SNIA NSF | NETWORKING STORAGE

Ethernet-Attached SSDs
Brilliant Idea or Storage Silliness?

Live Webcast
March 17, 2020
10:00 am PT

Today's Presenters

Moderator:
Ted Vojnovich
Lenovo

Presenter:
Mark Carlson
Kioxia

Presenter:
Rob Davis
Mellanox

Contrarian:
John F. Kim
Mellanox

SNIA-At-A-Glance

SNIA NSF | NETWORKING STORAGE

SNIA-at-a-Glance

185
industry leading organizations

2,000
active contributing members

50,000
IT end users & storage pros worldwide

Learn more: **snia.org/technical** **@SNIA**

Technologies We Cover

- Ethernet
- iSCSI
- NVMe-oF
- InfiniBand
- Fibre Channel, FCoE
- Hyperconverged (HCI)
- Storage protocols (block, file, object)
- Virtualized storage
- Software-defined storage

SNIA NSF | NETWORKING STORAGE

SNIA Legal Notice

- The material contained in this presentation is copyrighted by the SNIA unless otherwise noted.
- Member companies and individual members may use this material in presentations and literature under the following conditions:
  - Any slide or slides used must be reproduced in their entirety without modification
  - The SNIA must be acknowledged as the source of any material used in the body of any document containing material from these presentations.
- This presentation is a project of the SNIA.
- Neither the author nor the presenter is an attorney and nothing in this presentation is intended to be, or should be construed as legal advice or an opinion of counsel. If you need legal advice or a legal opinion please contact your attorney.
- The information presented herein represents the author's personal opinion and current understanding of the relevant issues involved. The author, the presenter, and the SNIA do not assume any responsibility or liability for damages arising out of any reliance on or use of this information.

NO WARRANTIES, EXPRESS OR IMPLIED. USE AT YOUR OWN RISK.

Agenda

SNIA NSF | NETWORKING STORAGE

- Brief history of storage access models
- Brief history of Ethernet as a storage network
- NVMe™ over Ethernet to drive – opportunities
  - Disaggregation & solution management
- NVMe over Ethernet to drive – use cases
- NVMe over Ethernet to drive – challenges / work to be done
- Debate: NVMe over Ethernet to drive:
  - Next step in evolution or solution looking for a problem to solve

The Evolution of Storage Networks

SNIA NSF | NETWORKING STORAGE

- Direct attached storage: Single host owns storage
- Storage Area Networks: Multiple hosts share storage
  - Avoid “silos” of storage and enables storage efficiencies
  - Examples include Fibre Channel & iSCSI storage networks
    - But require “Storage Controllers” to front storage
- Hyperscale: DAS storage on commodity systems
  - Special software manages many hyperscale nodes in a solution
- Industry moving to NVMe / NVMe-oF™ technology
  - Now, systems AND devices on native Ethernet as a Storage Network

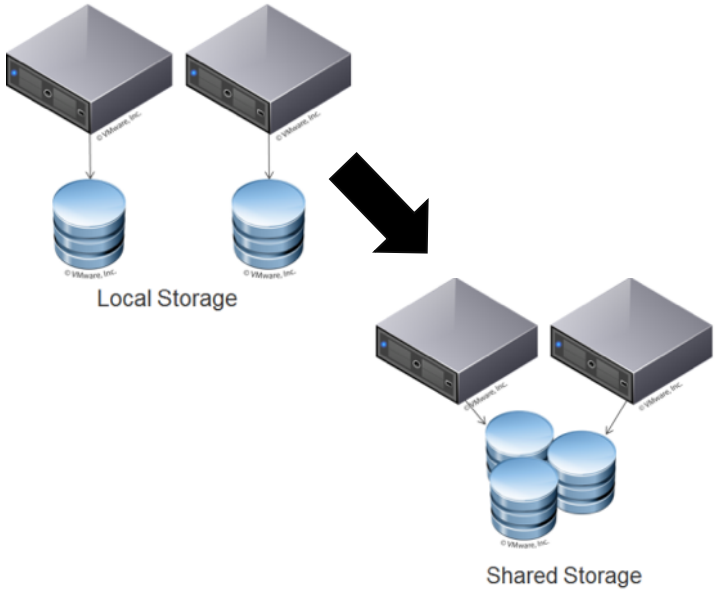
The Ethernet as a Storage Network

SNIA NSF | NETWORKING STORAGE

- Initially, just a transport
  - End points performed all the storage services (iSCSI)
- Use of Ethernet matured: Specialized protocols
  - Key/value protocol to access data in mainframe context
  - Object protocol to access massive amounts of unstructured data
- Now, NVMe over Ethernet: Storage in a queuing paradigm
  - High performance / low latency / few or no processing blockages
  - No longer gated by transaction paradigm (wait for ACK)
- Next step, NVMe over Ethernet to the drive
  - Removes “Storage Controller” processing blockage

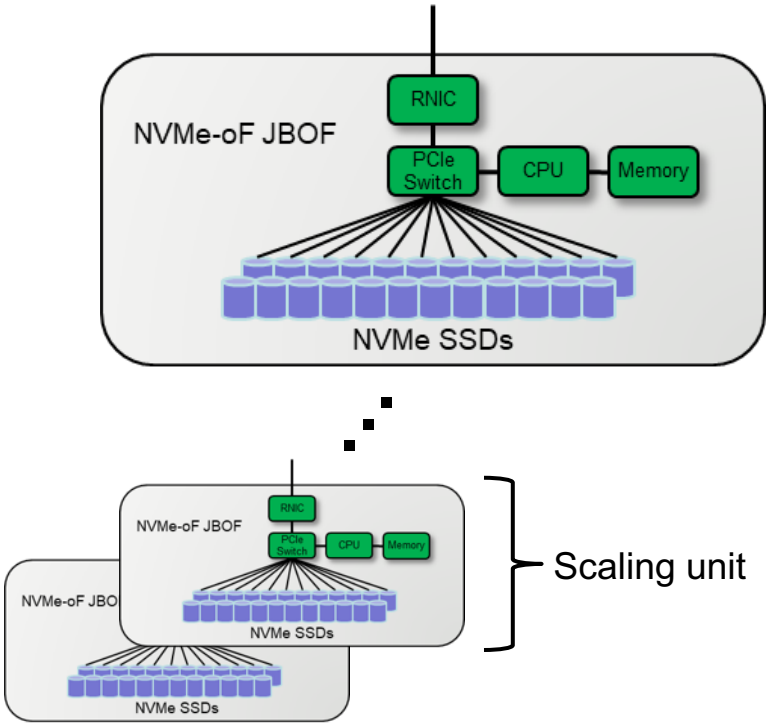
NVMe over Fabrics (NVMe-oF)

- Sharing NVMe based storage across a Network
  - Better utilization: capacity, rack space, power
  - Better scalability: management, fault isolation
- NVMe-oF standard at NVMe.org
  - 50+ contributors
  - Version 1.0 released in 2016
  - Fabrics: Ethernet, InfiniBand, Fibre Channel
- Products now in the market from most major storage system vendors



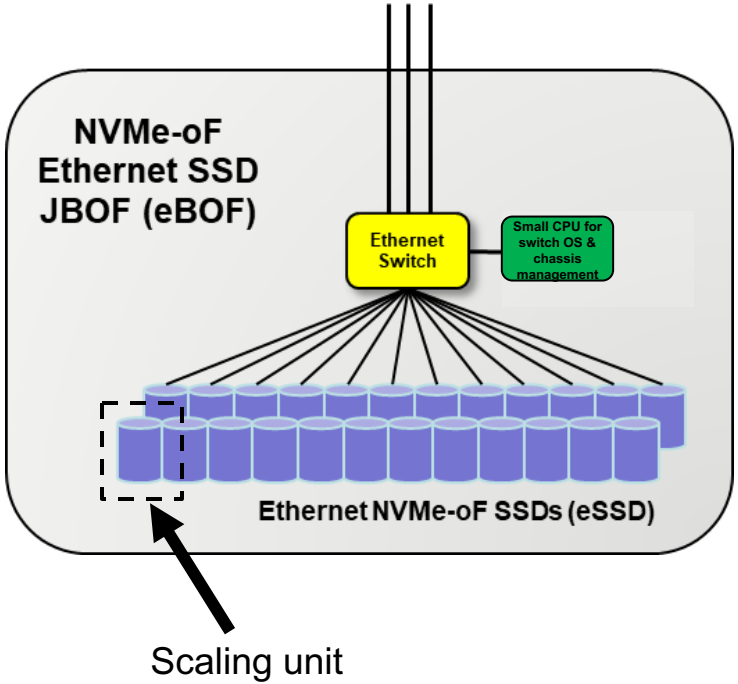
NVMe-oF Storage Targets Today

- Systems terminate the NVMe-oF connection and use PCIe based SSDs internally
  - SSDs behind an array/JBOF controller
- Performance Limits
  - SSD performance increasing faster than CPU NVMe-over-Ethernet-to-drive use cases
  - NIC performance
  - Latency - Store and Forward architecture
- Cost – CPU, SOC/rNICs, Switches, Memory don't scale well to match increasing SSD performance



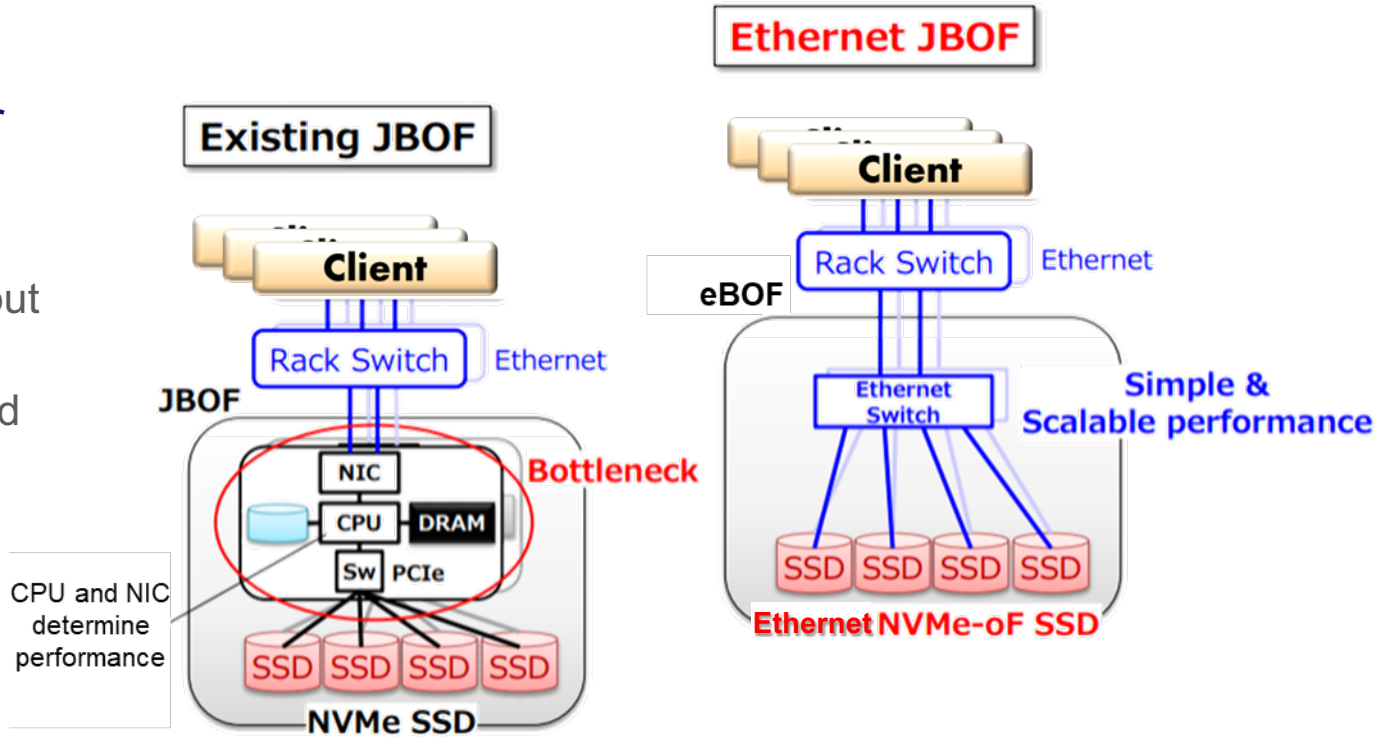
NVMe-oF Ethernet SSDs

- With NVMe-oF termination on the drive itself, controller functionality is now distributed
  - Scaling point becomes a single drive in an inexpensive enclosure
  - Enables eBOFs (Ethernet-attached Bunch Of Flash)
    - Power, cooling, SSDs, and an Ethernet Switch
- Does this make each drive more expensive?
  - Maybe initially, but now customer buys their “controller” incrementally, as needed for new capacity
  - Efficiencies of scale now are applied to controller functionality
  - Lower cost/bandwidth and cost/IOPS



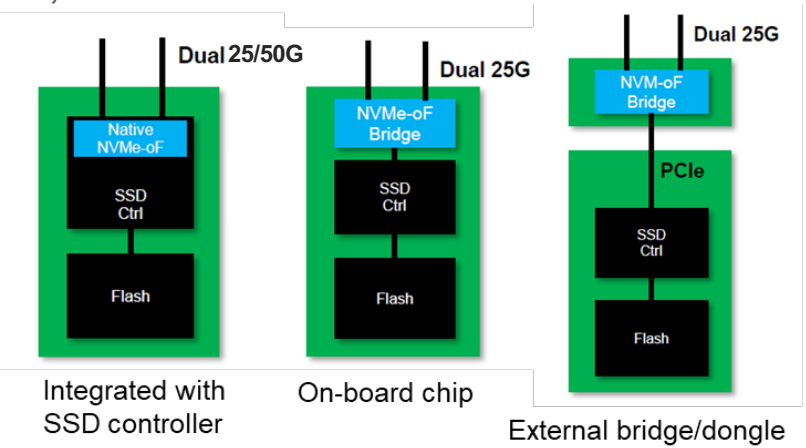
JBOF CPU/NIC Complex can be a Bottleneck

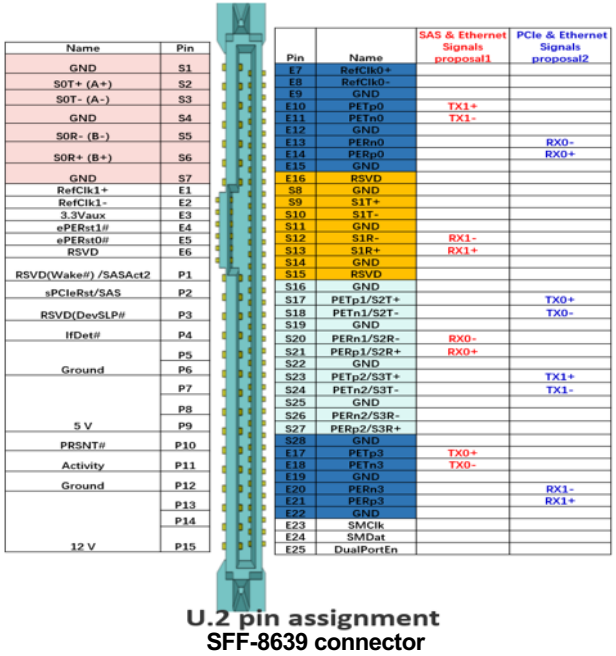
- SSD throughput increasing faster than network bandwidth
  - SSD throughput will triple
  - Network speed only doubles



eSSDs

- Different eSSD designs today
- Some will support multiple interfaces and protocols
  - Ethernet, PCIe, SAS, SATA
  - RoCE, TCP





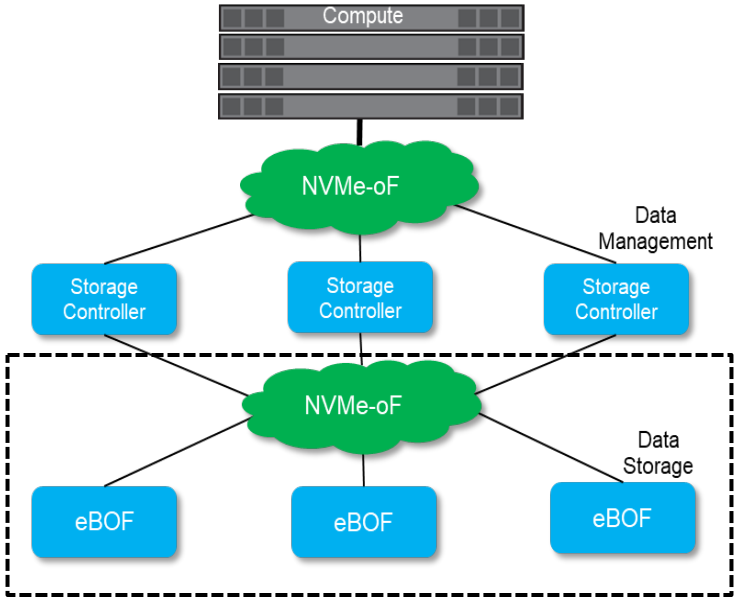
Name	Pin
GND	S1
S0T+ (A+)	S2
S0T- (A-)	S3
GND	S4
S0R- (B-)	S5
S0R+ (B+)	S6
GND	S7
RefClk1+	E1
RefClk1-	E2
3.3Vaux	E3
ePERst1#	E4
ePERst0#	E5
RSVD	E6
RSVD(Wake#) /SASAct2	P1
sPCIeRst/SAS	P2
RSVD(DevSLP#	P3
IfDet#	P4
	P5
Ground	P6
	P7
	P8
5 V	P9
PRSNT#	P10
Activity	P11
Ground	P12
	P13
	P14
12 V	P15

Pin	Name	SAS & Ethernet Signals proposal1	PCIe & Ethernet Signals proposal2
E7	RefClk0+		
E8	RefClk0-		
E9	GND		
E10	PETp0	TX1+	
E11	PETn0	TX1-	
E12	GND		
E13	PERn0		RX0-
E14	PERp0		RX0+
E15	GND		
E16	RSVD		
S8	GND		
S9	S1T+		
S10	S1T-		
S11	GND		
S12	S1R-	RX1-	
S13	S1R+	RX1+	
S14	GND		
S15	RSVD		
S16	GND		
S17	PETp1/S2T+		TX0+
S18	PETn1/S2T-		TX0-
S19	GND		
S20	PERn1/S2R-	RX0-	
S21	PERp1/S2R+	RX0+	
S22	GND		
S23	PETp2/S3T+		TX1+
S24	PETn2/S3T-		TX1-
S25	GND		
S26	PERn2/S3R-		
S27	PERp2/S3R+		
S28	GND		
E17	PETp3	TX0+	
E18	PETn3	TX0-	
E19	GND		
E20	PERn3		RX1-
E21	PERp3		RX1+
E22	GND		
E23	SMClk		
E24	SMDat		
E25	DualPortEn		

U.2 pin assignment
SFF-8639 connector

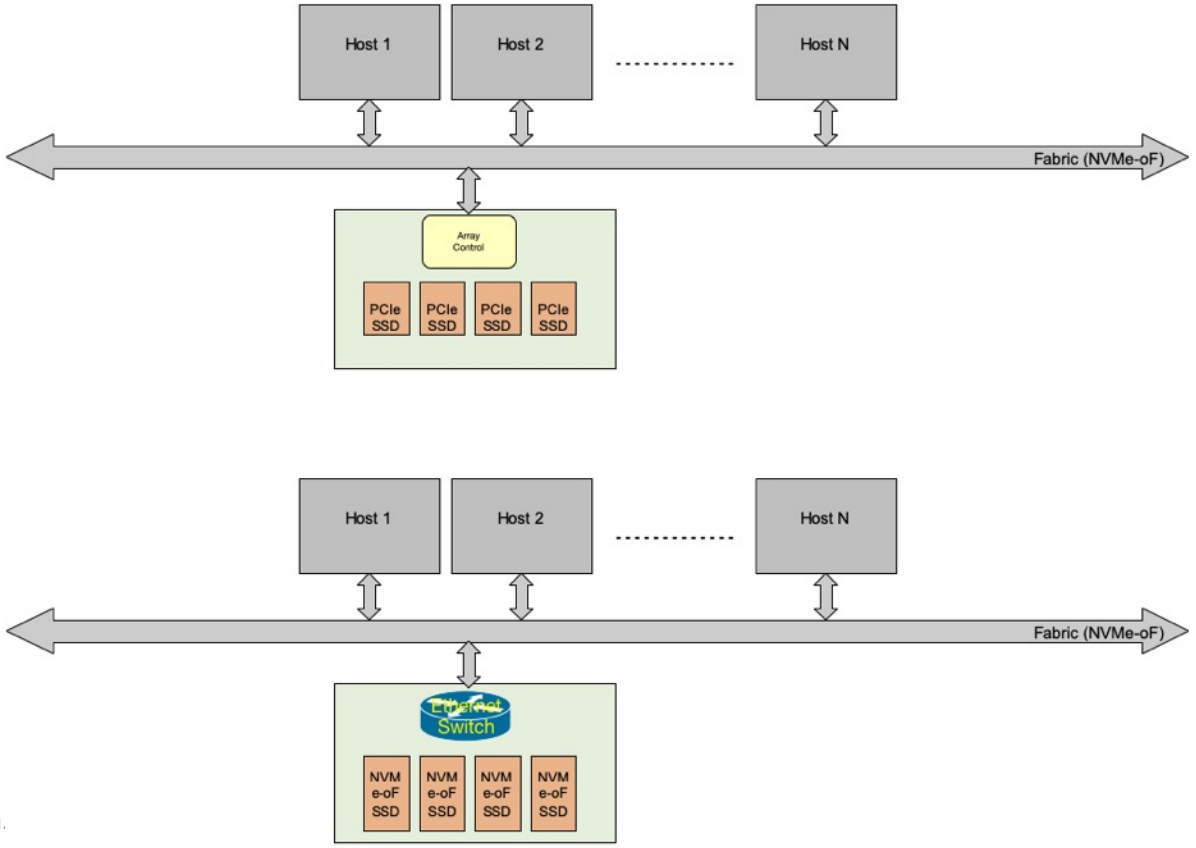
Use Case: Behind the Controller

- Scale storage capacity with large pools of disks
  - Many NVMe SSDs in many enclosures
  - PCIe only scales so far and at JBOF increments
- Using eSSDs allows much higher scaling
  - Still hiding individual SSD management from users
- Data services in the storage controllers ➔ value add
  - Orchestration between hosts and large pools of disks
    - Whole disks or slices of disks that provide massive pools effectively
  - Robust data protection schemes / distributed solution controllers



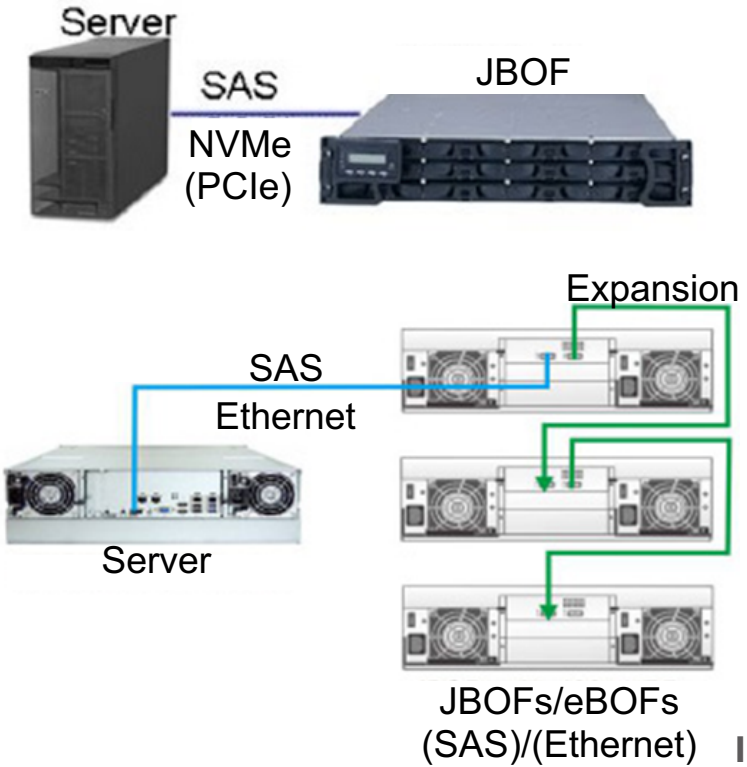
Use Case: Disaggregated SSD Storage

- Today: Array controller handles conversion from NVMe-oF to PCIe based drives
- With eSSD: Ethernet drives only require an Ethernet Switch and fit into an eBOF for power and cooling



Use Case: DAS Capacity Expansion

- Today:
  - Server's SAS controller has expansion port to external SAS JBOF
  - Or external PCIe port to NVMe JBOF
- With eSSD:
  - Unlike SAS, it is difficult to extend PCIe, but easy to extend Ethernet
  - Cost savings by removing SAS infrastructure from the Server

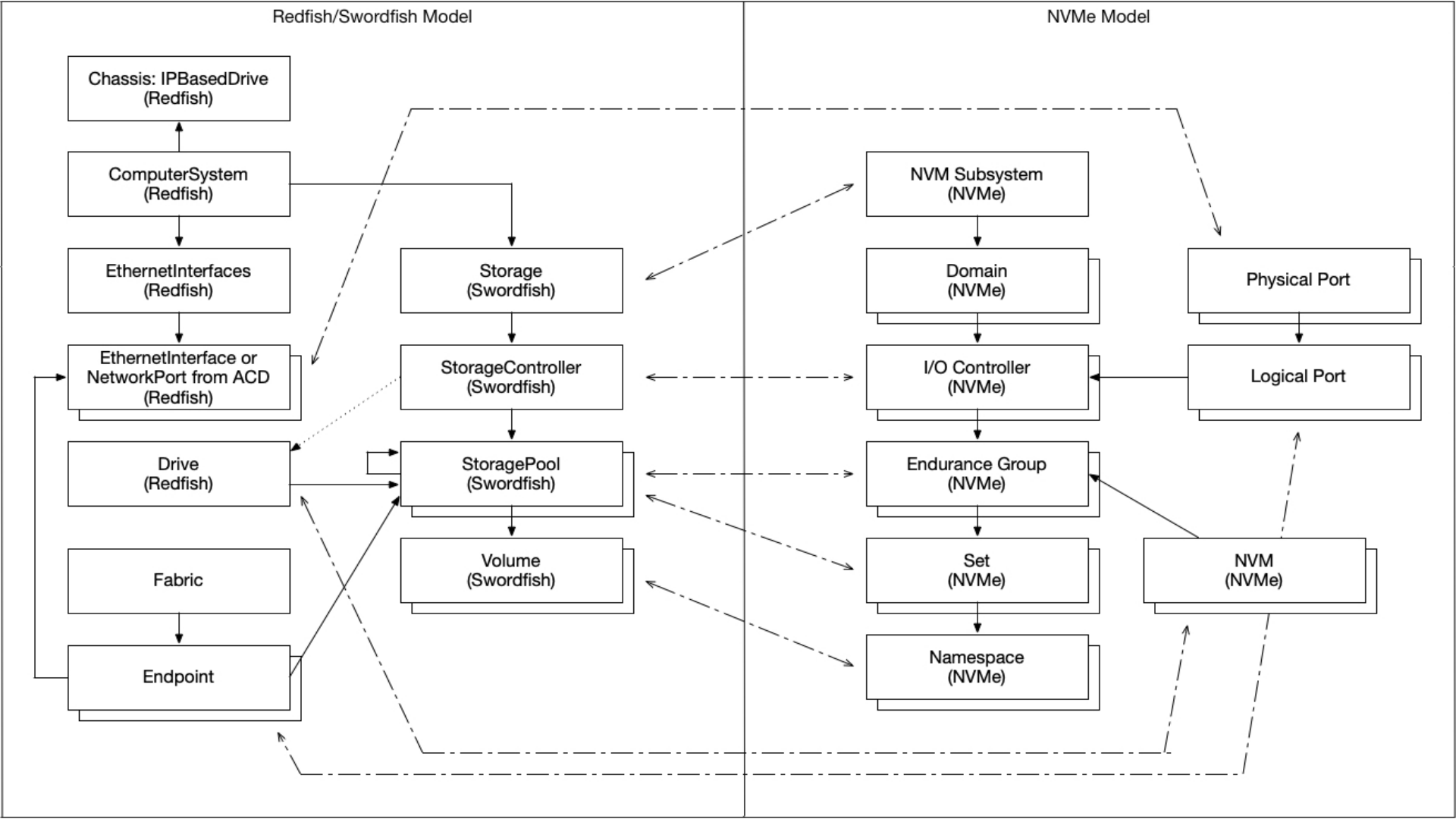


SNIA Native NVMe-oF Drive Specification

- Discover and Configure: the drives, their interfaces, the speeds, the management capabilities
- Connectors
  - Some connectors may need to configure the PHY signals based on the type of drive interface
  - Survivability and mutual detection is important
- Pin-outs
  - For common connectors and form factors
- NVMe-oF integration
  - Discovery controllers / Admin controllers
- Management
  - Through Ethernet/TCP for Datacenter-wide management

Management

- Scale out orchestration of 10's of thousands of drives possible by using a RESTful API such as DTMF Redfish™
- Redfish/SNIA Swordfish™ follow a principal that each element report it's own management information
  - Follow links in higher level management directly to the drive's management endpoint
  - HTTP/TCP/Ethernet based
- NVMe-oF Drive Interoperability Profile
  - Mock up to start
  - Push new models through Swordfish contributions
  - Publish Interoperability Profile at DMTF
- Map the profile to NVMe & NVMe-MI properties and actions


Redfish/Swordfish Model
NVMe Model
Chassis: IPBasedDrive (Redfish)
ComputerSystem (Redfish)
EthernetInterfaces (Redfish)
EthernetInterface or NetworkPort from ACD (Redfish)
Drive (Redfish)
Fabric
Endpoint
Storage (Swordfish)
StorageController (Swordfish)
StoragePool (Swordfish)
Volume (Swordfish)
NVM Subsystem (NVMe)
Domain (NVMe)
I/O Controller (NVMe)
Endurance Group (NVMe)
Set (NVMe)
Namespace (NVMe)
Physical Port
Logical Port
NVM (NVMe)

Data Services

- Modern storage system controllers also implement data services
  - Dedup, Compression, Replication, Encryption, etc.
- Data services software (SDS) can be run anywhere in the network on commodity hardware
  - Hyperscaler approach: roll your own
  - Enterprise approach: licensed software
- Some of these services are envisioned to move into drives
  - Computational Storage

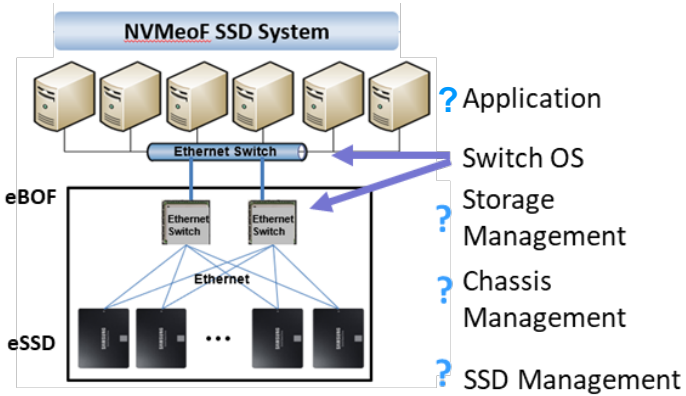
Computational Storage Future

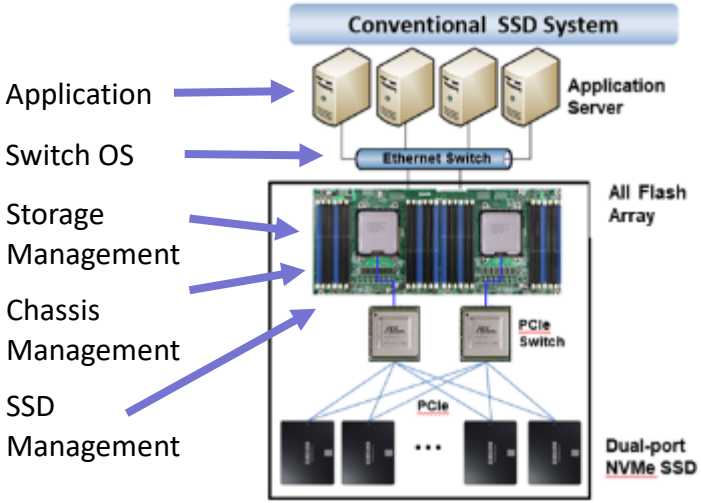
- Opportunity to move the computational tasks to the data where it lives
  - Queries and searches can be parallelized across multiple devices
- But limited if just offloading a single host (i.e. by PCIe)
- High likelihood that NVMe will be extended to accommodate the Computational Storage functions
- Distributing computational storage across the network via Ethernet allows it to be globally shared
  - Perhaps via CXL in the future
- SNIA is a first mover in Computational Storage standards

But Then There's Our Villain

But Wait… Concerns?

- Where is the storage software?
- How do I provision the storage?
- Does my application need to be modified?
- Where is the data protection?



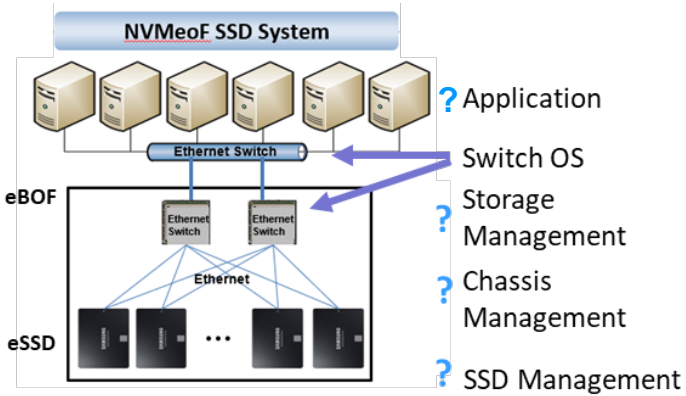


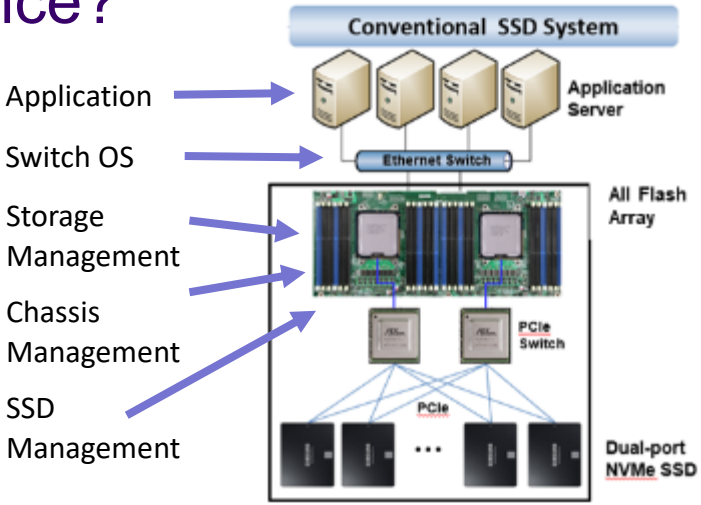
eSSD Use Case is Key

- Back-end scale-out: No problem!
  - Features/management still on controller
- Distributed storage software: Probably fine
  - Large, controlled and closed environment
  - Storage features distributed across many servers
  - Ideal for key-value store or computational storage
- Standard enterprise storage: Not ready yet!
  - Infrastructure not ready yet to consume eSSD safely
  - Software to provision, manage, secure, and protect must live somewhere

More E-SSD Concerns

- What about balancing performance?
- Now I need more switches!
- Who enforces security?





Areas of Debate

- What are Pros/Cons of NVMe over Ethernet to the drive?
  - Next logical step or just another experiment
- What are Pros/Cons of NVMe over Ethernet to the drive solutions?
  - Problems solved vs. inhibitors
- Ultimately, is this a pervasive or niche solution?
  - What will be the “killer App” for NVMe over Ethernet to the drive
  - Simply a better storage model, or needs computational storage, etc. to make sense?

Summary

SNIA NSF | NETWORKING STORAGE

- Ethernet as a storage network continues to mature
- NVMe over Ethernet continues to mature
- NVMe over Ethernet to drive offers new capabilities
  - Flexibility, massive scaling, elimination of solution “choke” points
- NVMe over Ethernet to drive has some current challenges
  - Orchestration, baseband drive functions
- Debate over the vision vs actual customer value
  - First movers will clear the “fog”

SNIA Technical Work

- Object Drive Technical Work Group
  - https://www.snia.org/object-drives
- Scalable Storage Management Technical Work Group
  - https://www.snia.org/tech_activities/standards/curr_standards/swordfish
- Computational Storage Technical Work Group
  - https://www.snia.org/computational

After This Webcast

SNIA NSF | NETWORKING STORAGE

- Please rate this webcast and provide us with feedback
- This webcast and a PDF of the slides will be posted to the SNIA Networking Storage Forum (NSF) website and available on-demand at www.snia.org/forums/nsf/knowledge/webcasts
- A full Q&A from this webcast, including answers to questions we couldn't get to today, will be posted to the SNIA-NSF blog: sniansfblog.org
- Follow us on Twitter @SNIANSF

Thank You