Benchmarks 29 May 2013 (c) napp-it.org

Hardware: SM X9 SRL-F, Xeon E5-2620 @ 2.00GHz, 65 GB RAM, 6 x IBM 1015 IT (Chenbro 50bay) OS: napp-it appliance v. 0.9c1, OmniOS stable (May 2013)

Disks:

- 5 Seagate SAS ST3146855SS, 146 GB, 15k/rpm,
- 1 Intel 320, 300 GB SSD (MLC),
- 1 ATP SATA II SSD 16 GB (SLC)
- 1 Winkom ML-X8480, 480 GB MLC
- 1 ZeusRAM 8GB SAS (DRAM)

Intension of these benchmarks:

- verify some basic dependencies
- only a overview, no interest in absolute values

- quick tests with small files, larger files are more accurat but not too different

What I read from the benchmarks

Test 1: Sequential performance vs number of vdevs/disks via dd

- Sequential values scales with number of vdevs/disks (about 100-130 MB/s per disk)

- even a single disk is fast enough for 1 GB network
- a fast SSD is as good or better than 4 enterprise 15k rpm SAS disks

OPS/s (fileserver benchmark)

- OPS/s scales with number of vdevs

- a fast SSD is as good or better than 4 enterprise 15k rpm SAS disks

OPS/s (webserver benchmark) - similar values with number of disks or SSD

Test 2: iSCSI vs SMB (sync disabled)

- iSCSI is similar to SMB regarding writes
- iSCSI is more than twice as fast compared to SMB regarding reads (needs some more tests)
- a fast SSD is as good or better than 4 enterprise 15k rpm SAS disks

Test 3: Async vs Sync Write

To check if a SSD is a good ZIL, set sync to always, create a volumebased iSCSI Target, run a Crystalmarbench and check 4k QD32 value

(With 32 concurrent small writes, this is the benchvalues that is of interest for a ZIL)

- Sync write perfomance is only 10-20% of async without dedicated ZIL !!!
- A ZIL build from a 3 years old enterprise class SLC SSD is mostly slower than without ZIL
- (this pool is build from fast disks, but a dedicated ZIL needs to be really fast or its useless)
- A Intel 320 SSD (quite often used because of the included supercap) is a quite good ZIL, You get up to 60% of the async values (at least with a larger 320, i used a 300 GB SSD)
- Only a DRAM based ZeusRAM is capable to deliver similar values like async write
- Some SSDs like newest SLC ones or a Intel S3700 are very good and much cheaper

Filebench: Randomwrite

Sync write values are quite bad, even with a ZeusRAM.

I suppose this is due the small 8 GB ZeusRAM (a ZIL needs to hold about 10s of writes, not ideal for a local benchmark) but a single 8 GB ZeusRAM should be ok for a single 10 GbE link (about 1 GB/s x 10s = less than 10 GB needed Zilsize).

Test 4: Async vs Sync on a SSD only pool

- sync write performance is up to 40% of the async performance
- a slow SSD as extra ZIL, even a SLC one is a very bad idea (although may increase durability of MLC SSD's)
- Even with a SSD only pool, a ZeusRAM is a good idea. (Up to 70% or asny values and increase durability of MLC SSD's)
- ZFS seems quite well when a Pool is nearly full (at least with benchmarks from small files. Performance with large files
- like ESXi VM's is a different thing from my experience, so try to stay below 70% fillrate)

The benchmarks

Test1: Use the Seagate in a Raid-0, test performance vs number of vdevs, sync: default (=disabled)

Remote tests are done from Windos via 10 GbE either via CIFS or iSCSI

Filebench, all Seagate SAS Disks in Raid-0, i do not check absolute values but differences plus dd write with 128GB, 2 MB blocks, writeonly, NAS-Tester http://www.808.dk/?code-csharp-nas-performance. Because of the large RAM-Cache, i check mainly write values, readvalues are mostly similar without cache.

Stage 1.1: (fileserver.f, 30s), Raid-0 (one basic 15k disk disk per vdev)

Disks 1 2 3	OPS 104987 ops 399095 ops 233414 ops,	OPS/s 3499.449 ops/s 3302.761 ops/s 7779.562 ops/s	(707/1415 r/w)	185.9mb/s	Latency 1634us cpu/op 428us cpu/op 1123us cpu/op	49.4ms latency 13.0ms latency 22.8ms latency	dd write 111 MB/s 229 MB/s 378 MB/s	NAS tester write 400 MB (Windows SMB) 143 MB/s 108 MB/s 117 MB/s
4 Stage 1	397243 ops .2: (webserver.f, 3	13238.229 ops/s 60s), Raid-0 (one b	(1203/2407 r/w) asic 15k disk per v		542us cpu/op	13.1ms latency	475 MB/s	176 MB/s

Disks	OPS	OPS/s	RW	Latency	
1	13605195 ops	453490.7 ops/s	(146287/14631 r/w) 2405.3mb/s	56us cpu/op	0.2ms latency
2	13658179 ops	455255.654 ops/s	s (146856/14688 r/w) 2414.6mb/s,	56us cpu/op	0.2ms latency
3	13595568 ops,	453166.862 ops/s	s (146182/14620 r/w) 2404.3mb/s,	56us cpu/op,	0.3ms latency
4	13553535 ops	451769.074 ops/s	s (145731/14575 r/w) 2396.3mb/s,	56us cpu/op,	0.2ms latency

Stage 2	.1: Compare to a	single SSD (480 G	GB), (fileserver.f)					
Disks	OPS	OPS/s	RW		Latency		dd write	NAS tester write 400 MB (Windows SMB)
1	633773 ops,	21123.501 ops/s	, (1920/3841 r/w),	509.5mb/s,	428us cpu/op,	8.1ms latency	470 MB/s	141 MB/s

Stage 2.2: Compare to a single SSD (480 GB), (webserver.f)

1 13649111 ops, 454954.630 ops/s (146759/14678 r/w), 2413.5mb/s, 56us cpu/op, 0.3ms latency

Test 2. iSCSI vs SMB, disks vs SSD, sync disabled, volume based LU

iSCSI Benchmark: Windows 7-64, 8GB RAM, 10 GbE via iSCSI Target (volumebased, 50 GB, 64k blocksize, thin prof., writeback cache enabled, NTFS formatted Pool from single Seagate disk via iSCSI Pool from 2 disks, 2 vdev=Raid-0 Pool from 3 disks, 3 vdevs in Raid 0 Pool from 4 disks, 4 vdevs in Raid 0

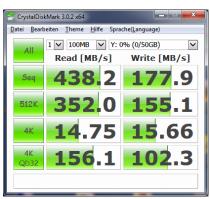
🚔 Untitled - ATTO Disk Benchmark _ 🗆 🕳 🗙 🚔 Untitled - ATTO Disk Benchmark _ 🗆 🗙 🚔 Untitled - ATTO Disk Benchmark _ 🗆 🗙 🚰 Untitled - ATTO Disk Benchmark - - X Untitled - ATTO Disk Benchmark <u>File View H</u>elp <u>File View H</u>elp <u>File View H</u>elp <u>File View H</u>elp Eile <u>V</u>iew <u>H</u>elp D 📽 🖬 🏯 🔂 🕂 🕈 🕅 0 📽 🖬 🛎 🖪 🔶 🕈 🕅 0 📽 🖬 🗸 🕂 🕈 🕅 0 📽 🖬 🖉 🖪 🕂 📍 😢 D 📽 🖬 🗸 🕀 🕈 🕺 Force Write Access Direct I/O -Force Write Access Direct I/0 • Force Write Access Direct I/O [-9-] Force Write Access Direct I/O Drive: [-y-] Drive: [-y-] Drive: [·y·] Drive: - Force Write Access [-\$-] Drive: C 1/0 Comparison C 1/0 Comparison C I/O Comparison C 1/0 Comparison Transfer Size: 0.5 💌 to 8192.0 💌 KB Transfer Size: 0.5 💌 jo 8192.0 💌 KB Transfer Size: 0.5 💌 to 8192.0 💌 KB Transfer Size: 0.5 💌 to 8192.0 💌 KB Transfer Size: 0.5 💌 to 8192.0 💌 KB Qverlapped I/O Total Length: 256 MB 💌 C Neither Total Length: 256 MB 💌 C Neither Total Length: 256 MB ▼ C Neither Total Length: 256 MB 💌 C Neither Total Length: 256 MB 👻 Queue Depth: 4 💌 Queue Depth: 4 -Queue Depth: 4 💌 Queue Depth: 4 💌 Controlled by Controlled b Controlled b Controlled by ontrolled <u>b</u>y Start Start Start Start -• • ¥ -<< Description >: << Description >> << Description >> << Description > << Description >> Test Results - Test Results Test Results - Test Results -Test Results write Read -Write Read Read Read Write Read = Write Read Write Read -Write Write I Read = Write Write Bead 4424 7205 13534 13906 26295 27136 46869 43596 88749 78223 144340 117135 143643 236377 148742 392254 144264 479742 116135 445894 122760 465265 0.5 1.0 2.0 4.0 4573 6495 14737 13975 27648 26557 46080 44390 5647 11235 21798 4551 14848 28318 7022 13940 4446 0.5 1.0 13245 2.0 2.0 💻 26046 2.0 22982 2.0 44390 72527 118870 39836 63015 34049 58370 45511 83658 44281 78610 4.0 84276 72527 121714 118870 191151 223665 125720 116851 196318 222564 115395 103198 16. 16.0 161 214493 210857 227276 358454 32 208835 380887 225090 440648 206071 351054 64 234094 377153 128 128.0 228576 480889 128 247094 501276 248822 438584 334707 454975 100 256.0 256 255983 442732 355543 453876 256 238080 368561 512. 512 512 402653 461682 336948 463962 368561 478281 417566 481498 203360 397682 228942 459424 1024. 129366 483667 1024 1024. 385314 465110 1024.0 119040 474058 129992 482580 103910 492542 2048 2048 357913 462819 409378 477218 2048 213467 477218 238080 477218 4096. 4096. 4096.0 4096. 4096.0 8192.0 8192.0 296286 471974 81921 8192 8192.0 100 150 200 250 300 350 400 450 500 Transfer Rate - MB / Sec 0 50 100 150 200 250 300 350 400 450 500 Transfer Rate - MB / Sec 0 50 100 150 200 250 300 350 400 450 500 Transfer Rate - MB / Sec 0 100 200 300 400 500 600 700 800 900 1000 Transfer Rate - MB / Sec 0 50 100 150 200 250 300 350 400 450 500 Transfer Rate - MB / Sec 0 50 For Help, press F1 or Help, press F1 or Help, press F1 r Help, press F1 Help, press F1

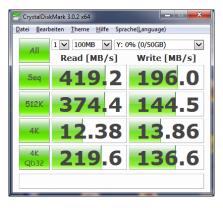
Drive Y: iSCSI 50 GB

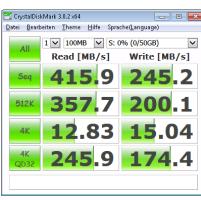
AI



atei <u>B</u> ear	beiten <u>T</u> heme <u>H</u>	ilfe Sprac	he(<u>L</u> anguage)	
All	1 🔽 100MB	Y: 0	% (0/50GB)	~
740	Read [ME	8/s]	Write [ME	8/s]
Seq	431	.5	172	.3
512K	381	.4	150	.2
4K	12.5	57	13.5	59
4K QD32	152	.5	118	.1







Pool from Single 480 GB SSD

_ D _X

Direct I/O

C Neither

C 1/0 Comparison

Overlapped I/O

Queue Depth: 4 💌

Start

Write Read

222947 467021 260990 432507 345625 449533

451694 460551 442128 458304

494811 467424

401506 465831

Drive Z: same Pool via SMB

CrystalDisk	<mark 3.0.2="" th="" x64<=""><th></th></mark>	
<u>D</u> atei <u>B</u> earb	eiten <u>T</u> heme <u>H</u> ilfe Spra	che(<u>L</u> anguage)
	1 🗸 100MB 🖌 Z: 0	0% (0/84GB) ✓
All	Read [MB/s]	Write [MB/s]
Seq	148.0	68. 67
512K	144.8	189 .3
4K	16 .23	14.22
4K QD32	164 .9	62. 35
1		

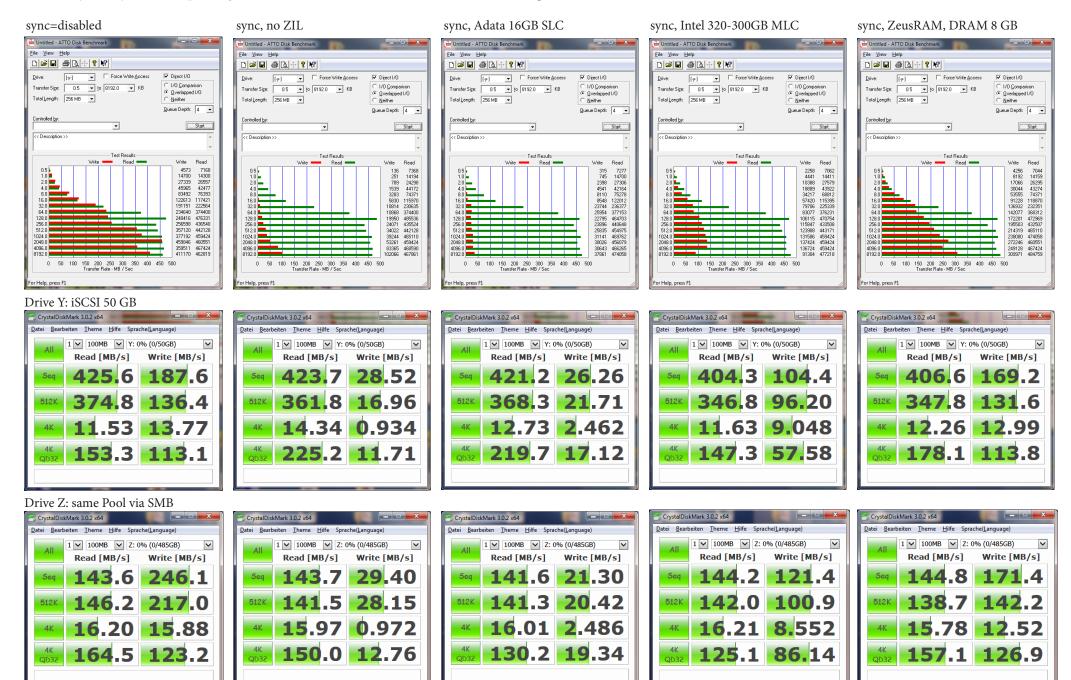
All		0% (0/84GB)
Seq	144.0	224 .9
512K	142.8	213.8
4K	16.28	15.96
4K	160 2	447 3
QD32	160 .2	117.2

	kMark 3.0.2 x64 beiten <u>T</u> heme <u>H</u> ilfe Spra	che(Language)
All		0% (0/84GB) 🔽 Write [MB/s]
Seq	146 .2	233.1
512K	144.5	113 .5
4K	16 .09	15 .78
4K QD32	166 .2	108 .0

		che(<u>L</u> anguage))% (0/84GB)
All	Read [MB/s]	Write [MB/s]
Seq	146 .5	243.4
512K	142.7	226.1
4K	15 .66	15 .66
4K QD32	161 .0	126.2

📑 CrystalDis	kMark 3.0.2 x64	
<u>D</u> atei <u>B</u> eart		che(<u>L</u> anguage)
All		0% (0/485GB)
	Read [MB/s]	Write [MB/s]
Seq	141 .7	235.4
512K	141.1	224.6
4K	16 .21	16 .21
4K QD32	164 .5	<mark>98.</mark> 68
11/1///////////////////////////////////		

Test 3. Async vs sync write depending on ZIL, Pool build from 5 x vdevs, each from a basic Seagate 15k/m disks (Raid-0)



Filebench randomwrite.f, 30s 44393.296 ops/s, 346.8mb/s,

Filebench randomwrite.f 30s 8808.833 ops/s, 68.8mb/s

Filebench randomwrite.f 30s 12240.467 ops/s, 95.6mb/s

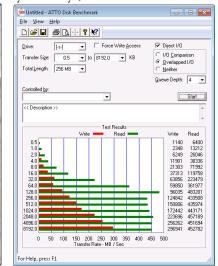
Filebench randomwrite.f 30s 2283.002 ops/s, 17.8mb/s

Filebench randomwrite.f 30s 4068.654 ops/s, 31.8mb/s

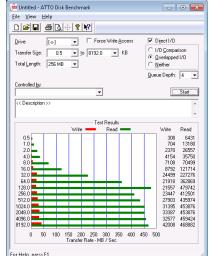
Test 4. Async vs sync write depending on ZIL on a SSD Pool, Pool build from 1 x vdev from a basic Winkom SSD 480 GB, important is the 4k QD32 value

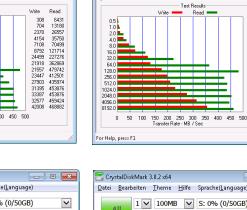
sync=disabled 🚔 Untitled - ATTO Disk Benchmark _ - × <u>File View H</u>elp □ 📽 🖬 🛎 📐 🕂 የ 🕺 Drive: [-\$-] ▼ Force Write Access Direct I/D C 1/0 Comparison Transfer Sige: 0.5 💌 to 8192.0 💌 KB Overlapped I/O Total Length: 256 MB 👻 C Neither Queue Depth: 4 💌 Controlled by: Start • < Description >> - Test Results Write Read • Write Read 4608 6767 14700 14300 28672 26046 45066 43274 79198 75832 127276 116260 209841 228384 0.5 2.0 🚍 40 8.0 64.0 210857 370791 222947 467021 260990 432507 256.0 345625 449533 451694 460551 512.0 1024. 2048.0 442128 458304 494811 467424 401506 465831 4096.0 0 50 100 150 200 250 300 350 400 450 500 Transfer Rate - MB / Sec For Help, press F1

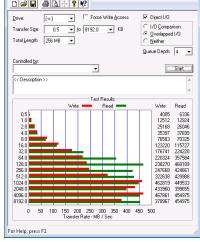
sync=always, no ZIL



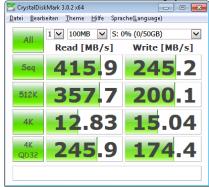






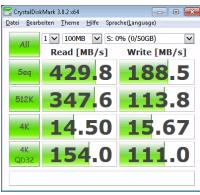


Drive S: iSCSI 50 GB, Pool empty



		iche(Language)
All	1 V 100MB V S: 0 Read [MB/s]	0% (0/50GB) Write [MB/s]
Seq	420.5	101 .9
512K	366.0	79. 78
4K	13 .16	6. 934
4K QD32	206 .4	45 .76

苦 CrystalDis	kMark 3.0.2 x64	
<u>D</u> atei <u>B</u> ear	beiten <u>T</u> heme <u>H</u> ilfe Spra	che(<u>L</u> anguage)
All	1 V 100MB V S: 0	0% (0/50GB)
	Read [MB/s]	Write [MB/s]
Seq	414.6	27 .36
512K	365.3	21 .64
4K	13 .52	2.421
4K QD32	158 .9	11.85



sync, ZeusRAM Dram ZIL

Transfer Size: 0.5 💌 to 8192.0 💌 KB

▼ Force Write Access

-

Direct 1/0

C Neither

C 1/0 Comparison

Overlapped I/O

Queue Depth: 4 💌

Start

Write Read

3712 6639 7278 10113 13115 25982 26295 38149 46869 71204 73107 112058

122012 224220 133218 366479

177165 483201

199255 425626 210537 438006

231409 435974 276737 453876

315065 446329

357913 462819

角 Untitled - ATTO Disk Benchmark

□ 🔎 🖬 🛎 🖪 🔂 🕂 💡 📢

[-\$-]

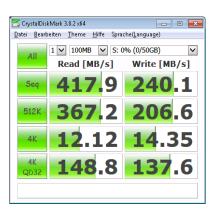
Total Length: 256 MB 👻

Eile ⊻iew <u>H</u>elp

Controlled by:

CC Description 33

Drive:

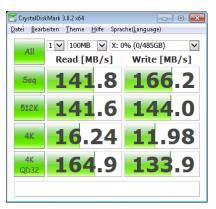


Drive X: same Pool via SMB

🔚 CrystalDis	kMark 3.0.2 ×64	- • •
<u>D</u> atei <u>B</u> earl	oeiten <u>T</u> heme <u>H</u> ilfe Spra	che(<u>L</u> anguage)
All		0% (0/485GB)
_	Read [MB/s]	Write [MB/s]
Seq	141.6	248.1
512K	142 .3	239.3
4K	16 .05	16 .13
4K QD32	164 .8	133.0
-		

😤 CrystalDiskMark 3.0.2 xt	64		
<u>D</u> atei <u>B</u> earbeiten <u>T</u> hen	ne <u>H</u> ilfe Sprad	:he(Language)	
All 1 🗸 100	MB 🔽 X: 0	% (0/485GB)	~
	[MB/s]	Write [MI	3/s]
Seq 14	1.6	119	.1
512K 14	2.6	109	.9
^{4K} 15	.71	7.71	L 3
4K 16	8. 0	62.4	12

		Iche(Language) 0% (0/485GB)
All	Read [MB/s]	Write [MB/s]
Seq	142.1	22 .96
512K	140 .2	20 .18
4K	16 .26	2 .497
4K QD32	162 .0	20.13

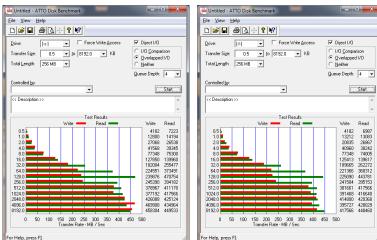


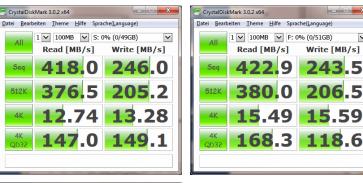
	kMark 3.0.2 x64 beiten <u>T</u> heme <u>H</u> ilfe Spra	che(Language)
All	1 V 100MB V X: Read [MB/s]	0% (0/485GB)
Seq	141.3	247.3
512K	138 .9	228 .1
4K	16 .07	16 .25
4K QD32	146 .8	127.0

Test 5: special configurations

4K

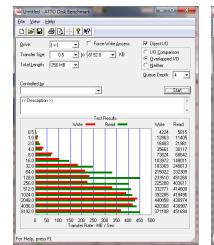
sync=off, iSCSI, volume LU, SSD sync=off, iSCSI, file LU, SSD

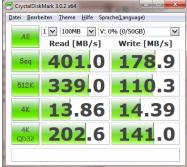






4 x vdevs, each from a basic disk 1 x vdev Z1 from 4 datadisks (4+1)





 \checkmark

~

Write [MB/s]

243.5

206.5

15.59

Filebench fileserver.f 13594.182 ops/s, (1236/2472 r/w), 327.4mb/s, 393us cpu/op, 12.8ms latency

Filebench randomrw.f 88637.352 ops/s, (86004/2634 r/w), 692.5mb/s, 13us cpu/op, 0.0ms latency

Filebench webserver.f 458002.397 ops/s, (147742/14777 r/w), 2430.2mb/s, 55us cpu/op, 0.3ms latency

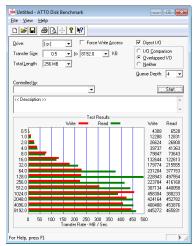
Filebench fileserver.f

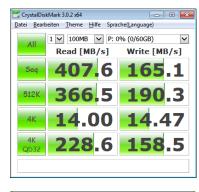
9352.514 ops/s, (850/1701 r/w), 224.4mb/s, 474us cpu/op, 18.9ms latency

Filebench randomrw.f 86419.294 ops/s, (83691/2728 r/w), 675.1mb/s, 17us cpu/op, 0.0ms latency

Filebench webserver.f 456351.152 ops/s, (147209/14723 r/w), 2420.4mb/s, 55us cpu/op, 0.3ms latency

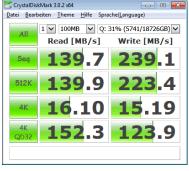
4 x Z2, each 7 disks RE4 5400rpm





iSCSI

SMB



Backup pool (green WD disks RE4)

dd: 1800 MB/s write, 4000 MB/s read fileserver.f

29950.846 ops/s, (2723/5446 r/w), 726.0mb/s, 604us cpu/op, 5.1ms lat

Question: Volume or Filebased Logical Units?

Volumbased LUs are minimal faster, but not as easy to handle compared to filebased LUs regarding copy/move/backup/restore from snap.

167.4 153.5

More vdevs or Raid-Z? (same amount of datadisks/poolsize)

50% faster on latency, r/w and cpu/op than the Raid-Z1.

If you look at sequential performance, they are similar, Z1 even slightly

faster. If you look at the fileserver-filebench, the multi-vdev option is up to

Untitled - ATTO Disk Benchmark

0 2 8 8 8 8 9 8

Total Length: 256 MB 💌

[·₂·] ▼ Force Write Access

0.5 • 10 8192.0 • KB

•

0 50 100 150 200 250 300 350 400 450 500 Transfer Rate - MB / Sec

Datei Bearbeiten Theme Hilfe Sprache(Language)

Read [MB/s]

405

367

5 🗸 100MB 🖌 Z: 0% (0/50GB)

0

3

13.52 13.85

Test Results

Read 1

Direct I/O

C I/O Comparison

○ <u>D</u>verlapped I/O ○ <u>N</u>either

Write

Queue Depth: 4 💌

Start

Read

218270 379011 225633 408846 245390 398094

240300 330034 398103 417566 426088 413887 390925 430974 439029 427056 400320 451504

409378 451694

- - X

Write [MB/s]

207.3

190.1

 $\mathbf{\vee}$

File View Help

Transfer Size:

0

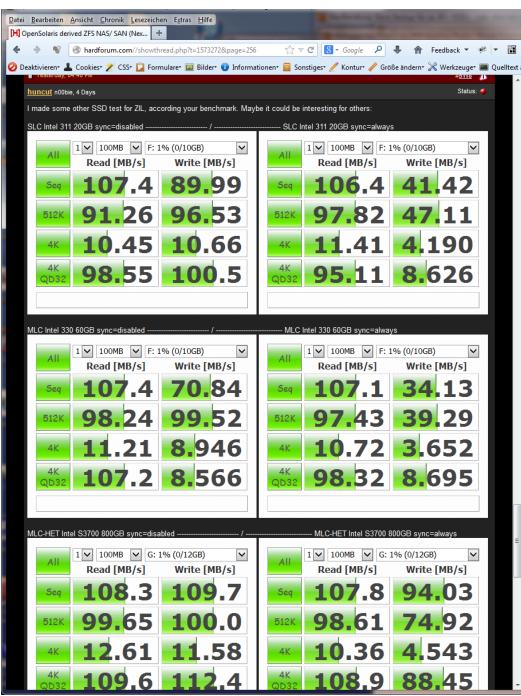
4096.0

r Help, press E

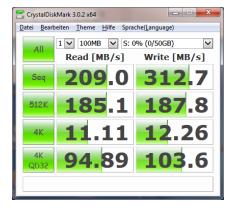
CrystalDiskMark 3.0.2 x64

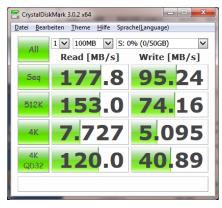
Drive:

More Benchmarks (sync vs async Performance - Is this a good Zil? Look mostly at 4k QD32 with sync=always and 32 concurrent small 4k writes

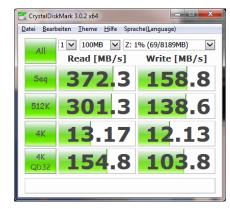


Winkom SSD 120 GB (SF1222, Intel SLC Nand, high IOPS)





10 GbE iSCSI, sync=always



10 GbE iSCSI, sync=always, best of all 4k QD32

10 Gbe iSCSI, sync=disabled

ZeusRAM (8 GB DRAM based)

All		% (69/8189MB)
	Read [MB/s]	Write [MB/s]
Seq	368.5	290.3
512K	348.9	248.3
4K	14 .74	15.03
4K QD32	209.1	169.4

10 Gbe iSCSI, sync=disabled

1GB network,