TOSHIBA
Leading Innovation

Enhancing NVMe-oF Capabilities Using Storage Abstraction

SNIA Storage Developer Conference

Yaron Klein and Verly Gafni-Hoek

February 2018

Outline

- NVMe SSD Overview
- NVMe-oF Overview
- A Typical NVMe-oF System Architecture
- Software Storage Abstraction Overview
- Abstraction Configurations for NVMe-oF Subsystems

NVMe SSD Overview

Faster	• IOPS and throughput; applications perform better
Quicker	• Low latency, being directly connected to the CPU
Design	• Command set created from the ground up for SSDs • NVMeoF as native fabric
Consistency	• Better performance consistency than SAS or SATA
Flexibility	• More form factors, power, # lanes, connectivity, client and enterprise

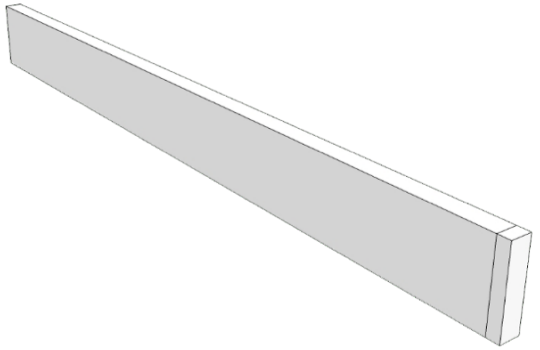
2.5” SFF

Add In Card

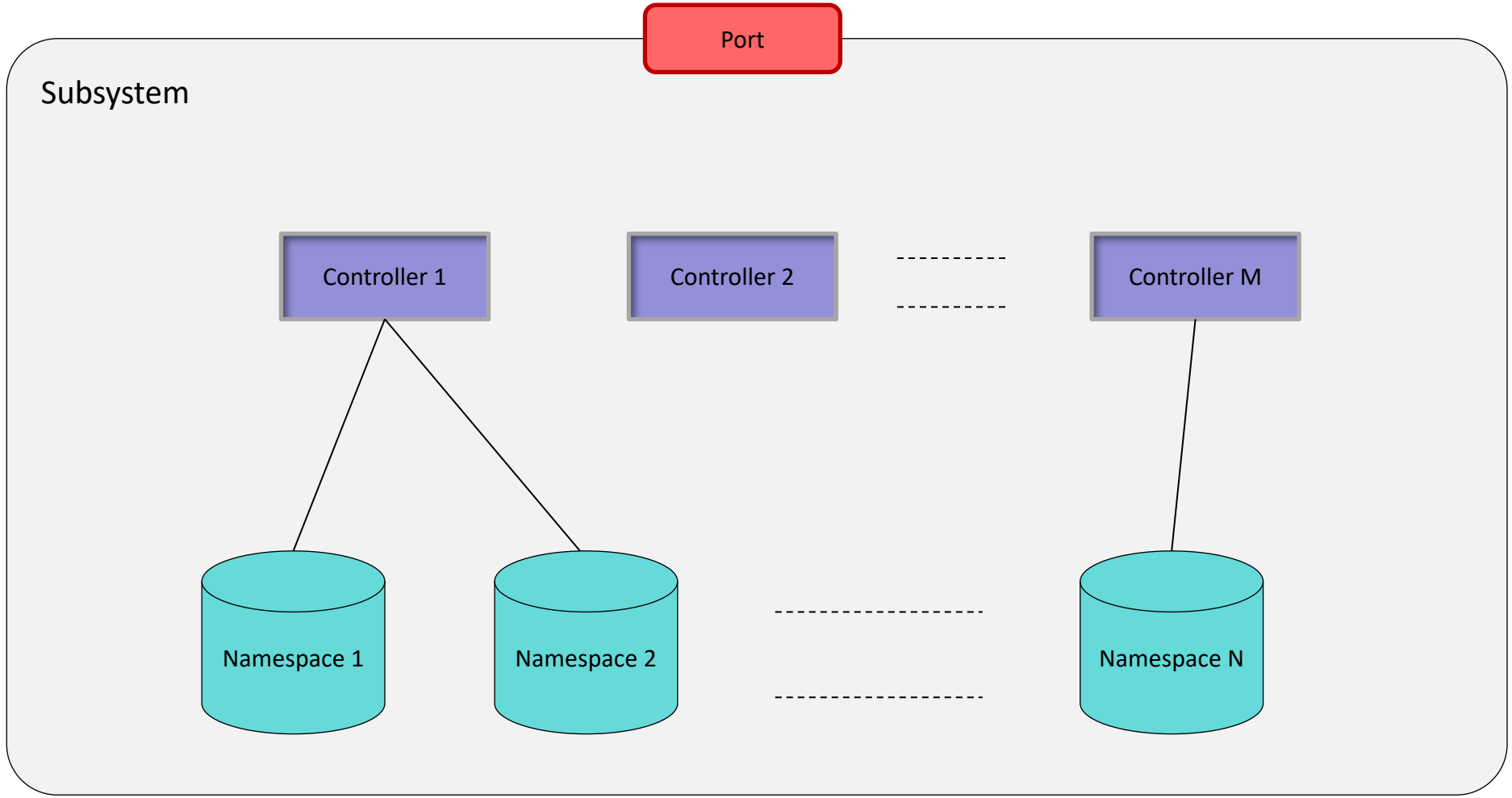
FHFL/FHHL/HHHL

M.2

Ruler/NGSFF



Subsystem - Controller - Namespace



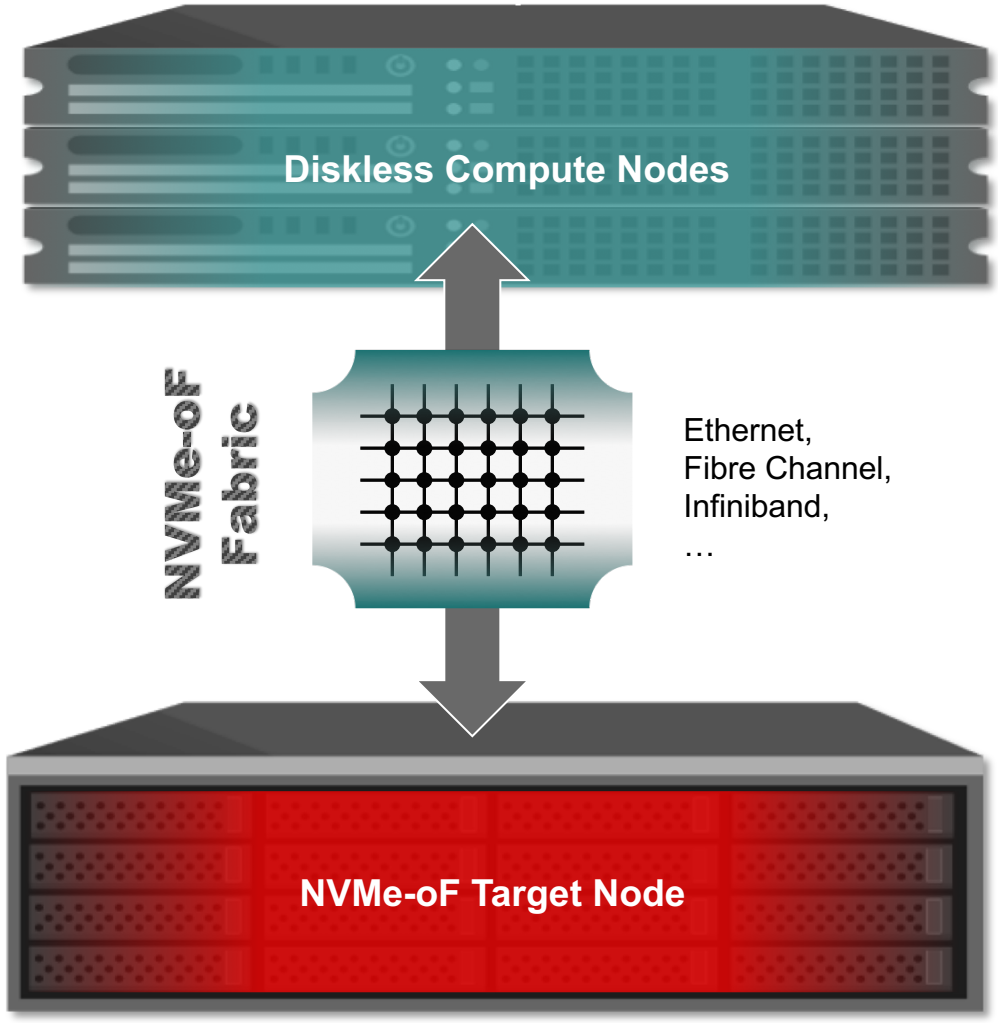
An NVM subsystem includes one or more controllers, one or more namespaces, one or more ports, a non-volatile memory storage medium, and an interface between the controller(s) and non-volatile memory storage medium.

What is NVMe over Fabrics (NVMeoF)?

Connects compute nodes to NVMe storage across the datacenter network

Preserves the performance and low latency of *native* NVMe

Uses remote direct memory access (RDMA), with bindings for several transport protocols

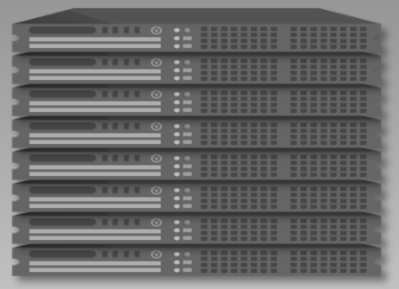


Why is NVMeoF important to the Datacenter?

Direct-attached SSDs

- Storage local to each compute node
- "One-size fits all" – leads to islands of stranded storage or compute power

Pooled Storage with NVMe-oF



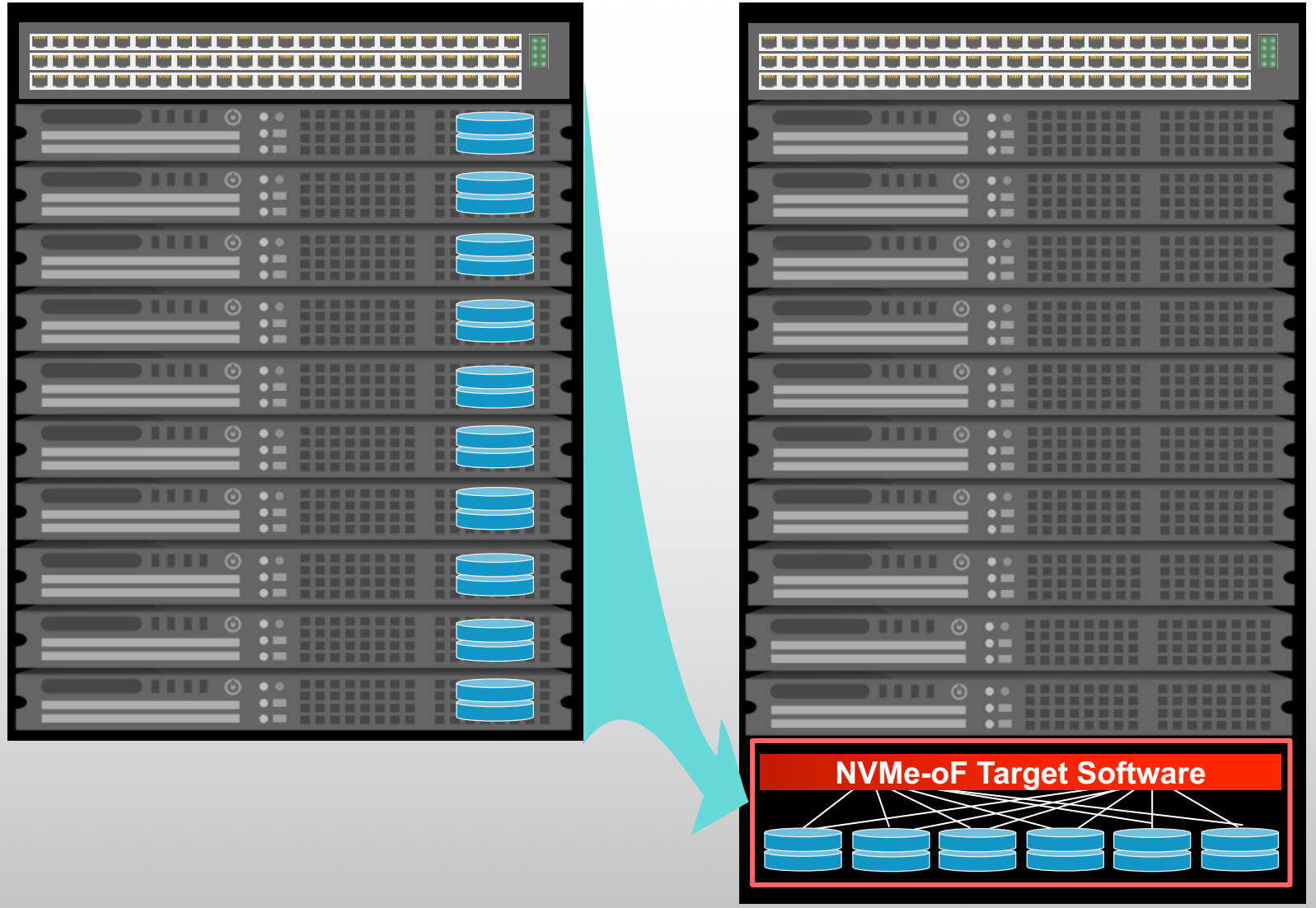
Diskless compute nodes

NVMe-oF storage node

- Disaggregates and shares fast NVMe storage at full performance
- Enables optimal allocation of storage capacity and performance to each node

Each job gets "just the right amount" of high performance, low latency storage

Super Fast Block Storage, Disaggregated and Abstracted



FAST
Near local NVMe throughput and latency

EFFICIENT
Allocate required capacity and grow on-demand. No stranded storage

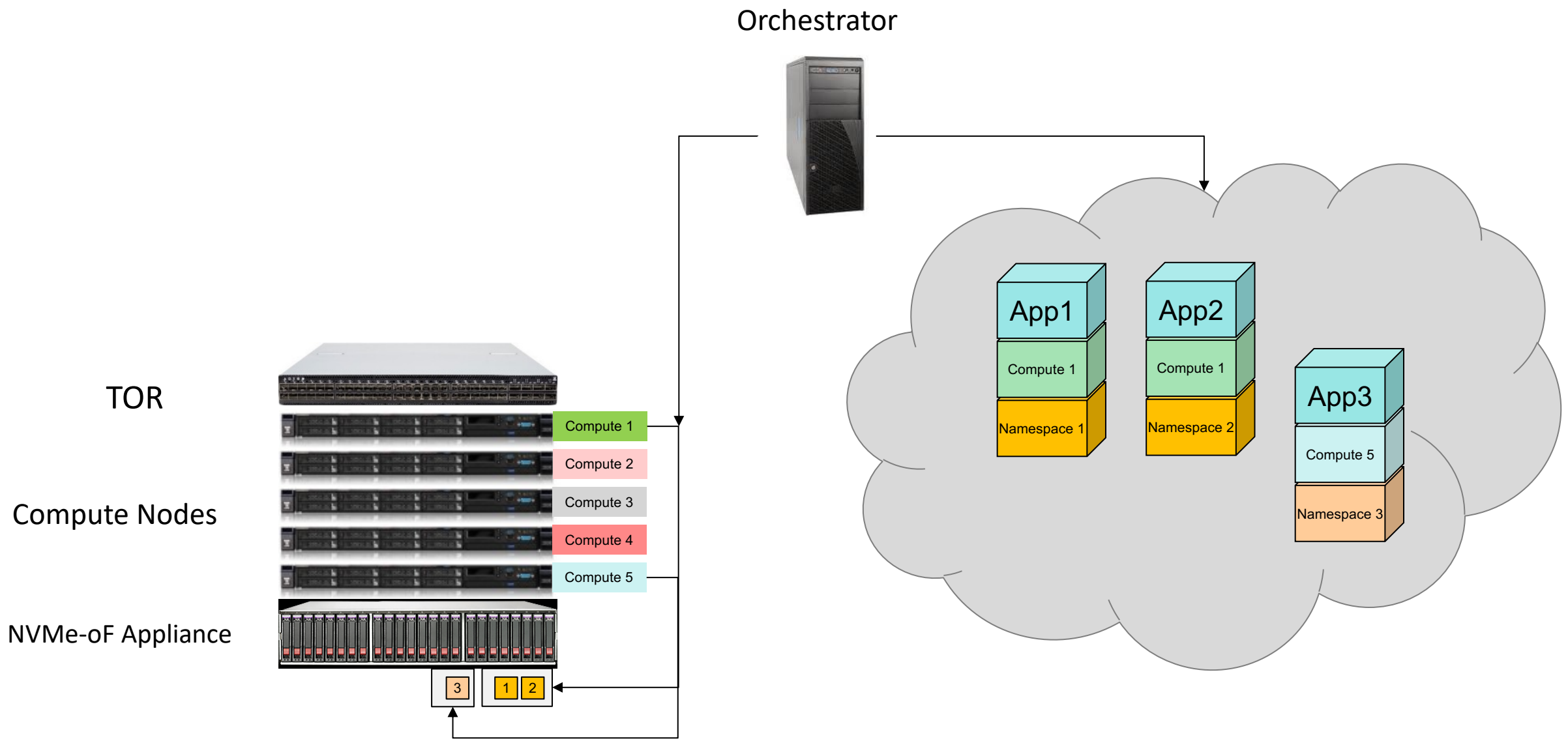
OPTIMAL SSD UTILIZATION
Share high capacity SSDs between multiple servers for optimal Watt/TB & optimal rack utilization

FLEXIBLE
Namespace abstraction hides physical drive complexity

MANAGEABLE
Connectors to fast-evolving orchestration, provisioning and telemetry tools

SAFE
Balanced SSD wear management

Orchestrating Virtual Environment

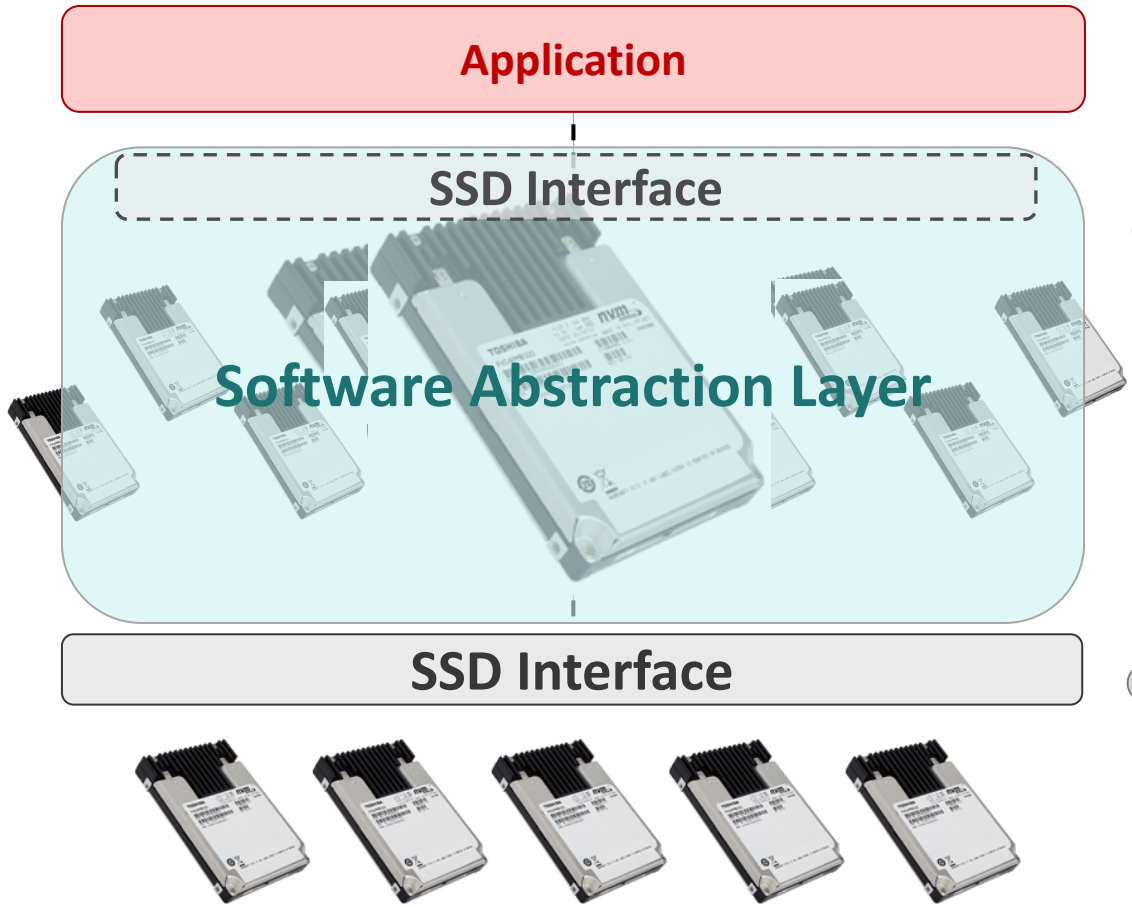


What is Storage Software Abstraction?

Add additional SW layer between the application and the SSD driver:

This layer decides which commands to implement or manipulate and which commands are forwarded to the SSD driver as is

This layer exposes SSDs in a completely different manner than the underlying physical SSDs



Application World Frontend

10 X 2TB SSD,
2 X 2TB SSDs

Physical World Backend

5 X 4 TB SSDs

NVMe-oF Storage Abstraction Advantages

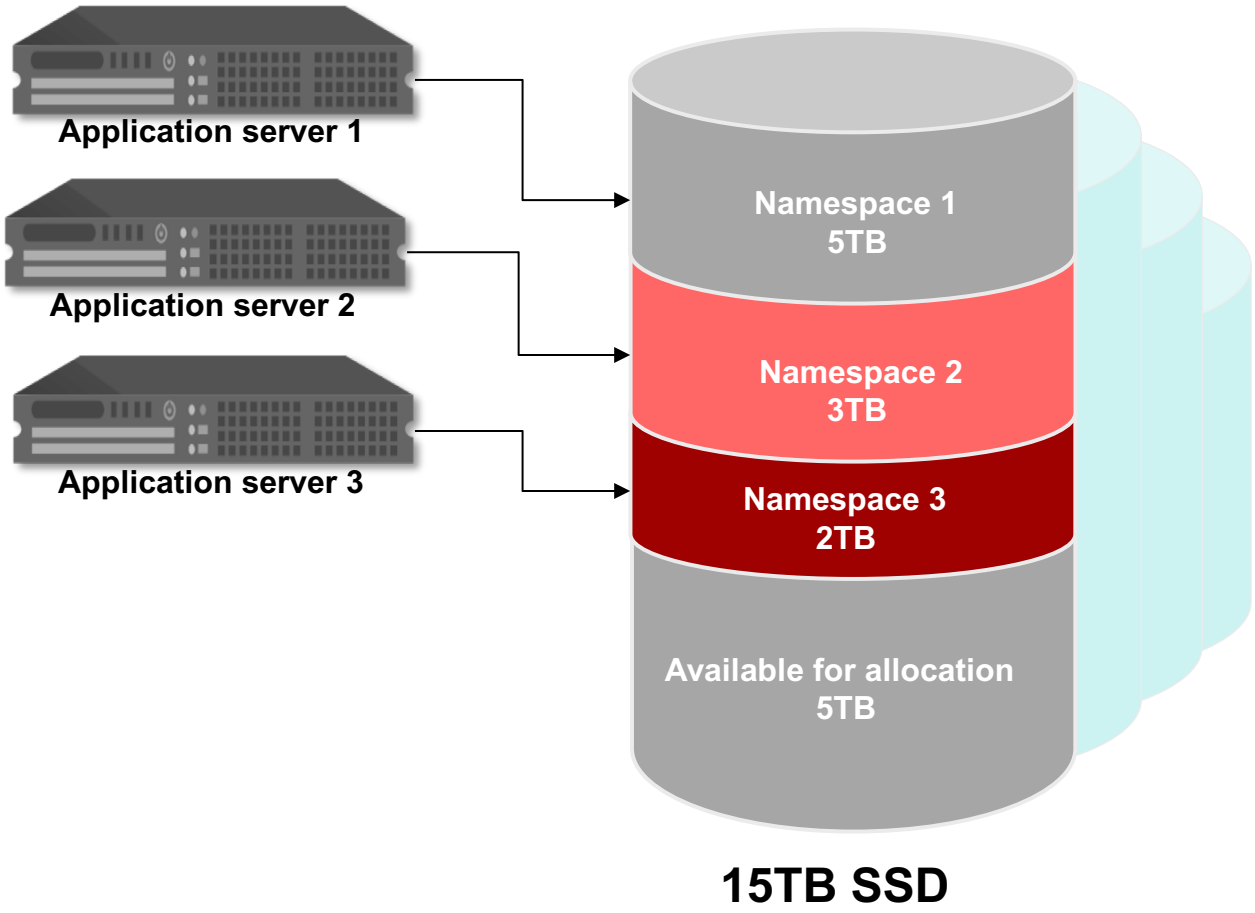
Flexible capacity allocation for different compute nodes enables optimizing large SSD's utilization

The largest capacity SSDs improve Watt/TB and deliver better space utilization in the rack

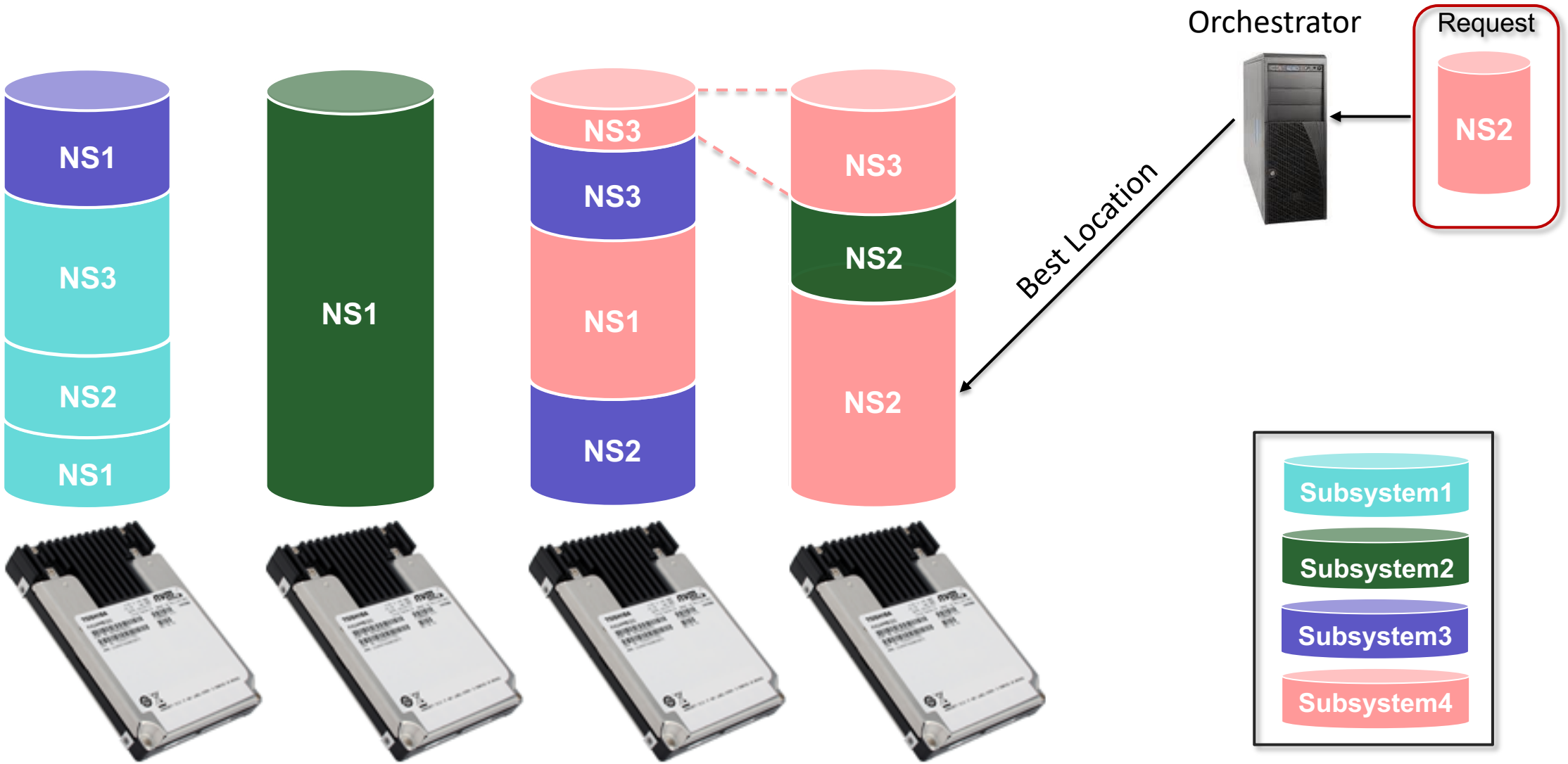
Ability to expose different storage layouts to different hosts upon demand

Can implement storage features that are not implemented by the SSDs FW

Increase namespaces per SSD:
Unlimited provisioning flexibility



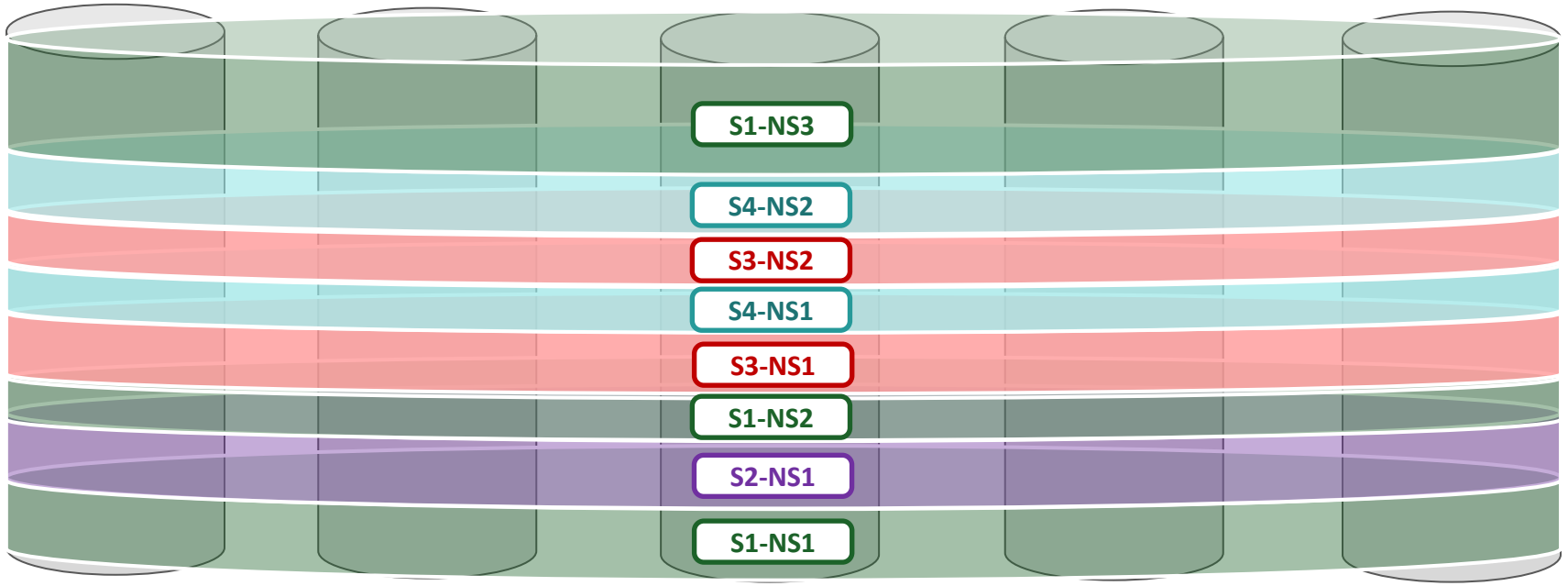
Abstracted Storage Pool

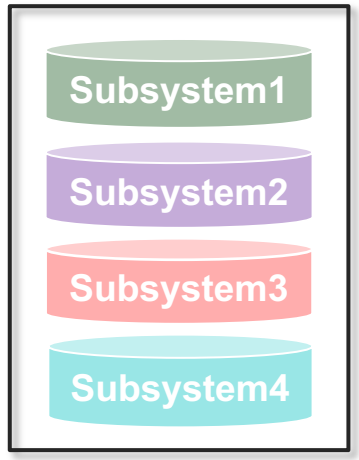


Abstracted Storage Pool Advantages

- Capacity utilization is more efficient
- Blast radius is minimized
- Practically no limitation on namespace size and granularity
- Supports smart placement of namespaces across the SSDs

Storage Abstraction Configurations – Striped Configuration



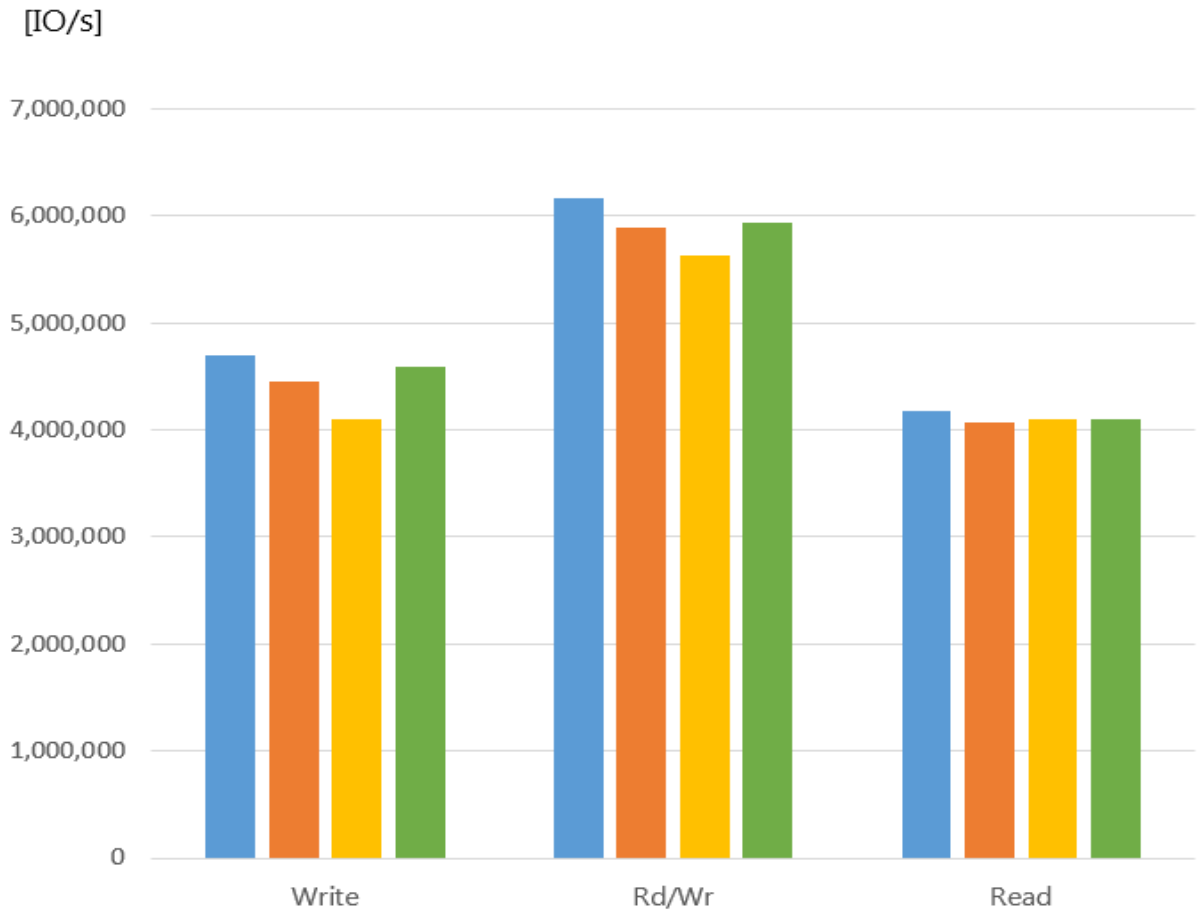


Striped Abstraction Advantages

Increased performance – IOs are sent in parallel to and from all SSDs:
Higher IO/s and lower latency

Balanced wear leveling by design

Abstraction doesn’t mean performance degradation



IO/s on different abstraction configurations, 4K command size

24 Directly Attached NVMe SSDs
24 NVMe-oF Subsystems w/o Abstraction
2 Striped Nvme-oF Subsystems over 12 NVMe SSDs each
24 Abstracted NVMe-oF Subsystems

Tested on a Toshiba NVMe-oF Solution with 24 NVMe SSDs and 2 X 100GB NICs

Thank You!

Q&A

Yaron Klein Yaron.Klein@taec.toshiba.com
Verly Gafni-Hoek Verly.Gafni@taec.toshiba.com

http://storage.toshiba.com/nvme-of-software

TOSHIBA

Leading Innovation