# bcachefs

Danny Robson

# Warning

bcachefs is experimental

Know your risk tolerance

This is simply my understanding as a user

# My Needs

- A single, high end, personal machine.
- Photography, development, databases.
- I’m currently using 14TB of 22TB.

# Previous system

# Previous system

1. XFS

# Previous system

1. XFS
2. lvm

# Previous system

1. XFS
2. lvm
3. dmcrypt

# Previous system

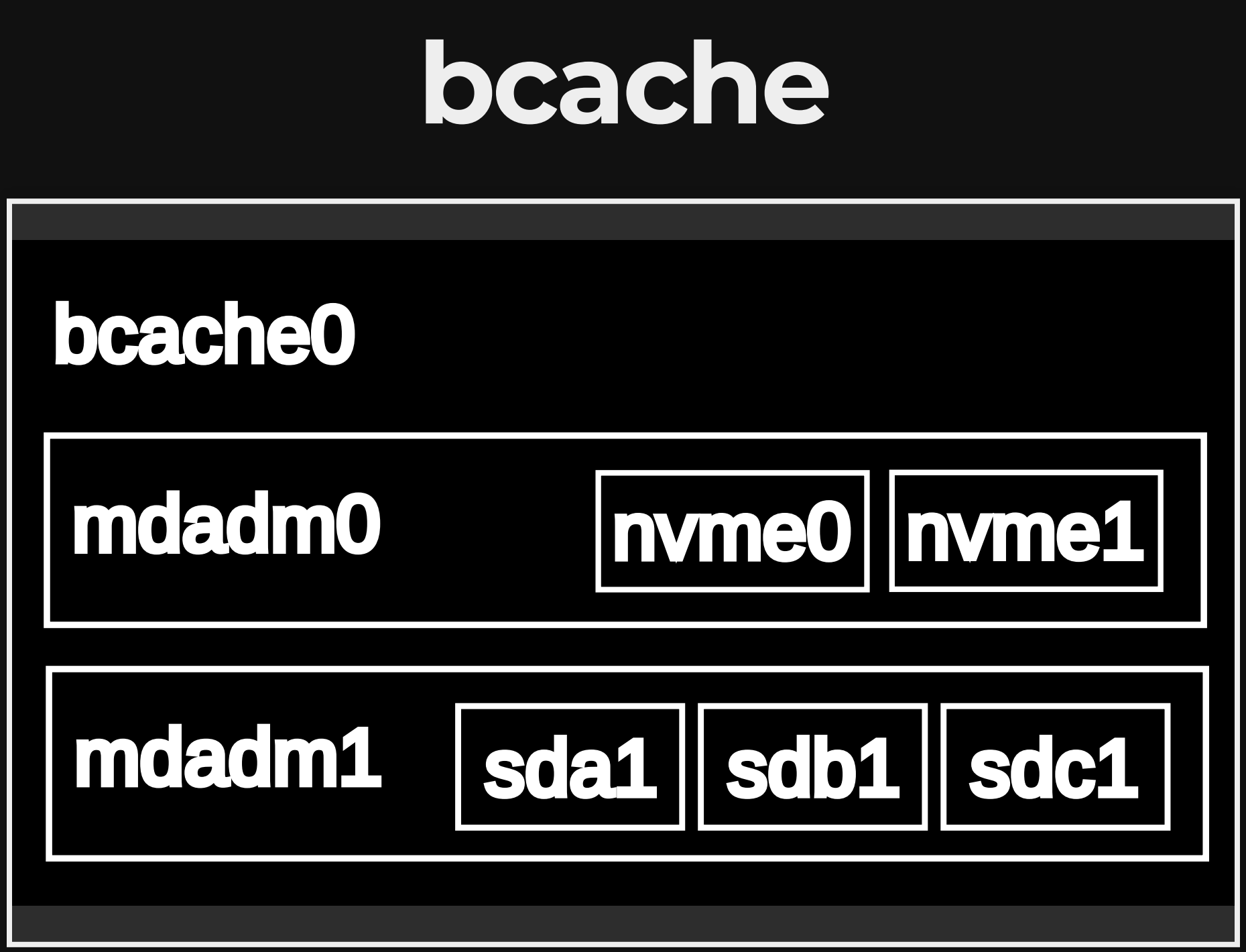
1. XFS
2. lvm
3. dmcrypt
4. mdadm

# Previous system

1. XFS
2. lvm
3. dmcrypt
4. mdadm
5. bcache

# bcache

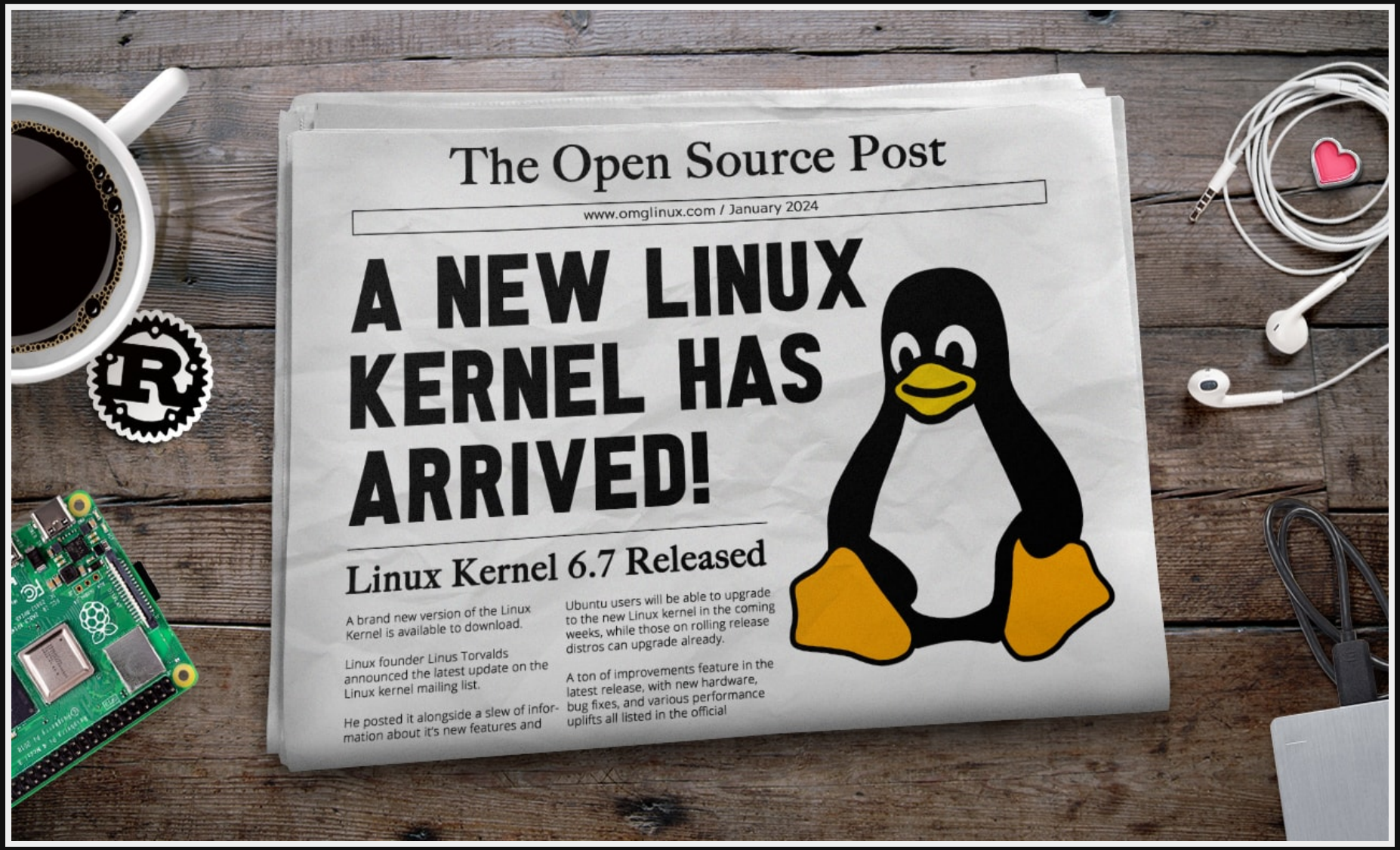
Use a faster block device to cache reads/writes to a slower block device


bcache
bcache0
mdadm0
nvme0
nvme1
mdadm1
sda1
sdb1
sdc1

# Issues

- Complex
- Difficult
- Inflexible
- Potentially poor performance

# Linux 6.7


The Open Source Post
www.omglinux.com / January 2024
A NEW LINUX KERNEL HAS ARRIVED!
Linux Kernel 6.7 Released
A brand new version of the Linux Kernel is available to download.
Linux founder Linus Torvalds announced the latest update on the Linux kernel mailing list.
He posted it alongside a slew of information about it's new features and
Ubuntu users will be able to upgrade to the new Linux kernel in the coming weeks, while those on rolling release distros can upgrade already.
A ton of improvements feature in the latest release, with new hardware, bug fixes, and various performance uplifts all listed in the official

# bcachefs

A modern copy-on-write, multi-device, filesystem

- caching
- compression
- redundancy
- snapshots
- encryption
- quotas

# bcachefs format

```
bcachefs format
    --label=ssd.ssd1 /dev/nvme0n1p1
    --label=hdd.hdd1 /dev/sda1
```

# bcachefs format

```
bcachefs format
    --label=ssd.ssd1 /dev/nvme0n1p1
    --label=hdd.hdd1 /dev/sda1
    --foreground_target=ssd
    --promote_target=ssd
    --background_target=hdd
```

# bcachefs format

```
bcachefs format
    --label=hdd.hdd1 /dev/sda1
    --label=hdd.hdd2 /dev/sda2
    --label=hdd.hdd3 /dev/sda3
    --replicas=2
```

# bcachefs format

```
bcachefs format
    --compression=lz4
    --encrypted
    --replicas=2
    --label=hdd.hdd1 /dev/sdc
    --label=hdd.hdd2 /dev/sdd
    --label=hdd.hdd3 /dev/sde
    --label=hdd.hdd4 /dev/sdf
    --discard
    --label=ssd.ssd1 /dev/sda
    --label=ssd.ssd2 /dev/sdb
    --foreground_target=ssd
    --promote_target=ssd
    --background_target=hdd
```

# mount

```
mount -t bcachefs /dev/sda1:/dev/sdb1:/dev/sdc1 /mnt
```

or

```
bcachefs mount UUID=aaaaaaaa-bbbb-cccc-dddd-eeeeeeeeeeee
```

# mount

```
#!/bin/bash

# Try to mount a bcachefs filesystem using only one known device
#   usage: mount.bcachefs ${DEVICE} ${MOUNTPOUNT}
#
# THIS WILL EXPLODE ONE DAY. DO NOT USE UNLESS YOU KNOW HOW TO BO

device="$1"; shift
mountpoint="$1"; shift

uuid=$(bcachefs show-super "${device}" | grep "External UUID")
uuid="${uuid##* }"

bcachefs mount "UUID=${uuid}" "${mountpoint}" $@
exit $?
```

# mount

Warning

unclean mount times for large arrays can be **long**.

# Adding disks

```
bcachefs device add
    --label hdd.hdd9
    /dev/sdi

bcachefs data rereplicate ${MOUNTPOINT}
```

# Removing disks

```
echo 2 > /sys/fs/bcachefs/${UUID}/options/metadata_replicas
bcachefs data rereplicate ${MOUNTPOINT}

bcachefs device set-state ${DEVICE} readonly
bcachefs device evacuate ${DEVICE}
```

# snapshots

```
bcachefs subvolume \
    create /home/danny/src

bcachefs subvolume snapshot \
    /home/danny/src \
    /var/lib/backup/src_$(date --iso-8601)
```

Highly efficient, theoretically supports thousands to millions.

# compression

```
echo 'none' > /sys/fs/bcachefs/${UUID}/options/compression
echo 'zstd' > /sys/fs/bcachefs/${UUID}/options/background_compres
```

# options: scope

- format
- mount
- runtime
- inode

# options: tools

- `bcachefs`, `mount`
- `sysfs`
- `xattr`

# options

```
bcachefs setattr
    --data_replicas=1
    /home/danny/.local/share/Steam/
```

# options

```
bcachefs setattr
    --compression lz4
    --background_compression zstd:15
    /home/danny/src
```

# options

```
setfattr
    -n bcachefs_effective.data_replicas
    -v 1
    /home/danny/inconsequential_file
```

# Downsides

- Will never compete with XFS on raw performance, but:
  - XFS can’t snapshot, detect errors, replicate, etc...
  - Tiered storage is quite effective

# Downsides

Noticably higher memory use than previously

```
echo 2 > /proc/sys/vm/drop_caches
```

# Downsides

Self healing, but lacks scrub

# Upcoming

- Bug fixes, performance, integration
- Quota rewrite for 6.9
- Erasure coding
- SMR and ZNS support

# More information

**homepage**

https://bcachefs.org

**readthedocs**

https://bcachefs-docs.readthedocs.io/en/latest/index.html

**arch**

https://wiki.archlinux.org/title/Bcachefs

**reddit**

https://www.reddit.com/r/bcachefs