STORAGE DEVELOPER CONFERENCE
SDC 22
Fremont, CA
September 12-15, 2022
BY Developers FOR Developers

A SNIA Event

Data Processing Unit as a Storage Initiator

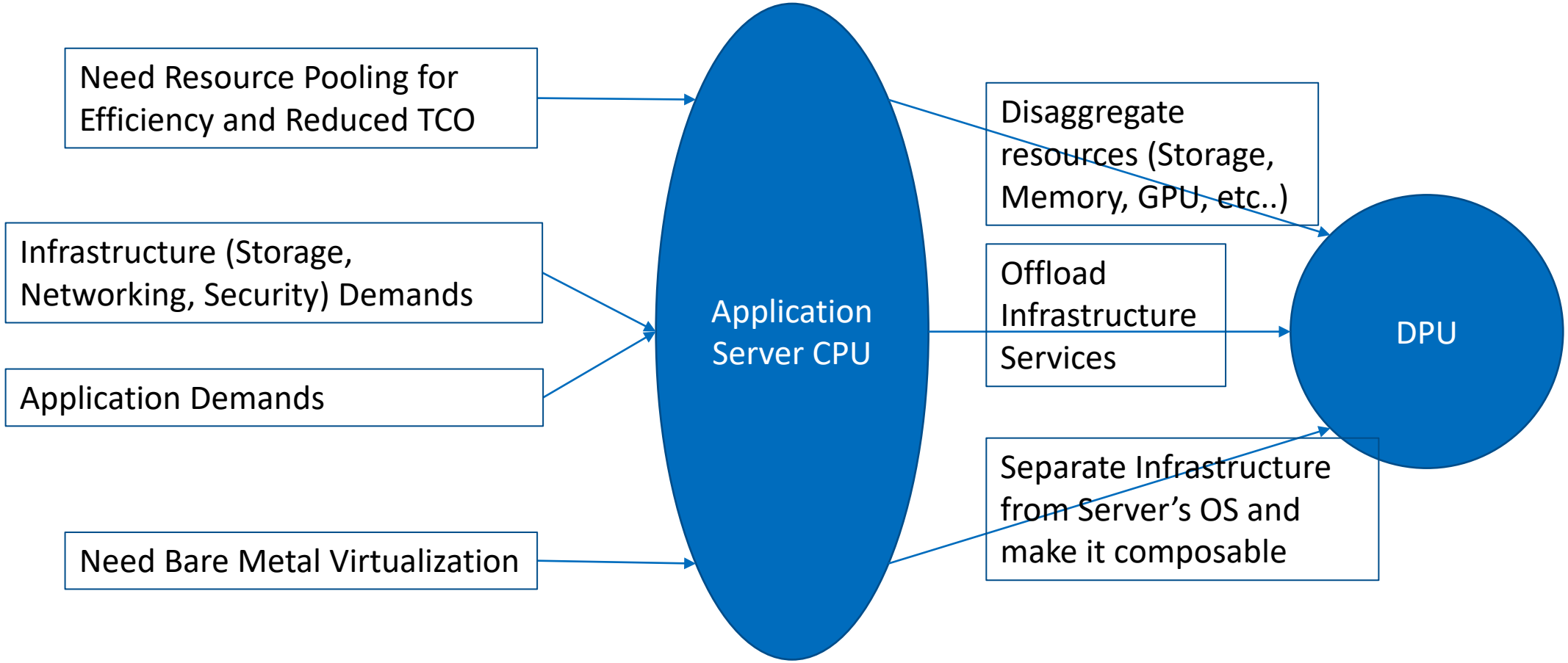
Enabling High Performance Storage Disaggregation and Bare Metal Virtualization

PratapaReddy Vaka
Sr. Director, Storage Software
Fungible Inc
pratapa.vaka@fungible.com

Agenda

- Need for DPU in Compute Nodes
- DPU as a Storage Initiator (SI)
- Integration with Orchestration Systems
- SI Performance

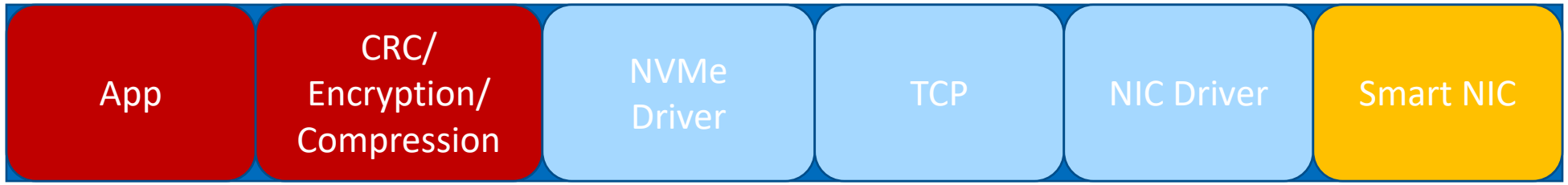
Why DPU in Compute Nodes?



FUNGIBLE

DPU as a Storage Initiator

- Storage Protocol Offloads (NVMe/TCP, NVMe/TLS, NVMe/RoCE, NVMe/TrueFabric™, etc…)
- Storage Data Processing Offloads (In line Encryption/Compression/ErasureCoding/CRC)



Software Storage Initiator

DPU as a Storage Initiator

FUNGIBLE

DPU as a Storage Initiator

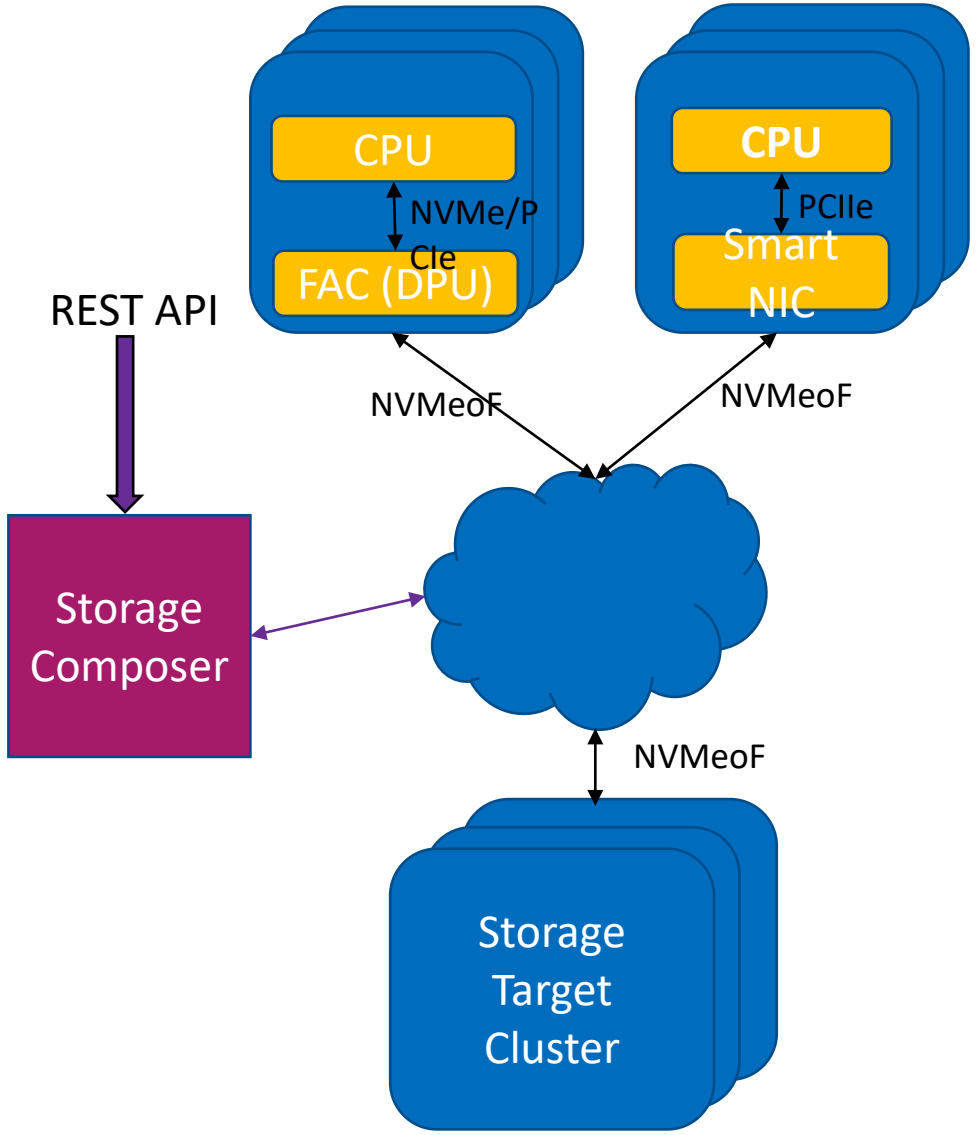
- Deliver high performance storage disaggregation
  - Remote Storage Looks Like Local Storage
  - Highly pipelined and async processing in DPU and in-line HW accelerators enable efficient use of resources
  - SR-IOV with multiple PFs and VFs enables bare metal performance for VMs
- Offload Storage Processing
  - Storage protocol and data processing (encryption and compression) offloads in addition to networking and security offloads save significant number of Server CPU cycles
- Enable Bare Metal Virtualization
  - Device Emulation and Composability of Storage and Networking without depending on Server's OS
  - Emulate large number of storage devices (For ex: NVMe namespaces)
- Enhance Storage Security
  - Support of Data Encryption, TLS v1.3, and In-band Auth with very little impact on performance
  - Hardware enforcement of storage security policies

NVMe/TCP Storage Initiator
(Fungible Accelerator Card)

FUNGIBLE

DPU as a Storage Initiator – More Benefits

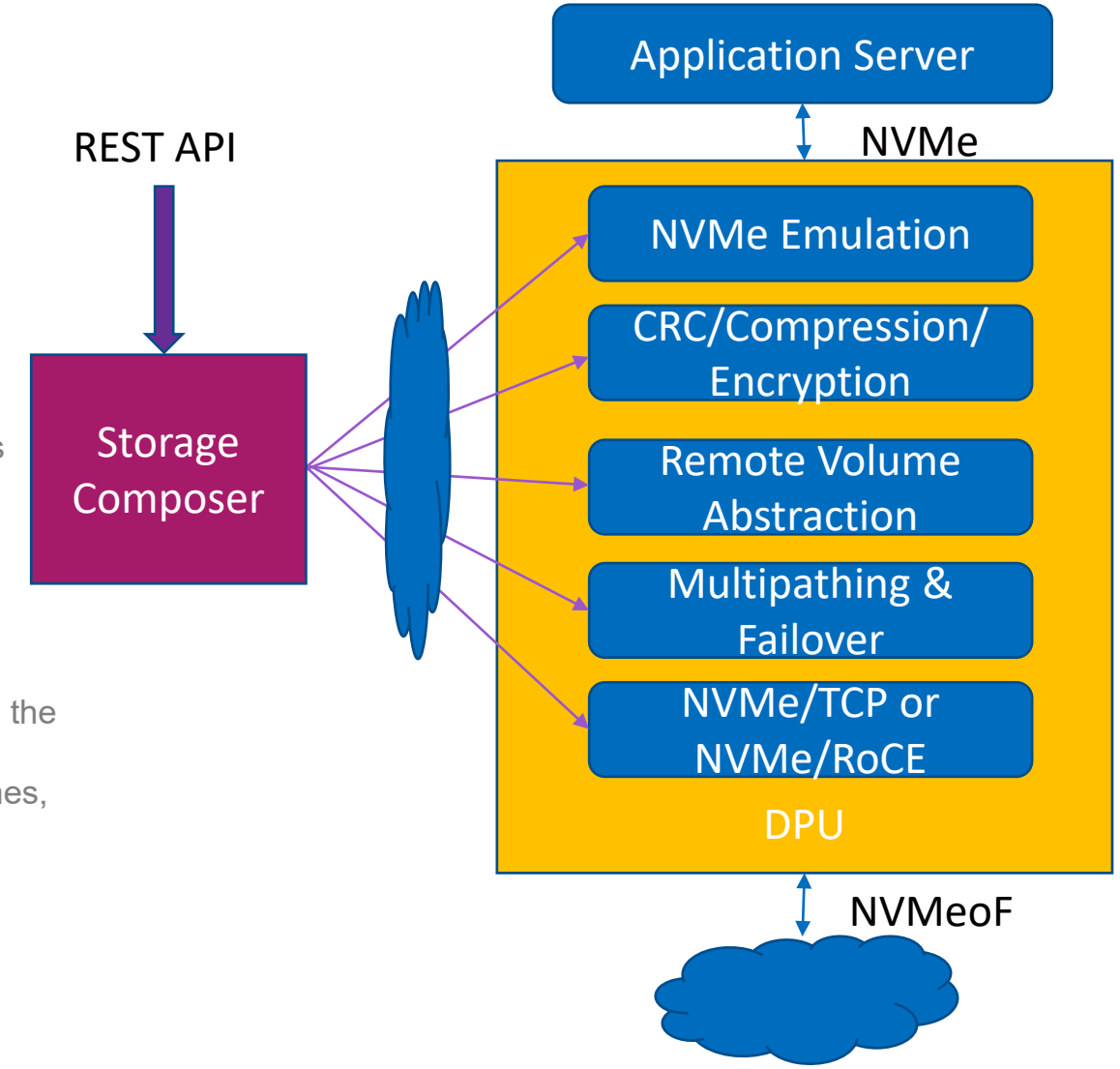
- Support Multi Tenancy with QoS
  - QoS (Guarantees and Limits) between Tenants to adhere to SLAs
- Programmable
  - Quickly roll out new or custom features
- High Availability
  - Handle Multipathing and Failover to Storage Targets
- Support End to End Data Integrity
  - Use of HW CRC Accelerators for performance
- Eco System
  - NVMe Standards based Protocols and Discovery: NVMe, NVMe/TCP, NVMe/RDMA, NVMe Discovery
  - REST based Intent API for the centralized and highly available composer for easy integration with any orchestration system
  - Integration with open-source orchestration systems like K8S and OpenStack
- Low Power
  - DPU's small size and footprint reduces power and cooling requirements
- Support Diskless Servers
  - Remote Boot of Bare Metal and VMs
- Reduce Network Traffic
  - Inline Compression, Client caching, Read direct from back-end node, etc...



FUNGIBLE

SI - Software Architecture

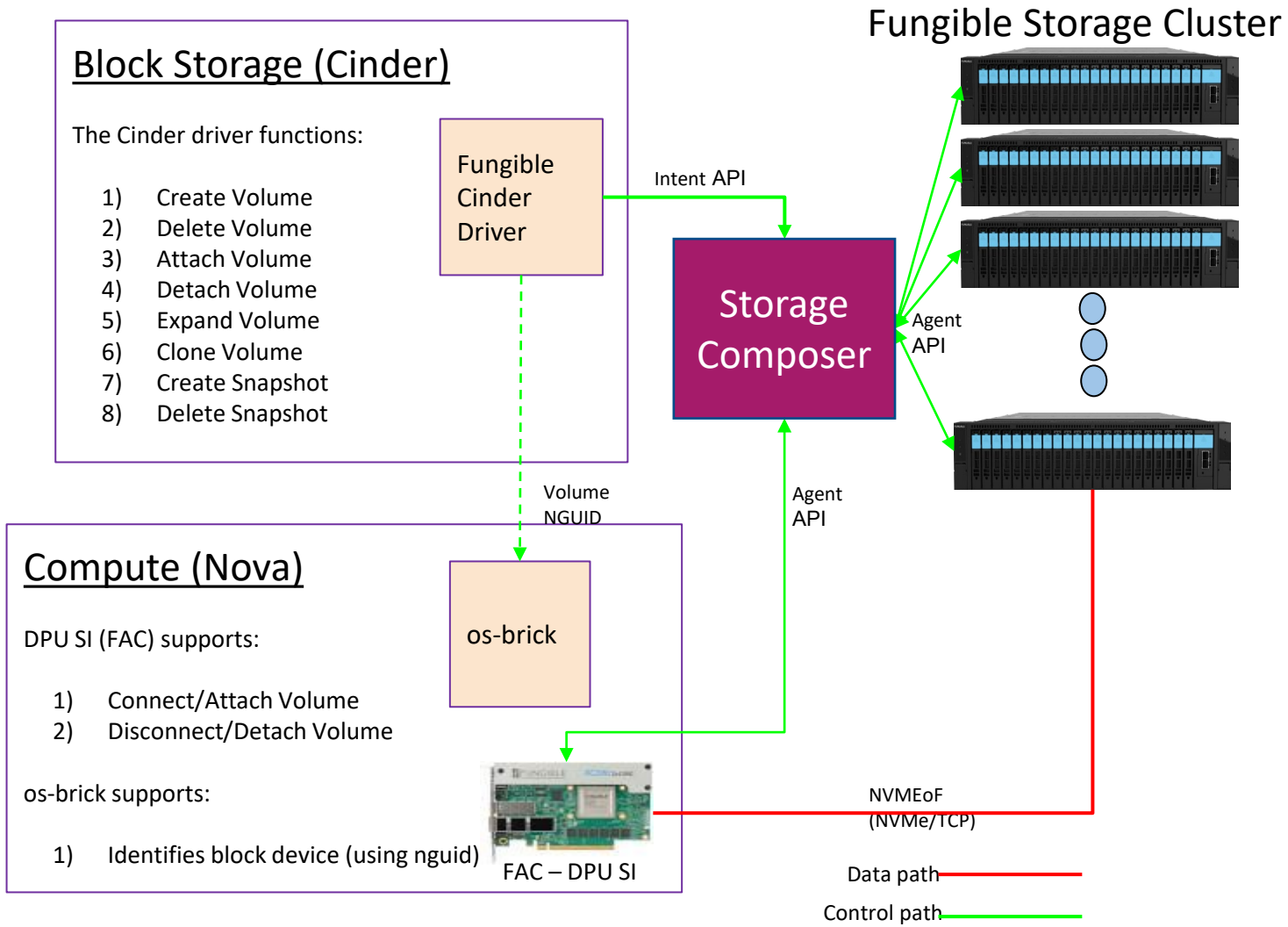
- Async and Run to Completion Model
  - Efficient messaging between processing threads
  - Minimize context switching overhead
  - No interrupt overhead
- Pipelining
  - Pipelining of IO processing among a few processer threads for efficiency
  - Placement of many flexible pipelines among the processor threads
- Zero Copy
  - The Storage Software never touches the data
  - Data moves through HW units (Networking, PCIe) and Inline Accelerators through efficient DMA engines
- In line HW accelerators
  - Efficient DMA handling, Pipelining, Multiple HW threads, Load balancing across HW threads, etc…
- Composable Data Plane
  - Allow the data plane to be composed by an orchestration system through the centralized Composer
  - Create/Delete/Attach/Detach/Mount/Unmount of NVMe controllers, volumes, and devices with right abstractions
  - Per Volume Composability of In-line Accelerators
- Lock Free and Cache Friendly
  - Processing thread serialization and run-to-completion
  - Prefetching cache lines ahead of picking the next run-to-completion processing handler



FUNGIBLE

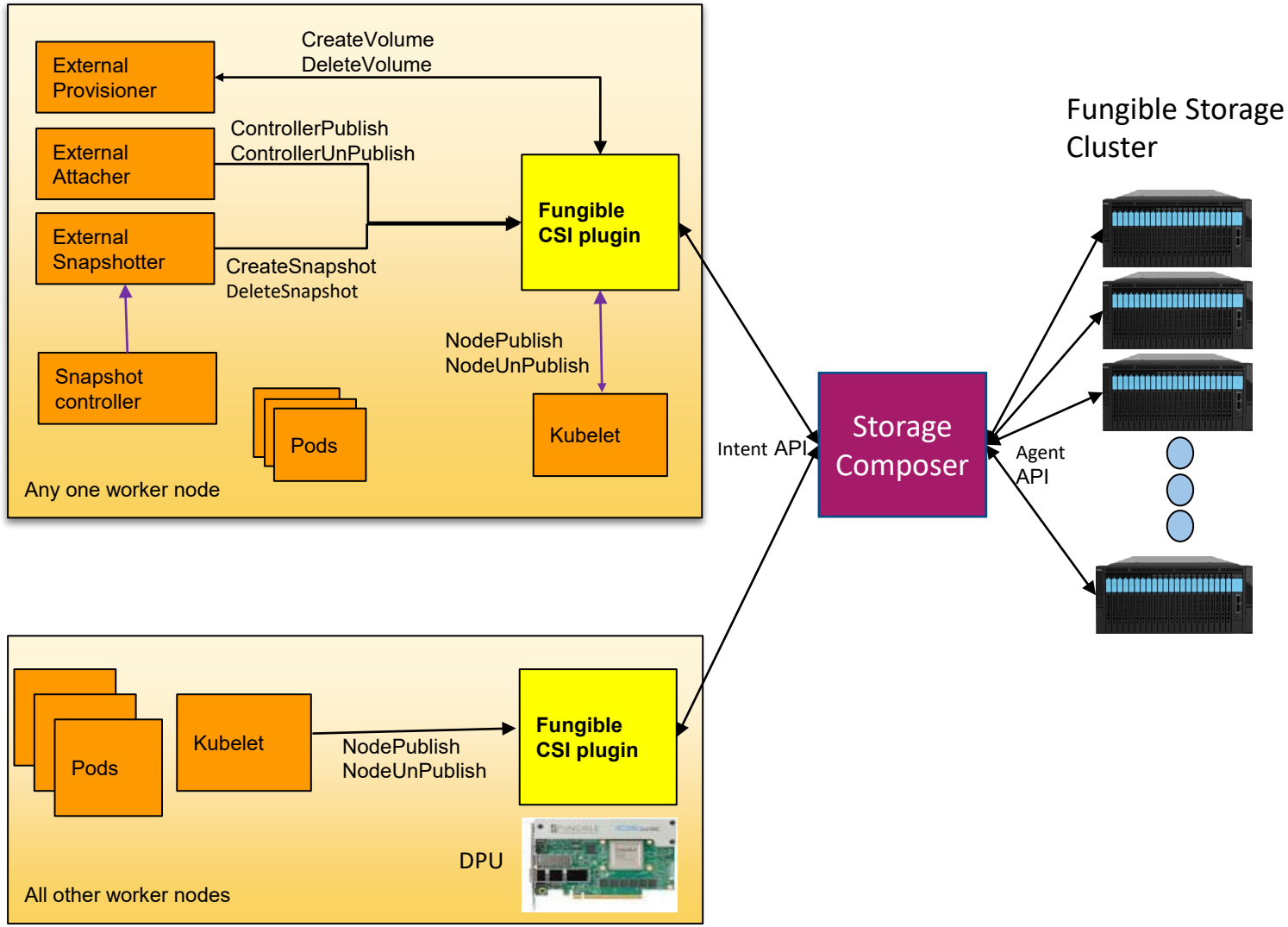
Integration with Open Stack

- Open Stack is a Cloud Operating System for orchestrating the cloud infrastructure (storage, networking, security, and other resources)
- Open Stack Cinder component virtualizes pools of block storage devices and provides end users with a consistent API to access different types of storage
- Open Stack allows Vendor Specific Cinder Drivers to communicate with the storage targets and initiators
- Cinder provides the information needed for Open Stack Nova to attach VMs to Volumes.

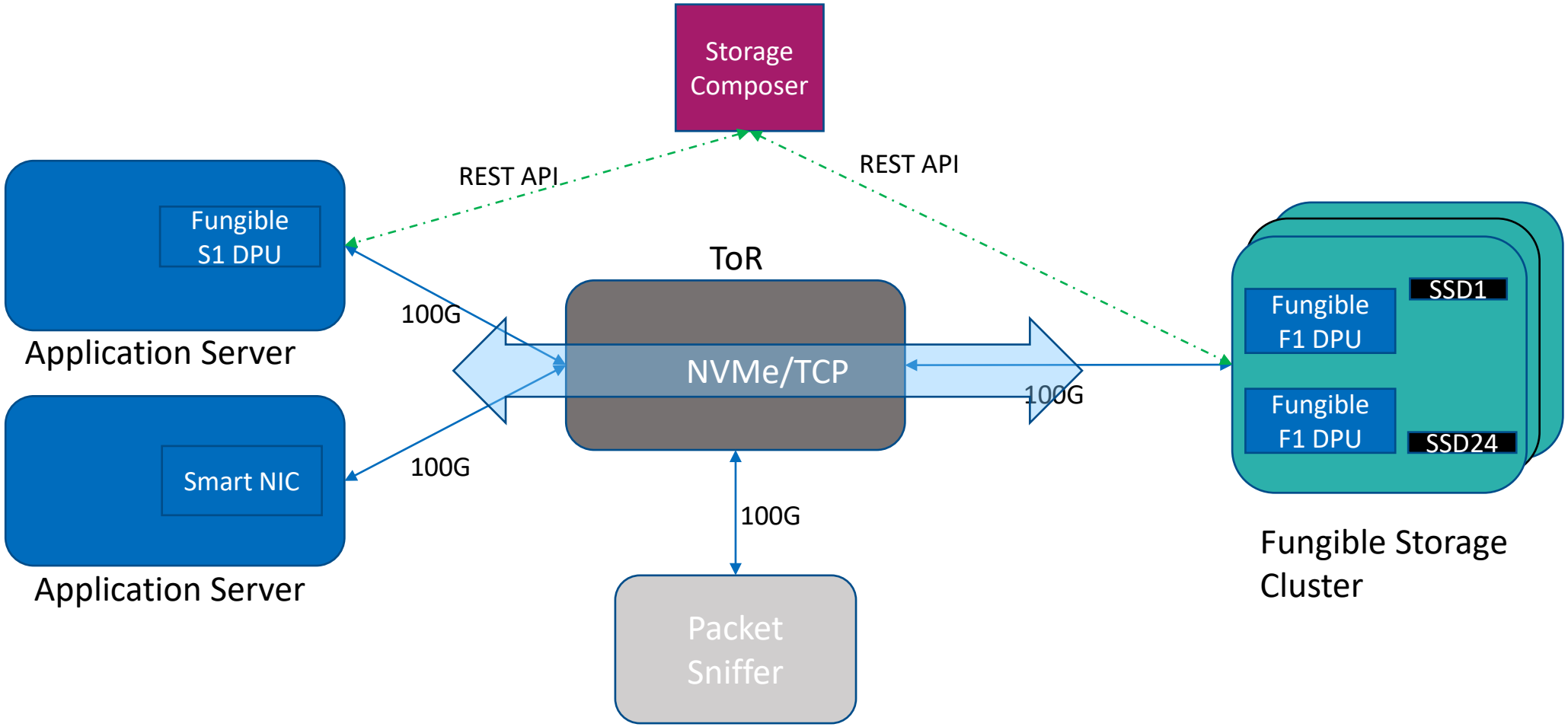


Integration with K8S – CSI Plugin

- Container Storage Interface (CSI) Specification for developing plugin for Container Orchestration (CO) systems
  - Defined gRPCs that CO can invoke for provisioning & attachment
- One plugin works across multiples COs
  - Kubernetes, Cloud Foundry, Mesos
- The CSI Plugin allows
  - Attaching/Detaching NVMe namespaces to the Containers
  - Creating/Deleting Remote Volumes
  - Defining New Storage Classes
- The Storage Composer creates an NVMe namespace (remote volume) in the SI, connects it to the Storage target via multiple paths (when multiple paths are available), and attaches it to a PCIe function.
- The Kubelet associates the NVMe namespace with a container by finding the info from Storage Composer via CSI plugin



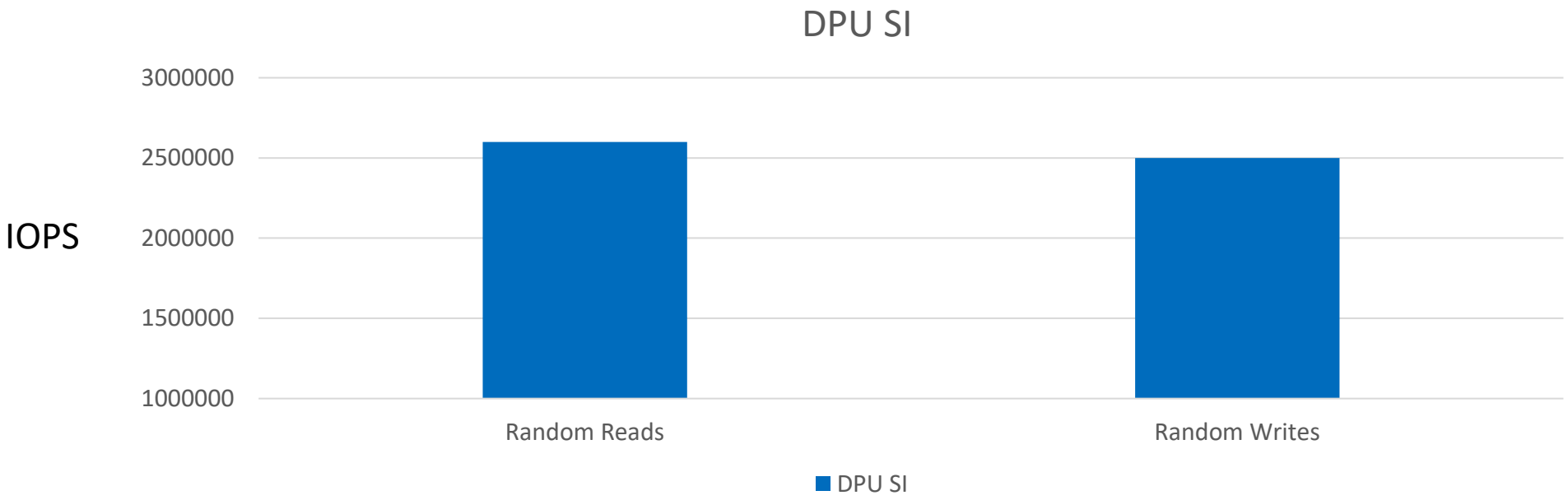
Storage Initiator – Benchmarking Setup



FUNGIBLE

Storage Initiator - Performance

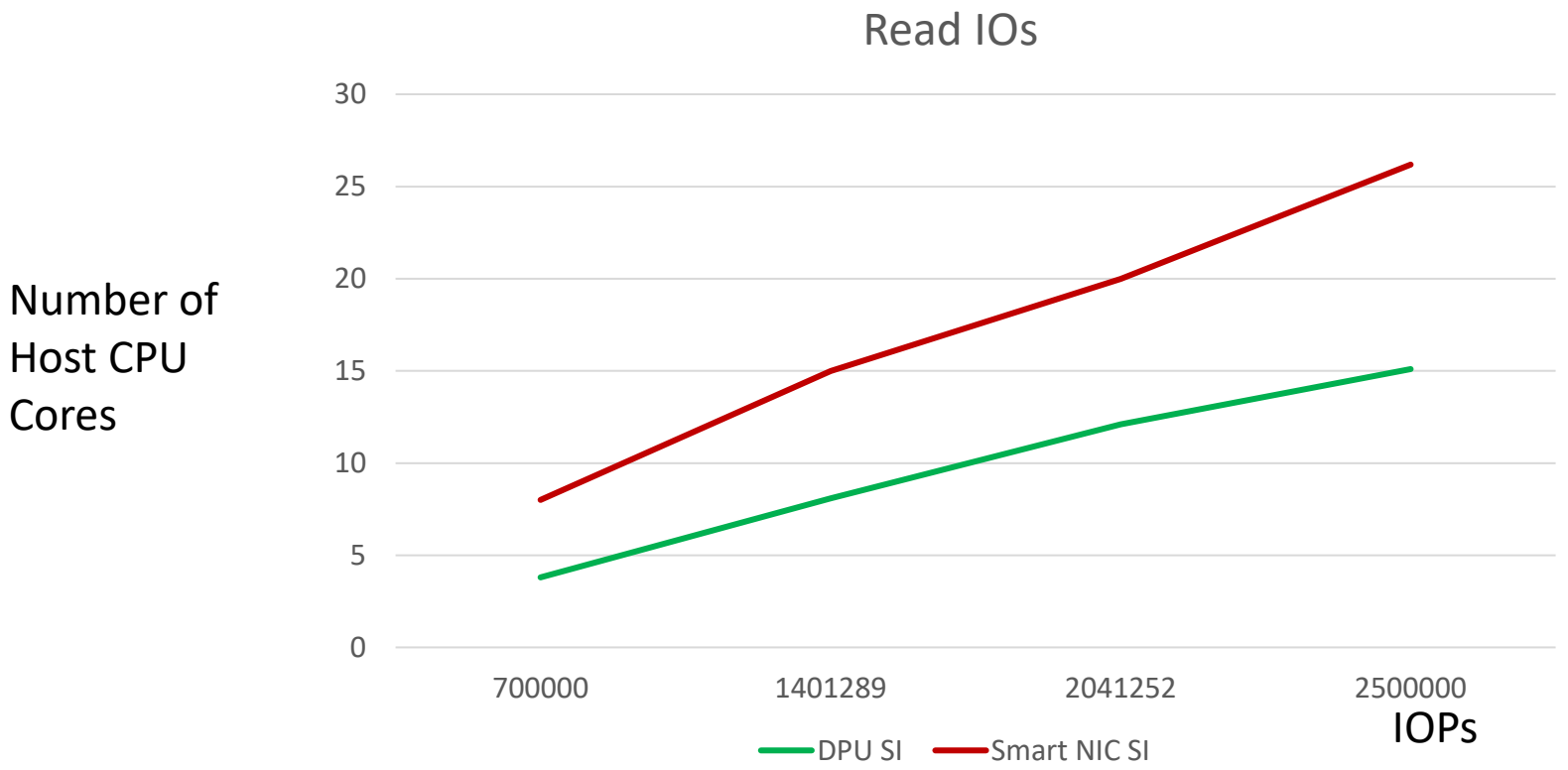
- Fungible FC200 delivers unparalleled performance of 2.5M+ Random Read/Write IOPs
- Delivers same bare metal performance to VMs using VF pass through



FUNGIBLE

Storage Initiator – Host CPU Utilization

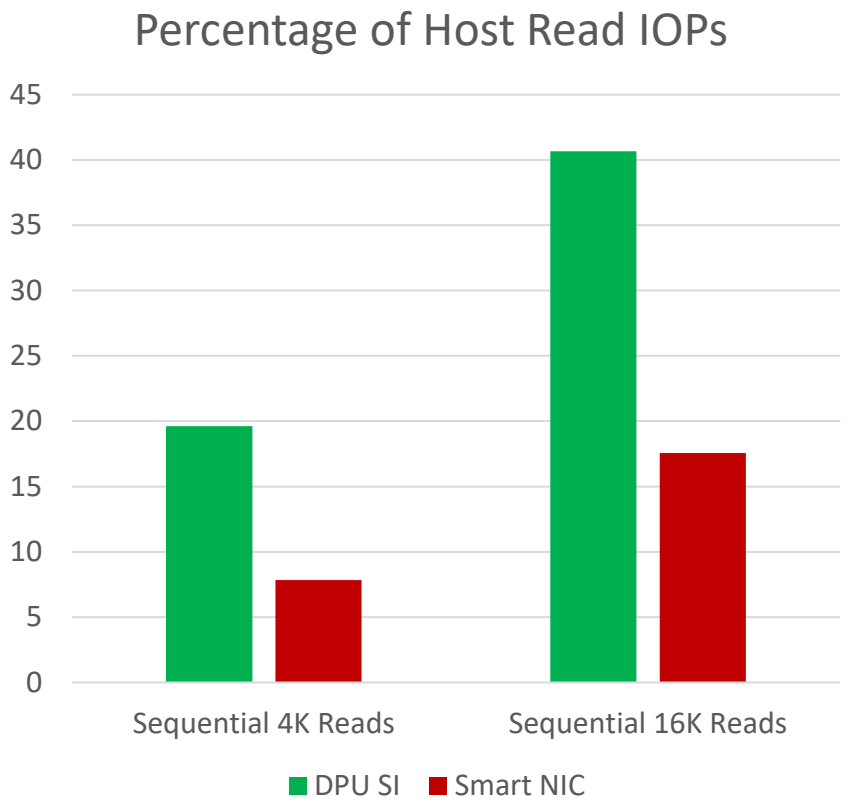
- Fungible FC200 saves ~10 CPU cores per ~2.5M IOPs with NVMe/TCP offloaded
- Host CPU cycle savings will be much more with Compression and Encryption also offloaded

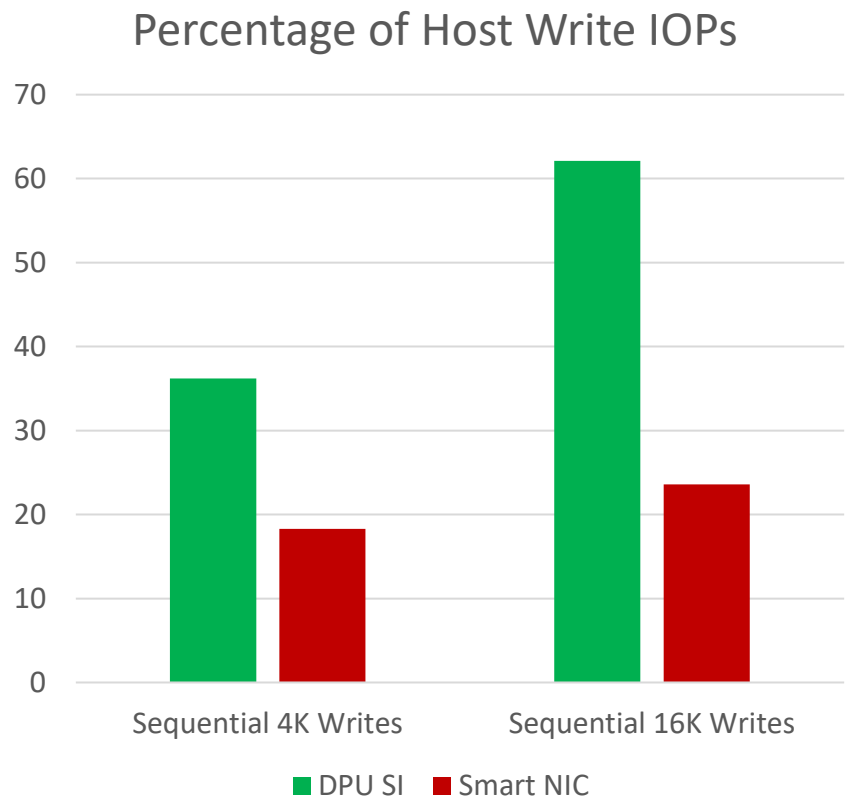


FUNGIBLE

Storage Initiator – VirtIO Performance

- Virt IO Performance nearly doubles when DPU-SI is used





FUNGIBLE

Other DPU Presentations from Fungible

- Next Generation Architecture For Scale-out Block Storage By Jaspal Kohli
- The Rise of DPU-based Storage Systems By Jaishankar (Jai) Menon

FUNGIBLE

Thank You!

Please take a moment to rate this session.

Your feedback is important to us.

FUNGIBLE

BACKUP

FUNGIBLE

Section Subtitle

Section Title

Section Subtitle

FUNGIBLE

Light Slide Title

- Bullets 1
  - Bullets 2
    - Bullets 3
      - Bullets 4
        - Bullets 5

FUNGIBLE

-
  - Bullets 2
    - Bullets 3
      - Bullets 4
        - Bullets 5

FUNGIBLE

Please take a moment to rate this session.

Your feedback is important to us.

FUNGIBLE