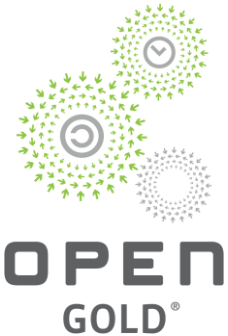
OPEN POSSIBILITIES.

Challenges and Effects of EDSFF-based NVMe-oF Storage Solution

OCP GLOBAL SUMMIT
NOVEMBER 9-10, 2021

Challenges and Opportunities of EDSFF based Storage Solution

Jungsoo Kim, System Architect, Samsung Electronics
Duckho Bae, Software Architect, Samsung Electronics



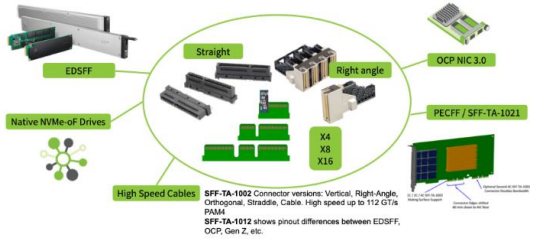
Agenda

- EDSFF Benefits
- Poseidon V1 – E1.S Reference System
- Poseidon V2 – E3.X Reference System
- Poseidon Storage OS – NVMe-oF Solution

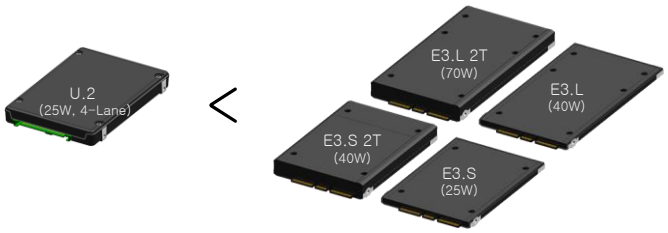
EDSFF – Benefits

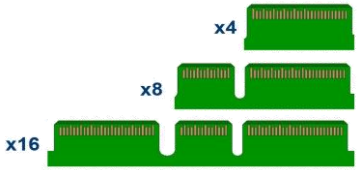
STORAGE

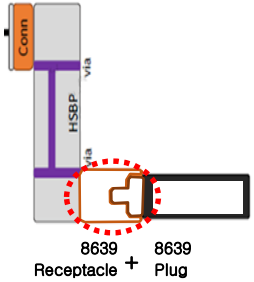
- Designed to overcome conventional device limitations
- Improves thermals, power, and scalability
- Various power options - 25W, 40W, 70W
- Various PCIe lane options - x2, x4, x8, x16
- Advantageous for high-speed interface (< PCIe Gen5)
- Built in LEDs, carrier-less design



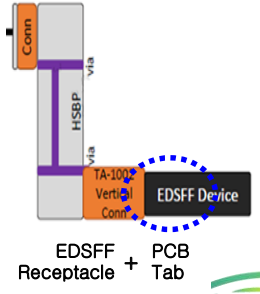
Source: https://www.snia.org/forums/cmsi/knowledge/formfactors







U.2



EDSFF

OPEN POSSIBILITIES.

OCP GLOBAL SUMMIT
NOVEMBER 9-10, 2021

EDSFF Reference System

STORAGE

- E1 and E3 based system can increase the performance and density
- Have more flexibilities than a traditional system

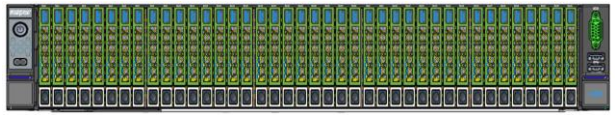
U.2 SSD x 10ea

VS.

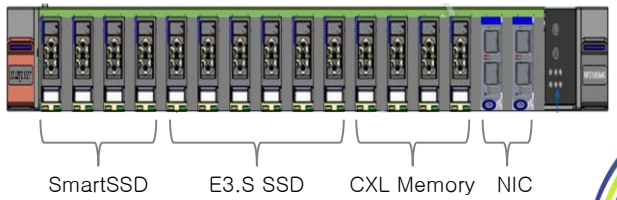
E1.S SSD x 32ea

U.2 SSD x 24ea

VS.



E3.S SSD x 40ea



OPEN POSSIBILITIES.

OCP GLOBAL SUMMIT
NOVEMBER 9-10, 2021

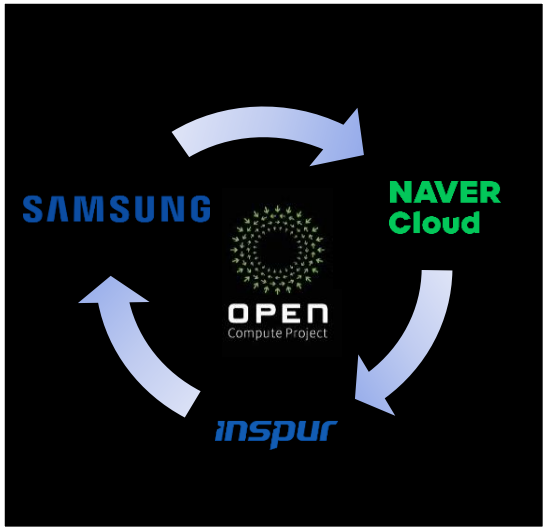
Poseidon Project

- OCP based industrial collaboration b/w “Component Vendor↔System Vendor ↔Data Center”
- Open-source HW & SW project to expand NVMe eco-system

Poseidon SW
Storage OS for NVMe-oF

GitHub

Poseidon HW
E1.S/E3.S Ref. System



OPEN POSSIBILITIES.

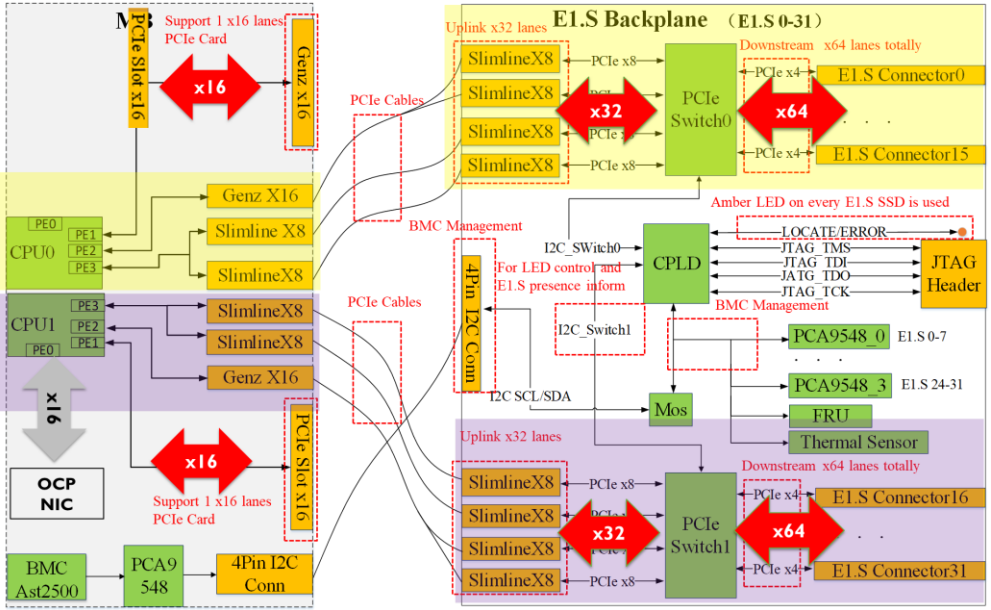
OCP GLOBAL SUMMIT
NOVEMBER 9-10, 2021

E1.S Reference System – PSD V1

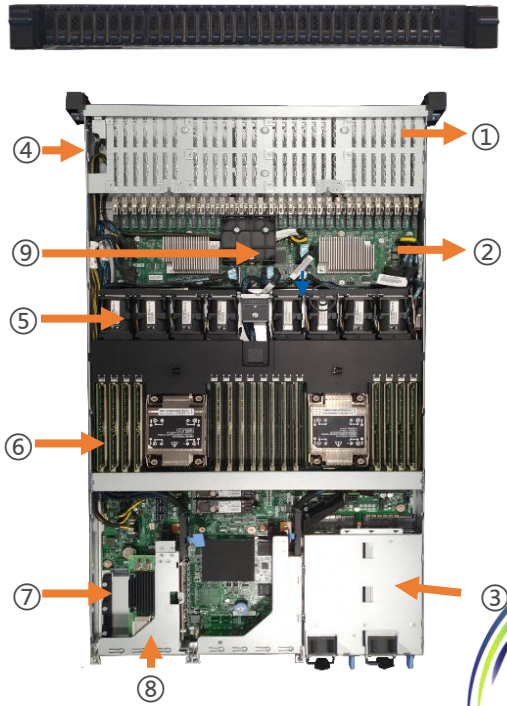
①	: E1.S SSD (5.9/8.01/9.5)	32ea
②	: 32 E1.S BP	1
③	: PSU	2ea
④	: IO Module	1ea
⑤	: FAN	8ea
⑥	: MB	1ea
⑦	: FHHL Card	2ea
⑧	: OCP NIC V3	1ea
⑨	: NVDIMM Power Module	2ea

Motherboard

Backplane



Front View

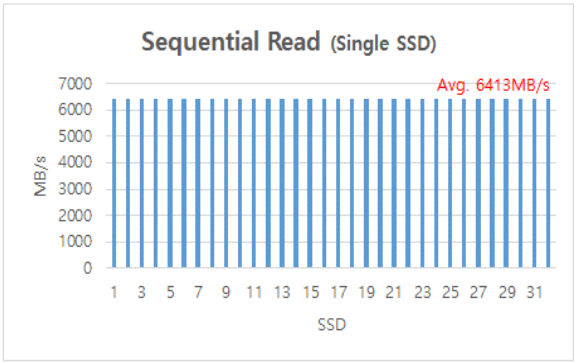


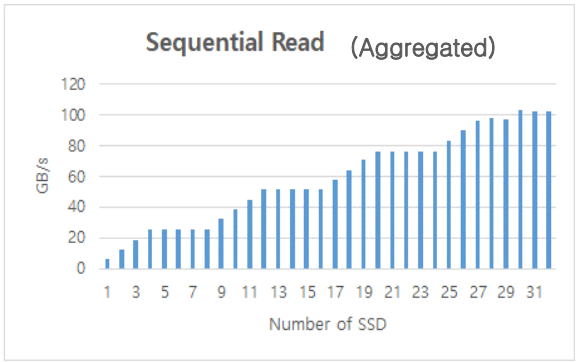
OPEN POSSIBILITIES.

OCP GLOBAL SUMMIT
NOVEMBER 9-10, 2021

IO Performance – Single vs Aggregated

- Each 32 SSDs shows stable IO performance around 6.4GB/s
- Aggregated IO performance achieved > 102GB/s
- Lanes b/w CPU↔PCIe Switch (x16) are saturated with 4 SSDs





*Theoretical B/W limit: 128GB/s

- Samsung PM9A3 Specification

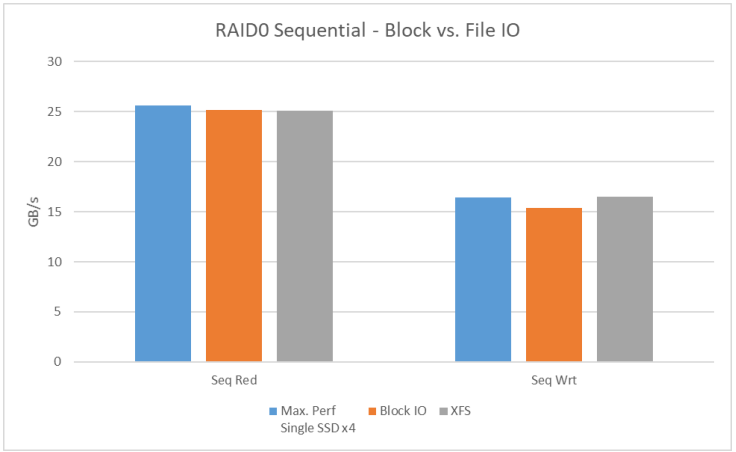
Form factor	E1.S
Capacity	960 GB, 1.92 TB, 3.82 TB, 7.68 TB
Sequential read	Up to 6,500 MB/s
Sequential write	Up to 3,200 MB/s
Random read	Up to 900,000 IOPS
Random write	Up to 150,000 IOPS
Physical Dimensions	31.5 x 111.49 x 5.9 mm
Power consumption	Read: <= 9.7W, Write: <= 11.7W
Host interface	PCIe Gen 4 x4

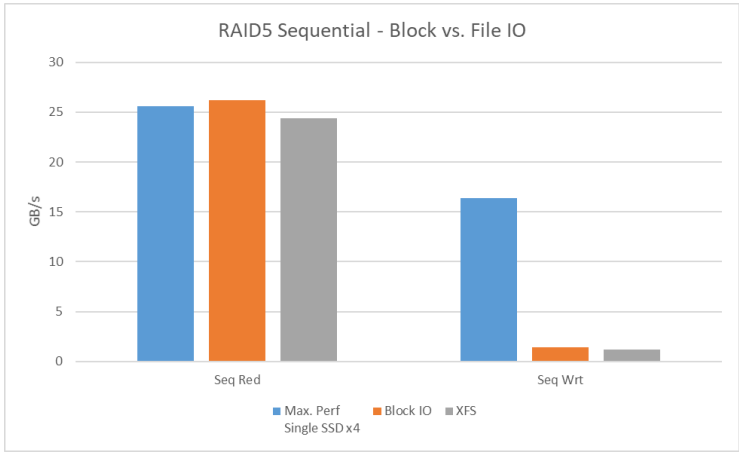
OPEN POSSIBILITIES.

OCP GLOBAL SUMMIT
NOVEMBER 9-10, 2021

IO Performance – Raid 0&5

- Compared Raid 0 & 5 performance of 4 x NVMe SSDs
- Observed significant performance degradation on Raid 5 write operation
- Further R&D would be need for S/W Raid on high density system





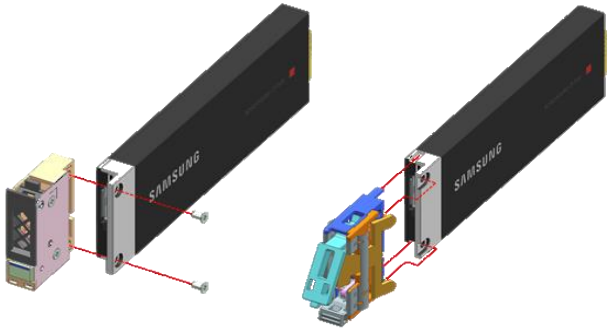
OS: Ubuntu 20.04.2 LTS
Kernel: Linux 5.4.0-65-generic
Num. of threads : 112 (56 * 2)
Model name : Intel Xeon Gold 6330N CPU @ 2.20GHz
fio: fio-3.16

OPEN POSSIBILITIES.

E1.S Tool-less Design

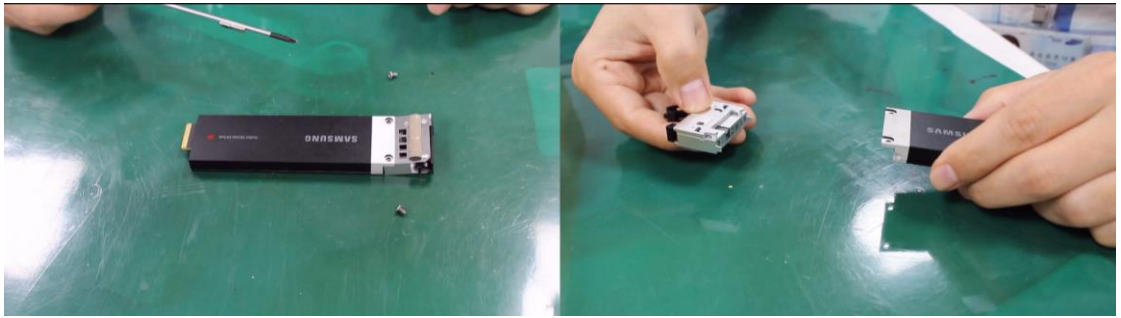
STORAGE

- DC customers want to improve serviceability in their datacenter by removing the screws
- E1.S + extension kit with screws are the only option in the market, and we developed the innovate new tool-less ext. kit design to satisfy the requirements



Current Design
– Screw type

Tool-less Design
– Clip type



Screw Type (40sec)

Tool-less Type (4sec)

Compatible w/ a screw type extension kit!

OPEN POSSIBILITIES.

E3 Reference System – PSD V2

STORAGE

- Designed to maximize the benefits of E3.x form factor
- Can configure the system according to application's needs

[High Density Storage]

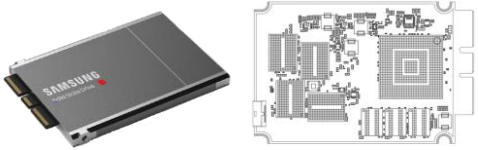
Type1: 24x E3.S 1T (PCIe x4)

[Memory Cache / Cloud]

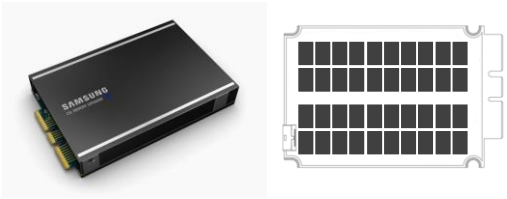
Type2: 8x E3.S 1T + 8x E3.S 2T (PCIe x8)

[AI / Video Transcoding]

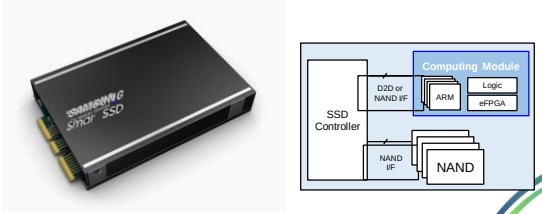
[E3.S Gen5 SSD]



[CXL Memory Expander]



[Smart SSD]



*Above E3 SmartSSD picture is for illustration purpose only

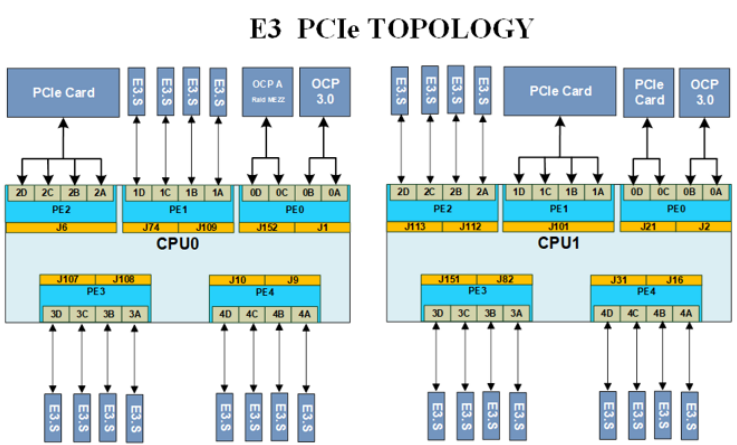
OPEN POSSIBILITIES.

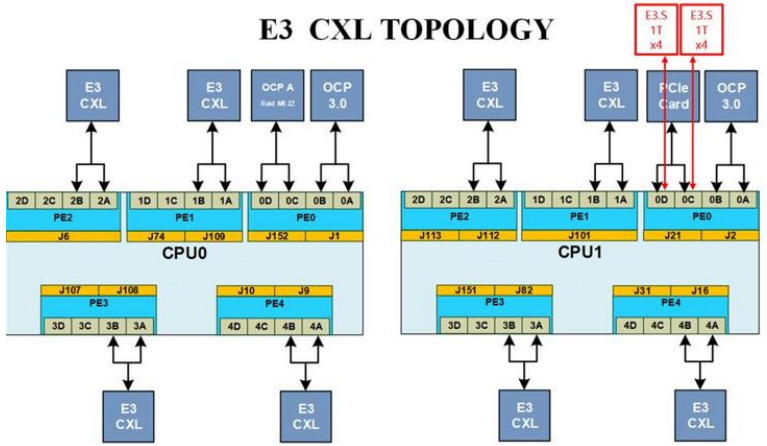
OCP GLOBAL SUMMIT
NOVEMBER 9-10, 2021

Poseidon V2 – PCIe/CXL Topology

STORAGE

- PCIe & CXL topology varies depending on the system mode



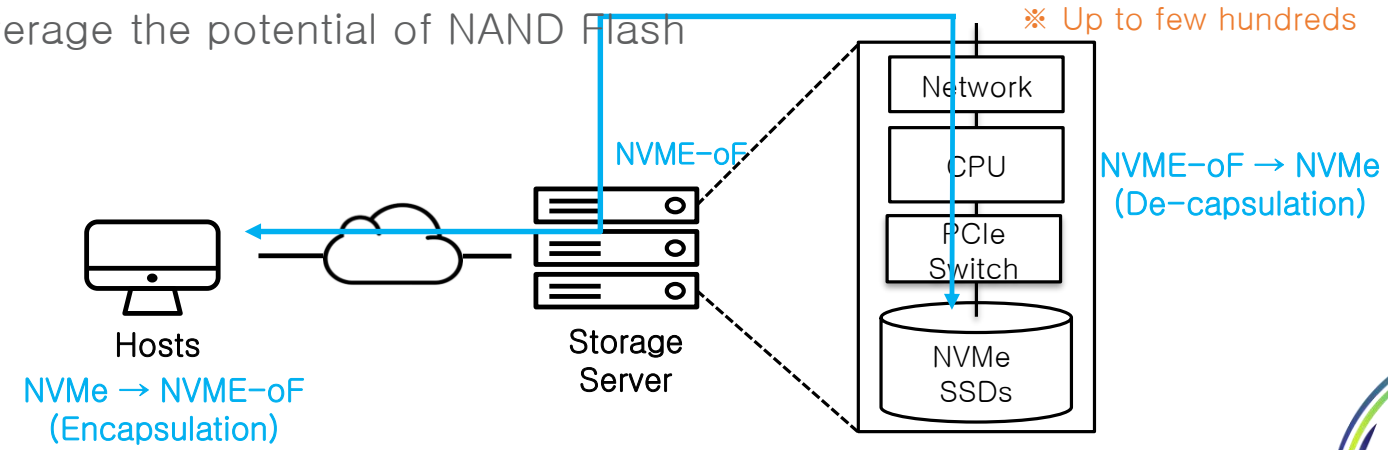


OPEN POSSIBILITIES.

OCP GLOBAL SUMMIT
NOVEMBER 9-10, 2021

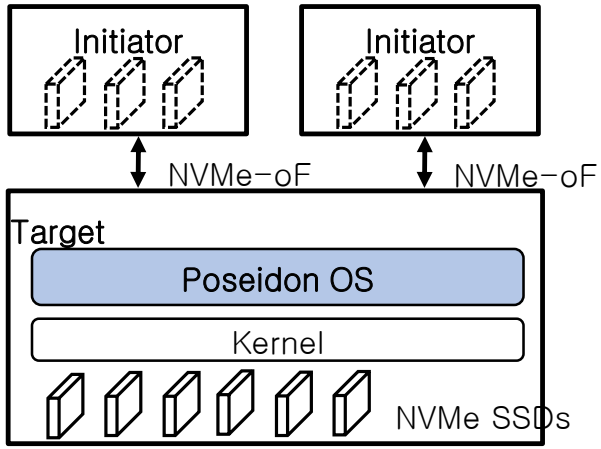
PoseidonOS

- Open-Source NVMe-oF Solution for Disaggregated Datacenter Storage Solution
  - Develop NVMe Technology and Eco-system
  - Improve Datacenter Storage Efficiency and Performance
  - Leverage the potential of NAND Flash

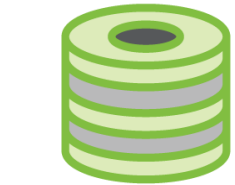


PoseidonOS Features

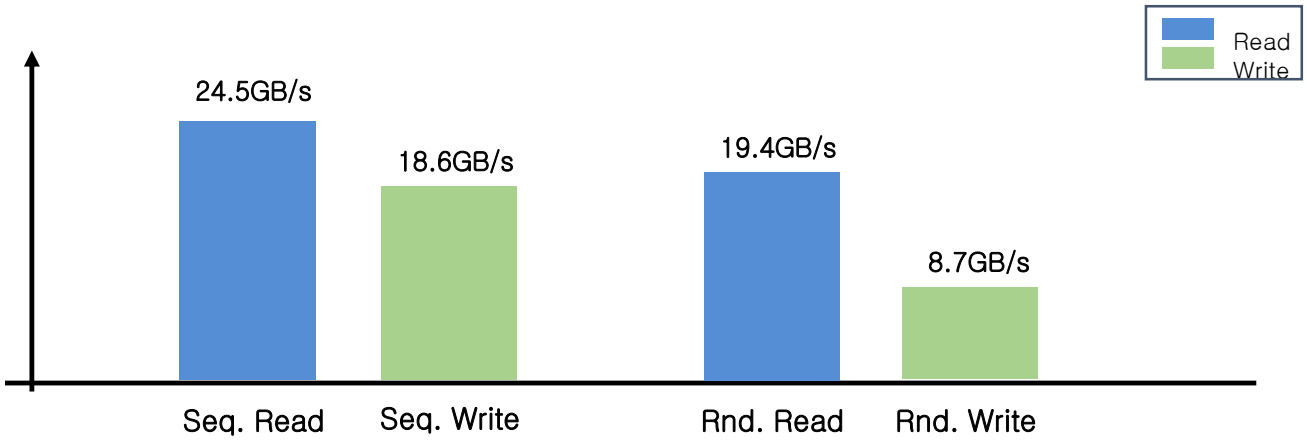
- User-space storage OS for NVMe-oF
- Provide PCIe Gen4 < performance via network
  - Up to 200GbE
- Support valuable storage features
  - NUMA-Aware, Volume Mgmt, Perf Throttling, SW RAID, …
- Easily integrate with upper orchestration layers
  - RESTful, CSI, …



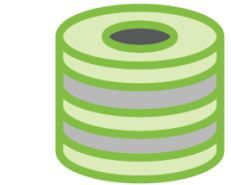
Performance Numbers

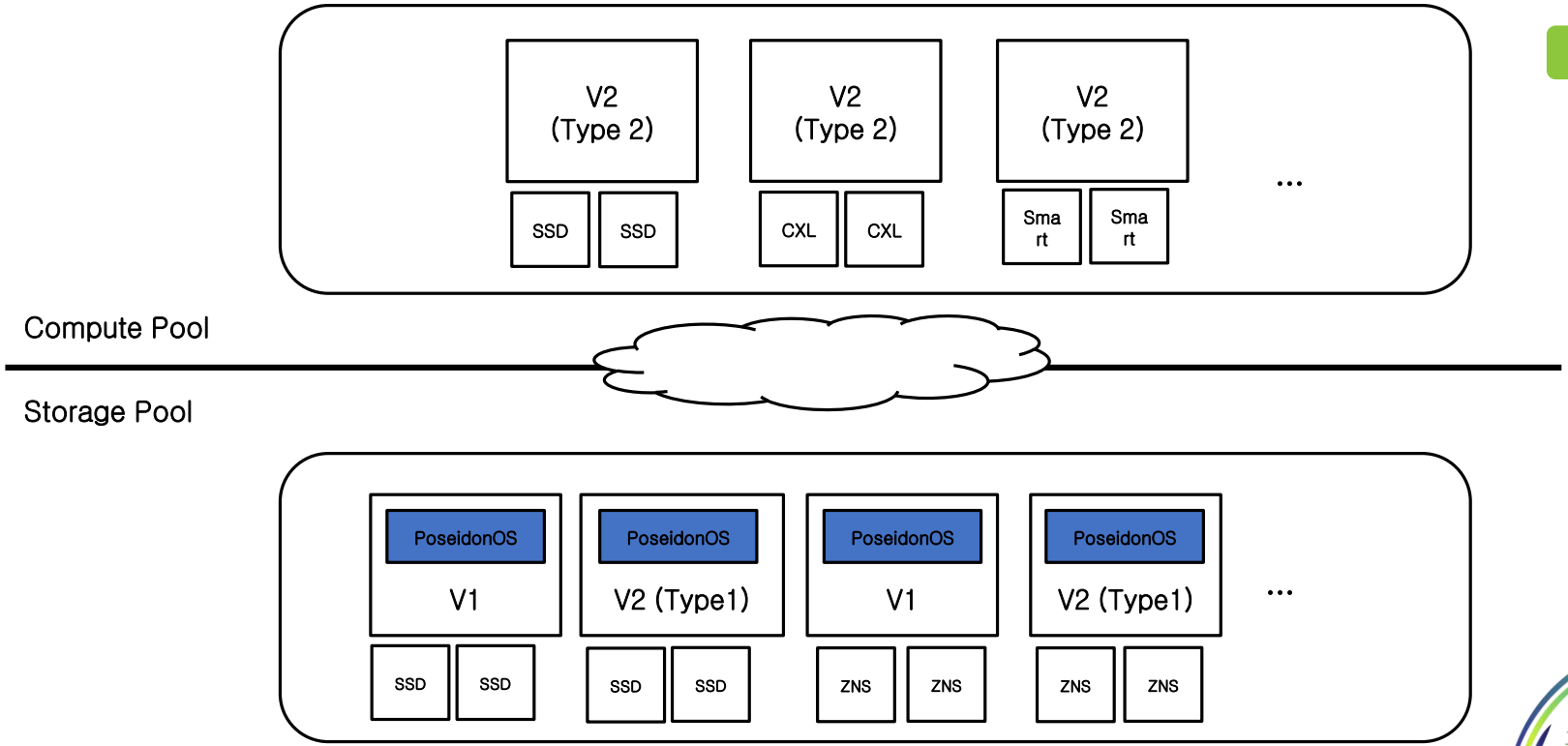


- Achieved up to 200GbE Performance via NVMe/TCP
- Random Write has room for improvement



Poseidon in Datacenter

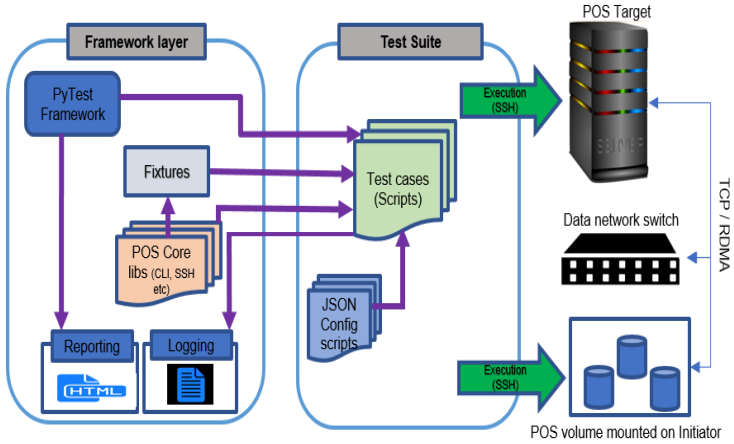




Trident : Test Tool for Poseidon OS

- Helps open source users quickly setup and explore all features of Poseidon OS (POS)
- Planned to be open sourced to compliment the user documentation provided with POS
- Enables community users to quickly validate their lab setup
- Along with POS, can also be deployed on Qemu/KVM VMs and emulated NVMe devices
- Includes Test cases that are light-weight and intuitive as they exploit features of pytest framework like fixtures and parametrization.

Test Framework and Lab



OPEN POSSIBILITIES.

Future Work

- Poseidon V2 contribution in 2022
- Tool-less design upgrade and datacenter adoption
- Support innovative devices (ex. ZNS, QLC)
- Support more features/provide developers toolkits
- Enable PCIe Gen5 performance
- Available at Github

  https://github.com/poseidonos/poseidonos

Call to Action

- Timeline for Contribution Availability
  - PM9A3 OCP SSD (`21 Q1) – OCP Inspired
  - Poseidon V1 (`21 Q3) – OCP Inspired
  - Tool-less SSD Design (`21 Q3) – OCP Inspired
  - Poseidon V2 (Target in `22 Q3)
- Timeline for Product/Facility Availability
- Link to Contribution DB/OCP Marketplace
- Samsung PM9A3 SSD complies with the Open Compute Project (OCP) NVMe Cloud SSD (https://www.samsung.com/semiconductor/ssd/pm9a3/)
- Poseidon OS github: https://github.com/poseidonos/poseidonos

Thank you!

NOVEMBER 9-10, 2021