



Flash Memory Summit

Pre-Conference Seminar B: SmartNIC/DPU Storage Solutions

Presented by: Rob Davis (NVIDIA), Joseph White (Dell/EMC), John Mao (VAST), Donpaul Stevens (AirMettle), Brad Reger (Foxconn/Ingrasys) and Jeff Feierfeil (Nebulon)

Agenda



Flash Memory Summit

- Teaser
 - Rob Davis – NVIDIA
- DPU for Storage: An Introduction
 - Joseph White – Dell/EMC
- Disaggregated Storage Architectures with DPUs and Next-Gen JBOFs
 - John Mao – VAST
- Interactive Analytics & AI made possible by DPUs
 - Donpaul Stevens – AirMettle
- 15 minute break
- DPUs Empower New Storage Architecture for NVMe-oF Targets
 - Brad Reger – Foxconn/Ingrasys
- How nebula uses DPU/SPUs in its "smartInfrastructure" solution
 - Jeff Feierfeil – Nebula
- Q/A and Panel Discussion
 - All

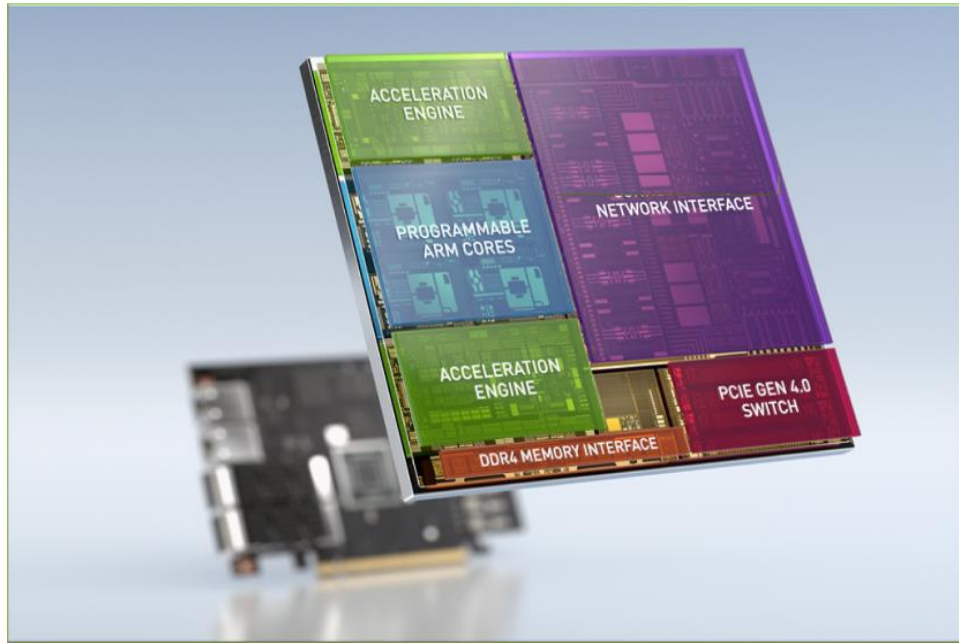


Teaser

Rob Davis

VP Storage Technology, NVIDIA Networking Platforms

DPU's are Designed to Accelerate Most Data Center Workloads



Offload

Accelerate

Isolate

Enterprise

Cloud Con
Bare Metal as



Flash Based Storage



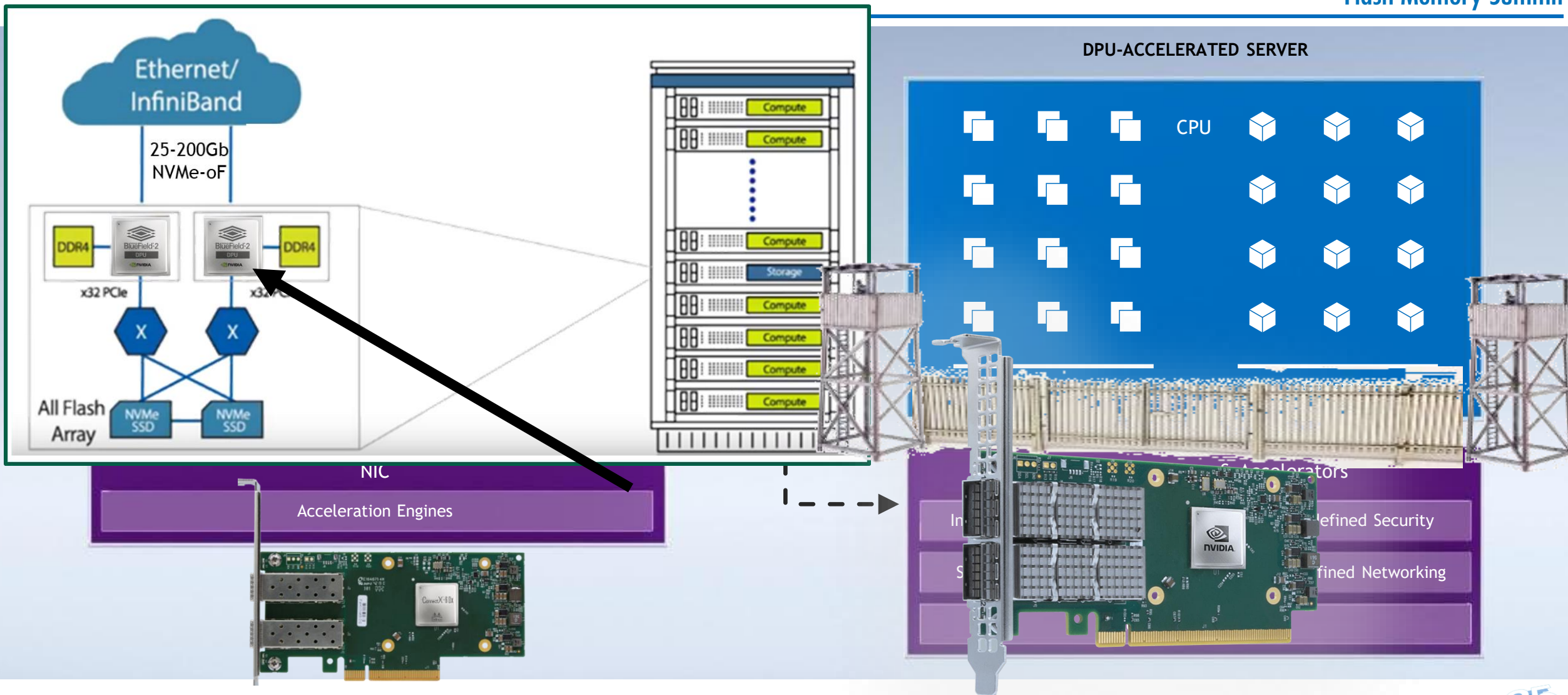
Cloud Gaming



5G
trial/Core/Edge



How: DPUs Offload, Accelerate and Secure Storage

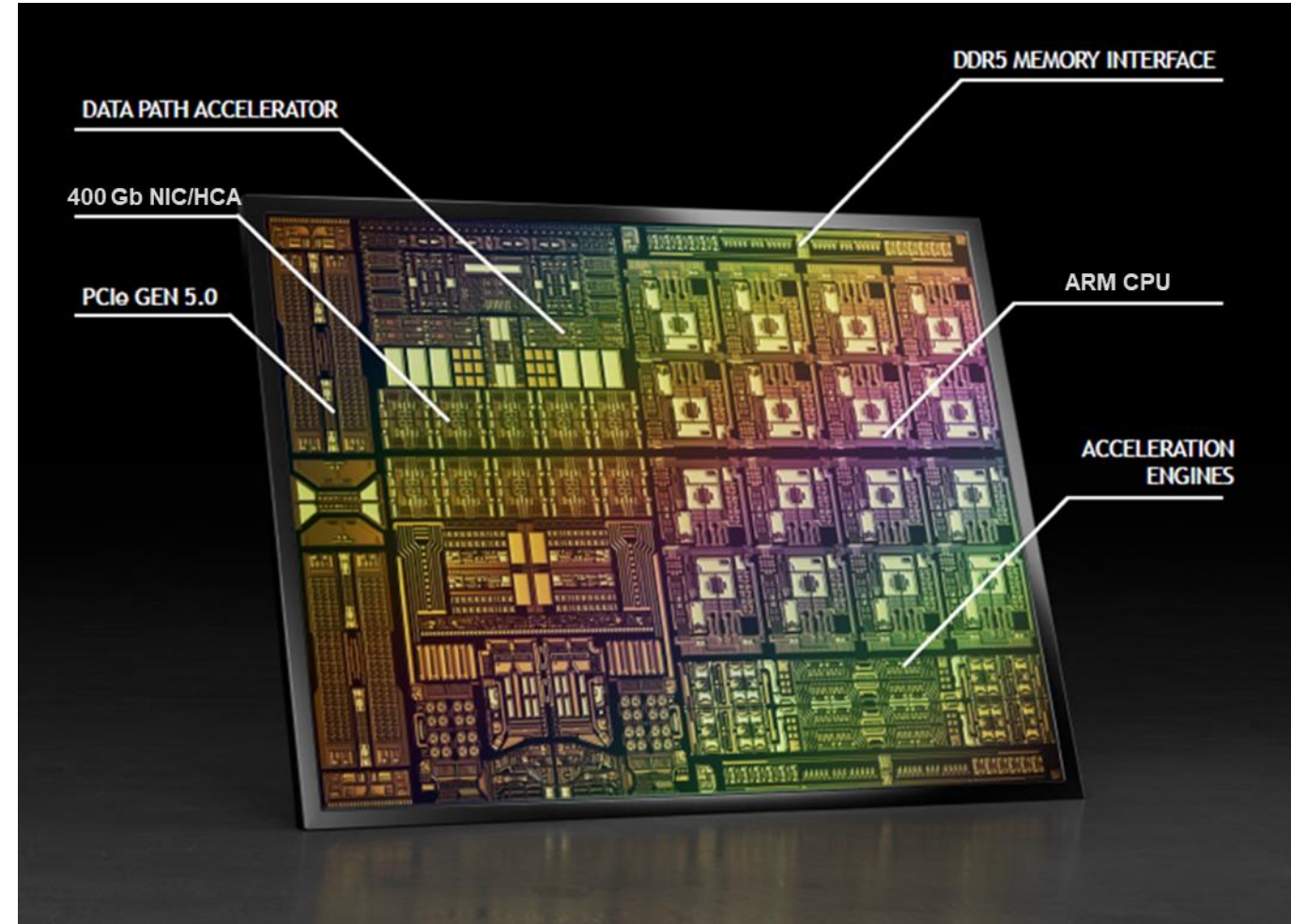


General DPU Architecture



Flash Memory Summit

- CPU
 - ARM
- HW Offloads
 - CPU cycle intensive or latency sensitive functions
- IO
 - PCIe, Ethernet, (IB)
- SDK
 - OPI



Some Storage Companies Implementing DPU Storage Solutions



AIC



AirMettle



DELL
Technologies



IBM



KIOXIA

mercury



nebulon



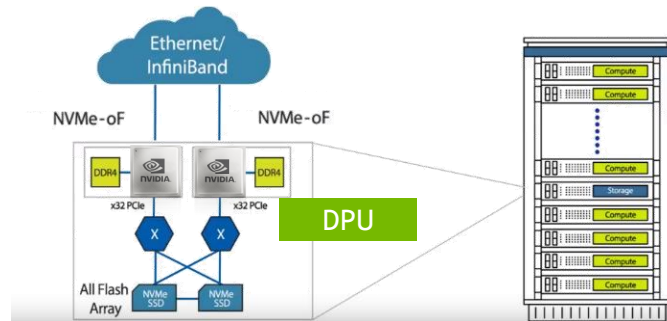
SAMSUNG



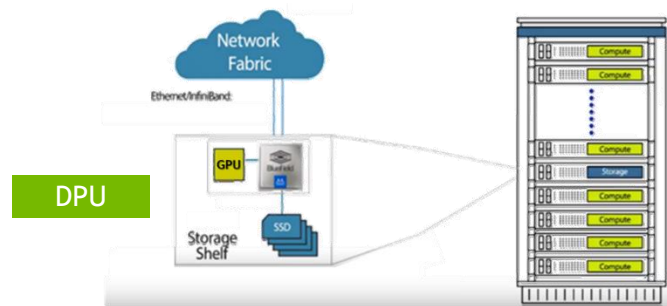
DPU Storage Use Cases



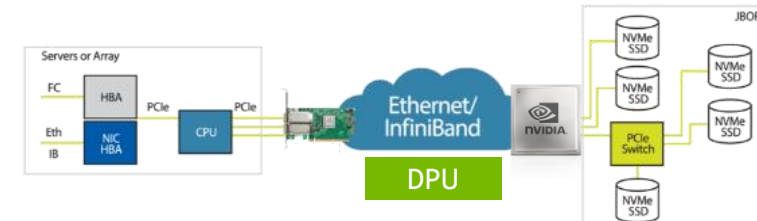
Flash Memory Summit



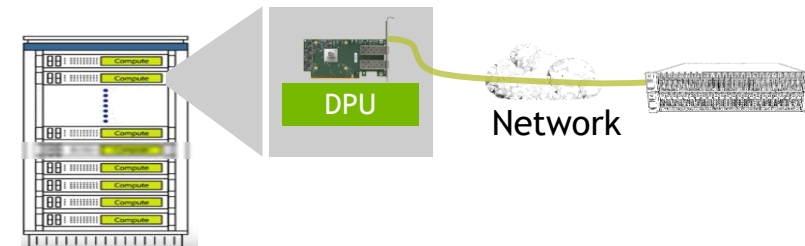
All Flash Array/JBOF for Storage Area Networks



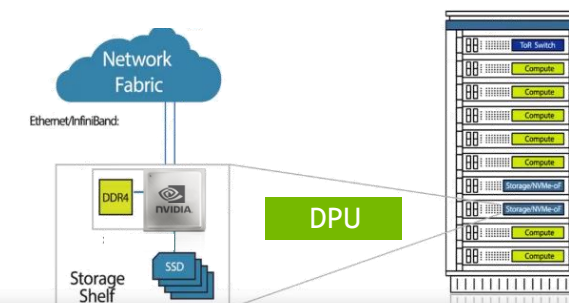
Computational Storage



Backend NVMe-oF Cluster



Server Based Storage / SDS



Compute and Storage Disaggregation



Flash Memory Summit

DPU for Storage: An Introduction

Joseph L White, Fellow/VP, Dell Technologies

Agenda



Flash Memory Summit

- DPU Terminology and Overview
- DPU Open Standards
- NVMe-oF Refresher
- DPUs for Storage and Storage Access
 - Offloads
 - Acceleration
 - JBOF/EBOF

DPU Terminology: xPU where 'x' stands for...



SmartNIC	SmartNIC (Intelligent NIC has also been used)
DPU	Data Processing Unit
IPU	Infrastructure Processing Unit
FAC	Function Accelerator Card
NAPU	Network Attached Processing Unit
NPU	Neural Processing Unit
TPU	Tensor Processing Unit
GPU	Graphics Processing Unit
CPU	Central Processing Unit
NPU	Network Processing Unit
APU	Application Processing Unit (GPU + CPU)

These are the DPU/xPUs
Industry still working through
common terminology

vector, matrix, and tensor
processing accelerators

General purpose compute

Networking

**These already
have good
definitions and
semantics**

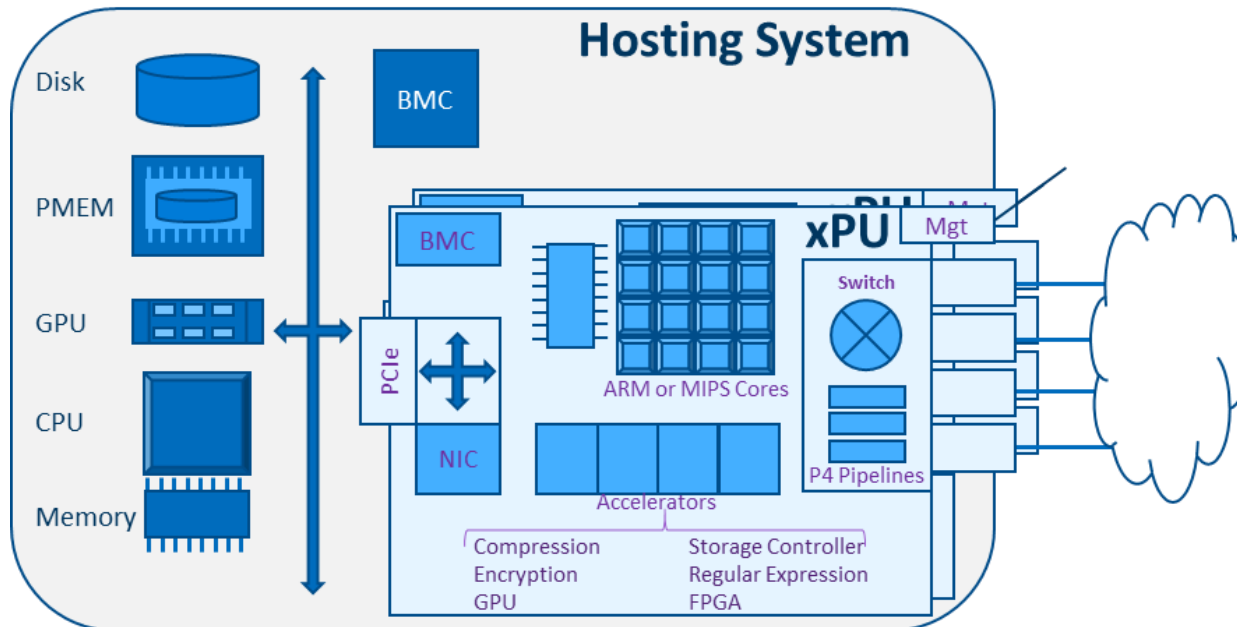
Devices with a tightly coupled combinations of CPU, xPU, GPU, etc. will exist.
The taxonomy gets interesting at that point...

DPU: Overall Architecture

DPU - Data Processing Unit (aka xPU)

Effectively a micro-server optimized for dataflow and packet processing providing accelerators, offload engines, & local services

Presents virtual functions to a host (looks like a NIC, GPU, etc)



DPU Internal Components

- General Purpose CPU Cores with Memory
- PCIe Interface
- Network Interfaces (Data and Management)
- Local Switching
- Accelerators & Offloads
- Programmable Pipelines
- Embedded BMC

Server Architecture

- DPU typically a built as a PCIe Card (>1 allowed)
- Other instantiations like switch embedded or standalone possible
- DPU presents conventional PCIe functions to hosting servers
- DPU can directly access PCIe Devices

DPU Operating System

- Linux (N flavors, Ubuntu/Debian is common)
- VMware
- proprietary

Common Tool Chains Apply

- System configuration and management
- Network configuration and management

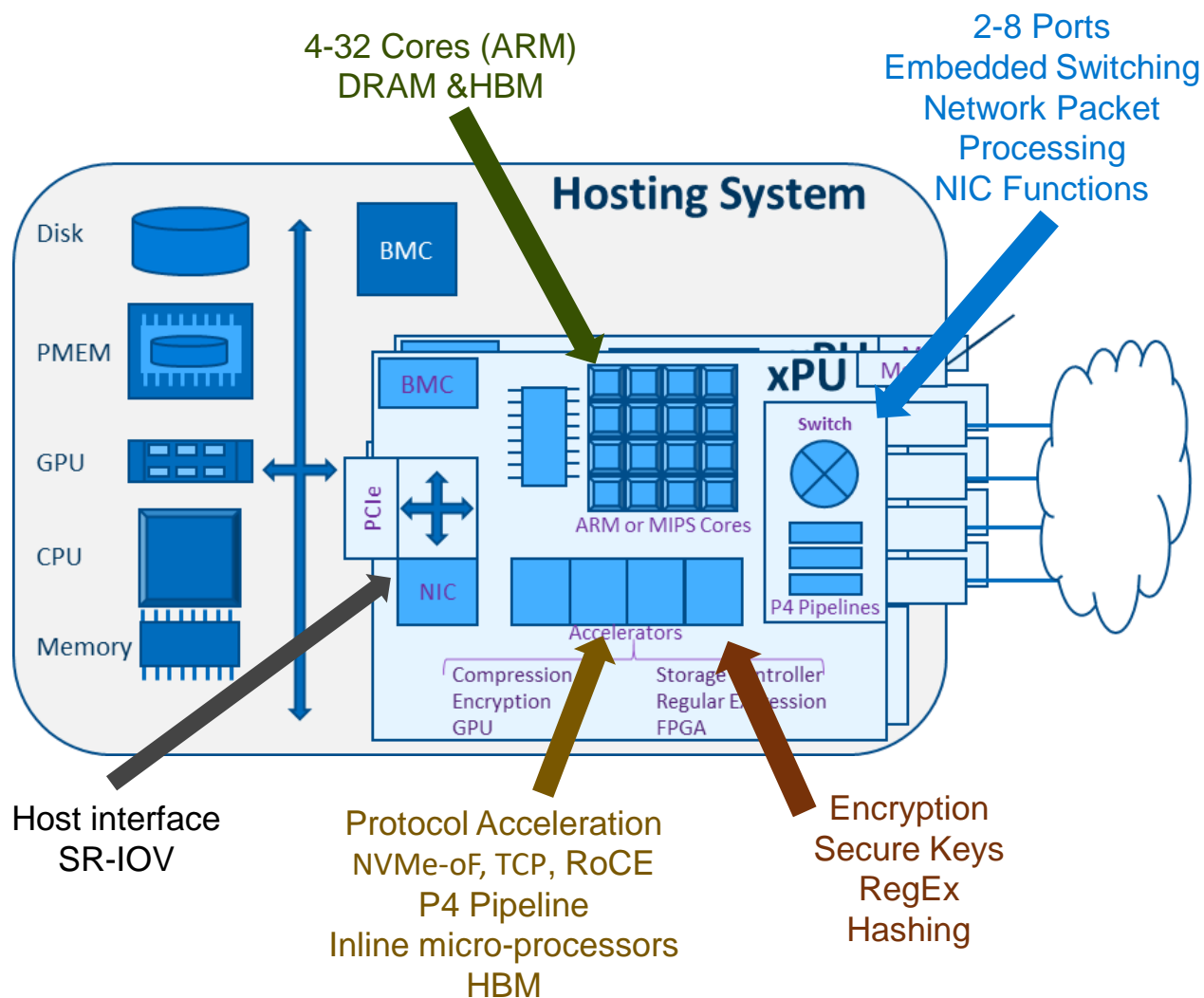
K8s

- container installation and management

DPU Functions



Flash Memory Summit



Core Acceleration & Offload Functions

- Network Switching
- Network Connectivity
- Gateway
- Storage Connectivity including NVMe/TCP, NVMe/RoCE
- Storage Services
- Expose Hosting System Resources
- Security (Firewall, DPI, Key Management, Intrusion Detection/Protection, Host Isolation)
- Telemetry Collection and Processing
- Control Plane Offload & Services
- Hypervisor
- CNF/NFV Hosting
- Provide Accelerators/Co-processing to Hosting system
- Boot and provisioning

Open Programmable Infrastructure Project



Community Driven Open Ecosystem for frameworks based on DPU enabled systems

- Originated as Diamond Bluff in late 2021
 - rapid growth in 2022, 20+ companies, 100+ individuals
- OPI Project is now a Linux Foundation project
 - <https://opiproject.org/>
 - <https://github.com/opiproject>
 - Announcement 21-JUN-2022

Current Working Areas

Organization & Administration + Legal/Governance

Vision Statement/Goals + requirements

Events, Outreach, Orientation

Provisioning and Platform Management

Open Programmable Infrastructure API and Behavioral Model

Use Case

Developer Platform/POC/Reference Platform

“Founding members of OPI include Dell Technologies, F5, Intel, Keysight Technologies, Marvell, NVIDIA and Red Hat with a growing number of contributors representing a broad range of leading companies in their fields ranging from silicon and device manufactures, ISVs, test and measurement partners, OEMs to end users.”

OPI Project Goals

- Create community-driven standards-based open ecosystem for DPU/IPU-like technologies
- Create vendor agnostic framework and architecture for DPU/IPU-based software stacks
- Reuse existing or define a set of new common APIs for DPU/IPU-like technologies when required
- Provide implementation examples to validate the architectures/APIs

OPI Project Deliverables

- Open Source Projects
- Specifications/Standards
- Reference Platforms
- Test Suites & Cases
- POC/Prototypes

DPU Open API Scope



Flash Memory Summit

• System

- Systems Management & Lifecycle
 - (Redfish, BMC, etc.)
- Monitoring, Metering, & Telemetry

• Operating System (Linux)

- Standard Linux Libraries and packages
- Container and Application Hosting
- Leverage commonly used APIs
 - DPDK, SPDK, EBPF

• Hardware (PCIe...)

- Virtual Function Mapping
- Offload Configuration

• Low Level (likely Vendor specific APIs)

- Micro-Code in Data Flow Processing Cores
- P4 Packet Processing Pipelines

• Vendor Unique API & SDK

- *These are NOT common/Open APIs*
- IPDK, DOCA, ASAP2, SNAP

• Storage

- Networked Storage
 - NVMe/TCP
 - NVMe/RoCE(RDMA)
- Storage Services
 - RAID/Erasure Coding/etc
- Compression
- SDXI Offload

• Networking

- **SONiC**
 - OpenConfig (includes BGP, etc)
 - SAI implementation by the DPU
- Policing and QoS and SLA
- Multi-tenant Overlay
- Host facing NIC Configurations
- OVS

• Gateway

- Connection Tracking
- Load Balancing
- NAT
- Tunnels

• Security

- Policy & Filters
- Crypto Offloads
- Secure Storage
 - keys, secrets, attestation, ...
- Key Management
- Network security offload
 - (TLS, IPSec)
- RegEx matching

NVMe-oF Storage Networking



Flash Memory Summit

Accelerating the evolution of storage connectivity with a modern, automated and secure NVMe IP SAN **Ecosystem**

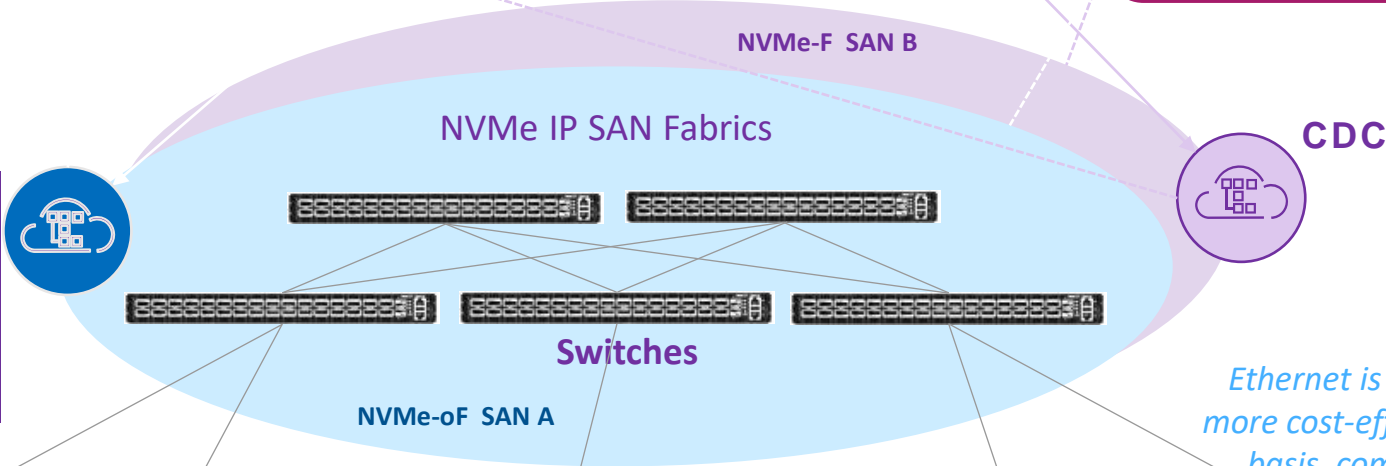


Central Discovery Controller

Discovery
Registration
NameService

ZoneService
State Change Notifications
Connectivity Management

Deploy as: VM
Containers
Switch Embedded



Ethernet is about an order of magnitude more cost-effective on a per unit bandwidth basis compared to alternative fabrics

Rack Servers

Chassis Servers

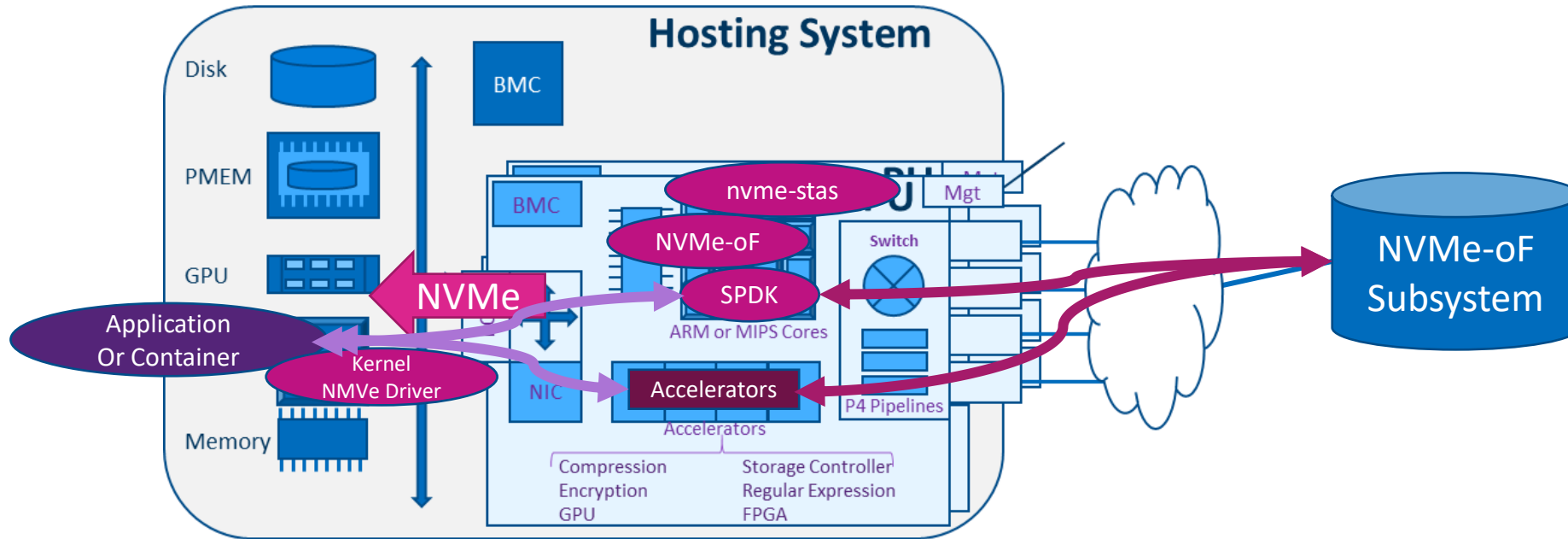
Storage Array

EBOF

JBOF



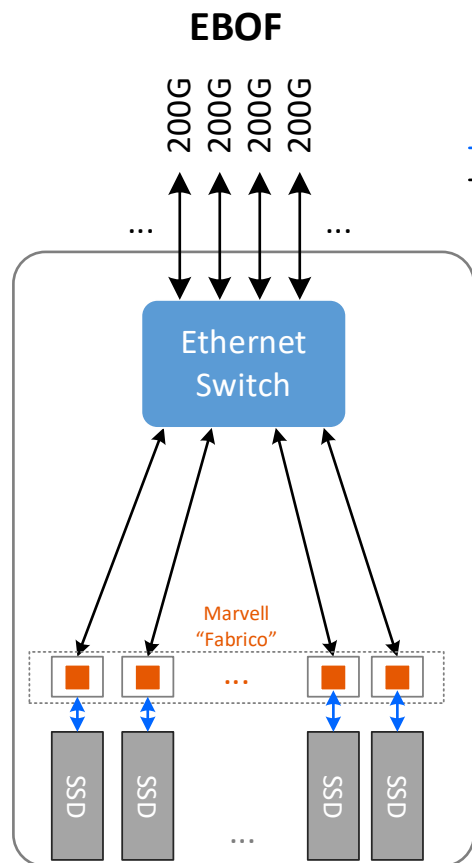
DPU NVMe-oF Host Offloads



NVMe-oF Host Offload for Network Storage Access

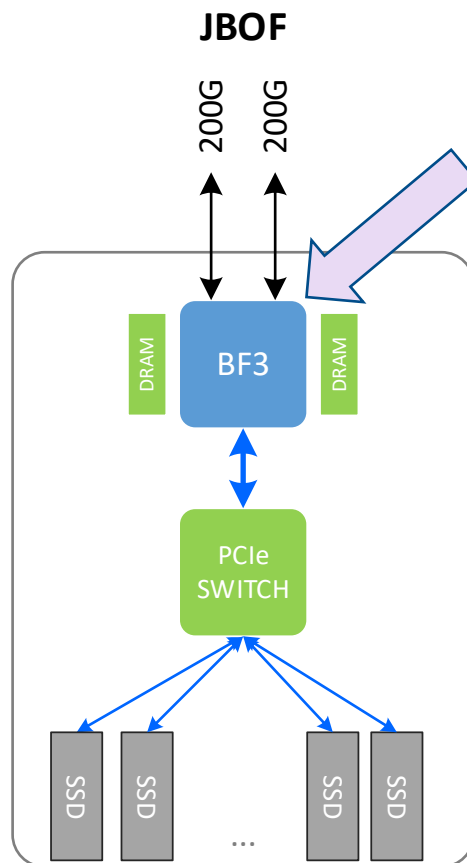
- Supports Bare-metal, VM, and Container deployments
- Encryption & HASH/CRC
- Optional: compression, RAID, networking integration

xBOF: Technology Options and Subsystem offloads

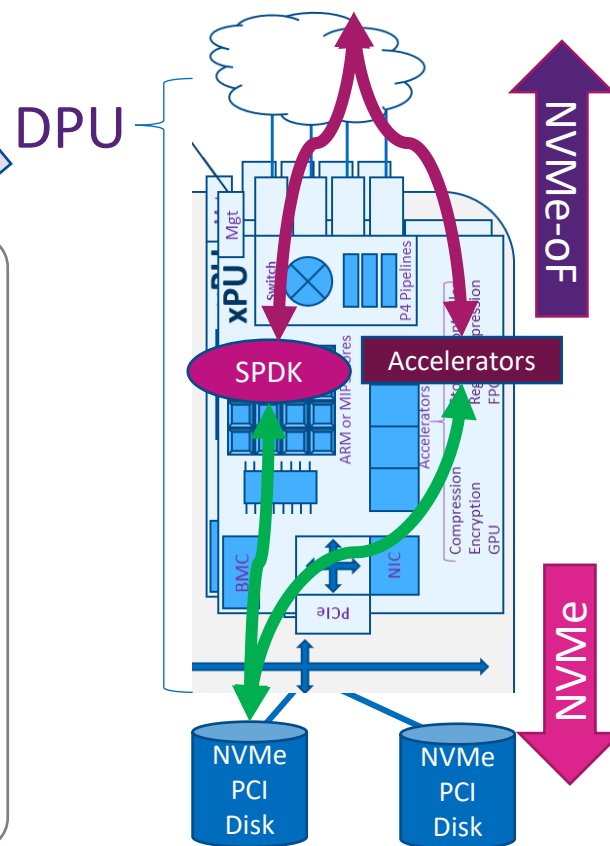


Marvell "Fabrics"
Access Module

— PCIe
— Ethernet



Nvidia "Bluefield"
Access Module



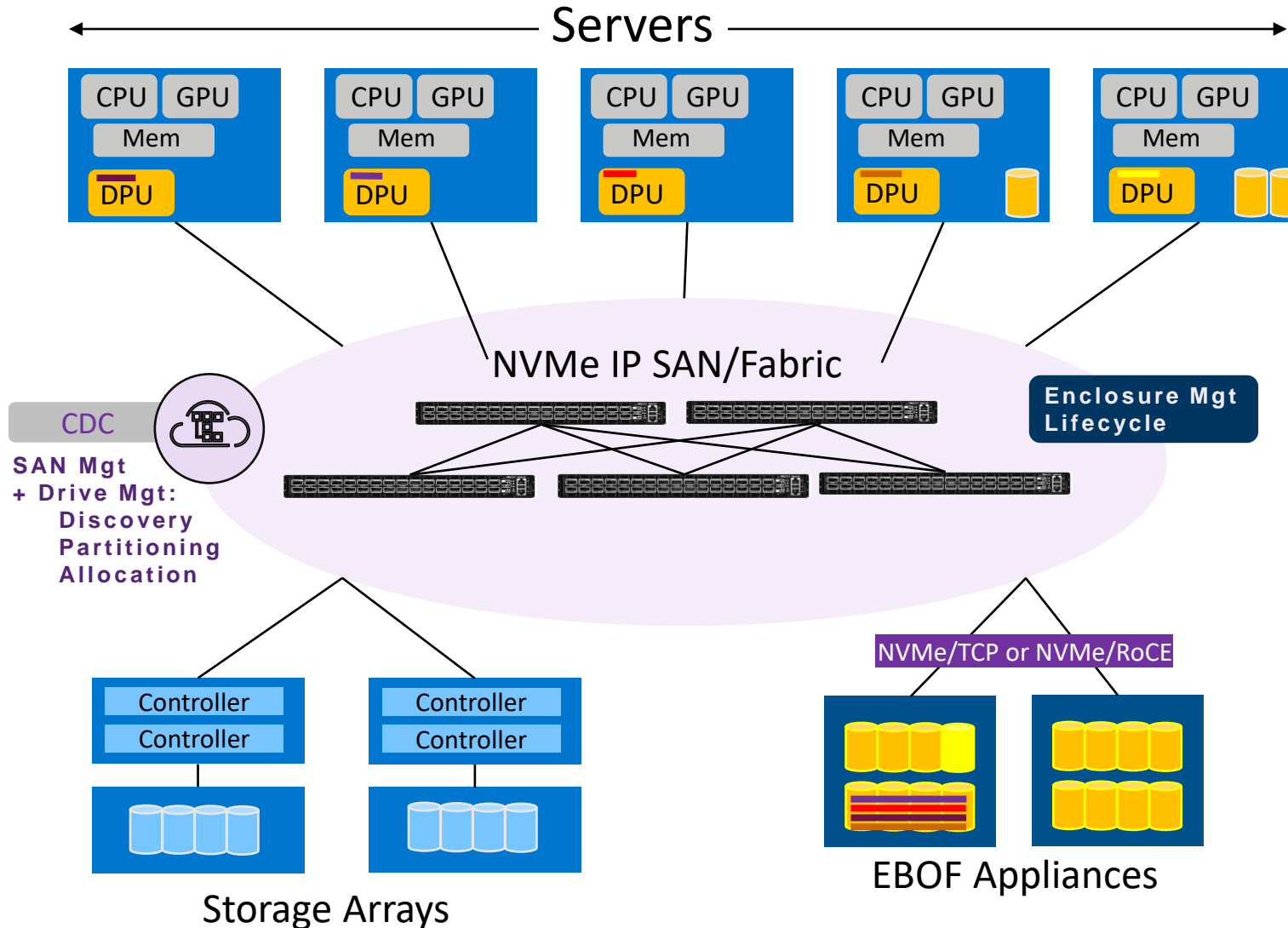
NVMe-oF Sub-system

- Compression
- Encryption
- HW/SW split depends on offloads available and protocol details
- Acceleration can be extended to storage arrays or Software Defined Storage

End to End DPUs & Storage



Flash Memory Summit



IP NVMe-oF SAN

- SFSS + SFS + OMNI



JBOF & EBOF & Arrays

- Disk Level Disaggregated Storage



DPU Offloads & Software Defined Storage

- NVMe/TCP & NVMe/RoCE
- RAID / Erasure Coding → drive aggregation
- Compression
- Data Traffic Encryption
- Software Defined Storage
- File System



Flash Memory Summit

Disaggregated Storage Architectures with DPUs and Next-Gen JBODs

John Mao

Global Head of Business Development

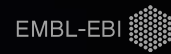
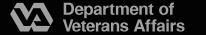
VAST Data



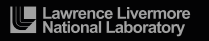
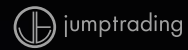
AQUATIC



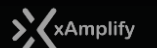
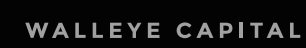
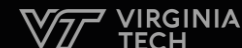
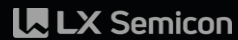
DARTMOUTH



FBI



EXABYTES OF VAST DATA



OUR MISSION

NO MORE *TIER*S

BREAKING TRADEOFFS TO MAKE STORAGE SIMPLE

ALL-NVMe PERFORMANCE

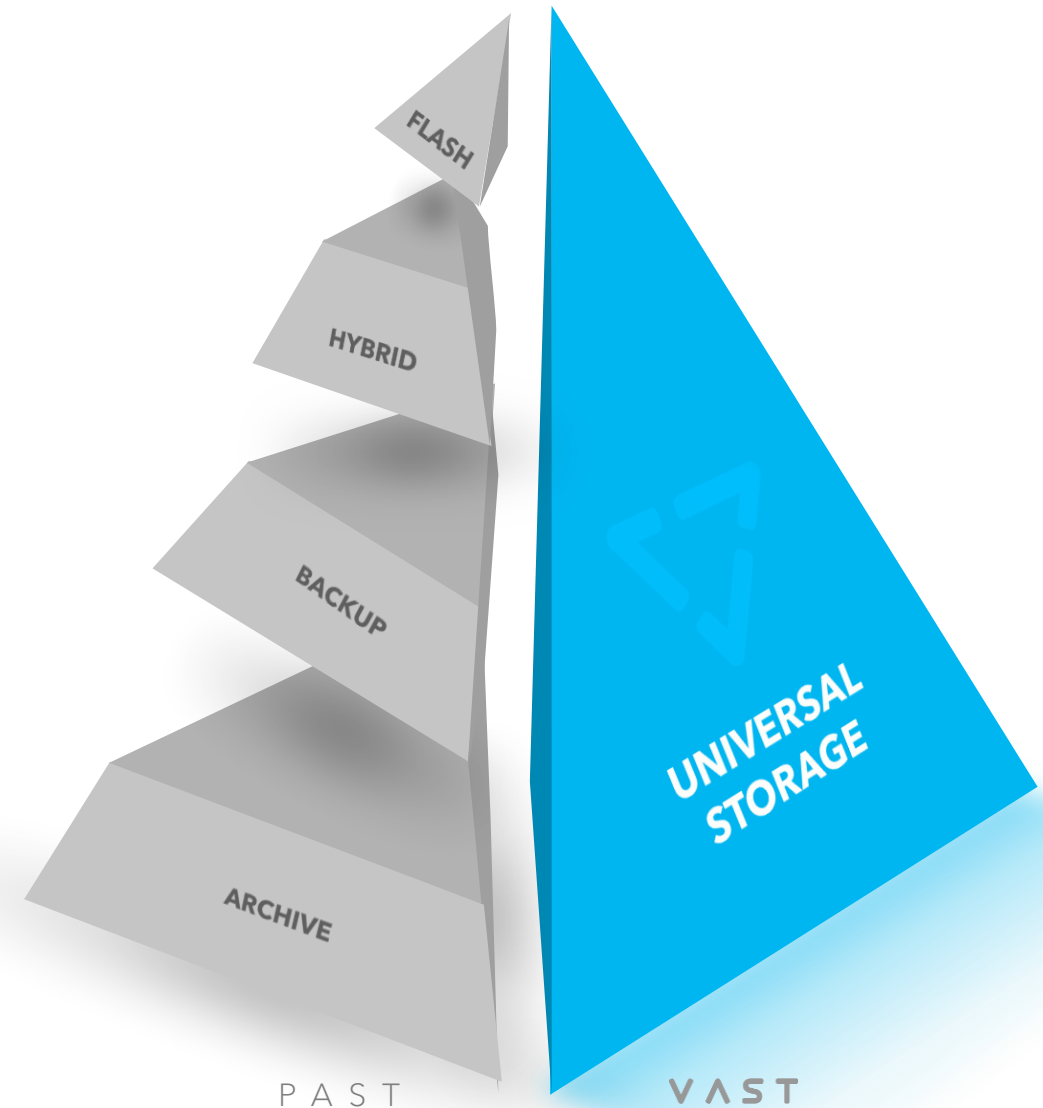
Eliminate application bottlenecks

EXABYTE SCALE, WITH ARCHIVE ECONOMICS

Eliminate the HDD with a single-tier flash cloud

ENTERPRISE NAS & OBJECT STORAGE

Standard interfaces with game-changing performance



V A S T

UNIVERSAL STORAGE

STARTING FROM A RENAISSANCE IN HARDWARE



NVME OVER FABRICS FOR DISAGGREGATION

The latency of DAS, over switched commodity networks.



QLC FLASH FOR COST SAVINGS

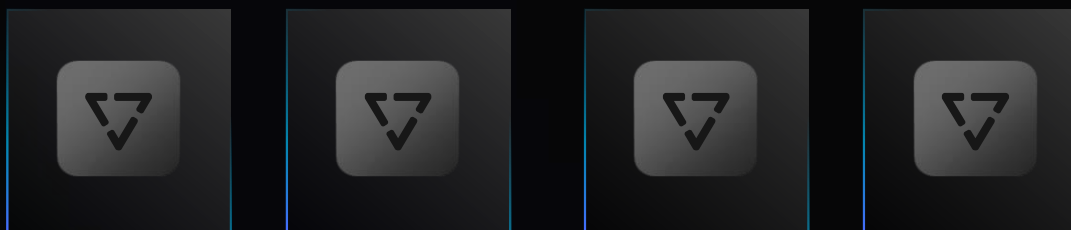
Low-cost, lower-endurance hyperscale flash.



STORAGE CLASS MEMORY LONGEVITY & EFFICIENCY

Enables write shaping to QLC and rich metadata.

DISAGGREGATED & SHARED-EVERYTHING



STATELESS LOGIC

Never Rebuild On Server Failure

NO CODE INTERDEPENDENCIES

Isolates Bugs & Bad Actors



DENSE QLC DRIVES

Saves Money

ENABLES WIDE RESILIENT STRIPES

36+4 To 146+4 To Save Money





VAST

VAST CBOX

- 4 x CNode Container Servers in 2U
- Up to 38GB/s read, 5.5GB/s write
- Up to 200K read IOPS, 90K write IOPS
- 800W Active

NVMe SWITCHES

- 16, 32 or 64 port switches
- Ethernet or InfiniBand options
- Large fabrics built from spine+leaf topology

VAST DBOX

- 675TB or 1,350TB JBOF enclosure, fault-tolerant
- Up to 42GB/s read, 5GB/s write
- Up to 450K read IOPS, 200K write IOPS
- 1.4kW Active

VAST & NVIDIA BlueField DPUs



INTRODUCING **CERES**

[seer-eez]





DESIGN POINTS

EASIER SERVICEABILITY

No more sliding systems in and out of racks, better cable management

GREATER MODULARITY

Lower entry-point pricing, lower-cost path to DBox HA

HIGHER DENSITY

Lay a foundation for even higher density drives

ADVANCED PERFORMANCE

Move to advanced versions of PCIe

RULERS ARE THE FUTURE



EDSFF E1.L
Long Ruler

FRONT-LOADED

Easy to service, lessens reliance on cable management

DEEPER THAN U.2

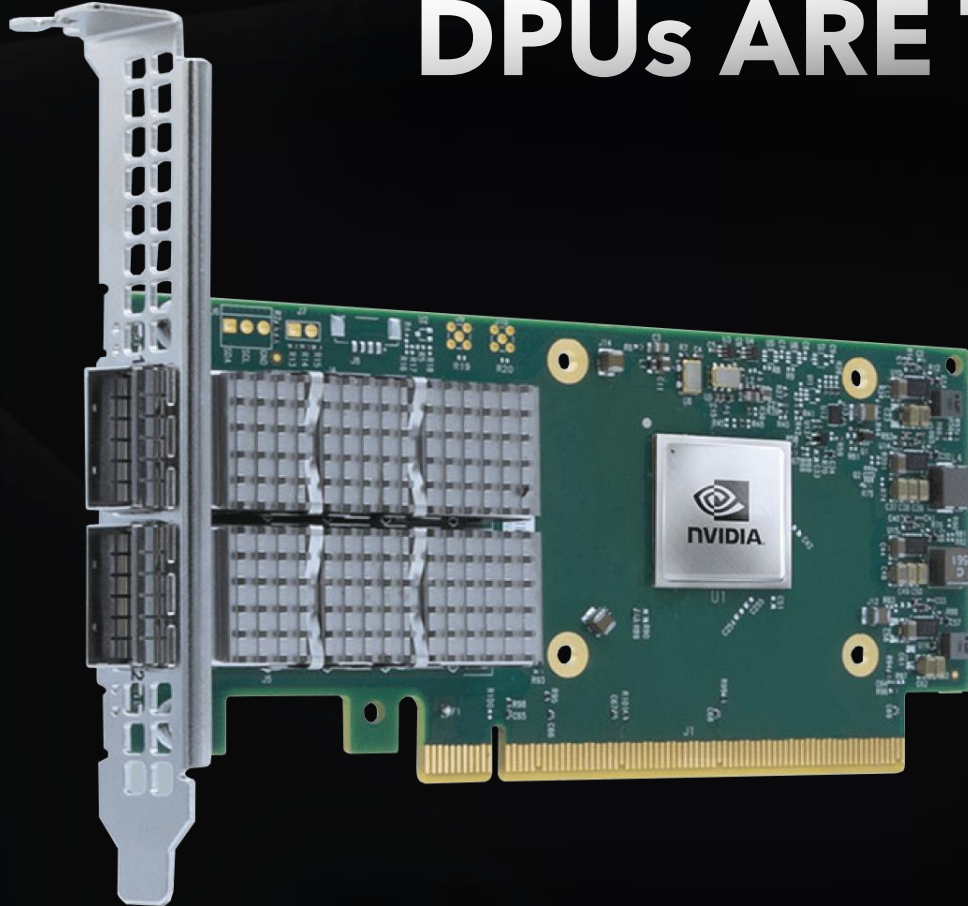
More drive footprint per rack unit

THERMALLY-EFFICIENT

Provides us some potential upside



DPU_s ARE THE FUTURE



BLUEFIELD-1
BY NVIDIA

INTEGRATED NIC & TARGET

Less (moving) parts, less cost

ARM PROCESSORS

Lower power per DBox

TEENY TINY

Makes for nice 1Uing



Up to 2 x NVIDIA BLUEFIELD-1 DPUs PER TRAY

4 x 100Gb ETH/IB per tray

8 x HOT-SWAP SCM DRIVES

Gen4 U.2 SSDs on Sleds

22 x HIGH-CAPACITY QLC RULERS

15TB or 30TB Gen4 NVMe E1.L SSDs
338TB or 675TB in 1U

REDUNDANT PCI GEN4 SWITCHES

Enables single-ported drive support

2 x HOTSWAP DNODE TRAYS

Dual I/O modules for high-availability

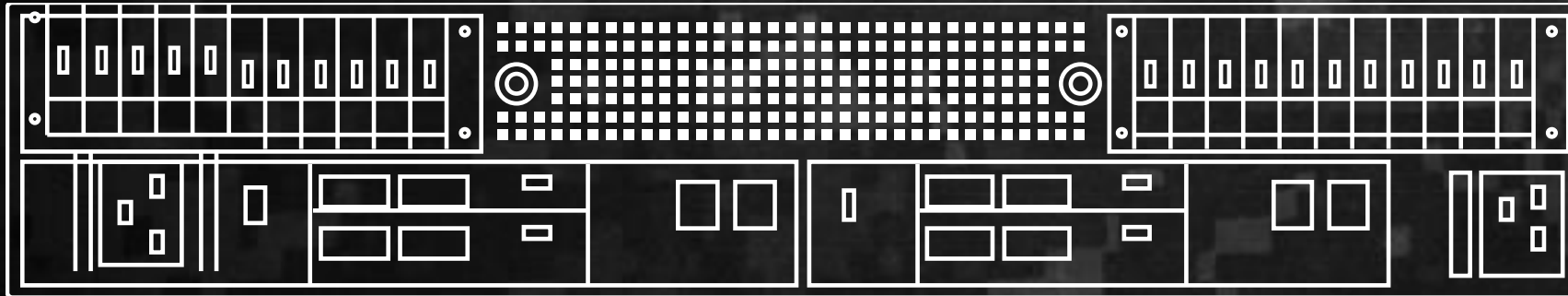


CERES
1U NVMe-oF JBOF

	AIC DF-3015	AIC DF-3030
CAPACITY: RAW/USABLE@SCALE	338TB / 299TB	675TB / 598TB
EFFECTIVE @ 3:1 REDUCTION	897TB	1.79PB
CONNECTIVITY	4x100Gb • 8x100Gb • IB/ETH	
R•W THROUGHPUT (GB/s)	Up to 64GB/s Reads • Up to 10GB/s Writes	
ACTIVE POWER (AI WORKLOADS)	500W	



8 x HOT-SWAP SCM DRIVES
Gen4 U.2 SSDs on Sleds



22 x HIGH-CAPACITY QLC RULERS
15TB or 30TB Gen4 NVMe SSDs

2 x BLUEFIELD-1 SMART NICs PER TRAY
4 x 100Gb ETH/IB per tray

mercury

CERES DBOX
1U NVMe-oF JBOF

Key Benefits of DPUs (for VAST)

- **Better Physical Density**
Full Active-Active IO Modules in 1RU → support for new ruler SSDs
- **Improved Performance (Density)**
Next-gen PCIe with faster networking → 2x performance density (over previous gen)
- **Lower (Hardware) TCO**
Eliminates “brawny” CPU cores and DRAM in IO Modules → lower cost and power

Future Plans & Considerations

- **Transition to BlueField-3 DPUs**
400Gb NDR and GbE support, security and more storage accelerators
- **Evaluate applicability in VAST's "Cnode" architecture**
End-to-end NVMe-oF advantages? Filesystem software accelerators?
- **Client/App Side DPUs + VAST Software**
Integrate VAST filesystem software into client DPUs for seamless IO scalability?
(ie. GPU servers equipped with DPUs supporting RDMA storage access)



THANK YOU



Flash Memory Summit

Interactive Analytics & AI

Flash Storage service made possible by DPUs

By: Donpaul Stephens

Founder & CEO,  AirMettle



AirMettle's Mission



Flash Memory Summit

- AirMettle is developing a real-time smart data lake solution that simplifies big data analytics and accelerates processing by an order of magnitude, or more.
- It is delivered as an intelligent data lake storage service which performs basic analytics tasks that:
 - Reduce network traffic
 - Improve data freshness,
 - Enable real-time operation.



AirMettle's Mission



Flash Memory Summit

- AirMettle is developing a real-time smart data lake solution that simplifies big data analytics and accelerates processing by an order of magnitude, or more.
- It is delivered as an intelligent data lake storage service which performs basic analytics tasks that:
 - Reduce network traffic
 - Improve data freshness,
 - Enable real-time operation.

DPU Supercharges it,
Let's see how

Flash storage platform options



Flash Memory Summit

Traditional Storage Chassis



1 Node



2 CPUs



24 DIMMs



48 Drives



Flash storage platform options



Flash Memory Summit

Traditional Storage Chassis



1 Node



2 CPUs



24 DIMMs



48 Drives



Hot

It's the latest rage (for roughly last decade)

Popular for Software Defined Storage

Flash storage platform options



Flash Memory Summit

Traditional Storage Chassis



1 Node



2 CPUs



24 DIMMs



48 Drives



- Dual-socket storage chassis
 - Amortizes CPU & DRAM over large # of SSDs
- Challenges:
 - Large fault-domain (48 SSDs!!!)
 - Limited network bandwidth & processing / SSD
 - Traditionally little to no analytics in storage
 - High power / heat density

Flash storage platform options



Flash Memory Summit

Traditional Storage Chassis



1 Node



2 CPUs



24 DIMMs

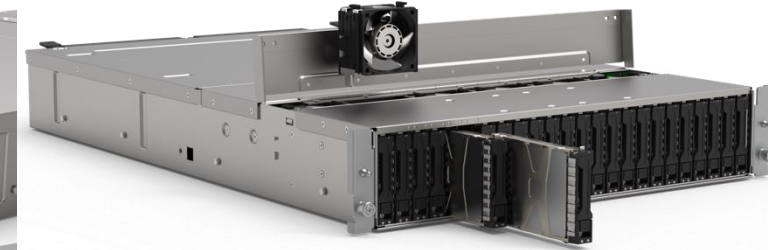
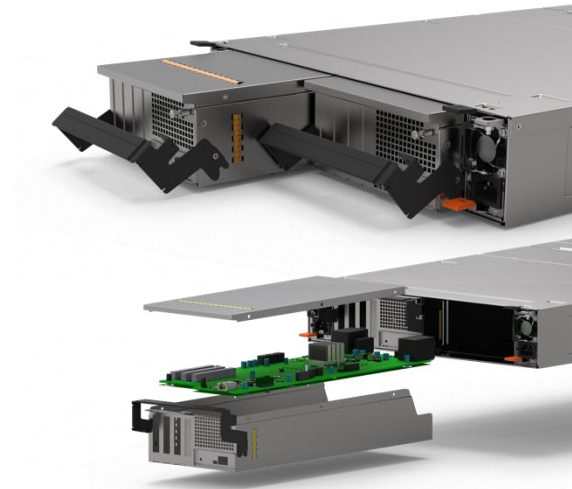


48 Drives



- Dual-socket storage chassis
 - Amortizes CPU & DRAM over large # of SSDs
- Challenges:
 - Large fault-domain (48 SSDs!!!)
 - Limited network bandwidth & processing / SSD
 - Traditionally little to no analytics in storage
 - High power / heat density

DPU empowered JBOF!



Flash storage platform options



Flash Memory Summit

Traditional Storage Chassis



1 Node



2 CPUs



24 DIMMs

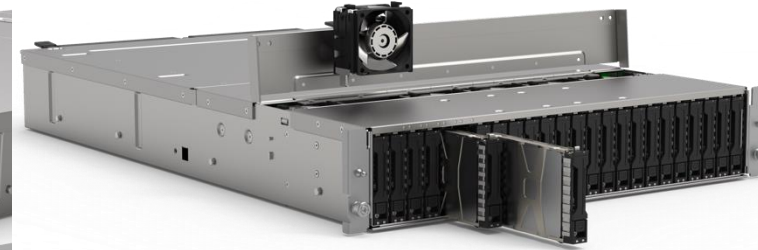
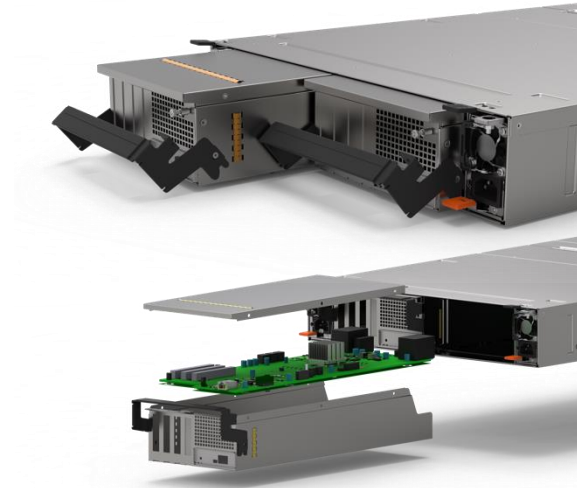


48 Drives



- Dual-socket storage chassis
 - Amortizes CPU & DRAM over large # of SSDs
- Challenges:
 - Large fault-domain (48 SSDs!!!)
 - Limited network bandwidth & processing / SSD
 - Traditionally little to no analytics in storage
 - High power / heat density

DPU empowered JBOF!



Cool

DPU & Storage Class Memory
Enable a new Era for Big Data

Flash storage platform options



Flash Memory Summit

Traditional Storage Chassis



1 Node



2 CPUs



24 DIMMs

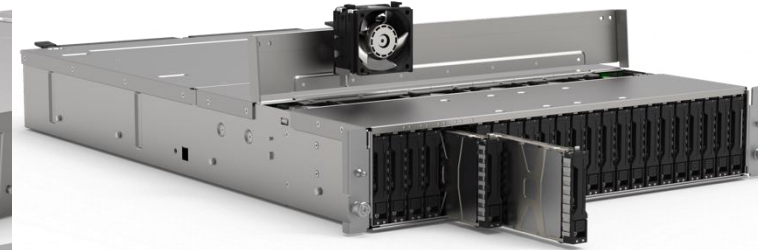
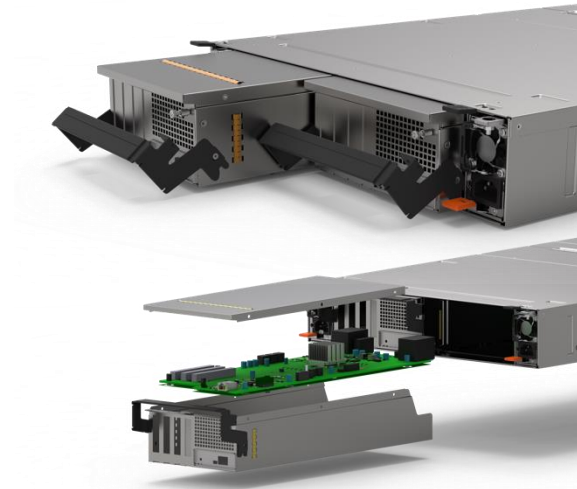


48 Drives



- Dual-socket storage chassis
 - Amortizes CPU & DRAM over large # of SSDs
- Challenges:
 - Large fault-domain (48 SSDs!!!)
 - Limited network bandwidth & processing / SSD
 - Traditionally little to no analytics in storage
 - High power / heat density

DPU empowered JBOF!



- NVMe-oF / “Just a Bunch of Flash” chassis
 - Neither traditional socket nor DRAM DIMMs
- Benefits:
 - Small fault-domain: 6 – 12 SSDs per DPU
 - High network bandwidth: low-latency
 - Maximum analytics performance
 - Optimum power density for conventional data centers



Look “under the hood”: Data Lakes today, and how we make better ones



Traditional Data Lake



Objects are internally partitioned
For storage in parallel

Data Lake

Traditional Data Lake:

Data generally arrives semi-structured

Comes
from
Everywhere



Objects are internally partitioned
For storage in parallel

Primarily
Semi-structured data

Data Lake

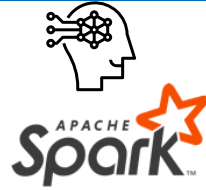
Traditional Data Lake:

Data must be moved to gain value from it



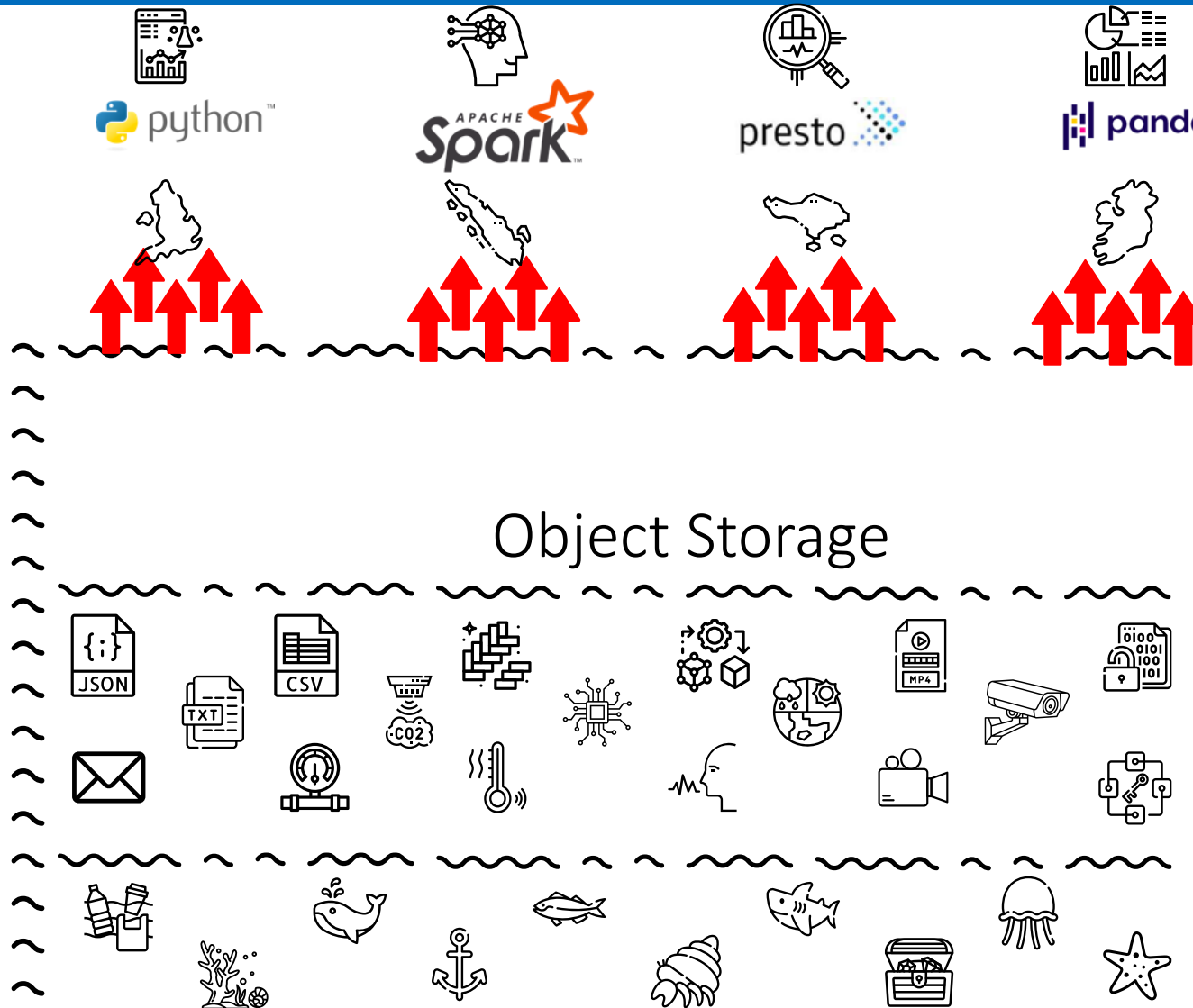
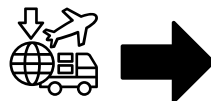
Flash Memory Summit

Analyzed
In Islands



Applications retrieve full objects*
To their own (small) clusters
for processing

Comes
from
Everywhere



Objects are internally partitioned
For storage in parallel

Primarily
Semi-structured data

Data Lake

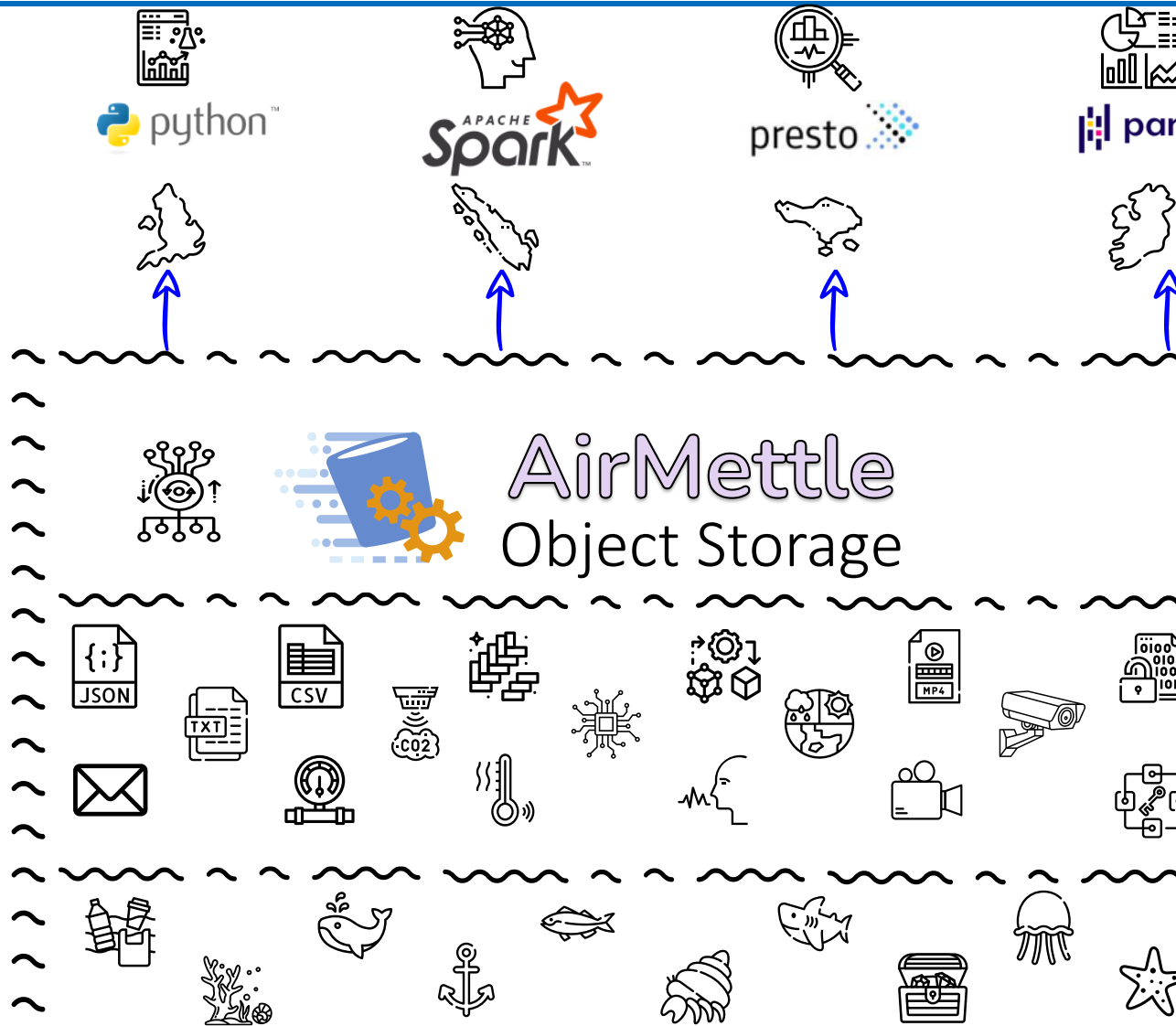
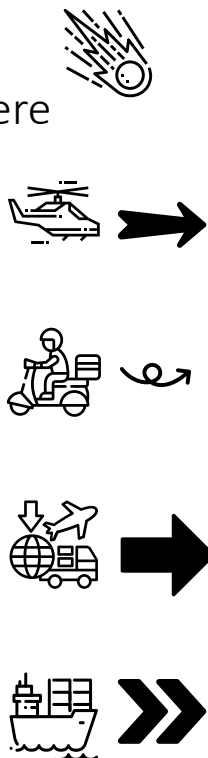
Smart Data Lake:

Load the desired subset or result, eliminate the need to move 90%+ of the data

Analyzed **Faster**
In Islands



Comes from
Everywhere



Retrieve what is needed in an
Immediately usable form

Objects are internally partitioned
For storage & **processing** in parallel

Primarily
Semi-structured data

Data Lake



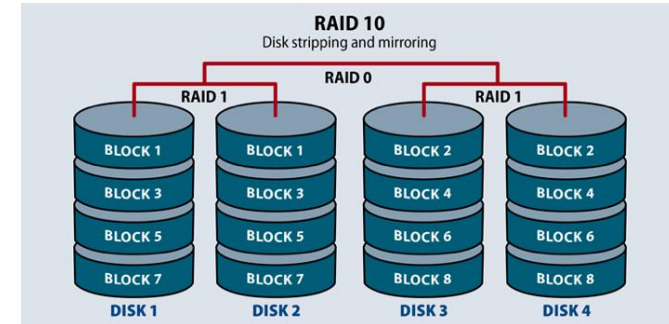
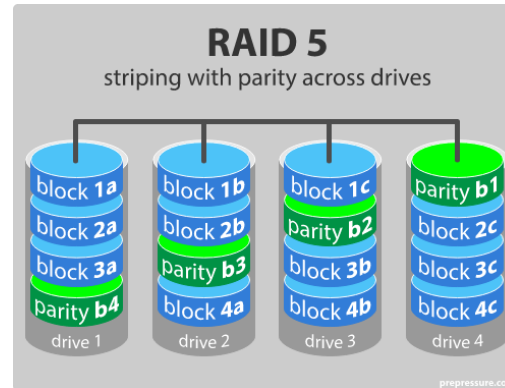
Processing in parallel within the storage service?

Legacy data protection makes this hard;
Let's see why

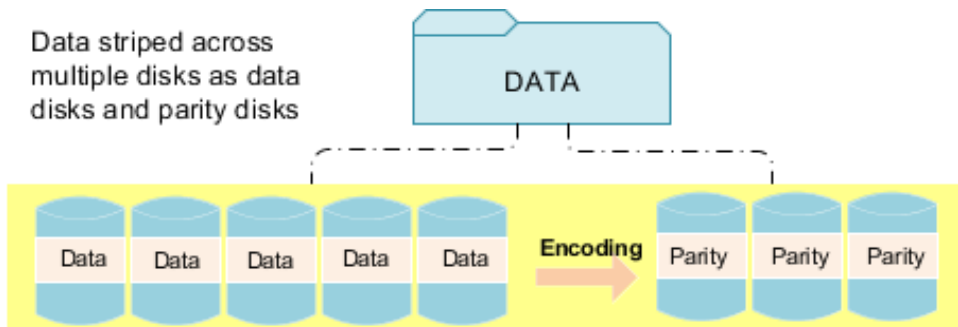
Data Protection 101:

How do storage solutions protect data?

RAID:



Erasure Coding:



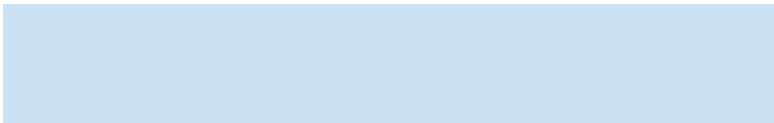
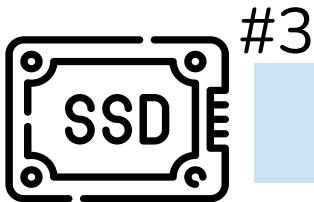
Data protection algorithms designed for HDD

What that means for data reliably placed in storage:

First 4 devices shown...

1	155190	7706	1	17	21168.23	0.04	0.02	N	O	3/13/96	2/12/96	3/22/96	DELIVER_IN_PERSON	TRUCK	egular_courts_above_the
1	67310	7311	2	36	45983.16	0.09	0.06	N	O	4/12/96	2/28/96	4/20/96	TAKE_BACK_RETURN	MAIL	ly_final_dependencies:_slyly_bold_
1	63700	3701	3	8	13309.6	0.1	0.02	N	O	1/29/96	3/5/96	1/31/96	TAKE_BACK_RETURN	REG_AIR	riously_regular _express_dep
1	2132	4633	4	28	28955.64	0.09	0.06	N	O	4/21/96	3/30/96	5/16/96	NONE	AIR	lites._fluffily_even_de
1	24027	1534	5	24	22824.48	0.1	0.04	N	O	3/30/96	3/14/96	4/1/96	NONE	FOB	_pending_foxes._slyly_re
1	15635	638	6	32	49620.16	0.07	0.02	N	O	1/30/96	2/7/96	2/3/96	DELIVER_IN_PERSON	MAIL	arefully_slyly_ex
2	106170	1191	1	38	44694.46	0	0.05	N	O	1/28/97	1/14/97	2/2/97	TAKE_BACK_RETURN	RAIL	ven_requests._deposits_breach_a

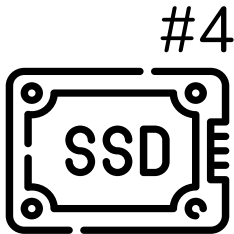
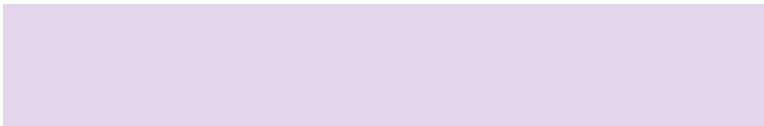
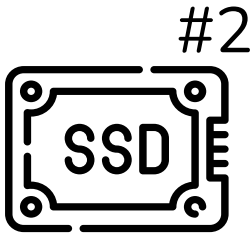
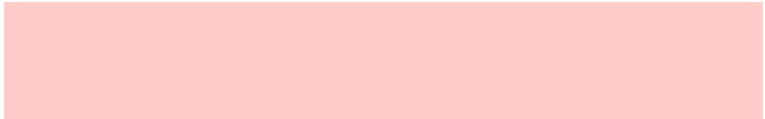
Simple Table:



Bytes of data divided evenly across SSDs!



Data protection and streaming performance!



Supports data protection algorithms designed for HDD!



What that means for data reliably placed in storage:

First 4 devices shown...

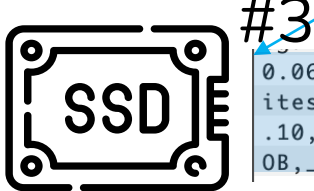
Simple Table:

1	155190	7706	1	17	21168.23	0.04	0.02	N	O	3/13/96	2/12/96	3/22/96	DELIVER_IN_PERSON	TRUCK	regular_courts_above_the
1	67310	7311	2	36	45983.16	0.09	0.06	N	O	4/12/96	2/28/96	4/20/96	TAKE_BACK_RETURN	MAIL	ly_final_dependencies:_slyly_bold_
1	63700	3701	3	8	13309.6	0.1	0.02	N	O	1/29/96	3/5/96	1/31/96	TAKE_BACK_RETURN	REG_AIR	riously_regular _express_dep
1	2132	4633	4	28	28955.64	0.09	0.06	N	O	4/21/96	3/30/96	5/16/96	NONE	AIR	lites._fluffily_even_de
1	24027	1534	5	24	22824.48	0.1	0.04	N	O	3/30/96	3/14/96	4/1/96	NONE	FOB	_pending_foxes._slyly_re
1	15635	638	6	32	49620.16	0.07	0.02	N	O	1/30/96	2/7/96	2/3/96	DELIVER_IN_PERSON	MAIL	arefully_slyly_ex
2	106170	1191	1	38	44694.46	0	0.05	N	O	1/28/97	1/14/97	2/2/97	TAKE_BACK_RETURN	RAIL	ven_requests._deposits_breach_a



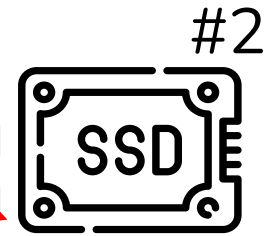
#1

1,155190,7706,1,17,21168.23,0.04,0.02,N,O,1996-03-13,1996-02-12,1996-03-22,DELIVER_IN_PERSON,TRUCK,regular_courts_above_the,1,67310,7311,2,36,45983.16,0.09,0.06,N,O,1996-04-12,1996-02-28,1996-04-20,TAKE_BACK_RETURN,MAIL,ly_final_dependencies:_slyly_bold_,1,63700,3701,3,8,13309.60,0.10,0.02,N,O,1996-01-29,1996-03-05,1996-01-31,TAKE_BACK_RETURN,REG_AIR,riously_regular|_express_dep,1,2132,4633,4,28,28955.64,0.09,0.06,N,O,1996-04-21,1996-03-30,1996-05-16,NONE,AIR,lites._fluffily_even_de,1,24027,1534,5,24,22824.48,0.10,0.04,N,O,1996-03-30,1996-03-14,1996-04-01,NONE,FOB,_pending_foxes._slyly_re,1,15635,638,6,32,49620.16,0.07,0.02,N,O,1996-01-30,1996-02-07,1996-02-03,DELIVER_IN_PERSON,MAIL,arefully_slyly_ex,2,106170,1191,1,38,44694.46,0.00,0.05,N,O,1997-01-28,1997-01-14,1997-02-02,TAKE_BACK_RETURN,RAIL,ven_requests._deposits_breach_a



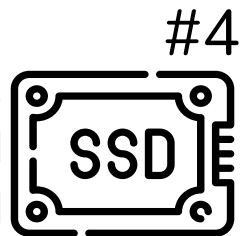
#3

0.06,N,O,1996-04-21,1996-03-30,1996-05-16,NONE,AIR,lites._fluffily_even_de,1,24027,1534,5,24,22824.48,0.10,0.04,N,O,1996-03-30,1996-03-14,1996-04-01,NONE,FOB,_pending_foxes._slyly_re,1,15635,638,6,32,49620.16,0.07,0.02,N,O,1996-01-30,1996-02-07,1996-02-03,DELIVER_IN_PERSON,MAIL,arefully_slyly_ex,2,106170,1191,1,38,44694.46,0.00,0.05,N,O,1997-01-28,1997-01-14,1997-02-02,TAKE_BACK_RETURN,RAIL,ven_requests._deposits_breach_a



#2

RETURN,MAIL,ly_final_dependencies:_slyly_bold_,1,63700,3701,3,8,13309.60,0.10,0.02,N,O,1996-01-29,1996-03-05,1996-01-31,TAKE_BACK_RETURN,REG_AIR,riously_regular|_express_dep,1,2132,4633,4,28,28955.64,0.09,0.06,N,O,1996-04-21,1996-03-30,1996-05-16,NONE,AIR,lites._fluffily_even_de,1,24027,1534,5,24,22824.48,0.10,0.04,N,O,1996-03-30,1996-03-14,1996-04-01,NONE,FOB,_pending_foxes._slyly_re,1,15635,638,6,32,49620.16,0.07,0.02,N,O,1996-01-30,1996-02-07,1996-02-03,DELIVER_IN_PERSON,MAIL,arefully_slyly_ex,2,106170,1191,1,38,44694.46,0.00,0.05,N,O,1997-01-28,1997-01-14,1997-02-02,TAKE_BACK_RETURN,RAIL,ven_requests._deposits_breach_a



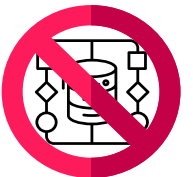
#4

16,0.07,0.02,N,O,1996-01-30,1996-02-07,1996-02-03,DELIVER_IN_PERSON,MAIL,arefully_slyly_ex,2,106170,1191,1,38,44694.46,0.00,0.05,N,O,1997-01-28,1997-01-14,1997-02-02,TAKE_BACK_RETURN,RAIL,ven_requests._deposits_breach_a

Bytes of data divided evenly across SSDs!



Data protection and streaming performance!



HDD-centric RAID/Erasure Coding prevent in-storage analytics

Partitioning, Processing, and Protecting data



Flash Memory Summit



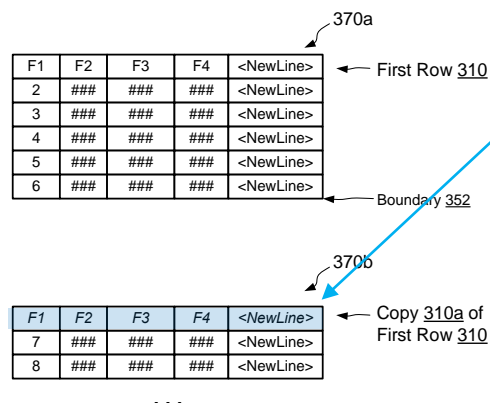
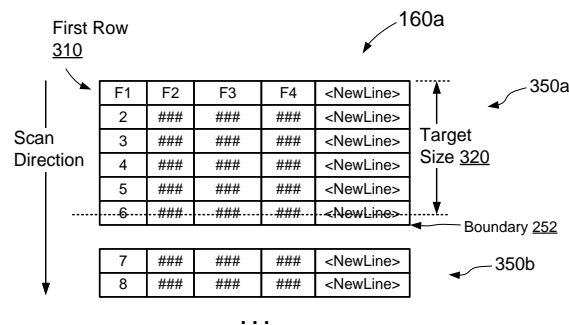
Docket #: 1215-001.001
Inventors: Donpaul Stephens and Neil Cohen
Title: PARTITIONING, PROCESSING, AND PROTECTING DATA
Page: 3 of 11

Docket #: 1215-001.001
Inventors: Donpaul Stephens and Neil Cohen
Title: PARTITIONING, PROCESSING, AND PROTECTING DATA
Page: 4 of 11



Patent Pending Application Figures:

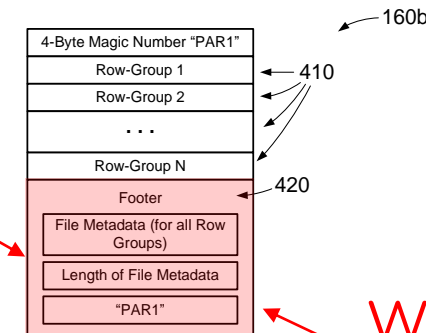
- Data is unchanged for client
- Each internal component can be processed in parallel



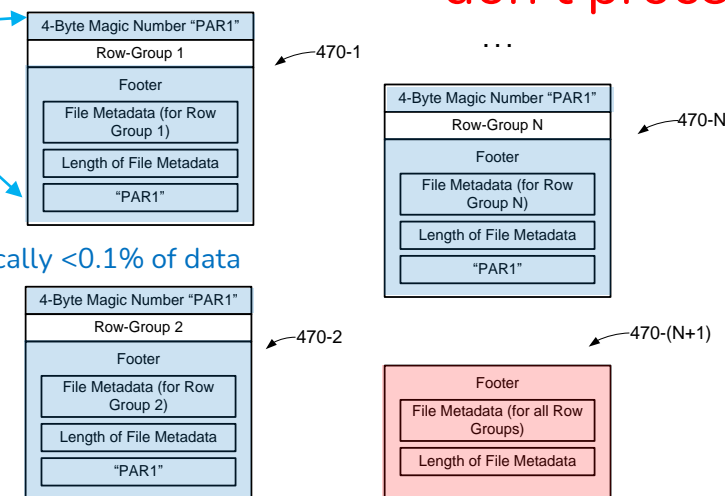
Object's own metadata

internal metadata

Not to scale!
Meta-data typically <0.1% of data



We store, but don't process!!



AirMettle internal metadata enables parallel in-storage analytics

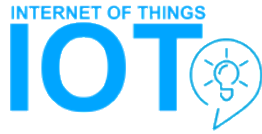
Accelerated analytics of classic tabular data

Industry Standard, Object Storage Select API



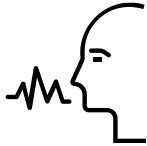
Flash Memory Summit

Security Information & Event Management



- Scan historical data to diagnose current events

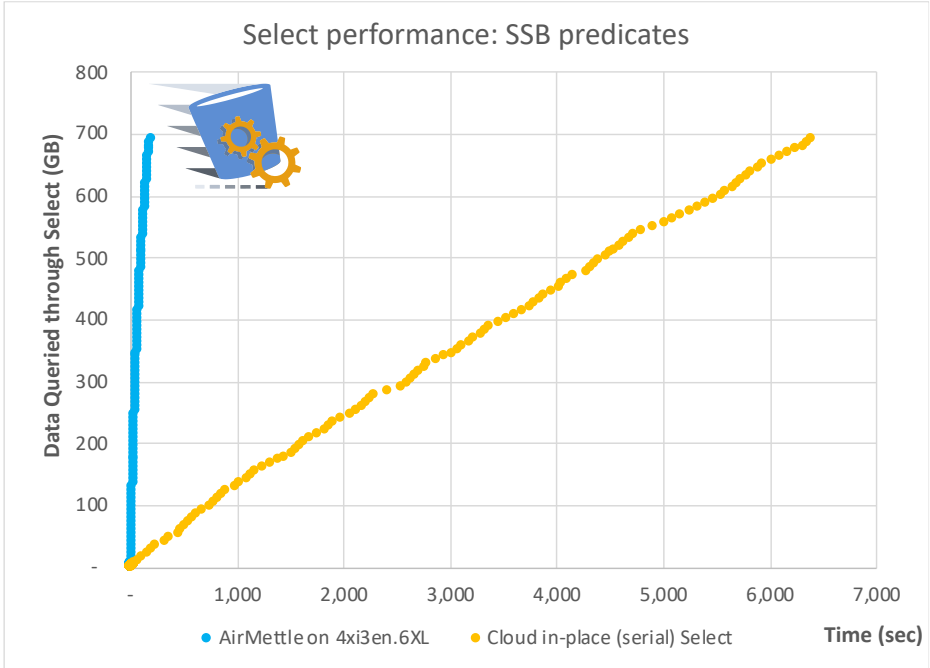
Natural Language Processing



- Find text with key-words

Validated with.  & 

Star Schema Benchmark
Utilized 223 Select queries to Object Storage:

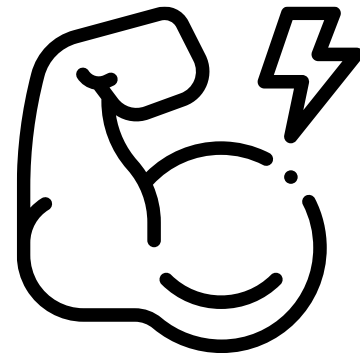


100 X faster

Under a minute vs. 1 hour 45min

Unprecedented speed of analysis: Directly from storage

No data warehouse required



DPU

Faster & Cheaper!

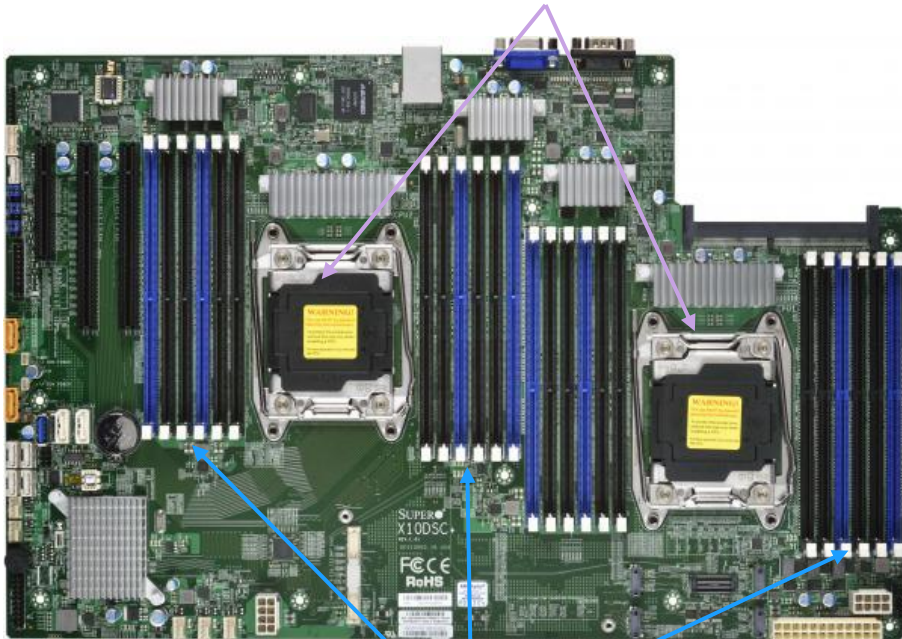
Honey, I shrunk the server

But how?



Flash Memory Summit

CPU sockets
for data processing

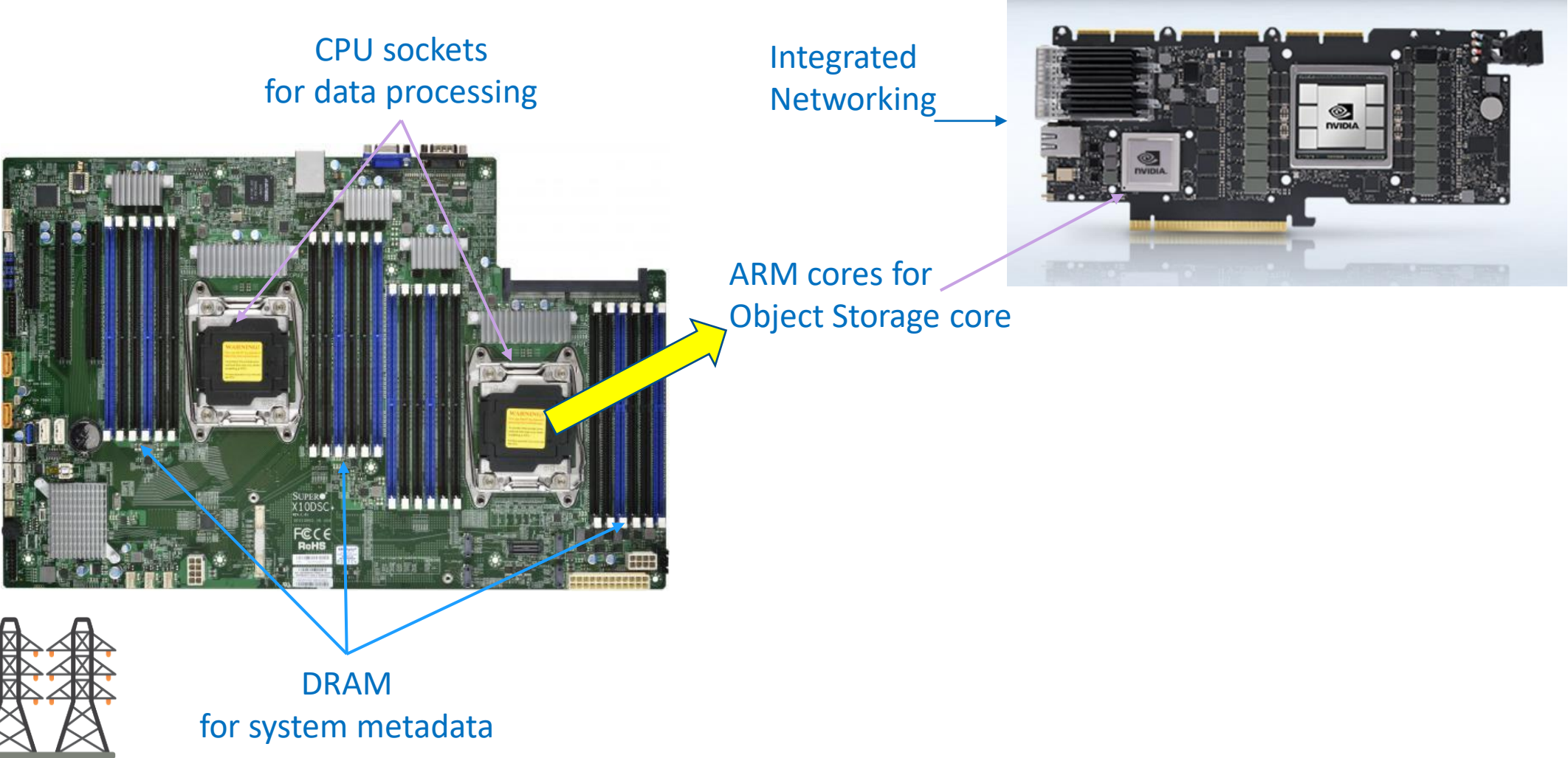


DRAM
for system metadata



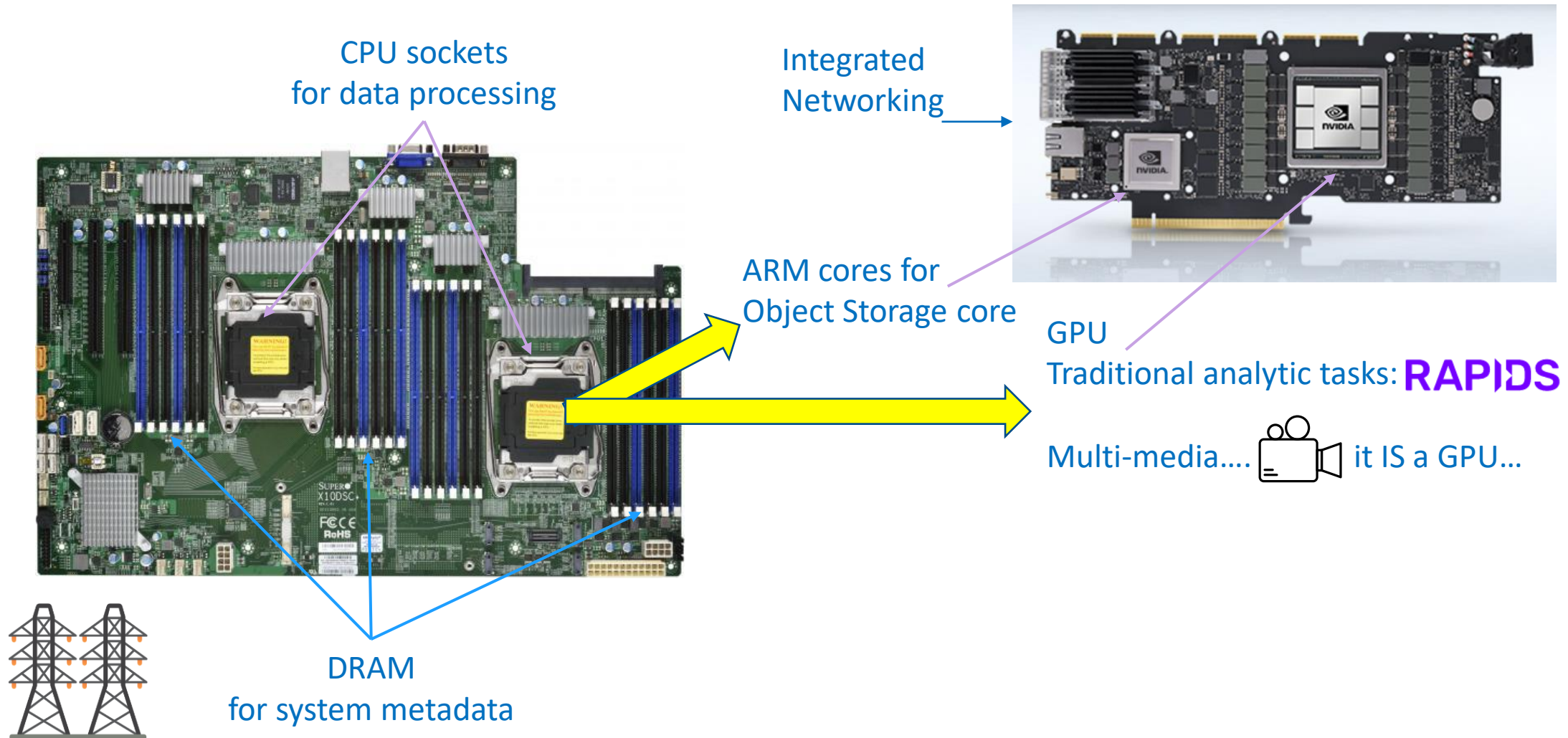
Honey, I shrunk the server

But how?



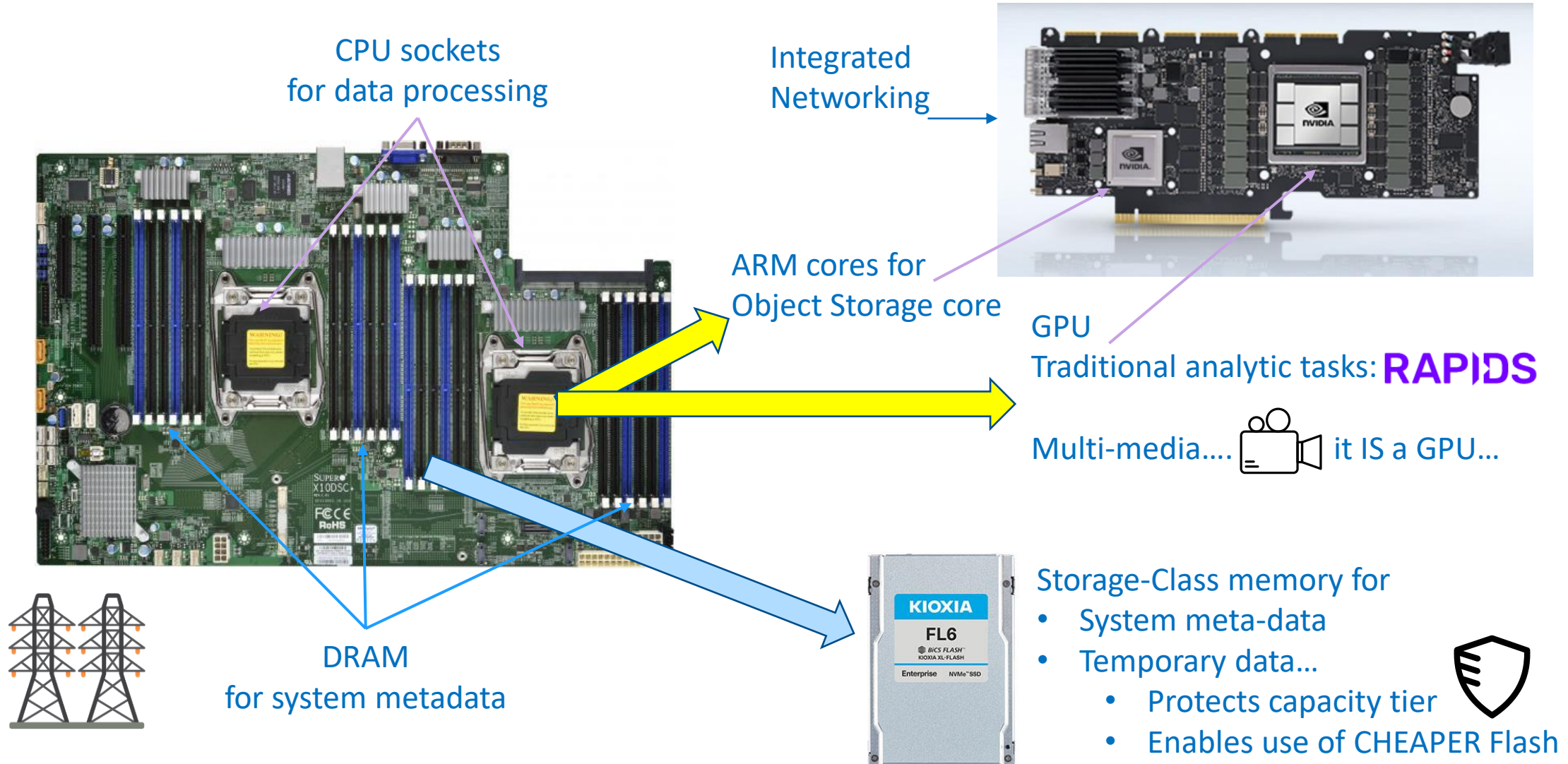
Honey, I shrunk the server

But how?



Honey, I shrunk the server

But how?



Hardware in production today from trusted suppliers?

Certainly, why else would I be talking to you?



Flash Memory Summit

NVMe-oF / “Just a Bunch of Flash” chassis



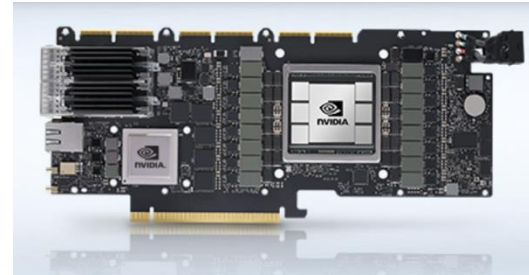
Hardware in production today from trusted suppliers?

Certainly, why else would I be talking to you?



Flash Memory Summit

Plug THIS card



NVMe-oF / “Just a Bunch of Flash” chassis



SSDs: Capacity & Storage Class

Hardware in production today from trusted suppliers?

Certainly, why else would I be talking to you?



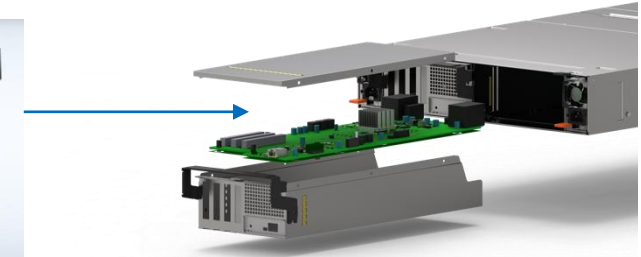
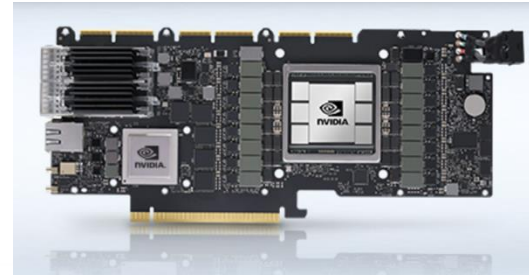
Flash Memory Summit

NVMe-oF / “Just a Bunch of Flash” chassis



SSDs: Capacity & Storage Class

Plug THIS card



In this carrier
In lieu of basic NVMe-OF card

Hardware in production today from trusted suppliers?

Certainly, why else would I be talking to you?



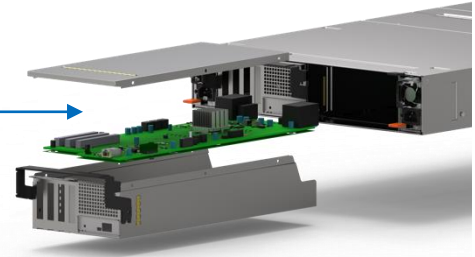
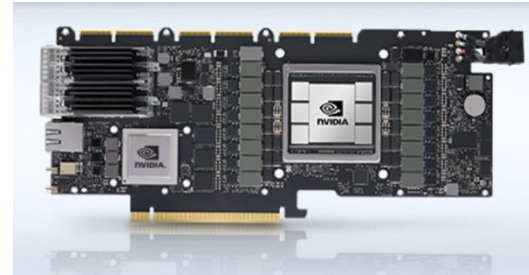
Flash Memory Summit

NVMe-oF / “Just a Bunch of Flash” chassis



SSDs: Capacity & Storage Class

Plug THIS card



In this carrier
In lieu of basic NVMe-OF card



Plug it in

Hardware in production today from trusted suppliers?

Certainly, why else would I be talking to you?



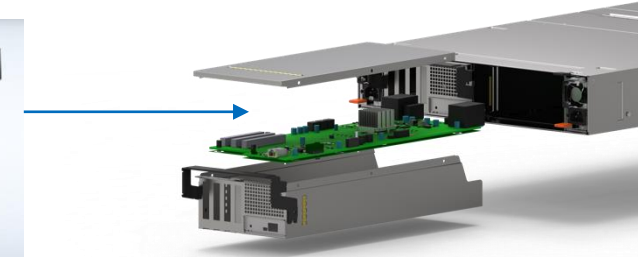
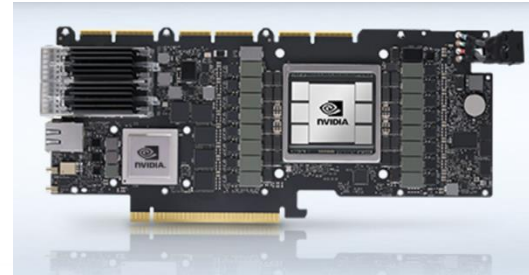
Flash Memory Summit

NVMe-oF / “Just a Bunch of Flash” chassis



SSDs: Capacity & Storage Class

Plug THIS card



In this carrier

In lieu of basic NVMe-OF card



Plug it in

Add Software





Flash Memory Summit

Thank you

Donpaul C. Stephens
donpaul@airmettle.com
Founder, AirMettle, Inc.
+1-646-872-2124



AirMettle



Break – 15 Minutes



Flash Memory Summit

DPU's Empower New Storage Architecture for NVMe-oF Targets

Brad Reger

Principal Architect

Ingrasys Technology (subsidiary of Foxconn)

Brad.Reger@ingrasys.com

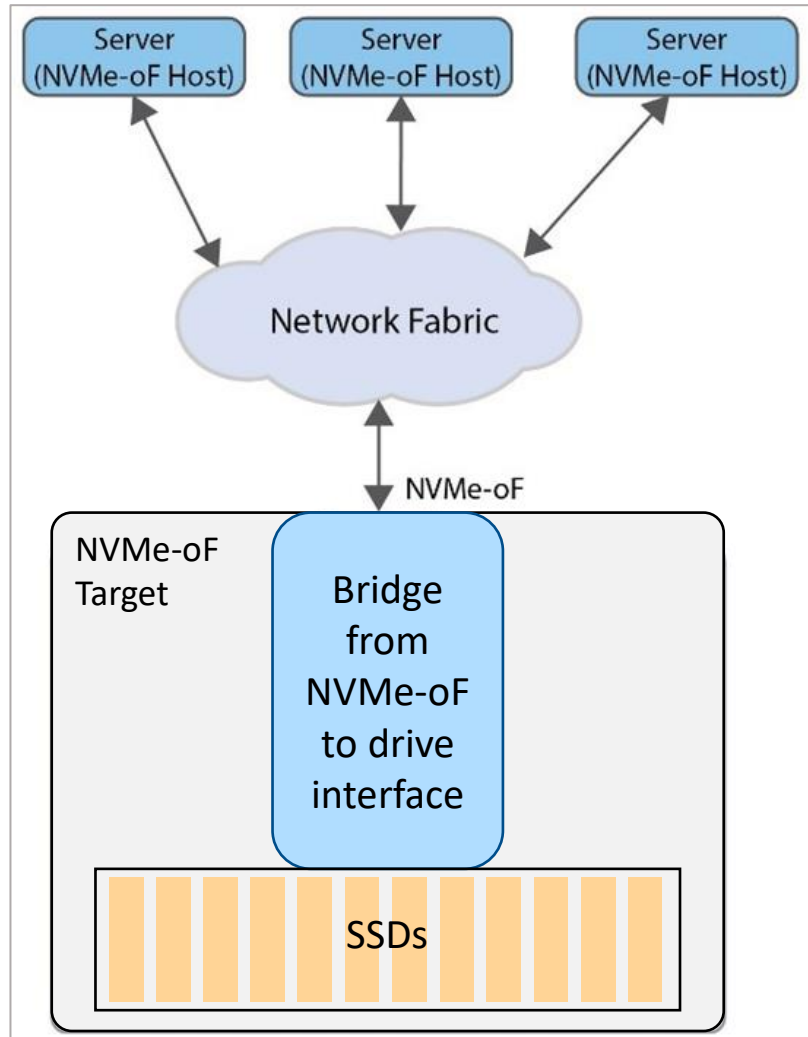
Agenda



Flash Memory Summit

- Shared, Network Storage
- NVMe-oF Targets
- DPUs for Storage Offloads
- Ingrasys CDI Proof of Concept
- Ingrasys EBOF
- Kioxia EM6 SSDs

Generic Compute, Network & NVMe-oF Storage



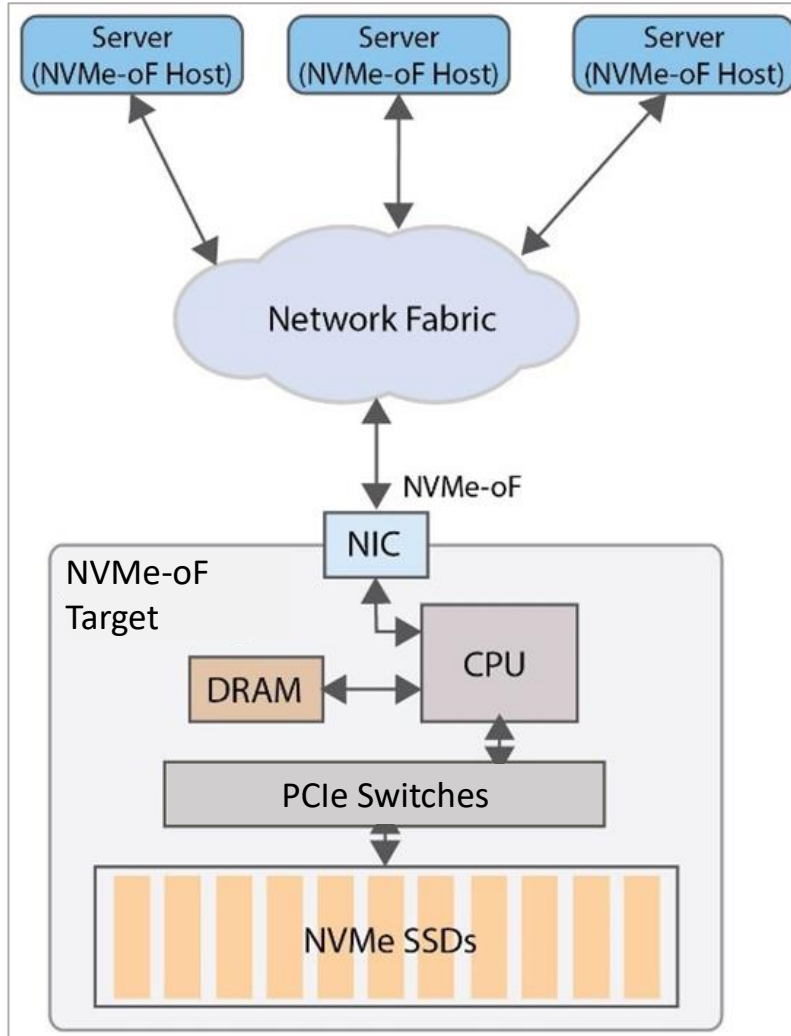
Shared, Network Storage Concepts

- Avoid storage silos with inherent stranded storage
- Disaggregate storage from compute servers into a large, shared pool to improve capacity utilization & storage efficiency
- Use NVMe-oF to flexibly allocate storage to any server on the network
- Develop high-RAS storage targets or simpler, replicated targets
- Detect server failures and remap storage to replacement server
- Scale compute power and storage capacity independently

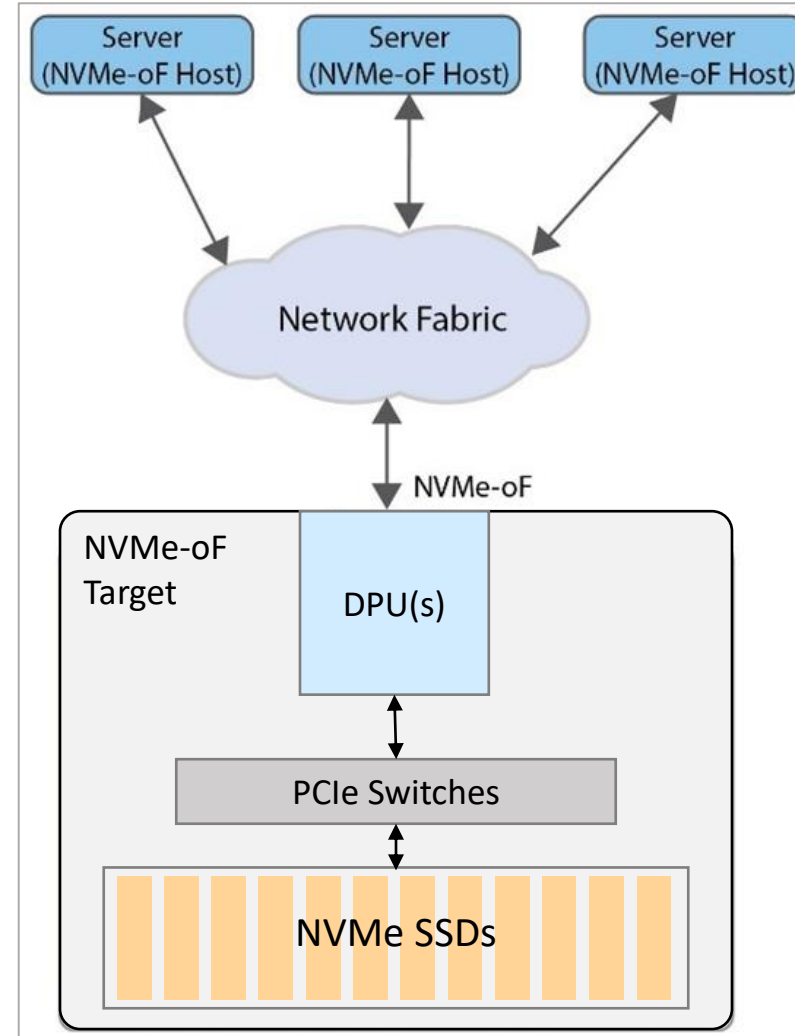
Three Ways to Build an NVMe-oF Target



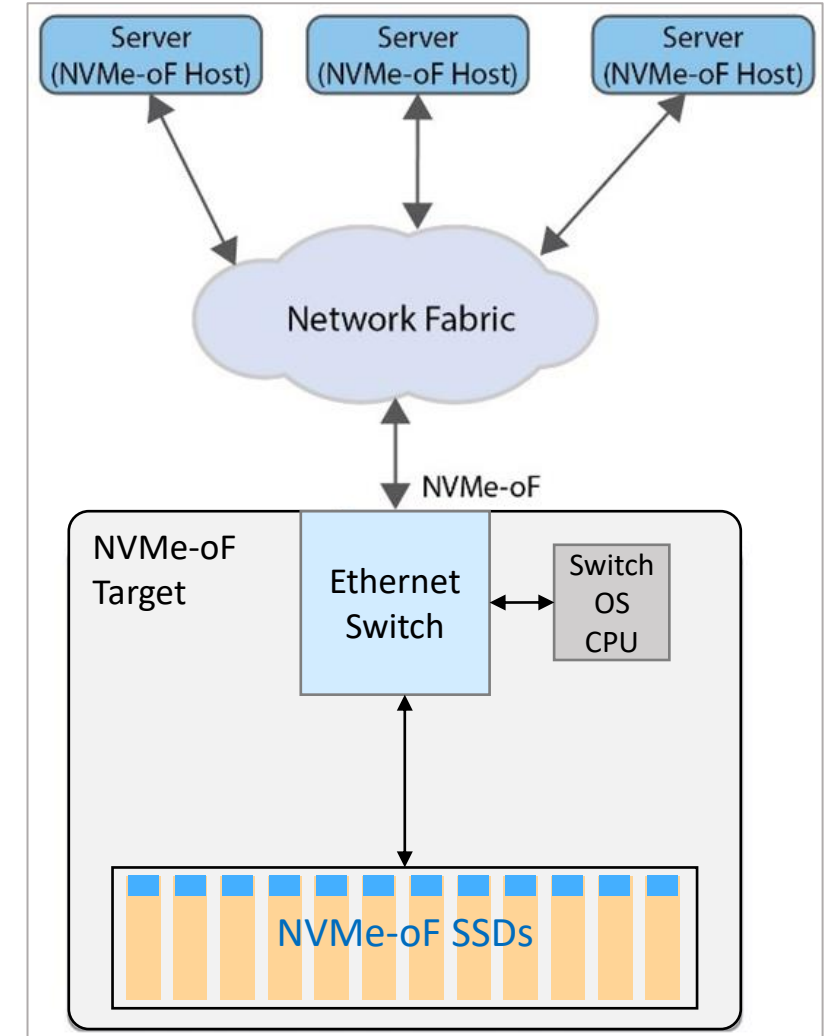
Flash Memory Summit



Legacy JBOF

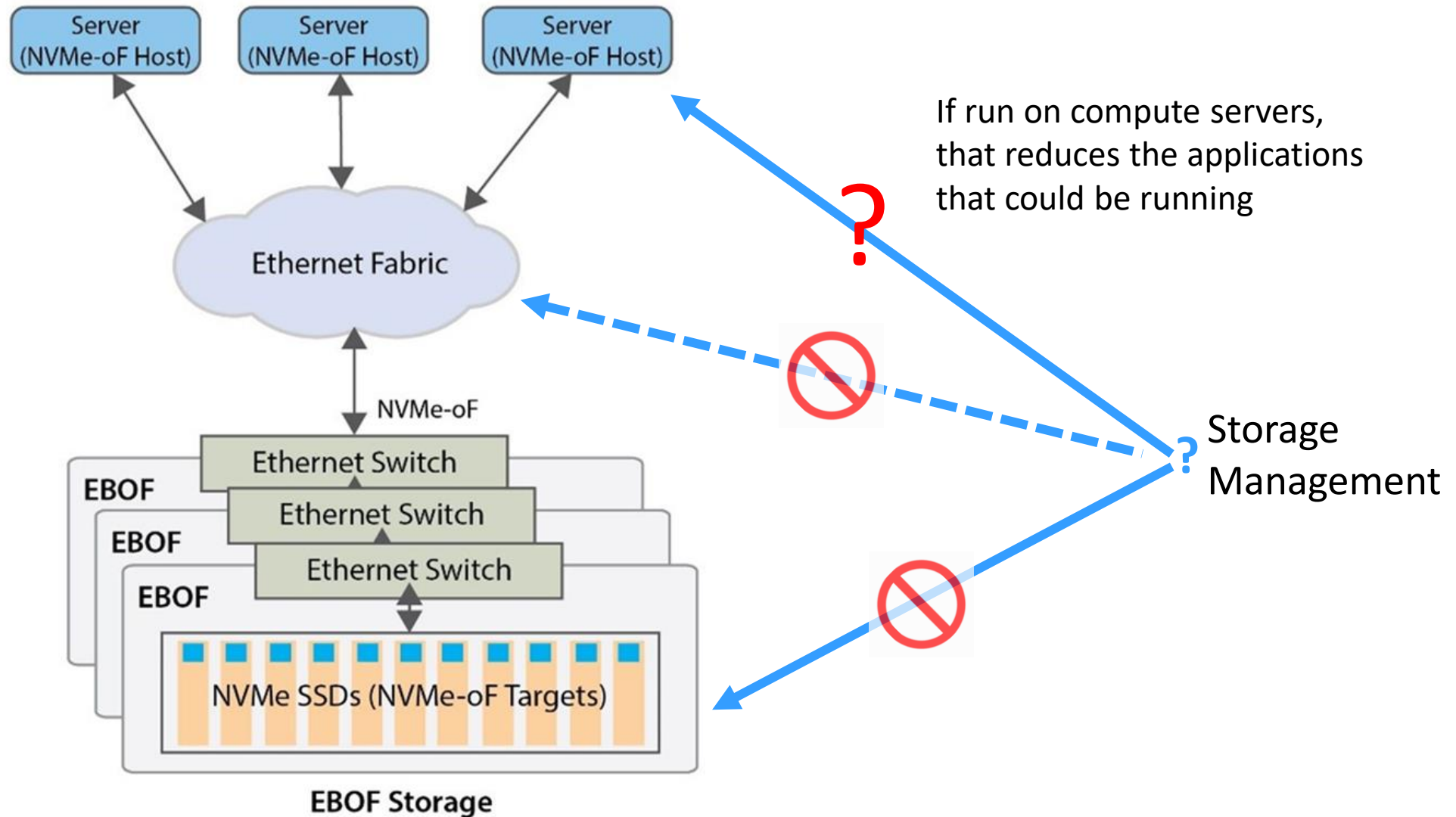


DPU-Based JBOF



Ethernet EBOF

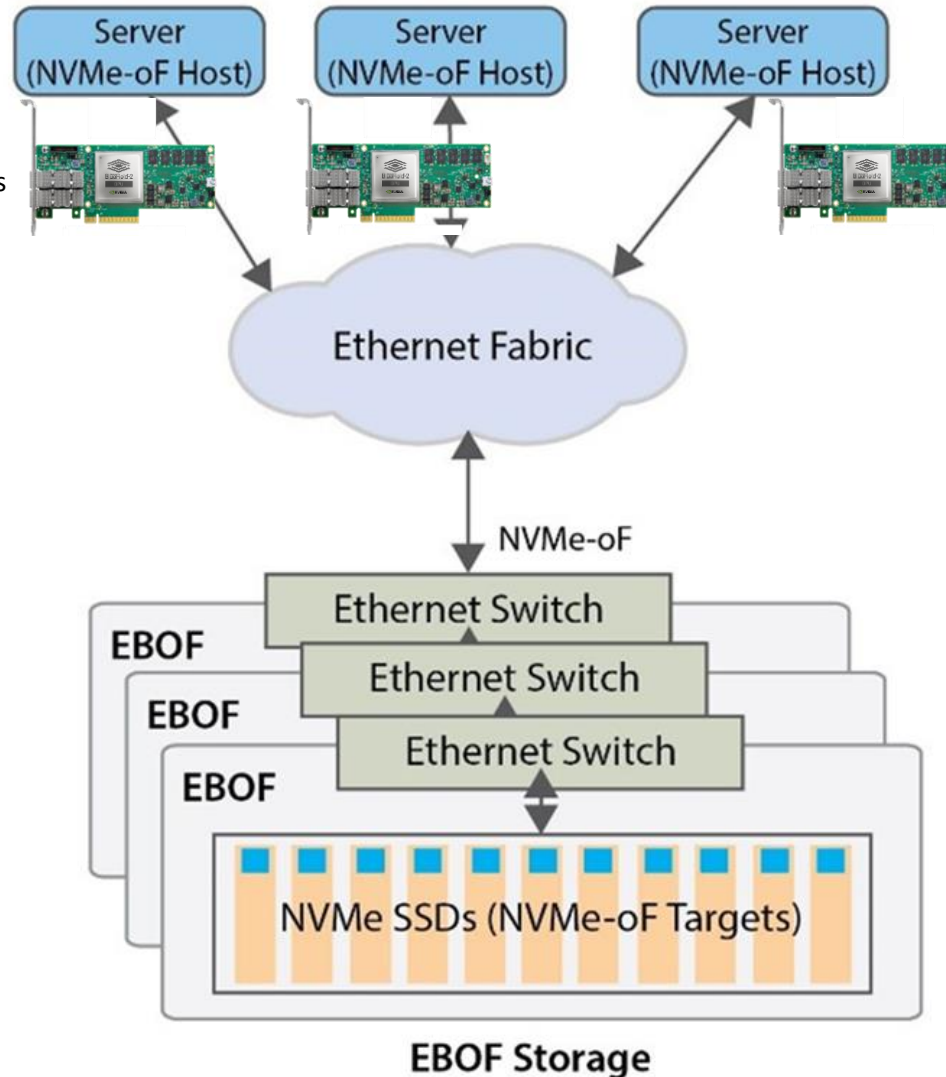
What About Storage Management (SDS)?



Best Location is a DPU in the Server

Compute Servers

- General Purpose Servers
- GPU/Accelerator Servers
- Scale-out Storage Nodes
- Enterprise Storage Controllers



•DPU Storage Management

- NVMe to NVMe-oF Bridging
- Multiprotocol
- Discovery
- Volume Management
- Compression/decompression
- Encryption for data-in-flight and at-rest
- High Availability support

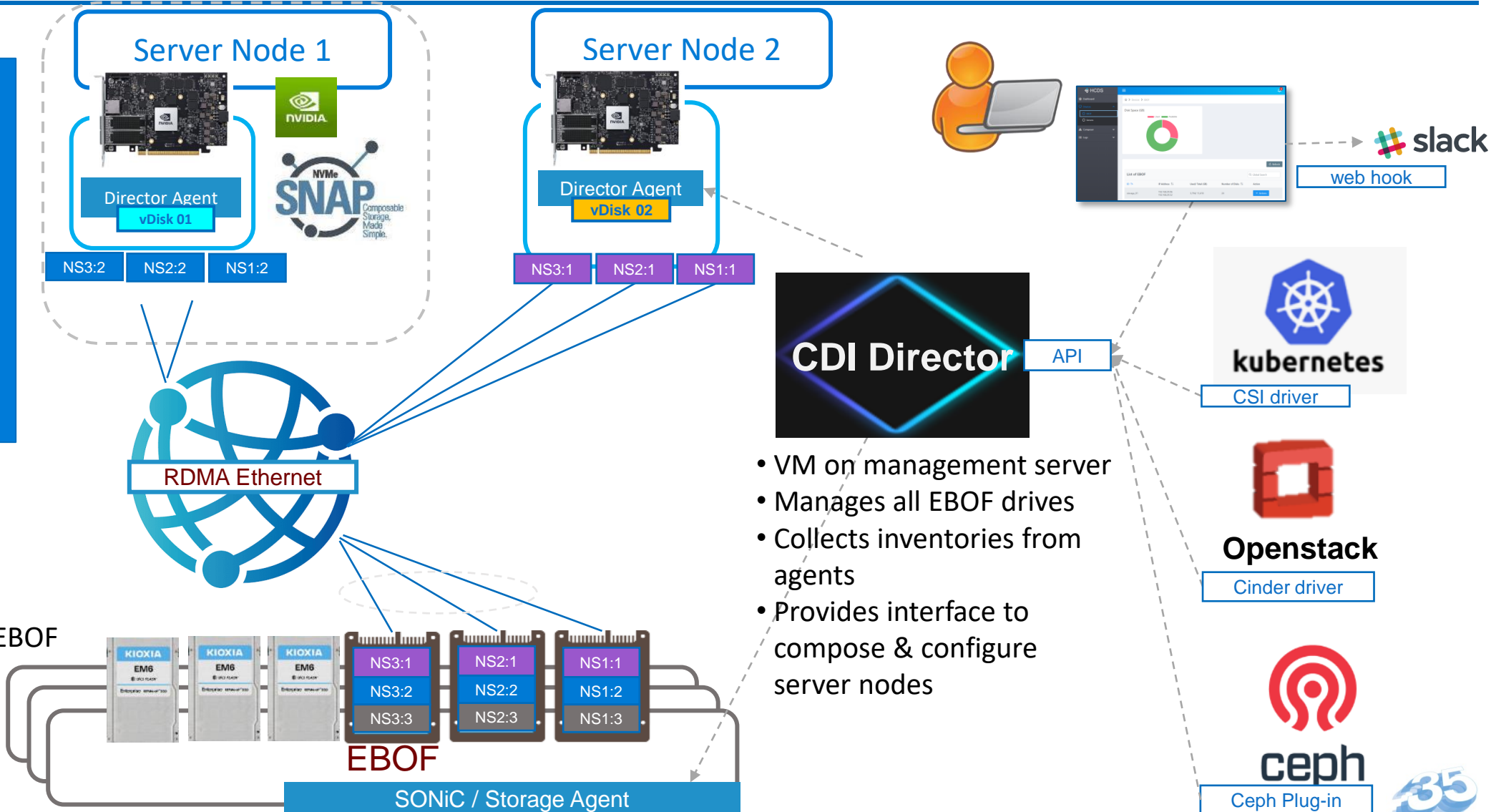
• Shared, Network Storage

Ingrasys CDI Software



Flash Memory Summit

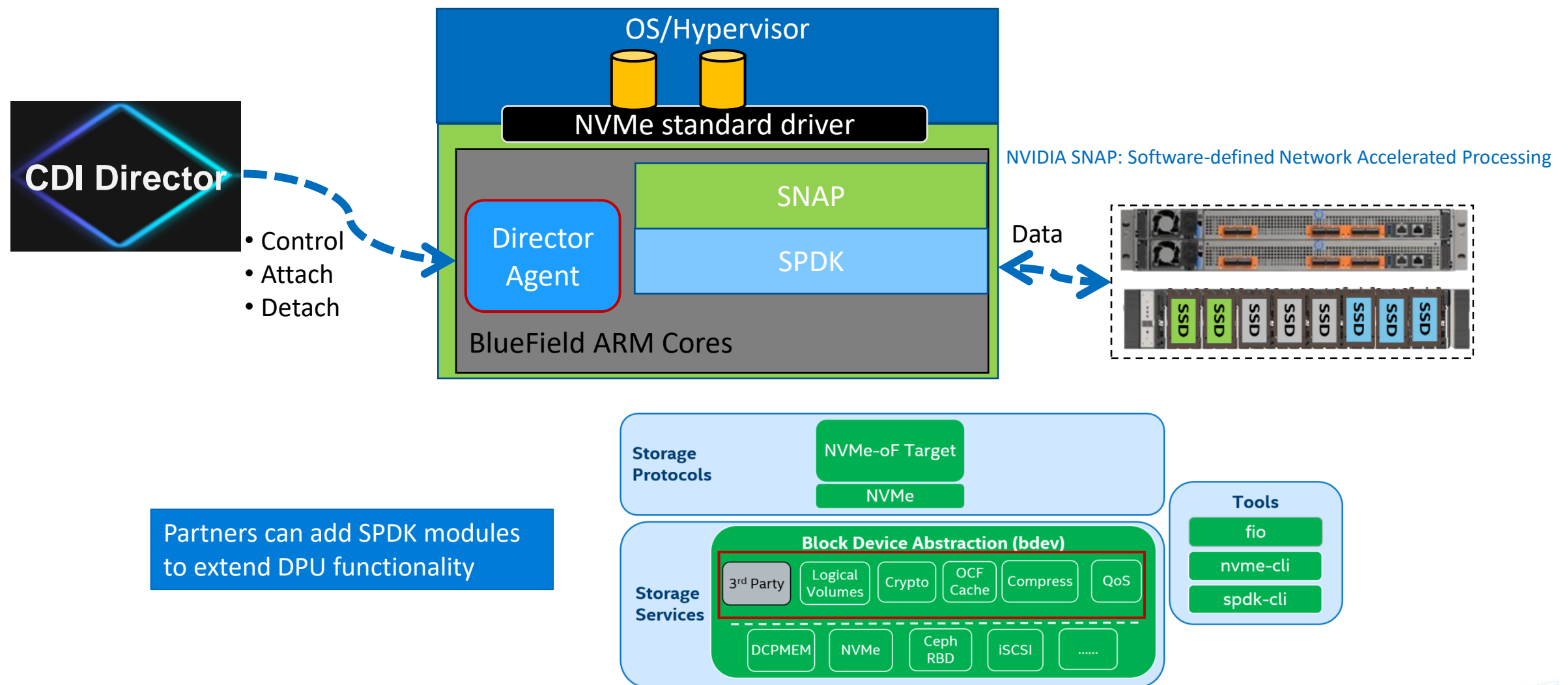
- Transparent NVMe-oF: EBOF drives appear to host OS as local NVMe drives
- Simplify host driver issues
- DPU offloads storage functions from host CPU



- VM on management server
- Manages all EBOF drives
- Collects inventories from agents
- Provides interface to compose & configure server nodes

- EBOFs can be:
- Starred off a TOR switch
 - Daisy-chained EBOF-to-EBOF

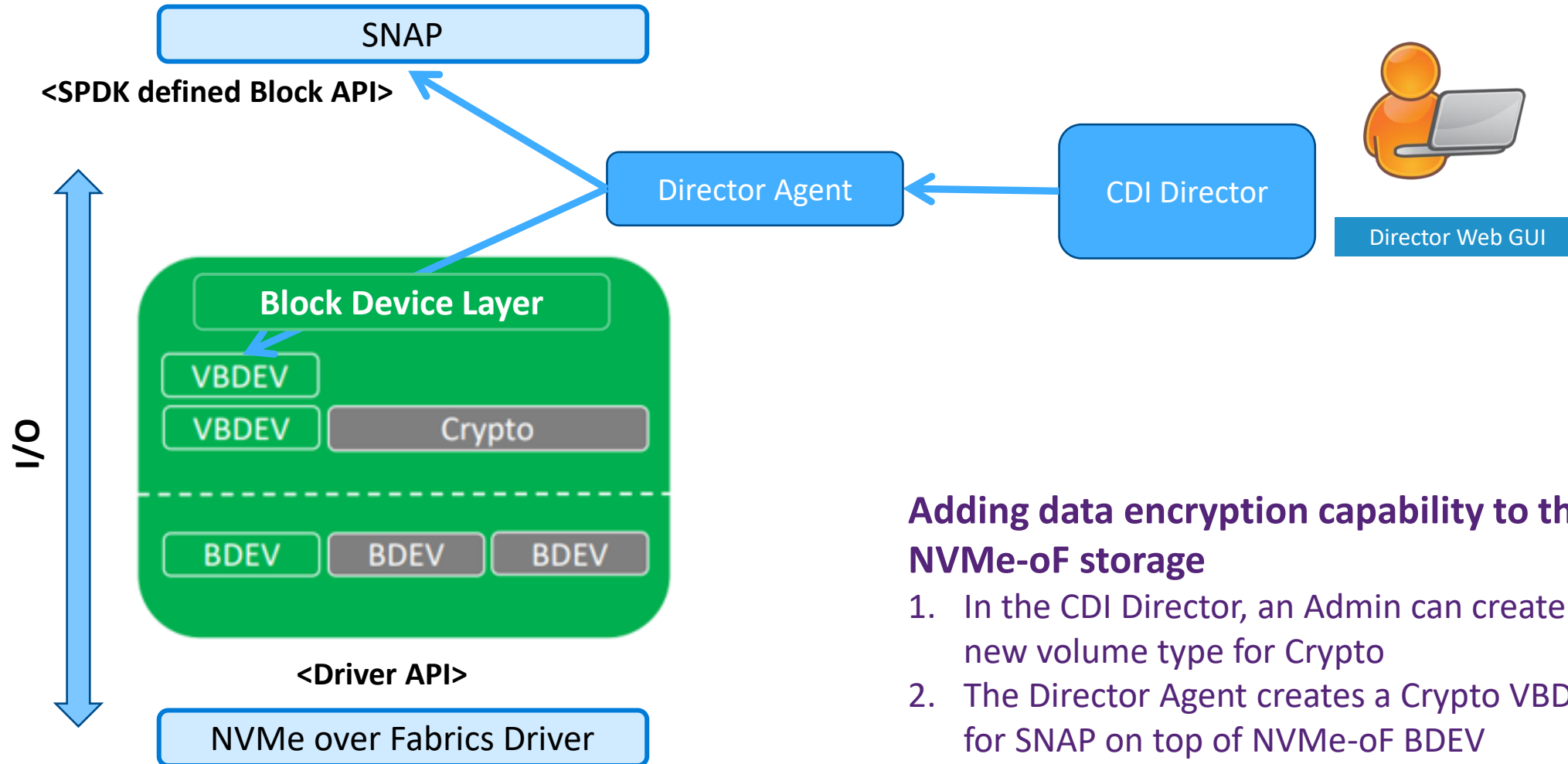
Software Architecture with DPU



BlueField-2 Extension – Data Encryption Example



Flash Memory Summit



Adding data encryption capability to the NVMe-oF storage

1. In the CDI Director, an Admin can create a new volume type for Crypto
2. The Director Agent creates a Crypto VBDEV for SNAP on top of NVMe-oF BDEV

Benefits of the EBOF, DPU, CDI Director Platform



Flash Memory Summit

- **High-Performance, Shared Storage Pool**
 - EBOFs enable full throughput from all drives at once
 - Multiple EBOFs connected to TOR switches can provide full throughput to all drives
 - Flexibly allocate & re-allocate storage capacity to any compute nodes; no stranded storage
 - EBOF internal switches run open-source SONiC OS
- **NVMe-oF Discovery and Connection**
 - Discover EBOF SSDs in the network
 - Allocate SSD namespaces to server DPUs
 - DPUs present namespaces or logical volumes to local NVMe driver
- **High Availability**
 - Servers can be assigned in a HA group, when one server encounter failures, attached NVMe-oF targets can fail-over to another server automatically.
- **Management**
 - 3rd party SW integration through CDI Director API

ES2000 EBOF



Flash Memory Summit



Enclosure Form Factor:

2U 19" EIA - 87mm x 447mm x 597mm (H x W x D)

Dual, Redundant Switch Modules

Dual, Redundant PSUs

N+1 Redundant Fans

Virtual Midplane; increased airflow

NVMe-oF/RoCE now; TCP coming

24x Drive Bays:

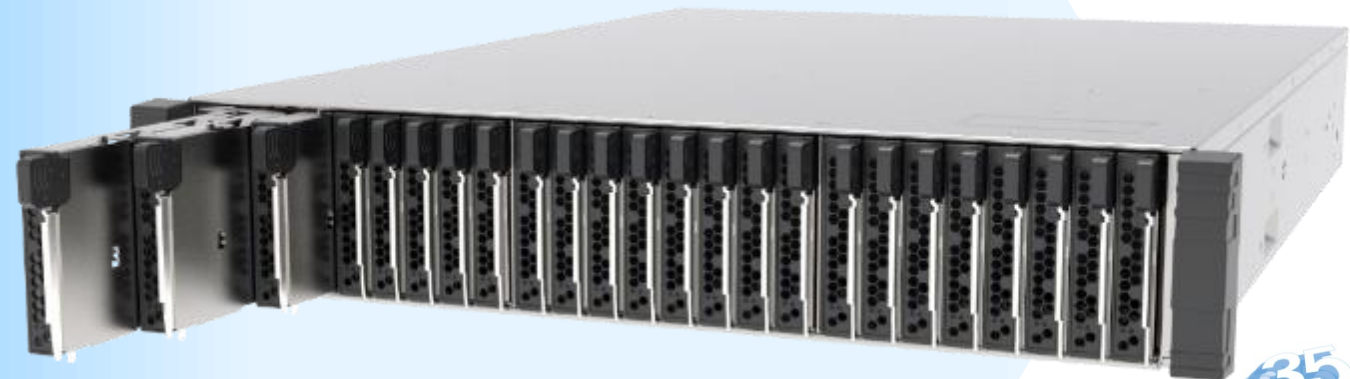
25G Ethernet NVMe-oF SSD

Std NVMe SSD w/ NVME-oF bridge interposer

U.2, U.3 (15mm)

E3.S/1T (7.5mm)

E3.L/1T (7.5mm)



Kioxia EM6 NVMe-oF™ SSD



Flash Memory Summit

- **Key Features**

- 2.5" SFF 15mm
- SFF-9639 Rev 2.1
 - Added Native NVMe-oF pinout
 - Dual 25 GBASE-KR Ethernet
- Marvell bridge SoC based NVMe-oF™ solution
- NVMe over Fabrics 1.1
- RoCEv2 transport
- 3840 / 7680 GB
- 1 DWPD
- Now shipping with ES2000 EBOF



© 2022 KIOXIA Corporation. All Rights Reserved.



Thank You!

Some Terms

Terms	Our Definitions
HDD	Hard Disk Drive; spinning magnetic media
SSD	Solid State Drive; typically flash media, could be Optane or other electronic media
Disk	Another name for an HDD (not an SSD)
Drive	Generic term for HDDs and SSDs
JBOD	Just a Bunch of <u>Drives</u> , but many use this for Just a Bunch Of <u>Disks</u>
JBOF	Just a Bunch of Flash (SSDs), could be attached to server via PCIe or Ethernet
EBOF	Ethernet-switched Bunch of Flash (SSDs)
Compute Server	A server whose primary job is to run end-user applications
GPU Server	A specialized compute server whose primary job is to run end-user applications on GPUs or other accelerators
Storage Server	A server whose primary job is to run "storage software" that virtualizes physical drives internally or externally attached to, and logically behind the storage server
Shared Storage	Generic term for storage servers accessed by multiple clients or compute servers (NAS, SAN, NVMe-oF, object, scale up, scale out, distributed, disaggregated, etc.)



Flash Memory Summit

How nebulon uses DPU/SPUs in its smartInfrastructure solution

Jeff Feierfeil

smartInfrastructure overview

primary use cases & role of DPU/SPU

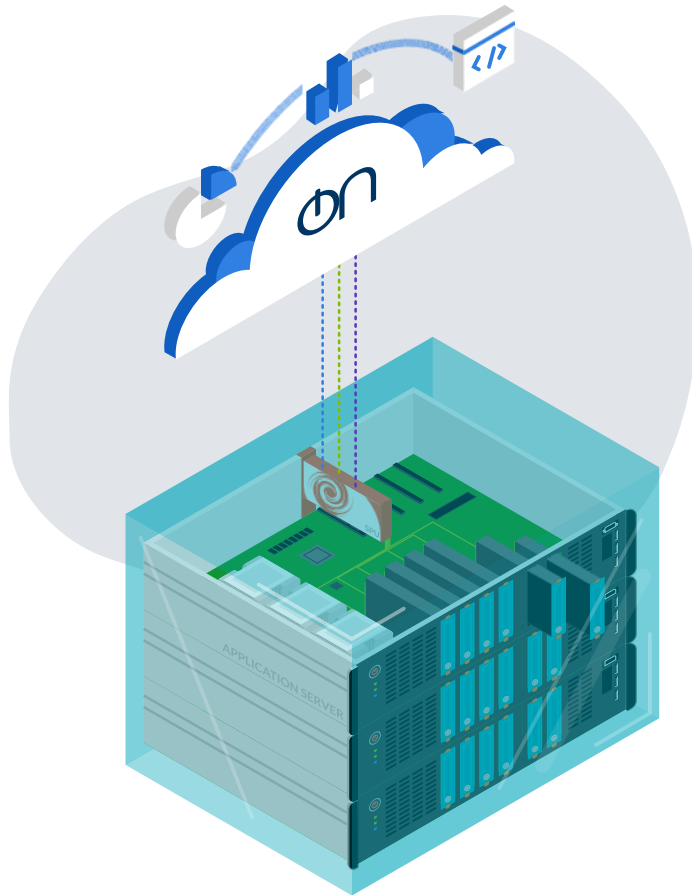
more information

nebulon **smart**Infrastructure

enabling hybrid cloud enterprises to deploy, manage & maintain on-premises application infrastructure at-scale, as simply and rapidly as in the public cloud



nebulon cloud operating platform for on-premises infrastructure



Nebulon ON Cloud

automation, updates & new features, provisioning

DPU-based Nebulon SPU

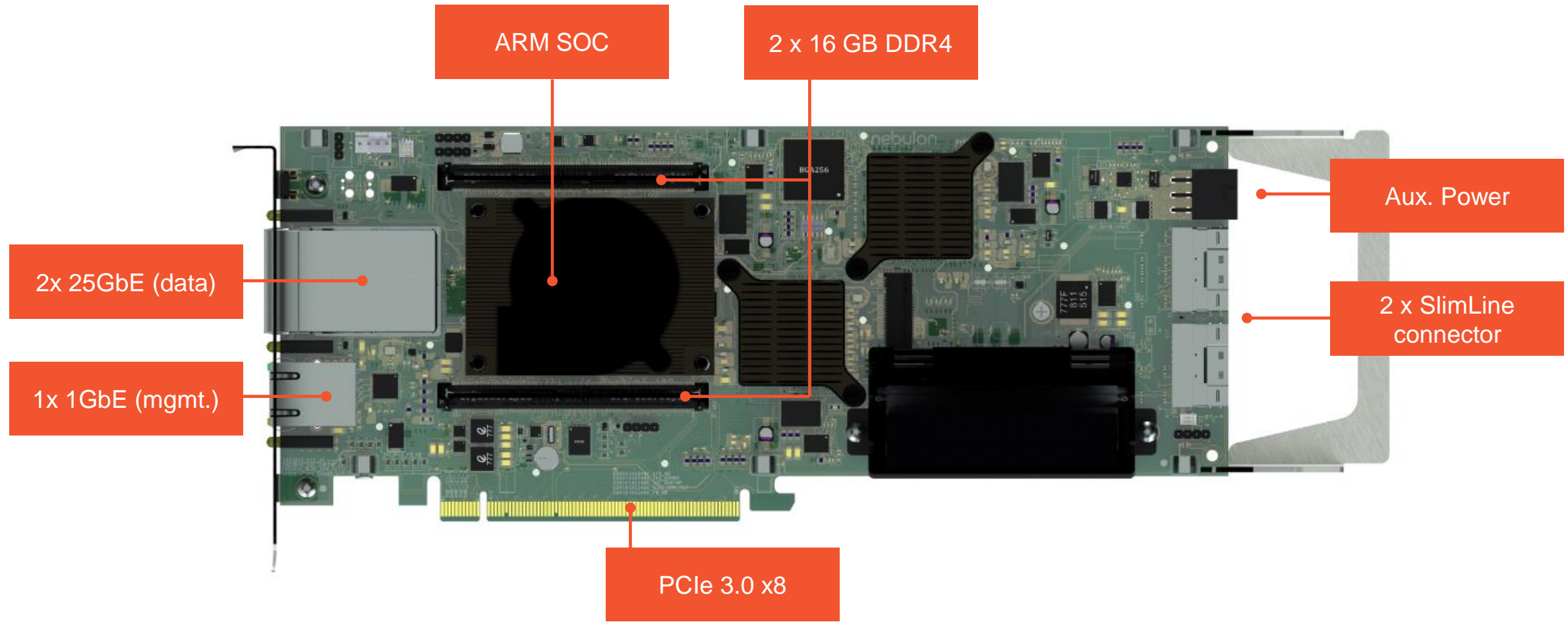
OS and data volumes, fully offloaded SDS, isolated domain

Nebulon Machine Images for x86 Servers

variety of applications, consistent deployments



Services Processing Unit (SPU) in detail





DPU solves critical challenges with on-premises infrastructure



1. rapid ransomware recovery

recover your infrastructure
in less than 4 minutes



2. securing the distributed edge

reduce footprint, secure the
edge, fleet manage all sites as
one

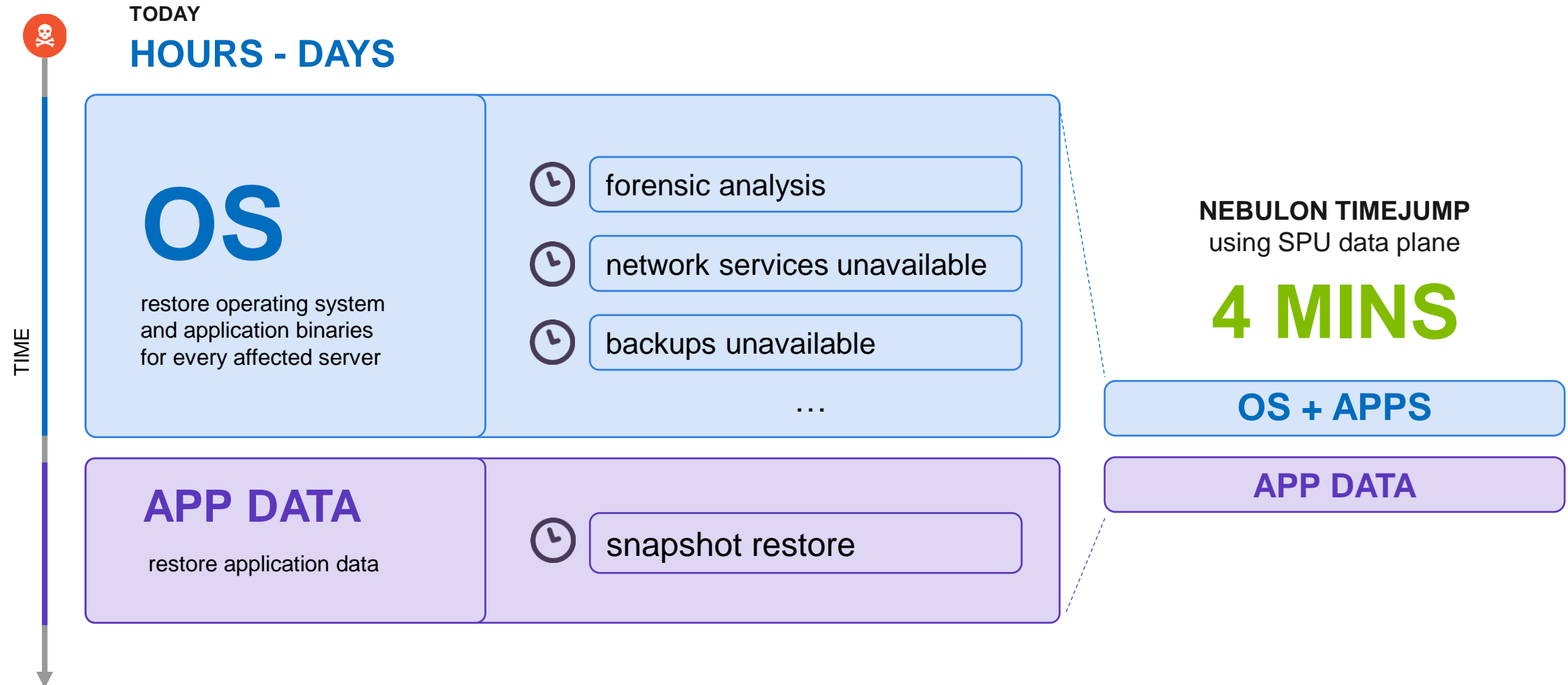


3. private cloud as-a-service

reduce hosting costs & extend
managed services off- and on-
premises



ransomware recovery without DPU is slow





DPU is a critical component for rapid ransomware recovery

isolated data protection services

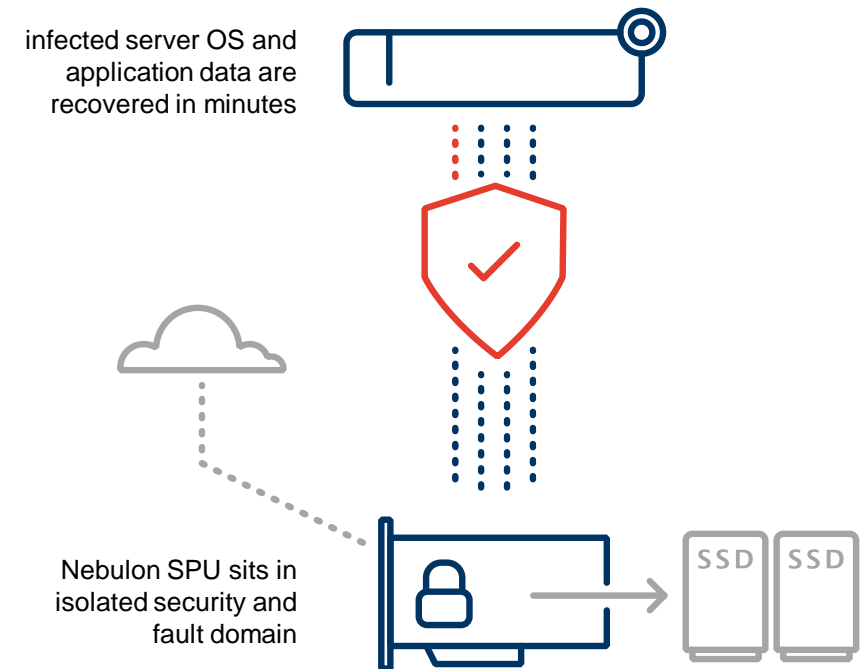
uses SPU in isolated security and fault domain for encryption and protection

recover application data & the operating system

immutable snapshot-protection of OS and application data

reduce cluster recovery time **from days to under 4 minutes**

push-button, API-accessible recovery of all affected servers





DPU is a critical component for securing the distributed edge



fleet-managed remote administration, including 'deep infrastructure' updates

remote management, control & automation of edge locations, instantiate infrastructure consistently across sites, one-click server-storage updates across all sites, including boot & data volumes



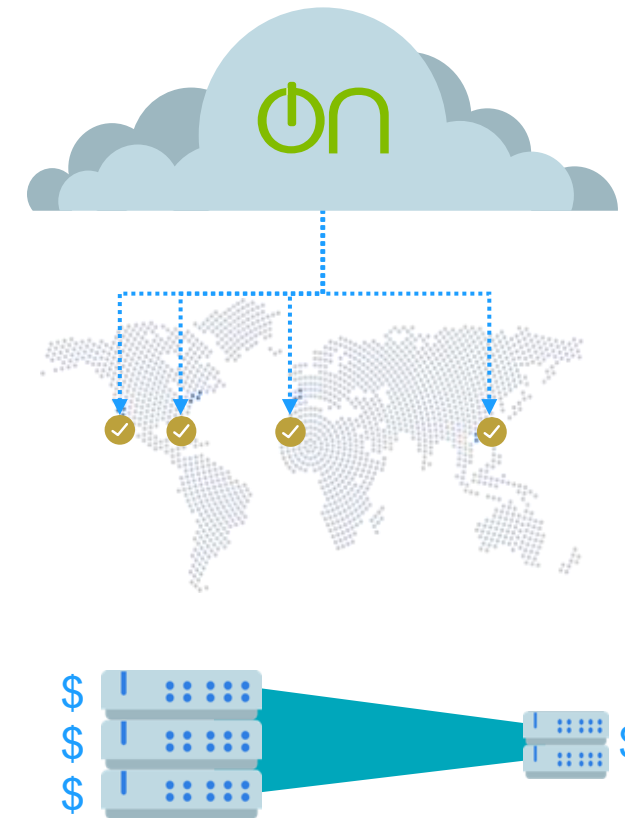
smaller configurations, better workload density reduces costs

reduce infrastructure costs and improve density with 2-node switchless configurations, no server overhead, common infrastructure with core



powerful edge security protection with 4-minute remote ransomware recovery

centralized security/permissions management, minimize security threats and provide timely firmware updates, always-on encryption, rapid ransomware recovery of OS and data





DPU is a critical component for private cloud as-a-service

grow revenue, reduce service delivery costs

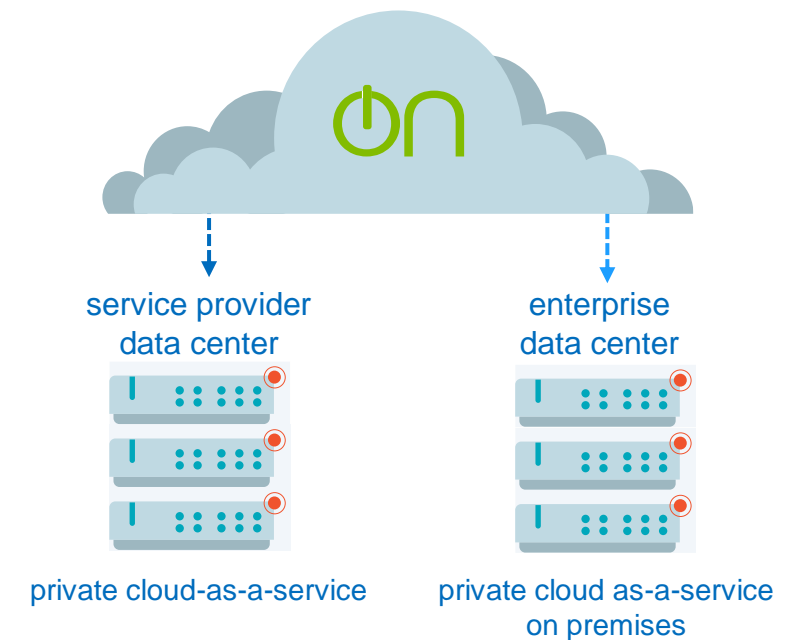
accelerated service provisioning and enhanced security for new services
improve asset utilization and optimize infrastructure management

minimize customer churn

protect workloads and SLAs by minimizing availability risks and security threats using SPU's fault isolation

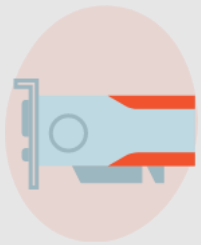
automate everything

fully automate your entire IT infrastructure at scale with a single API endpoint, consistent API version across all services, independent of services



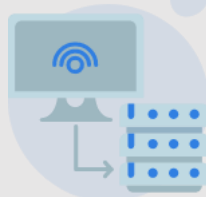


why DPU/SPU in smartInfrastructure?



ZERO- DEPENDENCY

- any OS or hypervisor
- isolated fault domains



ZERO-TOUCH

- remote fleet management through secure, embedded SPU



ZERO-DRIFT

- SPU protects OS configuration and allows freezing its state



ZERO-TRUST

- end-to-end HW crypto authentication, encryption



Thank you & learn more

interested in learning more about
smartInfrastructure?
visit our solutions page.

<https://nebulon.com/solutions/>

experience smartInfrastructure in
action or try it yourself!
request a demo.

nebulon.com/request-a-demo/



Q/A Panel

Rob Davis (NVIDIA), Joseph White(Dell/EMC), John Mao(VAST), Donpaul Stevens(AirMettle), Brad Reger(Foxconn/Ingrasys) and Jeff Feierfeil(Nebulon)



Thank You!