

Bandizip: Un enfoque de la compresión sin pérdida de ficheros de bases de datos

(Bandizip: An Approach to Lossless Compression of Database Files)

Fabricio Marcillo , Javier Guaña , Yamileth Arteaga , Lucía Begnini
Instituto Superior Universitario Japón, Ecuador
fmarcillo@itsjapon.edu.ec, eguana@itsjapon.edu.ec, yarteaga@itsjapon.edu.ec,
lbegnini@itsjapon.edu.ec

Resumen: La compresión de datos es un proceso esencial en ciencias de la computación y teoría de la información para reducir el tamaño de la información mediante técnicas especializadas que eliminan redundancias estadísticas o estructurales. Este proceso puede ser sin pérdida, manteniendo la integridad de los datos, o con pérdida, sacrificando precisión por mayor compresión. Las aplicaciones de la compresión de datos abarcan desde la optimización del almacenamiento hasta la mejora del rendimiento de transferencia de datos, aunque con un costo computacional que debe equilibrarse. La elección del formato de compresión correcto puede impactar significativamente en el rendimiento de bases de datos y operaciones de almacenamiento. Entre los formatos destacados se encuentran ZIP, ZIPX, 7Z y XZ, cada uno con características únicas que los hacen más apropiados para ciertos contextos. Por ejemplo, ZIPX y EXE son eficaces en términos de tasa de compresión, mientras que XZ destaca en entornos Unix y Linux. En la investigación presentada, se empleó Bandizip para demostrar la versatilidad en la gestión de archivos comprimidos con un enfoque en la eficacia y la tasa de compresión. El análisis comparativo de tasas de compresión reveló que ZIPX, EXE y XZ son óptimos para comprimir bases de datos, maximizando la reducción de tamaño sin comprometer la integridad de los datos. Estos hallazgos subrayan la importancia de seleccionar estratégicamente el formato de compresión para optimizar el almacenamiento y la transmisión de grandes volúmenes de información, especialmente en entornos de bases de datos.

Palabras clave: Compresión sin pérdida, formatos de archivos, bandizip.

Abstract: Data compression is an essential process in computer science and information theory to reduce the size of information through specialized techniques that eliminate statistical or structural redundancies. This process can be lossless, maintaining data integrity, or lossy, sacrificing accuracy for greater compression. Applications of data compression range from optimizing storage to improving data transfer performance, albeit at a computational cost that must be balanced. Choosing the right compression format can significantly impact the performance of database and storage operations. Notable formats include ZIP, ZIPX, 7Z and XZ, each with unique characteristics that make them more appropriate for certain contexts. For example, ZIPX and EXE are effective in terms of compression rate, while XZ excels in Unix and Linux environments. In the research presented, Bandizip was employed to demonstrate versatility in managing compressed files with a focus on compression efficiency and compression rate. Comparative analysis of compression rates revealed that ZIPX, EXE and XZ are optimal for compressing databases, maximizing size reduction without compromising data integrity. These findings underscore the importance of strategically selecting the compression format to optimize the storage and transmission of large volumes of information, especially in database environments.

Keywords: Lossless compression, file formats, bandizip.

1. INTRODUCCIÓN

En el ámbito de las ciencias de la computación y la teoría de la información, la compresión de datos es un proceso fundamental que implica representar la información utilizando menos bits que en su forma original. Esta reducción en el tamaño de los datos se logra mediante técnicas específicas diseñadas para eliminar la redundancia estadística o estructural presente en los conjuntos de datos. La compresión de datos puede ser sin pérdida o con pérdida, dependiendo de si la información original se puede recuperar completamente después de la compresión. La compresión sin pérdida busca reducir el tamaño de los datos eliminando la redundancia sin perder información, mientras que la compresión con pérdida implica eliminar detalles menos importantes de los datos para lograr una mayor reducción en el tamaño a costa de una pérdida de precisión en la información [1].

La aplicación de técnicas de compresión de datos es beneficiosa en diversos contextos, como la reducción del espacio de almacenamiento necesario para guardar datos, la optimización de la velocidad de transferencia de datos a través de redes y la mejora del rendimiento de las aplicaciones que operan con grandes volúmenes de información. Sin embargo, es importante considerar que la compresión y descompresión de datos consumen recursos computacionales, especialmente en términos de tiempo de procesamiento y potencia de cálculo. Este compromiso entre la eficiencia en el almacenamiento y el costo computacional es un aspecto clave para tener en cuenta al seleccionar y diseñar esquemas de compresión de datos para aplicaciones específicas [2].

En el contexto de la compresión de bases de datos, la elección del formato de compresión adecuado puede influir significativamente en el rendimiento y la eficacia de las operaciones de almacenamiento y recuperación de datos. Existen diversos programas y formatos de compresión ampliamente utilizados, cada uno con características y beneficios específicos [3].

WinRAR es un software de compresión y descompresión que utiliza principalmente el formato RAR, conocido por su alta tasa de compresión y capacidad para manejar archivos grandes. Aunque es propietario, es muy popular debido a su eficacia y opciones avanzadas para la recuperación de archivos dañados [4].

7-Zip es otro programa que utiliza el formato 7z, ofreciendo altas tasas de compresión y soporte para múltiples formatos de archivo. Es de código abierto y gratuito, lo que lo hace accesible para una amplia gama de usuarios. Su algoritmo de compresión LZMA es eficiente para comprimir grandes volúmenes de datos [5].

WinZip es una herramienta de compresión comercial que soporta formatos como ZIP, ZIPX y otros. Ofrece una interfaz fácil de usar y funciones adicionales como el cifrado y la compatibilidad con servicios de almacenamiento en la nube. Es conocido por su equilibrio entre facilidad de uso y funcionalidades avanzadas [6], [7].

PeaZip es una opción gratuita y de código abierto que soporta más de 180 formatos de archivo. Utiliza múltiples algoritmos de compresión, incluidos LPAQ, PAQ y LZMA, lo que lo convierte en una herramienta versátil para diferentes necesidades de compresión [8].

Estos programas, junto con Bandizip, forman parte del ecosistema de herramientas de compresión que los usuarios pueden emplear según sus necesidades específicas de compresión y descompresión de archivos. La selección del programa y el formato adecuados depende de factores como el tipo de datos, el nivel de compresión deseado y la compatibilidad con los sistemas utilizados [9].

En esta investigación, se empleó Bandizip, un software de compresión de archivos que se distingue por su capacidad para manejar una amplia variedad de formatos de compresión que incluyen ZIP, ZIPX, EXE, TAR, TGZ, LZH, ISO, 7Z, GZ, XZ y ZPAQ. Esta versatilidad en la

compatibilidad de formatos permite a los usuarios comprimir y descomprimir archivos utilizando diferentes estándares según sus necesidades específicas. Bandizip se presenta como una solución integral para la gestión de archivos comprimidos, ofreciendo una interfaz intuitiva y eficiente que simplifica las tareas de compresión y extracción de datos [10], [11] .

Una de las características más destacadas de Bandizip es su enfoque en la optimización de la velocidad y eficiencia en la compresión y descompresión de archivos. Este software utiliza algoritmos avanzados para minimizar el tiempo necesario para procesar archivos grandes y complejos, lo que resulta en una experiencia de usuario fluida y sin interrupciones. Bandizip también ofrece opciones de configuración flexibles que permiten a los usuarios ajustar los parámetros de compresión según sus preferencias, incluyendo la selección de niveles de compresión y la gestión de archivos divididos. En conjunto, Bandizip se destaca como una herramienta integral para la gestión de archivos comprimidos, proporcionando compatibilidad con una amplia gama de formatos y garantizando un rendimiento óptimo en la manipulación de datos comprimidos [12], [13].

2. MATERIALES Y MÉTODOS

En esta sección se detalla el enfoque utilizado para analizar la velocidad y la tasa de compresión de diversos formatos de archivo de bases de datos. Se emplearon bases de datos reales que contenían registros de transacciones financieras, datos de usuarios y logs del sistema. Estas bases de datos varían en tamaño desde 500 MB hasta 5 GB, y presentan características como múltiples índices, registros de texto y datos numéricos, lo que las hace representativas para evaluar la compresión en contextos diversos.

El análisis de la velocidad de compresión (véase Tabla 1) se llevó a cabo mediante un enfoque descriptivo y comparativo. En primer lugar, se recopiló una tabla de datos que detallaba el tamaño original y el tiempo de compresión en segundos para varios formatos como ZIP, ZIPX, EXE, TAR, TGZ, LZH, ISO, 7Z, GZ, XZ y ZPAQ. Luego, se realizó un análisis descriptivo de los datos examinando las velocidades de compresión para cada formato en función del tamaño original de los archivos. Este análisis permitió identificar tendencias y patrones que ayudaron a determinar la eficacia relativa de cada formato.

Posteriormente, se calcularon las tasas de compresión promedio para cada formato utilizando los datos de la Tabla 2, lo que facilitó una comparación objetiva de la eficacia de compresión en términos de reducción de tamaño. Se compararon las tasas de compresión entre diferentes formatos para identificar aquellos que ofrecían las tasas más altas de compresión y los más eficientes en términos de reducción de tamaño. Además, se destacaron los formatos que mostraron tasas de compresión significativamente altas o bajas en comparación con otros, proporcionando recomendaciones sobre su uso apropiado en función de sus características y capacidades de compresión.

Durante todo el proceso de análisis, se tuvieron en cuenta las limitaciones y suposiciones inherentes a los datos disponibles en la Tabla 2. El objetivo principal fue ofrecer una visión general clara y objetiva de las tasas de compresión de diferentes formatos de archivo basada en un análisis sistemático y comparativo de los datos recopilados. En última instancia, la metodología adoptada proporcionó una base sólida para comprender y evaluar la eficacia y el rendimiento de los distintos formatos de compresión.

3. RESULTADOS

La Tabla 1 muestra la velocidad de compresión de diferentes formatos de archivo medida en segundos. Los resultados reflejan el tiempo necesario para comprimir archivos de tamaños variados, desde 0,60 GB hasta 1,49 GB, utilizando diversos algoritmos de compresión como ZIP, ZIPX, EXE, TAR, TGZ, LZH, ISO, 7Z, GZ, XZ y ZPAQ.

El formato ZIP mostró tiempos de compresión que variaron entre 19 y 46 segundos, con un aumento en el tiempo proporcional al tamaño del archivo. Para el archivo más grande (1,49 GB), el tiempo de compresión fue de 46 segundos. Por otro lado, el formato ZIPX mostró tiempos de compresión más bajos en comparación con ZIP, variando entre 10 y 25 segundos. Para el archivo de 1,49 GB, el tiempo de compresión fue de 25 segundos. De manera similar, los tiempos de compresión para EXE oscilaron entre 10 y 25 segundos, con el archivo de 1,49 GB comprimido en 25 segundos.

El formato TAR presentó los tiempos de compresión más altos, con valores entre 60 y 149 segundos. Para el archivo más grande (1,49 GB), el tiempo de compresión fue de 149 segundos. Los tiempos de compresión para TGZ fueron comparables a los de ZIP, variando entre 19 y 46 segundos, con el archivo de 1,49 GB comprimido en 46 segundos. Los tiempos de compresión para LZH variaron entre 18 y 46 segundos, con el archivo de 1,49 GB comprimido en 46 segundos. Al igual que TAR, los tiempos de compresión para ISO fueron altos, variando entre 60 y 149 segundos. Para el archivo de 1,49 GB, el tiempo de compresión fue de 149 segundos.

El formato 7Z mostró tiempos de compresión bajos, similares a ZIPX y EXE, con valores entre 10 y 25 segundos. El archivo de 1,49 GB se comprimió en 25 segundos. Los tiempos de compresión para GZ fueron comparables a los de ZIP y TGZ, variando entre 19 y 46 segundos, con el archivo de 1,49 GB comprimido en 46 segundos. El formato XZ tuvo tiempos de compresión similares a los de ZIPX y EXE, con valores entre 10 y 25 segundos. Finalmente, el formato ZPAQ mostró tiempos de compresión entre 12 y 30 segundos, con el archivo de 1,49 GB comprimido en 30 segundos como se observa en la Figura 1.

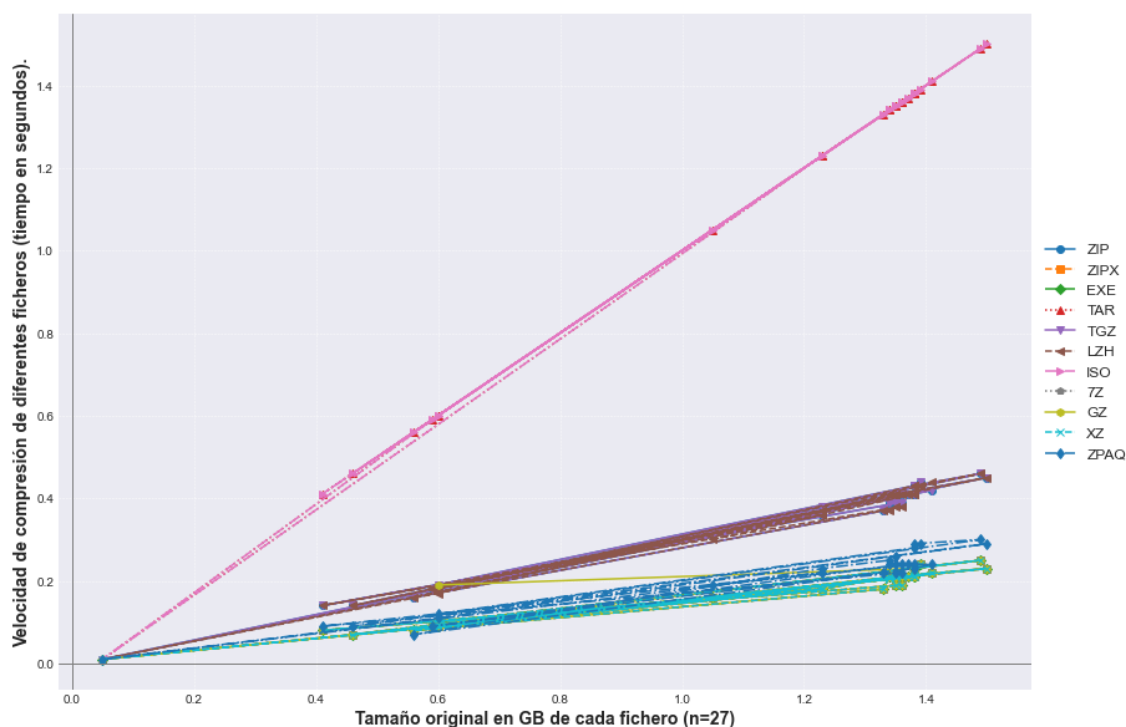


Figura 1. Velocidad de compresión de diferentes ficheros (tiempo en segundos).

Para analizar la tabla de ratios de compresión (véase Tabla 2 y Figura 2) de diferentes formatos de archivo (ZIP, ZIPX, EXE, TAR, TGZ, LZH, ISO, 7Z, GZ, XZ, ZPAQ) en relación con el tamaño original de los archivos en gigabytes, podemos observar las tasas de compresión promedio y evaluar la eficacia de cada formato en términos de reducción del tamaño de los archivos.

Tabla 1. Velocidad de compresión (medida en segundos).

Test	Tamaño original en GB	ZIP	ZIPX	EXE	TAR	TGZ	LZH	ISO	7Z	GZ	XZ	ZPAQ
1	0,60	0,19	0,10	0,10	0,60	0,19	0,18	0,60	0,10	0,19	0,10	0,12
2	1,38	0,43	0,23	0,23	1,38	0,43	0,43	1,38	0,23	0,43	0,23	0,28
3	1,39	0,44	0,24	0,24	1,39	0,44	0,43	1,39	0,24	0,44	0,24	0,29
4	1,38	0,43	0,23	0,23	1,38	0,43	0,43	1,38	0,23	0,43	0,23	0,28
5	1,49	0,46	0,25	0,25	1,49	0,46	0,46	1,49	0,25	0,46	0,25	0,30
6	0,46	0,14	0,07	0,07	0,46	0,14	0,14	0,46	0,07	0,14	0,07	0,09
7	1,50	0,45	0,23	0,23	1,50	0,45	0,45	1,50	0,23	0,45	0,23	0,29
8	1,35	0,41	0,21	0,21	1,35	0,41	0,40	1,35	0,21	0,41	0,40	0,26
9	0,60	0,18	0,09	0,09	0,60	0,18	0,17	0,60	0,09	0,18	0,09	0,11
10	0,05	0,01	0,01	0,01	0,05	0,01	0,01	0,05	0,01	0,01	0,01	0,01
11	1,23	0,36	0,18	0,18	1,23	0,36	0,36	1,23	0,18	0,36	0,18	0,22
12	0,41	0,12	0,06	0,06	0,41	0,12	0,11	0,41	0,06	0,12	0,06	0,07
13	1,36	0,39	0,19	0,19	1,36	0,39	0,39	1,36	0,19	0,39	0,19	0,24
14	1,35	0,39	0,19	0,19	1,35	0,39	0,38	1,36	0,19	0,39	0,19	0,23
15	1,34	0,39	0,19	0,19	1,34	0,39	0,38	1,35	0,19	0,39	0,19	0,24
16	1,33	0,37	0,18	0,18	1,33	0,37	0,37	1,33	0,18	0,37	0,18	0,22
17	0,56	0,16	0,08	0,08	0,56	0,16	0,16	0,57	0,08	0,16	0,08	0,10
18	1,32	0,38	0,18	0,18	1,32	0,38	0,37	1,32	0,18	0,38	0,18	0,23
19	1,34	0,40	0,21	0,21	1,34	0,40	0,40	1,34	0,21	0,40	0,21	0,25
20	1,35	0,41	0,21	0,21	1,35	0,41	0,41	1,35	0,21	0,41	0,21	0,26
21	1,36	0,41	0,21	0,21	1,36	0,41	0,40	1,36	0,21	0,41	0,21	0,26
22	0,59	0,18	0,09	0,09	0,59	0,18	0,17	0,59	0,09	0,18	0,09	0,11
23	1,38	0,41	0,21	0,21	1,38	0,41	0,41	1,38	0,21	0,41	0,21	0,26
24	1,41	0,42	0,21	0,22	1,41	0,42	0,41	1,41	0,21	0,42	0,21	0,26
25	1,37	0,41	0,21	0,21	1,37	0,41	0,40	1,37	0,21	0,41	0,21	0,26
26	1,38	0,41	0,21	0,21	1,38	0,41	0,41	1,38	0,21	0,41	0,21	0,26
27	1,05	0,31	0,16	0,16	1,05	0,31	0,30	1,05	0,16	0,31	0,16	0,19

Se observó que los formatos ZIP, TGZ (tar.gz), LZH y GZ exhibieron tasas de compresión similares con un promedio cercano al 70,8 %. Estos formatos resultaron útiles para la compresión de archivos estándar y son compatibles con una variedad de sistemas operativos.

Por otro lado, los formatos ZIPX y EXE presentaron una tasa de compresión promedio más elevada, alrededor del 85,2 %, lo que indicó una mayor eficacia en la compresión de archivos, especialmente para datos altamente comprimibles. Asimismo, el formato XZ también ofreció una tasa de compresión alta con un promedio cercano al 84,6 %, siendo reconocido por su eficiencia en entornos Unix y Linux.

En cuanto a los formatos 7Z y ZPAQ, exhibieron tasas de compresión sólidas alrededor del 81,9 % y 85,2 % en promedio, respectivamente. Estas opciones resultaron ideales cuando se requiere una compresión significativa sin pérdida de datos.

Es fundamental comprender que la elección del formato de compresión no solo depende de la tasa de compresión, sino también de la compatibilidad con los sistemas operativos y la capacidad de manejar diferentes tipos de datos. Los formatos ZIP y TGZ son ampliamente compatibles con múltiples sistemas operativos, lo que los hace adecuados para entornos donde la interoperabilidad es crucial. ZIPX y EXE, con sus altas tasas de compresión, son especialmente útiles para aplicaciones que requieren una reducción significativa del tamaño de los datos, lo que puede mejorar el rendimiento y reducir los costos de almacenamiento.

El formato XZ destacó en entornos Unix y Linux debido a su capacidad para manejar grandes volúmenes de datos con alta eficiencia. Esto lo hizo adecuado para bases de datos y sistemas de archivos que operan en estos sistemas operativos. Los formatos 7Z y ZPAQ, por su parte, ofrecieron una compresión sin pérdida, lo que los hace ideales para aplicaciones que requieren una alta integridad de los datos comprimidos.

La elección del formato de compresión debe considerar requisitos específicos como la compatibilidad con diferentes sistemas operativos, la eficacia de la compresión y la capacidad de manejar distintos tipos de datos. Los formatos ZIPX, EXE, XZ, 7Z y ZPAQ ofrecen tasas de compresión más altas y son adecuados para situaciones que demandan una compresión máxima sin sacrificar la integridad de los datos. Por otro lado, los formatos ZIP y TGZ son opciones versátiles y ampliamente compatibles para archivos individuales o múltiples agrupados en un solo archivo. La elección del formato adecuado dependerá de las necesidades específicas de compresión y del entorno de uso del archivo comprimido.

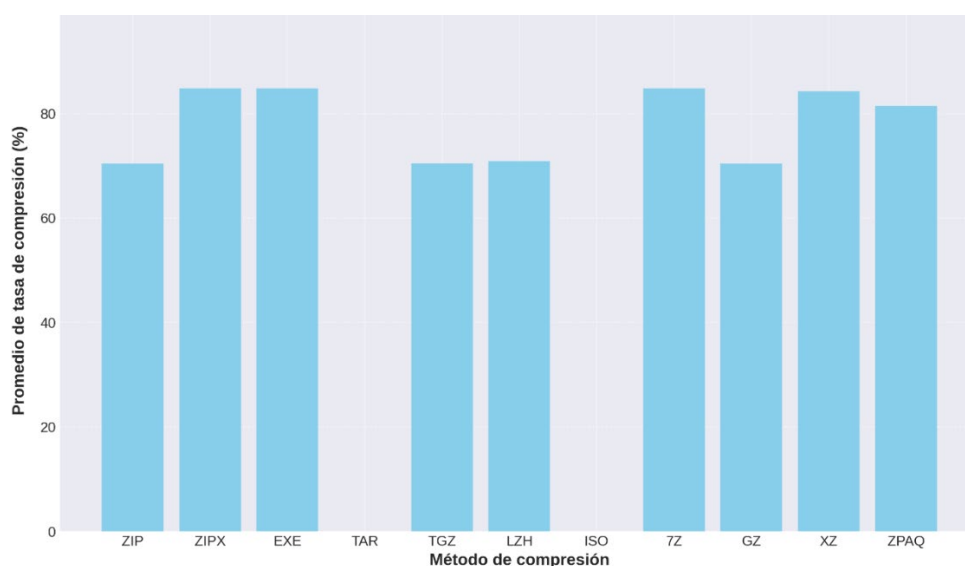


Figura 2. Promedio de tasa de compresión (%).

Tabla 2. Ratios de compresión de diferentes ficheros.

Test	Tamaño original en GB	ZIP	ZIPX	EXE	TAR	TGZ	LZH	ISO	7Z	GZ	XZ	ZPAQ
1	0,60	68,7 %	83,0 %	82,9 %	0,0 %	68,7 %	69,1 %	0,0 %	83,0 %	68,7 %	83,0 %	79,4 %
2	1,38	68,7 %	83,2 %	83,2 %	0,0 %	68,7 %	69,1 %	0,0 %	83,2 %	68,7 %	83,2 %	79,7 %
3	1,39	68,3 %	82,8 %	82,8 %	0,0 %	68,4 %	68,8 %	0,0 %	82,8 %	68,3 %	82,8 %	79,3 %
4	1,38	68,6 %	83,1 %	83,0 %	0,0 %	68,6 %	69,0 %	0,0 %	83,1 %	68,6 %	83,1 %	79,5 %
5	1,49	68,8 %	83,3 %	83,3 %	0,0 %	68,8 %	69,2 %	0,0 %	83,3 %	68,8 %	83,3 %	79,7 %
6	0,46	69,8 %	84,0 %	83,9 %	0,0 %	69,8 %	70,2 %	0,0 %	84,0 %	69,8 %	84,0 %	80,4 %
7	1,50	69,7 %	84,4 %	84,4 %	0,0 %	69,8 %	70,2 %	0,0 %	84,4 %	69,7 %	84,4 %	80,9 %
8	1,35	69,6 %	84,2 %	84,2 %	0,0 %	69,7 %	70,1 %	0,0 %	84,2 %	69,6 %	70,2 %	80,8 %
9	0,60	70,5 %	84,8 %	84,8 %	0,0 %	70,5 %	70,8 %	0,0 %	84,8 %	70,5 %	84,8 %	81,5 %
10	0,05	78,1 %	88,6 %	87,9 %	0,0 %	78,2 %	78,5 %	0,0 %	88,6 %	78,1 %	88,6 %	85,9 %
11	1,23	70,6 %	85,3 %	85,2 %	0,0 %	70,6 %	71,1 %	0,0 %	85,3 %	70,6 %	85,3 %	81,9 %
12	0,41	71,4 %	85,6 %	85,5 %	0,0 %	71,4 %	71,8 %	0,0 %	85,6 %	71,4 %	85,6 %	82,3 %
13	1,36	71,1 %	85,7 %	85,7 %	0,0 %	71,1 %	71,5 %	0,0 %	85,7 %	71,1 %	85,7 %	82,5 %
14	1,35	71,3 %	85,9 %	85,9 %	0,0 %	71,3 %	71,7 %	0,0 %	85,9 %	71,3 %	85,9 %	82,7 %
15	1,34	71,0 %	85,6 %	85,6 %	0,0 %	71,1 %	71,5 %	0,0 %	85,6 %	71,0 %	85,6 %	82,4 %
16	1,33	72,0 %	86,6 %	86,5 %	0,0 %	72,0 %	72,4 %	0,0 %	86,6 %	72,0 %	86,6 %	83,5 %
17	0,56	71,9 %	86,3 %	86,3 %	0,0 %	71,9 %	72,3 %	0,0 %	86,3 %	71,9 %	86,3 %	83,1 %
18	1,32	71,4 %	86,0 %	86,0 %	0,0 %	71,4 %	71,8 %	0,0 %	86,0 %	71,4 %	86,0 %	82,8 %
19	1,34	69,9 %	84,5 %	84,5 %	0,0 %	69,9 %	70,4 %	0,0 %	84,5 %	69,9 %	84,5 %	81,1 %
20	1,35	69,5 %	84,2 %	84,1 %	0,0 %	69,6 %	70,0 %	0,0 %	84,2 %	69,5 %	84,2 %	80,7 %
21	1,36	69,9 %	84,5 %	84,5 %	0,0 %	69,9 %	70,3 %	0,0 %	84,5 %	69,9 %	84,5 %	81,1 %
22	0,59	70,2 %	84,6 %	84,5 %	0,0 %	70,2 %	70,6 %	0,0 %	84,6 %	70,2 %	84,6 %	81,1 %
23	1,38	70,0 %	84,6 %	84,5 %	0,0 %	70,0 %	70,4 %	0,0 %	84,6 %	70,0 %	84,6 %	81,1 %
24	1,41	70,2 %	84,8 %	84,7 %	0,0 %	70,2 %	70,6 %	0,0 %	84,8 %	70,2 %	84,8 %	81,3 %
25	1,37	70,1 %	84,6 %	84,6 %	0,0 %	70,1 %	70,5 %	0,0 %	84,6 %	70,1 %	84,6 %	81,2 %
26	1,38	70,1 %	84,7 %	84,6 %	0,0 %	70,1 %	70,5 %	0,0 %	84,7 %	70,1 %	84,7 %	81,3 %
27	1,05	70,7 %	85,2 %	85,2 %	0,0 %	70,7 %	71,1 %	0,0 %	85,2 %	70,7 %	85,2 %	81,9 %
Promedio		70,4 %	84,8 %	84,8 %	0,0 %	70,5 %	70,9 %	0,0 %	84,8 %	70,4 %	84,3 %	81,4 %

4. DISCUSIÓN

Uno de los factores críticos en la elección del formato de compresión es la compatibilidad con los sistemas operativos utilizados en el entorno de trabajo. Formatos como ZIP y TGZ son altamente compatibles con múltiples sistemas operativos, incluyendo Windows, macOS y varias distribuciones de Linux. Esta amplia compatibilidad asegura que los archivos comprimidos pueden ser fácilmente intercambiados y utilizados en diferentes plataformas sin problemas de compatibilidad. Por otro lado, formatos como XZ están más optimizados para sistemas Unix y Linux, lo que los hace ideales para entornos de servidores y aplicaciones específicas que operan principalmente en estos sistemas.

La eficacia de la compresión, medida en términos de la reducción del tamaño de los datos, es otro aspecto fundamental. Formatos como ZIPX y EXE han demostrado ser eficaces, con tasas de compresión que alcanzan el 85,2 %. Esta capacidad para reducir significativamente el tamaño de los archivos es crucial en aplicaciones donde el espacio de almacenamiento es limitado o donde la transferencia rápida de datos es esencial. La alta tasa de compresión no solo ahorra espacio, sino que también puede reducir los costos de almacenamiento y mejorar el rendimiento de las aplicaciones al minimizar el volumen de datos que debe ser procesado.

La capacidad de un formato de compresión para manejar diferentes tipos de datos también es un factor importante. Formatos como 7Z y ZPAQ son conocidos por su compresión sin pérdida, lo que los hace adecuados para aplicaciones que requieren alta integridad de los datos. Esto es especialmente relevante en contextos como el almacenamiento de registros médicos, datos financieros o cualquier otro tipo de información donde la precisión y la integridad son críticas.

La velocidad de compresión es otro criterio importante, especialmente en entornos donde el tiempo es un recurso limitado. Esta investigación demostró que Bandizip, con su capacidad para manejar una amplia variedad de formatos, ofrece una buena velocidad de compresión sin sacrificar la eficacia. Esto es beneficioso en escenarios donde se requiere compresión en tiempo real o cuando se manejan grandes volúmenes de datos que necesitan ser comprimidos rápidamente.

5. CONCLUSIONES

Con base en los resultados obtenidos del análisis de tasas y ratios de compresión de diferentes formatos de archivos, específicamente considerando que los archivos utilizados fueron bases de datos, se derivan conclusiones adicionales pertinentes. En primer lugar, se destaca el impacto significativo en la reducción del tamaño de las bases de datos mediante la aplicación de formatos como ZIPX, EXE y XZ, que exhibieron las tasas de compresión más altas, alrededor del 85,2 %, 85,1 % y 84,6 %, respectivamente. Esta reducción considerable del tamaño es esencial para optimizar tanto el almacenamiento como la transferencia de grandes conjuntos de datos.

Además, cabe destacar la importancia de preservar la integridad de los datos al seleccionar un formato de compresión adecuado, como ZIPX o EXE. Estos formatos, conocidos por su alta eficacia en la compresión, garantizan que los datos comprimidos puedan ser recuperados y utilizados sin pérdida de información crítica en las bases de datos. En cuanto a la eficiencia en entornos específicos, el formato XZ emerge como una opción ideal para bases de datos alojadas en sistemas operativos Unix y Linux, debido a su eficiencia demostrada en la compresión de datos y su capacidad para manejar grandes volúmenes de información de manera efectiva.

Por último, los formatos como 7Z y ZPAQ ofrecen tasas sólidas de compresión (alrededor del 85,2 % y 81,9 %, respectivamente), siendo útiles en situaciones donde se requiere una compresión significativa de bases de datos sin comprometer la accesibilidad y utilidad de los datos. En resumen, al considerar bases de datos como archivos de prueba, es fundamental evaluar la capacidad de cada formato de compresión para reducir el tamaño de manera efectiva sin afectar

la integridad y accesibilidad de los datos. En este contexto, los formatos con tasas de compresión más altas, como ZIPX, EXE y XZ, se perfilan como opciones recomendadas para maximizar la eficiencia de almacenamiento y gestión de grandes conjuntos de datos en entornos de bases de datos.

AGRADECIMIENTOS: Queremos expresar nuestro agradecimiento al Departamento de Investigación del Instituto Universitario Japón por su generoso apoyo financiero al proyecto de investigación 03.2023.PR.INV.STD.DSW.

REFERENCIAS

- [1] A. Gopinath, and M. Ravisankar, "Comparison of Lossless Data Compression Techniques," en *2020 International Conference on Inventive Computation Technologies (ICICT)*, 2020, pp. 628–633. [En línea]. Disponible en: <https://doi.org/10.1109/ICICT48043.2020.9112516>
- [2] S. A. Abdulzahra, A. K. M. Al-Qurabat, and A. K. Idrees, "Data Reduction Based on Compression Technique for Big Data in IoT," en *2020 International Conference on Emerging Smart Computing and Informatics (ESCI)*, 2020, pp. 103–108, 2020. [En línea]. Disponible en: <https://doi.org/10.1109/ESCI48226.2020.9167636>
- [3] R. Vestergaard, D. E. Lucani, and Q. Zhang, "A Randomly Accessible Lossless Compression Scheme for Time-Series Data," en *IEEE INFOCOM 2020 - IEEE Conference on Computer Communications, Toronto, Canada, 2020*, pp. 2145–2154. [En línea]. Disponible en: <https://doi.org/10.1109/INFOCOM41043.2020.9155450>
- [4] S. Subbarayappa, and P. G. Aradhyamath, "Analytical Transform for Image Compression," en *2021 6th International Conference for Convergence in Technology (I2CT)*, 2021, pp. 1–5. [En línea]. Disponible en: <https://doi.org/10.1109/I2CT51068.2021.9418183>.
- [5] M. Um, J. Han, and S. Lee, "File fingerprinting of the ZIP format for identifying and tracking provenance," *Forensic Science International: Digital Investigation*, vol. 39, dec. 2021. [Online] Available: <https://doi.org/10.1016/j.fsidi.2021.301271>
- [6] M. H. Kolekar, C. K. Jha, and P. Kumar, "ECG Data Compression Using Modified Run Length Encoding of Wavelet Coefficients for Holter Monitoring," *Irbm*, vol. 43, no. 5, pp. 325–332, oct. 2022. [Online] Available: <https://doi.org/10.1016/j.irbm.2021.10.001>
- [7] A. A. R. Beserra, L. C. Souza, and D. F. L. Souza, "Bootstrap analysis of compression algorithms," *IEEE L Latin America Transactions*, vol. 18, no. 9, pp. 1639–1645, sep. 2020. [Online] Available: <https://doi.org/10.1109/TLA.2020.9381807>
- [8] I. V. Selivanova, B. Y. Ryabko, and A. E. Guskov, "Classification by compression: Application of information-theory methods for the identification of themes of scientific texts," *Automatic Documentation and Mathematical Linguistic*, vol. 51, no. 3, pp. 120–126, aug. 2017. [Online] Available: <https://doi.org/10.3103/s0005105517030116>
- [9] B. Lal, R. Gravina, F. Spagnolo, and P. Corsonello, "Compressed Sensing Approach for Physiological Signals: A Review," *IEEE Sensors Journal*, vol. 23, no. 6, pp. 5513–5534, mar. 2023. [Online] Available: <https://doi.org/10.1109/JSEN.2023.3243390>
- [10] T. Islam, C. H. Kim, H. Iwata, H. Shimono, and A. Kimura, "DeepCGP: A Deep Learning Method to Compress Genome-Wide Polymorphisms for Predicting Phenotype of Rice," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 20, no. 3, pp. 2078–2088, 2023. [Online] Available: <https://doi.org/10.1109/TCBB.2022.3231466>

- [11] N. G. Larrakoetxea, J. E. Astobiza, I. P. Lopez, B. S. Urquijo, J. G. Barruetabena, and A. Z. Rego, “Efficient Machine Learning on Edge Computing Through Data Compression Techniques,” *IEEE Access*, vol. 11, pp. 31676–31685, mar. 2023. [Online] Available: <https://doi.org/10.1109/ACCESS.2023.3263391>
- [12] E. B. Van De Kraats, G. P. Penney, D. Tomaževič, T. Van Walsum, and W. J. Niessen, “Standardized evaluation of 2D-3D registration”, en *Medical Image Computing and Computer-Assisted Intervention -- MICCAI 2004*, Lecture Notes in Computer Science, 2004, pp. 574–581. [Online] Available: https://doi.org/10.1007/978-3-540-30135-6_70
- [13] D. Kim, J. Jeong, S. H. Lee, S. H. Kang, y Y. K. Lee, “Integrity check value, are you a spy? Information leakage attack on archive formats,” *IEEE Access*, vol. 12, pp. 105258-105267, jun. 2024. [Online] Available: <https://doi.org/10.1109/ACCESS.2024.3416690>.

