SDC 20

BY Developers FOR Developers

Storage Developer Conference
September 22-23, 2020

Smart Storage Adapter for Composable Architectures

Rémy GAUGUEY
Sr Software Architect
KALRAY

Kalray at SDC20

Kalray is well represented this year at SDC with 4 sessions! Please have a look.

- **A NVMe-oF Storage Diode for Classified Data Storage**
  Jean-Baptiste Riaux, Sr Field Application Engineer
- **High-performance RoCE/TCP Solutions for End-to-end NVMe-oF Communication**
  Jean-François Marie, Chief Solution Architect
- **Next Generation Datacenters Require Composable Architecture Enablers and Programmable Intelligence**
  Jean-François Marie, Chief Solution Architect
- **Smart Storage Adapter for Composable Architectures**
  Rémy Gauguey, Sr Software Architect

Kalray at SDC20

Kalray is well represented this year at SDC with 4 sessions! Please have a look.

- **A NVMe-oF Storage Diode for Classified Data Storage**
  Jean-Baptiste Riaux, Sr Field Application Engineer
- **High-performance RoCE/TCP Solutions for End-to-end NVMe-oF Communication**
  Jean-François Marie, Chief Solution Architect
- **Next Generation Datacenters Require Composable Architecture Enablers and Programmable Intelligence**
  Jean-François Marie, Chief Solution Architect
- **Smart Storage Adapter for Composable Architectures**
  Rémy Gauguey, Sr Software Architect

Abstract

Abstract

Smart Storage Adapter for Composable Architectures

The variety of architectures, use-cases and workloads to be managed by Data Center appliances is increasing. It is driving a need for storage and compute disaggregation, while at the same time forcing IT pros to simplify Data Center management and move to hyperconverged infrastructure. However the HCI approach results in siloing of storage that leads to capacity waste and scalability issue.

This paper describes how Kalray's fully programmable Smart Storage adapter leverages NVMe-oF technology to offload servers from heavy storage disaggregation task, and pave the way toward a fully Composable Infrastructure.

The Presenter

About the Presenter

Rémy Gauguey is a Senior Software Architect at Kalray, for the Data Center Business Unit. He has more than 25 years of experience in the high tech industry, with strong expertise in SoCs, RTOS and high performance packet processing.

He develops advanced architectures for composable infrastructure, leveraging the MPPA® manycore technology from Kalray.

Rémy has been previously developing his expertise at Conexant, Mindspeed Technologies and the CEA labs. He holds several patents in the fields of software architecture and packet processing.

Smart Storage Adapter for Composable Architectures

THE DATA PROCESSING UNIT REVOLUTION

In the Data-Centric Era

Scale-out data center & micro-services based applications

Network traffic explosion
East-West traffic, multi-tenant, overlays...

Data Storage Capacity explosion
Storage spread across servers / disaggregation

Multi-tenant and **security** threat
Cryptography everywhere (storage, network…)

More and more **complex** data processing
AI, analytics …

General purpose CPU and OS inefficiencies

~25% of the servers **power** spent in data centric computation
Storage stack, network stack, crypto…

General Purpose CPUs inefficient for data centric computation
But Single threaded user applications

THE DATA PROCESSING UNIT REVOLUTION

In the Data-Centric Era

Scale-out data center & micro-services based applications

Network traffic explosion
East-West traffic, multi-tenant, overlays...

Data Storage Capacity explosion
Storage spread across servers / disaggregation

Multi-tenant and **security** threat
Cryptography everywhere (storage, network…)

More and more **complex** data processing
AI, analytics …

General purpose CPU and OS inefficiencies

~25% of the servers **power** spent in data centric computation
Storage stack, network stack, crypto…

General Purpose CPUs inefficient for data centric computation
But Single threaded user applications

Data Processing Unit

Need for a new class of processing accelerator for these predominantly data-centric processing tasks!

FUTURE DATA CENTER INFRASTRUCTURE **CHALLENGES**

Towards Composable Infrastructure

❶ HCI

(Hyper Converged Infrastructure)

- Reduce complexity and hardware sprawl
- Reduce costs
- Increase agility and scalability

❷ Disaggregation

- Larger and larger datasets generated by Containerized applications and VMs
- Large diversity of application workloads

❸ Composable

- Any HW can be plugged into the system and expose new services to the others

FUTURE DATA CENTER INFRASTRUCTURE **CHALLENGES**

Towards Composable Infrastructure

➊ HCI
(Hyper Converged Infrastructure)

- Reduce complexity and hardware sprawl
- Reduce costs
- Increase agility and scalability

➋ Disaggregation

- Larger and larger datasets generated by Containerized applications and VMs
- Large diversity of application workloads

➌ Composable

- Any HW can be plugged into the system and expose new services to the others

HCI 2.0 architecture is a solution for HCI/Disaggregation …

FUTURE DATA CENTER INFRASTRUCTURE **CHALLENGES**

Towards Composable Infrastructure

➊ HCI
(Hyper Converged Infrastructure)

- Reduce complexity and hardware sprawl
- Reduce costs
- Increase agility and scalability

➋ Disaggregation

- Larger and larger datasets generated by Containerized applications and VMs
- Large diversity of application workloads

➌ Composable

- Any HW can be plugged into the system and expose new services to the others

HCI 2.0 architecture is a solution for HCI/Disaggregation …

… BUT ⚠

- Additional load on HCI cluster CPU by SW disaggregation
- Additional load on HCI cluster interconnect
- Storage Disaggregation is complex and expensive
- Clusters scalability limitation
- **HCI does not enable COMPOSABILITY**

FUTURE DATA CENTER INFRASTRUCTURE **CHALLENGES**

Towards Composable Infrastructure

➊ HCI
(Hyper Converged Infrastructure)

- Reduce complexity and hardware sprawl
- Reduce costs
- Increase agility and scalability

➋ Disaggregation

- Larger and larger datasets generated by Containerized applications and VMs
- Large diversity of application workloads

➌ Composable

- Any HW can be plugged into the system and expose new services to the others

HCI 2.0 architecture is a solution for HCI/Disaggregation …

… BUT ⚠

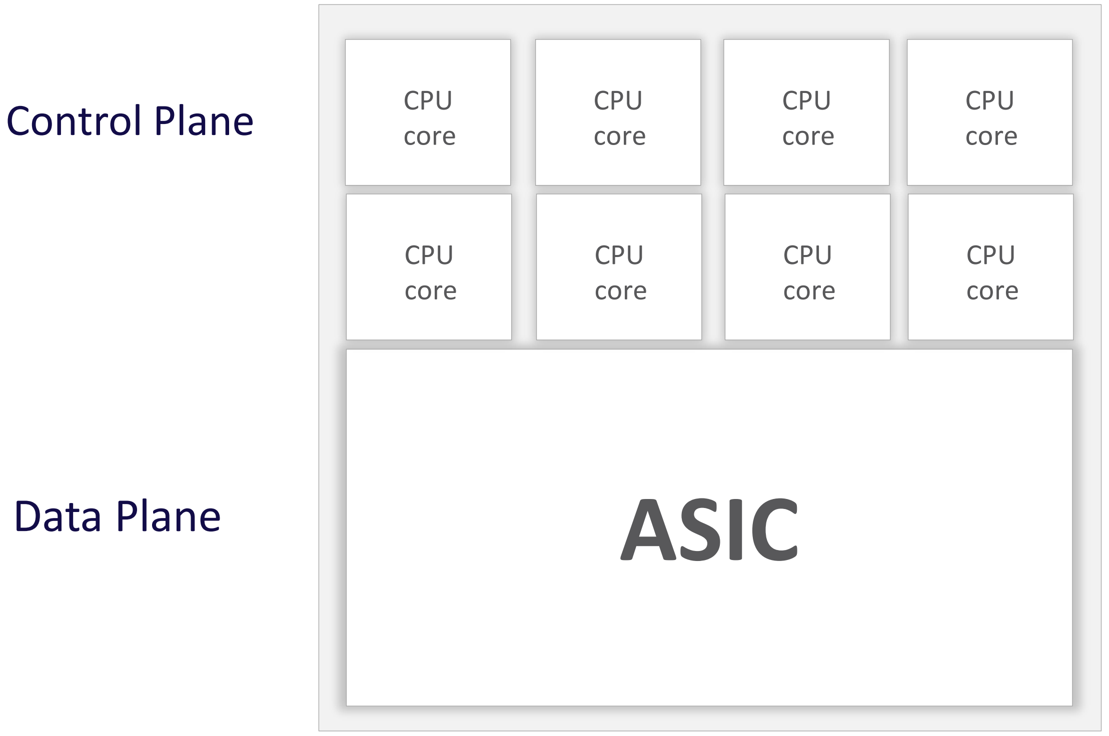
- Additional load on HCI cluster CPU by SW disaggregation
- Additional load on HCI cluster interconnect
- Storage Disaggregation is complex and expensive
- Clusters scalability limitation
- **HCI does not enable COMPOSABILITY**

Need a new approach for a truly COMPOSABLE infrastructure!

COOLIDGE™: THE ULTIMATE I/O PROCESSOR

Why Coolidge is a Revolution vs Competition ?

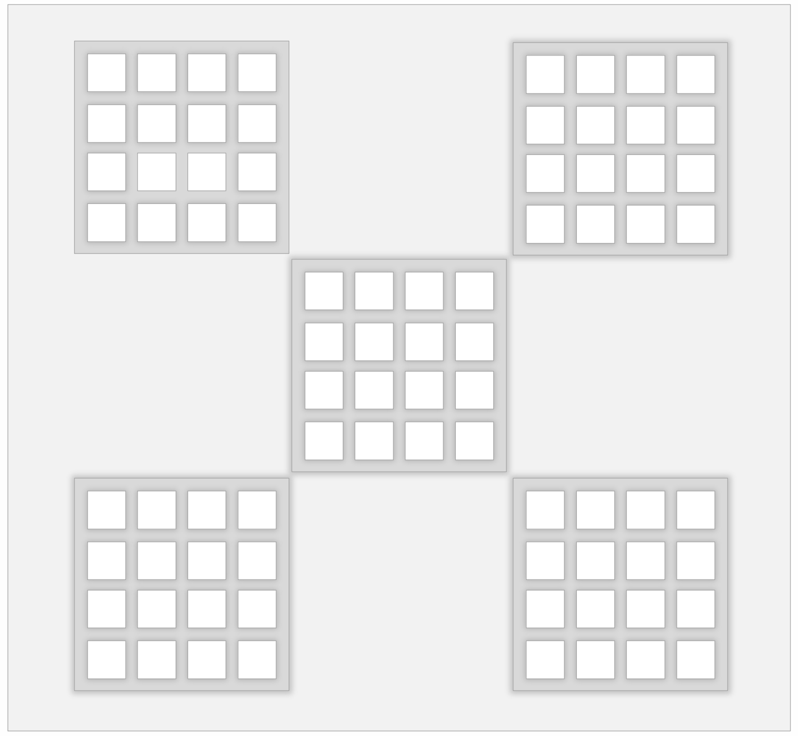
“SmartNic” Usual Approach



CONS ✗

- A few power hungry RISC CPU cores
- CPU flexibility limited to control plane
- Data plane is “hardwired” – No new services / no possible evolution!

Kalray’s MPPA®3 Coolidge™



80 highly efficient VLIW independent **CPU** cores, gathered into **5 clusters**, running at **1.2GHz,** connected to high speed fabrics & high speed interfaces.

PROS ✓

</> Fully programmable

Control Plane / Mgt Plane – Linux – 16 cores
Data Plane - 64 cores

Power efficiency
25W Typ

Top Performance Any workload
200KDMIPs, 25TOPS

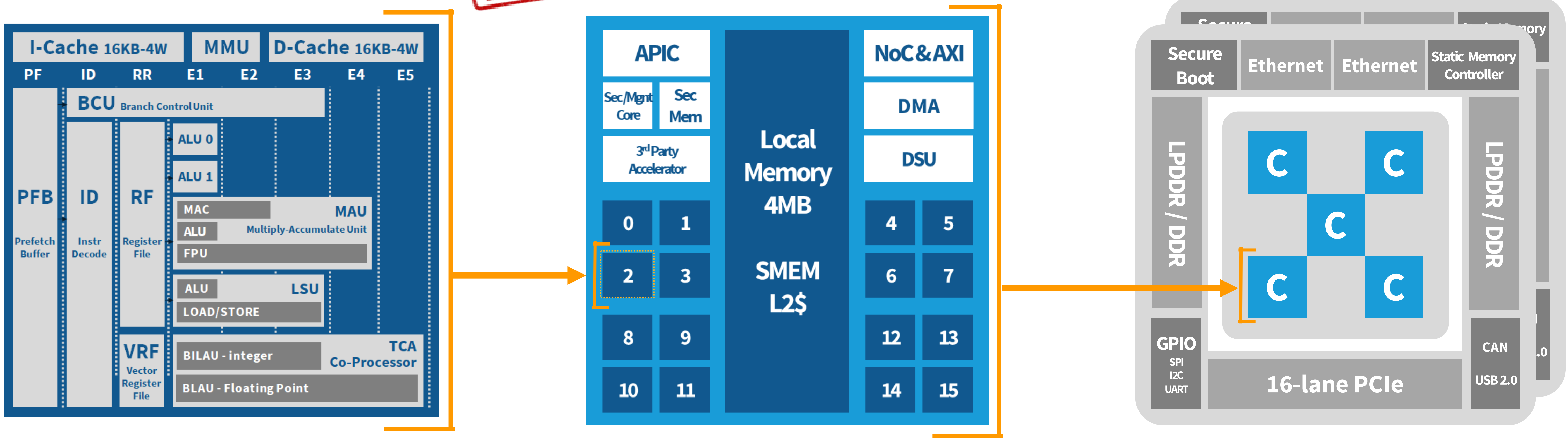
High Speed I/O
2x100Gbps,PCIGen4,DDR4

Functional Isolation & Safety
Secure Islands, Encrypt/Decrypt, Secure Boot

MPPA® COOLIDGE ARCHITECTURE

The I/O Processor for Next Gen Intelligent Systems

PATENTED



3RD GENERATION KALRAY CORE

- VLIW 64-bit core
- 6-issue VLIW architecture
- MMU + I&D cache (16KB+16KB)
- 16-bit/32-bit/64-bit IEEE 754-2008 FPU
- Vision/CNN Co-processor (TCA)

CLUSTER

Architecture

- 16 cores
- 1 safety/security dedicated core
- 600 to 1200 MHz

Memory

- L1 cache coherency (configurable)
- 4MB configurable memory (L2 cache)
- 256 bits / bandwidth up to 614GB/s)

MULTI CLUSTER ARCHITECTURE

5 Clusters: 80 cores + 80 co-processors

- Load Balancer / Packet Parser
- 2x100Gbps Ethernet
- PCIGen4
- DDR4 - 3200

AXI Bus + NoC Bus

- L2 refill in DDR and direct access to DDR from clusters
- DMA-based highly efficient data connection

DATA CENTRIC COMPUTATION

Workloads and Requirements

DATA CENTRIC WORKLOAD CHARACTERISTICS	DATA PROCESSING UNIT REQUIREMENTS
High parallelism Many stateless or stateful contexts : TCP/IP, TLS, IPsec sessions , NVMe queues	**Manycore (MIMD) architecture**
Short temporal data locality Complex memory hierarchy L1/L2/L3 not well suited	**Large on chip memory (TCM)** - With large bandwidth - Simple and deterministic memory subsystem
I/O intensive High IOPS and GB/s, low latency	**- Optimized interconnect** High bandwidth, low latency & deterministic on chip **- High speed interfaces**
Computational intensive Inline AI inference, analytics, crypto, erasure coding…	**- Floating Point Unit** **- AI acceleration** **- Cryptographic acceleration** **- Erasure Coding acceleration**
Variability and flexibility Programmability / flexibility (C, C++, standard APIs)	**- C / C++ programmable data plane** **- Standard APIs**

DATA CENTRIC COMPUTATION

MPPA®3 Coolidge™ is the Perfect Fit

DATA CENTRIC WORKLOAD CHARACTERISTICS	DATA PROCESSING UNIT REQUIREMENTS	KALRAY'S MPPA®3 COOLIDGE™
High parallelism Many stateless or stateful contexts : TCP/IP, TLS, IPsec sessions , NVMe queues	**Manycore (MIMD) architecture** ✓	**- 80 VLIW cores @ 1.2 GHz** **- 5 Clusters x16 cores**
Short temporal data locality Complex memory hierarchy L1/L2/L3 not well suited	**Large on chip memory (TCM)** - With large bandwidth - Simple and deterministic memory subsystem	
I/O intensive High IOPS and GB/s, low latency	**- Optimized interconnect** High bandwidth, low latency & deterministic on chip **- High speed interfaces**	
Computational intensive Inline AI inference, analytics, crypto, erasure coding…	**- Floating Point Unit** **- AI acceleration** **- Cryptographic acceleration** **- Erasure Coding acceleration**	
Variability and flexibility Programmability / flexibility (C, C++, standard APIs)	**- C / C++ programmable data plane** **- Standard APIs**	

DATA CENTRIC COMPUTATION

MPPA®3 Coolidge™ is the Perfect Fit

DATA CENTRIC WORKLOAD CHARACTERISTICS	DATA PROCESSING UNIT REQUIREMENTS	KALRAY'S MPPA®3 COOLIDGE™
High parallelism Many stateless or stateful contexts : TCP/IP, TLS, IPsec sessions , NVMe queues	**Manycore (MIMD) architecture**	✓ **- 80 VLIW cores @ 1.2 GHz** **- 5 Clusters x16 cores**
Short temporal data locality Complex memory hierarchy L1/L2/L3 not well suited	**Large on chip memory (TCM)** - With large bandwidth - Simple and deterministic memory subsystem	✓ **- 20 MB TCM** **- 5 isolated clusters with $L2**
I/O intensive High IOPS and GB/s, low latency	**- Optimized interconnect** High bandwidth, low latency & deterministic on chip **- High speed interfaces**	
Computational intensive Inline AI inference, analytics, crypto, erasure coding…	**- Floating Point Unit** **- AI acceleration** **- Cryptographic acceleration** **- Erasure Coding acceleration**	
Variability and flexibility Programmability / flexibility (C, C++, standard APIs)	**- C / C++ programmable data plane** **- Standard APIs**	

DATA CENTRIC COMPUTATION

MPPA®3 Coolidge™ is the Perfect Fit

DATA CENTRIC WORKLOAD CHARACTERISTICS	DATA PROCESSING UNIT REQUIREMENTS	KALRAY'S MPPA®3 COOLIDGE™
High parallelism Many stateless or stateful contexts : TCP/IP, TLS, IPsec sessions , NVMe queues	**Manycore (MIMD) architecture**	✓ **- 80 VLIW cores @ 1.2 GHz** **- 5 Clusters x16 cores**
Short temporal data locality Complex memory hierarchy L1/L2/L3 not well suited	**Large on chip memory (TCM)** - With large bandwidth - Simple and deterministic memory subsystem	✓ **- 20 MB TCM** **- 5 isolated clusters with $L2**
I/O intensive High IOPS and GB/s, low latency	**- Optimized interconnect** High bandwidth, low latency & deterministic on chi **- High speed interfaces**	✓ **- High perf. NoC** **- 2x100 Gbps Ethernet** **- PCIe x16 Gen4 (RC/EP)**
Computational intensive Inline AI inference, analytics, crypto, erasure coding…	**- Floating Point Unit** **- AI acceleration** **- Cryptographic acceleration** **- Erasure Coding acceleration**	
Variability and flexibility Programmability / flexibility (C, C++, standard APIs)	**- C / C++ programmable data plane** **- Standard APIs**	

DATA CENTRIC COMPUTATION

MPPA®3 Coolidge™ is the Perfect Fit

DATA CENTRIC WORKLOAD CHARACTERISTICS	DATA PROCESSING UNIT REQUIREMENTS		KALRAY'S MPPA®3 COOLIDGE™
High parallelism Many stateless or stateful contexts : TCP/IP, TLS, IPsec sessions , NVMe queues	**Manycore (MIMD) architecture**	✓	**- 80 VLIW cores @ 1.2 GHz** **- 5 Clusters x16 cores**
Short temporal data locality Complex memory hierarchy L1/L2/L3 not well suited	**Large on chip memory (TCM)** - With large bandwidth - Simple and deterministic memory subsystem	✓	**- 20 MB TCM** **- 5 isolated clusters with $L2**
I/O intensive High IOPS and GB/s, low latency	**- Optimized interconnect** High bandwidth, low latency & deterministic on chi **- High speed interfaces**	✓	**- High perf. NoC** **- 2x100 Gbps Ethernet** **- PCIe x16 Gen4 (RC/EP)**
Computational intensive Inline AI inference, analytics, crypto, erasure coding…	**- Floating Point Unit** **- AI acceleration** **- Cryptographic acceleration** **- Erasure Coding acceleration**	✓	**- Up to 1.15 TFLOPs (SP)** **- Up to 25 TOPs (8bits)/4.2TFLOPS (HP) for AI** **- 100 Gbps + Crypto acc.** **- Line rate Reed Solomon**
Variability and flexibility Programmability / flexibility (C, C++, standard APIs)	**- C / C++ programmable data plane** **- Standard APIs**		

DATA CENTRIC COMPUTATION
MPPA®3 Coolidge™ is the Perfect Fit

DATA CENTRIC WORKLOAD CHARACTERISTICS	DATA PROCESSING UNIT REQUIREMENTS		KALRAY'S MPPA®3 COOLIDGE™
High parallelism Many stateless or stateful contexts : TCP/IP, TLS, IPsec sessions , NVMe queues	**Manycore (MIMD) architecture**	✓	**- 80 VLIW cores @ 1.2 GHz** **- 5 Clusters x16 cores**
Short temporal data locality Complex memory hierarchy L1/L2/L3 not well suited	**Large on chip memory (TCM)** - With large bandwidth - Simple and deterministic memory subsystem	✓	**- 20 MB TCM** **- 5 isolated clusters with $L2**
I/O intensive High IOPS and GB/s, low latency	**- Optimized interconnect** High bandwidth, low latency & deterministic on chi **- High speed interfaces**	✓	**- High perf. NoC** **- 2x100 Gbps Ethernet** **- PCIe x16 Gen4 (RC/EP)**
Computational intensive Inline AI inference, analytics, crypto, erasure coding…	**- Floating Point Unit** **- AI acceleration** **- Cryptographic acceleration** **- Erasure Coding acceleration**	✓	**- Up to 1.15 TFLOPs (SP)** **- Up to 25 TOPs (8bits)/4.2TFLOPS (HP) for AI** **- 100 Gbps + Crypto acc.** **- Line rate Reed Solomon**
Variability and flexibility Programmability / flexibility (C, C++, standard APIs)	**- C / C++ programmable data plane** **- Standard APIs**	✓	**- Linux, OpenDataPlane** **- SPDK BDEVs, NVMe** **- OpenCL**

Kalray Smart Storage Adapter Solutions

MPPA®

The Processor at the Heart of Intelligent Systems

KALRAY SMART STORAGE ADAPTER SOLUTION
K200 / K200-LP & ACS SDK

K200 & K200-LP
manufactured by wistron

2 Form Factors
- FHHL (Full Height) - K200 - Single Slot
- HHHL (Low Profile) - K200-LP
  Single or Double Slots

Manycore Architecture
- 80 VLIW cores @ 1.2 Ghz
- 5 Clusters x16 cores

High Speed Ethernet
- 2x100GbE / 8x25 GbE

Certified NVMe-oF Stack
- NVMe-oF 1.1 (Target, Intiator)
- RoCE v1/v2, TCP

Advanced SSD interface
- PCIe-Gen4
- NVMe 1.1 to 1.4 SSDs
  No need for CMB
- Dual port SSD support

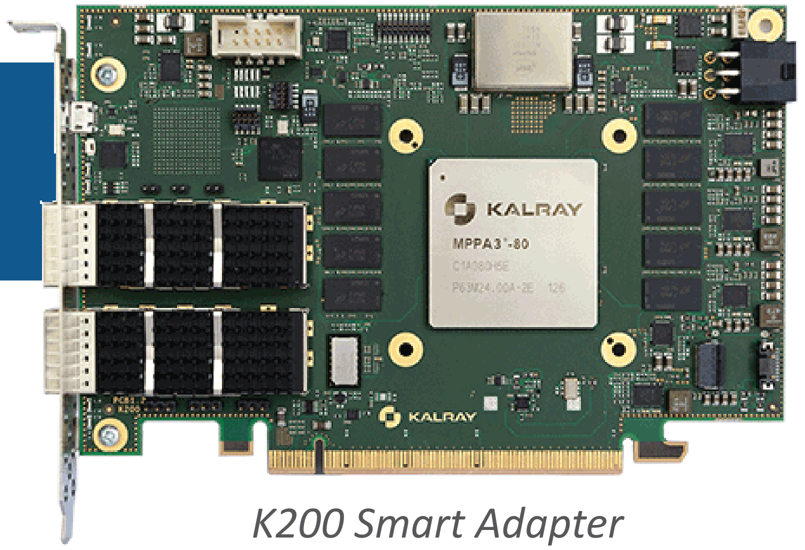
2 Modes
- Stand-alone
- Host CPU co-processor
  / "host-agnostic" support

Agnostic Host Support
- NVMe Driver

DDR-3200
- 8GB to 32GB

H/W Accelerators
- Encryption / Decryption
- Hashing (SHA-256, SHA-3)
- Erasure Coding

Low Power
- 35W (single slot)
- 65W (double slot)



K200 Smart Adapter

Key figures (per card)
- Random R/W RoCE: **4-6 MIOPS**
- Random R/W TCP: **2-4 MIOPS**
- Sequential R/W (RoCE&TCP): **25GB/s**
- Latency (RoCE/TCP): **10 /30 usec**

SDK
AccessCore®
Open Software & Tools

Open Software Environment
- Linux / SPDK Control Plane (16 Cores)
- Fully Programmable Data Plane (64 Cores)
- Storage, Network and Compute Services
  (AI,DSP,NVMe,NVMe-oF,ROCE,TCP, RAID, de-dup,..)

Agnostic Host Support
- NVMe Driver

+ Extra compute available
- @ 3MIOPS, 50% cores available !
- Storage Services (RAID, de-dedup ...)
- AI
- Analytics ...

KALRAY SMART STORAGE ADAPTER SOLUTION

Simplified Integration into any System

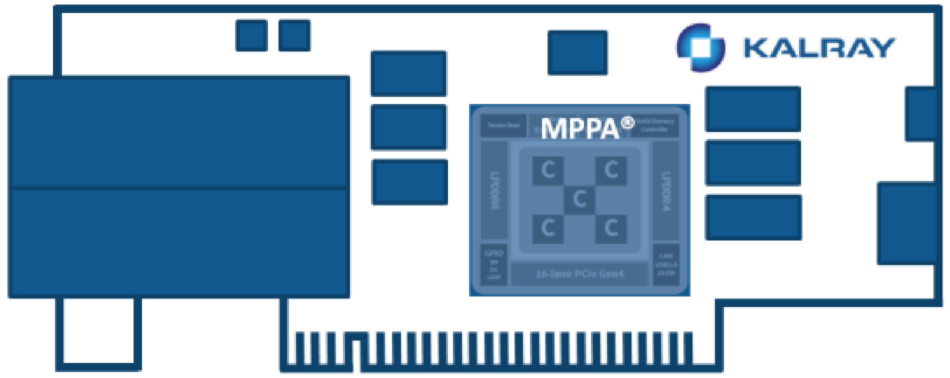
AccessCore ™ Storage framework on MPPA® to deliver Data Services

SDK

iol Certified nvm EXPRESS®

Standard
NVMe-oF
TCP & RoCE

Storage cluster interconnect



Management through GbE

SNIA Swordfish · DMTF Redfish

Ease of integration on a x86 node via SR-IOV ***NVMe Emulation***

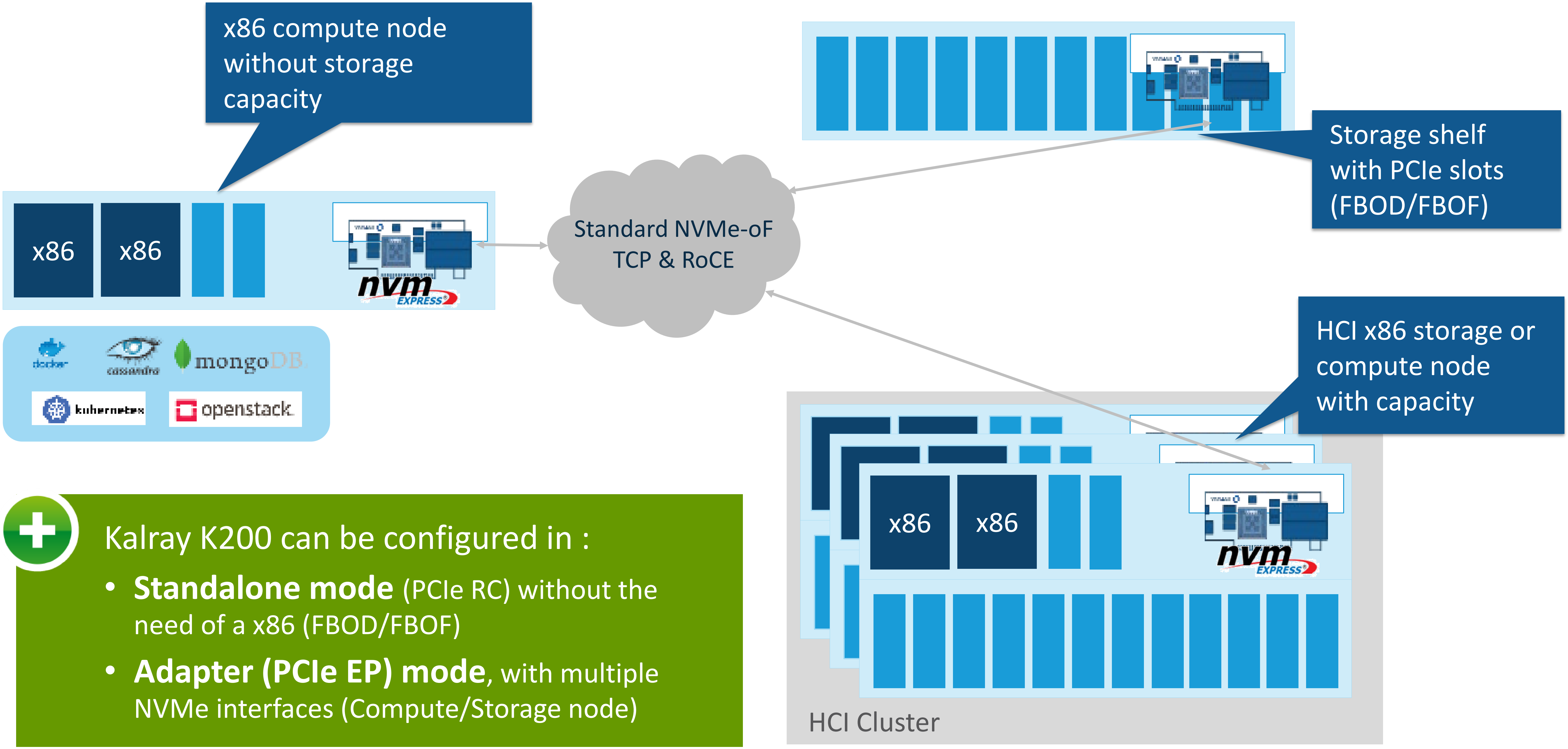
nvm EXPRESS®

Drives access through PCIe *(RC or P2P)*

nvm EXPRESS® · Serial Attached SCSI

KALRAY SMART STORAGE ADAPTER

Where Does It Fit?



x86 compute node without storage capacity

Storage shelf with PCIe slots (FBOD/FBOF)

Standard NVMe-oF TCP & RoCE

HCI x86 storage or compute node with capacity

Kalray K200 can be configured in :

- **Standalone mode** (PCIe RC) without the need of a x86 (FBOD/FBOF)
- **Adapter (PCIe EP) mode**, with multiple NVMe interfaces (Compute/Storage node)

EXAMPLE: LYMMA JBOF REFERENCE PLATFORM

White Label NVMe-oF (RoCE/TCP) JBOF

Hyper Optimized JBOF (no x86)

- JBOF Chassis:
  - Stand-alone
  - 2U – 1200W Redundant
  - 24 U.2 NVMe SSDs
  - 6xPCIe Gen3 x16
- Kalray Smart Controller Cards
  - 2 to 6 Cards
- BMC chip – AST2500 (ASpeed)
- 1Gbps management interface

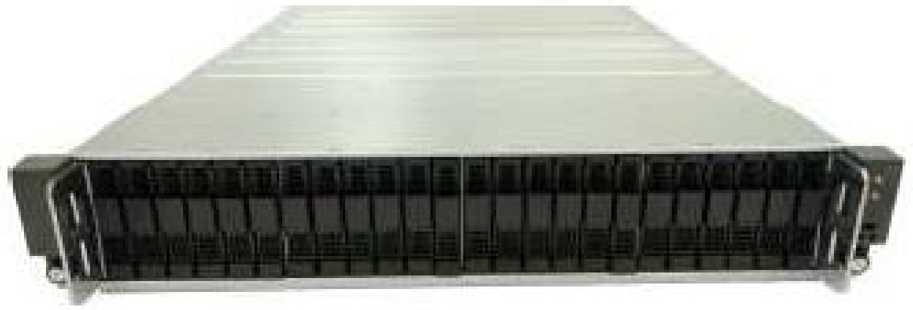
wistron

NVMe SSDs

Redundant Power

System Cooling FANs

PCIe Card Cages 12



Kalray AccessCore® Storage (ACS) Framework

SDK

MPPA®

The Processor at the Heart of Intelligent Systems

SDK **AccessCore®**
Open Software & Tools

ACCESSCORE FOR STORAGE & NETWORKING
ACS4.x Architecture Highlights

PROGRAMMABILITY

- Full programmability on data, control & management planes
  - Control & Management plane : Linux (typical : 1 Cluster - 16 cores)
  - Data plane : Cluster OS (light POSIX OS) (typical: 1 to 4 Clusters – 16 to 64 cores)

EFFICIENCY

- Run to completion full dataplane
  - From network functions to NVMe stack on light OS cores
- True inline processing
  - No need for x86 pre/post processing

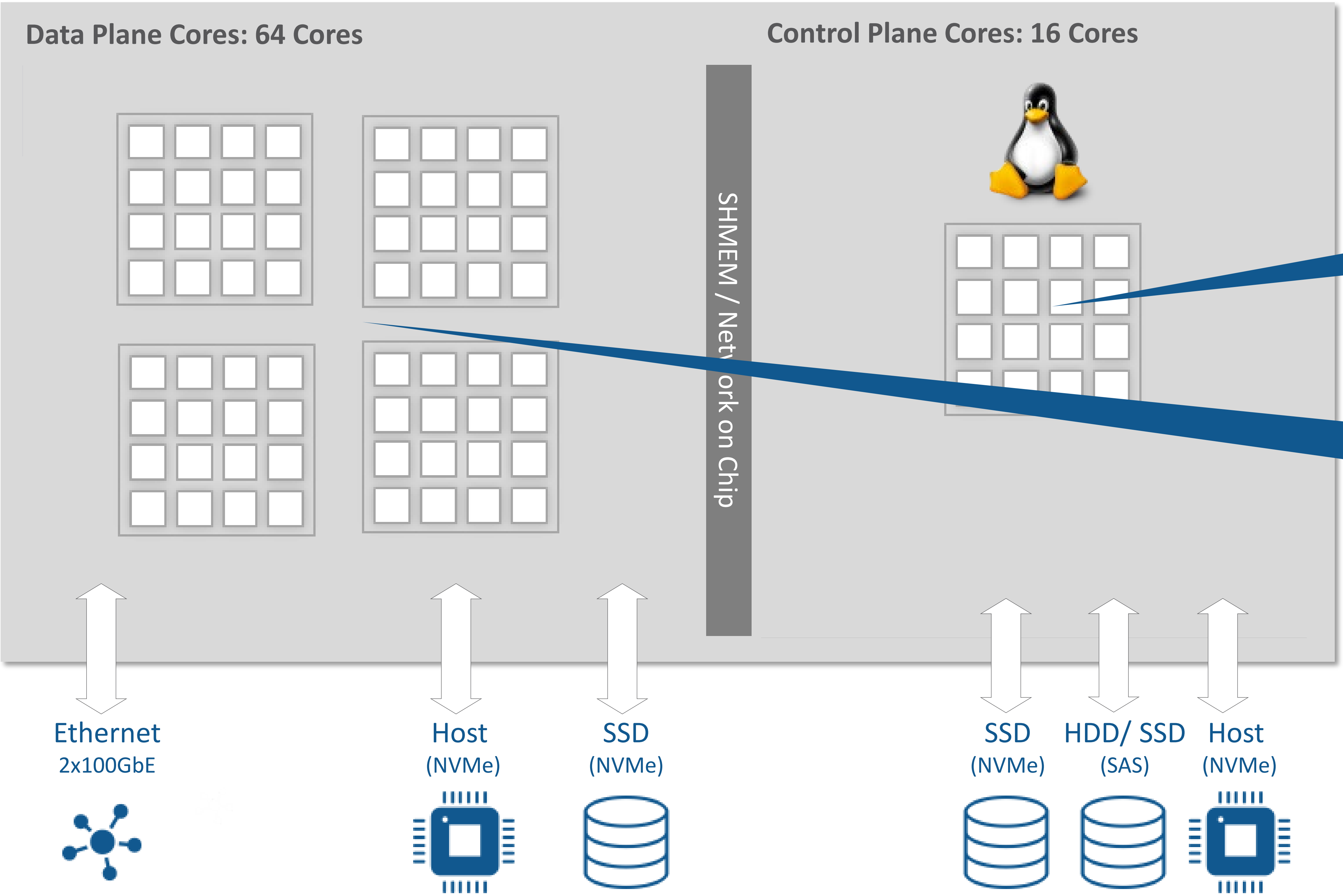
STANDARDIZED

- Hardware interfaces
  - NVMe emulation
- Software APIs & tool chain
  - Linux APIs: SPDK, virtio, ibverbs …
  - Data plane APIs: sockets, SPDK nvme lib, SPDK BDEV, ODP
  - Librairies : ISA-L, Buildroot, binutils

SDK AccessCore®
Open Software & Tools

ACS4.x ARCHITECTURE

A Fully Flexible Software Environment



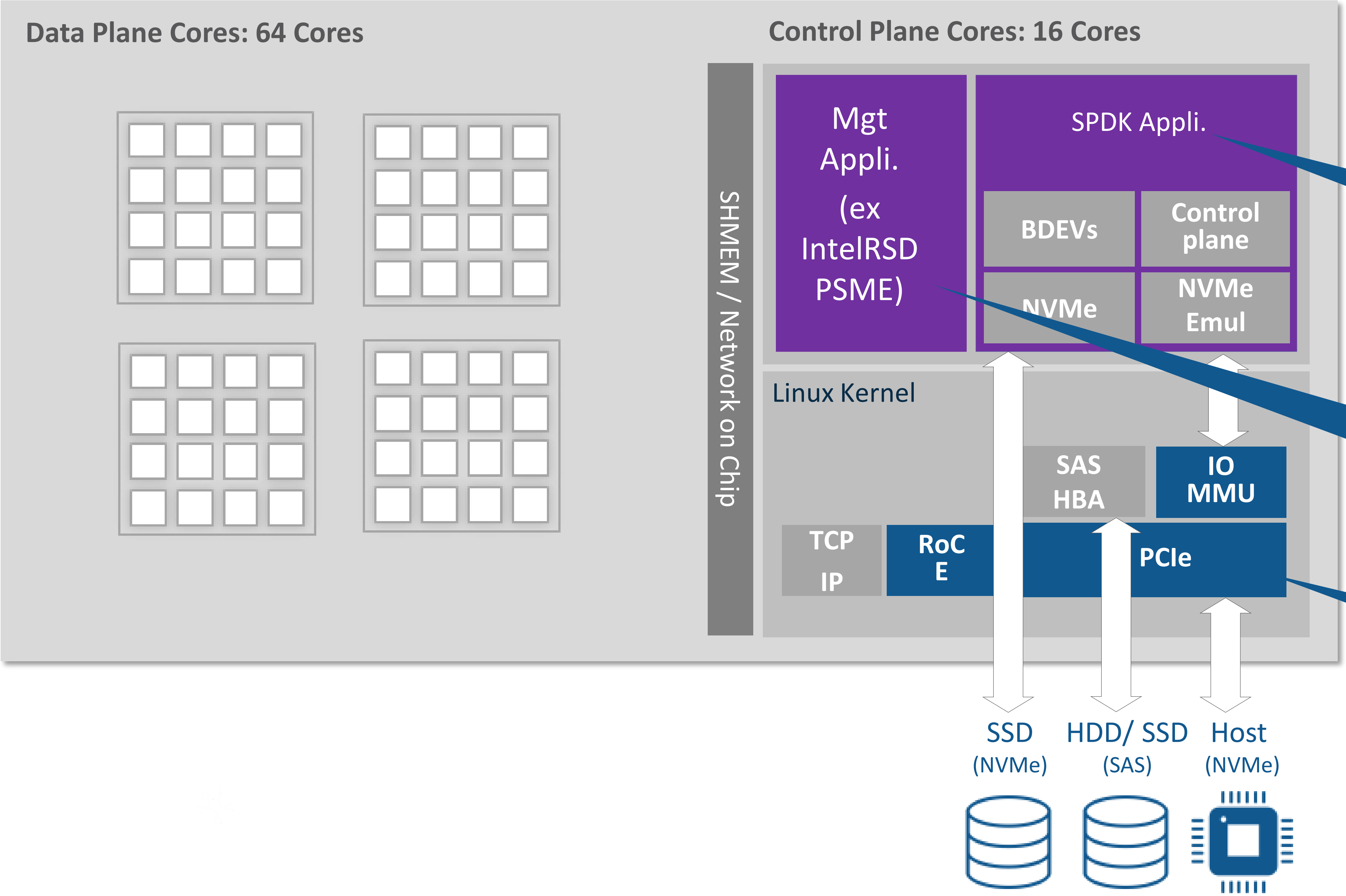
Linux on control plane cluster (16 cores)

ClusterOS on 4 data plane clusters (64 cores)

Legend: Kalray Drivers/Firmware | Kalray Frameworks | 3rd party Stack | Custom / Appli Code

ACS4.x ARCHITECTURE

A Fully Flexible Software Environment



SPDK on Linux user space for control plane ONLY

Redfish/Swordfish support via IntelRSD

PCIe P2P : PCIe RC emulation to control x86 attached devices (NVMe, SAS HBA)

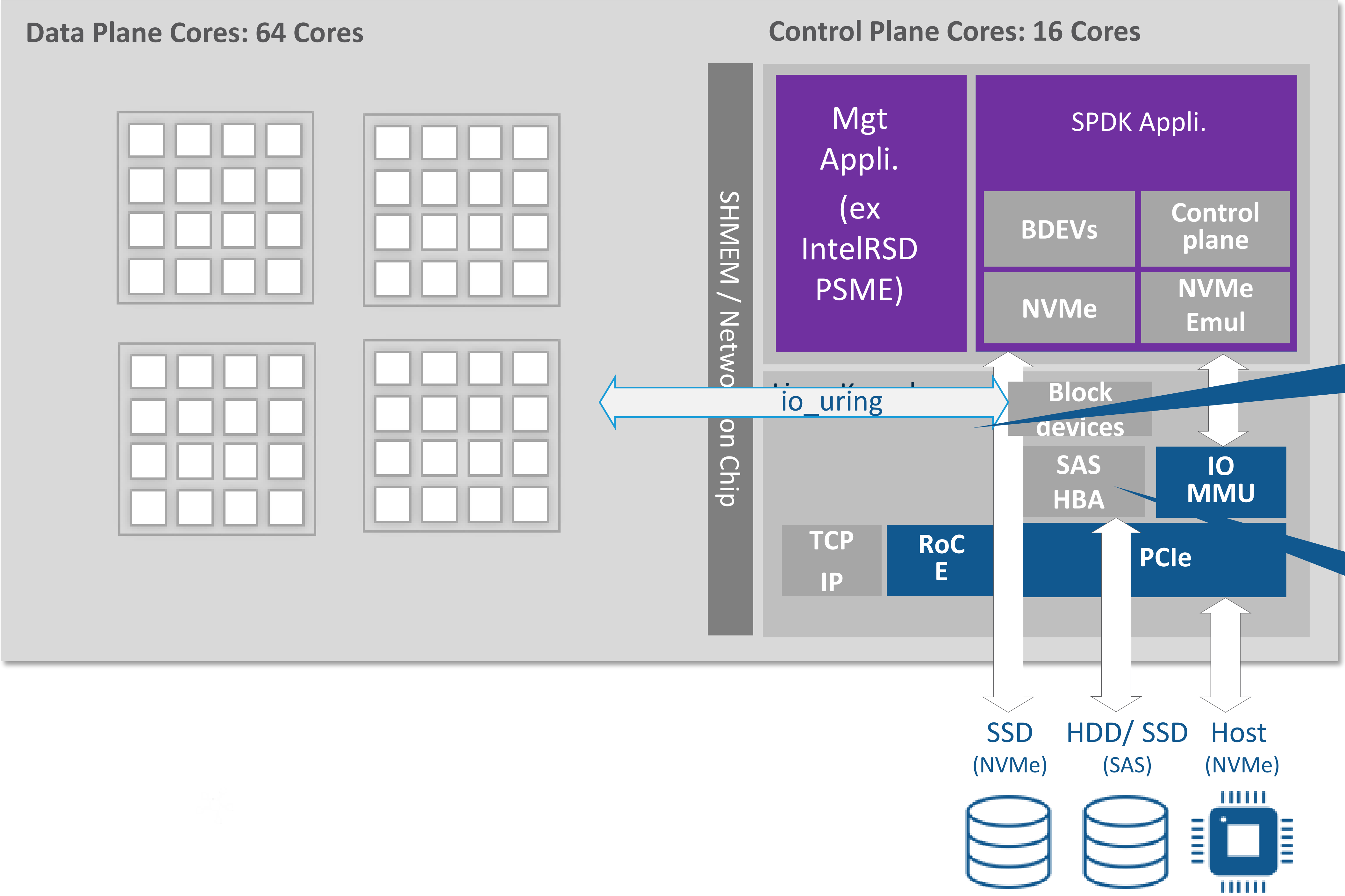
Legend

🟦	Kalray Drivers/Firmware
🟩	Kalray Frameworks
🟦	3rd party Stack
🟪	Custom / Appli Code

SDK AccessCore® Open Software & Tools

ACS4.x ARCHITECTURE

A Fully Flexible Software Environment



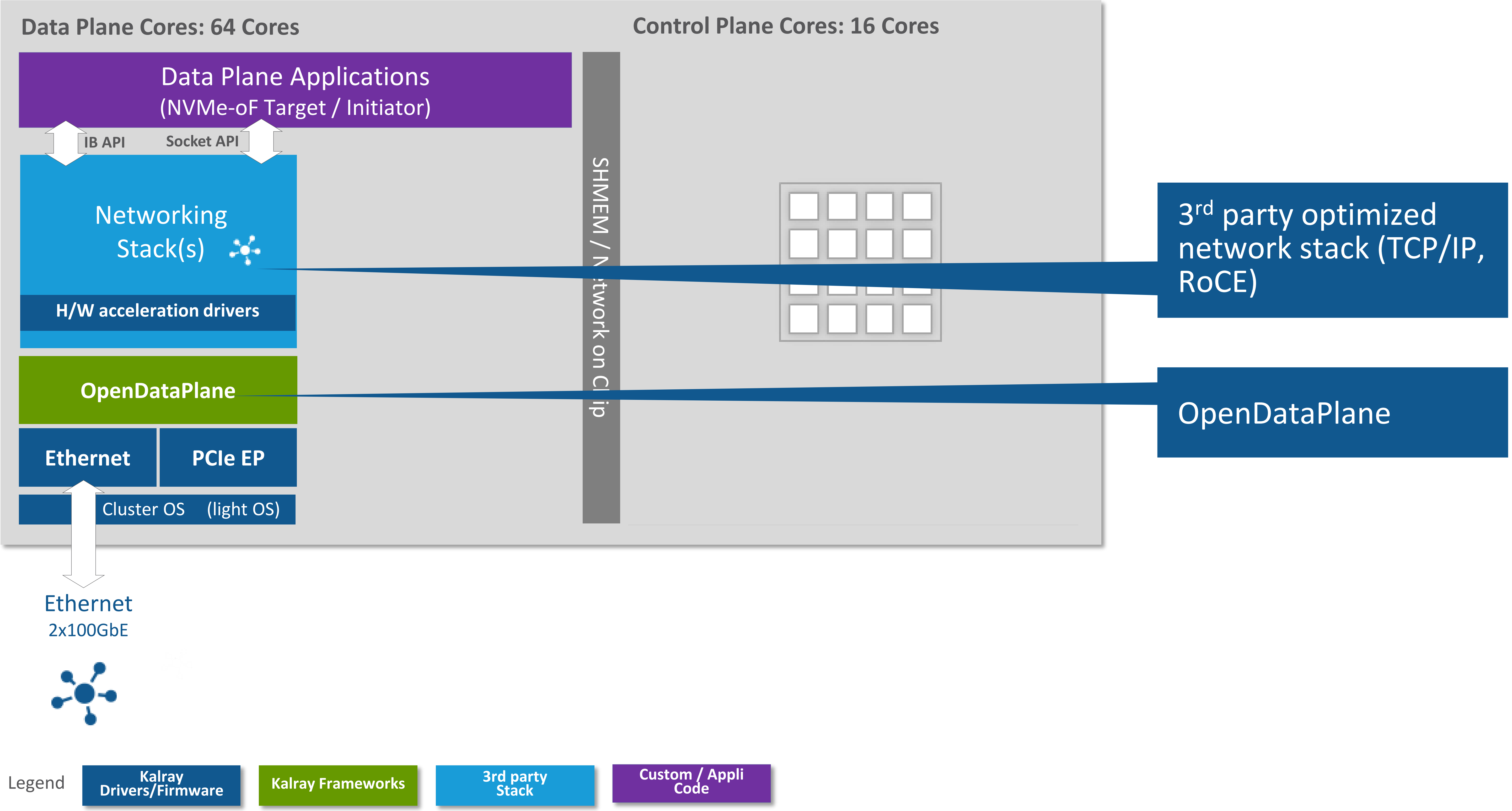
ONLY for SAS : Data plane comm. using io_uring (1MIOPs @4K)

SAS HBA driver for external add in card

Legend	Kalray Drivers/Firmware	Kalray Frameworks	3rd party Stack	Custom / Appli Code

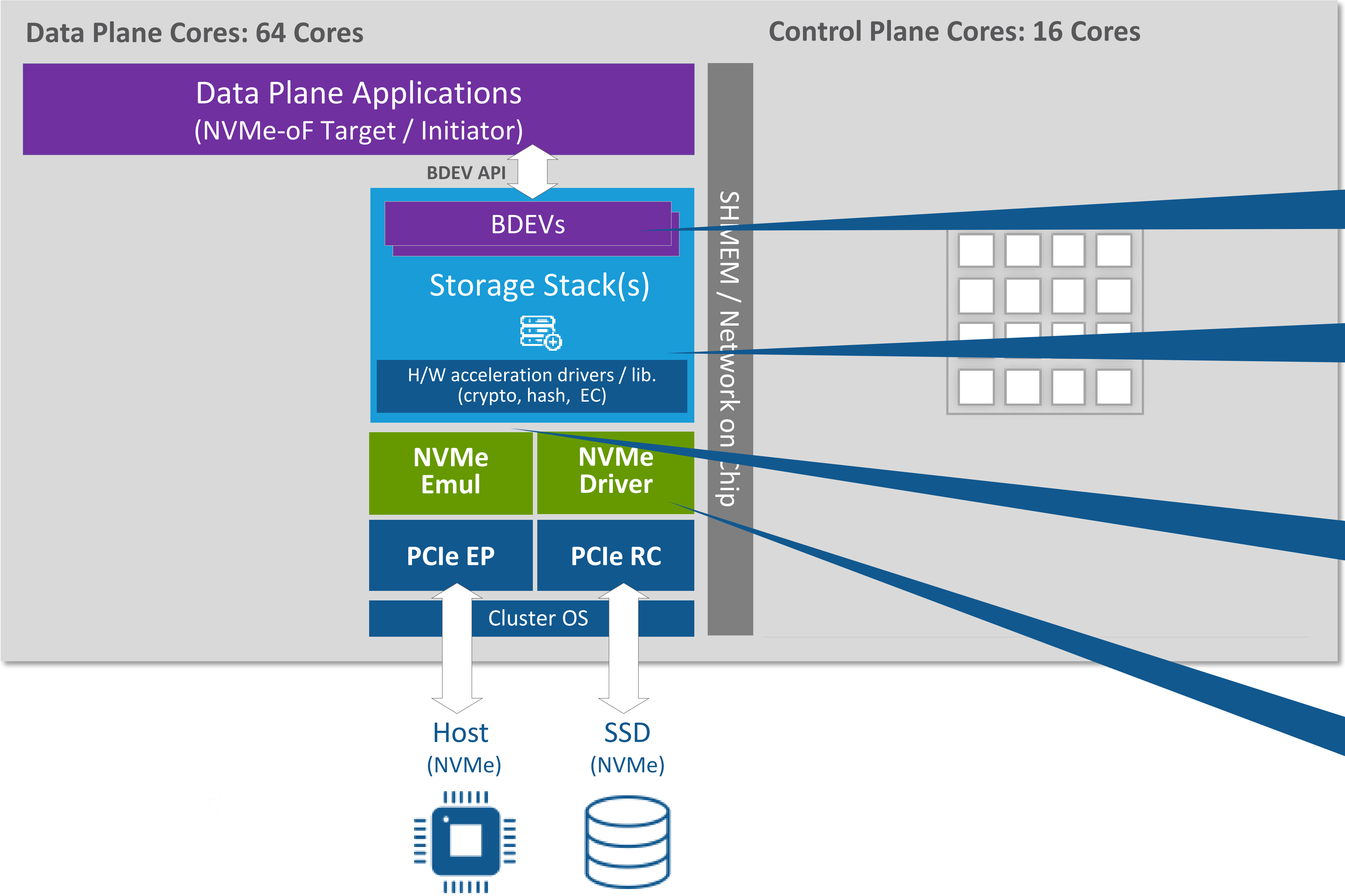
SDK AccessCore® Open Software & Tools

ACS4.x ARCHITECTURE
A Fully Flexible Software Environment



ACS4.x ARCHITECTURE

A Fully Flexible Software Environment



SPDK on Cluster OS BDEVs porting over data plane

ISA-L compliant lib for Erasure Coding, deflate

NVMe emulation (up to 256 controllers with SR-IOV)

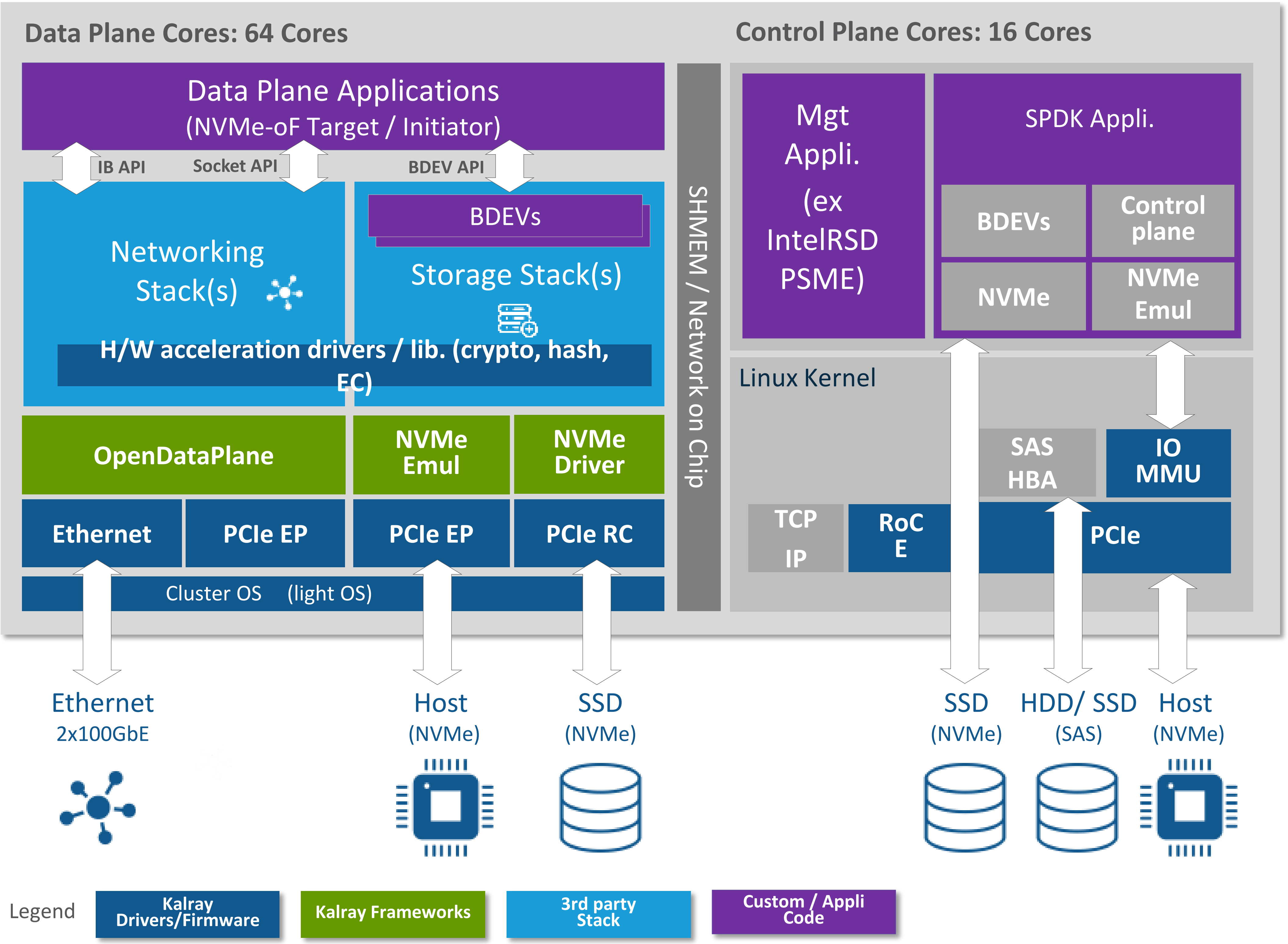
NVMe I/O queues data access

Legend

Kalray Drivers/Firmware	Kalray Frameworks	3rd party Stack	Custom / Appli Code

ACS4.x ARCHITECTURE

A Fully Flexible Software Environment



- A complete & modular software framework
- Based on an optimized SPDK for both data plane **AND** control plane
- Open to partners

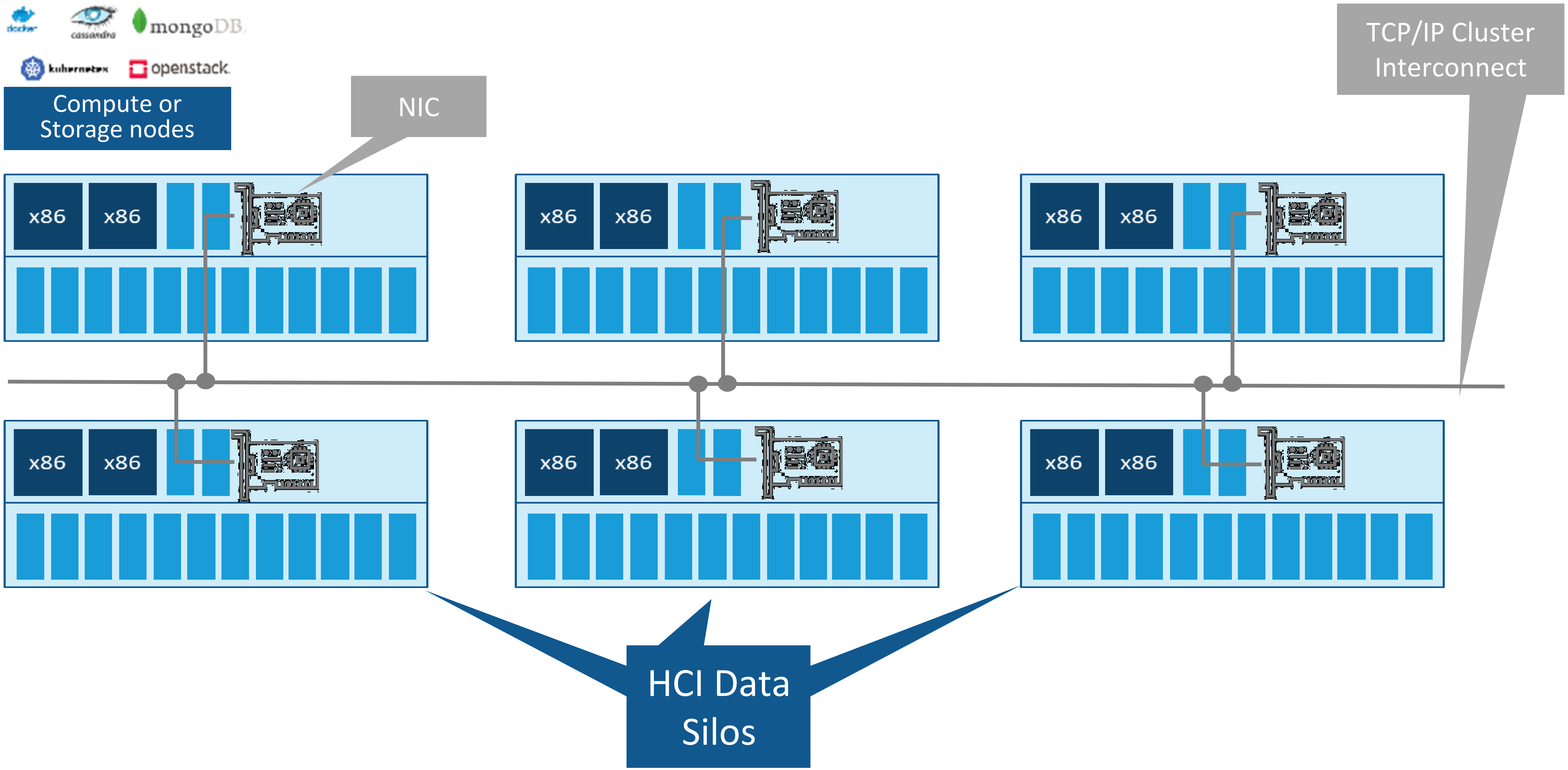
USE CASE: Future HyperConverged Infrastructure

MPPA®

The Processor at the Heart
of Intelligent Systems

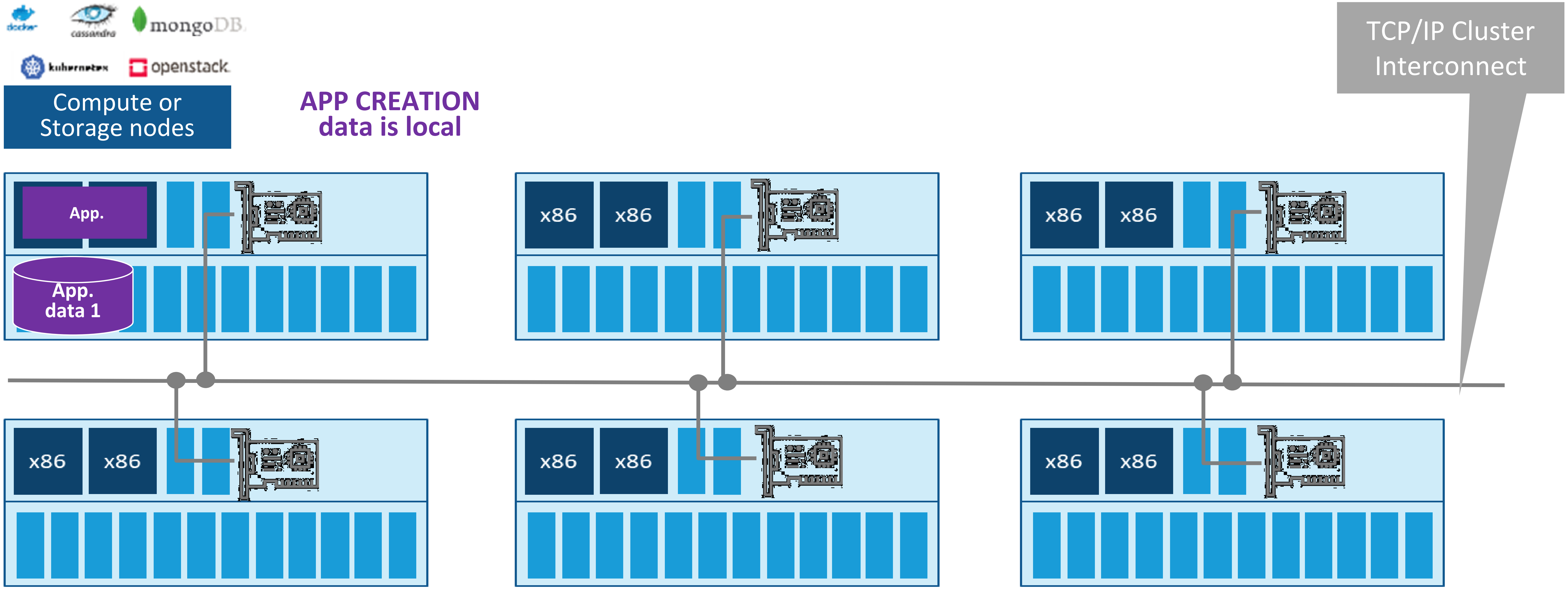
CURRENT HCI TOPOLOGY

HCI Silos Imposes Disaggregation by Software



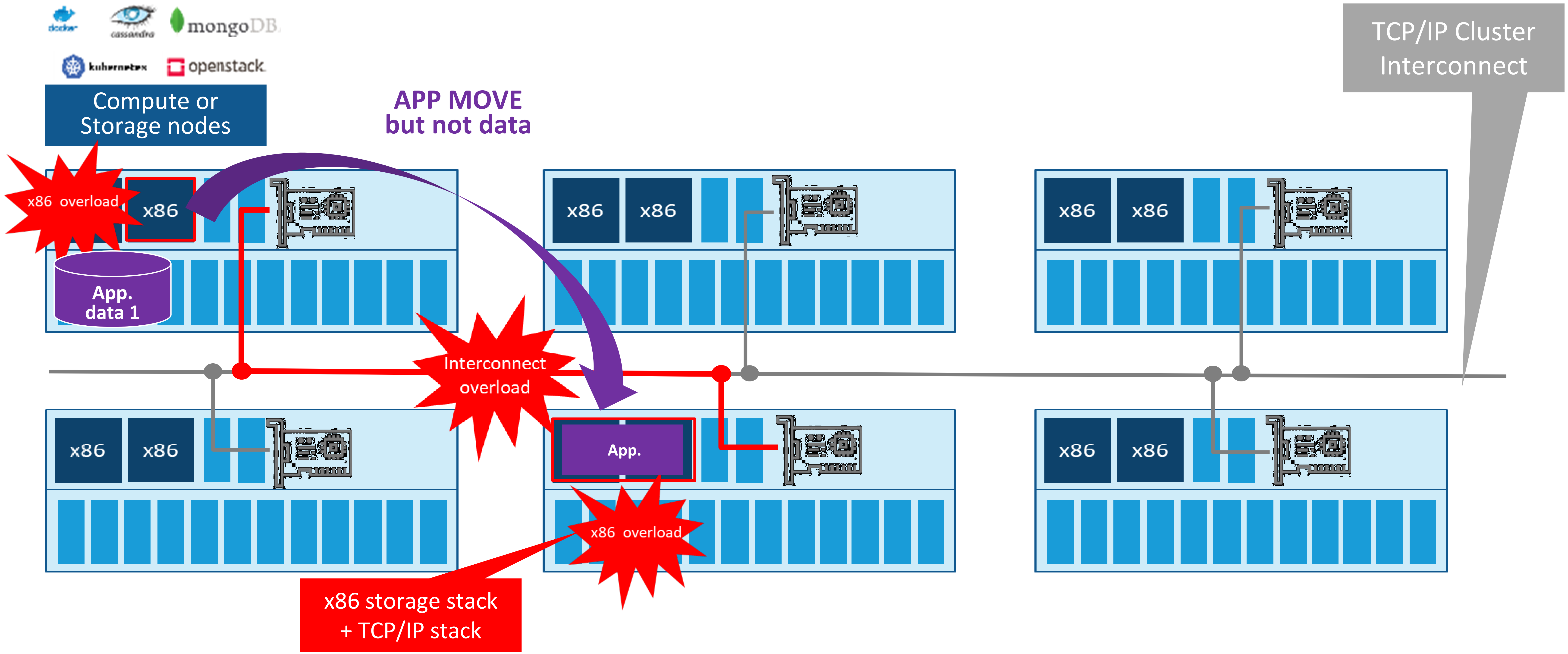
CURRENT HCI TOPOLOGY

VMs / Services are Spread across HCI Silos



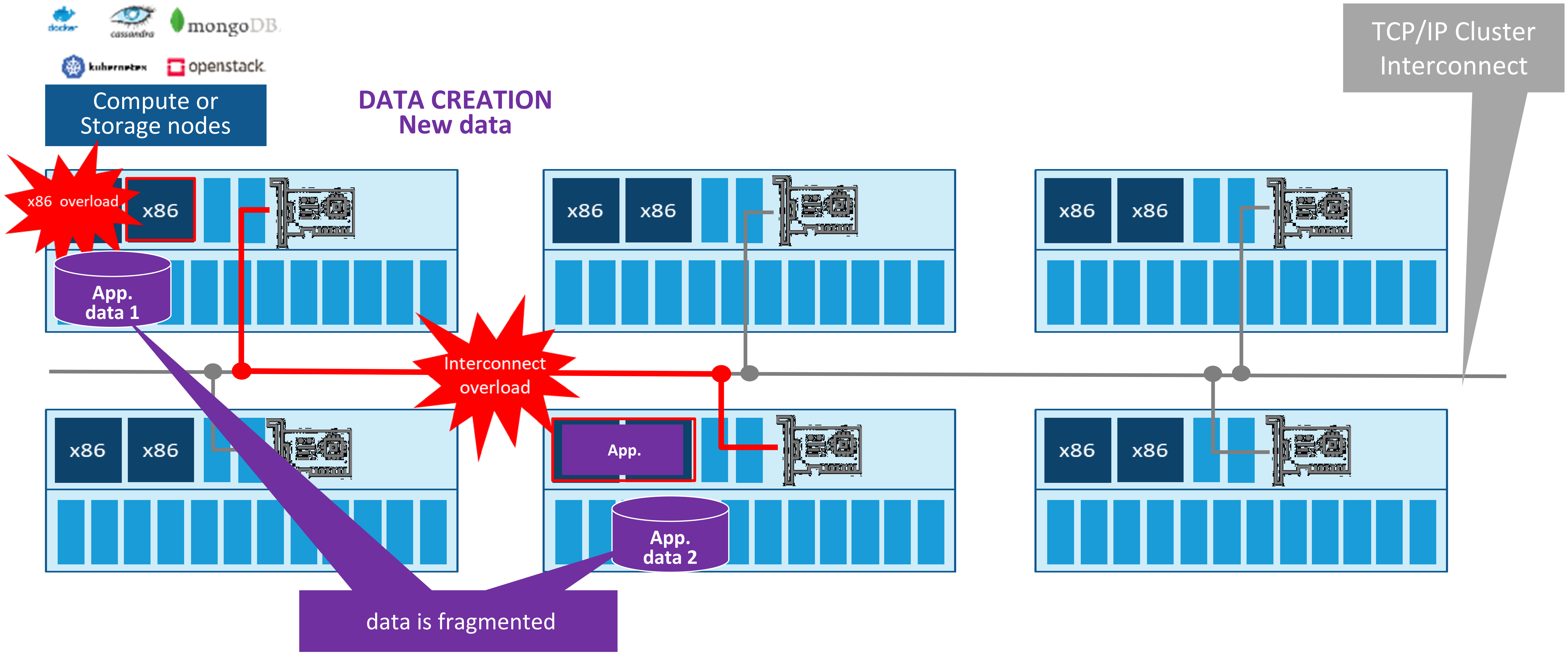
CURRENT HCI TOPOLOGY

Impact: Overload of Interconnect and CPUs



CURRENT HCI TOPOLOGY

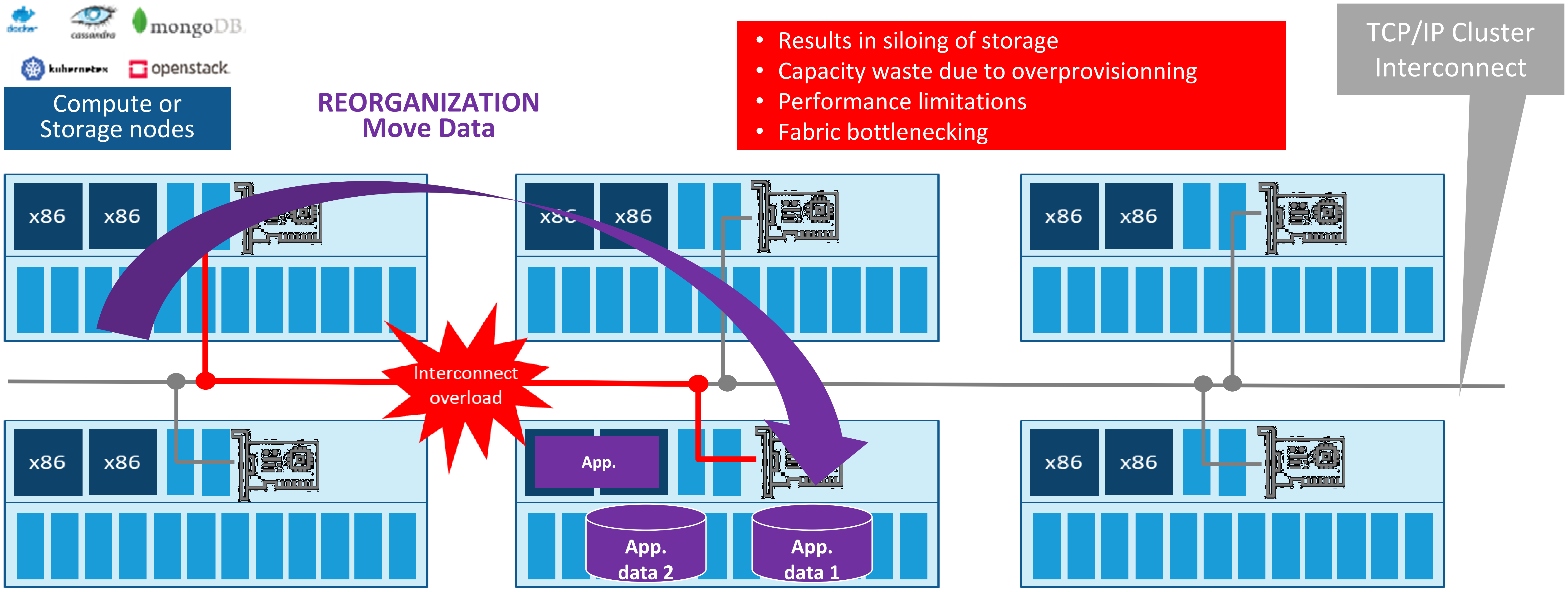
Fragmented Data Set



CURRENT HCI TOPOLOGY

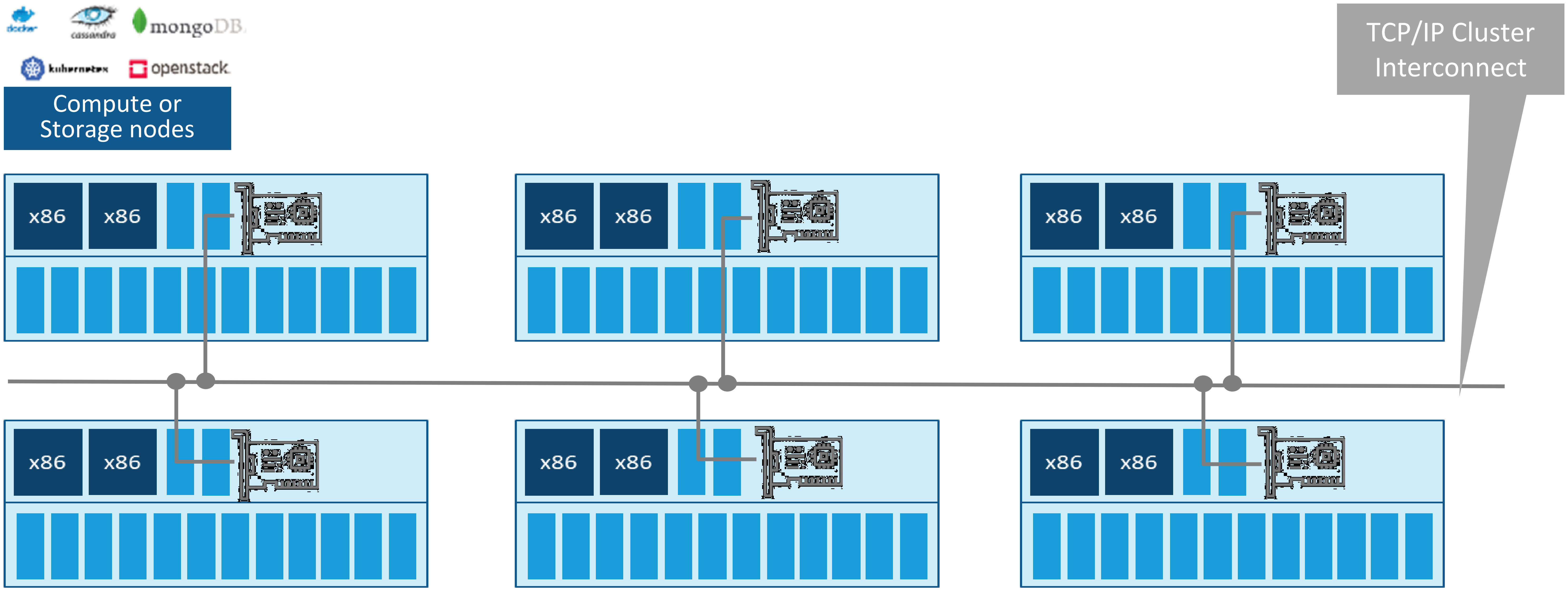
Data Relocation is Compulsory

- Results in siloing of storage
- Capacity waste due to overprovisionning
- Performance limitations
- Fabric bottlenecking



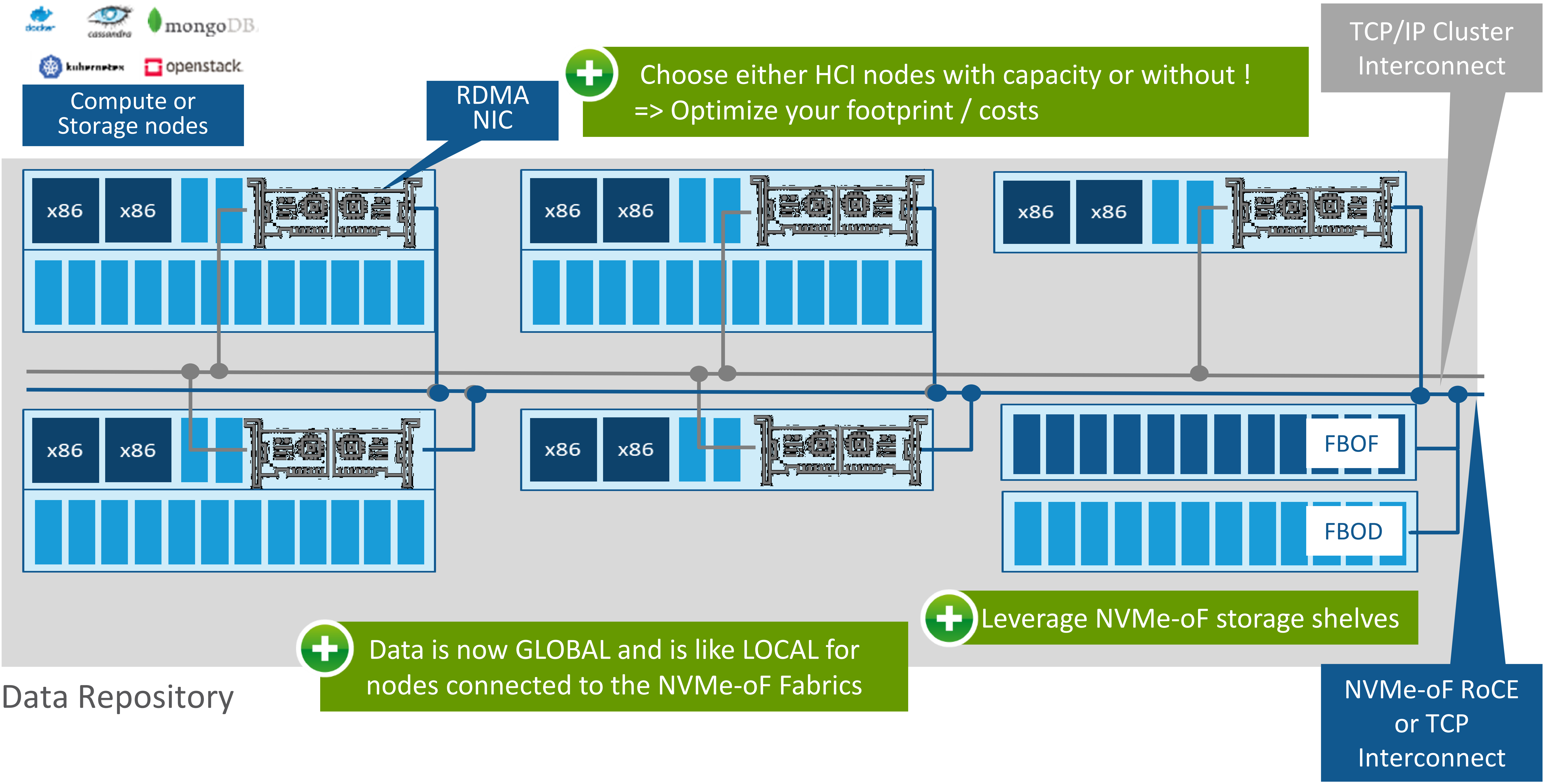
HCI WITH NVME-OF

NVMe-oF Removes Some Boundaries



HCI WITH NVMe-oF

Leveraging NVMe-oF Storage Appliances



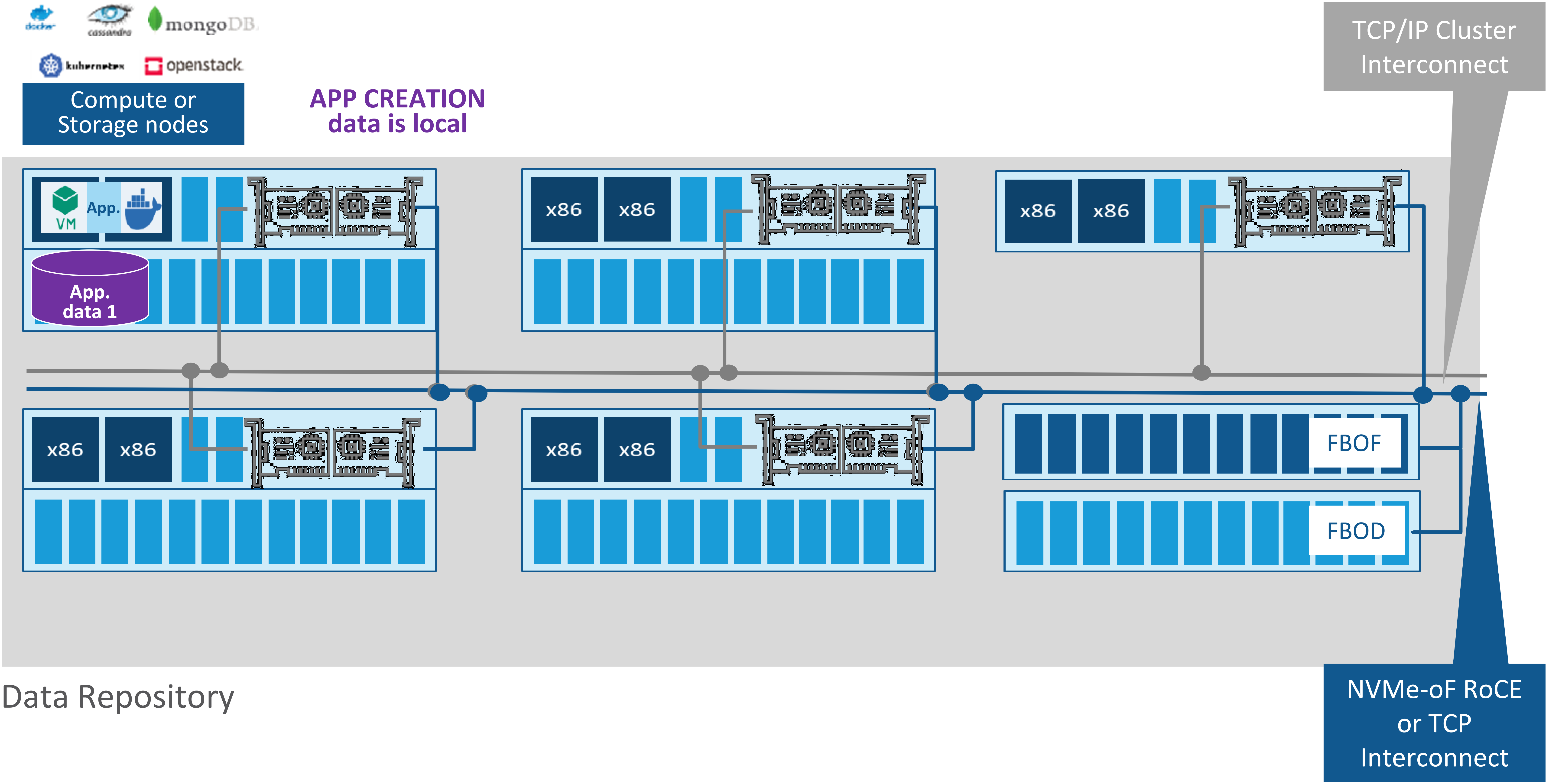
Choose either HCI nodes with capacity or without !
=> Optimize your footprint / costs

Leverage NVMe-oF storage shelves

Data is now GLOBAL and is like LOCAL for nodes connected to the NVMe-oF Fabrics

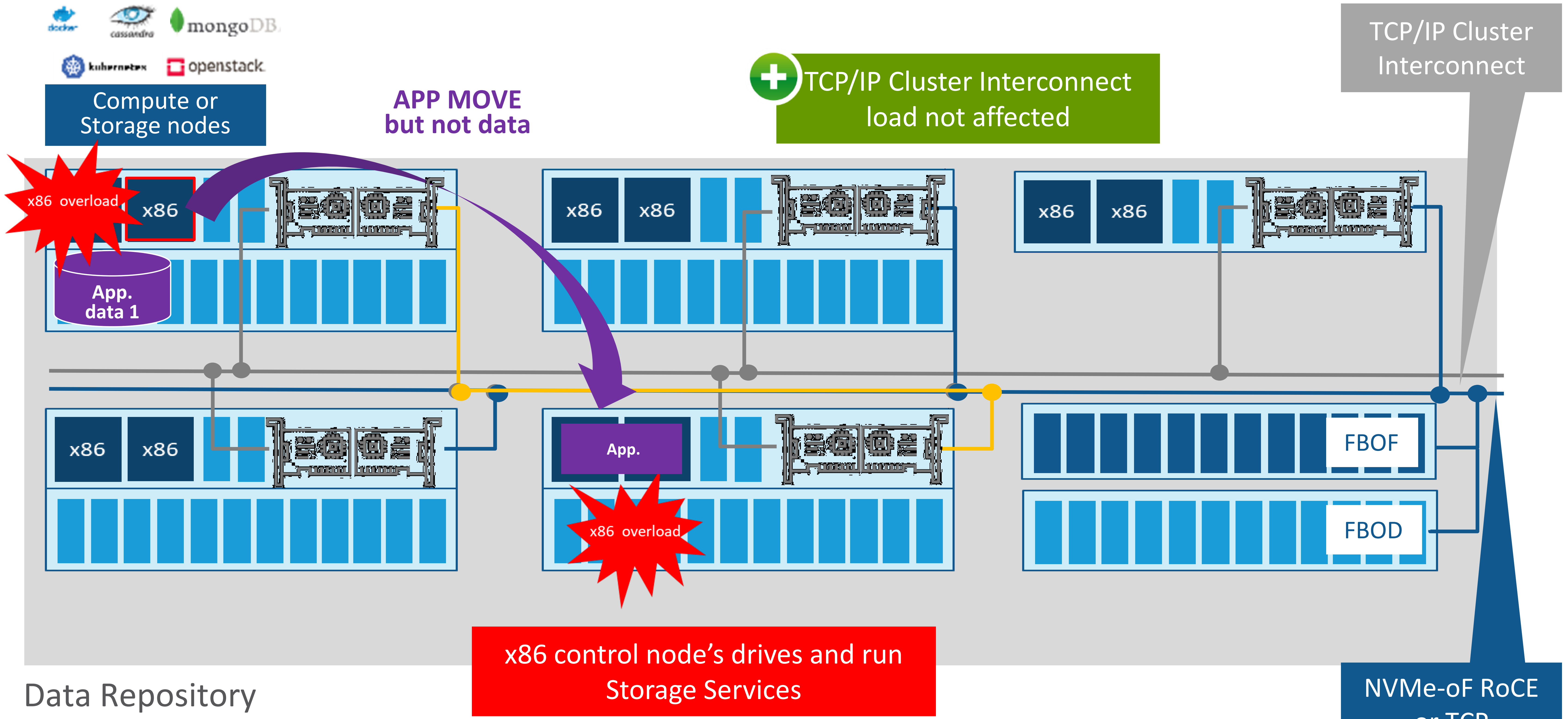
HCI WITH NVMe-oF

NVMe-oF Removes Some Boundaries



HCI WITH NVMe-oF

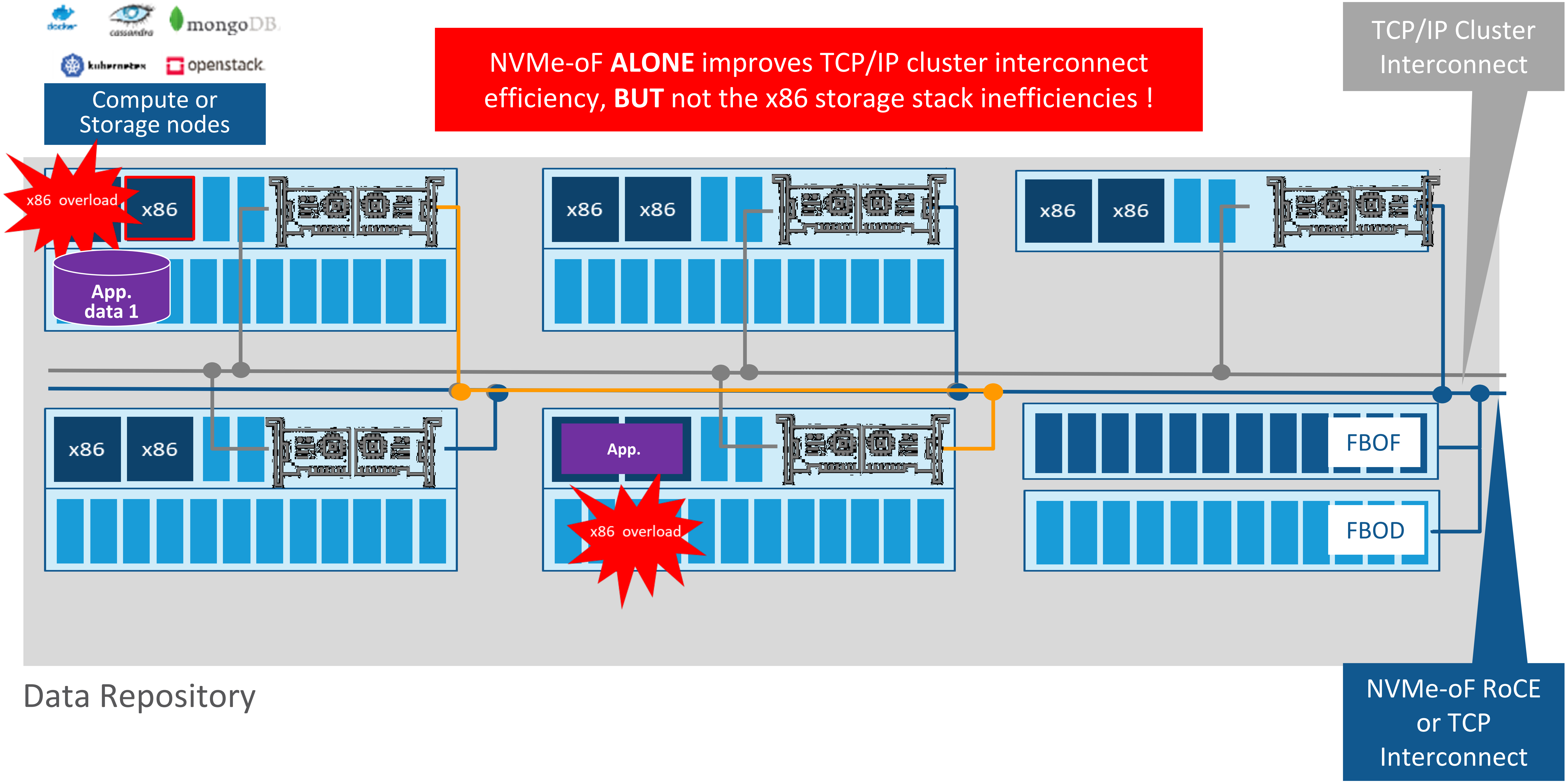
NVMe-oF Removes Interconnect Overload but ...



HCI WITH NVMe-oF

NVMe-oF Limitations

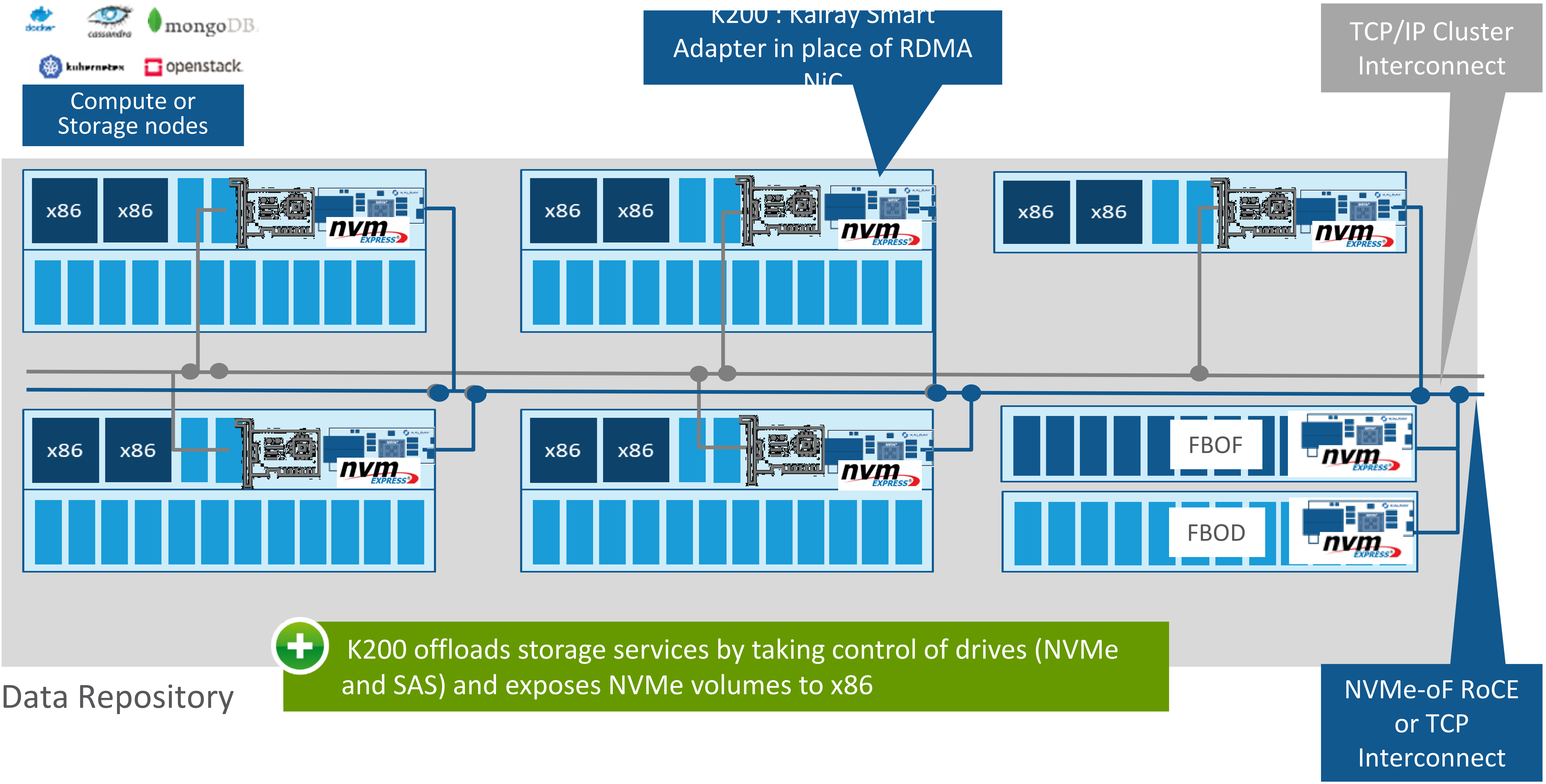
NVMe-oF **ALONE** improves TCP/IP cluster interconnect efficiency, **BUT** not the x86 storage stack inefficiencies !



Data Repository

HCI WITH KALRAY

Kalray Smart Adapter Enables New HCI Topology



Compute or Storage nodes

K200 : Kalray Smart Adapter in place of RDMA NiC

TCP/IP Cluster Interconnect

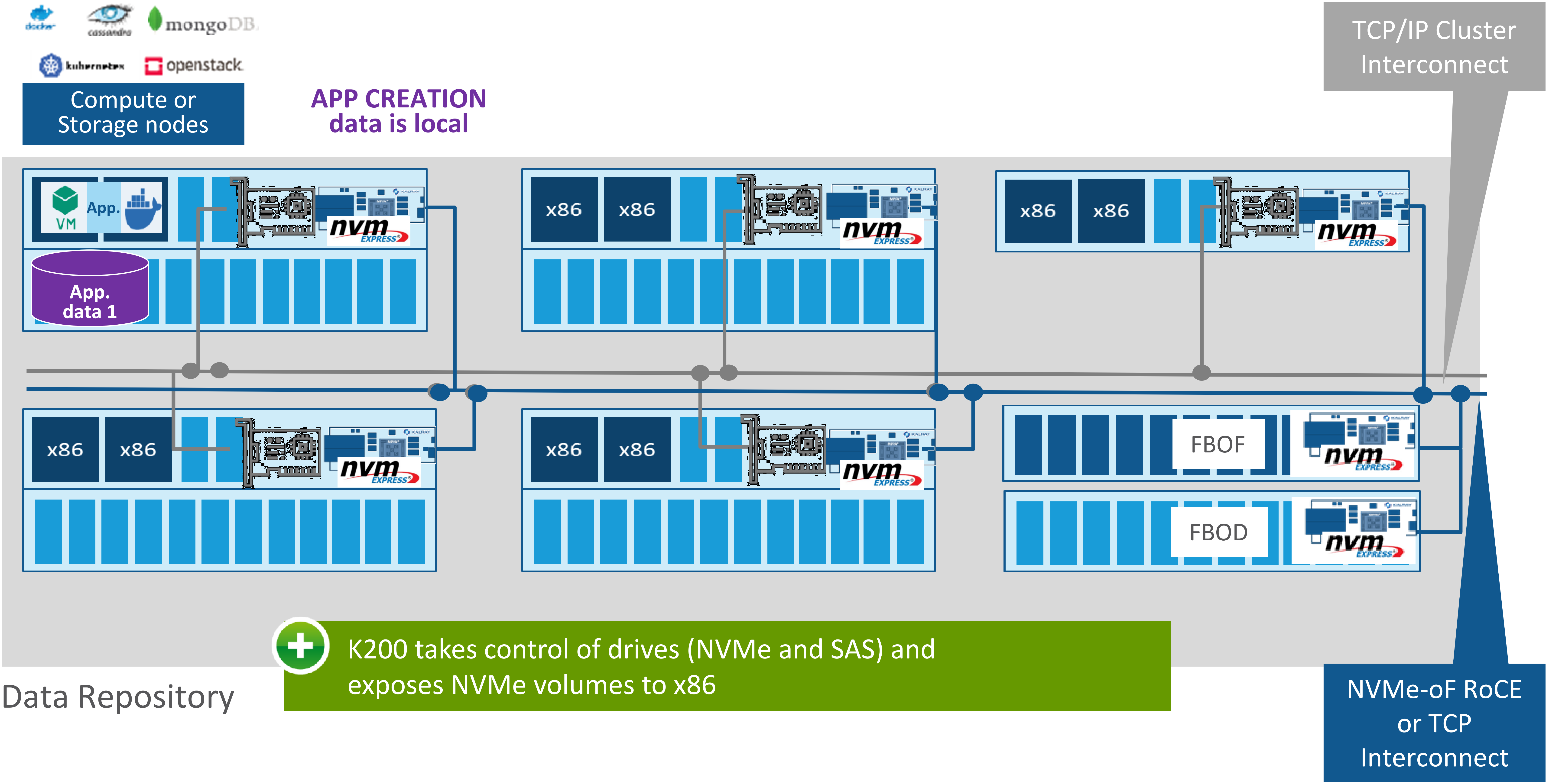
NVMe-oF RoCE or TCP Interconnect

Data Repository

K200 offloads storage services by taking control of drives (NVMe and SAS) and exposes NVMe volumes to x86

HCI WITH KALRAY

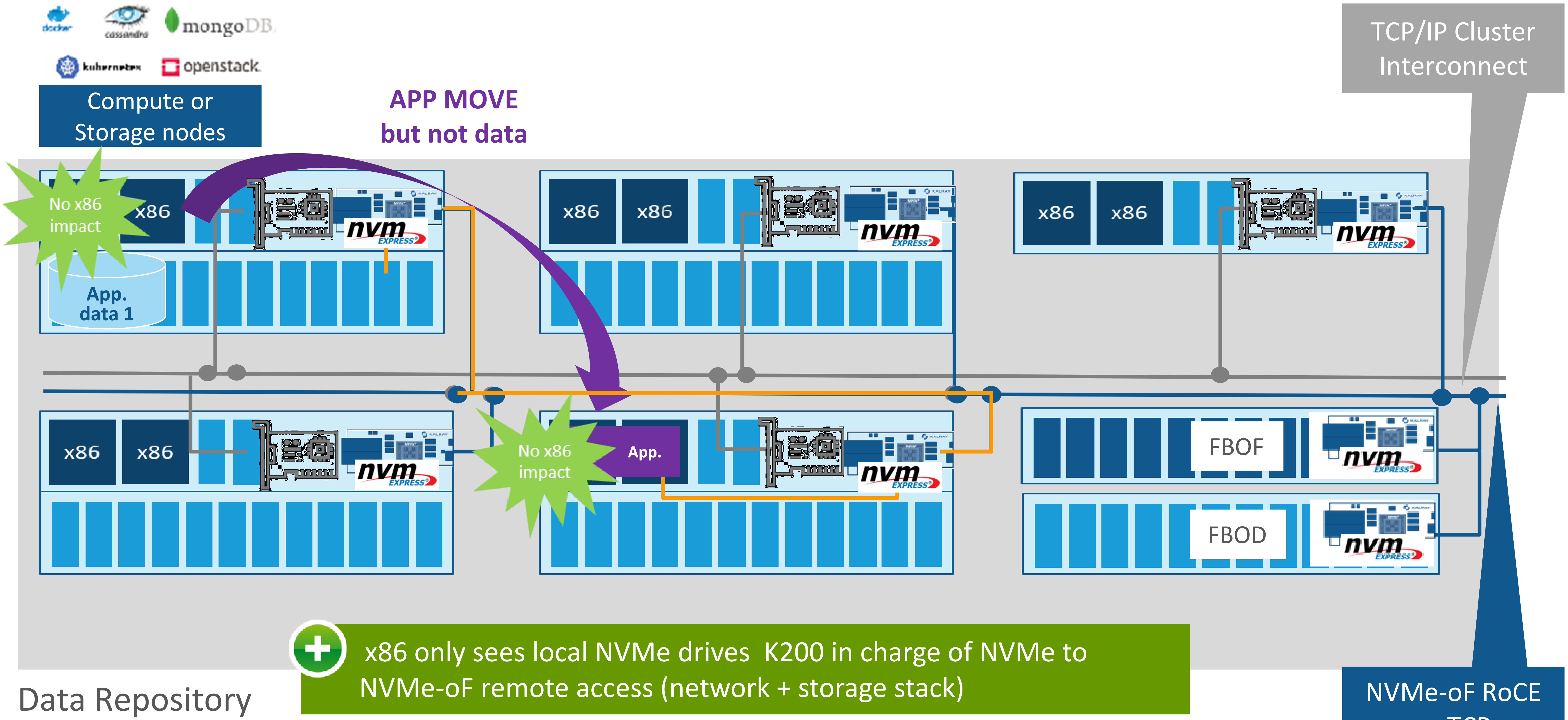
Drives are Under Kalray Storage Adapter Control



K200 takes control of drives (NVMe and SAS) and exposes NVMe volumes to x86

HCI WITH KALRAY

Nodes Only See Local NVMe Drives

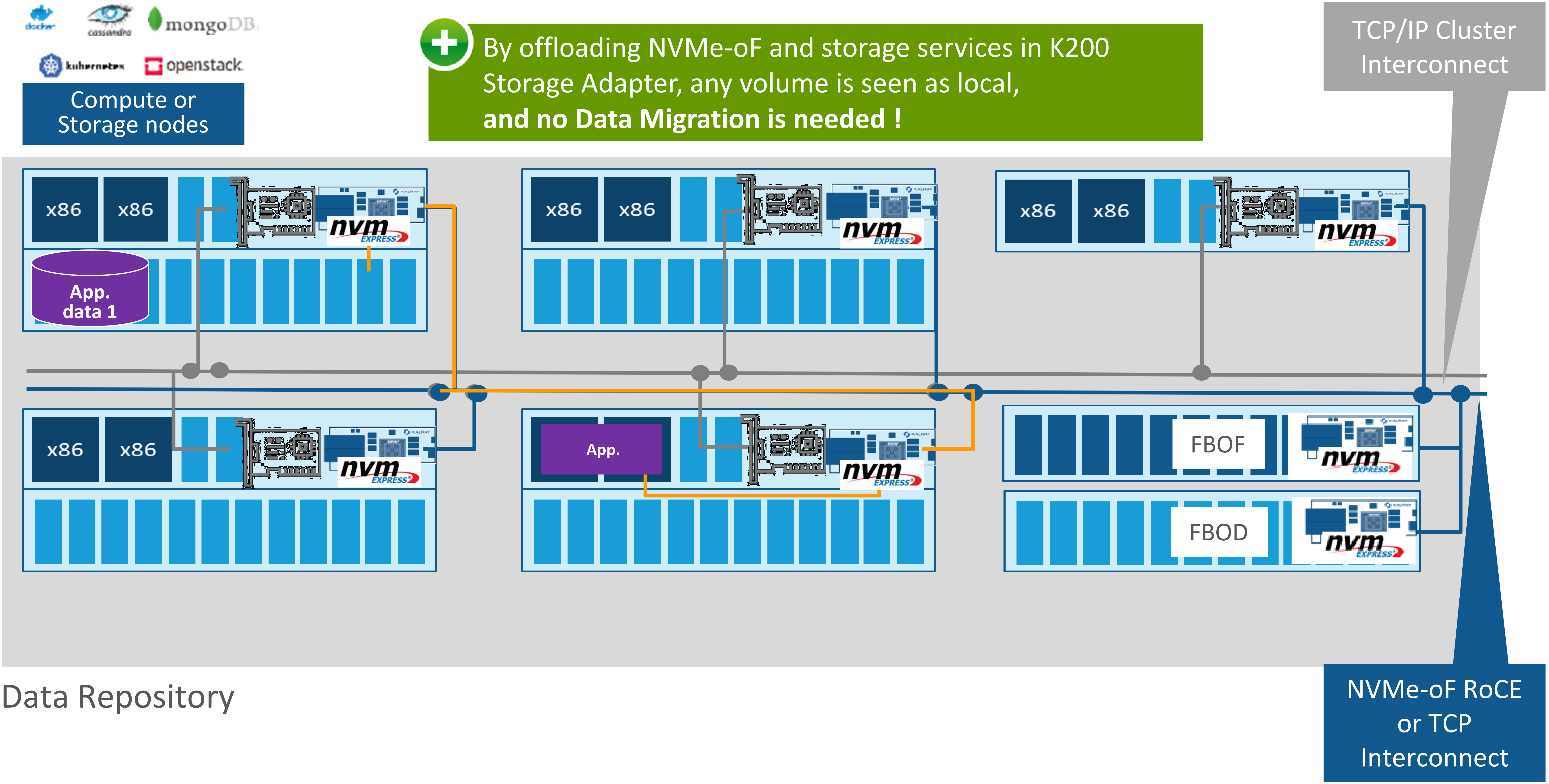


x86 only sees local NVMe drives K200 in charge of NVMe to NVMe-oF remote access (network + storage stack)

HCI WITH KALRAY

No More Data Migration Required

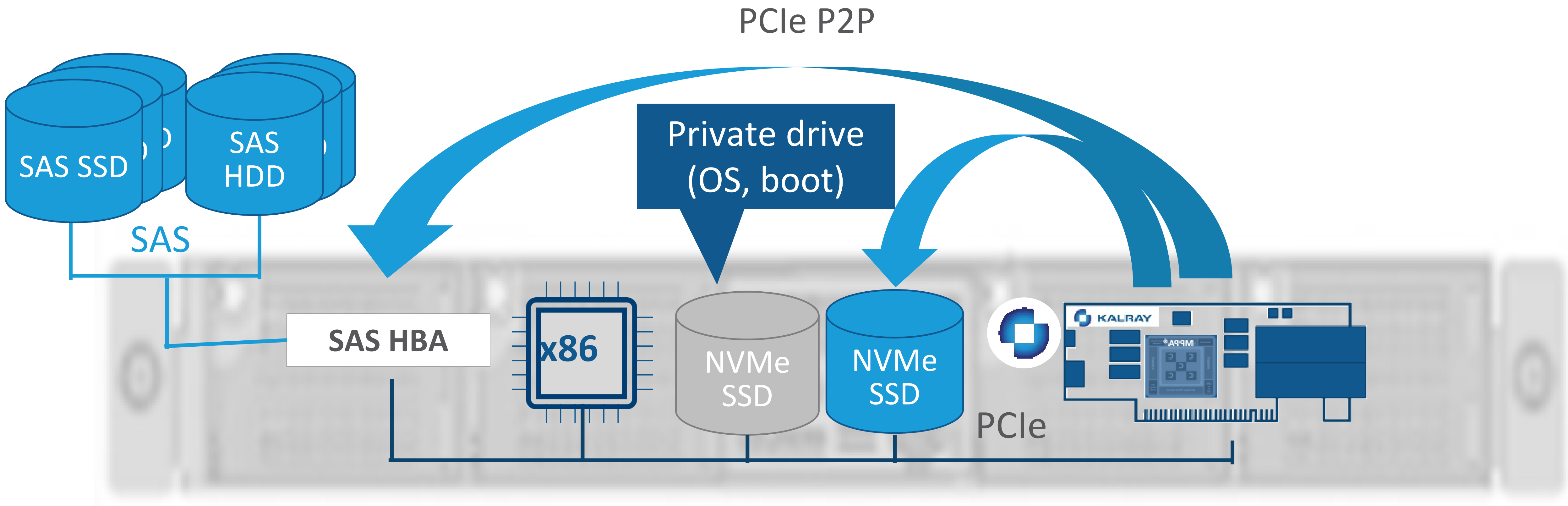
By offloading NVMe-oF and storage services in K200 Storage Adapter, any volume is seen as local, **and no Data Migration is needed !**



Data Repository

COMPOSABLE ARCHITECTURE WITH KALRAY ADAPTERS

x86 Node System Architecture

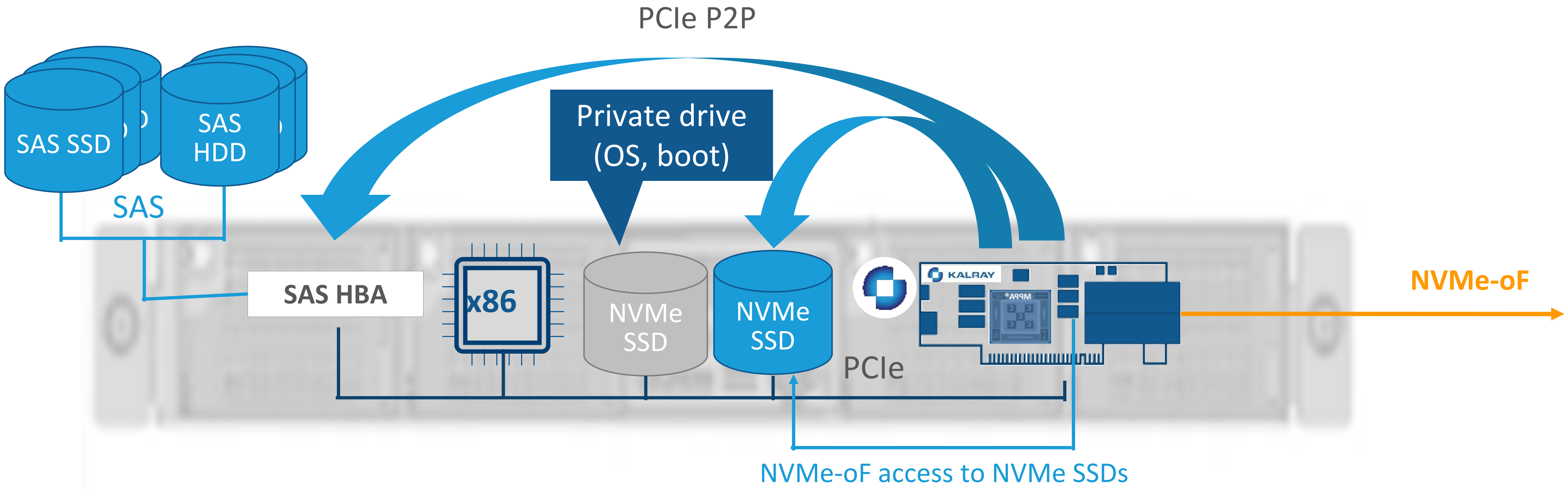


K200 Smart Adpater offloads x86 from NVMe-oF & storage services

- NVMe-oF Remote access to node's drives (NVMe / SAS) without x86 involvement
- Local access to node's drive (NVMe / SAS) via storage adapter (NVMe emulation)
- Storage Services added by Kalray Adapter :
  - Caching
  - Distributed Erasure Coding
  - High avalability

COMPOSABLE ARCHITECTURE WITH KALRAY ADAPTERS

x86 Node System Architecture

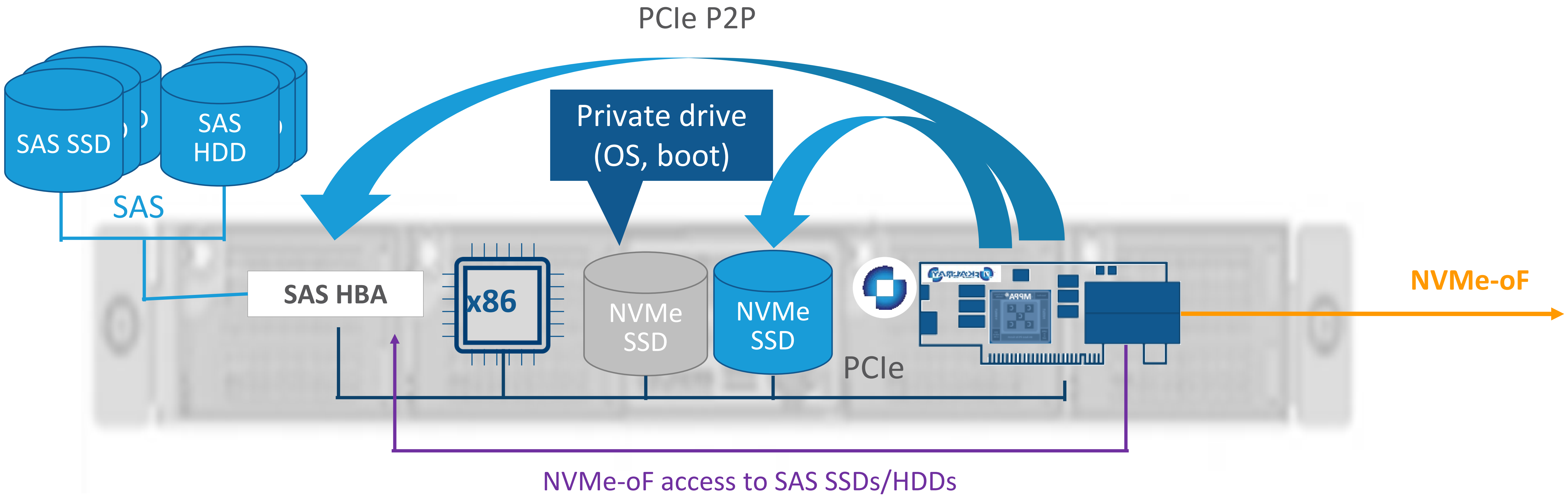


K200 Smart Adpater offloads x86 from NVMe-oF & storage services:

- NVMe-oF Remote access to node's drives (NVMe / SAS) without x86 involvement
- Local access to node's drive (NVMe / SAS) via storage adapter (NVMe emulation)
- Storage Services added by Kalray Adapter :
  - Caching
  - Distributed Erasure Coding
  - High avalability

COMPOSABLE ARCHITECTURE WITH KALRAY ADAPTERS

x86 Node System Architecture

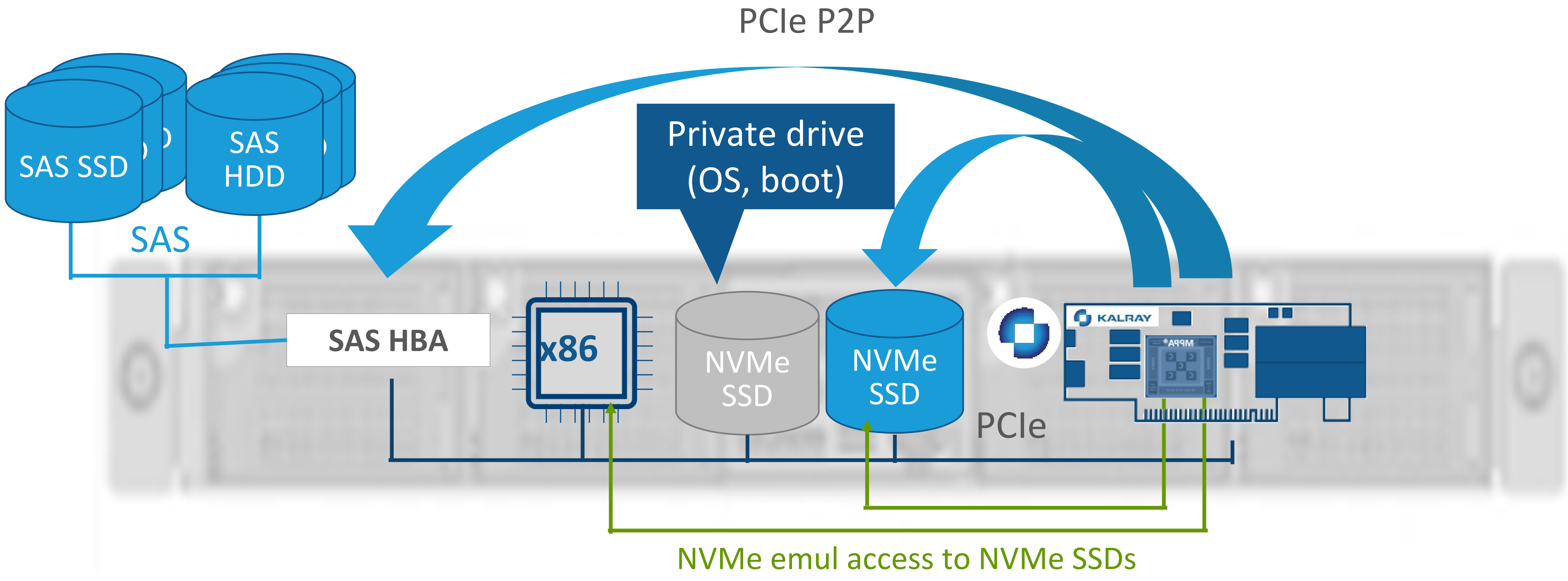


K200 Smart Adpater offloads x86 from NVMe-oF & storage services:

- NVMe-oF Remote access to node’s drives (NVMe / SAS) without x86 involvement
- Local access to node’s drive (NVMe / SAS) via storage adapter (NVMe emulation)
- Storage Services added by Kalray Adapter :
  - Caching
  - Distributed Erasure Coding
  - High avalability

COMPOSABLE ARCHITECTURE WITH KALRAY ADAPTERS

x86 Node System Architecture

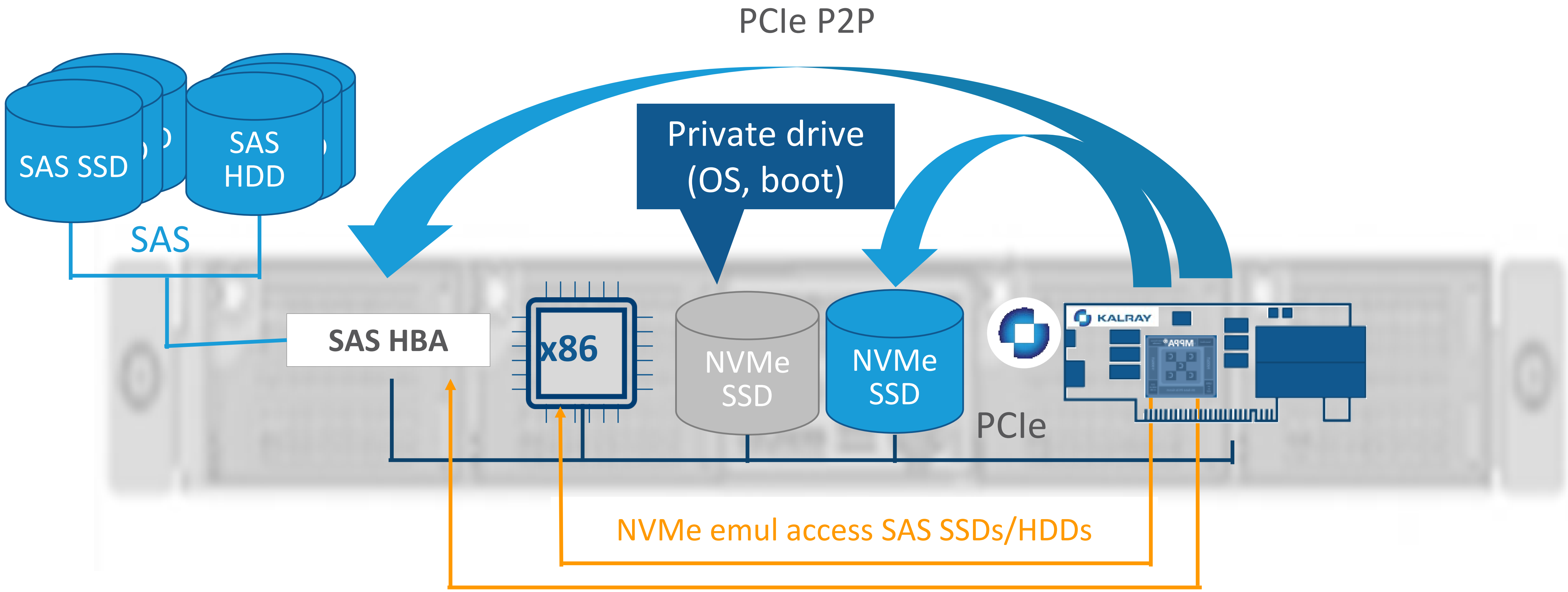


K200 Smart Adpater offloads x86 from NVMe-oF & storage services:

- NVMe-oF Remote access to node's drives (NVMe / SAS) without x86 involvement
- Local access to node's drive (NVMe / SAS) via storage adapter (NVMe emulation)
- Storage Services added by Kalray Adapter :
  - Caching
  - Distributed Erasure Coding
  - High avalability

COMPOSABLE ARCHITECTURE WITH KALRAY ADAPTERS

x86 Node System Architecture



K200 Smart Adpater offloads x86 from NVMe-oF & storage services:

- NVMe-oF Remote access to node’s drives (NVMe / SAS) without x86 involvement
- Local access to node’s drive (NVMe / SAS) via storage adapter (NVMe emulation)
- Storage Services added by Kalray Adapter :
  - Caching
  - Distributed Erasure Coding
  - High avalability

Conclusion

MPPA®

The Processor at the Heart

of Intelligent Systems

TOWARD A TRUE & EFFICIENT COMPOSABLE DISAGGREGATED INFRASTRUCTURE

HIGHER PERFORMANCE

- Leverage Kalray cards performance and exploit full NVMe SSD capabilities
- Offload x86 from heavy storage stacks

LOWER COST

- Switch to a true **C**omposable **D**isaggregated **I**nfrastructure with commodity components
- Optimize HCI nodes efficiency

FULLY FLEXIBLE

- Fully programmable data plane
- Data Plane additional storage services based on SPDK framework (EC, caching…)

FUTURE PROOF

- Leverage standard NVMe-oF protocols
- Compliant with other NVMe-oF appliances
- Ease of in-the-field update

Please take a moment
to rate this session.

Your feedback matters to us.