SNIA CMSI | COMPUTE, MEMORY, AND STORAGE

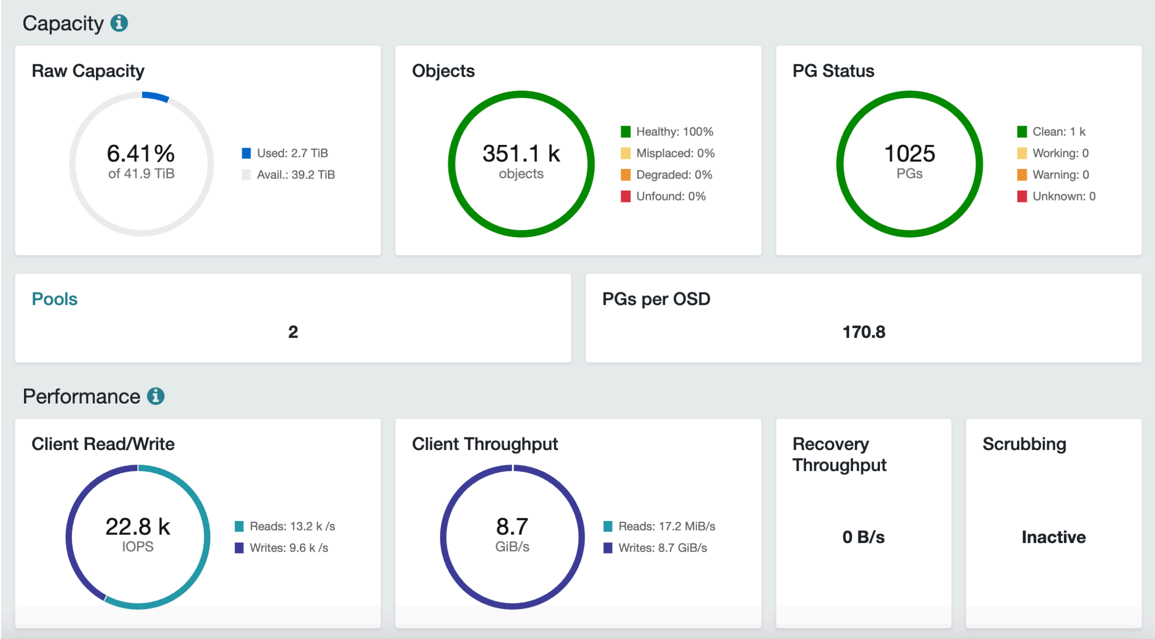
Leveraging Computational Storage for Cost Efficiency: TCO Case Study

SNIA CMSI

Presented by:
Jonmichael Hands, Chia Network

Agenda

- Why CS usecase – compression
- Case study – Ceph, zstandard vs CS
- TCO comparison



TCO in Computational Storage - Compression

- Data reduction increases effective capacity, and reduces TCO $ / TBe proportionally – but these techniques can be done with open source software on CPUs
- The SNIA Storage TCO model currently has a field for compression
- Compression and performance are relevant together, as generally there is a tradeoff of IOPS and/or CPU utilization
    - E.g. vSAN only able to turn on compression with SSDs
- Easy to show benefit of computational storage in synthetic workloads, raw disk io, but much harder with filesystems
- TCO reduction from CS can be shown for reduction in CPU, server consolidation. ISO performance TCO with CPU compression would require even more resources.

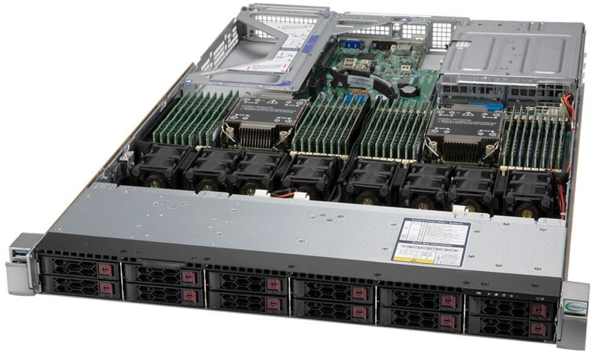
System Configuration

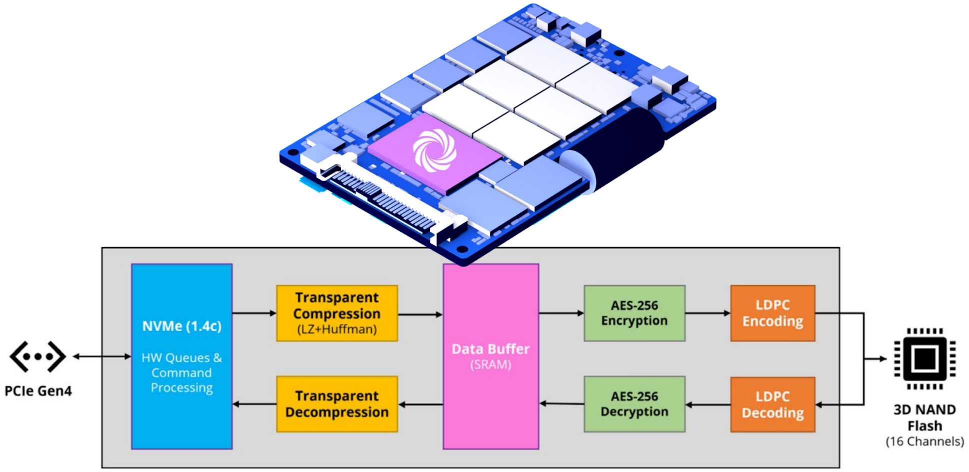
System Config

- Single Node Cluster, Ice Lake Server
- Supermicro Ultra SuperServer SYS-120U-TNR
- 2x Intel Xeon Gold 6338 CPU
- 512GB DDR4 @3200MHz, 32x 16GB DIMMs
- 12x ScaleFlux CSD 3000 3.84TB

Software Config

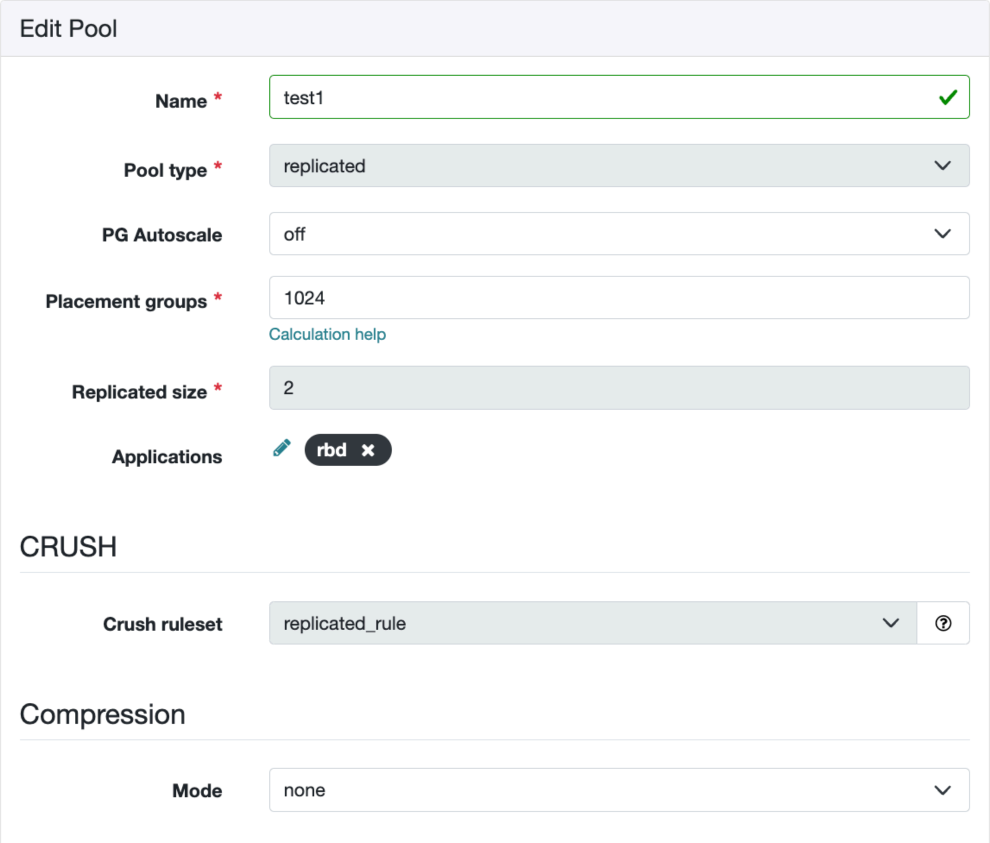
- Ubuntu 23.04, Kernel 6.2.0-20-generic
- ceph version 17.2.6 quincy
  - 2x replication
  - 40x RBD of 1TB each
- fio-3.33
  - 128k sequential read, write, random read, QD 128
  - Buffer compress = 60 (average 2:1 compression ratio)



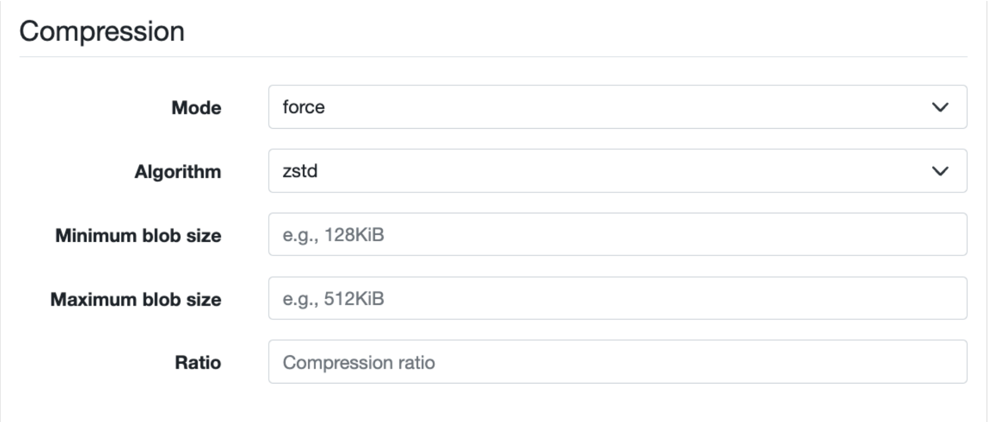


Test cases

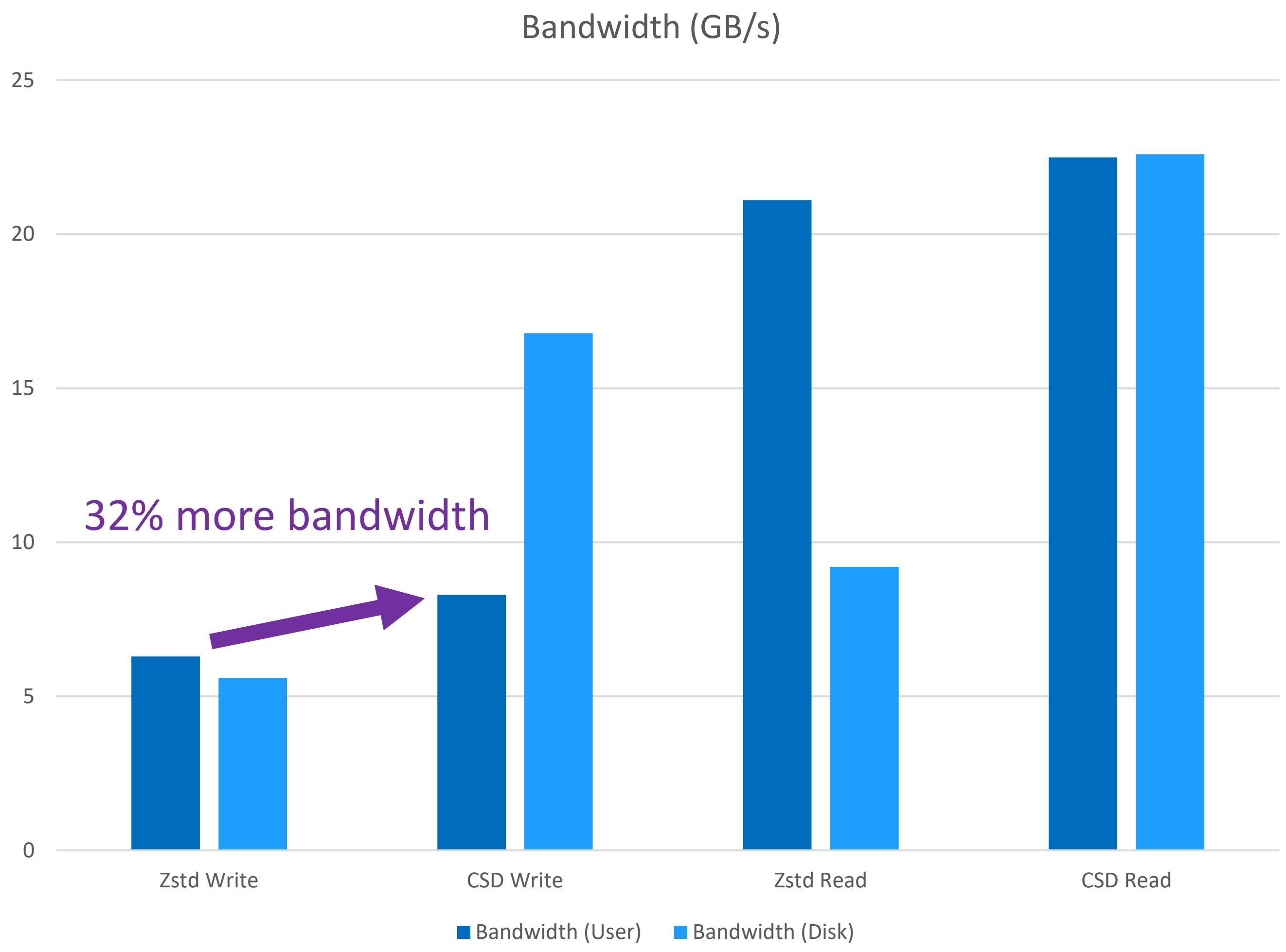
Test Case 2 – no compression (done on drive)

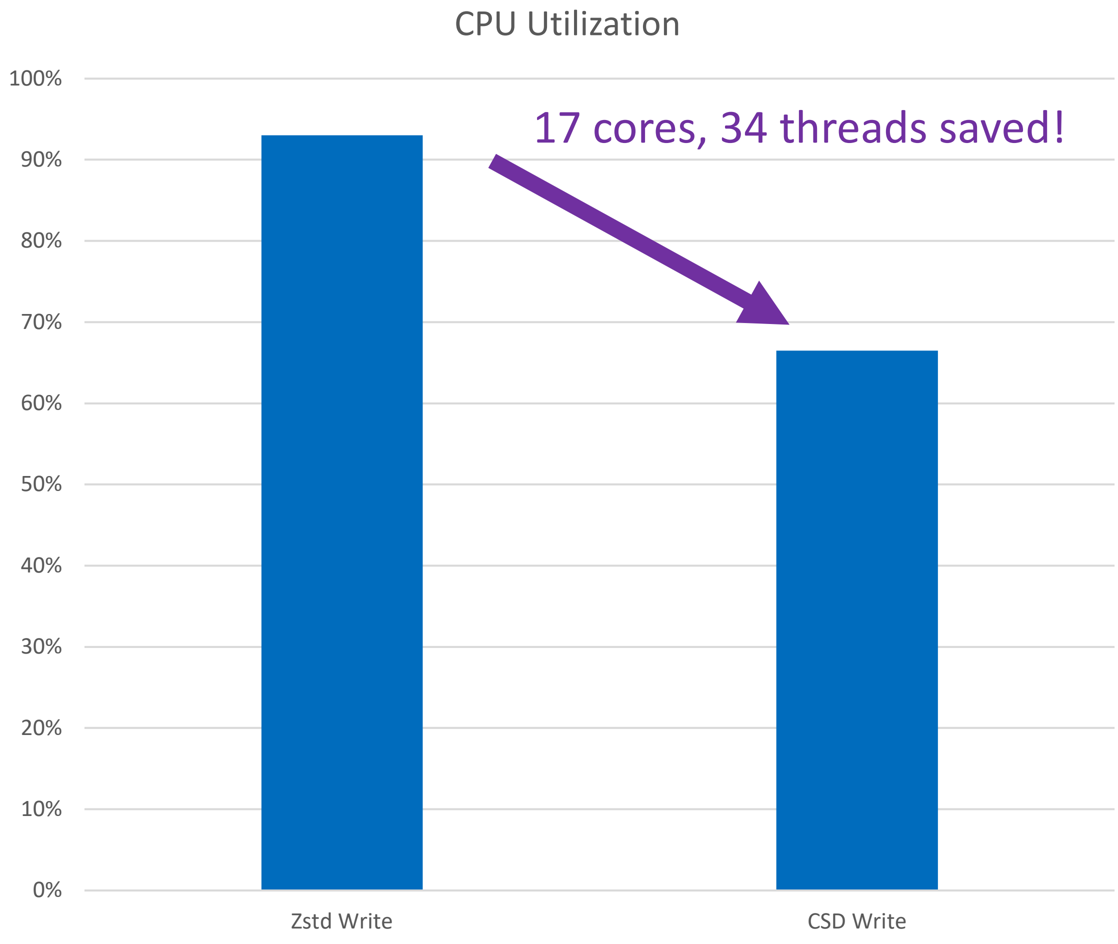


Test Case 1 – Zstandard force



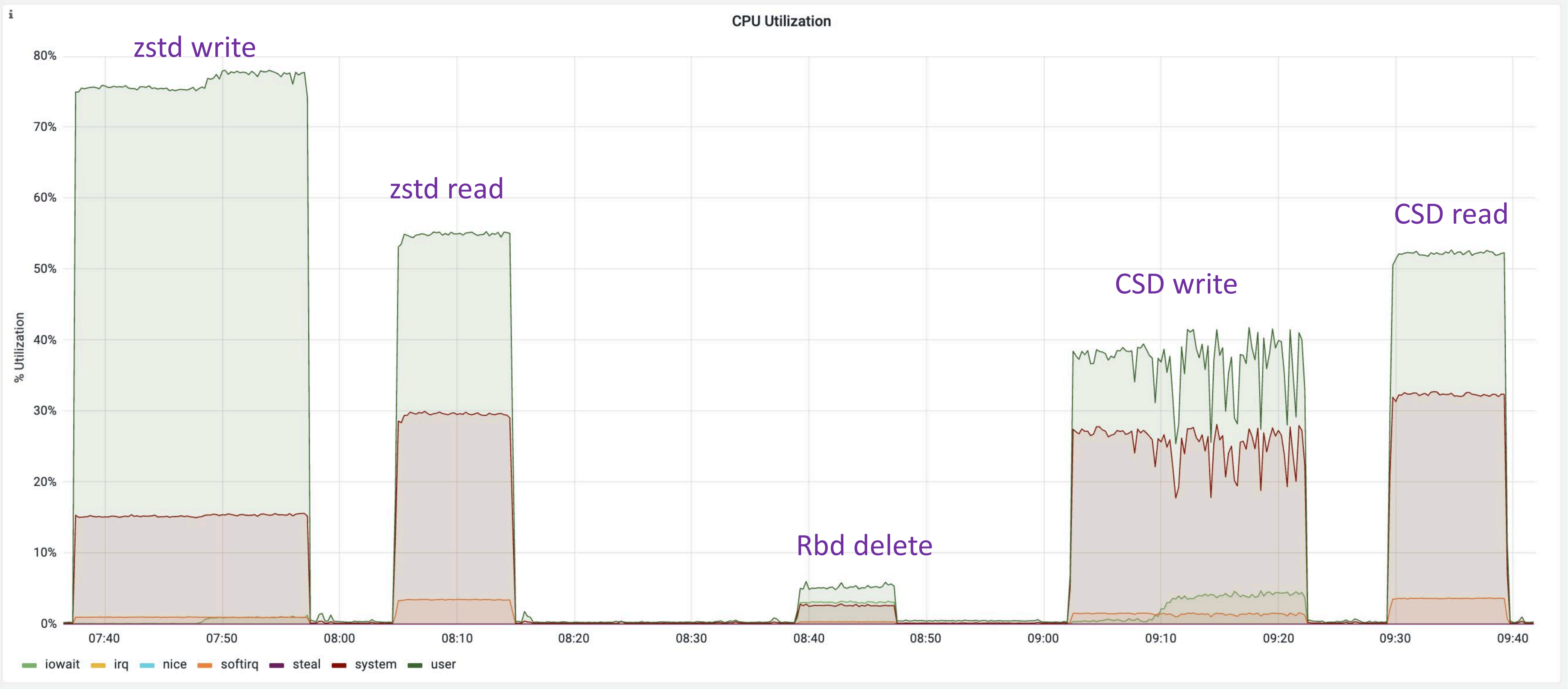
Impact of inline compression on CPU and bandwidth



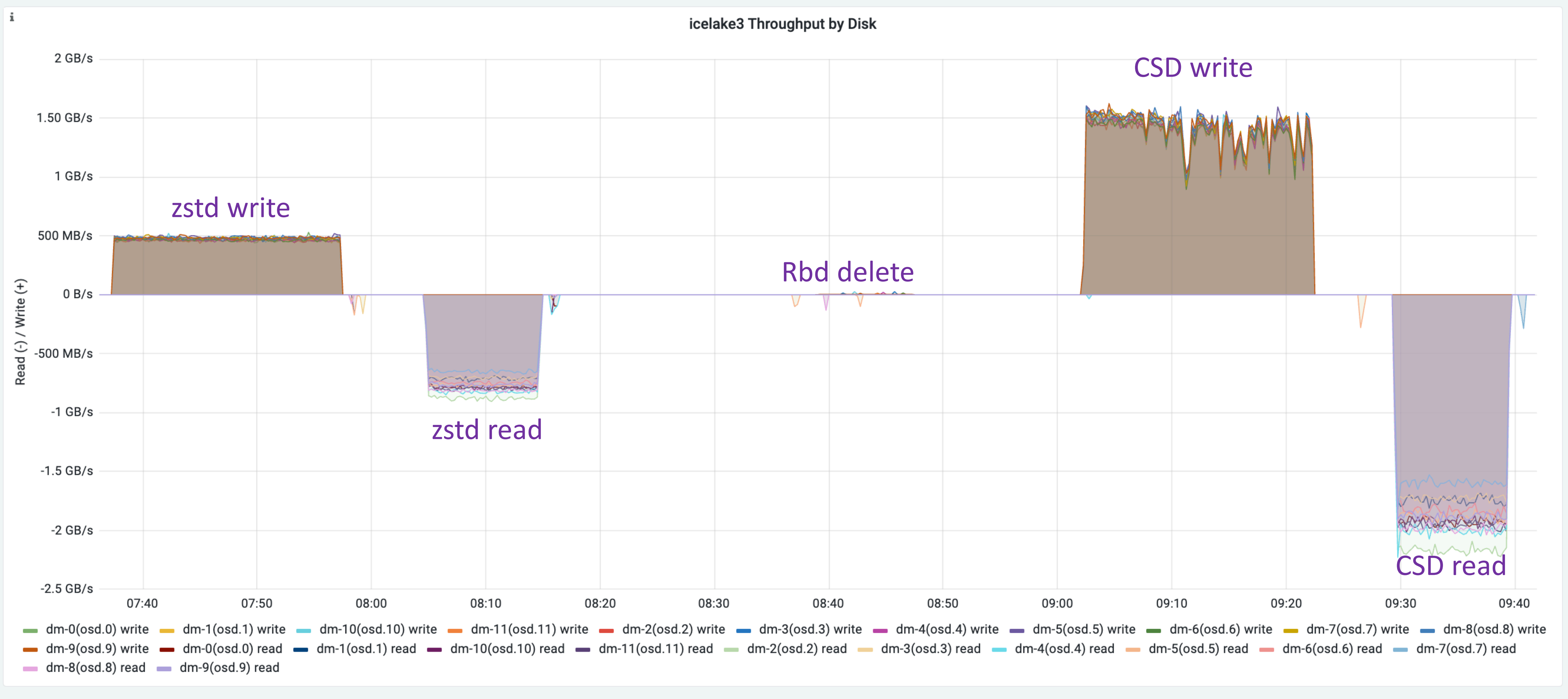


Raw data

Test Case	Bandwidth App (GB/s)	Bandwidth Disk (GB/s)	CPU Utilization	Compression Ratio
Zstd Write	6.3	5.6	93%	2.285
Zstd Read	21.1	9.2	87.8%	
CSD Write	8.3	16.79	66.5%	2
CSD Read	22.5	22.6	88.1%	



CPU Utilization
zstd write
zstd read
Rbd delete
CSD write
CSD read
80%
70%
60%
50%
40%
30%
20%
10%
0%
% Utilization
07:40
07:50
08:00
08:10
08:20
08:30
08:40
08:50
09:00
09:10
09:20
09:30
09:40
iowait
irq
nice
softirq
steal
system
user



icelake3 Throughput by Disk
2 GB/s
1.50 GB/s
1 GB/s
500 MB/s
0 B/s
-500 MB/s
-1 GB/s
-1.5 GB/s
-2 GB/s
-2.5 GB/s
Read (-) / Write (+)
07:40
07:50
08:00
08:10
08:20
08:30
08:40
08:50
09:00
09:10
09:20
09:30
09:40
zstd write
zstd read
Rbd delete
CSD write
CSD read
dm-0(osd.0) write
dm-1(osd.1) write
dm-10(osd.10) write
dm-11(osd.11) write
dm-2(osd.2) write
dm-3(osd.3) write
dm-4(osd.4) write
dm-5(osd.5) write
dm-6(osd.6) write
dm-7(osd.7) write
dm-8(osd.8) write
dm-9(osd.9) write
dm-0(osd.0) read
dm-1(osd.1) read
dm-10(osd.10) read
dm-11(osd.11) read
dm-2(osd.2) read
dm-3(osd.3) read
dm-4(osd.4) read
dm-5(osd.5) read
dm-6(osd.6) read
dm-7(osd.7) read
dm-8(osd.8) read
dm-9(osd.9) read

TCO Reduction – Save CPU Cores

Intel® Xeon® Gold 6338 Processor

Essentials

Product Collection	3rd Generation Intel® Xeon® Scalable Processors
Code Name	Products formerly Ice Lake
Vertical Segment	Server
Processor Number	6338
Lithography	10 nm
Recommended Customer Price	$2990.00

CPU Specifications

Total Cores	32
Total Threads	64
Max Turbo Frequency	3.20 GHz
Processor Base Frequency	2.00 GHz
Cache	48 MB
Intel® UPI Speed	11.2 GT/s
Max # of UPI Links	3
TDP	205 W

Intel® Xeon® Gold 5318N Processor

Essentials

Product Collection	3rd Generation Intel® Xeon® Scalable Processors
Code Name	Products formerly Ice Lake
Vertical Segment	Server
Processor Number	5318N
Lithography	10 nm
Recommended Customer Price	$1602.00

CPU Specifications

Total Cores	24
Total Threads	48
Max Turbo Frequency	3.40 GHz
Intel SpeedStep® Max Frequency	3.40 GHz
Processor Base Frequency	2.10 GHz
Cache	36 MB
Intel® UPI Speed	11.2 GT/s
Max # of UPI Links	3
TDP	150 W

SNIA Storage TCO Model Results

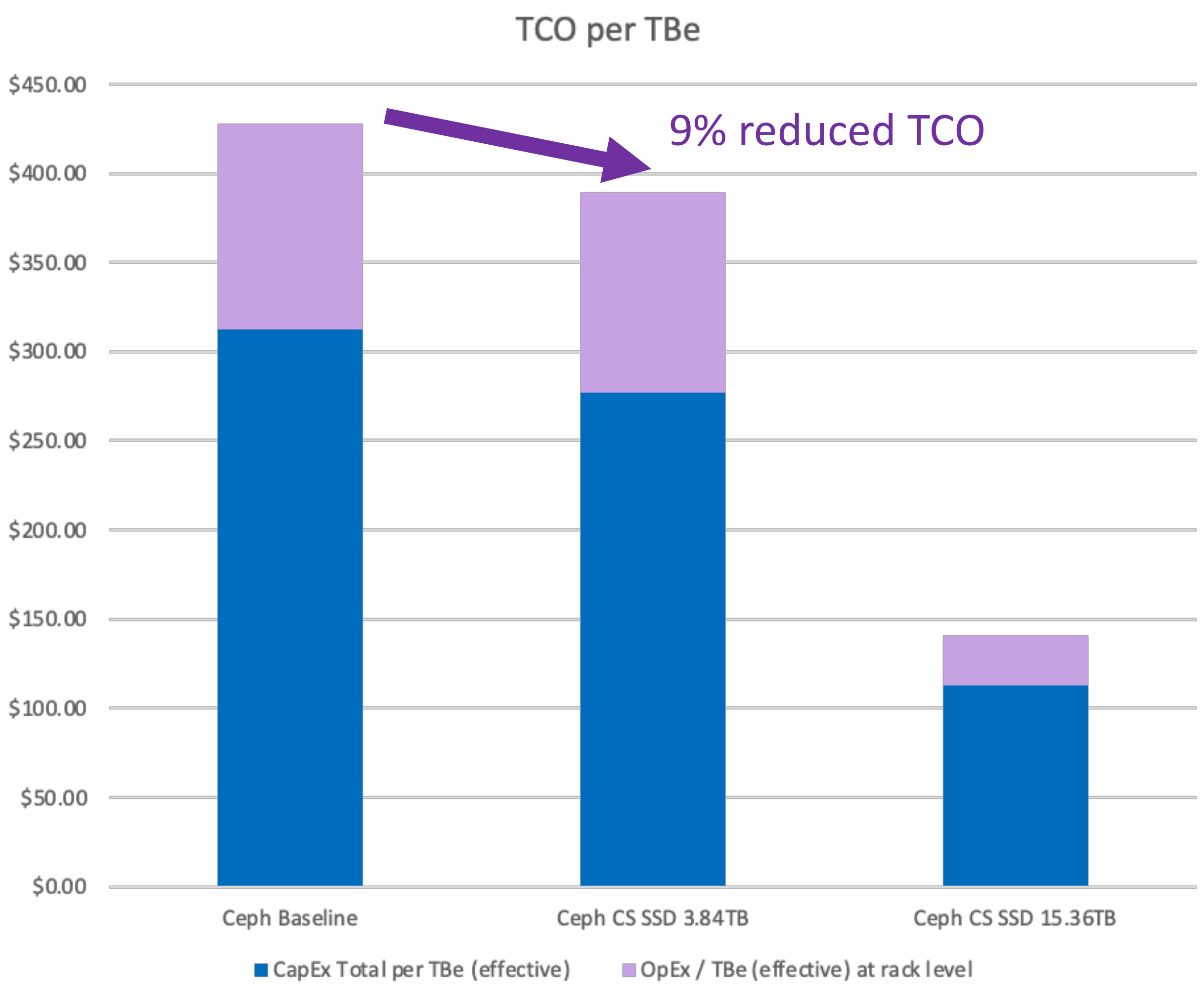


Figure 5. TCO Results with CPU sku reduction

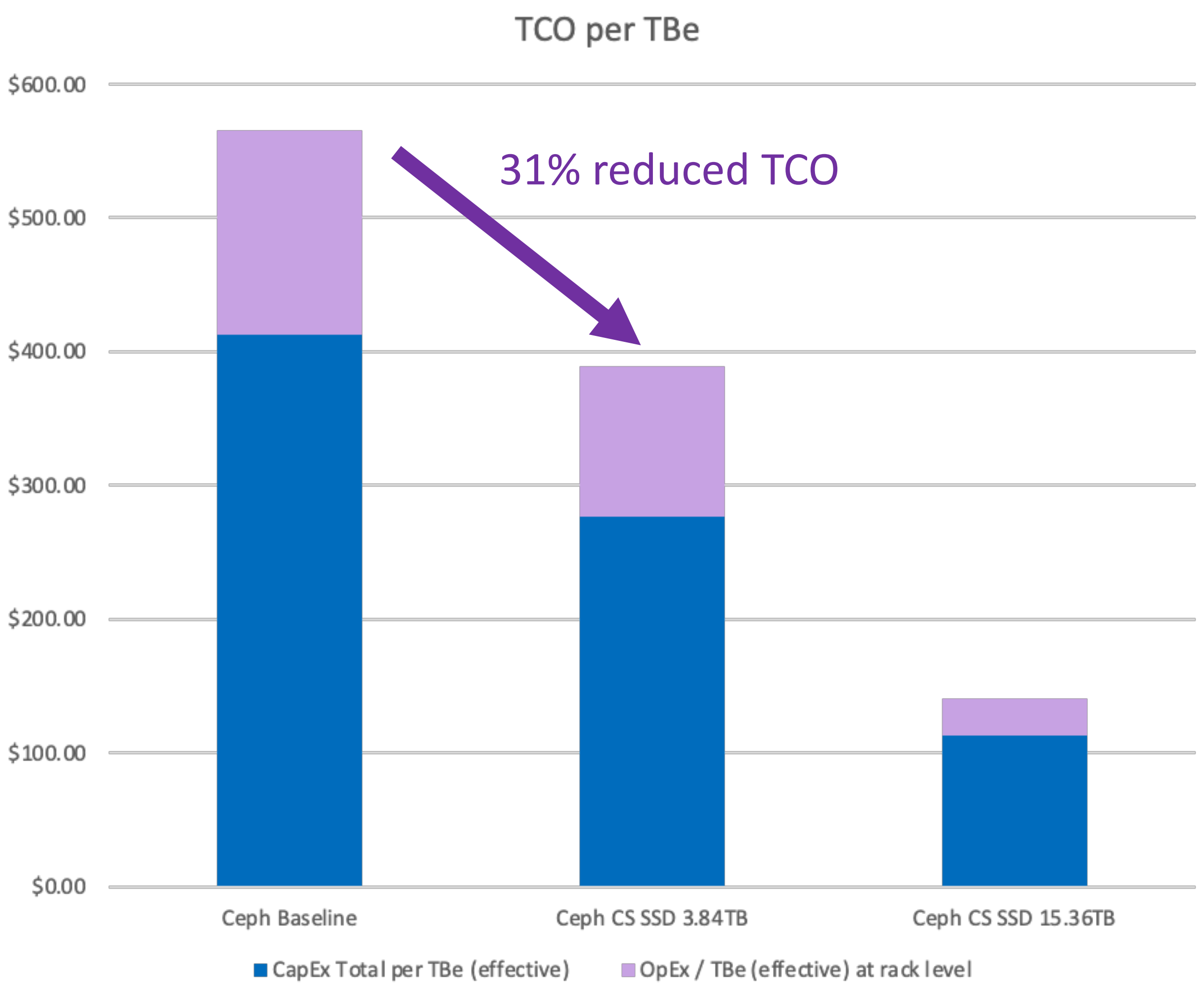


Figure 6. TCO Results with CPU sku reduction + performance

Backup

Software Config

FIO

```
[global]
ioengine=rbd
direct=1
bs=128k
iodepth=128
rw=write
runtime=1200
pool=test1
time_based
buffer_compress_percentage=60
group_reporting

[rbd_image_1]
rbdname=rbd_image_1

[rbd_image_2]
rbdname=rbd_image_2

...

[rbd_image_40]
rbdname=rbd_image_40
```

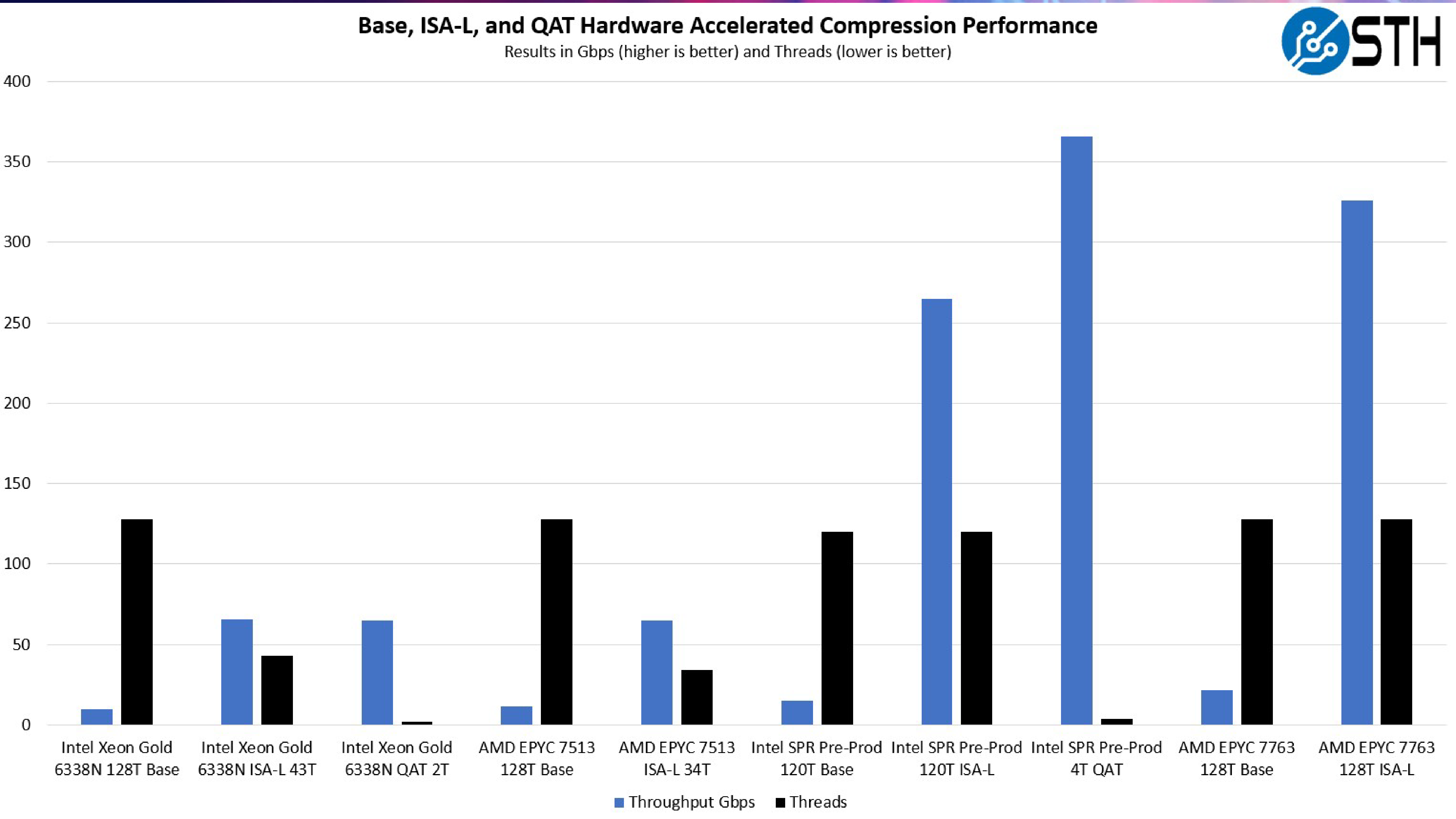
Ceph

```
bdev_enable_discard: True
mgr_max_pg_num_change: 32768
mon_max_pg_per_osd:
mon: 32768
osd: 32768
mon_osd_max_creating_pgs
mon: 32768
osd: 32768
mon_osd_max_initial_pgs
osd: 32768
```


TCO Model Config

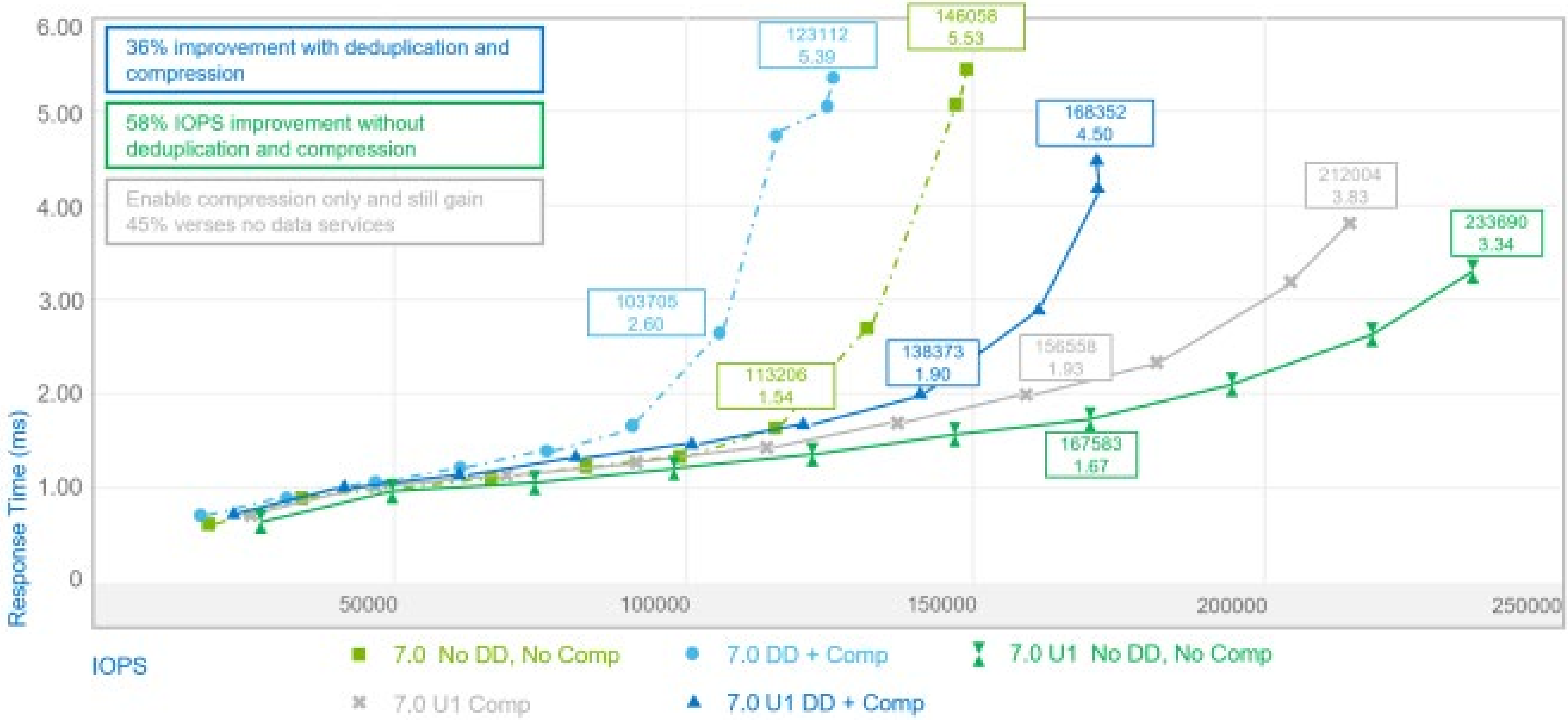
Drive Attributes			
Drive Name	Ceph Baseline	Ceph CS SSD 3.84TB	Ceph CS SSD 15.36TB
Capacity (GB)	3840	3840	15360
Power Active (W)	15	15	15
Power Idle (W)	3.5	3.5	3.5
AFR %	0.44%	0.44%	0.44%
ASP ($/GB)	$0.070	$0.070	$0.070
ASP ($)	$268.80	$268.80	$1,075.20
Workload			
Deployment Term (years)	5	5	5
Error encoding / replication	1.5	1.5	1.5
Performance multiplier	1.32	1	1
Capacity utilization	90%	90%	90%
Duty Cycle (active vs idle)	30%	30%	30%
Data Reduction Ratio (Compressi	45%	50%	50%
CapEx			
Capacity Per Rack (TB)	276.48	276.48	1105.92
CapEx Storage (Per Rack)	$19,354	$19,354	$77,414
CapEx Rack	$94,708	$72,500	$72,500
CapEx Total (Per Rack)	$114,062	$91,854	$149,914
CapEx Total per TB raw	$413	$332	$136
CapEx Total per TBe (effective)	$412.55	$276.85	$112.96
CapEx per TBe per month of depl	$6.88	$4.61	$1.88
OpEx			
Power Storage (Active/Idle/Duty	500.4	500.4	500.4
Power Max (Rack Limit, W)	5230	4430	4430
OpEx Storage ($ deployment tern	$3,068	$3,068	$3,068
OpEx Rack ($ deplyment term)	$25,448	$20,542	$20,542
OpEx Drive Failures / Replacemer	$158	$158	$158
OpEx Rackspace / Cooling	$13,500	$13,500	$13,500
OpEx Total per Rack	$42,175	$37,269	$37,269
OpEx / TB raw at rack level	$152.54	$134.80	$33.70
OpEx / TBe (effective) at rack leve	$152.54	$112.33	$28.08
OpEx / TBe (effective) per month	$2.54	$1.87	$0.47
Total Cost of Ownership			
TCO (OpEx + CapEx) per TBe	$565.09	$389.19	$141.05
TCO per TBe per month of deploy	$9.42	$6.49	$2.35
Capacity per rack (PB)	0.3	0.3	1.1
Effective Capacity per rack (PBe)	0.3	0.3	1.3

Rack and System Attributes	Baseline	CSD
Data center cost / rack space per m	$ 225	$ 225
Rack cost	$ -	$ -
Server cost	$ 11,651	$ 8,875
JBOD cost	$ -	$ -
Switch cost	$ 1,500	$ 1,500
Server Power (W)	500	$ 400
JBOD Power (W)	0	0
Switch Power (W)	150	150
Server RU	1	1
JBOD RU	0	0
Switch RU	1	1
Drives per server	12	12
Drives per JBOD	0	0
Servers per rack	6	6
JBODs per rack	0	0
Utility server per rack (no storage)	2	2
Totals		
Cost, CapEx per Rack	$ 94,708	$ 72,500
Power, W per Rack	4150	3350
Size, RU per Rack	9	9
Fixed		
Electricity Cost per kWhr	$ 0.14	$ 0.14
Data Center PUE	1	1
drive replacement cost	$ 100.00	$ 100.00


Base, ISA-L, and QAT Hardware Accelerated Compression Performance
Results in Gbps (higher is better) and Threads (lower is better)
STH
400
350
300
250
200
150
100
50
0
Intel Xeon Gold 6338N 128T Base
Intel Xeon Gold 6338N ISA-L 43T
Intel Xeon Gold 6338N QAT 2T
AMD EPYC 7513 128T Base
AMD EPYC 7513 ISA-L 34T
Intel SPR Pre-Prod 120T Base
Intel SPR Pre-Prod 120T ISA-L
Intel SPR Pre-Prod 4T QAT
AMD EPYC 7763 128T Base
AMD EPYC 7763 128T ISA-L
Throughput Gbps
Threads

Massive improvements for database workloads (RDBMS)

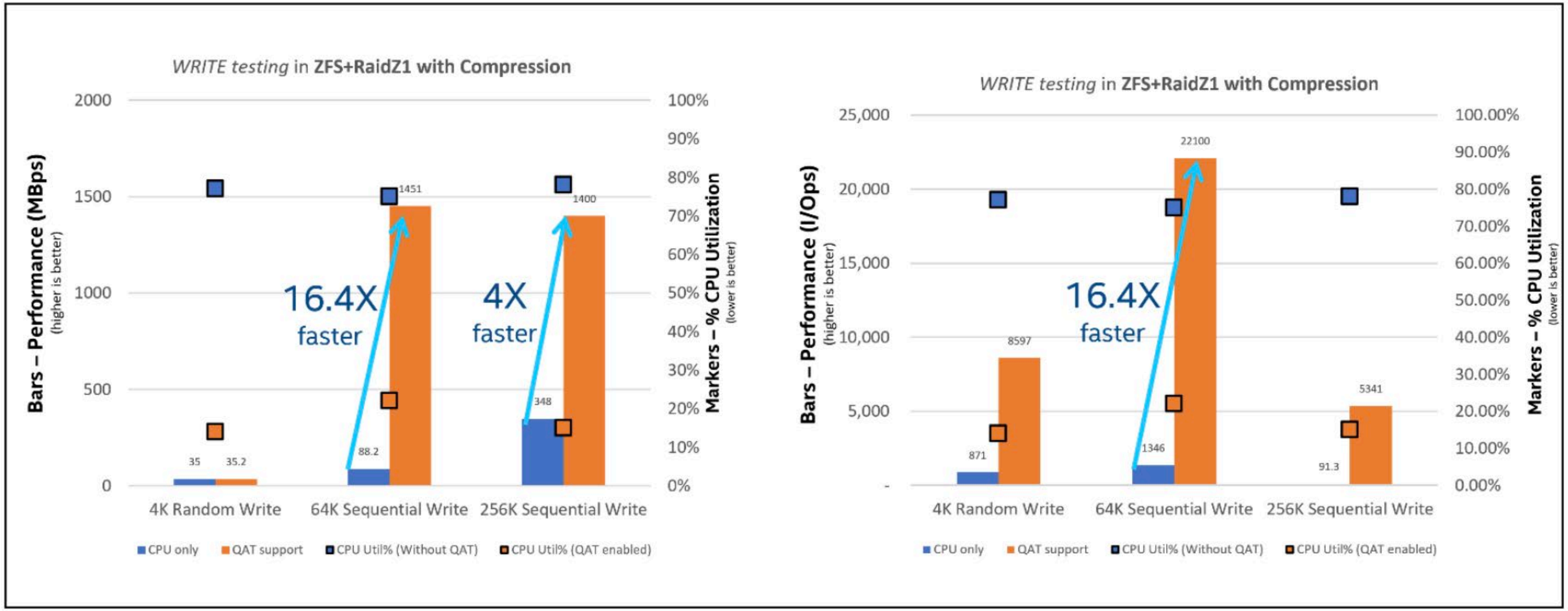
Smoother and more predictable latency



https://infohub.delltechnologies.com/l/harnessing-the-performance-of-dell-emc-vxrail-7-0-100-a-lab-based-performance-analysis/conclusion-3-vsan-compression-only-is-nearly-penalty-free

Example Performance Comparison

[1]Using Intel® QuickAssist Adapter C62x Series for Hardware-based Data Compression, Free CPU Computing Resources, Improve Data Compression and IO Performance in NAS ZFS's SW RaidZ1



Performance comparison of NAS SW-RAID with ZFS compression based on Intel® Quick Assist Technology

https://www.intel.com/content/www/us/en/architecture-and-technology/wanyou-high-density-storage-nas-solution.html

