Technologia NVMe staje się coraz bardziej popularna, w środowisku testowym przeprowadzono następujące testy w oparciu o platformę SYS-1029U-TN10RT, wykorzystując możliwośći Intel VROC. Testy przeprowadzono na dystrybucjach Linux Centos, Linux Red Hat 7.4 oraz Linux Red Hat 7.5.

Konfiguracja 1.1

Serwer SYS-1029U-TN10RT-SKL4116-256GB-NVMe

Lp	Nazwa	Opis	Jedn	Ilość
1	SYS-1029U-TN10RT	Dual Xeon Scalable LGA3647, Up to 3TB DDR4; 2xPCI-E x16; 2x10GBase-T Intel X540; 10x2.5" NVMe, PWS-R 1000W Titanium	szt.	1
2	P4X-SKL4116-SR3HQ	Intel® Xeon® Scalable Skylake Silver 4116 Processor 12C/24T 16.5M Cache, 2.10 GHz	szt.	2
3	MEM-DR432L-SL01-ER29	32GB DDR4-2933 2Rx4 LP ECC RDIMM,HF,RoHS	szt.	8
4	HDS-IUN2-SSDPE2KX010T8	Intel DC P4510 1TB NVMe PCIe 3.0 3D TLC 2.5" 15mm 1DWPD, FW VDV10131	szt.	10
5	AOC-VROCPREMOD	INTEL VROC PREMIUM, RAID0,1,5,10,HF,ROHS not only for Intel SSDs, doing raid 0,1,5,10	szt.	1

Szczegółowa konfiguracja 1029U-TN10RT

Procesor - Intel Xeon Scalable 4116 – 2 sztuki, każdy:

- of Cores 12
- of Threads 24
- Processor Base Frequency 2.10 GHz
- Max Turbo Frequency 3.00 GHz
- Cache 16.5 MB L3

Pamięć RAM - MEM-DR432L-SL01-ER29
8x 32GB DDR4 2933MHz ECC REG DIMM – 256GB

Dyski NVMe - HDS-IUN2-SSDPE2KX010T8 – 10 sztuk

- Sequential Read (up to) 2850 MB/s
- Sequential Write (up to) 1100 MB/s
- Random Read (100% Span) 465000 IOPS
- Random Write (100% Span) 70000 IOPS

Intel VROC Premium – RAID 0,1,5,10

Wyniki testów przed optymalizacją

Linux Centos 7 10xNVMe RAID5

RAID 5 z 10 dysków NVMe WRITE

```
[root@centos]# fio --name=writefile --size=100G --filesize=100G --filename=disk_test.bin --bs=1M --nrfiles=1 --direct=1 --sync=0 --
randrepeat=0 --rw=write --refill_buffers --end_fsync=1 --iodepth=200 --ioengine=libaio
writefile: (g=0): rw=write, bs=(R) 1024KiB-1024KiB, (W) 1024KiB-1024KiB, (T) 1024KiB-1024KiB, ioengine=libaio, iodepth=200
fio-3.1

  WRITE: bw=1511MiB/s (1585MB/s), 1511MiB/s-1511MiB/s (1585MB/s-1585MB/s), io=100GiB (107GB), run=67750-67750msec
Disk stats (read/write):
```

RAID 5 z 10 dysków NVMe READ

[root@centos]# fio --time_based --name=benchmark --size=100G --runtime=30 --filename=disk_test.bin --ioengine=libaio --randrepeat=0 --iodepth=128 --direct=1 --invalidate=1 --verify=0 --verify_fatal=0 --numjobs=4 --rw=randread --blocksize=4k --group_reporting
benchmark: (g=0): rw=randread, bs=(R) 4096B-4096B, (W) 4096B-4096B, (T) 4096B-4096B, ioengine=libaio, iodepth=128

READ: bw=1723MiB/s (1807MB/s), 1723MiB/s-1723MiB/s (1807MB/s-1807MB/s), io=50.5GiB (54.2GB), run=30001-30001msec
Disk stats (read/write):

Wyniki testów Red Hat 7.5 6xNVMe RAID 5, 4xNVMe RAID 0

md124 : active raid0 nvme6n1[3] nvme7n1[2] nvme8n1[1] nvme9n1[0]
md126 : active raid5 nvme0n1[5] nvme1n1[4] nvme2n1[3] nvme3n1[2] nvme4n1[1] nvme5n1[0]

RAID 5 6xNVMe WRITE

[root@rh7.5]# fio --name=writefile --size=100G --filesize=100G --filename=disk_test.bin --bs=1M --nrfiles=1 --direct=1 --sync=0 --randrepeat=0 --rw=write --refill_buffers --end_fsync=1 --iodepth=200 --ioengine=libaio
writefile: (g=0): rw=write, bs=(R) 1024KiB-1024KiB, (W) 1024KiB-1024KiB, (T) 1024KiB-1024KiB, ioengine=libaio, iodepth=200

WRITE: bw=1572MiB/s (1648MB/s), 1572MiB/s-1572MiB/s (1648MB/s-1648MB/s), io=100GiB (107GB), run=65149-65149msec

RAID 0 4xNVMe WRITE

[root@rh7.5 raid0]# fio --name=writefile --size=100G --filesize=100G --filename=disk_test.bin --bs=1M --nrfiles=1 --direct=1 --sync=0 --randrepeat=0 --rw=write --refill_buffers --end_fsync=1 --iodepth=200 --ioengine=libaio
writefile: (g=0): rw=write, bs=(R) 1024KiB-1024KiB, (W) 1024KiB-1024KiB, (T) 1024KiB-1024KiB, ioengine=libaio, iodepth=200

WRITE: bw=2411MiB/s (2528MB/s), 2411MiB/s-2411MiB/s (2528MB/s-2528MB/s), io=100GiB (107GB), run=42467-42467msec

RAID 5 6xNVMe READ

[root@rh7.5]# fio --time_based --name=benchmark --size=100G --runtime=30 --filename=disk_test.bin --ioengine=libaio --randrepeat=0 --iodepth=128 --direct=1 --invalidate=1 --verify=0 --verify_fatal=0 --numjobs=4 --rw=randread --blocksize=4k –group_reporting

READ: bw=1621MiB/s (1700MB/s), 1621MiB/s-1621MiB/s (1700MB/s-1700MB/s), io=47.5GiB (51.0GB), run=30001-30001msec

RAID 0 4xNVMe READ

[root@rh7.5 raid0]# fio --time_based --name=benchmark --size=100G --runtime=30 --filename=disk_test.bin --ioengine=libaio --randrepeat=0 --iodepth=128 --direct=1 --invalidate=1 --verify=0 --verify_fatal=0 --numjobs=4 --rw=randread --blocksize=4k –group_reporting

READ: bw=1882MiB/s (1974MB/s), 1882MiB/s-1882MiB/s (1974MB/s-1974MB/s), io=55.1GiB (59.2GB), run=30001-30001msec

RAID 5 6xNVMe WRITE

[root@rh7.5]# fio --time_based --name=benchmark --size=100G --runtime=30 --filename=disk_test.bin --ioengine=libaio --randrepeat=0 --iodepth=128 --direct=1 --invalidate=1 --verify=0 --verify_fatal=0 --numjobs=4 --rw=randwrite --blocksize=4k –group_reporting

WRITE: bw=207MiB/s (218MB/s), 207MiB/s-207MiB/s (218MB/s-218MB/s), io=6225MiB (6527MB), run=30006-30006msec

RAID 0 4xNVMe WRITE

[root@rh7.5 raid0]# fio --time_based --name=benchmark --size=100G --runtime=30 --filename=disk_test.bin --ioengine=libaio --randrepeat=0 --iodepth=128 --direct=1 --invalidate=1 --verify=0 --verify_fatal=0 --numjobs=4 --rw=randwrite --blocksize=4k --group_reporting

WRITE: bw=441MiB/s (463MB/s), 441MiB/s-441MiB/s (463MB/s-463MB/s), io=12.9GiB (13.9GB), run=30001-30001msec

Podsumowanie:

Test	Parameters			RAID 5		RAID 0
	Block size	**# of jobs**	**IOD**	**6 NVMe**	**10 NVMe**	**4 NVMe**
Seq write (GB/s)	1024KiB	1	200	1.648	1.029	2.528
Rand read (KIOPS)	4kB	4	128	417	378	483
Rand write (KIOPS)	4kB	4	128	60.1	37.1	114

Wyniki testów po optymalizacji

Wyniki testów Red Hat 7.5 6xNVMe RAID 5, 4xNVMe RAID 0

Dla pojedynczego dysku:

*[root@rh7.5]# fio --name=dummy --size=50G --runtime=20 --filename=/dev/nvme0n1 --ioengine=libaio --direct=1 --rw=randread -bs=4k** --iodepth=64* ***--numjobs=8*** *--group_reporting*
dummy: (g=0): rw=randread, bs=(R) 4096B-4096B, (W) 4096B-4096B, (T) 4096B-4096B, ioengine=libaio, iodepth=64

READ: bw=2517MiB/s (2640MB/s), 2517MiB/s-2517MiB/s (2640MB/s-2640MB/s), io=49.2GiB (52.8GB), run=20003-20003msec

RAID 0 4xNVMe READ

[root@rh7.5]# fio --name=dummy --size=50G --runtime=20 --filename=/dev/md124p1 --ioengine=libaio --direct=1 --rw=randread -bs=4k** --iodepth=64 **--numjobs=48** --group_reporting
dummy: (g=0): rw=randread, bs=(R) 4096B-4096B, (W) 4096B-4096B, (T) 4096B-4096B, ioengine=libaio, iodepth=64

READ: bw=7208MiB/s (7558MB/s), 7208MiB/s-7208MiB/s (7558MB/s-7558MB/s), io=141GiB (151GB), run=20007-20007msec

RAID 5 6xNVMe RAID 5

[root@rh7.5]# fio --name=dummy --size=50G --runtime=20 --filename=/dev/md126p1 --ioengine=libaio --direct=1 --rw=randread -bs=4k** --iodepth=64 **--numjobs=48** --group_reporting
dummy: (g=0): rw=randread, bs=(R) 4096B-4096B, (W) 4096B-4096B, (T) 4096B-4096B, ioengine=libaio, iodepth=64

READ: bw=6116MiB/s (6413MB/s), 6116MiB/s-6116MiB/s (6413MB/s-6413MB/s), io=47.0GiB (51.5GB), run=8035-8035msec

Podsumowanie

Test	Parameters			RAID 5	RAID 0
	Block Size	**# of jobs**	**IOD**	**6 NVMe**	**4 NVMe**
Rand Read (KIOPS)	4kB	48	64	1566	1845
Rand Read (KIOPS)	4kB	16	64	1220	1690
Rand Read (KIOPS)	4kB	8	64	735	1098

Obciążenie CPU

Linux 3.10.0-957.21.3.el7.x86_64 18.06.2019 _x86_64_ (48 CPU)

	CPU	%usr	%nice	%sys	%iowait	%irq	%soft	%steal	%guest	%gnice	%idle
21:28:46	all	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	100,00
21:28:47	all	0,00	0,00	0,04	0,00	0,00	0,00	0,00	0,00	0,00	99,96
21:28:48	all	0,00	0,00	0,02	0,00	0,00	0,00	0,00	0,00	0,00	99,98
21:28:49	all	0,06	0,00	0,21	0,00	0,00	0,00	0,00	0,00	0,00	99,73
21:28:50	all	9,01	0,00	21,70	0,00	0,00	0,02	0,00	0,00	0,00	69,27
21:28:51	all	17,91	0,00	46,79	1,01	0,00	0,00	0,00	0,00	0,00	34,30
21:28:52	all	18,04	0,00	47,28	1,01	0,00	0,00	0,00	0,00	0,00	33,68
21:28:53	all	18,19	0,00	47,19	0,02	0,00	0,00	0,00	0,00	0,00	34,59
21:28:54	all	18,33	0,00	46,53	0,94	0,00	0,00	0,00	0,00	0,00	34,21
21:28:55	all	18,15	0,00	46,39	0,50	0,00	0,02	0,00	0,00	0,00	34,93
21:28:56	all	17,76	0,00	46,38	0,45	0,00	0,00	0,00	0,00	0,00	35,40
21:28:57	all	18,08	0,00	45,98	0,00	0,00	0,00	0,00	0,00	0,00	35,94
21:28:58	all	17,99	0,00	45,82	0,14	0,00	0,02	0,00	0,00	0,00	36,03
21:28:59	all	17,96	0,00	45,27	0,02	0,00	0,00	0,00	0,00	0,00	36,74
21:29:00	all	17,95	0,00	45,26	0,39	0,00	0,02	0,00	0,00	0,00	36,38
21:29:01	all	18,13	0,00	46,16	0,14	0,00	0,00	0,00	0,00	0,00	35,57
21:29:02	all	17,93	0,00	45,57	0,23	0,00	0,00	0,00	0,00	0,00	36,27
21:29:03	all	17,66	0,00	45,08	0,00	0,00	0,00	0,00	0,00	0,00	37,26
21:29:04	all	17,92	0,00	44,68	0,00	0,00	0,02	0,00	0,00	0,00	37,39
21:29:05	all	17,58	0,00	44,78	0,14	0,00	0,00	0,00	0,00	0,00	37,50
21:29:06	all	17,47	0,00	45,00	0,16	0,00	0,00	0,00	0,00	0,00	37,37
21:29:07	all	17,49	0,00	44,39	0,16	0,00	0,00	0,00	0,00	0,00	37,97
21:29:08	all	17,80	0,00	43,95	0,00	0,00	0,00	0,00	0,00	0,00	38,25
21:29:09	all	17,70	0,00	44,37	0,02	0,00	0,02	0,00	0,00	0,00	37,88
21:29:10	all	8,98	0,00	22,02	0,04	0,00	0,02	0,00	0,00	0,00	68,93
21:29:11	all	0,00	0,00	0,02	0,00	0,00	0,00	0,00	0,00	0,00	99,98
Średnia:	**all**	**13,50**	**0,00**	**34,35**	**0,20**	**0,00**	**0,01**	**0,00**	**0,00**	**0,00**	**51,94**

Konfiguracja 1.2

Serwer SYS-1029U-TN10RT-SKL8176-768GB-NVMe

Lp	Nazwa	Opis	Jedn	Ilość
1	SYS-1029U-TN10RT	Dual Xeon Scalable LGA3647, Up to 3TB DDR4; 2xPCI-E x16; 2x10GBase-T Intel X540; 10x2.5" NVMe, PWS-R 1000W Titanium	szt.	1
2	P4X-SKL8176-SR37A	SKL-SP 8176 28C/56T 2.1G 38.5M 10.4GT UPI Pokaż	szt.	2
3	MEM-DR432L-SL03-ER26	32GB DDR4-2666 2Rx4 LP ECC RDIMM,HF,RoHS	szt.	24
4	HDS-IUN2-SSDPE2KX040T8	Intel DC P4510 4TB NVMe PCIe 3.1 3D TLC 2.5" 15mm 1DWPD	szt.	10
5	AOC-VROCPREMOD	INTEL VROC PREMIUM, RAID0,1,5,10,HF,ROHS not only for Intel SSDs, doing raid 0,1,5,10	szt.	1

Szczegółowa konfiguracja 1029U-TN10RT
Procesor - Intel Xeon Scalable 4116 – 2 sztuki, każdy:

- of Cores 28
- of Threads 56
- Processor Base Frequency 2.10 GHz
- Max Turbo Frequency 3.80 GHz
- Cache 38.5 MB L3

Pamięć RAM - MEM-DR432L-SL01-ER29
8x 32GB DDR4 2933MHz ECC REG DIMM - 256GB

Dyski NVMe - HDS-IUN2-SSDPE2KX040T8 – 10 sztuk

- Sequential Read (up to) 3000 MB/s
- Sequential Write (up to) 2900 MB/s
- Random Read (100% Span) 636500 IOPS
- Random Write (100% Span) 111500 IOPS

Intel VROC Premium – RAID 0,1,5,10

Podsumowanie (po optymalizacji)

Red Hat 7.4 3xNVMe RAID 5, 8xNVMe RAID 5, 4xNVMe RAID 0

Test	Parameters			RAID 5		RAID 0
	Block size	# of jobs	IOD	3 NVMe	8 Nvme	4 NVMe
Seq write (GB/s)	128KB	1	128	1.267	1.23	3.517
		8	128	1	1.347	3.254
Rand read (KIOPS)	4KB	16	128	1530	1631	2739
Rand write (KIOPS)	4kB	16	128	74.8	40.1	297
		8	128	71.9	41.9	484