KALRAY

White Paper

SMART STORAGE ADAPTER SOLUTIONS

Upgrade Your Data Center to Composable Disaggregated Architecture

CONTENTS

1

INTRODUCTION

Composability is on the critical path to enterprise data center transformation. It enables on-premises enterprises to be more agile, scalable, and cost-effective via a software-defined model. However, to accelerate the move from hyperconverged infrastructure (HCI) to disaggregated composability, a new approach is needed.

KALRAY

Wanted: A New Approach to Deliver Composability

Data processing and analytics are driving digital transformation and edge computing. As enterprises strive to turn ever-increasing volumes of data into actionable insights, they need to accelerate the deployment of IT resources while simultaneously maximizing efficiency and optimizing costs.

Whether delivering scalable IT resources at the network edge or deploying agile infrastructure in the data center, the hyperconverged, disaggregated, and composable infrastructure spectrum offers a seamless path to success.

Or does it?

Hyperconverged Infrastructure (HCI)

Tightly coupled compute, storage and network components allows cost-effective resources to be easily deployed and scaled as needed.

- Reduces complexity
- Eliminates hardware sprawl
- Increases agility and scalability

Disaggregated Infrastructure

Disaggregation allows customers to purchase, deploy, and scale storage and compute independently of one another for increased cost savings.

- Supports diverse workloads
- Enables cost-effective storage
- Scaling for rapid data growth

Composable Disaggregated Infrastructure (CDI)

Composability allows disparate IT resources to be collected into pools and programmatically composed into virtual servers on the fly.

- Mix and match hardware
- No reconfiguration required
- Optimizes on-demand resources
- Match QoS requirements to workload

HCI 2.0

Since standard hyperconverged systems feature an aggregated infrastructure, users can't linearly scale individual resources. If a specific workload only needs additional storage, the data center must purchase both compute and storage, wasting money that could be allocated elsewhere.

Storage vendors are addressing this problem in HCI 2.0, disaggregating compute and storage so users can acquire only the resources they need, when they need them. While this may appear to bridge the gap between hyperconverged and disaggregated infrastructure - and pave the way towards composability - it introduces a set of new challenges.

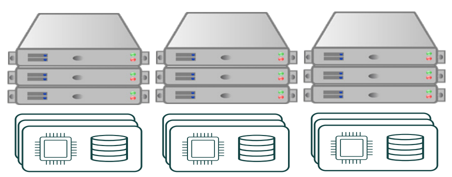
Storage aggregation is complicated and expensive. With extra software layers required to run on top of each HCI node, additional loads are placed on the x86 cluster CPUs and interconnect. Even though some hardware vendors are attempting to offset the performance impact by moving more intelligence into their network interface cards (NICs), the lack of programmability means that it's like using a Band-Aid instead of a plaster cast for a significant break.

1.1 Enabling Composability Through Disaggregation

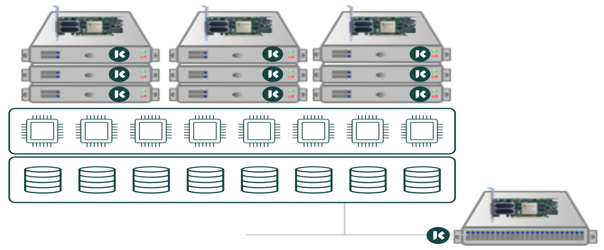
Composability is an architectural design principle that abstracts compute, storage, and networking resources from their physical locations and allows infrastructure to be managed by software. A highly composable system provides discrete components that can be selected and assembled in various combinations to provide specific QoS (Quality of Service) to match user requirements using software-defined intelligence.

Eliminating the need to configure and deploy environments meeting the needs of each workload, composable disaggregated infrastructure provides a pool of fluid resources that can be allocated on the fly to meet the unique demands of any application. It allows IT departments to create an agile, cost-effective data center with optimal application performance, eliminating overprovisioning and underutilization. As a result, on-premises infrastructure can be provisioned as quickly and painlessly as public cloud resources.

Standard HCI



Disagregated HCI enabling Composability



As shown in the graphic above, standard HCI architecture comprises tightly coupled components designed to deliver maximum performance. However, as the following section highlights, siloed storage often results in fabric bottlenecking, leading to diminished efficiency due to resource overprovisioning to maintain performance.

On the other hand, disaggregation allows IT architects to mix and match server, storage, and network components from different vendors - including storage adapters specifically designed to maximize performance and reduce overhead. Resources can then be configured to create pools from which infrastructure can be dynamically composed according to the demands of each workload.

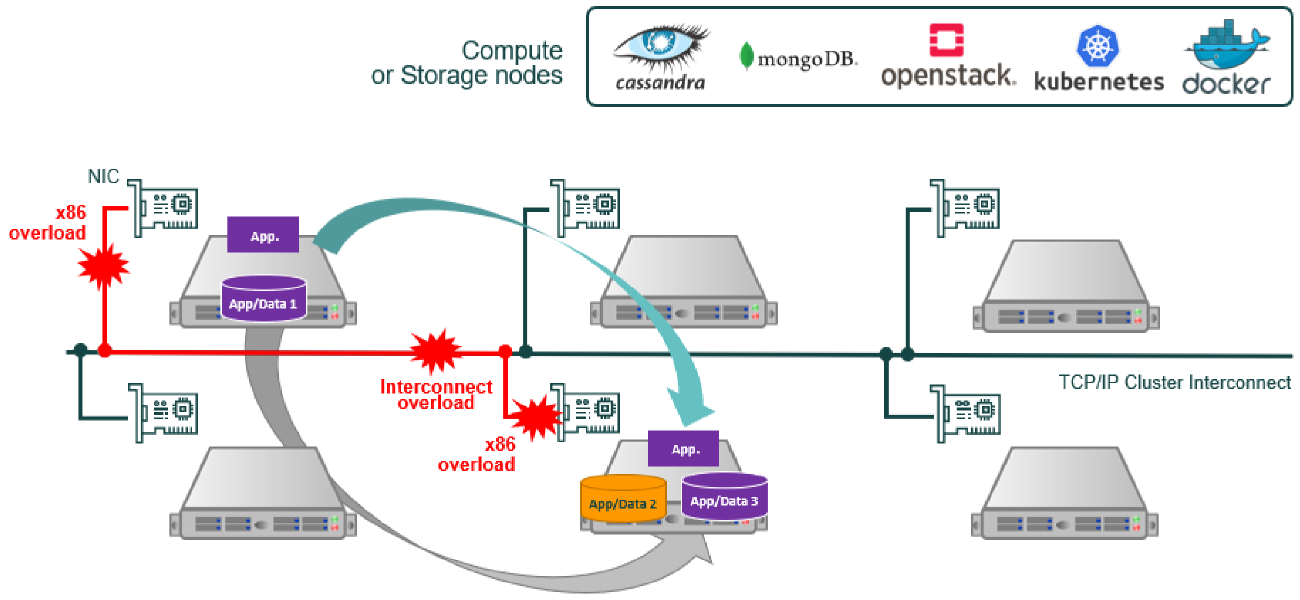
1.2 Siloed Storage, Fabric Bottlenecking, and Overprovisioning

With current topologies, HCI data silos impose disaggregation by using a software layer to manage the storage stack. Virtual machines, containers, and services are spread across nodes using TCP/IP Cluster Interconnect. Data is created locally within the same node in response to the application's data needs, as seen in the graphic below.

However, VMs and containers often need to be relocated to alternate nodes to meet performance and availability SLAs. During migration, the application is moved to a secondary node while still accessing the original node's primary data.

This not only overloads the TCP/IP interconnect but also the x86 storage stack on both nodes. In many cases, when new data is created, it is located on the secondary node, resulting in data fragmentation and fabric bottlenecking. The additional overhead triggers the need for the original dataset to be moved to the new node, embedding siloed storage and resulting in capacity wastage due to overprovisioning.

Current HCI Topology: Overloaded CPUs and Interconnects



1. The application creates data locally
2. During migration, the application is moved but not the data
3. New data is created on the new node, resulting in fragmentation
4. A reorganization is required to move the original data

Adding NVMeoF/TCP or NVMeoF/RoCE to the TCP/IP Cluster Interconnect eliminates the interconnect overload by offering faster and more efficient connectivity between storage and servers - as well as reduced application host server CPU utilization. But it doesn't eliminate the problem entirely. Since the storage services - such as logical volume management, data protection, and virtualization - still run on the x86 compute nodes, they will still experience a degree of overload, resulting in storage stack inefficiencies, poor application utilization, and decreased data center ROI.

What we need is a better, smarter way to deliver the full potential of composability.

2

SMART STORAGE ADAPTERS

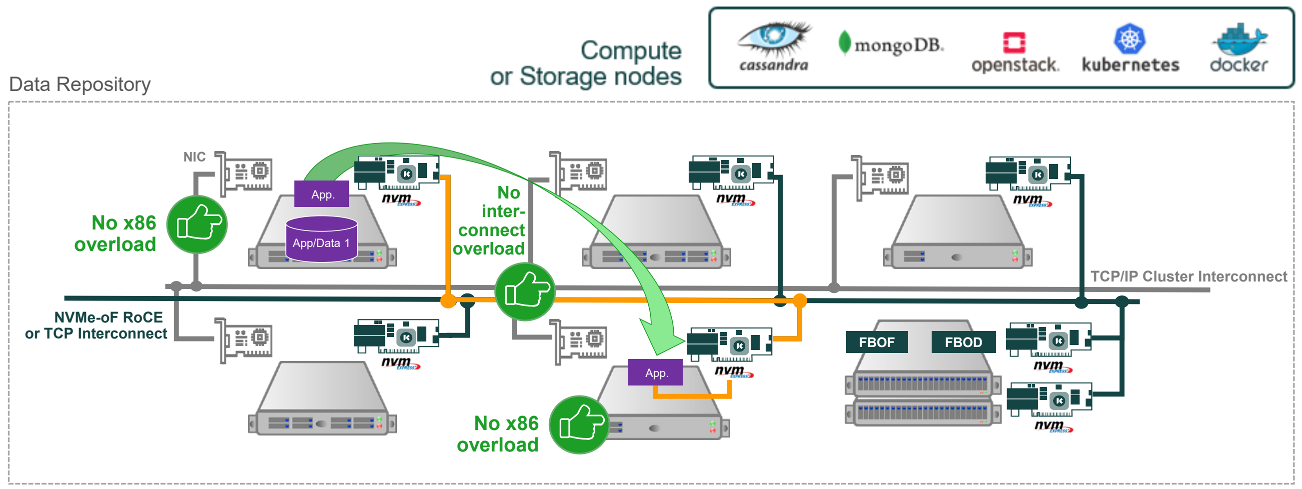
Enabling New Composable Topologies

Enabling Highly-Efficient Composable Disaggregated Infrastructure

Certified for NVM Express™ over Fabrics (NVMe-oF) with TCP (NVMe™/TCP) protocol (and soon for NVM Express over Fabrics (NVMe-oF) with RDMA (NVMe/RoCE) protocol), Kalray's K200-LP™ Coolidge™- based smart storage adapter solution is a fullyprogrammable, smart storage adapter delivering high performance while offloading demanding data services. Delivering up to 1.5 MIOPS (Millions of I/O Operations Per Second), the K200-LP™ offers new possibilities in terms of storage performance for data-intensive applications such as AI, AR, data analytics, and IoT.

Addressing the shortcomings of NVMe-oF and TCP/IP Cluster Interconnects, the K200-LP™ offloads storage services by taking control of NVMe devices and exposing NVMe volumes to the x86 processors via a standard NVMe driver. It implements a standard function name NVMe Emulation (see graphic below). As a result, the x86 processors see all drives as local, eliminating both x86 and interconnect bottlenecks and the need for data migration.

HCI with Kalray: K200-LP™ Smart Adapters Enable New HCI Topology



The x86 only sees local NVMe drives, while the K200-LP™ looks after the NVMe to NVMe-oF remote access (network + storage stack)

1. The application creates data
2. During migration, the application is moved but not the data
3. New data is created on the new node, but since data is global, no fragmentation occurs
4. A reorganization is required to move the original data

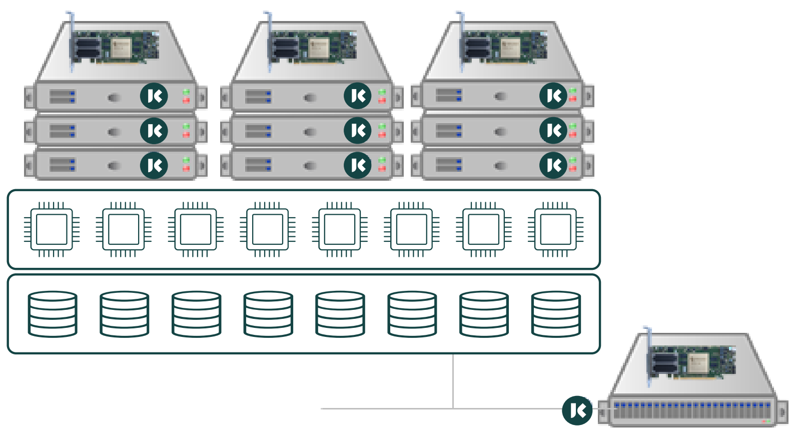
Competitive solutions are limited in delivering the performance needed for demanding algorithms such as data protection and footprint reduction. In contrast, the K200-LP™ leverages the capabilities of the 3rd generation MPPA® (Massively Parallel Processor Array) Coolidge™ Intelligent Processors to run the necessary protocols in real-time with zero degradation in server bandwidth. A standard PCIe card supporting NVMe™/TCP and NVMe™/RoCE via the AccessCore® Storage SDK, the K200-LP™ smart storage adapter is highly configurable to meet the composability needs of a broad range of workloads and use cases.

2.1 Transform your HCI Infrastructure into Composable Disaggregated Infrastructure (CDI) by Simply Adding Kalray Smart Adapter Cards

By its very definition, transformation involves significant change. However, what we propose is a revolt against traditional transformation with a seamless approach to composable disaggregated infrastructure that's fast and cost-effective, leveraging existing infrastructure investments and application stacks.

Simply plugging the Kalray K200-LP™ Smart Storage Adapter Solution card into your existing HCI nodes (illustrated on the right with the Kalray logo) creates a highly efficient and cost effective storage fabric. It transforms your infrastructure from one that's siloed to one that's open, disaggregated, and fully composable.

Composable Disaggregated Infrastrucuture (CDI) with Kalray Smart Cards

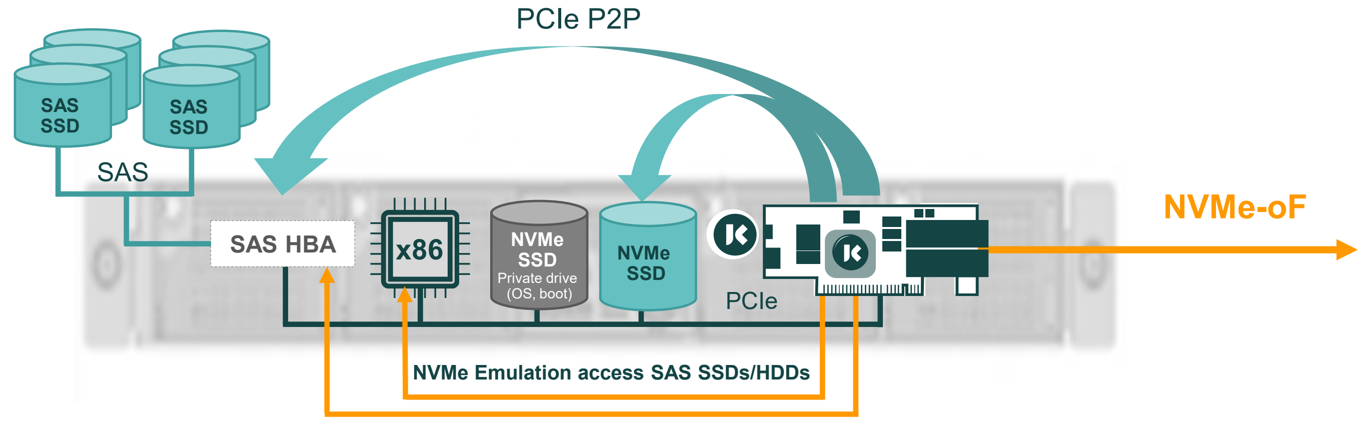


The term HCI generally refers to the physical hardware (server, storage, and network) and the software layer managed by the hypervisor. However, Kalray's architecture is not limited to those virtualized solutions, encompassing bare metal, containers, and microservices.

The Kalray architecture acts as an NVMe storage layer underlying all software layers, delivering increased flexibility for the vast majority of applications. As a standards-based reference platform it offers full interoperability with:

- Kubernetes Container Storage Interface (CSI) and Redfish/Swordfish management interfaces
- Existing NVMe-oF components via Kalray's certified NVMe-oF NVMe-TCP compatibility
- Operating system interfaces using NVMe Emulation (see below)

Composable Disaggregated Architecture with Kalray Smart Storage Adapters - x86 Node System Architecture

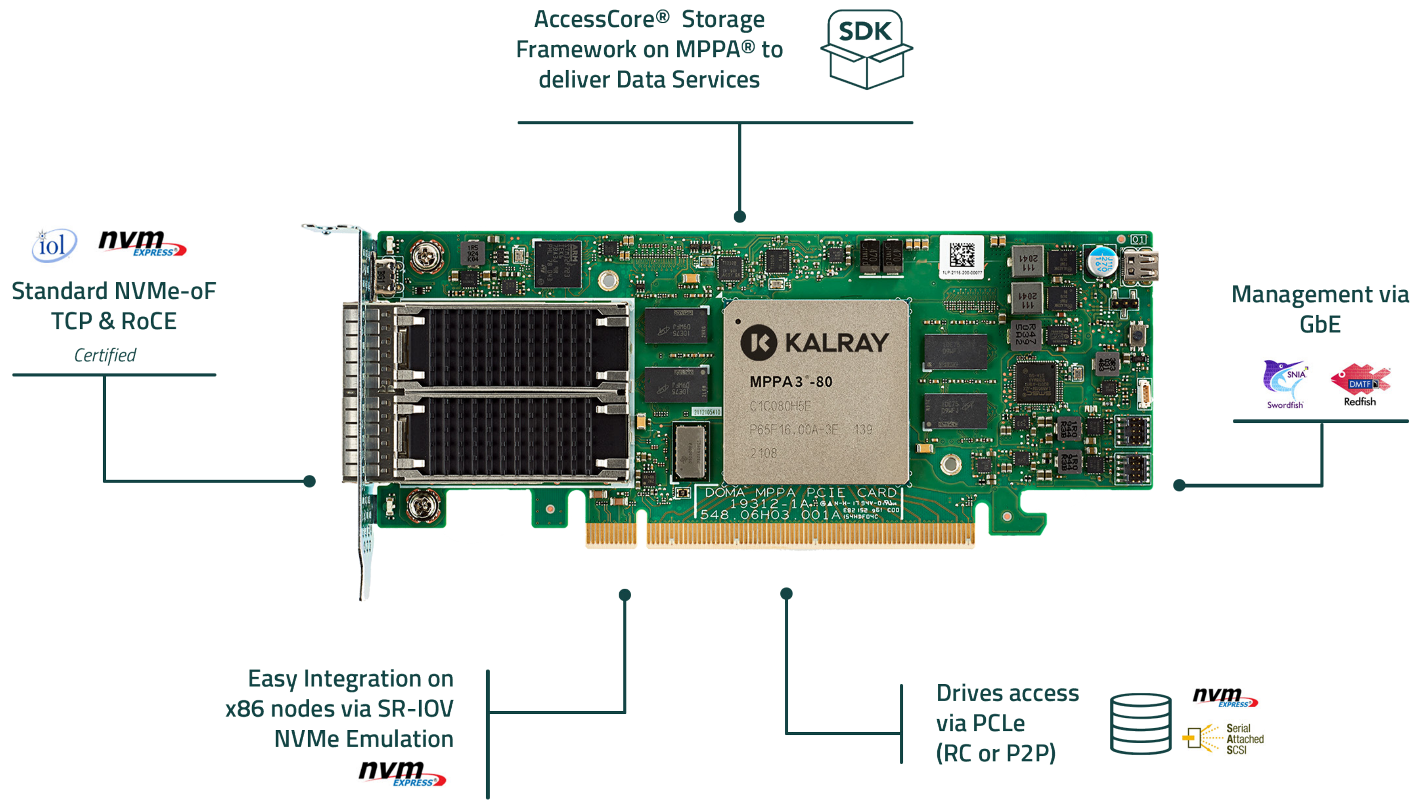


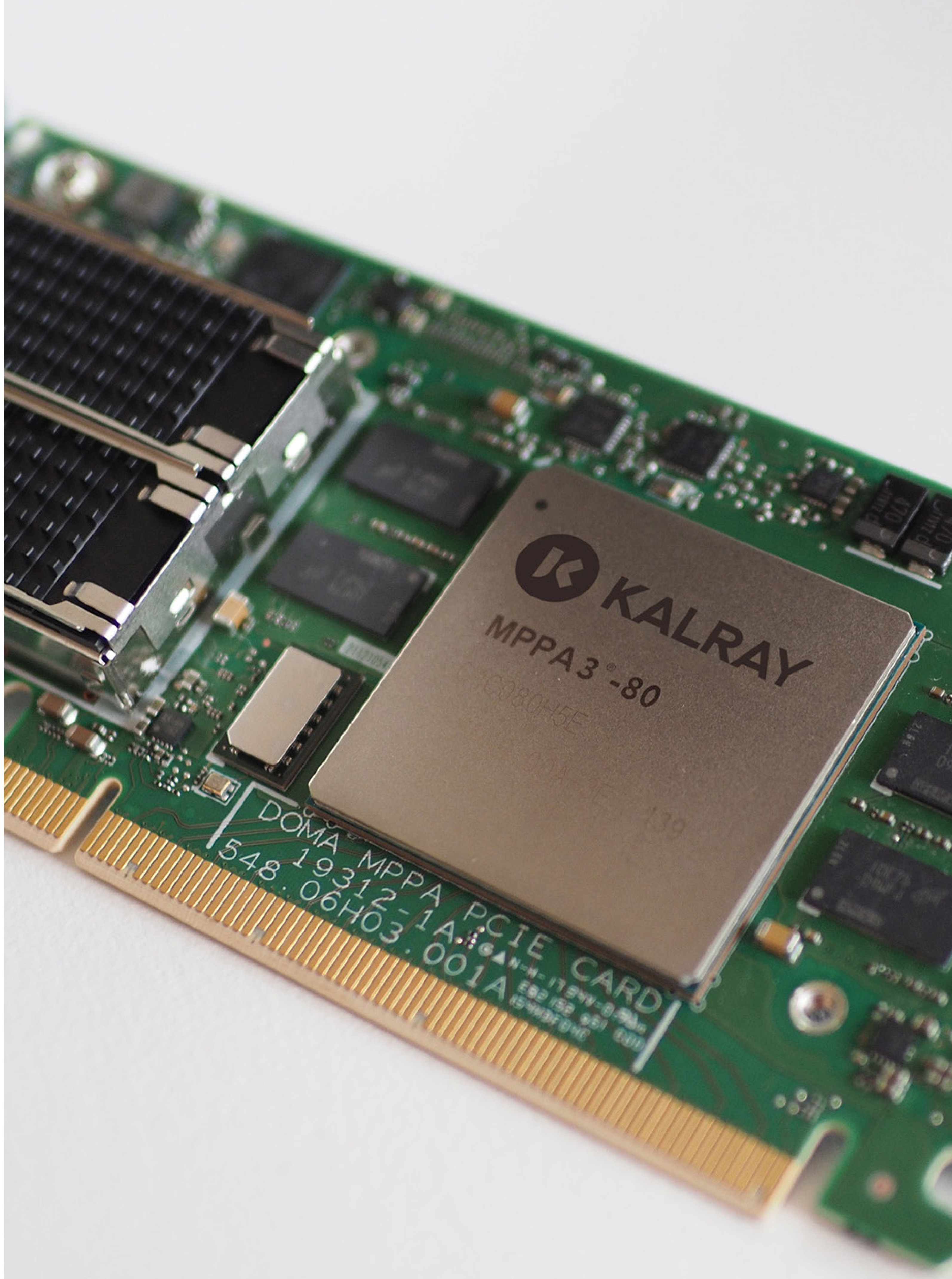
Kalray's CDI architecture ensures the delivery of enhanced data center-ready storage services, including:

- Global data repository
- Scale-out support
- Storage services offload from x86 (incl. RAID)
- Increased CPU utilization of 20% or more
- Seamless integration with Kubernetes, VMWare, and other open-sourceplatforms
- Compatibility with any standard HCI hardware
- Support for heterogeneous servers and nodes

Fast, Easy Integration

Build Next-Gen Storage Today with Kalray Smart Storage Adapter





3

THE MPPA® DPU PROCESSOR ARCHITECTURE

Leveraging Kalray's Massively Parallel Processor Array

The MPPA® DPU Processor Architecture

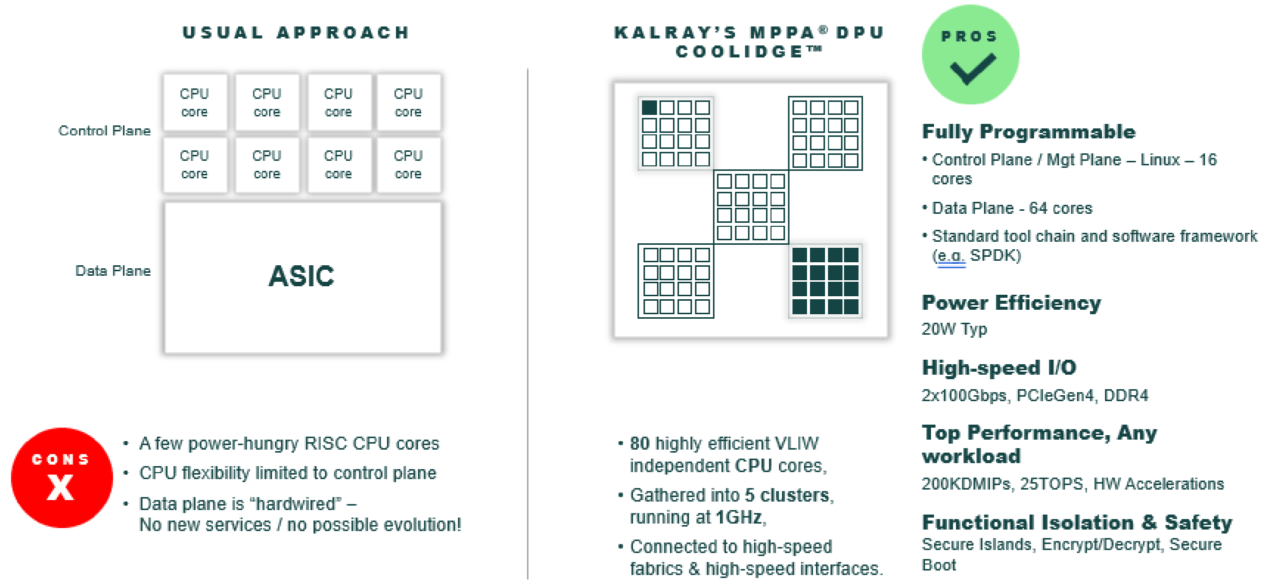
3.1 Massively Parallel Architecture

K200-LP™ Smart Storage Adapters leverage the power and programmability of Kalray's manycore architecture. Instead of a hard-wired data plane and a few power-hungry RISC cores, Kalray's 3rd generation MPPA® 3 Coolidge™ delivers 80 highly efficient VLIW independent CPU-cores running at 1.2GHz, and connected to high-speed fabrics and interfaces.

Configured as five clusters - one cluster for the control plane and four for the data plane - the Coolidge processor delivers a performance of up to 25 TOPzS for any workload, along with full programmability for composable infrastructures. Each cluster (comprised of 16 cores and 16 acceleration co-processors) provides complete functional isolation and safety for the communication fabric.

3.2 Delivering Unparalleled Performance

Designed as a standard PCIe card supporting both NVMe™/TCP and NVMe™/RoCE, the K200-LP™ takes advantage of the Coolidge processor's support for standard industry interfaces - including PCIe Gen4 x16 and 2 x 100G Ethernet to deliver best-in-class performance of up to 1.5 MIOPS (Millions of I/O Operations Per Second). When coupled with the latest PCIe Gen4 x86 processor, MPPA® can also be used as an accelerator to deliver full-duplex bandwidth of up to 200 Gbit/s.



4

FULLY PROGRAMMABLE

Kalray AccessCore® Storage (ACS) Framework

KALRAY

Accelerating Results with a Fully Flexible Software Environment

Using standardized programming tools, libraries, and languages, Kalray's AccessCore® Storage (ACS) Framework makes it easy for any software programmer to access and program the MPPA® DPU processor for optimal results.

4.1 Enhanced Programmability

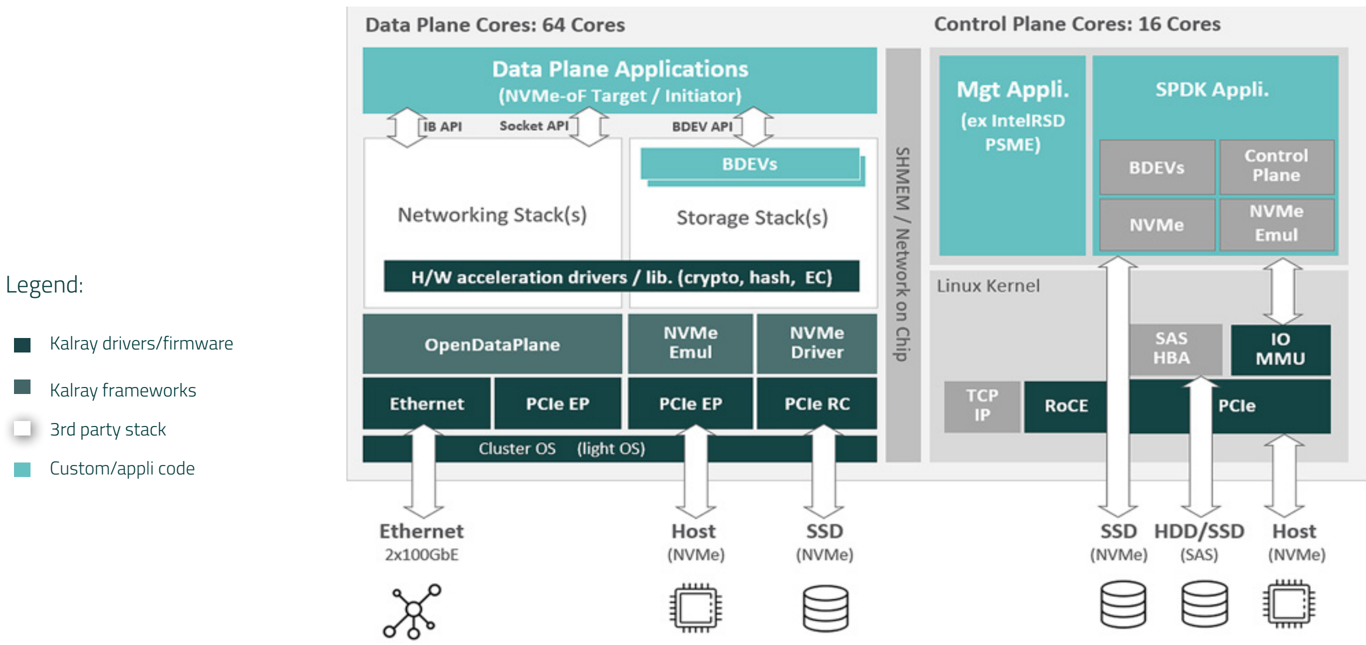
Kalray offers full programmability of control, data, and management planes using standard languages (C/C++), open APIs, and tools. Incorporating a set of highly optimized libraries, ACS provides a full-featured, modular software framework to program an MPPA® DPU processor. It uses an optimized Storage Performance Development Kit (SPDK) for writing scalable, high-performance, usermode storage applications for both data and control planes.

4.2 Architectural Highlights

Standardization: AccessCore® Storage uses standard hardware interfaces for NVMe Emulation and Linux and data plane APIs and libraries for rapid programmability and investment protection.

Programmability: ACS provides full programmability of the 16-core (one cluster) control and management planes using Linux, and the 16-64 core (1-4 clusters) data plane using ClusterOS (a light POSIX OS).

Efficiency: Eliminating the need for x86 pre- and post-processing, AccessCore® Storage provides true inline processing and full programmability of the data plane—from network functions to the NVMe stack on light OS cores.



K200-LP™ MAIN FEATURES

Smart Storage Adapter Solution

Form Factors

- HHHL (Low Profile) - K200-LP™

Manycore Architecture

- 80 VLIW cores @ 1.2 GHz
- 5 Clusters x16 cores

High-Speed Ethernet

- 2x100GbE / 8x25 Gbe

Certified NVMe-oF Stack

- NVMe-oF 1.0 to 1.1
- RoCE v1/v2, TCP

Advanced SSD interface

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

Two 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)

Performance (per card)

- Random R/W RoCE: 1.5 MIOPS
- Random R/W TCP: 1.5 MIOPS
- Sequential R/W (RoCE & TCP): 12 GB/s
- Latency (RoCE/TCP): 20/50 usec

SDK

AccessCore® Storage 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

CONCLUSION

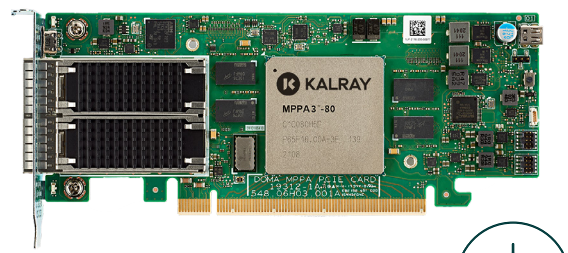
Kalray's K200-LP™ Smart Storage Adapter delivers on the promise of hyperconverged, disaggregated, and composable infrastructure. With fully programmable control and data planes, NVMe emulation, and MPPA® DPU architecture.

Deploying Kalray's K200-LP™ Smart Storage Adapter with flash memory-based SSDs and hyper-fast communication protocols - such as NVMe™/TCP or NVMe™/RoCE - provides the breakthrough solution you need for massive scalability, increased performance, and optimized costs with in your data center.

NVMe-oF

Higher Performance
Offload x86 from heavy storage stacks and leverage the performance of the K200-LP™ Smart Storage Adapter to

Investment Protection
Leverage standard NVMe-oF protocols for simplifying onsite upgrades and integration with other NVMe-oF appliances.



Increased Flexibility
Take advantage of a fully programmable data plane with additional storage services based on the SPDK framework for asynchronous, polled-mode, shared-nothing server applications.

Lower Costs
Switch to a true disaggregated composable infrastructure with commodity components for increased efficiency and reduced costs.

THE AUTHOR

Rémy Gauguey

Senior Software Architect,
Datacenter Business Line, Kalray

A Senior Software Architect in Kalray's Data Center Business Unit, Rémy Gauguey has more than 25 years of experience in the high-tech industry, specializing in SoCs, RTOS, and high-performance packet processing. Leveraging the Kalray MPPA® manycore technology, Rémy develops advanced architectures for composable disaggregated infrastructure.

Rémy previously developed his expertise at Conexant, Mindspeed Technologies, and the French Atomic Energy Commission (Commissariat à l'énergie atomique et aux énergies alternatives). He holds several patents in the fields of software architecture and packet processing.

ABOUT KALRAY

Kalray (Euronext Growth Paris: ALKAL) is a leading provider of hardware and software technologies and solutions for high-performance, data centric computing markets, from cloud to edge.

Kalray provides a full range of products to enable smarter, more efficient, and energy-wise data-intensive applications and infrastructures. Its offers include its unique patented DPU (Data Processing Unit) processors and acceleration cards as well as its leading-edge software-defined storage and data management offers. Separated or in combination, Kalray's high-performance solutions allow its customers to improve the efficiency of data centers or design the best solutions in fast-growing sectors such as AI, Media & Entertainment, Life Sciences, Scientific Research, Edge Computing, Automotive and others.

Founded in 2008 as a spin-off of the well-known French CEA research lab, with corporate and financial investors such as Alliance Venture (Renault-Nissan-Mitsubishi), NXP Semiconductors or Bpifrance, Kalray is dedicated through technology, expertise, and passion to offer more: more for a smart world, more for the planet, more for customers and developers.

www.kalrayinc.com
contact@kalrayinc.com

KALRAY

THE POWER OF MORE

Intelligent Data Processing,
From Cloud to Edge

Kalrayinc.com