

Instalación

Instalamos debian 11

Añadimos sources:

```
/etc/apt/sources.list
```

```
deb http://deb.debian.org/debian/ bullseye main contrib non-free
deb-src http://deb.debian.org/debian/ bullseye main main contrib non-free

deb http://security.debian.org/debian-security bullseye-security main
contrib
deb-src http://security.debian.org/debian-security bullseye-security main
contrib
```

Instalamos paquetes necesarios

```
apt-get update
```

```
apt install linux-headers-"$(uname -r)" linux-image-amd64
apt install zfs-dkms zfsutils-linux
```

Habilitamos servicios de ZFS y reiniciamos

```
systemctl enable zfs.target
systemctl enable zfs-import cache
systemctl enable zfs-mount
systemctl enable zfs-import.target
systemctl enable zfs-import-scan
systemctl enable zfs-share
```

Miramos los pool de zfs (no habrá ninguno)

```
zpool list
```

Miramos el id de los discos, en mi caso tengo 4 discos de 8tb:

```
ls -l /dev/disk/by-id
```

```
lrwxrwxrwx 1 root root 9 Dec 12 23:46 ata-
Samsung_SSD_850_EVO_120GB_S21UNSBG127800D -> ../../sdg
lrwxrwxrwx 1 root root 10 Dec 12 23:46 ata-
Samsung_SSD_850_EVO_120GB_S21UNSBG127800D-part1 -> ../../sdg1
lrwxrwxrwx 1 root root 10 Dec 12 23:46 ata-
Samsung_SSD_850_EVO_120GB_S21UNSBG127800D-part2 -> ../../sdg2
lrwxrwxrwx 1 root root 10 Dec 12 23:46 ata-
Samsung_SSD_850_EVO_120GB_S21UNSBG127800D-part5 -> ../../sdg5
```

```
lrwxrwxrwx 1 root root 9 Dec 12 23:46 ata-ST8000DM004-2CX188_WCT37KFE ->
../..sdb
lrwxrwxrwx 1 root root 9 Dec 12 23:46 ata-ST8000DM004-2CX188_WCT38QL5 ->
../..sdf
lrwxrwxrwx 1 root root 9 Dec 12 23:46 ata-ST8000DM004-2CX188_WCT38Y6B ->
../..sdc
lrwxrwxrwx 1 root root 9 Dec 12 23:46 ata-ST8000DM004-2CX188_ZCT1KCM9 ->
../..sda
lrwxrwxrwx 1 root root 9 Dec 12 23:46 usb-HP_iLO_Internal_SD-
CARD_000002660A01-0:0 -> ../..sde
lrwxrwxrwx 1 root root 10 Dec 12 23:46 usb-HP_iLO_Internal_SD-
CARD_000002660A01-0:0-part1 -> ../..sde1
lrwxrwxrwx 1 root root 9 Dec 12 23:46 usb-
Verbatim_STORE_N_GO_07930DA70298-0:0 -> ../..sdd
lrwxrwxrwx 1 root root 9 Dec 12 23:46 wwn-0x5000c500c520c2f7 -> ../..sda
lrwxrwxrwx 1 root root 9 Dec 12 23:46 wwn-0x5000c500cf855885 -> ../..sdc
lrwxrwxrwx 1 root root 9 Dec 12 23:46 wwn-0x5000c500cf875876 -> ../..sdf
lrwxrwxrwx 1 root root 9 Dec 12 23:46 wwn-0x5000c500cf8fda13 -> ../..sdb
lrwxrwxrwx 1 root root 9 Dec 12 23:46 wwn-0x5002538da01151be -> ../..sdg
lrwxrwxrwx 1 root root 10 Dec 12 23:46 wwn-0x5002538da01151be-part1 ->
../..sdg1
lrwxrwxrwx 1 root root 10 Dec 12 23:46 wwn-0x5002538da01151be-part2 ->
../..sdg2
lrwxrwxrwx 1 root root 10 Dec 12 23:46 wwn-0x5002538da01151be-part5 ->
../..sdg5
```

Son los discos ata-ST8000DM004-XXXXX_XXXXX que corresponden a sdb, sdf, sdc y sda. Cogemos los wwn-XXXXXX que son:

```
wwn-0x5000c500c520c2f7
wwn-0x5000c500cf855885
wwn-0x5000c500cf875876
wwn-0x5000c500cf8fda13
```

Cremos el pool en el directorio dades

```
zpool create -o ashift=12 -O acltype=posixacl -O compression=lz4 -O
relatime=on -O xattr=sa dades raidz1 wwn-0x5000c500c520c2f7
wwn-0x5000c500cf855885 wwn-0x5000c500cf875876 wwn-0x5000c500cf8fda13
```

Ya está creado:

```
zpool status
```

```
pool: dades
state: ONLINE
scan: none requested
config:
```

```
NAME
```

```
STATE
```

```
READ WRITE CKSUM
```

dades	ONLINE	0	0	0
raidz1-0	ONLINE	0	0	0
wwn-0x5000c500c520c2f7	ONLINE	0	0	0
wwn-0x5000c500cf855885	ONLINE	0	0	0
wwn-0x5000c500cf875876	ONLINE	0	0	0
wwn-0x5000c500cf8fda13	ONLINE	0	0	0

errors: No known data errors

Tunning

Voy a decir que la swap no salte al no ser que use el 90% porque zfs coge mucha memoria. Por defecto vienen al 40%

```
sysctl vm.swappiness
vm.swappiness = 60
```

Añado al final de:

```
/etc/sysctl.conf
vm.swappiness=10
```

Como tengo 16Gb de RAM, le digo que use 12gbs, Por defecto tiene 7.7Gb

```
arc_summary | grep "Target size "
Target size (adaptive):                100.0 %      7.7 GiB
```

Añadimos a la configuración de zfs 12Gb que son $12 \times 1024 \times 1024 \times 1024 = 12884901888$

```
/etc/modprobe.d/zfs.conf
options zfs zfs_arc_max=12884901888 zfs_prefetch_disable=1
```

Miramos la caché. Instalamos sdparm

```
apt-get install sdparm
```

Miramos un disco:

```
sdparm /dev/sda
```

```
/dev/sda: ATA          ST8000DM004-2CX1  0001
Read write error recovery mode page:
AWRE          1  [cha: n, def:  1]
ARRE          0  [cha: n, def:  0]
PER           0  [cha: n, def:  0]
Caching (SBC) mode page:
IC            0  [cha: n, def:  0]
```

```
WCE          1 [cha: y, def: 1]
RCD          0 [cha: n, def: 0]
Control mode page:
TST          0 [cha: n, def: 0]
SWP          0 [cha: n, def: 0]
```

En caching, dentro de WCE tiene que tener def a 1 como lo tiene:

```
WCE          1 [cha: y, def: 1]
```

Reiniciamos y comprobamos los valores:

```
sysctl vm.swappiness
vm.swappiness = 10
```

```
arc_summary | grep "Target size "
      Target size (adaptive):                100.0 %    12.0 GiB
```

Cambio de disco

Primero lo marco offline:

```
zpool offline dades wwn-0x5000c500c520c2f7
```

Vemos como queda el pool:

```
zpool status
```

```
pool: dades
state: DEGRADED
status: One or more devices has been taken offline by the administrator.
        Sufficient replicas exist for the pool to continue functioning in a
        degraded state.
action: Online the device using 'zpool online' or replace the device with
        'zpool replace'.
scan: none requested
config:
```

NAME	STATE	READ	WRITE	CKSUM
dades	DEGRADED	0	0	0
raidz1-0	DEGRADED	0	0	0
wwn-0x5000c500c520c2f7	OFFLINE	0	0	0
wwn-0x5000c500cf855885	ONLINE	0	0	0
wwn-0x5000c500cf875876	ONLINE	0	0	0
wwn-0x5000c500cf8fda13	ONLINE	0	0	0

```
errors: No known data errors
```

Apagamos el servidor y cambiamos el disco, en este caso era el sda. Miramos el nuevo id

```
ls -l /dev/disk/by-id| grep sda
```

```
lrwxrwxrwx 1 root root 9 Dec 12 21:26 ata-ST8000AS0002-1NA17Z_Z840BE4L ->
.././sda
lrwxrwxrwx 1 root root 10 Dec 12 21:26 ata-ST8000AS0002-1NA17Z_Z840BE4L-
part1 -> .././sda1
lrwxrwxrwx 1 root root 9 Dec 12 21:26 wwn-0x5000c50087752765 -> .././sda
lrwxrwxrwx 1 root root 10 Dec 12 21:26 wwn-0x5000c50087752765-part1 ->
.././sda1
```

Al hacer el replace puede que tengamos que forzar si el disco ya pertenecía a un raid, se hace con -f:

```
zpool replace -f dades wwn-0x5000c500c520c2f7 wwn-0x5000c50087752765
```

Miramos el estado. Como había poco contenido dice que tardará 43 segundos, que es lo que ha tardado:

```
zpool status
```

```
pool: dades
state: DEGRADED
status: One or more devices is currently being resilvered. The pool will
continue to function, possibly in a degraded state.
action: Wait for the resilver to complete.
scan: resilver in progress since Sat Dec 12 21:28:46 2020
22.0G scanned at 3.15G/s, 3.08G issued at 450M/s, 22.0G total
782M resilvered, 13.97% done, 0 days 00:00:43 to go
config:
```

NAME	STATE	READ	WRITE	CKSUM	
dades	DEGRADED	0	0	0	
raidz1-0	DEGRADED	0	0	0	
replacing-0	DEGRADED	0	0	0	
wwn-0x5000c500c520c2f7	OFFLINE	0	0	0	
wwn-0x5000c50087752765	ONLINE	0	0	0	(resilvering)
wwn-0x5000c500cf855885	ONLINE	0	0	0	
wwn-0x5000c500cf875876	ONLINE	0	0	0	
wwn-0x5000c500cf8fda13	ONLINE	0	0	0	

```
errors: No known data errors
```

Lo volvemos a mirar y ya está online:

```
zpool status
```

```
pool: dades
state: ONLINE
scan: resilvered 13.2G in 0 days 00:02:04 with 0 errors on Sat Dec 12
```

22:14:11 2020

config:

NAME	STATE	READ	WRITE	CKSUM
dades	ONLINE	0	0	0
raidz1-0	ONLINE	0	0	0
wwn-0x5000c50087752765	ONLINE	0	0	0
wwn-0x5000c500cf855885	ONLINE	0	0	0
wwn-0x5000c500cf875876	ONLINE	0	0	0
wwn-0x5000c500cf8fda13	ONLINE	0	0	0

errors: No known data errors

NFS

Instalamos NFS:

```
apt install nfs-kernel-server
```

Compartimos recurso:

```
zfs set sharenfs=no_subtree_check,no_root_squash,async,rw=@192.168.1.0/24
dades
```

Lo comprobamos:

```
zfs get sharenfs dades
```

NAME	PROPERTY	VALUE
SOURCE		
dades	sharenfs	no_subtree_check,no_root_squash,async,rw=@192.168.1.0/24
local		

Desde otro servidor (tenemos que tener instalado nfs-common)

```
apt-get install nfs-common
```

```
showmount -e 192.168.1.76
Export list for 192.168.1.76:
/dades      192.168.1.0/24
```

Instalación FTP

Lo haremos con docker. Primero instalamos docker:

```
sudo apt-get install \
  apt-transport-https \
  ca-certificates \
  curl \
  gnupg-agent \
  software-properties-common
```

```
curl -fsSL https://download.docker.com/linux/debian/gpg | sudo apt-key add -
```

```
apt-key fingerprint 0EBFCD88
```

```
add-apt-repository \
  "deb [arch=amd64] https://download.docker.com/linux/debian \
  buster \
  stable"
```

```
apt-get update
apt-get install docker-ce docker-ce-cli containerd.io
```

Docker compose:

```
sudo curl -L
"https://github.com/docker/compose/releases/download/1.27.4/docker-compose-$(
uname -s)-$(uname -m)" -o /usr/local/bin/docker-compose
sudo chmod +x /usr/local/bin/docker-compose
```

```
docker run --rm --name ftp_server -d -p 21:21 -p 21000-21010:21000-21010 -v
/dades/jandro:/ftp/jandro -e USERS="user|password" -e
ADDRESS=ftp.lobo99.info delfer/alpine-ftp-server
```

Pruebas de velocidad

Las hacemos con la aplicación fio, la instalamos:

```
apt-get install fio
```

Local

Escritura 260 IOPS

Creamos un fichero de 16gbs:

```
cd /dades
fio --size=16G --name=create --filename=fio_file --bs=1M --nrfiles=1 --
direct=0 --sync=0 --randrepeat=0 --rw=write --refill_buffers --end_fsync=1 -
-iodepth=200 --ioengine=libaio --fallocate=none
```

```

create: (g=0): rw=write, bs=(R) 1024KiB-1024KiB, (W) 1024KiB-1024KiB, (T)
1024KiB-1024KiB, ioengine=libaio, iodepth=200
fio-3.21
Starting 1 process
create: Laying out IO file (1 file / 16384MiB)
Jobs: 1 (f=1): [F(1)][100.0%][eta 00m:00s]
create: (groupid=0, jobs=1): err= 0: pid=198903: Sat Dec 12 22:44:59 2020
  write: IOPS=260, BW=261MiB/s (273MB/s)(16.0GiB/62844msec); 0 zone resets
    slat (usec): min=417, max=24224, avg=3121.52, stdev=1499.31
    clat (usec): min=10, max=1442.9k, avg=698492.10, stdev=205388.14
    lat (msec): min=3, max=1448, avg=701.62, stdev=206.26
    clat percentiles (msec):
      | 1.00th=[ 201], 5.00th=[ 243], 10.00th=[ 451], 20.00th=[ 600],
      | 30.00th=[ 651], 40.00th=[ 684], 50.00th=[ 701], 60.00th=[ 726],
      | 70.00th=[ 760], 80.00th=[ 827], 90.00th=[ 936], 95.00th=[ 1011],
      | 99.00th=[ 1301], 99.50th=[ 1435], 99.90th=[ 1435], 99.95th=[ 1435],
      | 99.99th=[ 1435]
    bw ( KiB/s): min=114459, max=991232, per=100.00%, avg=288007.45,
stdev=101346.60, samples=115
    iops        : min= 111, max= 968, avg=281.17, stdev=98.98, samples=115
    lat (usec)   : 20=0.01%
    lat (msec)   : 4=0.01%, 10=0.01%, 20=0.02%, 50=0.04%, 100=0.07%
    lat (msec)   : 250=5.21%, 500=7.21%, 750=54.96%, 1000=26.78%, 2000=5.69%
    cpu          : usr=9.03%, sys=16.29%, ctx=168107, majf=0, minf=15
    IO depths    : 1=0.1%, 2=0.1%, 4=0.1%, 8=0.1%, 16=0.1%, 32=0.2%,
>=64=99.6%
    submit       : 0=0.0%, 4=100.0%, 8=0.0%, 16=0.0%, 32=0.0%, 64=0.0%,
>=64=0.0%
    complete     : 0=0.0%, 4=100.0%, 8=0.0%, 16=0.0%, 32=0.0%, 64=0.0%,
>=64=0.1%
    issued rwts: total=0,16384,0,0 short=0,0,0,0 dropped=0,0,0,0
    latency     : target=0, window=0, percentile=100.00%, depth=200

Run status group 0 (all jobs):
  WRITE: bw=261MiB/s (273MB/s), 261MiB/s-261MiB/s (273MB/s-273MB/s),
io=16.0GiB (17.2GB), run=62844-62844msec

```

Nos quedamos con la línea 7 que dice que son 260 IOPS

```
write: IOPS=260, BW=261MiB/s (273MB/s)(16.0GiB/62844msec); 0 zone resets
```

Provem 4 processos de lectura y escriptura:

Lectura 418 IOPS:

```

fio --time_based --name="$(hostname).randread" --size=16G --runtime=30 --
filename=fio_file --ioengine=libaio --randrepeat=0 --iodepth=128 --direct=1
--invalidate=1 --verify=0 --verify_fatal=0 --numjobs=4 --rw=randread --
blocksize=8k --group_reporting

```

```
nas.randread: (g=0): rw=randread, bs=(R) 8192B-8192B, (W) 8192B-8192B, (T)
```



```

8192B-8192B, ioengine=libaio, iodepth=128
...
fio-3.21
Starting 4 processes
Jobs: 4 (f=4): [r(4)][100.0%][r=3496KiB/s][r=437 IOPS][eta 00m:00s]
nas.randread: (groupid=0, jobs=4): err= 0: pid=269572: Sat Dec 12 22:46:28
2020
    read: IOPS=418, BW=3346KiB/s (3426kB/s)(98.1MiB/30026msec)
        slat (usec): min=8, max=391677, avg=9551.50, stdev=17207.97
        clat (usec): min=10, max=2366.9k, avg=1164430.92, stdev=264881.17
            lat (msec): min=7, max=2384, avg=1173.98, stdev=266.34
        clat percentiles (msec):
            | 1.00th=[ 300], 5.00th=[ 844], 10.00th=[ 944], 20.00th=[ 1020],
            | 30.00th=[ 1070], 40.00th=[ 1099], 50.00th=[ 1150], 60.00th=[ 1183],
            | 70.00th=[ 1234], 80.00th=[ 1284], 90.00th=[ 1418], 95.00th=[ 1687],
            | 99.00th=[ 2089], 99.50th=[ 2140], 99.90th=[ 2165], 99.95th=[ 2198],
            | 99.99th=[ 2333]
        bw ( KiB/s): min= 928, max= 5600, per=100.00%, avg=3437.96,
stdev=220.98, samples=220
        iops      : min= 116, max= 700, avg=429.75, stdev=27.62, samples=220
        lat (usec) : 20=0.03%
        lat (msec) : 10=0.02%, 20=0.01%, 50=0.04%, 100=0.25%, 250=0.56%
        lat (msec) : 500=1.02%, 750=1.23%, 1000=13.63%, 2000=81.79%,
>=2000=1.42%
        cpu       : usr=0.08%, sys=0.86%, ctx=4704, majf=0, minf=1073
        IO depths : 1=0.1%, 2=0.1%, 4=0.1%, 8=0.3%, 16=0.5%, 32=1.0%,
>=64=98.0%
        submit    : 0=0.0%, 4=100.0%, 8=0.0%, 16=0.0%, 32=0.0%, 64=0.0%,
>=64=0.0%
        complete  : 0=0.0%, 4=100.0%, 8=0.0%, 16=0.0%, 32=0.0%, 64=0.0%,
>=64=0.1%
        issued rwts: total=12558,0,0,0 short=0,0,0,0 dropped=0,0,0,0
        latency   : target=0, window=0, percentile=100.00%, depth=128

Run status group 0 (all jobs):
    READ: bw=3346KiB/s (3426kB/s), 3346KiB/s-3346KiB/s (3426kB/s-3426kB/s),
io=98.1MiB (103MB), run=30026-30026msec

```

Esriptura (409 IOPS):

```

fio --time_based --name="$(hostname).randwrite" --size=16G --runtime=30 --
filename=fio_file --ioengine=libaio --randrepeat=0 --iodepth=128 --direct=1
--invalidate=1 --verify=0 --verify_fatal=0 --numjobs=4 --rw=randwrite --
blocksize=8k --group_reporting

```

```

nas.randwrite: (g=0): rw=randwrite, bs=(R) 8192B-8192B, (W) 8192B-8192B, (T)
8192B-8192B, ioengine=libaio, iodepth=128

```

```

...
fio-3.21
Starting 4 processes
Jobs: 4 (f=4): [w(4)][100.0%][w=1729KiB/s][w=216 IOPS][eta 00m:00s]

```

```

nas.randwrite: (groupid=0, jobs=4): err= 0: pid=269629: Sat Dec 12 22:47:54
2020
  write: IOPS=409, BW=3277KiB/s (3355kB/s)(96.1MiB/30025msec); 0 zone resets
    slat (usec): min=12, max=859027, avg=9751.32, stdev=28939.98
    clat (usec): min=10, max=2770.9k, avg=1202311.14, stdev=522285.60
    lat (msec): min=10, max=2782, avg=1212.06, stdev=524.35
    clat percentiles (msec):
      | 1.00th=[ 215], 5.00th=[ 642], 10.00th=[ 735], 20.00th=[ 818],
      | 30.00th=[ 877], 40.00th=[ 927], 50.00th=[ 995], 60.00th=[ 1099],
      | 70.00th=[ 1368], 80.00th=[ 1703], 90.00th=[ 2089], 95.00th=[ 2232],
      | 99.00th=[ 2500], 99.50th=[ 2601], 99.90th=[ 2668], 99.95th=[ 2702],
      | 99.99th=[ 2769]
    bw ( KiB/s): min= 432, max= 7168, per=100.00%, avg=3419.70,
stdev=401.51, samples=216
    iops       : min= 54, max= 896, avg=427.37, stdev=50.21, samples=216
    lat (usec)  : 20=0.03%
    lat (msec)  : 20=0.04%, 50=0.14%, 100=0.25%, 250=0.70%, 500=0.97%
    lat (msec)  : 750=9.44%, 1000=39.76%, 2000=36.13%, >=2000=12.54%
    cpu         : usr=0.12%, sys=0.89%, ctx=4456, majf=0, minf=61
    IO depths   : 1=0.1%, 2=0.1%, 4=0.1%, 8=0.3%, 16=0.5%, 32=1.0%,
>=64=98.0%
    submit      : 0=0.0%, 4=100.0%, 8=0.0%, 16=0.0%, 32=0.0%, 64=0.0%,
>=64=0.0%
    complete    : 0=0.0%, 4=100.0%, 8=0.0%, 16=0.0%, 32=0.0%, 64=0.0%,
>=64=0.1%
    issued rwts: total=0,12298,0,0 short=0,0,0,0 dropped=0,0,0,0
    latency    : target=0, window=0, percentile=100.00%, depth=128

Run status group 0 (all jobs):
  WRITE: bw=3277KiB/s (3355kB/s), 3277KiB/s-3277KiB/s (3355kB/s-3355kB/s),
io=96.1MiB (101MB), run=30025-30025msec

```

Remoto

SCP:

```

5.6G 105.0MB/s 00:54
16GB 104.2MB/s 02:37

```

NFS: Montamos el recurso:

```
mount 192.168.1.76:/dades /mnt/nfs/
```

Copiamos los ficheros:

```
time cp fitxer /mnt/nfs
```

```
5.6G real 0m53.395s
```

16GB real 2m33.965s

FTP:

Desde una conexión 100Mbps/100Mbps

1.6Gb (1718558720 bytes) sent in 132.96 secs (12.3269 MB/s)

5.5Gb (5864260625 bytes) sent in 453.54 secs (12.3311 MB/s)

Desde una conexión 552Mbps/573Mbps

1Gb (1073741824 bytes) sent in 23.7 secs (45279.08 Kbytes/sec)

5Gb (5368709120 bytes) sent in 110 secs (48596.40 Kbytes/sec)

From:

<http://wiki.legido.com/> - **Legido Wiki**

Permanent link:

<http://wiki.legido.com/doku.php?id=informatica:microservers:raid>

Last update: **2020/12/16 00:05**

