The Linux Storage Stack Diagram (Linux Kernel 6.9)

(version 6.9 2024-05-12, outlines the Linux storage stack as of Linux Kernel version 6.9)

Applications (processes)

mmap (anonym. pages)

read (2) write (2) open (2) stat (2) chmod (2) ... (2)

malloc

NVMe over TCP NVMe over RDMA NVMe over Fibre Channel

Fibre Channel over Ethernet ISCSI Fibre Channel FireWire USB Virtual Host SCSI RDMA

LIO

tcm_loop tcm_fc iscsi_target_mod tcm_qla2xxx sbp_target tcm_usb_gadget tcm_vhos ib_srpt

target_core_mod

target_core_file

target_core_iblock

target_core_user

target_core_rd

target_core_pscsi

memory userspace

VFS

loop

Direct I/O (O_DIRECT)

Block-based FS
ext4 btrfs xfs f2fs
gfs2 ocfs2 ntfs3
bcachefs iso9660
erofs squashfs ...

Stackable FS
ecryptfs overlayfs

Network FS
smbfs ceph
nfs coda
scsilayout
...
fuse

fscache

userspace (e.g. sshfs)

network

Pseudo FS
proc sysfs
pipefs usbfs
futexfs ...

Special purpose FS
tmpfs
ramfs
devtmpfs

Raw flash
ubifs jffs2

UBI MTD

Page cache

nvmet
nvme-tcp nvme-rdma nvme-fc nvmet-loop
nvmet

struct bio
- operation & flags
- target device
- bvec list
- bvec_iter (sector, size, index)

BIOs (block I/Os)

(optional) Devices on top of "normal" block devices

device mapper (dm)

Production use
dm-cache dm-clone dm-crypt dm-ebs dm-era dm-integrity
dm-linear dm-mirror dm-raid dm-stripe dm-switch dm-thin
dm-unstripe dm-verity dm-vdo dm-writecache dm-zoned

Debugging use
dm-delay dm-dust
dm-flakey
dm-zero

drbd
md (RAID)
bcache

stackable

Block Layer

blk_mq

I/O scheduler
kyber bfq mq_deadline

Request based drivers

BIO based drivers

read-only access

/dev/loop*

Request-based dm targets
dm-multipath

/dev/ublkb* /dev/rbd* /dev/vd* /dev/mmcblk*p* /dev/mtdblock*ro UBI MTD

NVMe Layer
/dev/nvme*n* /dev/ng*

/dev/nullb* /dev/n64cart

loop

/dev/zram* /dev/nbd* /dev/dasd* /dev/rssd* /dev/ubiblock*

ublkb rbd virtio_blk mmc mtdblock ubi

NVMe transport drivers
nvme-tcp
nvme (pci) nvme-rdma
nvme-fc nvme-loop

/dev/rd* /dev/pmem

null_blk n64cart

brd pmem

userspace

zram nbd dasd mtip32xx

sysfs (transport attrib.)

network
memory

MTD (Memory Technology Device)

infiband/RoCE/iWarp
network
fibre Channel HBA driver

memory

SCSI mid Layer
SCSI upper level drivers
/dev/sd* /dev/sr*
/dev/st* ...

Transport Classes
scsi_transport_fc ..._sas ...

SCSI low level drivers
tcm_loop aacraid vmware_pvscsi
libata smartpqi megaraid_sas ... virtio_scsi
ahci ... mpt3_sas hyperv_storage
ata_piix mpi3mr iscsi_tcp

network

Physical Devices
ATA controller LSI HBA LSI RAID Hyper-V virtio_pci SPI-NAND PCIe NVMe
Adaptec RAID Broadcom MPI3.0 HBA/RAID IBM DASD SPI-NOR N64 cartridge
Adaptec SmartHBA/RAID VMware SD-/MMC-Card Micron PCIe Raw NAND Persistent memory