

Implementación y Evaluación de un Sistema RAID para Comparar su Rendimiento con un Disco Individual

Johana Catalina Ríos Torres
Universidad de Antioquia
jcatalina.rios@udea.edu.co

Juan Guillermo Preciado Zapata
Universidad de Antioquia
guillermo.preciado@udea.edu.co

<https://github.com/jcriostorres/RAID-System-Performance-Evaluation.git>

I. INTRODUCCIÓN

En el contexto actual de crecimiento exponencial de los datos, la optimización y protección del almacenamiento se ha convertido en un tema crucial tanto en el ámbito académico como en la industria. Este proyecto tiene como objetivo la implementación y evaluación de un sistema RAID (Redundant Array of Independent Disks), utilizando una computadora con recursos limitados (Intel con 4GB de RAM y Ubuntu 24.04), cuatro discos duros mecánicos y un adaptador PCI para su conexión. Este enfoque permite profundizar en conocimientos prácticos y teóricos relacionados con sistemas de almacenamiento, persistencia de datos y pruebas de rendimiento, mientras se exploran configuraciones RAID 0, 1 y 5.

La motivación para el desarrollo de este proyecto radica en comprender y analizar el impacto del RAID en la capacidad, velocidad, redundancia y confiabilidad del sistema de almacenamiento, comparándolo con el rendimiento de un disco mecánico tradicional de un terabyte. Esto proporciona una base sólida para aplicar conceptos fundamentales de sistemas operativos, como la gestión de dispositivos de entrada/salida, la consistencia de datos y la implementación de sistemas de archivos.

Para evaluar el desempeño de las configuraciones RAID, se llevaron a cabo pruebas detalladas de lectura y escritura secuencial y aleatoria utilizando herramientas como **KDiskMark**. Posteriormente, los datos recopilados fueron analizados utilizando métodos estadísticos como el análisis de varianza (ANOVA) implementado en un notebook de Python, lo que permitió validar los resultados de manera rigurosa. La metodología combinó la implementación práctica del sistema RAID con pruebas de rendimiento y análisis estadístico, proporcionando un enfoque integral al tema.

Los resultados obtenidos demuestran el impacto de cada configuración RAID en términos de velocidad y redundancia, destacando las fortalezas y limitaciones de cada nivel en escenarios con recursos limitados. Este trabajo no solo refuerza los conocimientos teóricos sobre almacenamiento, sino que

también aporta aplicaciones prácticas para la toma de decisiones en sistemas reales, como servidores y centros de datos.

II. MARCO TEÓRICO

La implementación de un sistema RAID (Redundant Array of Independent Disks) requiere comprender diversos conceptos teóricos relacionados con el almacenamiento, la gestión de datos y la arquitectura de sistemas operativos. A continuación, se describen los elementos más relevantes:

1. RAID: Conceptos Fundamentales

RAID es una tecnología que combina múltiples discos duros para mejorar el rendimiento, la redundancia o ambos. Fue diseñado para abordar dos desafíos principales del almacenamiento:

- **Rendimiento:** Mejorar las velocidades de lectura y escritura mediante la distribución de datos en múltiples discos (striping).
- **Redundancia:** Garantizar la integridad de los datos al replicarlos o calcular paridad, permitiendo la recuperación en caso de fallos de hardware.

Niveles de RAID

- **RAID 0:** Distribuye los datos entre varios discos para mejorar el rendimiento, pero no ofrece redundancia. Su capacidad es igual a la suma de los discos disponibles.
- **RAID 1:** Duplica los datos en dos discos (mirroring), proporcionando alta redundancia, pero sacrificando capacidad.
- **RAID 5:** Distribuye los datos y la paridad entre tres o más discos, combinando rendimiento y redundancia. Puede tolerar la falla de un disco sin pérdida de datos.

2. Sistemas de Entrada/Salida y Almacenamiento

Los sistemas RAID interactúan directamente con el sistema operativo para gestionar los dispositivos de entrada/salida (I/O). En Linux, herramientas como mdadm permiten

implementar RAID por software, mientras que adaptadores PCI facilitan la conexión de discos a sistemas con hardware limitado. Este aspecto del proyecto conecta con los temas de **gestión de dispositivos, buffering, y programación de discos**.

3. Persistencia y Consistencia de Datos

La persistencia de los datos en RAID se garantiza mediante técnicas como la duplicación (RAID 1) o la paridad (RAID 5). Estas estrategias permiten mantener la consistencia de los datos en caso de fallos. La gestión eficiente de paridad y datos distribuidos es un desafío clave para los sistemas operativos y de almacenamiento.

4. Herramientas para la Evaluación de Rendimiento

El rendimiento de los sistemas RAID puede evaluarse mediante herramientas específicas:

- **KDiskMark:** Una interfaz gráfica para pruebas de rendimiento de discos, útil para analizar patrones de lectura/escritura secuenciales y aleatorios. Estas herramientas permiten analizar aspectos clave como IOPS (operaciones por segundo) y consistencia en el rendimiento.

5. Métodos Estadísticos para el Análisis de Datos

El análisis de los datos obtenidos en las pruebas de rendimiento requiere herramientas estadísticas para validar los resultados:

- **ANOVA (Análisis de Varianza):** Permite comparar el impacto de las diferentes configuraciones RAID (0, 1, 5) sobre el rendimiento. Este método evalúa la significancia estadística de las diferencias en velocidades de lectura/escritura y redundancia.
- **Notebook en Python:** Facilita el análisis estadístico y la visualización de datos para identificar tendencias y patrones en el rendimiento de los sistemas.

III. METODOLOGIA

1. Búsqueda de Información y Preparación de Recursos

Objetivo:

Recopilar información técnica y teórica sobre sistemas RAID, configuraciones de almacenamiento, y herramientas de medición de rendimiento.

Acciones:

- Revisión bibliográfica de documentación técnica y estándares RAID.
- Selección de discos duros y recursos computacionales compatibles para realizar pruebas.

Justificación:

Garantizar un marco teórico robusto para fundamentar las decisiones experimentales y técnicas.

2. Implementación de Conexiones de Discos para

Configurar RAID

Objetivo:

Configurar físicamente los discos y conexiones necesarias para implementar RAID 0, RAID 1, RAID 5 y un disco único como referencia.

Acciones:

- Uso de múltiples discos duros de capacidades similares conectados al sistema mediante interfaces SATA y controladores compatibles.
- Configuración de los arreglos RAID utilizando herramientas nativas de Linux (e.g., mdadm).

Justificación:

RAID se implementó como estándar en sistemas Linux para garantizar la compatibilidad y la replicabilidad del experimento.

3. Configuración de los RAIDs en Linux

Objetivo:

Configurar cada arreglo RAID (RAID 0, RAID 1 y RAID 5) en un entorno Linux.

Acciones:

- Creación y verificación de cada configuración RAID utilizando mdadm.
- Montaje y formateo de las unidades para pruebas uniformes.
- Pruebas iniciales para validar el correcto funcionamiento de cada configuración.

Justificación:

Linux ofrece herramientas flexibles y detalladas para la configuración y monitoreo de RAID, permitiendo mayor control del experimento.

3. Pruebas de Rendimiento

Objetivo:

Medir el rendimiento de cada configuración RAID y del disco único.

Acciones:

- Uso de la herramienta KDiskMark para realizar pruebas de lectura/escrituras secuenciales (SEQ1M) y aleatorias (RND4K).
- Generación de múltiples mediciones por configuración para garantizar la validez estadística de los datos.

Justificación:

Las pruebas de rendimiento permiten evaluar el throughput en diferentes escenarios representativos de cargas de trabajo reales.

4. Extracción y Transformación de Datos

Objetivo:

Organizar y procesar los datos de rendimiento para su análisis estadístico.

Acciones:

- Extracción manual de los resultados generados por KDiskMark.
- Transformación y limpieza de los datos para su integración en un archivo CSV estructurado.
- Identificación y corrección de posibles inconsistencias o errores manuales.

Justificación:

La organización y estructuración de los datos son fundamentales para realizar un análisis estadístico preciso y reproducible.

5. Análisis Estadístico y Evaluación

Objetivo:

Determinar la significancia estadística de las diferencias de rendimiento entre las configuraciones RAID y el disco único.

Acciones:

- Uso de ANOVA de dos vías con interacción para analizar los efectos de las configuraciones de almacenamiento y el tipo de prueba en el throughput.
- Verificación de supuestos estadísticos (e.g., normalidad de residuos, homogeneidad de varianzas) y ajuste del modelo según sea necesario.
- Interpretación de los resultados para identificar configuraciones óptimas.

Justificación:

ANOVA permite evaluar de manera robusta las interacciones entre múltiples factores (tipo de almacenamiento y tipo de prueba).

6. Redacción del Informe

Finalmente, se elaboró un informe detallado que incluye la metodología, los resultados y las conclusiones sobre la utilidad y rendimiento de las configuraciones RAID en distintos escenarios.

IV. IMPLEMENTACIÓN

1. Preparación del Sistema

Hardware Utilizado:

Una computadora con recursos limitados equipada con:

- 4 discos duros.
- Tarjeta controladora PCI con puertos SATA adicionales para conectar los discos.
- 4 cables SATA y un cable de poder SATA-Molex para la alimentación de los discos.
- Un USB booteable con Ubuntu para instalar y ejecutar el sistema operativo.

Software Utilizado:

- Sistema operativo Linux Ubuntu instalado en un disco dedicado para gestionar las configuraciones RAID.
- mdadm: Herramienta principal de Linux utilizada para crear y administrar las configuraciones RAID.
- KDiskMark: Herramientas de benchmarking para medir el rendimiento en distintas configuraciones.

2. Implementación de las Configuraciones RAID

Configuración de RAID:



Se implementaron las configuraciones RAID 0, RAID 1 y RAID 5, así como una configuración de disco único como referencia.

Los arreglos RAID se crearon y gestionaron utilizando mdadm, asegurando su correcta configuración mediante comandos de Linux:

- Creación de cada arreglo.

RAID 0

- Requisitos: Mínimo 2 dispositivos de almacenamiento.
- Beneficio principal: Rendimiento en términos de lectura/escritura y capacidad.
- Aspectos a tener en cuenta: Asegúrese de tener copias de seguridad funcionales. Una falla en un solo dispositivo destruirá todos los datos de la matriz.

Comprobación de los discos instalados

```

juang@juang-G41T-R7:~$ lsblk -o NAME,SIZE,FSTYPE,TYPE,MOUNTPOINT
NAME        SIZE FSTYPE TYPE MOUNTPOINT
loop0       55,4M loop /snap/core18/2846
loop1       73,9M loop /snap/core22/1663
loop2        4K loop /snap/bare/5
loop3        74,3M loop /snap/core22/1564
loop4       269,8M loop /snap/firefox/4793
loop5        10,7M loop /snap/firmware-updater/127
loop6        11,1M loop /snap/firmware-updater/147
loop7       505,1M loop /snap/gnome-42-2204/176
loop8        91,7M loop /snap/gtk-common-themes/1535
loop9       289,8M loop /snap/kde-frameworks-5-core18/35
loop10      107,5M loop /snap/kdiskmark/99
loop11       10,5M loop /snap/snap-store/1173
loop12       38,8M loop /snap/snapd/21759
loop13      580K loop /snap/snapd-desktop-integration/178
loop14       568K loop /snap/snapd-desktop-integration/253
sda         149,1G disk
├─sda1       1M part
├─sda2      93,5G ext4 part
├─sda3      55,2G ext4 part /
├─sdb       931,5G disk
├─sdb1      931,5G ext4 part
├─sdc       465,8G disk
├─sdc1      465,8G ext4 part
├─sdd       931,5G disk
├─sdd1      931,5G ext4 part
└─sdd1      931,5G ext4 part
juang@juang-G41T-R7:~$

```

Creación y verificación - RAID

```

juan@juan-G41T-M7:~$ sudo mdadm --create --verbose /dev/md0 --level=0 --raid-devices=2 /dev/sdb /dev/sdd
mdadm: chunk size defaults to 512K
mdadm: partition table exists on /dev/sdb
mdadm: partition table exists on /dev/sdd but will be lost or
      meaningless after creating array
mdadm: partition table exists on /dev/sdd
mdadm: partition table exists on /dev/sdd but will be lost or
      meaningless after creating array
Continue creating array? y
mdadm: Defaulting to version 1.2 metadata
mdadm: array /dev/md0 started.
juan@juan-G41T-M7:~$ cat /proc/mdstat
Personalities : [raid0]
md0 : active raid0 sdd[1] sdb[0]
      1953269544 blocks super 1.2 512k chunks

```

Creación del sistema de archivos

```

juan@juan-G41T-M7:~$ sudo mkfs.ext4 -F /dev/md0
mke2fs 1.47.0 (5-Feb-2023)
Descartando los bloques del dispositivo: hecho
Se está creando un sistema de archivos con 488315136 bloques de 4k y 122885376 nodos.1
UUID del sistema de archivos: 3545b172-9203-42ad-8a57-a0c9f0da8b5
Respaldo del superbloque guardados en los bloques:
      32768, 98304, 163840, 229376, 294912, 819200, 884736, 1685632, 2654208,
      4096000, 7962624, 11239424, 20480000, 23887872, 71663616, 78675968,
      102400000, 214990848
Reservando las tablas de grupo: hecho
Escribiendo las tablas de nodos: hecho
Creando el fichero de transacciones (262144 bloques): hecho
Escribiendo superbloques y la información contable del sistema de archivos: hecho

```

Montaje del sistema de archivos

```

juan@juan-G41T-M7:~$ sudo mkdir -p /mnt/md0
juan@juan-G41T-M7:~$ ls
Descargas Documentos Escritorio Ingeniería Música Plantillas Público snap Videos
juan@juan-G41T-M7:~$ df -h | grep '/mnt/md0'
juan@juan-G41T-M7:~$ df -h | grep '/mnt/md0'
juan@juan-G41T-M7:~$ mount | grep '/mnt/md0'
juan@juan-G41T-M7:~$ sudo mount /dev/md0 /mnt/md0
juan@juan-G41T-M7:~$ df -h | grep '/mnt/md0'
/dev/md0      1.0T    28K    1.7T    1% /mnt/md0
juan@juan-G41T-M7:~$
$ ficheros      Tamaño Usados  Disp Us%o Montado en
/dev/sda3       54G    11G    42G   20% /
/dev/md0        1.0T    28K    1.7T    1% /mnt/md0
juan@juan-G41T-M7:~$

```

Guardar el diseño de la matriz

```

juan@juan-G41T-M7:~$ sudo mdadm --detail --scan | sudo tee -a /etc/mdadm/mdadm.conf
ARRAY /dev/md0 metadata=1.2 UUID=20b839e8:1497b8ab:17d3c882:885428fe
juan@juan-G41T-M7:~$ sudo update-initramfs -u
update-initramfs: Generating /boot/initrd.img-6.8.0-47-generic
juan@juan-G41T-M7:~$ echo '/dev/md0 /mnt/md0 ext4 defaults,nofail,discard 0 0' | sudo tee -a /etc/fstab
/dev/md0 /mnt/md0 ext4 defaults,nofail,discard 0 0
juan@juan-G41T-M7:~$

```

- Verificación y monitoreo del estado de los discos.

RAID 0 Funcionando

```

juan@juan-G41T-M7:~$ sudo mkdir -p /mnt/raid
juan@juan-G41T-M7:~$ sudo mount /dev/md0 /mnt/raid
mount: /mnt/raid: el dispositivo especial /dev/md0 no existe.
dmesg(1) may have more information after failed mount system call.
juan@juan-G41T-M7:~$ sudo mount /dev/md0 /mnt/raid
juan@juan-G41T-M7:~$ ls /mnt/raid
escritura_aleatoria.0.0  escritura_aleatoria.2.0  lost+found  testfile2
escritura_aleatoria.1.0  escritura_aleatoria.3.0  testfile

```

Raid 0, falla luego de retirado un disco

```

juan@juan-G41T-M7:~$ ls /mnt/raid
juan@juan-G41T-M7:~$ sudo mdadm --detail /dev/md0
[sudo] contraseña para juan:
/dev/md0:
      Version : 1.2
      Raid Level : raid0
      Total Devices : 1
      Persistence : Superblock is persistent

      State : inactive
      Working Devices : 1

      Name : juan-G41T-M7:0 (local to host juan-G41T-M7)
      UUID : 20b839e8:1497b8ab:17d3c882:885428fe
      Events : 0

      Number Major Minor RaidDevice State
       -   -   -   -   -   -
        0   8   32   -   -   removed

```

Información perdida

```

juan@juan-G41T-M7:~$ sudo mdadm --detail /dev/md0
[sudo] contraseña para juan:
mdadm: cannot open /dev/md0: No such file or directory
juan@juan-G41T-M7:~$ ls /mnt/raid

```

RAID 1

- Requisitos: Mínimo 2 dispositivos de almacenamiento.
- Beneficio principal: Redundancia entre dos dispositivos de almacenamiento.
- Cosas a tener en cuenta: dado que se mantienen dos copias de los datos, solo se podrá utilizar la mitad del espacio en disco.

Comprobación de los discos instalados (ver img Raid 0)

Creando la matriz y estado del Raid

```

juan@juan-G41T-M7:~$ sudo mdadm --create --verbose /dev/md0 --level=1 --raid-devices=2 /dev/sdb /dev/sdd
[sudo] contraseña para juan:
mdadm: Note: this array has metadata at the start and
      may not be suitable as a boot device. If you plan to
      store '/boot' on this device please ensure that
      your boot-loader understands md/v1.x metadata, or use
      --metadata=0.90
mdadm: partition table exists on /dev/sdd
mdadm: partition table exists on /dev/sdb but will be lost or
      meaningless after creating array
mdadm: size set to 976630464K
mdadm: automatically enabling write-intent bitmap on large array
Continue creating array? Y
mdadm: Defaulting to version 1.2 metadata
mdadm: array /dev/md0 started.

```

Creación del sistema de archivos (ver img Raid 0)

Montaje del sistema de archivos (ver img Raid 0)

Guardar el diseño de la matriz (ver img Raid 0)

- Verificación y monitoreo del estado de los discos.

RAID 1 Funcionando

```

      1      8      32      1
juan@juan-G41T-M7:~$ ls /mnt/md0
14.txt comando.txt lost+found
juan@juan-G41T-M7:~$

```

Fallando un disco duro del Raid 1 (Se puede ver que la información sigue visible)

```

juan@juan-G41T-M7:~$ sudo mdadm --fail /dev/md0 /dev/sdb
mdadm: set /dev/sdb faulty in /dev/md0
juan@juan-G41T-M7:~$ sudo mdadm --detail /dev/md0
/dev/md0:
      Version : 1.2
      Creation Time : Fri Nov 22 21:15:27 2024
      Raid Level : raid1
      Array Size : 976630464 (931.39 GiB 1000.07 GB)
      Used Dev Size : 976630464 (931.39 GiB 1000.07 GB)
      Raid Devices : 2
      Total Devices : 2
      Persistence : Superblock is persistent

      Intent Bitmap : Internal

      Update Time : Sat Nov 23 17:27:18 2024
      State : clean, degraded
      Active Devices : 1
      Working Devices : 1
      Failed Devices : 1
      Spare Devices : 0

      Consistency Policy : bitmap

      Name : juan-G41T-M7:0 (local to host juan-G41T-M7)
      UUID : 7f9b5ac6:8c7e80b4:ecf74a7e:ee9f63e8
      Events : 6316

      Number Major Minor RaidDevice State
       -   -   -   -   -   -
        0   8   0   0   removed
        1   8   32  1   active sync  /dev/sdc
        0   8   16  -   faulty    /dev/sdb

```

Información funcional

```

      Number Major Minor RaidDevice State
       -   -   -   -   -   -
        0   8   0   0   removed
        1   8   32  1   active sync  /dev/sdc
        0   8   16  -   faulty    /dev/sdb
juan@juan-G41T-M7:~$ ls /mnt/md0
14.txt comando.txt lost+found
juan@juan-G41T-M7:~$

```

RAID 5

- Requisitos: Mínimo de 3 dispositivos de almacenamiento.
- Beneficio principal: Redundancia con mayor capacidad utilizable.

- Aspectos a tener en cuenta: si bien la información de paridad se distribuye, se utilizará la capacidad de un disco para la paridad. RAID 5 puede tener un rendimiento muy bajo cuando se encuentra en un estado degradado.

Comprobación de los discos instalados (ver img Raid 0)

Creando la matriz y estado del Raid

```
juan@juan-G41T-M7:~$ sudo mdadm --create --verbose /dev/md0 --level=5 --raid-devices=3 /dev/sdd /dev/sdb /dev/sdc
[sudo] contraseña para juan:
mdadm: layout defaults to left-symmetric
mdadm: layout defaults to left-symmetric
mdadm: chunk size defaults to 512K
mdadm: /dev/sdd appears to contain an ext2fs file system
size=976762584K mtimes=Sat Nov 23 18:24:12 2024
mdadm: partition table exists on /dev/sdb
mdadm: partition table exists on /dev/sdb but will be lost or
meaningless after creating array
mdadm: partition table exists on /dev/sdc
mdadm: partition table exists on /dev/sdc but will be lost or
meaningless after creating array
mdadm: size set to 488254464K
mdadm: automatically enabling write-intent bitmap on large array
mdadm: largest drive (/dev/sdd) exceeds size (488254464K) by more than 1K
Continue creating array? y
mdadm: Defaulting to version 1.2 metadata
mdadm: array /dev/md0 started.
```

Creación del sistema de archivos (ver img Raid 0)

Montaje del sistema de archivos (ver img Raid 0)

Guardar el diseño de la matriz

```
juan@juan-G41T-M7:~$ cat /proc/mdstat
Personalities : [raid0] [raid1] [raid6] [raid5] [raid4] [raid0]
md0 : active raid5 sdc[3] sdb[1] sdd[0]
      976588928 blocks super 1.2 level 5, 512k chunk, algorithm 2 [3/2] [UU]
      [>.....] recovery = 1.9% (9308308/488254464) finish=236.2min speed=33784K/sec
      bitmap: 2/4 pages [8KB], 65536KB chunk

unused devices: <none>

juan@juan-G41T-M7:~$ sudo mdadm --detail --scan | sudo tee -a /etc/mdadm/mdadm.conf
ARRAY /dev/md0 metadata=1.2 spares=1 UUID=4fa035bd:32253b2d:9e95d141:76af37bc
juan@juan-G41T-M7:~$ sudo update-initramfs -u
update-initramfs: Generating /boot/initrd.img-6.8.0-49-generic
juan@juan-G41T-M7:~$ echo "/dev/md0 /mnt/md0 ext4 defaults,nofail,discard 0 0" | sudo tee -a /etc/fstab
/dev/md0 /mnt/md0 ext4 defaults,nofail,discard 0 0
```

- Verificación y monitoreo del estado de los discos.

RAID 5 Funcionando

```
juan@juan-G41T-M7:~$ ls /mnt/md0
comando.txt lost+found raid5
```

Fallando un disco duro del Raid 5 (Se puede ver que la información sigue visible y dos discos activos de los 3 usados)

```
juan@juan-G41T-M7:~$ mdadm --fail /dev/md0 /dev/sdc
mdadm: error opening /dev/md0: Permission denied
juan@juan-G41T-M7:~$ sudo mdadm --fail /dev/md0 /dev/sdc
[sudo] contraseña para juan:
mdadm: set /dev/sdc faulty in /dev/md0
juan@juan-G41T-M7:~$ sudo mdadm --detail /dev/md0
/dev/md0:
  Version : 1.2
  Creation Time : Sat Nov 23 20:12:48 2024
  Raid Level : raid5
  Array Size : 976588928 (931.27 GiB 999.95 GB)
  Used Dev Size : 488254464 (465.64 GiB 499.97 GB)
  Raid Devices : 3
  Total Devices : 3
  Persistence : Superblock is persistent

  Intent Bitmap : Internal

  Update Time : Sun Nov 24 06:56:51 2024
  State : clean, degraded
  Active Devices : 2
  Working Devices : 2
  Failed Devices : 1
  Spare Devices : 0

  Layout : left-symmetric
  Chunk Size : 512K
  Consistency Policy : bitmap

  Name : juan-G41T-M7:0 (local to host juan-G41T-M7)
  UUID : 4fa035bd:32253b2d:9e95d141:76af37bc
  Events : 5990

  Number Major Minor RaidDevice State
    0   8   48   0   active sync  /dev/sdd
    1   8   16   1   active sync  /dev/sdb
    -   0   0    -   removed
    3   8   32   -   faulty   /dev/sdc
```

Información funcional

```
juan@juan-G41T-M7:~$ ls /mnt/md0
comando.txt lost+found raid5
```

- Los discos fueron formateados y montados en el sistema para realizar pruebas uniformes. (Este ejercicio se realizó en todos los Raid Instalados)

1) Desmontar el Raid

```
juan@juan-G41T-M7:~$ sudo umount /mnt/md0
```

2) Se detiene el RAID

```
juan@juan-G41T-M7:~$ sudo mdadm --stop /dev/md0
mdadm: stopped /dev/md0
```

3) Se elimina el RAID

```
juan@juan-G41T-M7:~$ sudo mdadm --remove /dev/md0
mdadm: error opening /dev/md0: No such file or directory
```

4) Se limpian los disco que fueron usados en el RAID

```
juan@juan-G41T-M7:~$ sudo mdadm --zero-superblock /dev/sdc
juan@juan-G41T-M7:~$ sudo mdadm --zero-superblock /dev/sdb
juan@juan-G41T-M7:~$ sudo mdadm --zero-superblock /dev/sdd
```

5) Se verifica el estado del montaje y se consulta de nuevo la RAID

```
juan@juan-G41T-M7:~$ ls /mnt/md0
juan@juan-G41T-M7:~$ sudo mdadm --detail /dev/md0
mdadm: cannot open /dev/md0: No such file or directory
```

3. Realización de Pruebas de Rendimiento

Pruebas Realizadas:

Para cada configuración (RAID 0, RAID 1, RAID 5, y Disco Único), se realizaron un total de 28 pruebas de rendimiento.

Las pruebas incluían:

- Operaciones de lectura y escritura secuenciales (SEQ1M).
- Operaciones de lectura y escritura aleatorias (RND4K).

Herramientas utilizadas:

- KDiskMark: Para pruebas detalladas de throughput (MB/s).
- hdparm: Para obtener métricas básicas de lectura y escritura.

	Read [MB/s]	Write [MB/s]
SEQ1M Q8T1	167,61	154,66
SEQ1M Q1T1	159,84	110,03
RND4K Q32T1	0,84	2,23
RND4K Q1T1	0,56	2,40

4. Extracción y Organización de Datos

Proceso de Extracción:

- Los datos generados por KDiskMark fueron recopilados manualmente debido a la falta de una opción de exportación directa a CSV.
- Los resultados de cada prueba se organizaron en un dataset estructurado, con columnas que especifican:
- Configuración del almacenamiento (Storage).
- Tipo de prueba realizada (Medicion).
- Rendimiento medido (Throughput_MB_s).

ID	Fecha	Prueba	Storage	Medicion	Throughput_MB_s
1	11/16/2024	Prueba-1	RAID-0	SEQ1M Q8T1 Read	167.61
329		Prueba-14	RAID-1	SEQ1M Q8T1 Read	94.889
330		Prueba-14	RAID-1	SEQ1M Q1T1 Read	60.153
331		Prueba-14	RAID-1	RND4K Q32T1 Read	0.652
332		Prueba-14	RAID-1	RND4K Q1T1 Read	0.316
333		Prueba-14	RAID-1	SEQ1M Q8T1 Write	42.737
334		Prueba-14	RAID-1	SEQ1M Q1T1 Write	38.842
335		Prueba-14	RAID-1	RND4K Q32T1 Write	0.404
336		Prueba-14	RAID-1	RND4K Q1T1 Write	0.288

Tabla 1 daset

Creación del Dataset:

Se incluyeron un total de 896 registros (28 pruebas por cada configuración y cada tipo de prueba).

5. Análisis Estadístico

Herramientas Utilizadas:

- El análisis de los datos se realizó en un Jupyter Notebook utilizando Python.
- Las bibliotecas principales incluyeron:
- pandas y numpy: Para manipulación de datos.
- matplotlib y seaborn: Para visualización.
- statsmodels y scipy: Para pruebas estadísticas.

Análisis Realizado:

Verificación de Normalidad:

- Se utilizaron histogramas y pruebas de Shapiro-Wilk para evaluar la distribución de los residuos del modelo.

Análisis de Varianza (ANOVA):

- Se aplicó un ANOVA de dos vías con interacción para analizar el efecto de las configuraciones RAID y el tipo de medición en el rendimiento.

Visualizaciones:

- Se generaron gráficos de interacción (Storage vs. Medición), diagramas de caja (boxplots) y distribuciones de los datos para interpretar los resultados.

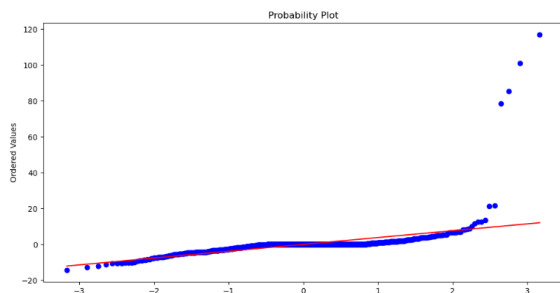


Figura 2 Gráfica de Probabilidad Normal para Evaluación de la Normalidad

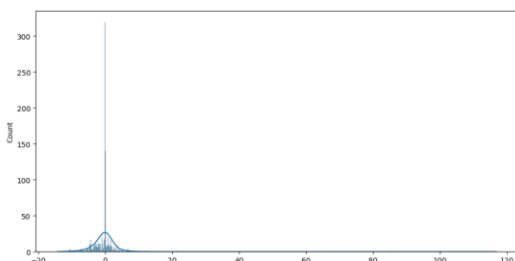


Figura 3 Histograma y Densidad de Datos

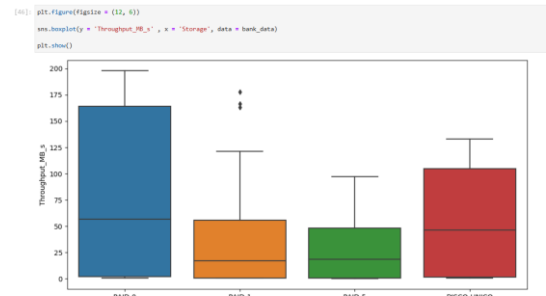


Figura 1 Análisis del Throughput (MB/s) Almacenamiento

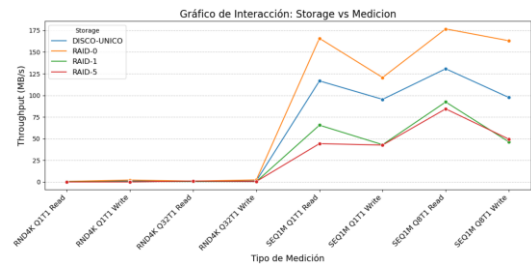


Figura 4 Comparación Throughput entre Almacenamiento y Tipos de Medición

7. Presentación de Resultados

Informe Final:

- Los resultados y análisis se documentaron en un informe con formato IEEE,
- **Repositorio en GitHub:**
- Los datos, scripts y resultados fueron publicados en un repositorio público para garantizar la transparencia y reproducibilidad:
- [Repositorio del Proyecto](#)

V. RESULTADOS

Durante la implementación de las configuraciones RAID en el sistema operativo Linux, se presentaron problemas de desconfiguración recurrentes. Estos problemas surgieron particularmente al reconfigurar cada RAID (RAID 0, RAID 1, RAID 5) y el disco único, lo cual resultó en retrasos significativos en la experimentación.

- Además, la lentitud general del sistema debido a la calidad de los discos utilizados fue un factor limitante en el rendimiento. Esto pudo afectar los tiempos de prueba y la consistencia en las mediciones.

- La extracción de los datos de rendimiento fue un proceso lento y tedioso debido a la falta de soporte de KDiskMark para exportar resultados en un formato adecuado (como CSV). La manipulación manual de los datos puede haber introducido errores de digitación o inconsistencias. Esto representa una posible fuente de error en el análisis de los resultados.
- Los resultados estadísticos se analizaron mediante un ANOVA de dos vías con interacción, que permitió determinar: RAID 0 presentó el mayor rendimiento, especialmente en operaciones secuenciales, mientras que RAID 1 y RAID 5 mostraron menores rendimientos debido a la redundancia.
- Las pruebas secuenciales (SEQ1M) tuvieron un rendimiento significativamente superior en comparación con las pruebas aleatorias (RND4K).

Comparación de Rendimiento RAID vs Disco Único				
Pruebas	Disco Único (MB/s)	RAID 0 (MB/s)	RAID 1 (MB/s)	RAID 5 (MB/s)
RND4K read-write	0.46 - 1.25	0.65 - 2.55	0.42 - 1.49	0.48 - 1.64
SEQ1M read-write	95 - 116	155 - 165	43 - 47	44 - 49

- Todos los resultados, datos obtenidos, y scripts de análisis fueron cargados al repositorio público de GitHub: <https://github.com/jcriostorres/RAID-System-Performance-Evaluation.git>
- Limitaciones del hardware: La calidad de los discos utilizados puede haber influido negativamente en los resultados de rendimiento.
- Errores de ETL: El traspaso manual de datos desde KDiskMark a archivos CSV puede haber introducido errores de digitación o datos inconsistentes.
- La normalidad de los residuos del modelo ANOVA no se cumplió completamente, como se verificó mediante pruebas gráficas (Q-Q plot) y la prueba de Shapiro-Wilk ($p=0.0$). Aunque esto no invalida el análisis, debido al tamaño grande de la muestra, podría afectar la interpretación de los intervalos de confianza y los valores p .

VI. CONCLUSIONES

- Los resultados obtenidos demuestran que **RAID 0** ofrece un notable incremento en las velocidades de lectura/escritura debido a la distribución de datos (striping), lo que lo hace ideal para aplicaciones donde el rendimiento es prioritario y la redundancia no es un requisito.
- **RAID 1**, aunque sacrifica la capacidad efectiva, proporciona una alta redundancia al duplicar los datos, siendo una opción confiable para entornos donde la seguridad de la información es crítica.

- **RAID 5** logra un equilibrio entre rendimiento y redundancia, permitiendo tolerancia a fallos con un menor impacto en la capacidad utilizable, aunque su rendimiento en escritura es menor debido al cálculo de la paridad.
- Herramientas como **KDiskMark**, y **hdparm** permitieron identificar patrones clave en el rendimiento de las configuraciones RAID, proporcionando datos detallados sobre velocidad, latencia y consistencia en las operaciones de entrada/salida.
- El análisis de varianza (ANOVA) en Python permitió validar estadísticamente las diferencias significativas en el rendimiento entre las configuraciones RAID, mostrando que las decisiones sobre qué nivel RAID utilizar deben basarse en los requisitos específicos de redundancia, capacidad y rendimiento.
- La implementación con hardware reciclado y software libre (Ubuntu 24.04 y mdadm) demostró que es posible construir sistemas RAID funcionales sin grandes inversiones, lo que resalta la viabilidad de soluciones de bajo costo en entornos académicos o domésticos.

VII. REFERENCIAS

- Silberschatz, A., Galvin, P. B., & Gagne, G. (2018). *Operating System Concepts* (10th ed.). Wiley.
- Linux RAID Wiki. (n.d.). *RAID setup* - ArchWiki.
- Bovet, D. P., & Cesati, M. (2005). *Understanding the Linux Kernel* (3rd ed.). O'Reilly Media.
- Mdadm - Managing RAID on Linux. (n.d.). *Linux Manual Pages*.
- ComputerNetworkingNotes. (n.d.). *How to Configure RAID in Linux Step by Step Guide*.
- LinuxBabe. (n.d.). *How to Set Up Software RAID 1 on an Existing Linux Distribution*.
- Seagate Technology. (n.d.). *An Introduction to RAID and How it Works*