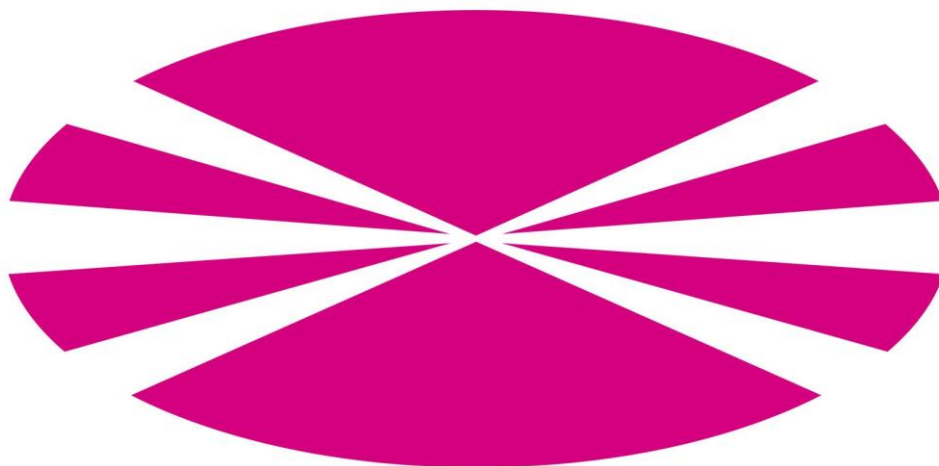


Estudio del rendimiento de diversas opciones para el soporte de sistemas de ficheros

Calidade na Xestión TIC

Maseda Dorado, Tomé



UNIVERSIDADE DA CORUÑA

Facultade de Informática

Universidade da Coruña

A Coruña, Spain

GUÍA DE DESPLIEGUE

Introducción

De las 16 opciones/combinaciones posibles se mostrará una guía de despliegue de una de ellas (para no ser repetitivos, el resto de despliegues serán similares cambiando algunas configuraciones que se explicarán brevemente a lo largo de la guía). En este caso vamos a montar un sistema de ficheros **ext4**, usando **RAID 5** para la redundancia de datos y **SSD** como tecnología de almacenamiento. Todos estos sistemas se desplegarán sobre instancias de AWS.

Construcción del RAID 5

En primer lugar, creamos una instancia nueva, para la que vamos a crear 3 volúmenes (SSD) desde el servicio de EBS de Amazon. (En caso de usar RAID 10 con 4 discos habría que montar 4 volúmenes, la configuración también puede variar si usamos HDD, habría que indicarlo en el *Tipo de volumen*).

The screenshot shows the AWS Management Console interface. The top navigation bar includes the AWS logo, 'Servicios', and a search bar. The left sidebar contains a navigation menu with categories like 'Panel de EC2', 'Eventos', 'Etiquetas', 'Límites', 'Instancias', 'Imágenes', 'Elastic Block Store', and 'Red y seguridad'. The main content area is titled 'Crear volumen' (Create volume). It features a table of existing volumes and a form to create a new volume.

Nombre	ID de volumen	Tamaño	Tipo de volumen	IOPS	Velocidad (I/s)	Snapshot	Creada	Zona de disponib.	Estado	Estado de alarm.	Información de conex.	Monitorizaci.	Estado de volu.	Cifrado
vol-057d1339	vol-057d1339	8 GiB	gp2	100	-	snap-0c3ac7f5	14 de junio de 2021	eu-west-3c	in-use	Algunos	+0865314294294311		Bien	No cifrado
vol-0a14c2b3	vol-0a14c2b3	8 GiB	gp2	100	-	snap-0c3ac7f5	26 de mayo de 2021	eu-west-3c	in-use	Algunos	+0a20b9a46da4c2102		Bien	No cifrado
vol-000a25d4	vol-000a25d4	8 GiB	gp2	100	-	snap-0c3ac7f5	26 de mayo de 2021	eu-west-3c	in-use	Algunos	+05c888bdf0fca053d		Bien	No cifrado

Crear volumen

Tipo de volumen: SSD de uso general (gp2) (Mín.: 1 GiB, máx.: 16384 GiB)

Tamaño (GiB): 1 (Mín.: 1 GiB, máx.: 16384 GiB)

IOPS: 100 / 3000 (Referencia de 3 IOPS por GiB con un mínimo de 100 IOPS, ampliable a 3000 IOPS)

Velocidad (MB/s): No aplicable

Zona de disponibilidad: eu-west-3c

ID de instancia: Seleccione una instancia

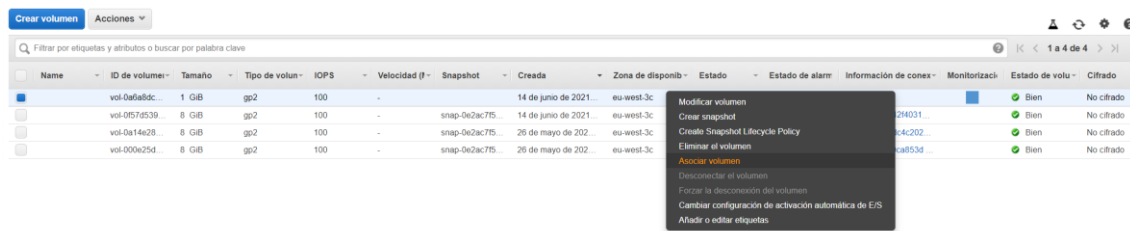
Cifrado: ☐ Cifrar este volumen

Clave: (127 caracteres como máximo) Valor: (255 caracteres como máximo)

Actualmente, este recurso no tiene etiquetas. Elija el botón Añadir una etiqueta o haga clic para añadir una etiqueta Nombre.

Añadir una etiqueta: 50 restantes (Hasta 50 etiquetas como máximo)

* Obligatorio Cancelar Crear volumen



Asociar volumen

Volumen ⓘ

vol-0a6a8dc312add55dc en eu-west-3c

Instancia ⓘ

en eu-west-3c

Dispositivo ⓘ

Dispositivos Linux: /dev/sdf through /dev/sdp

Nota: los kernels de Linux más recientes pueden cambiar el nombre de sus dispositivos a /dev/xvdf through /dev/xvdp internamente, aunque el nombre introducido aquí (y mostrado en las detalles) sea /dev/sdf through /dev/sdp.

Cancelar

Adjuntar

Name	ID de volumen	Tamaño	Tipo de volun	IOPS	Velocidad (I/s)	Snapshot	Creada	Zona de disponib	Estado	Estado de alarm	Información de conex	Monitorizaci	Estado de volu	Cifrado
vol-0271bbd...	vol-0271bbd...	1 GiB	gp2	100	-		14 de junio de 2021...	eu-west-3c	in-use	Ninguno	i-06853142b42f4031...		Bien	No cifrado
vol-0a04bbb...	vol-0a04bbb...	1 GiB	gp2	100	-		14 de junio de 2021...	eu-west-3c	in-use	Ninguno	i-06853142b42f4031...		Bien	No cifrado
vol-0a6a8dc...	vol-0a6a8dc...	1 GiB	gp2	100	-		14 de junio de 2021...	eu-west-3c	in-use	Ninguno	i-06853142b42f4031...		Bien	No cifrado
vol-057d539...	vol-057d539...	8 GiB	gp2	100	-	snap-0e2ac7f5...	14 de junio de 2021...	eu-west-3c	in-use	Ninguno	i-06853142b42f4031...		Bien	No cifrado

Una vez montados los volúmenes nos conectamos a la instancia.

Dado que después vamos a usar un sistema de ficheros remoto, para evitar confusiones y errores, vamos a cambiarle el nombre a la instancia:

```
sudo hostnamectl set-hostname FSlocal
```

Comprobamos que los volúmenes que hemos asociado están disponibles:

```
fdisk -l
```

```

Disk /dev/loop1: 55.46 MiB, 58142720 bytes, 113560 sectors
Units: sectors of 1 * 512 = 512 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 512 bytes / 512 bytes

Disk /dev/loop2: 70.39 MiB, 73797632 bytes, 144136 sectors
Units: sectors of 1 * 512 = 512 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 512 bytes / 512 bytes

Disk /dev/loop3: 32.28 MiB, 33841152 bytes, 66096 sectors
Units: sectors of 1 * 512 = 512 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 512 bytes / 512 bytes

Disk /dev/xvda: 8 GiB, 8589934592 bytes, 16777216 sectors
Units: sectors of 1 * 512 = 512 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 512 bytes / 512 bytes
Disklabel type: dos
Disk identifier: 0x5198cb0

Device      Boot Start      End  Sectors  Size Id Type
/dev/xvda1  *        2048 16777182 16775135   8G 83 Linux

Disk /dev/xvdf: 1 GiB, 1073741824 bytes, 2097152 sectors
Units: sectors of 1 * 512 = 512 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 512 bytes / 512 bytes

Disk /dev/xvdg: 1 GiB, 1073741824 bytes, 2097152 sectors
Units: sectors of 1 * 512 = 512 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 512 bytes / 512 bytes

Disk /dev/xvdh: 1 GiB, 1073741824 bytes, 2097152 sectors
Units: sectors of 1 * 512 = 512 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 512 bytes / 512 bytes
ubuntu@FSlocal:~$ █

```

Los discos se ven como discos sin formatear, para usarlos habrá que darles un formato, para cada volumen crearemos una partición que ocupe todo el disco de tipo 'Linux raid autodetec':

```
sudo fdisk /dev/xvdf
```

```

ubuntu@FSlocal:~$ sudo fdisk /dev/xvdf

Welcome to fdisk (util-linux 2.34).
Changes will remain in memory only, until you decide to write them.
Be careful before using the write command.

Device does not contain a recognized partition table.
Created a new DOS disklabel with disk identifier 0x055460f0.

Command (m for help): n
Partition type
   p   primary (0 primary, 0 extended, 4 free)
   e   extended (container for logical partitions)
Select (default p):

Using default response p.
Partition number (1-4, default 1):
First sector (2048-2097151, default 2048):
Last sector, +/-sectors or +/-size{K,M,G,T,P} (2048-2097151, default 2097151):

Created a new partition 1 of type 'Linux' and of size 1023 MiB.

Command (m for help): █

```

```

Command (m for help): t
Selected partition 1
Hex code (type L to list all codes): L

 0 Empty                24 NEC DOS               81 Minix / old Lin   bf Solaris
 1 FAT12                 27 Hidden NTFS Win    82 Linux swap / So  c1 DRDOS/sec (FAT-
 2 XENIX root            39 Plan 9              83 Linux             c4 DRDOS/sec (FAT-
 3 XENIX usr             3c PartitionMagic     84 OS/2 hidden or   c6 DRDOS/sec (FAT-
 4 FAT16 <32M           40 Venix 80286        85 Linux extended   c7 Syrix
 5 Extended              41 PPC PReP Boot     86 NTFS volume set  da Non-FS data
 6 FAT16                 42 SFS                87 NTFS volume set  db CP/M / CTOS / .
 7 HPFS/NTFS/exFAT      4d QNX4.x             88 Linux plaintext  de Dell Utility
 8 AIX                   4e QNX4.x 2nd part   8e Linux LVM        df BootIt
 9 AIX bootable         4f QNX4.x 3rd part   93 Amoebs           e1 DOS access
 a OS/2 Boot Manag     50 OnTrack DM        94 Amoebs BBT       e3 DOS R/O
 b W95 FAT32            51 OnTrack DM6 Aux  9f BSD/OS           e4 SpeedStor
 c W95 FAT32 (LBA)     52 CP/M              a0 IBM Thinkpad hi  ea Rufus alignment
 e W95 FAT16 (LBA)     53 OnTrack DM6 Aux  a5 FreeBSD         eb BeOS fs
 f W95 Ext'd (LBA)     54 OnTrackDM6        a6 OpenBSD         ee GPT
10 OPUS                 55 EZ-Drive          a7 NeXTSTEP        ef EFI (FAT-12/16/
11 Hidden FAT12         56 Golden Bow        a8 Darwin UFS       f0 Linux/PA-RISC b
12 Compaq diagnost     5c Priam Edisk        a9 NetBSD           f1 SpeedStor
14 Hidden FAT16 <3     61 SpeedStor         ab Darwin boot      f4 SpeedStor
16 Hidden FAT16         63 GNU HURD or Sys  af HFS / HFS+       f2 DOS secondary
17 Hidden HPFS/NTF     64 Novell Netware    b7 BSDI fs          fb VMware VMFS
18 AST SmartSleep      65 Novell Netware    b8 BSDI swap        fc VMware VMKCORE
1b Hidden W95 FAT3     70 DiskSecure Mult  bb Boot Wizard hid fd Linux raid auto
1c Hidden W95 FAT3     75 PC/IX             bc Acronis FAT32 L fe LANstep
1e Hidden W95 FAT1     80 Old Minix         be Solaris boot    ff BBT

Hex code (type L to list all codes): fd
Changed type of partition 'Linux' to 'Linux raid autodetect'.

Command (m for help): w
The partition table has been altered.
Calling ioctl() to re-read partition table.
Syncing disks.

```

Se repiten estos pasos para los volúmenes /dev/xvdg y /dev/xvdh.

Una vez formateados los volúmenes se ven así:

```
Disk /dev/xvda: 8 GiB, 8589934592 bytes, 16777216 sectors
Units: sectors of 1 * 512 = 512 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 512 bytes / 512 bytes
Disklabel type: dos
Disk identifier: 0x5198cbc0

Device      Boot Start      End  Sectors  Size Id Type
/dev/xvda1 *    2048 16777182 16775135   8G 83 Linux

Disk /dev/xvdf: 1 GiB, 1073741824 bytes, 2097152 sectors
Units: sectors of 1 * 512 = 512 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 512 bytes / 512 bytes
Disklabel type: dos
Disk identifier: 0x055460f0

Device      Boot Start      End  Sectors  Size Id Type
/dev/xvdf1   2048 2097151 2095104 1023M fd Linux raid autodetect

Disk /dev/xvdg: 1 GiB, 1073741824 bytes, 2097152 sectors
Units: sectors of 1 * 512 = 512 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 512 bytes / 512 bytes
Disklabel type: dos
Disk identifier: 0xale71fd8

Device      Boot Start      End  Sectors  Size Id Type
/dev/xvdg1   2048 2097151 2095104 1023M fd Linux raid autodetect

Disk /dev/xvdh: 1 GiB, 1073741824 bytes, 2097152 sectors
Units: sectors of 1 * 512 = 512 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 512 bytes / 512 bytes
Disklabel type: dos
Disk identifier: 0x8c17dbd9

Device      Boot Start      End  Sectors  Size Id Type
/dev/xvdh1   2048 2097151 2095104 1023M fd Linux raid autodetect
ubuntu@FSlocal:~$
```

Ahora desplegaremos un RAID 5 con las particiones de los volúmenes, se usan los 3 volúmenes como dispositivos activos:

```
sudo mdadm --create /dev/md0 --level=5 --raid-devices=3 /dev/xvdf1 /dev/xvdg1
/dev/xvdh1
```

Si se usase un RAID 10 habría que realizar los pasos anteriores para dos discos, luego repetirlos para los otros dos discos. Para cada pareja de discos habría que ejecutar mdadm con level=1.

Con los dos dispositivos resultantes (dos RAIDs 1) habría que ejecutar mdadm con level=0 para montar un RAID 0 con los dos RAIDs 1 anteriores, eso sería un RAID 10.

```

ubuntu@FSlocal:~$ sudo mdadm --create /dev/md0 --level=5 --raid-devices=3 /dev/xvdf1 /dev/xvdg1 /dev/xvdh1
mdadm: Fail create md0 when using /sys/module/md_mod/parameters/new_array
mdadm: Defaulting to version 1.2 metadata
mdadm: array /dev/md0 started.
ubuntu@FSlocal:~$ cat /proc/mdstat
Personalities : [raid6] [raid5] [raid4]
md0 : active raid5 xvdh1[3] xvdg1[1] xvdf1[0]
      2091008 blocks super 1.2 level 5, 512k chunk, algorithm 2 [3/3] [UUU]

unused devices: <none>
ubuntu@FSlocal:~$ sudo mdadm --detail /dev/md0
/dev/md0:
       Version : 1.2
  Creation Time : Mon Jun 14 08:41:52 2021
     Raid Level : raid5
   Array Size : 2091008 (2042.00 MiB 2141.19 MB)
  Used Dev Size : 1045504 (1021.00 MiB 1070.60 MB)
   Raid Devices : 3
  Total Devices : 3
 Persistence : Superblock is persistent

   Update Time : Mon Jun 14 08:42:08 2021
        State : clean
   Active Devices : 3
 Working Devices : 3
 Failed Devices : 0
  Spare Devices : 0

    Layout : left-symmetric
   Chunk Size : 512K

Consistency Policy : resync

       Name : FSlocal:0 (local to host FSlocal)
       UUID : ede0e9d0:b22a7d18:d0455250:d0f595e0
       Events : 18

   Number   Major   Minor   RaidDevice State
    0         202       81         0     active sync   /dev/xvdf1
    1         202       97         1     active sync   /dev/xvdg1
    3         202      113         2     active sync   /dev/xvdh1
ubuntu@FSlocal:~$

```

Ahora tenemos el dispositivo `/dev/md0` que es un disco virtual que está creado por encima de los volúmenes que normalmente son discos físicos (en este caso no, porque trabajamos con instancias de AWS, pero en una máquina real sería lo habitual).

Ahora crearemos una partición de tipo Linux en el RAID 5 (fdisk) y crearemos un sistema de ficheros de tipo EXT4 (mkfs):

```
ubuntu@FSlocal:~$ sudo fdisk /dev/md0

Welcome to fdisk (util-linux 2.34).
Changes will remain in memory only, until you decide to write them.
Be careful before using the write command.

Device does not contain a recognized partition table.
Created a new DOS disklabel with disk identifier 0x790ceb9e.

Command (m for help): n
Partition type
   p   primary (0 primary, 0 extended, 4 free)
   e   extended (container for logical partitions)
Select (default p):

Using default response p.
Partition number (1-4, default 1):
First sector (2048-4182015, default 2048):
Last sector, +/-sectors or +/-size{K,M,G,T,P} (2048-4182015, default 4182015):

Created a new partition 1 of type 'Linux' and of size 2 GiB.

Command (m for help): w
The partition table has been altered.
Calling ioctl() to re-read partition table.
Syncing disks.

ubuntu@FSlocal:~$ sudo fdisk -l /dev/md0
Disk /dev/md0: 1.102 GiB, 2141192192 bytes, 4182016 sectors
Units: sectors of 1 * 512 = 512 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 524288 bytes / 1048576 bytes
Disklabel type: dos
Disk identifier: 0x790ceb9e

   Device      Boot  Start      End  Sectors  Size Id Type
  /dev/md0p1             2048 4182015 4179968    2G 83 Linux
ubuntu@FSlocal:~$
```

El sistema de ficheros se crea sobre la partición que acabamos de crear /dev/md0p1. (Si el tipo del sistema de ficheros fuese xfs habría que usar *mkfs.xfs /dev/md0p1*)

```
ubuntu@FSlocal:~$ sudo mkfs.ext4 /dev/md0p1
mke2fs 1.45.5 (07-Jan-2020)
Creating filesystem with 522496 4k blocks and 130816 inodes
Filesystem UUID: e740fcd6-fa67-49cb-b002-c3c02897299c
Superblock backups stored on blocks:
    32768, 98304, 163840, 229376, 294912

Allocating group tables: done
Writing inode tables: done
Creating journal (8192 blocks): done
Writing superblocks and filesystem accounting information: done

ubuntu@FSlocal:~$
```


Ahora montamos el sistema de ficheros en una ruta del sistema para comprobar que está disponible para su uso, si fuese un sistema de ficheros XFS habría que indicarlo con la opción `-t` (`mount -t xfs /dev/md0p1 /var/raid`):

```
sudo mkdir -p /var/raid
sudo mount /dev/md0p1 /var/raid
df -h
```

```
ubuntu@FSlocal:~$ sudo mkdir -p /var/raid
ubuntu@FSlocal:~$ sudo mount /dev/md0p1 /var/raid
ubuntu@FSlocal:~$ df -h
```

Filesystem	Size	Used	Avail	Use%	Mounted on
/dev/root	7.7G	1.3G	6.5G	17%	/
devtmpfs	484M	0	484M	0%	/dev
tmpfs	490M	0	490M	0%	/dev/shm
tmpfs	98M	856K	98M	1%	/run
tmpfs	5.0M	0	5.0M	0%	/run/lock
tmpfs	490M	0	490M	0%	/sys/fs/cgroup
/dev/loop0	34M	34M	0	100%	/snap/amazon-ssm-agent/3552
/dev/loop1	56M	56M	0	100%	/snap/core18/1997
/dev/loop2	71M	71M	0	100%	/snap/lxd/19647
/dev/loop3	33M	33M	0	100%	/snap/snapd/11588
tmpfs	98M	0	98M	0%	/run/user/1000
/dev/md0p1	2.0G	6.0M	1.9G	1%	<u>/var/raid</u>

```
ubuntu@FSlocal:~$
```

Ahora hacemos el montaje persistente configurándolo en el fichero `/etc/fstab`. Para ello primero obtenemos el UUID de la partición con el comando `blkid`:

```
ubuntu@FSlocal:~$ sudo blkid
/dev/sda1: LABEL="cloudimg-rootfs" UUID="e8070c31-bfee-4314-a151-d1332dc23486" TYPE="ext4" PARTUUID="5198cbc0-01"
/dev/loop0: TYPE="squashfs"
/dev/loop1: TYPE="squashfs"
/dev/loop2: TYPE="squashfs"
/dev/loop3: TYPE="squashfs"
/dev/md0p1: UUID="e740fcd6-fa67-49cb-b002-c3c02897299c" TYPE="ext4" PARTUUID="790ceb9e-01"
/dev/svd1: UUID="ede0e9d0-b22a-7d18-d045-5250d0f595e0" UUID_SUB="c6a7ef4b-4f37-ff38-9812-214925c8c678" LABEL="FSlocal:0" TYPE="linux_raid_member" PARTUUID="055460f0-01"
/dev/svdh1: UUID="ede0e9d0-b22a-7d18-d045-5250d0f595e0" UUID_SUB="718e288d-c05c-c0f4-095e-68d2bafbadca" LABEL="FSlocal:0" TYPE="linux_raid_member" PARTUUID="a1e71fd8-01"
/dev/svdh1: UUID="ede0e9d0-b22a-7d18-d045-5250d0f595e0" UUID_SUB="2dc914ab-2d88-a147-a787-365079c4bfa5" LABEL="FSlocal:0" TYPE="linux_raid_member" PARTUUID="8c17dbd9-01"
ubuntu@FSlocal:~$
```

Copiamos el UUID y añadimos la siguiente línea al fichero `/etc/fstab` (si el sistema de ficheros fuese de tipo XFS habría que indicar `xfs` en vez de `ext4`):

```
UUID= e740fcd6-fa67-49cb-b002-c3c02897299c /var/raid ext4 defaults 0 0
```

```
GNU nano 4.8 /etc/fstab
LABEL=cloudimg-rootfs / ext4 defaults,discard 0 1
UUID= e740fcd6-fa67-49cb-b002-c3c02897299c /var/raid ext4 defaults 0 0
```

¡OJO! Si después reconstruimos el sistema de ficheros para usar XFS el UUID de la partición cambiará.

Comprobamos que se monta correctamente desmontando el sistema de ficheros y ejecutando `mount -a` (que monta todos los sistemas de ficheros del fichero `/etc/fstab`):

```
ubuntu@FSlocal:~$ df -h
Filesystem      Size  Used Avail Use% Mounted on
/dev/root        7.7G  1.3G  6.5G  17% /
devtmpfs         484M    0  484M   0% /dev
tmpfs            490M    0  490M   0% /dev/shm
tmpfs            98M   864K   98M   1% /run
tmpfs            5.0M    0   5.0M   0% /run/lock
tmpfs            490M    0  490M   0% /sys/fs/cgroup
/dev/loop0       34M   34M    0 100% /snap/amazon-ssm-agent/3552
/dev/loop1       56M   56M    0 100% /snap/core18/1997
/dev/loop2       71M   71M    0 100% /snap/lxd/19647
/dev/loop3       33M   33M    0 100% /snap/snapd/11588
tmpfs            98M    0   98M   0% /run/user/1000
/dev/md0p1       2.0G  6.0M  1.9G   1% /var/raid
ubuntu@FSlocal:~$ sudo umount /var/raid
ubuntu@FSlocal:~$ df -h
Filesystem      Size  Used Avail Use% Mounted on
/dev/root        7.7G  1.3G  6.5G  17% /
devtmpfs         484M    0  484M   0% /dev
tmpfs            490M    0  490M   0% /dev/shm
tmpfs            98M   864K   98M   1% /run
tmpfs            5.0M    0   5.0M   0% /run/lock
tmpfs            490M    0  490M   0% /sys/fs/cgroup
/dev/loop0       34M   34M    0 100% /snap/amazon-ssm-agent/3552
/dev/loop1       56M   56M    0 100% /snap/core18/1997
/dev/loop2       71M   71M    0 100% /snap/lxd/19647
/dev/loop3       33M   33M    0 100% /snap/snapd/11588
tmpfs            98M    0   98M   0% /run/user/1000
ubuntu@FSlocal:~$ mount -a
mount: only root can use "--all" option
ubuntu@FSlocal:~$ sudo mount -a
ubuntu@FSlocal:~$ df -h
Filesystem      Size  Used Avail Use% Mounted on
/dev/root        7.7G  1.3G  6.5G  17% /
devtmpfs         484M    0  484M   0% /dev
tmpfs            490M    0  490M   0% /dev/shm
tmpfs            98M   864K   98M   1% /run
tmpfs            5.0M    0   5.0M   0% /run/lock
tmpfs            490M    0  490M   0% /sys/fs/cgroup
/dev/loop0       34M   34M    0 100% /snap/amazon-ssm-agent/3552
/dev/loop1       56M   56M    0 100% /snap/core18/1997
/dev/loop2       71M   71M    0 100% /snap/lxd/19647
/dev/loop3       33M   33M    0 100% /snap/snapd/11588
tmpfs            98M    0   98M   0% /run/user/1000
/dev/md0p1       2.0G  6.0M  1.9G   1% /var/raid
ubuntu@FSlocal:~$
```

En este punto ya tendríamos el soporte para sistema de ficheros local montado y podríamos proceder a analizarlo, pero la parte de análisis la dejamos para un momento posterior, antes voy a enseñar en esta guía como hacer que este servicio sea exportable.

Compartición de datos con SSHFS

Para hacer el servicio exportable usaremos otra instancia desde la que accederemos a los datos en remoto.

Como se mencionó previamente, para evitar confusiones y errores, vamos a cambiarle el nombre a esta segunda instancia también:

```
sudo hostnamectl set-hostname FSremoto
```

A continuación, instalamos la utilidad sshfs en las dos instancias usando apt install.

```
sudo apt install sshfs
```

Para poder usar SSHFS entre las instancias de **FSlocal** y **FSremoto**, tenemos que realizar un intercambio de claves entre ellas. Como usuario **root**, creo una clave público/privada en **FSremoto**:

```
ssh-keygen
```

Obtengo la parte pública de la clave y la copio:

```
cat .ssh/id_rsa.pub
```

Guardo esa clave pública en la instancia **FSlocal**:

```
nano /root/.ssh/authorized_keys
```

Usando SSHFS, monto el directorio **/var/raid** de la instancia **FSlocal** en el mismo directorio de la instancia **FSremoto**.

```
sshfs -o nonempty -o allow_other -o IdentityFile=/root/.ssh/id_rsa  
root@172.31.37.236:/var/raid /var/raid
```

```
root@FSremoto:~# sshfs -o nonempty -o allow_other -o IdentityFile=/root/.ssh/id_rsa root@172.31.37.236:/var/raid /var/raid
root@FSremoto:~# df -h
Filesystem              Size  Used Avail Use% Mounted on
/dev/root                7.7G  1.5G  6.3G  19% /
devtmpfs                484M    0  484M   0% /dev
tmpfs                   490M    0  490M   0% /dev/shm
tmpfs                   98M    0   98M   1% /run
tmpfs                   5.0M    0   5.0M   0% /run/lock
tmpfs                   490M    0  490M   0% /sys/fs/cgroup
/dev/loop0              34M   34M    0 100% /snap/amazon-ssm-agent/3552
/dev/loop1              56M   56M    0 100% /snap/core18/1997
/dev/loop2              71M   71M    0 100% /snap/lxd/19647
/dev/loop3              33M   33M    0 100% /snap/snapd/11588
tmpfs                   98M    0   98M   0% /run/user/1000
root@172.31.37.236:/var/raid 2.0G  6.0M  1.9G   1% /var/raid
```

El soporte de sistema de ficheros remoto ya estará montado, a partir de ahora el tráfico de disco en el directorio **/var/raid** de **FSremoto** se traducirá en tráfico de red para **FSremoto** y tráfico de disco para **FSlocal**.

Otras configuraciones

Configuración utilizada para los volúmenes HDD

Crear volumen

Tipo de volumen

HDD de rendimiento optimizado (st1)

Tamaño (GiB)

125

(Min.: 125 GiB, máx.: 16384 GiB)

IOPS

No aplicable

Velocidad (MB/s)

5/31

(Base de referencia: 40 MB/s por TiB)

Zona de disponibilidad*

eu-west-3b

ID de instantánea

Seleccione una instantánea

Cifrado

☐

Cifrar este volumen

Clave

(127 caracteres como máximo)

Valor

(255 caracteres como máximo)

Actualmente, este recurso no tiene etiquetas

Elija el botón **Añadir una etiqueta** o [haga clic para añadir una etiqueta Nombre](#)

Añadir una etiqueta

50 restantes

(Hasta 50 etiquetas como máximo)

* Obligatorio

Cancelar

Crear volumen

Ejemplo (resumen) de creación de un RAID 10

```
ubuntu@FSlocal:~$ sudo mdadm --create /dev/md1 --level=1 --raid-devices=2 /dev/xvdf1 /dev/xvdg1
mdadm: /dev/xvdf1 appears to be part of a raid array:
  level=raid5 devices=3 ctime=Mon Jun 14 08:41:52 2021
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: /dev/xvdg1 appears to be part of a raid array:
  level=raid5 devices=3 ctime=Mon Jun 14 08:41:52 2021
Continue creating array? y
mdadm: Defaulting to version 1.2 metadata
mdadm: array /dev/md1 started.
ubuntu@FSlocal:~$ sudo mdadm --create /dev/md2 --level=1 --raid-devices=2 /dev/xvdh1 /dev/xvdi1
mdadm: /dev/xvdh1 appears to be part of a raid array:
  level=raid5 devices=3 ctime=Mon Jun 14 08:41:52 2021
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
Continue creating array? y
mdadm: Defaulting to version 1.2 metadata
mdadm: array /dev/md2 started.
ubuntu@FSlocal:~$
```

Devuelve un warning porque algunos volúmenes se estaban usando para un RAID 5 pero se crean los dos RAIDs 1 sin problema.

```

ubuntu@FSlocal:~$ sudo mdadm --detail /dev/md1
/dev/md1:
    Version : 1.2
    Creation Time : Mon Jun 14 13:41:55 2021
    Raid Level : raid1
    Array Size : 1046528 (1022.00 MiB 1071.64 MB)
    Used Dev Size : 1046528 (1022.00 MiB 1071.64 MB)
    Raid Devices : 2
    Total Devices : 2
    Persistence : Superblock is persistent

    Update Time : Mon Jun 14 13:42:11 2021
    State : clean
    Active Devices : 2
    Working Devices : 2
    Failed Devices : 0
    Spare Devices : 0

Consistency Policy : resync

    Name : FSlocal:1 (local to host FSlocal)
    UUID : 48c94108:d5cb0429:4fec52b8:043878d2
    Events : 17

    Number Major Minor RaidDevice State
       0   202    81        0   active sync  /dev/xvdf1
       1   202    97        1   active sync  /dev/xvdg1

```

```

ubuntu@FSlocal:~$ sudo mdadm --detail /dev/md2
/dev/md2:
    Version : 1.2
    Creation Time : Mon Jun 14 13:42:20 2021
    Raid Level : raid1
    Array Size : 1046528 (1022.00 MiB 1071.64 MB)
    Used Dev Size : 1046528 (1022.00 MiB 1071.64 MB)
    Raid Devices : 2
    Total Devices : 2
    Persistence : Superblock is persistent

    Update Time : Mon Jun 14 13:42:36 2021
    State : clean
    Active Devices : 2
    Working Devices : 2
    Failed Devices : 0
    Spare Devices : 0

Consistency Policy : resync

    Name : FSlocal:2 (local to host FSlocal)
    UUID : 57cac92f:d66edc10:70ac8aee:8e2af3b0
    Events : 17

    Number Major Minor RaidDevice State
       0   202   113        0   active sync  /dev/xvdh1
       1   202   129        1   active sync  /dev/xvdi1

```

Por último se crea el RAID 10 (RAID 0 con dos RAID 1).

```

ubuntu@FSlocal:~$ sudo mdadm --create /dev/md0 --level=0 --raid-devices=2 /dev/md1 /dev/md2
mdadm: Defaulting to version 1.2 metadata
mdadm: array /dev/md0 started.
ubuntu@FSlocal:~$ sudo mdadm --detail /dev/md0
/dev/md0:
   Version : 1.2
  Creation Time : Mon Jun 14 13:46:05 2021
    Raid Level : raid0
   Array Size : 2088960 (2040.00 MiB 2139.10 MB)
  Raid Devices : 2
 Total Devices : 2
 Persistence : Superblock is persistent

   Update Time : Mon Jun 14 13:46:05 2021
     State : clean
   Active Devices : 2
 Working Devices : 2
 Failed Devices : 0
  Spare Devices : 0

    Layout : -unknown-
   Chunk Size : 512K

Consistency Policy : none

           Name : FSlocal:0 (local to host FSlocal)
          UUID : 88250382:4d5b194b:061ded41:e0a02327
         Events : 0

   Number   Major   Minor   RaidDevice State
     0         9         1         0     active sync   /dev/md1
     1         9         2         1     active sync   /dev/md2

```

Ahora habría que crear el sistema de ficheros como en un RAID 5.

GUÍA DE ANALISIS

Para cada una de las 16 opciones repetiremos este procedimiento (el ejemplo mostrado es el de la guía de despliegue SSD – RAID 5 – EXT4). Si el análisis se está realizando sobre un sistema de ficheros de soporte local estresaremos el sistema de ficheros del directorio `/var/raid` de la instancia **FSlocal**, si se está realizando sobre un sistema de ficheros de soporte remoto, se hará lo mismo pero en la instancia **FSremoto**.

Estresaremos el sistema de ficheros (montado en `/var/raid`) de la instancia correspondiente (en este caso, **FSlocal**) con *fio* pasándole un tamaño de fichero de 1GB:

```
cd /var/raid
```

```

sudo fio --randrepeat=1 --ioengine=libaio --direct=1 --gtod_reduce=1 --name=test --
filename=random_read_write.fio --bs=4k --iodepth=64 --size=1G --readwrite=randrw --
rwmixread=75

```

En base a la salida de la ejecución del comando *fio* sacaremos las siguientes métricas:

Ancho de banda de lectura

```

read: IOPS=4696, BW=18.3MiB/s (19.2MB/s)(768MiB/41838msec)
bw (  KiB/s): min=17224, max=54752, per=100.00%, avg=18791.48, stdev=4024.51, samples=83
iops        : min= 4306, max=13688, avg=4697.83, stdev=1006.13, samples=83

```

Ancho de banda de escritura

```
write: IOPS=1569, BW=6276KiB/s (6427kB/s)(256MiB/41838msec); 0 zone resets  
bw ( KiB/s): min= 5712, max=18328, per=99.99%, avg=6275.59, stdev=1354.20, samples=83  
iops      : min= 1428, max= 4582, avg=1568.86, stdev=338.55, samples=83
```

Operaciones de entrada/salida de lectura

```
read: IOPS=4696, BW=18.3MiB/s (19.2MB/s)(768MiB/41838msec)  
bw ( KiB/s): min=17224, max=54752, per=100.00%, avg=18791.48, stdev=4024.51, samples=83  
iops      : min= 4306, max=13688, avg=4697.83, stdev=1006.13, samples=83
```

Operaciones de entrada salida de escritura

```
write: IOPS=1569, BW=6276KiB/s (6427kB/s)(256MiB/41838msec); 0 zone resets  
bw ( KiB/s): min= 5712, max=18328, per=99.99%, avg=6275.59, stdev=1354.20, samples=83  
iops      : min= 1428, max= 4582, avg=1568.86, stdev=338.55, samples=83
```

Para cada una de estas métricas se cogerá el valor medio (avg) devuelto por fio.

MÉTRICAS DE RENDIMIENTO EN SISTEMAS DE FICHEROS LOCALES

Para cada una de las 8 configuraciones posibles se mostrarán las siguientes métricas:

- Ancho de banda de lectura (BW read)
- Ancho de banda de escritura (BW write)
- Operaciones de entrada/salida de lectura (IOPS read)
- Operaciones de entrada/salida de escritura (IOPS write)

Todas las métricas obtenidas serán resultado de tres ejecuciones del comando fio, de las que se obtendrá la media de las tres ejecuciones para cada métrica.

Se mostrarán de la siguiente forma:

<metrica>: (<resultado1>-<resultado2>-<resultado3>) <resultado> KiB/s

SSD-RAID5-EXT4

BW read: (18791.48-18781.95-18795.11) **18789.51 KiB/s**

BW write: (6275.59-6270.55-6275.87) **6274.00 KiB/s**

IOPS read: (4697.83-4695.47-4698.77) **4697.36 KiB/s**

IOPS write: (1568.86-1567.64-1568.94) **1568.48 KiB/s**

SSD-RAID5-XFS

BW read: (18788.34-18792.73-18789.14) **18790.07 KiB/s**

BW write: (6273.70-6275.46-6273.10) **6274.09 KiB/s**

IOPS read: (4697.08-4698.18-4697.27) **4697.51 KiB/s**

IOPS write: (1568.41-1568.86-1568.25) **1568.51 KiB/s**

SSD-RAID10-EXT4

BW read: (30465.94-30481.33-30472.41) **30473.23 KiB/s**

BW write: (10179.14-10183.57-10181.29) **10181.33 KiB/s**

IOPS read: (7616.47-7620.25-7618.06) **7618.26 KiB/s**

IOPS write: (2544.76-2545.84-2545.31) **2545.3 KiB/s**

SSD-RAID10-XFS

BW read: (30486.10-30481.73-30479.31) **30482.38 KiB/s**

BW write: (10186.02-10184.31-10182.96) **10184.43 KiB/s**

IOPS read: (7621.51-7620.39-7619.78) **7620.56 KiB/s**

IOPS write: (2546.49-2546.08-2545.73) **2546.10 KiB/s**

HDD-RAID5-EXT4

BW read: (252.49-247.09-265.06) **254.88 KiB/s**

BW write: (97.34-97.67-110.11) **101.71 KiB/s**

IOPS read: (63.10-61.76-66.25) **63.70 KiB/s**

IOPS write: (24.31-24.40-27.52) **25.41 KiB/s**

HDD-RAID5-XFS

BW read: (253.57-256.44-262.12) **257.37 KiB/s**

BW write: (102.50-95.99-100.61) **99.7 KiB/s**

IOPS read: (63.37-64.09-65.51) **64.32KiB/s**

IOPS write: (25.61-23.98-25.13) **24.91 KiB/s**

HDD-RAID10-EXT4

BW read: (262.10-260.31-256.38) **259.60 KiB/s**

BW write: (91.51-89.95-89.25) **90.24 KiB/s**
IOPS read: (65.43-65.03-64.02) **64.83 KiB/s**
IOPS write: (22.78-22.43-22.23) **22.48 KiB/s**

HDD-RAID10-XFS

BW read: (268.08-269.68-260.63) **266.13 KiB/s**
BW write: (90.99-93.23-91.19) **91.80 KiB/s**
IOPS read: (67.00-67.42-65.15) **66.52 KiB/s**
IOPS write: (22.72-23.30-22.79) **22.94 KiB/s**

MÉTRICAS DE RENDIMIENTO EN SISTEMAS DE FICHEROS REMOTOS (SSHFS)

Para las 8 configuraciones posibles se seguirá exactamente la misma estructura que para el soporte local de sistemas de ficheros

SSD-RAID5-EXT4

BW read: (2245.68-2518.82-2567.29) **2443.93 KiB/s**
BW write: (751.32-841.35-857.51) **816.72 KiB/s**
IOPS read: (561.35-629.64-641.75) **610.91 KiB/s**
IOPS write: (187.76-210.27-214.31) **204.11 KiB/s**

SSD-RAID5-XFS

BW read: (2621.73-2525.36-2450.78) **2532.62 KiB/s**
BW write: (875.77-843.51-818.55) **845.94 KiB/s**
IOPS read: (655.37-631.28-612.63) **633.09 KiB/s**
IOPS write: (218.88-210.82-204.58) **211.43 KiB/s**

SSD-RAID10-EXT4

BW read: (2596.61-2618.65-2600.65) **2605.30 KiB/s**
BW write: (868.77-874.63-868.67) **870.69 KiB/s**

IOPS read: (649.08-654.6-650.1) **651.26 KiB/s**
IOPS write: (217.13-218.6-217.11) **217.61 KiB/s**

SSD-RAID10-XFS

BW read: (2546.13-2576.32-2510.63) **2544.36 KiB/s**
BW write: (850.43-860.47-838.61) **849.84 KiB/s**
IOPS read: (636.46-644.01-627.60) **636.02 KiB/s**
IOPS write: (212.54-215.06-209.59) **212.40 KiB/s**

HDD-RAID5-EXT4

BW read: (44.56-47.54-42.59) **44.90 KiB/s**
BW write: (20.84-22.32-19.09) **20.75 KiB/s**
IOPS read: (11.04-11.78-10.54) **11.12 KiB/s**
IOPS write: (5.11-5.46-4.66) **5.08 KiB/s**

HDD-RAID5-XFS

BW read: (56.33-51.62-41.25) **49.73 KiB/s**
BW write: (23.94-23.92-20.83) **22.90 KiB/s**
IOPS read: (13.95-12.80-10.20) **12.32 KiB/s**
IOPS write: (5.86-5.88-5.10) **5.61 KiB/s**

HDD-RAID10-EXT4

BW read: (132.86-143.28-146.21) **140.78 KiB/s**
BW write: (46.96-50.05-50.59) **49.2 KiB/s**
IOPS read: (22.63-35.70-36.43) **31.59 KiB/s**
IOPS write: (11.64-12.40-12.52) **12.19 KiB/s**

HDD-RAID10-XFS

BW read: (666.68-618.29-614.15) **633.04 KiB/s**
BW write: (230.18-206.80-205.35) **214.11 KiB/s**

IOPS read: (166.60-154.49-153.45) **158.18 KiB/s**

IOPS write: (57.44-51.59-51.23) **53.42 KiB/s**

Datos gráficamente explicados

Para exponer gráficamente los datos cree dos tablas, una en la que se analizan los sistemas de ficheros locales para cada una de las configuraciones y otra en la que se evalúan los sistemas de ficheros remotos para cada una de las configuraciones, ya que no tendría sentido compararlas entre ellas (existe un retardo de red que afecta en gran medida a la eficiencia de las configuraciones).

Sistema de ficheros local

Configuración ▾	BW Read ▾	BW Write ▾	IOPS Read ▾	IOPS Write ▾
SSD-RAID5-EXT4	18789.51	6274.00	4697.36	1568.48
SSD-RAID5-XFS	18790.07	6274.09	4697.51	1568.51
SSD-RAID10-EXT4	30473.23	10181.33	7618.26	2545.30
SSD-RAID10-XFS	30482.38	10184.43	7620.56	2546.10
HDD-RAID5-EXT4	254.88	101.71	63.70	25.41
HDD-RAID5-XFS	257.37	99.7	64.32	24.91
HDD-RAID10-EXT4	259.60	90.24	64.83	22.48
HDD-RAID10-XFS	266.13	91.80	66.52	22.94

Se puede apreciar que el tipo de sistema de ficheros no influye prácticamente en el rendimiento de este. Por otra parte, vemos que hay una gran diferencia de rendimiento de RAID 5 a RAID 10 usando SSDs, sin embargo, usando HDDs no. Por último apreciar la GRAN DIFERENCIA de rendimiento entre usar discos SSD y usar discos HDD.

Sistema de ficheros remoto

Configuración ▾	BW Read ▾	BW Write ▾	IOPS Read ▾	IOPS Write ▾
SSD-RAID5-EXT4	2443.93	816.72	610.91	204.11
SSD-RAID5-XFS	2532.62	845.94	633.09	211.43
SSD-RAID10-EXT4	2605.30	870.69	651.26	217.61
SSD-RAID10-XFS	2544.36	849.84	636.02	212.40
HDD-RAID5-EXT4	44.90	20.75	11.12	5.08
HDD-RAID5-XFS	49.73	22.90	12.32	5.61
HDD-RAID10-EXT4	140.78	49.2	31.59	12.19
HDD-RAID10-XFS	633.04	214.11	158.18	53.42

Las diferencias a destacar son prácticamente las mismas, solo que en este caso hubo más diferencia de rendimiento entre los RAIDs que usan HDD que los RAIDs que usan SSD (que prácticamente tuvieron el mismo rendimiento). Por último, destacar que la diferencia de rendimiento entre EXT4 y XFS para RAID10 con HDD, se debe a un fallo de exactitud, para hacer estas pruebas se están testeando los datos solo 3 veces y, en ciertos casos, nos podemos encontrar con datos anómalos.

Decisión tomada fundamentada en los datos

No hay una decisión clara, todo depende de con qué objetivo querramos montar el soporte de sistemas de ficheros:

El tipo de sistema de ficheros lo considero indiferente en estos casos al tener un rendimiento prácticamente equivalente, esa decisión la basaría en el sistema operativo donde se vaya a instalar el soporte y utilizaría el tipo de sistema de ficheros que más se adecúe.

¿Buscamos un soporte de sistema de ficheros que sea eficiente y en el que vamos a guardar datos poco importante?

Entonces la tecnología de almacenamiento a escoger es SSD y el array de redundancia de discos RAID 10, sin duda. RAID 10 es más eficiente que RAID 5 pero menos tolerante a fallos.

¿Buscamos un soporte de sistema de ficheros que sea eficiente y en el que vamos a guardar datos MUY IMPORTANTES?

Entonces seguiría escogiendo SSD como tecnología de almacenamiento pero usaría RAID 5 para la redundancia de datos (mayor tolerancia a fallos).

¿Buscamos un soporte de sistema de ficheros que no tiene porque ser muy eficiente y en el que vamos a guardar una cantidad muy grande de datos?

Entonces usaría HDD, el rendimiento se vería altamente penalizado, pero si no nos importa y nos lo podemos permitir, el coste del almacenamiento se va a rebajar mucho usando discos HDD. Para la redundancia de datos usaría RAID 5 por seguridad.

Respecto a las diferencias entre un soporte de sistemas de ficheros local y remoto, se ve claramente que hacer un sistema de ficheros remoto penaliza mucho el rendimiento, los datos tienen que viajar a través de la red y esto ralentiza todas las escrituras y lecturas mucho, por tanto, siempre que pudiese usaría un sistema de ficheros local.

Anexo vídeos

- test_ssd RAID10 xfs local
- test_ssd RAID10 sshfs
- test_hdd RAID5 ext4 sshfs

Los tres vídeos son una muestra de como obtuve las métricas para las configuraciones correspondientes.