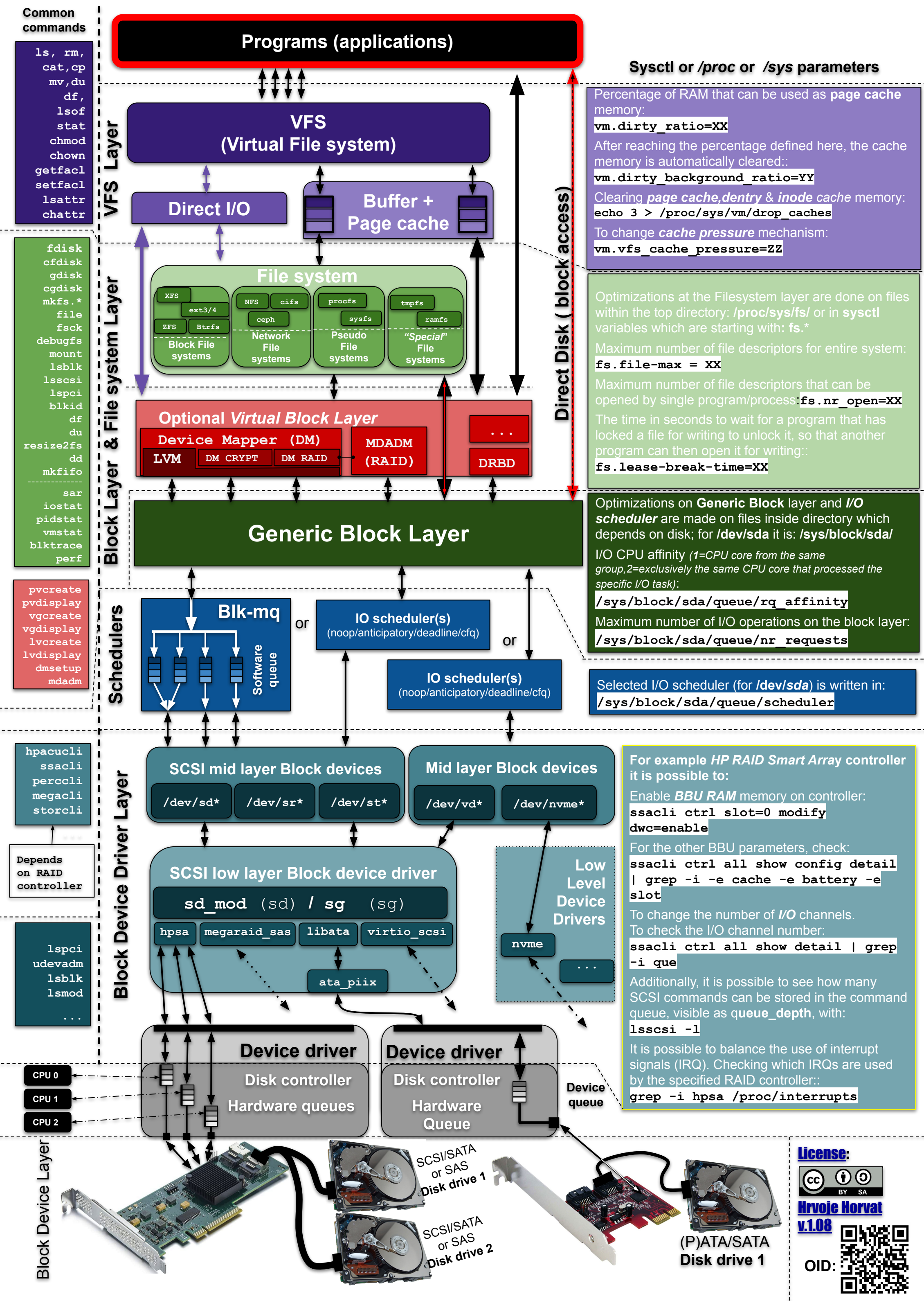
Common commands
ls, rm, cat,cp mv,du df, lsof stat chmod chown getfacl setfacl lsattr chattr
fdisk cfdisk gdisk cgdisk mkfs.* file fsck debugfs mount lsblk lsscsi lspci blkid df du resize2fs dd mkfifo
sar iostat pidstat vmstat blktrace perf
pvcreate pvdisplay vgcreate vgdisplay lvcreate lvdisplay dmsetup mdadm
hpacucli ssacli perccli megacli storcli
Depends on RAID controller
lspci udevadm lsblk lsmod ...
Programs (applications)
VFS Layer
VFS (Virtual File system)
Direct I/O
Buffer + Page cache
File system & File system Layer
File system
XFS
ext3/4
ZFS
Btrfs
Block File systems
NFS
cifs
ceph
Network File systems
procfs
sysfs
Pseudo File systems
tmpfs
ramfs
"Special" File systems
Optional Virtual Block Layer
Device Mapper (DM)
LVM
DM CRYPT
DM RAID
MDADM (RAID)
...
DRBD
Direct Disk (block access)
Block Layer
Generic Block Layer
Schedulers
Blk-mq
Software queue
or
IO scheduler(s)
(noop/anticipatory/deadline/cfq)
or
IO scheduler(s)
(noop/anticipatory/deadline/cfq)
Block Device Driver Layer
SCSI mid layer Block devices
/dev/sd*
/dev/sr*
/dev/st*
Mid layer Block devices
/dev/vd*
/dev/nvme*
SCSI low layer Block device driver
sd_mod (sd) / sg (sg)
hpsa
megaraid_sas
libata
virtio_scsi
ata_piix
Low Level Device Drivers
nvme
...
Device driver
Disk controller
Hardware queues
Device driver
Disk controller
Hardware Queue
Device queue
CPU 0
CPU 1
CPU 2
Block Device Layer
SCSI/SATA or SAS Disk drive 1
SCSI/SATA or SAS Disk drive 2
(P)ATA/SATA Disk drive 1
Sysctl or /proc or /sys parameters
Percentage of RAM that can be used as page cache memory:
vm.dirty_ratio=XX
After reaching the percentage defined here, the cache memory is automatically cleared::
vm.dirty_background_ratio=YY
Clearing page cache,dentry & inode cache memory:
echo 3 > /proc/sys/vm/drop_caches
To change cache pressure mechanism:
vm.vfs_cache_pressure=ZZ
Optimizations at the Filesystem layer are done on files within the top directory: /proc/sys/fs/ or in sysctl variables which are starting with: fs.*
Maximum number of file descriptors for entire system:
fs.file-max = XX
Maximum number of file descriptors that can be opened by single program/process:fs.nr_open=XX
The time in seconds to wait for a program that has locked a file for writing to unlock it, so that another program can then open it for writing::
fs.lease-break-time=XX
Optimizations on Generic Block layer and I/O scheduler are made on files inside directory which depends on disk; for /dev/sda it is: /sys/block/sda/
I/O CPU affinity (1=CPU core from the same group,2=exclusively the same CPU core that processed the specific I/O task):
/sys/block/sda/queue/rq_affinity
Maximum number of I/O operations on the block layer:
/sys/block/sda/queue/nr_requests
Selected I/O scheduler (for /dev/sda) is written in:
/sys/block/sda/queue/scheduler
For example HP RAID Smart Array controller it is possible to:
Enable BBU RAM memory on controller:
ssacli ctrl slot=0 modify dwc=enable
For the other BBU parameters, check:
ssacli ctrl all show config detail | grep -i -e cache -e battery -e slot
To change the number of I/O channels.
To check the I/O channel number:
ssacli ctrl all show detail | grep -i que
Additionally, it is possible to see how many SCSI commands can be stored in the command queue, visible as queue_depth, with:
lsscsi -l
It is possible to balance the use of interrupt signals (IRQ). Checking which IRQs are used by the specified RAID controller::
grep -i hpsa /proc/interrupts
License:
Hrvoje Horvat
v.1.08
OID: