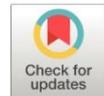


Mejora del sistema de almacenamiento de un servidor de backups mediante la evaluación de los sistemas de duplicación Opendedup (SDFS) Y ZFS

Improvement of the storage system of a backup server through the evaluation of the Opendedup duplication systems (SDFS) and ZFS

- ¹ Marco Vinicio Ramos Valencia  <https://orcid.org/0000-0003-3033-2404>
Escuela Superior Politécnica de Chimborazo, Facultad de Informática y Electrónica.
Riobamba, Ecuador
vi_ramos@esepoch.edu.ec
- ² Natalia Patricia Layedra Larrea  <https://orcid.org/0000-0003-1017-1746>
Escuela Superior Politécnica de Chimborazo, Facultad de Informática y Electrónica.
Riobamba, Ecuador
nlayedra@esepoch.edu.ec
- ³ Miguel Ángel Duque Vaca  <https://orcid.org/0000-0001-9442-2420>
Escuela Superior Politécnica de Chimborazo, Facultad de Informática y Electrónica.
Riobamba, Ecuador
m_duque@esepoch.edu.ec
- ⁴ Oscar Danilo Gaviláñez Álvarez  <https://orcid.org/0000-0002-7245-5640>
Escuela Superior Politécnica de Chimborazo, Facultad de Informática y Electrónica.
Riobamba, Ecuador
oscar.gavilanez@esepoch.edu.ec



Artículo de Investigación Científica y Tecnológica

Enviado: 15/05/2022

Revisado: 20/06/2022

Aceptado: 01/07/2022

Publicado: 29/07/2022

DOI: <https://doi.org/10.33262/ap.v4i3.248>

Cítese:

Ramos Valencia, M. V., Layedra Larrea, N. P., Duque Vaca, M. Ángel, & Gaviláñez Álvarez, O. D. (2022). Mejora del sistema de almacenamiento de un servidor de backups mediante la evaluación de los sistemas de duplicación Opendedup (SDFS) Y ZFS . AlfaPublicaciones, 4(3), 209–225. <https://doi.org/10.33262/ap.v4i3.248>



ALFA PUBLICACIONES, es una Revista Multidisciplinar, **Trimestral**, que se publicará en soporte electrónico tiene como **misión** contribuir a la formación de profesionales competentes con visión humanística y crítica que sean capaces de exponer sus resultados investigativos y científicos en la misma medida que se promueva mediante su intervención cambios positivos en la sociedad. <https://alfapublicaciones.com>

La revista es editada por la Editorial Ciencia Digital (Editorial de prestigio registrada en la Cámara Ecuatoriana de Libro con No de Afiliación 663) www.celibro.org.ec



Esta revista está protegida bajo una licencia Creative Commons Attribution Non Commercial No Derivatives 4.0 International. Copia de la licencia: <http://creativecommons.org/licenses/by-nc-nd/4.0/>

Palabras**claves:**

Sistemas de
duplicación,
Bacula, SDFS,
ZFS, Backups,
VirtualBox.

Keywords:

Duplication
systems,
Bacula, SDFS,
ZFS, Backups,
VirtualBox.

Resumen

Introducción. Un servidor de backups tiene un programa dedicado a la recuperación y organización de información para el aseguramiento de esta, procurando prevenir la pérdida de información por motivos tanto internos como externos. **Objetivo.** Mejorar el sistema de almacenamiento de un servidor de Backups mediante la evaluación de los sistemas de duplicación Opendedup (SDFS) Y ZFS. **Metodología.** Para el desarrollo se elaboró un caso de prueba para evaluar los sistemas propuestos. Al implementar herramientas de licencia independientes como Virtualbox, Bacula, etc., la instalación de soluciones de deduplicación se realiza bajo una distribución virtual Linux Mint 19.1 con Virtualbox, instalada en una sola máquina física, y también un servidor Bacula en el que se realizan las copias de seguridad. se creará y almacenará en nuestro propio entorno con la duplicación de SDFS y ZFS. **Resultados.** El sistema SDFS tuvo un cumplimiento del 50% en la comparación inicial frente al 81% de ZFS, además SDFS obtuvo 4 puntos de uso en la comparación a posterior del funcionamiento, mientras que ZFS solo tuvo 12 puntos de uso, por último, SDFS logró 2 puntos en comparación de funcionalidades superado por los 5 puntos de ZFS. **Conclusión.** ZFS ofrece un mejor rendimiento en términos de optimización del almacenamiento en el servidor de respaldo, a pesar de las dificultades encontradas durante la instalación, el método de deduplicación es superior a la ofrecida por SDSF en función de la capacidad de caché. Recomendamos crear copias de seguridad en formato de archivo incremental, porque con este tipo de datos el porcentaje de excepciones duplicadas es más pronunciado, con referencia a los datos subyacentes.

Abstract

Introduction. Introduction. A backup server has a program dedicated to the recovery and organization of information to secure it, trying to prevent the loss of information for both internal and external reasons. **Objective.** Improve the storage system of a Backup server by evaluating Opendedup duplication systems (SDFS) and ZFS. **Methodology.** For the development, a test case was elaborated to evaluate the proposed systems. When implementing standalone license tools like Virtualbox, Bacula, etc., the installation of deduplication solutions is done under a virtual

Linux Mint 19.1 distribution with Virtualbox, installed on a single physical machine, and also a Bacula server on which the deduplication is done. Backups. it will be created and stored in our own environment with SDFS and ZFS mirroring. **Results.** The SDFS system had 50% compliance in the initial comparison against 81% for ZFS, in addition SDFS obtained 4 points of use in the comparison after operation, while ZFS only had 12 points of use, finally SDFS increased 2 points in comparison of functionalities surpassed by the 5 points of ZFS. **Conclusion.** ZFS offers better performance in terms of storage optimization on the backup server, despite the difficulties encountered during installation, the deduplication method is superior to that offered by SDSF based on cache capacity. We recommend creating backups in incremental file format because with this type of data the percentage of duplicate exceptions is more pronounced, relative to the underlying data.

Introducción

La deduplicación de datos es una tecnología de reducción de datos que tiene como objetivo eliminar los datos redundantes (duplicados) en los sistemas de almacenamiento de datos, almacenando solo una copia de cada elemento, con el objetivo de reducir el espacio de almacenamiento, ahorrar energía y consumir ancho de banda (Álvarez, 2019).

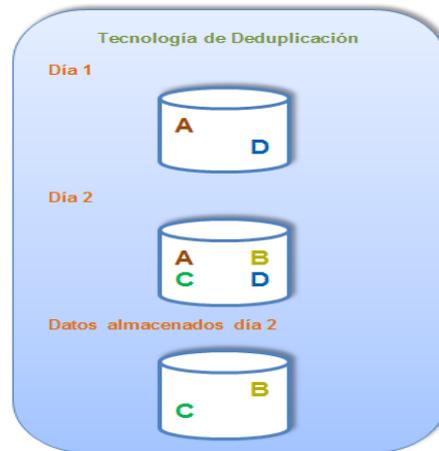
La tecnología redundante se basa en un índice que indica la ubicación de cada bloque de datos almacenado en el archivo. El programa que realiza esta operación busca cada nuevo dato ingresado en el sistema entre todos los datos previamente almacenados y almacena estos datos solo si no coincide con ningún dato ya almacenado (Lortu, 2019).

El proceso del sistema de deduplicación consiste almacenar una única copia de los datos almacenados en el disco duro y reemplazar las demás por un indicador a dicha copia. Al momento que un nuevo dato ingresa en el sistema éste es comparado con los datos almacenados previamente y será guardado únicamente si no coincide con los demás datos como se ve en la figura 1 (Mancheno, 2015).

Una ratio de deduplicación se define como la capacidad de optimización de datos en un sistema de almacenamiento que posee una solución de deduplicación. La ratio expresa la relación entre la cantidad de datos antes del proceso de deduplicación frente a la cantidad de datos después del proceso de deduplicación (Jiménez, 2019)

Figura 1

Proceso del sistema de deduplicación

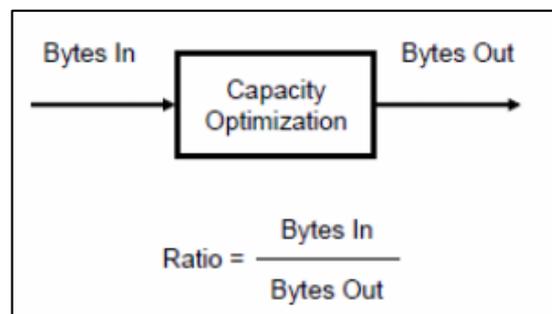


Fuente: Mancheno (2015)

Se calcula tomando la capacidad total de datos que se examinarán en busca de deduplicación y dividiéndola para la capacidad total de datos deduplicados (Whitehouse, 2021). A continuación, en la figura 2 se muestra el proceso para calcular las ratios de deduplicación:

Figura 2

Espacio reducción de las ratios de deduplicación



Fuente: Jiménez (2019)

Esta ratio se expresa generalmente en términos n:1 (p.e. 10:1 o 20:1). También puede utilizarse el porcentaje de reducción de espacio que puede definirse mediante la expresión: Porcentaje = $1 - (1/\text{Ratio})$. Si se toma el ejemplo de una ratio 10:1, esto quiere decir que la cantidad de espacio que necesita el total de datos deduplicados es 10 veces menor que la cantidad de espacio requerida para almacenar los mismos datos sin deduplicar (Jiménez, 2019).

Cuando las ratios de deduplicación son más grandes indica que los porcentajes de ahorro de capacidad también crecen, no obstante, este crecimiento empieza a ser lento cuando las ratios crecen a partir de un cierto punto. En la Tabla 1 se puede observar cómo se va dando esta conducta (Domínguez, 2020).

Tabla 1

Relación de la reducción de ratios y porcentajes

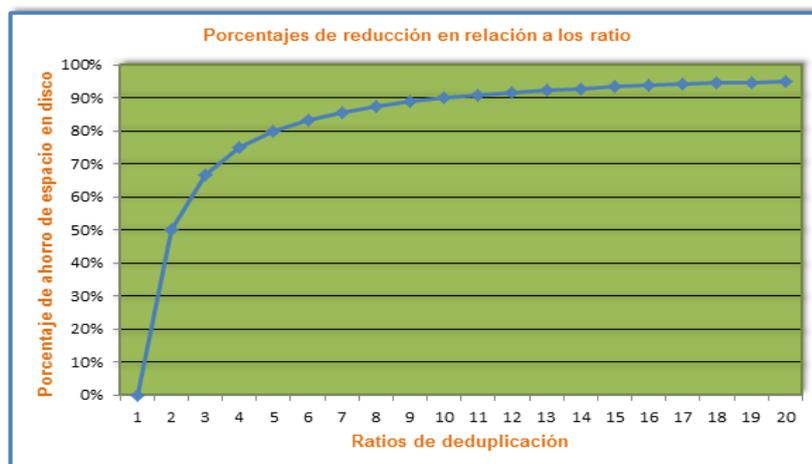
| RATIOS | PORCENTAJES = $(1 - \frac{1}{\text{Ratio}})$ |
|--------|--|
| 2:1 | $1/2 = 50\%$ |
| 5:1 | $4/5 = 80\%$ |
| 10:1 | $9/10 = 90\%$ |
| 20:1 | $19/20 = 95\%$ |
| 100:1 | $99/100 = 99\%$ |
| 500:1 | $499/500 = 99.8\%$ |

Fuente: Jiménez (2019)

Como se muestra arriba, el porcentaje aumenta lentamente con el aumento de las ratios de deduplicación. En la figura 3, se puede comprobar el comportamiento de la curva a medida que aumenta la ratio, y su crecimiento es lento a partir de ratios superiores a 5. La cantidad total de datos redundantes en el sistema de almacenamiento muestra el límite de las ratios de reducción que la solución de deduplicación puede lograr:

Figura 3

Porcentajes de reducción en relación con las ratios



Fuente: Jiménez (2009)

Opendedup (SDFS)

Es un sistema de ficheros que permite deduplicación en línea. Es un sistema que permite un ahorro de más del 90% del espacio utilizado, elimina archivos redundantes de más de un Petabyte, su rendimiento supera 1Gbps, entre otros (Cloudadmins, 2020).

Las principales características de Opendedup son (Opendedup, 2022):

- Deduplicación Variable y Fija, deduplica datos estructurados y no estructurados.
- Reducción de Utilización de almacenamiento - La deduplicación puede reducir la utilización del almacenamiento hasta en un 90% -95%
- Escalabilidad - Opendedup puede deduplicar enormes cantidades de datos a través de hasta 126 nodos independientes
- Alta disponibilidad - Todos los datos de bloque puede ser replica a través de hasta 7 nodos
- Replicación - deduplicación eficiente replicación cuenta
- Apoyo VMWare - Trabaja con VMS - puede deduplicar en 4k bloque de tamaño. Esto es necesario para deduplicar máquinas virtuales con eficacia
- Almacenamiento flexibles - los datos deduplicados pueden ser almacenados localmente, en la red a través de múltiples nodos, o en la nube.
- Global Inline deduplicación - comparte datos Opendedup únicas entre volúmenes en tiempo real.
- Configuración independiente o agrupada.

Sistema de Archivos ZFS

Fue creado por Sun Microsystems y actualmente es propiedad de Oracle Corporation. Una de las principales ventajas de ZFS es que permite la deduplicación a nivel de clúster integrada, identificando y eliminando bloques de datos redundantes, optimizando así el espacio en disco necesario para almacenar dichos datos (Mugarza, 2020).

ZFS descartará cualquier bloque de datos que es idéntico a un bloque ya escrito, mientras se mantiene una referencia para que siempre pueda reproducir el mismo bloque cuando se lee. Como buena práctica se recomienda revisar cual será el espacio aproximado que se obtendrá de la implementación del sistema deduplicado (Oracle, 2022):

- Probarlo con algunos datos reales.
- Con el comando `zdb -S` se puede obtener una estimación o simulación
- Si el servidor va orientado al almacenamiento de máquinas virtuales, se deberían evaluar preguntas como ¿Cuántas máquinas son compatibles? ¿Con qué frecuencia están parcheados? ¿Qué tan probable es que las personas apliquen el mismo software / parches / datos a sus máquinas? ¿Cuántos GB de datos, es

probable que esto genere? ¿Puede llegar a un caso de prueba representativo, después de todo para hacer la conjetura más apropiada?

El ahorro de almacenamiento esperado viene dado por la siguiente ecuación.

Almacenamiento Total - (total de almacenamiento / relación de deduplicación esperado)
= Se esperaba Ahorro Almacenamiento

Como un ejemplo ficticio, se tiene una agrupación de almacenamiento de 10 TB a ser utilizado para el almacenamiento de imágenes de máquinas virtuales en un escenario de escritorio virtual. En una prueba rápida, se crea un pool de 1 TB y se copian algunos datos de VM existentes a la misma, lo que produjo una relación dedup de 2. Esto significa que sólo se necesitan unos 5 TB de capacidad para almacenar los 10 TB de datos gracias a la deduplicación, por lo tanto, se ahorraría 5 TB de almacenamiento en disco (Starwind Software, 2020). Se supone que 1TB de disco tiene un costo de 100 dólares, por lo cual el ahorro por la tecnología de deduplicación es 500 dólares en este caso (Steffens, 2021).

Métodos de almacenamiento de deduplicación

Deduplicación de longitud variable

Es un algoritmo de deduplicación de datos que desglosa un sistema de archivos en segmentos de datos de longitud variable de subarchivo con el fin de determinar segmentos exclusivos y repetitivos. Esto reduce dramáticamente el almacenamiento de respaldo durante el proceso de respaldo y recuperación (EMC, 2021).

La deduplicación de longitud variable es un método avanzado de desglose de un flujo de datos a través de puntos de apoyo sensibles al contexto. Este método de segmentación inteligente de subarchivos proporciona mayor eficiencia de almacenamiento para datos redundantes, independientemente de donde se hayan insertado los nuevos datos. Como el nombre sugiere, la longitud de segmentos varía, con lo que se consiguen mayores proporciones de deduplicación (EMC, 2021).

Deduplicación de longitud fija

Para determinar segmentos únicos repetitivos, la deduplicación de longitud fija es un algoritmo de deduplicación de datos que desglosa un sistema de archivos en segmentos de datos de longitud fija de subarchivo. La deduplicación de longitud variable es un enfoque más avanzado, que fija segmentos de longitud variable según sus patrones de datos interiores. Esto resuelve el problema de cambio de posición de los datos del enfoque de bloques de tamaño fijo (EMC, 2021).

Actualmente, con el aumento del uso de Internet, tenemos dos grandes grupos en los que se divide la información: información online (embebida) y offline (fuera de línea).

Asimismo, dado que la deduplicación de datos es una tecnología orientada al almacenamiento de datos, está sujeta a estas limitaciones y características (Hewlett Packard, 2020).

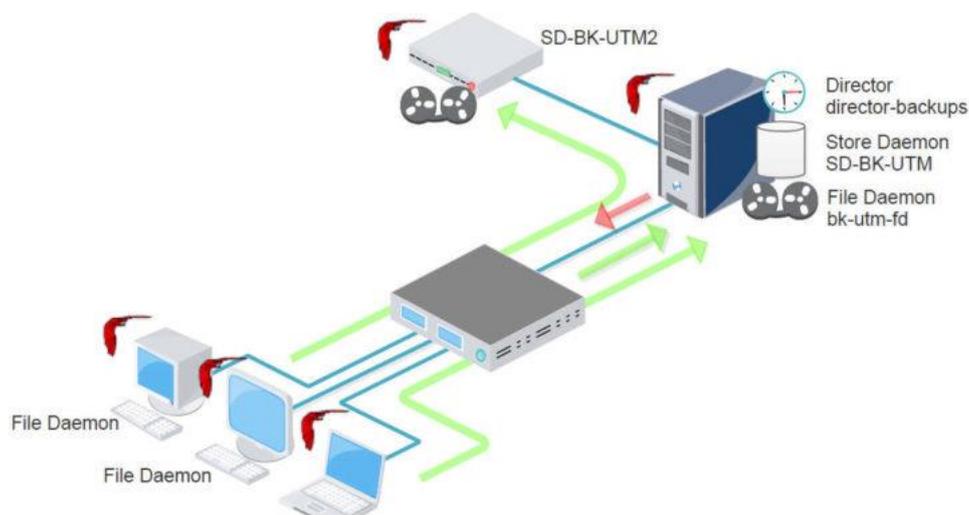
Entendiéndose por deduplicación offline, a ser él mismo capaz de identificar, preparar, deduplicar y manejar los datos existentes, no así a ser utilizado por una aplicación para almacenarlos, pues esta consta como un servidor instalado localmente, por ejemplo, Bacula (Mancheno, 2015).

Servidor de almacenamiento Bacula

Bacula es un sistema de administración de respaldo de red distribuido. Consta de tres componentes llamados Director (Dir), Storage Daemon (SD) y File Daemon (FD). cabeza de demonio, como muestra la figura 4.

Figura 4.

Esquema deduplicación Bacula



Fuente: Espitia (2021)

El director gestiona la programación de copias de seguridad (fecha y hora), así como la lista de clientes que solicitan copias de seguridad, con la capacidad de incluir o excluir archivos según su extensión. También es responsable de coordinar la comunicación entre los clientes y Storage Daemons. almacenamiento oculto (Espitia, 2021).

El Servicio de almacenamiento (SD) es responsable de administrar los medios de almacenamiento en los que se almacenan las copias de seguridad. Este componente recibe los archivos enviados directamente por el cliente (File Daemon) y los convierte en copias de seguridad almacenadas en los discos. El tercer componente, File Daemon, es el último

componente que se ejecuta en clientes (computadoras con Windows, Linux y Mac OSX) (Espitia, 2021).

Este proceso recibe instrucciones del administrador, como copiar carpetas y archivos y enviarlos a la tarjeta SD especificada. Como sistema distribuido, esta solución puede extenderse a múltiples clientes si es necesario y múltiples demonios de almacenamiento si es necesario. Este esquema distribuido no permite un equilibrio de carga óptimo de las copias de seguridad, que en ocasiones pueden superar los 800 GB en un solo día. Por lo tanto, debe tener una red que funcione lo mejor que pueda a 1 Gbps y un buen dispositivo en el lado SD (Espitia, 2021).

En general, Bacula es una solución de copia de seguridad potente, completa, flexible y escalable para copias de seguridad de red que merece un entorno profesional. Además, al ser un sistema de respaldo de información, encaja perfectamente en el tema de la seguridad informática (Espitia, 2021).

Metodología

Esta sección se enfocará en determinar los parámetros básicos que permitan elegir un sistema de archivos adecuado para el desempeño futuro y el correcto funcionamiento del servidor de respaldo, optimizando recursos, como el espacio en disco o el uso de datos de Internet (Mancheno, 2015). Estos parámetros se generan a partir de investigaciones sobre temas anteriores, tales como:

- Establecimiento arquitectónico de la interacción de diferentes sistemas de archivos en una misma computadora.
- Selección del servidor a utilizar para mejorar el proceso de extracción de copias de seguridad.
- Definir las pautas recomendadas para la implementación de servidores en Linux.

Medidas para elegir un sistema de archivos no duplicado para la implementación en el servidor de copia de seguridad

Estos parámetros han sido considerados de acuerdo con recomendaciones de varios autores y expertos en sistemas de deduplicación, como también si existe o no la característica en un primer paso comparativo teórico como lo muestra la tabla 2.

Tabla 2

Parámetros comparación sistemas de duplicación

| INDICADOR | CONCEPTO |
|----------------|--|
| Página Oficial | Hay un sitio web para el proyecto en sí, que proporcionará información o enlaces de descarga y soporte para el uso del software. |

Tabla 2

Parámetros comparación sistemas de duplicación (continuación)

| INDICADOR | CONCEPTO |
|---|---|
| Documentación en español/Inglés | Los documentos, archivos, enlaces o cualquier información útil compartida con el Software, o en su lugar de origen, están en español/inglés. |
| Foros/ Blogs | Un espacio de comunicaciones y mensajes en línea donde se pueden hacer y responder preguntas sobre el proyecto. |
| Ratio de deduplicación | El valor porcentual indica la relación de tamaño relativo entre el tamaño del disco original y el tamaño del disco resultante. |
| Tiempo en generar la copia de seguridad inicial | La medida en segundos es hacer la primera copia de seguridad del archivo. |
| Porcentaje de incremento por MB | Aumente el tamaño en megabytes entre la primera y la segunda copia de seguridad (las primeras copias de seguridad incrementales). |
| Taza de MB/s de escritura | La velocidad en MB/s a la que se escriben los datos en el disco duro de destino, durante la deduplicación. |
| Existencia de paquetes | Un paquete de software desarrollado por una empresa o comunidad y destinado a un sistema operativo o distribución en particular. |
| Método de almacenamiento | El algoritmo o algoritmos mediante los cuales se procesan los datos almacenados y cómo se organizan esos datos. Indica dónde y cómo permitir que se almacenen los datos |
| Flexibilidad de almacenamiento | Trata acerca del lugar y la manera que permite el almacenamiento de datos una tecnología. |
| Soporte de deduplicación on-line. | Se permite el almacenamiento remoto de datos. |
| Soporte de deduplicación off-line. | Habilita el almacenamiento local de archivos. |
| Soporte de Compresión | Indica si se puede aplicar un algoritmo de compresión además de eliminar datos redundantes para optimizar el espacio. |

Fuente: (Mancheno,2015)

Con la finalidad de comprender en forma clara y precisa el modelo y la interpretación de los resultados se ha realizado un estudio en tres partes:

Inicial: En esta etapa se realizará una comparativa teórica de las características y funcionalidades de los sistemas ZFS y SDFS de acuerdo con la siguiente ponderación descrita en la tabla 3.

Tabla 3

Escala Cuantitativa / Cualitativa

| Tipo | Valorización | | | | |
|--------------|---------------|------------|---------------|----------|--------------|
| Porcentaje | 0 | 25 | 50 | 75 | 100 |
| Cuantitativa | 0 | 1 | 2 | 3 | 4 |
| Cualitativa | Nada adecuado | Inadecuado | Poco adecuado | Adecuado | Muy adecuado |

Posterior: En esta etapa realizaremos la comparativa practica en un escenario dispuesto, como lo muestra la figura 4, para la prueba de los demás parámetros descritos en la Tabla 2 y el comportamiento de las características y funcionalidades de los sistemas ZFS y SDFS (Mancheno, 2015). En otras palabras, se tomarán las medidas en base a los resultados obtenidos durante la instalación y configuración del ambiente, para concluir que método nos brindará propiedades óptimas en nuestro servidor de respaldo.

Debido a la naturaleza de los datos, se decidió medir el uso comparativo entre los dos, según la tabla 3, que representan el uso más alto y su funcionalidad en el sistema de archivos correspondiente, recordemos que Optimizar = Utilizar menos recursos.

Funcionamiento: Para esta parte realizaremos la ponderación utilizando el método de la regla 3, considerando que el parámetro 1 será medido en base a los resultados presentados en (Besembel, Rivero 2019), donde realizaron un estudio comparativo entre las técnicas de almacenamiento de información, luego declarar una tabla hash es la forma más eficiente de almacenar datos.

Resultados

Inicial. - Un factor extremadamente importante es la presencia de un paquete personalizado para la distribución que se utilizará. Su presencia supondría un valor de 4, aunque no tenerlo sería de 2, porque eso supondría un pequeño aumento de tensión, como se verifica en la tabla 4y figura 5. Tener una web oficial se valorará con 4 puntos si está en nuestro idioma y 3 puntos si está en inglés, 2 puntos si está en otro idioma y 0 si no existe. En cuanto a los documentos, se les dará un máximo de 3 puntos por estar en español, ya que son menos importantes que los parámetros anteriores. Para foros o blogs, también obtendrán un máximo de 3 puntos por las razones explicadas anteriormente.

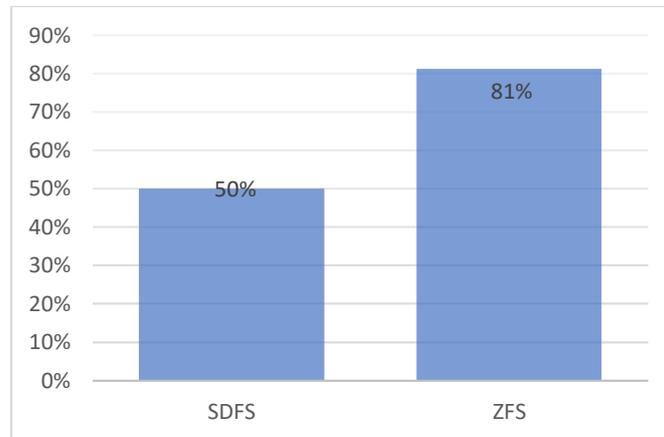
Tabla 4

Resultados Iniciales Sistemas Deduplicación

| Parámetro de Evaluación | Sistemas de Archivos Deduplicados | | |
|----------------------------------|-----------------------------------|-----|-----------|
| | SDFS | ZFS | Pond. Max |
| Página oficial. | 3 | 3 | 4 |
| Documentación en español/ inglés | 3 | 3 | 4 |
| Foro/Blog | 0 | 3 | 4 |
| Existencia de paquetes | 2 | 4 | 4 |
| Totales | 8 | 13 | 16 |
| | 50% | 81% | 100% |

Figura 5

Resultados iniciales entre SDFS y ZFS



Posterior. -Procedemos a calcular los pesos para cada línea por separado, teniendo en cuenta que el parámetro 1 y 3 representan el uso del espacio y el parámetro 2 representa el tiempo, y es necesario optimizar los recursos, por lo que en los pesos se considerarán mejores cuando sean inferiores. Por otro lado, el parámetro 4 representa la velocidad de escritura de datos en el disco duro de la computadora, que tiene una relación inversa entre este parámetro y el tiempo dedicado a una tarea, que es la mejor configuración al ser mayor, los resultados se muestran en la tabla 5 y figura 6 respectivamente.

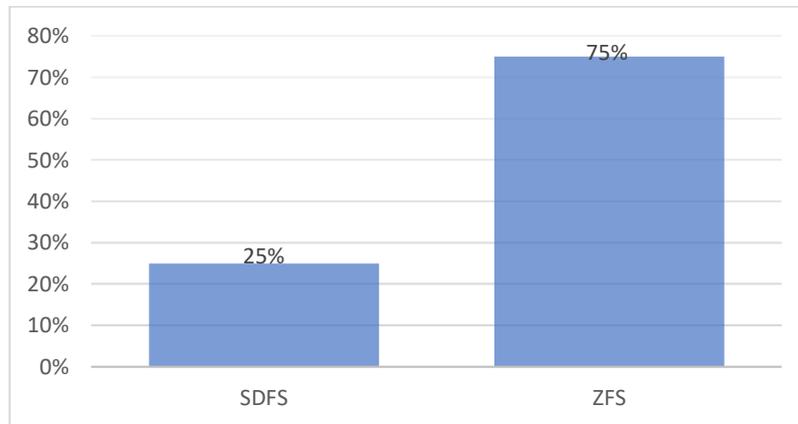
Tabla 5

Resultados Iniciales Sistemas Deduplicación

| Parámetro | Datos Funcionamiento | | Datos Ponderados | | |
|--|----------------------|---------|------------------|-----|-----------|
| | SDFS | ZFS | SDFS | ZFS | Pond. Max |
| Ratio de deduplicación (Gb) | 9.6 | 4.7 | 1 | 3 | 4 |
| Tiempo en generar el backup inicial (segundos) | 89.6 | 22 | 1 | 4 | 4 |
| Porcentaje de incremento por MB | 871.71% | 196.97% | 1 | 4 | 4 |
| Taza de MB/s de escritura | 4960.3 | 4960.3 | 1 | 1 | 4 |
| | | Totales | 4 | 12 | 16 |
| | | | 25% | 75% | 100% |

Figura 6

Resultados posteriores entre SDFS y ZFS



Funcionamiento. - Como podemos observar en la Tabla 6 y figura 7, ZFS tiene un mejor método de almacenamiento y mejor algoritmo de deduplicación off-line que el sistema SDFS, además en las pruebas realizadas la compresión fue mejorada con grandes cantidades de almacenamiento.

Es importante resaltar que tiene mayor facilidad realizar la instalación SSDS en una partición ZFS montada, lo que requiere el proceso del sistema operativo en otro disco sea de manera forzada, sin embargo, una vez realizada la instalación de ZFS presenta mejoras significativas en la deduplicación.

Tabla 6

Resultados Funcionamiento Sistemas Deduplicación

| Parámetro | Datos Funcionamiento | | Datos Ponderados | |
|----------------------------------|--|--|------------------|------|
| | SDFS | ZFS | SDFS | ZFS |
| Método de Almacenamiento. | Bloque de deduplicación variable o estático | Hash de 256 bits | 0 | 1 |
| Flexibilidad De Almacenamiento. | Local, en red, en nodos múltiples, en la nube. | Local, en red, en nodos múltiples, en la nube. | 1 | 1 |
| Soporte de Inline Deduplicación | Si | Si | 1 | 1 |
| Soporte de Offline Deduplicación | No | Si | 0 | 1 |
| Soporte de Compresión | No | Si (LZ4) | 0 | 1 |
| Totales | | | 3 | 5 |
| | | | 60% | 100% |

Fuente: Mancheno (2015)

Figura 7*Resultados posteriores entre SDFS y ZFS*

Resultados finales

Durante y después de la medición inicial, ZFS proporciona más facilidad para grabar, en términos del soporte que ofrece, también muestra una gran ventaja sobre SDFS. Finalmente, la optimización del espacio resultante entre ZFS, que es la mejor opción para eliminar la redundancia, se comparará con el tamaño real de los archivos afectados.

En este caso, se realiza un nuevo trabajo de copia de seguridad de un total de 7,5 GB de archivos, que son imágenes ISO del mismo sistema operativo, lo que los hace aptos para la deduplicación específica de los datos. Ya para finalizar, la copia de seguridad resultante tiene un tamaño de 4,9 GB, lo que supone un gran ahorro en la deduplicación de la información.

Conclusiones

- Un sistema de archivos redundantes ahorra recursos al eliminar la información duplicada en el disco, realizar copias de seguridad con Bacula en un formato de archivo y aprovechar de manera eficiente las características de un sistema de archivos redundantes, lo que hace que la entrada sea mucho más fácil de navegar.
- SDFS ofrece capacidades muy limitadas, aunque está muy bien equilibrado con la facilidad de instalación, mientras que ZFS ofrece un alto rendimiento, también bien equilibrado con la dificultad de realizar una instalación forzada del sistema que se ejecuta en el disco preparado.
- El sistema SDFS tuvo un 50% de cumplimiento en la comparación inicial contra un 81% para ZFS, además SDFS obtuvo 4 puntos de uso en la comparación después de la operación, mientras que ZFS solo tuvo 12 puntos de uso,

- Finalmente, SDFS aumentó 2 puntos en comparación de funcionalidades superadas por los 5 puntos de ZFS

Referencias bibliográficas

- Alvarez, C. (2019). Conceptos básicos: Deduplicación y sistemas de respaldo. *NetApp*.
<http://www.netapp.com/es/communities/tech-ontap/es-tot-bb-depublication.aspx>
- Besembel I y Rivero D. (2019). Estructuras lineales de datos. http://sistemas.ing.ula.ve/pr3/unidad_3/tema6/pdf/estructuras_nolinealesdatos.pdf.
- CloudAdmins. (2020). Opendedup: Deduplicación en línea y Deduplicación fuera de línea. <http://www.cloudadmins.org/opendedup-deduplicacion-en-linea/>
- Dominguez, M. (2020) La deduplicación en el backup. <http://www.whitebearsolutions.com/la-deduplicacion-en-el-backup/>
- EMC. (2021). Glosario de EMC: Deduplicación de longitud variable <http://mexico.emc.com/corporate/glossary/variable-length-deduplication.htm>.
- Espitia Torrez, E. (2021). Implementación de un sistema de backups con Bacula versión 7.4.3 sobre Ubuntu MATE. Espacio Tecnológico. spaciotecnologico.co/bacula/
- Hewlett Packard. (2020). La deduplicación: un almacenamiento más eficaz. 2009. <http://h30458.www3.hp.com/es/es/smb/736096.html>
- Jimenez, F. (2019) Técnicas de deduplicación de datos y aplicación en librerías virtuales. http://oa.upm.es/1803/1/PFC_FRANCISCO_JAVIER_JIMENEZ_PATRICIO.pdf
- Lortu. (2019) ¿Qué es la deduplicación?. <http://www.lortu.es/Technology.aspx>
- Mancheno Samaniego, A. (2015). Evaluación de los sistemas de deduplicación Opendedup (SDFS) Y ZFS para optimizar el sistema de almacenamiento en un servidor de backups. <http://dspace.esoch.edu.ec/handle/123456789/4579>
- Mugarza, M. (2020). Virtualización con software libre: Virtualización del almacenamiento. <http://www.mikelmugarza.com/?p=219>
- Opendedup. (2022) Opendedup. <https://opendedup.org/odd/>
- Oracle. (2021). Propiedad de eliminación de datos duplicados de ZFS. http://docs.oracle.com/cd/E37929_01/html/E36655/gbscy.html#gjhbo

Starwind Software.(2020) Data Deduplication methods: File-Level vs Block-Level vs Byte-Level deduplication. <http://www.starwindsoftware.com/file-level-vs-block-level-vs-byte-level-deduplication>

Steffens, H.(2021) ¿Qué es la deduplicación de datos y cómo beneficia a las empresas?. <http://liacolombia.com/2010/06/%C2%BFque-es-la-desduplicacion-de-datos-y-como-beneficia-a-las-empresas/>

Whitehouse, L. (2021) Cuándo y como utilizar la tecnología de deduplicación de datos en el safeguard de disco. <http://searchdatacenter.techtarget.com/es/consejo/Comprender-las-ratios-de-deduplicacion-de-datos-en-los-sistemas-de-copia-de-seguridad>



El artículo que se publica es de exclusiva responsabilidad de los autores y no necesariamente reflejan el pensamiento de la **Revista Alfa Publicaciones**.



El artículo queda en propiedad de la revista y, por tanto, su publicación parcial y/o total en otro medio tiene que ser autorizado por el director de la **Revista Alfa Publicaciones**.



Indexaciones

