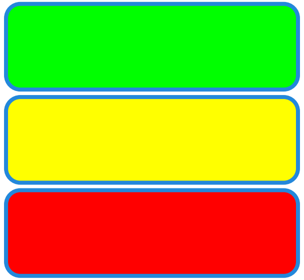
IT Lightning Talk

RAID and filesystem alignment

RAID?

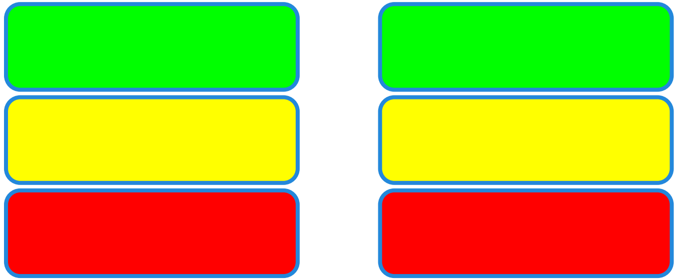
Redundant array of inexpensive disks

Simple: divide the disk in"chunks".

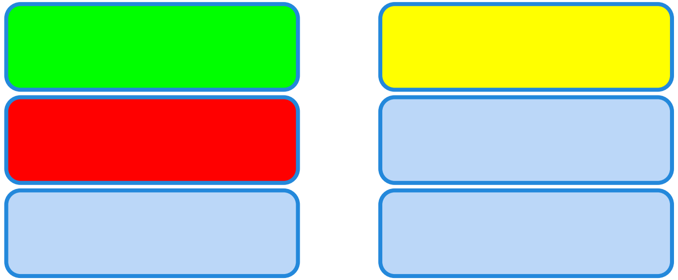


Play with chunks

Mirror



Strip



Checksum (I don't care, I need performance)

Expected performance (2 disks)?

Mirror:

- 1 stream @ W
- 1-2 streams @ R

Strip:

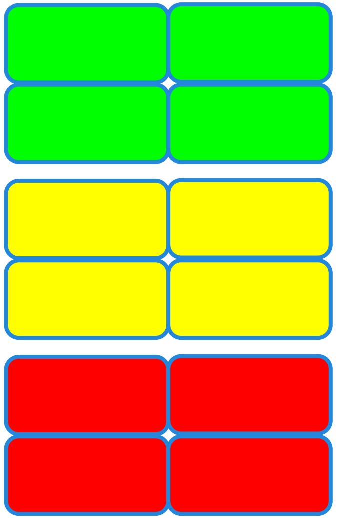
- 1 stream @ 2W
- 1 stream @ 2R

W: nominal write speed
R: nominal read speed

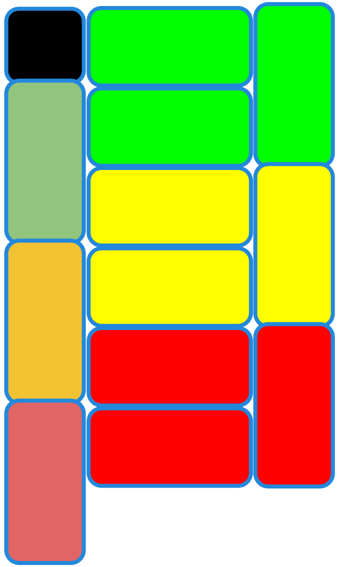
Expected performance

RAID is the layer below your FS

A filesystem is divided in blocks



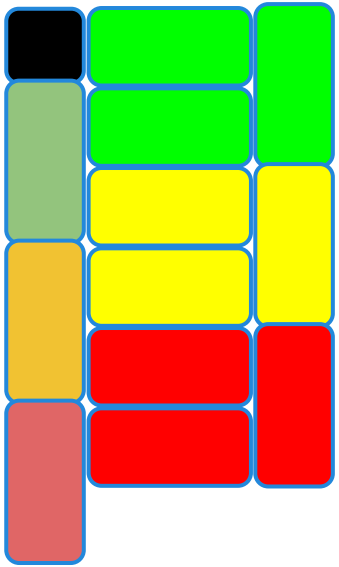
Non aligned file system



write a block on 2 chunks:

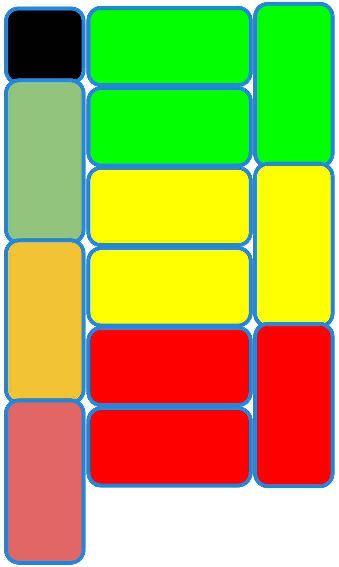
- read the 2 chunks
- move the head back
- rewrite the 2 chunks

Non aligned file system



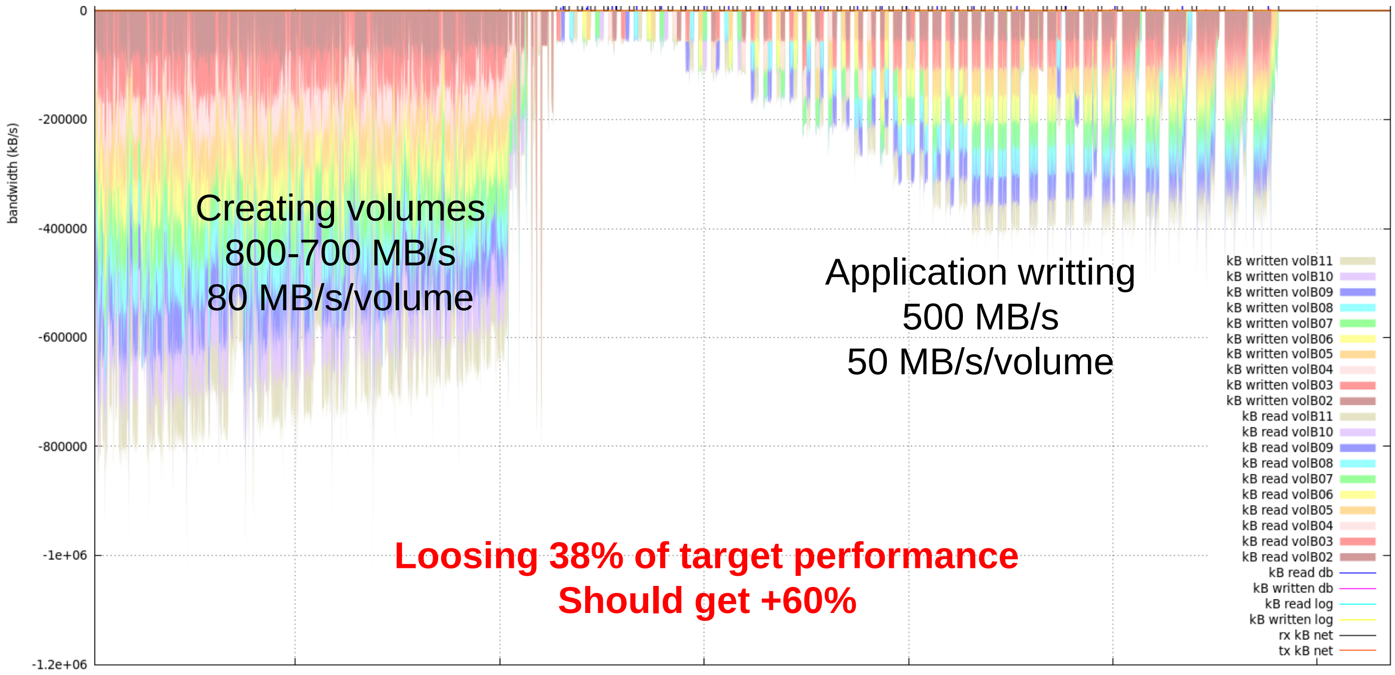
4K block 256K chunk
(80MB/s write 130MB/s read
seek 4ms)
4K write time = 512K read time
+ seek time + 512K write = 3.8
+ 4 + 6.3 = 14.1 ms
This is 0.28MB/s...

Real life case



10 X (2 disks in RAID1E)
expected write speed:
10 X 80MB/s
expected read speed:
10 X 260MB/s

Measurements



Creating volumes
800-700 MB/s
80 MB/s/volume

Application writting
500 MB/s
50 MB/s/volume

Loosing 38% of target performance
Should get +60%

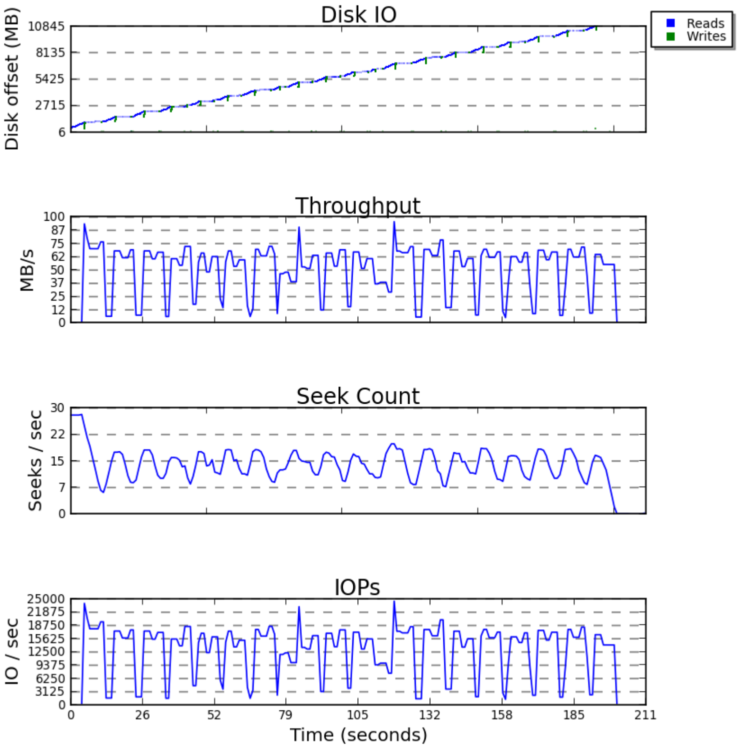
Seekwatcher

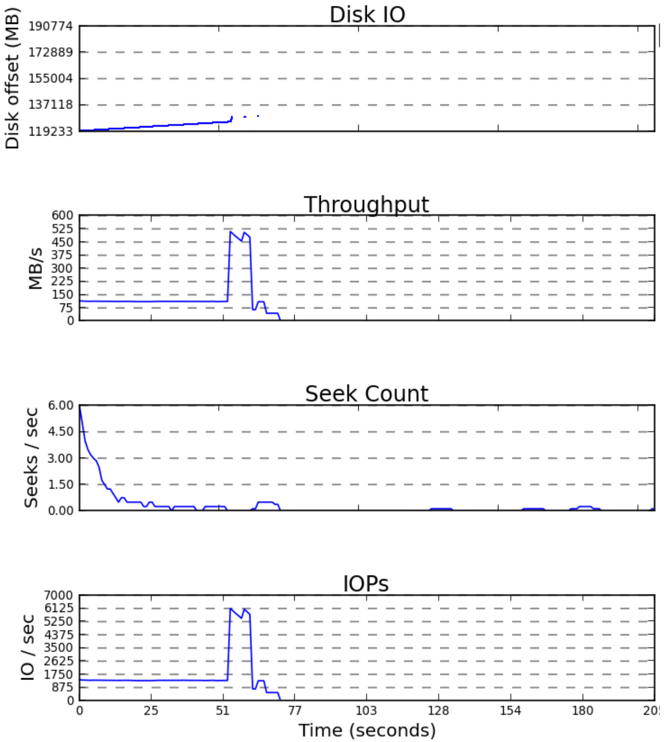
Collect traces with blktrace on the server

See those on your laptop !

Pictures or better : videos !! (aka disk pr0n)

Seekwatcher writes





Pr0n

Missing video (thank you Powerpoint!!!)

If WMP can read it you may not be able to insert it!!!

See disk pr0n on my social blog

Advices

Benchmark one disk => target performance

Create RAID and FS

Benchmark application IOs

Compare with the target performance

Prefer automatically aligned FS (XFS,windows FS) or use strides and stripe-width (ext4)

Advices

Align partitions on disk sectors (or forget about those)

Align paritions, chunks, FS on SSD erase block size

RAID managed by FS (ZFS, BTRFS?)

To go further

Pointers and more on my blog on social

Like

+1