes, Carlos Martin	**Tipo de Prueba**	Observación
Fecha Inicio	07/11/2022	**Fecha Final**	21/11/2022
Variable	**Indicador**	**Medida**	**Fórmula**
Gestión de almacenamiento de datos	Costo de Almacenamiento	U.S. $	$RCO = \frac{Costo\_Orig}{Costo\_Dedup}$

ITEM	FECHA	COSTO DE ALMACENAMIENTO (PRE_TEST)	FECHA	COSTO DE ALMACENAMIENTO (POST_TEST)
1				

Dr. Marlon Acuña Benites
DNI: 42097456
Ing. de Sistemas / Investigador

Anexo 4: Carta de presentación

UNIVERSIDAD CÉSAR VALLEJO

"Año del Fortalecimiento de la Soberanía Nacional"

Lima, 25 de octubre de 2022
Carta P. 1074-2022-UCV-VA-EPG-F01/J

Ing.
Viviana Claudia TABA SOKEI
Gerente General
EV Seguridad S.A.C

De mi mayor consideración:

Es grato dirigirme a usted, para presentar a TORRES PAREDES, Carlos Martin; identificado con DNI N° 07025329 y con código de matrícula N° 7002755404; estudiante del programa de MAESTRÍA EN INGENIERÍA DE SISTEMAS CON MENCIÓN EN TECNOLOGÍAS DE LA INFORMACIÓN quien, en el marco de su tesis conducente a la obtención de su grado de MAESTRO, se encuentra desarrollando el trabajo de investigación titulado:

Aplicación de algoritmos criptográficos en la gestión de almacenamiento de datos en una empresa de seguridad electrónica, Lima 2023

Con fines de investigación académica, solicito a su digna persona otorgar el permiso a nuestro estudiante, a fin de que pueda obtener información, en la institución que usted representa, que le permita desarrollar su trabajo de investigación. Nuestro estudiante investigador TORRES PAREDES, Carlos Martin asume el compromiso de alcanzar a su despacho los resultados de este estudio, luego de haber finalizado el mismo con la asesoría de nuestros docentes.

Agradeciendo la gentileza de su atención al presente, hago propicia la oportunidad para expresarle los sentimientos de mi mayor consideración.

Atentamente,

ESCUELA DE POSGRADO UCV JEFATURA CAMPUS LIMA NORTE

Dra. Estrella A. Esquiagola Aranda
Jefa
Escuela de Posgrado UCV
Filial Lima Campus Los Olivos

Anexo 5: Carta de aceptación

CCTV | Cámaras IP | Control de Acceso - Asistencia | Alarmas | Incendio
e-mail: evseguridad.consultor@gmail.com CEL: 961221675 RPM: 945177000

Lima, 03 de noviembre del 2022.
Ref.: Carta P. 1074-2022-UCV-VA-EPG-F01/J

Ingeniero:
Carlos Martin Torres Paredes
Maestría en Ingeniería de Sistemas
Universidad César Vallejo

De nuestra consideración:

Tenemos el agrado de dirigirnos a Usted para confirmar la recepción de su carta de presentación 1074-2022-UCV-VA-EPG-F01/J, que pone en nuestro conocimiento la investigación "Aplicación de algoritmos criptográficos en la gestión de datos en una empresa de seguridad electrónica, Lima 2023" que viene realizando como parte de los requerimientos del programa de Maestría en Ingeniería de Sistemas de la Universidad "César Vallejo", y para la cual solicita facilidades de acceso a información de los métodos operativos y de almacenamiento de datos de nuestra empresa.

Con relación a su solicitud, esta ha sido evaluada con resultado positivo, ya que consideramos que es de interés mutuo, para el campo académico y de la industria, el desarrollo de propuestas de mejora de las técnicas operativas y de la competitividad empresarial. En este sentido se le brindarán las facilidades solicitadas de acceso a nuestros métodos y datos operativos.

Asimismo, le deseamos una culminación exitosa del proyecto y desde ya quedamos pendientes de los resultados que este genere, los cuales redundarán en beneficio de nuestra representada y del sector en general.

Atentamente,

Viviana Taba Sokei
Gerente General
EV Seguridad S.A.C.

ING. VIVIANA TABA
GERENTE GENERAL
EV SEGURIDAD S.A.C.

Anexo 6: Constancia de idioma

CENTRO DE IDIOMAS
UNIVERSIDAD CÉSAR VALLEJO

CID- 2022-02-LN-1444

CONSTANCIA

La Jefa Nacional del Centro de Idiomas
de la Universidad César Vallejo

Hace Constar

Que, el(la) Sr(a). **TORRES PAREDES, CARLOS MARTIN**; estudiante del Programa de **MAESTRÍA EN INGENIERÍA DE SISTEMAS CON MENCIÓN EN TECNOLOGÍAS DE LA INFORMACIÓN** de la Universidad César Vallejo – Lima Los Olivos; con código N° 7002755404, ha aprobado el curso de 200 horas INGLÉS POSGRADO EXTRACURRICULAR, obteniendo la nota de 17 (diecisiete)/20, lo que equivale al Nivel A2 del MCER.

Se expide la presente constancia a solicitud de la parte interesada para los fines que estime conveniente.

Los Olivos, 10 de octubre de 2022

Atentamente,

Dra. Erica De Paz Berrospi
Jefatura Nacional del Centro de Idiomas
Universidad César Vallejo

Anexo 7: Aspectos Administrativos

A7.1 Recursos y presupuesto

A7.1.1 Recursos Humanos

Bajo este rubro estamos considerando los gastos operativos tales como transporte y la valoración de las horas hombre empleadas para el desarrollo de la investigación en actividades como recopilación de datos, procesamiento, análisis y documentación de resultados. En la siguiente tabla se aprecia el resumen de estos gastos.

Tabla A7.1

Recursos humanos – Presupuesto

Concepto	Descripción	Monto
Recopilación	Preparación de datos	S/. 4000.00
Movilidad y transporte	Movilidad	S/. 400.00
Procesamiento	Análisis y procesamiento de datos	S/. 5000.00
	Total	S/. 9400.00

A7.1.2 Recursos de Hardware (Equipos)

Bajo este apartado consideramos la adquisición de equipos necesarios para el proyecto, tales como cámaras, grabadores y computador personal. La tabla muestra el resumen de estos costos.

Tabla A7.2

Recursos de hardware - Presupuesto

Concepto	Descripción	Monto
Equipo de video	Cámaras y grabadores de frames	S/. 700.00
Equipo informático	Computador de escritorio (Core i5, 16 GB RAM, 500 GB SSD)	S/. 3000.00
	Total	S/. 3700.00

A7.1.3 Recursos de Software

Se ha considerado el uso de dos tipos de herramientas de software para el desarrollo del proyecto y el software para pruebas, que es de acceso libre, y el software de análisis estadístico, SPSS, que se trabajará con licencia educativa. LA siguiente tabla muestra el resumen de estos costos.

Tabla A7.3

Software - Presupuesto

Concepto	Descripción	Monto
Software de pruebas	Entorno de desarrollo-pruebas	S/. 100.00
Software de análisis	Licencia SPSS, software estadístico	S/. 300.00
	Total	S/. 400.00

A7.1.4 Presupuesto

Básicamente es el resultado de sumar los presupuestos parciales de las categorías previas: recursos humanos, recursos de hardware y recursos de software.

Tabla A7.4

Presupuesto general

Presupuestos parciales	Monto
Recursos humanos	S/. 9400.00
Recursos de hardware (Equipamiento)	S/. 3700.00
Recursos de Software	S/. 400.00
Total	S/. 13500.00

A7.2 Financiamiento

En la presente investigación se busca aplicar técnicas criptográficas que permitan mejorar la eficiencia de la gestión del almacenamiento de datos. Para tal fin se han empleado los recursos que se detalla en los puntos previos y los costos de desarrollo y análisis fueron asumidos por el maestrista que ha desarrollado la investigación.

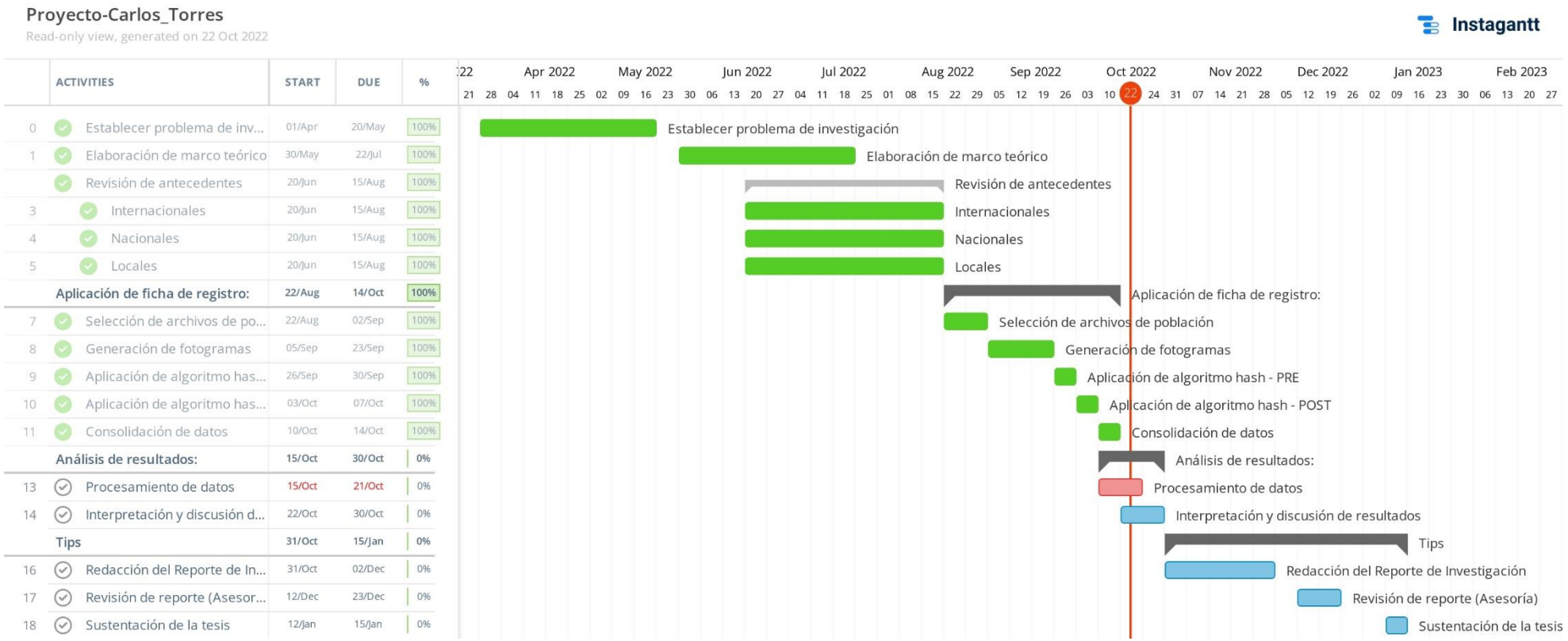
Tabla A7.5

Financiamiento de la investigación

Entidad Financiera	Monto	Porcentaje
Maestrista (Autofinanciado)	S/. 13500.00	100 %

A7.3 Cronograma de ejecución

Se presenta el Diagrama de Gantt del Proyecto



UNIVERSIDAD CÉSAR VALLEJO

ESCUELA DE POSGRADO

MAESTRÍA EN INGENIERÍA DE SISTEMAS CON MENCIÓN EN TECNOLOGÍAS DE LA INFORMACIÓN

Declaratoria de Autenticidad del Asesor

Yo, ACUÑA BENITES MARLON FRANK, docente de la ESCUELA DE POSGRADO MAESTRÍA EN INGENIERÍA DE SISTEMAS CON MENCIÓN EN TECNOLOGÍAS DE LA INFORMACIÓN de la UNIVERSIDAD CÉSAR VALLEJO SAC - LIMA NORTE, asesor de Tesis Completa titulada: "Aplicación de Algoritmos Criptográficos en la Gestión de Almacenamiento de Datos en una Empresa de Seguridad Electrónica, Lima 2023", cuyo autor es TORRES PAREDES CARLOS MARTIN, constato que la investigación tiene un índice de similitud de 16.00%, verificable en el reporte de originalidad del programa Turnitin, el cual ha sido realizado sin filtros, ni exclusiones.

He revisado dicho reporte y concluyo que cada una de las coincidencias detectadas no constituyen plagio. A mi leal saber y entender la Tesis Completa cumple con todas las normas para el uso de citas y referencias establecidas por la Universidad César Vallejo.

En tal sentido, asumo la responsabilidad que corresponda ante cualquier falsedad, ocultamiento u omisión tanto de los documentos como de información aportada, por lo cual me someto a lo dispuesto en las normas académicas vigentes de la Universidad César Vallejo.

LIMA, 05 de Enero del 2023

Apellidos y Nombres del Asesor:	**Firma**
ACUÑA BENITES MARLON FRANK **DNI:** 42097456 **ORCID:** 0000-0001-5207-9353	Firmado electrónicamente por: MACUNABE el 05-01-2023 14:16:19

Código documento Trilce: TRI - 0510256

INVESTIGA UCV