KALRAY

STORAGE DEVELOPER CONFERENCE
SDC 21
BY Developers FOR Developers

Unlocking the Potential of NVMe-oF & Software Defined Storage Thanks to Programmable DPUs

Sébastien Le Duc
Software Engineering Director
Kalray

www.kalrayinc.com

ABSTRACT

ABSTRACT

During the last two decades, the data center world has been moving to a “Software Defined Everything” paradigm. This has been taken care of mostly by hypervisors running on the x86 up to recently.

In parallel, a new communication protocol to interface with SSDs has been specified from the ground-up, allowing to fully exploit the levels of parallelism and performances of all-flash storage: NVMe, and NVMe-oF. NVMe-oF promises to enable the performances of direct attached all-flash storage with the flexibility an TCO savings of shared storage. To fully unlock the benefits of NVMe-oF while keeping the software defined paradigm, we believe a new kind of processor is needed: the Data Processing Unit, or DPU.

THE PRESENTER

Sébastien Le Duc
Software Engineering Director

Sébastien started his professional career in 1998 at STMicroelectronics where he worked on tools and architecture for a proprietary VLIW DSP.

He then joined ST-Ericsson in 2006 where he spent 8 years managing Multimedia Software development teams. He continued his career back at STMicroelectronics as Lead Software Architect on set-top box products.

Together with his technical background that ranges from low-level software development to middleware integration, his team management and product development experience, Sébastien brings outstanding Software Engineering leadership to Kalray.

Agenda

1. **Kalray in a Nutshell**
2. NVMe-oF Background
3. What is a DPU?
4. Use Cases
   - **#1**: NVMe-oF All-Flash-Array for SSD Disaggregation
   - **#2**: Storage Adapter for SSD Disaggregation
   - **#3**: Storage Adapter for Hyper-Converged Infrastructure
5. Conclusion

KALRAY IN A NUTSHELL

Kalray offers a new type of processor targeting the booming market of intelligent systems.

A Global Presence

- France (Grenoble, Sophia-Antipolis)
- USA (Los Altos, CA)
- Japan (Yokohama)
- Canada (Partner)
- China (Partner)
- South Korea (Partner)

Leader in Manycore Technology

3rd generation of MPPA® processor

~€100m R&D investment

30 Patent families

Industrial investors

NXP

RENAULT NISSAN MITSUBISHI

SAFRAN

MBDA

EURONEXT

- Public Company (ALKAL)
- Support from European Govts
- Working with 500 fortune companies

*Financial investors: CEA Investissement, Bpifrance, ACE, INOCAP Gestion, Pengpai

INTELLIGENT SYSTEMS / EDGE COMPUTING

At the Heart of Next Decade Industry

Next Gen. Embedded Systems

Next Gen. Data Center

Compute and AI Intensive Critical Systems

MPPA® Processors

PCIe Cards & Modules

Acceleration Solutions for Storage, Networking and Compute

Agenda

1. Kalray in a Nutshell
2. **NVMe-oF Background**
3. What is a DPU?
4. Use Cases
   - **#1**: NVMe-oF All-Flash-Array for SSD Disaggregation
   - **#2**: Storage Adapter for SSD Disaggregation
   - **#3**: Storage Adapter for Hyper-Converged Infrastructure
5. Conclusion

NVMe-oF BACKGROUND

2011

First NVM Express specification released

- Open logical device interface specification for accessing non-volatile storage attached via the PCIe bus
- Designed to capitalize on the low latency and internal parallelism of solid-state storage devices

2016

First NVMe over Fabrics specification released

- Extends the NVMe command set using a transport protocol over a network
- NVMe/FC, NVMe/RoCE, NVMe/iWARP, NVMe/IB
- Enables the benefits of NVMe technology to be realized at a much larger scale

2018

NVMe/TCP specification released

- Allows NVMe usage using existing network infrastructure

NVMe-oF: WHY NOW?

Storage market faces a dramatic change with growth in SSD adoption

- SSDs are improving at a much faster rate than CPUs … leading to inefficient processing and wasted storage

The two “standard” approaches are highly problematic

- Software Defined approaches based on virtualization techniques cannot sustain the required speed and performance or with massive impact on the overall system performance and cost
- Legacy x86-based Flash Array solutions can not sustain the required speed and performance at relevant price points

OMDIA

Dennis Hahn
Senior Analyst, Cloud & Data Center at OMDIA*

> “NVMe-oF JBOF is just starting to experience an uptake in on-premises enterprise DC segments for its ultra-high performance. **It is ramping aggressively for use by hyper converged infrastructure and in data intensive applications.”**

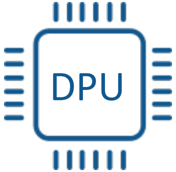
KALRAY

Agenda

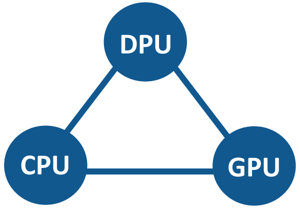
1. Kalray in a Nutshell
2. NVMe-oF Background
3. **What is a DPU?**
4. Use Cases
   - **#1**: NVMe-oF All-Flash-Array for SSD Disaggregation
   - **#2**: Storage Adapter for SSD Disaggregation
   - **#3**: Storage Adapter for Hyper-Converged Infrastructure
5. Conclusion

KALRAY

WHAT IS A DPU?



A new class of programmable processor specialized in running datacenter infrastructure services



The 3rd socket in data centers alongside the CPU and GPU

Networking
NFV, vSwitch, NAT, …

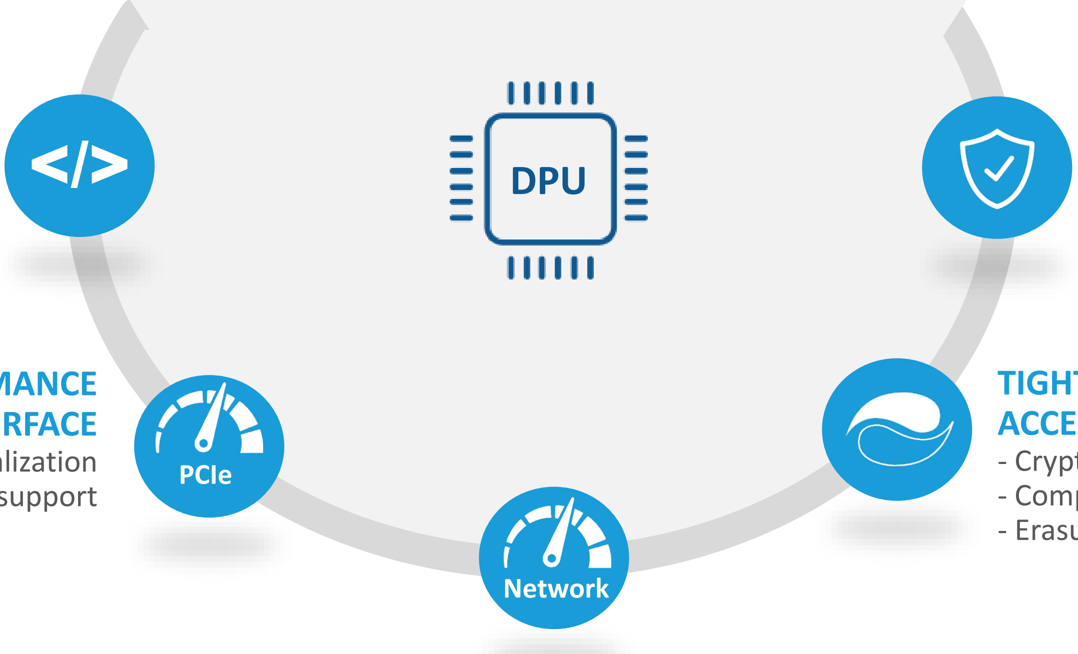
Storage
NVMe-oF, compression, deduplication, encryption, …

Security
Firewall, encryption, IPsec, …

Accelerate Software-Defined Datacenter Infrastructure Services

WHAT IS A DPU?

Key Features



FULLY PROGRAMMABLE
Management plane
control plane
and data plane

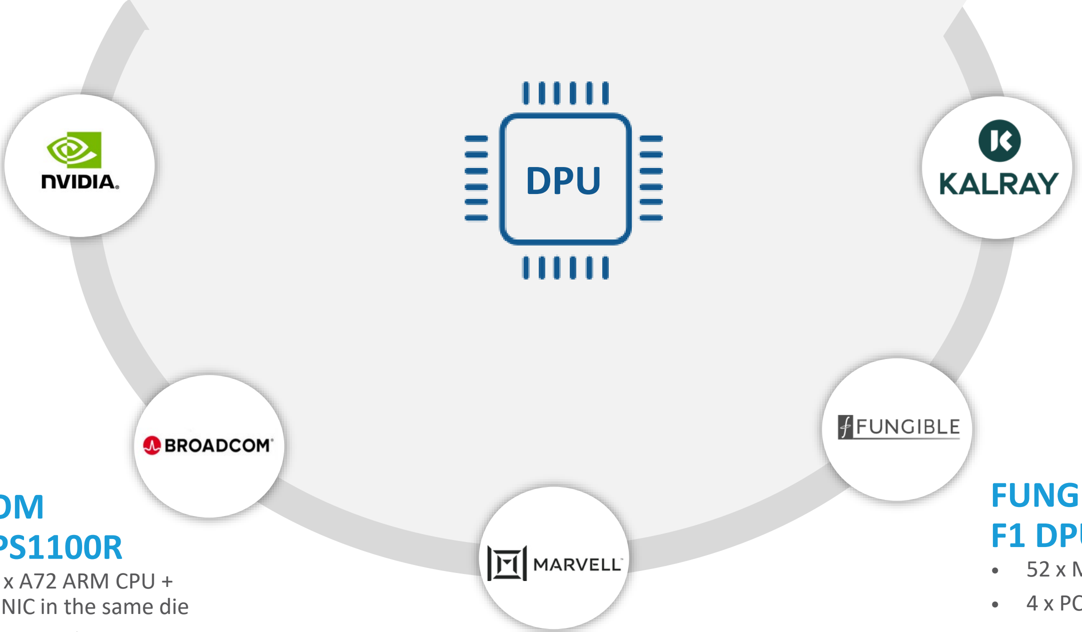
SECURITY
Root of trust, secure boot, secure firmware upgrades

HIGH PERFORMANCE PCIE INTERFACE
SR-IOV for virtualization support

TIGHTLY COUPLED INLINE ACCELERATORS
- Crypto accelerators (IPsec, TLS)
- Compression (storage)
- Erasure Coding

HIGH PERFORMANCE NETWORK INTERFACES
- Packet parsing / matching / dispatching
- RDMA support
- TCP acceleration (RSS, LRO, checksums, …)

CURRENT DPU LANDSCAPE



NVIDIA BLUEFIELD 2

- Essentially 8 x A72 ARM CPU + ConnectX6 DX NIC in the same die
- PCIe Gen4 x 16, 2 x 100GbE

KALRAY MPPA® Coolidge™

See next slide…

BROADCOM Stingray PS1100R

- Essentially 8 x A72 ARM CPU + NetExtreme NIC in the same die
- PCIe Gen3, 1 x 100GbE

FUNGIBLE F1 DPU

- 52 x MIPS64 CPUs
- 4 x PCIe GEN4 x 16, 8 x 100GbE

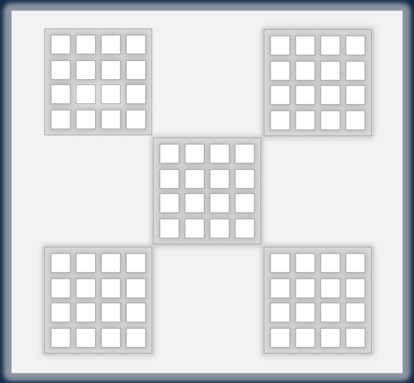
MARVELL LIQUIDIO III

- 24 x ARM CPUs
- PCIe GEN4 x 16, 2 x 100GbE
- More targeted to SDN than SDS

KALRAY MPPA®3 Coolidge™: AN ADVANCED DPU

Kalray's MPPA®3 Coolidge™

80 highly efficient VLIW independent **CPU** cores, gathered into **5 clusters**, running at **1GHz**



Power efficiency

20W Typ

High Speed I/O

2x100Gbps Eth, PCIe Gen4,DDR4

Tightly Coupled Accelerators

10 Crypto accelerators, Erasure coding acceleration, advanced programmable DMAs

Security

Root of trust, secure boot, secure vault

</> Fully programmable

Control & Mgmt Plane : 16 cores SMP CPU running Linux
Data Plane: 64 cores running SPDK

Agenda

1. Kalray in a Nutshell
2. NVMe-oF Background
3. What is a DPU?
4. **Use Cases**
   - **#1: NVMe-oF All-Flash-Array for SSD Disaggregation**
   - **#2**: Storage Adapter for SSD Disaggregation
   - **#3**: Storage Adapter for Hyper-Converged Infrastructure
5. Conclusion

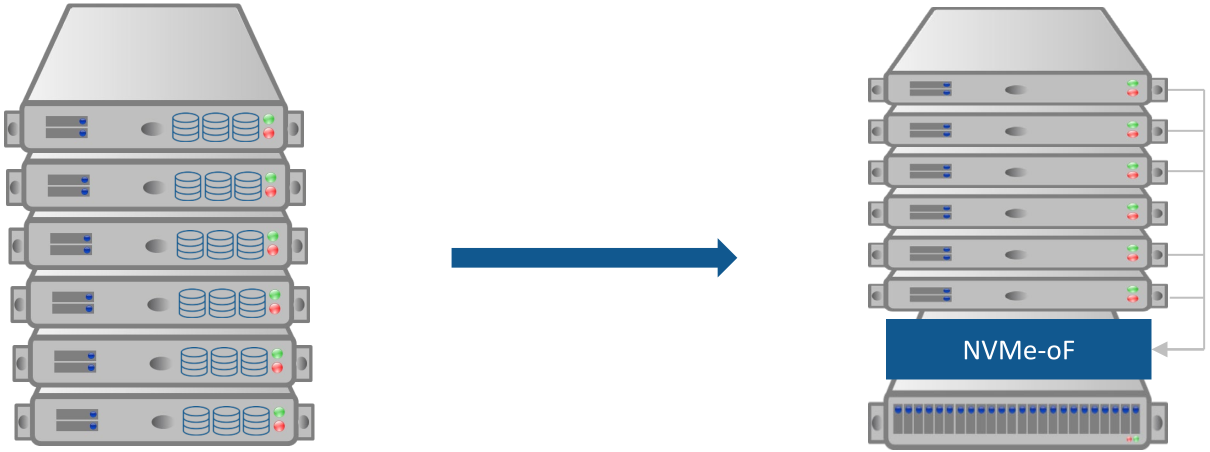
KALRAY

USE CASE #1

NVMe-oF All Flash Array for SSD Disaggregation

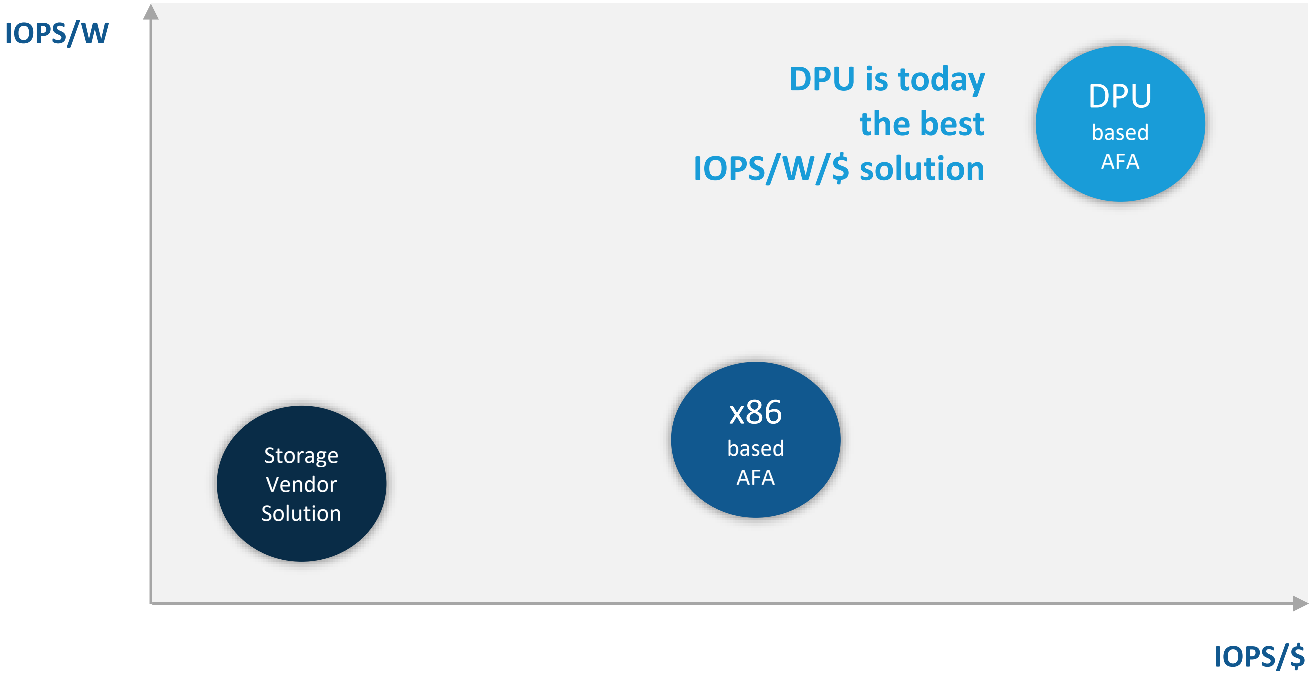
Direct-Attached SSDs

Disaggregated SSDs



- Servers can be composed on-the-fly with just the storage requested
- Workloads can move and re-attach to their volumes
- Scale storage and compute independently

USE CASE #1
NVMe-oF All Flash Array Solutions



DPU is today the best IOPS/W/$ solution

USE CASE #1
Data Services Enabled by DPU Usage

Basic Services

- TCP termination
- Logical Volumes
- Thin Provisioning
- Snapshots

Data Reduction

- Compression
- De-duplication
- Zero-detection

Data Availability

- RAID
- Erasure Coding

Data Security

- Encryption at rest
- Encryption in motion

DPUs are purpose-built for running data services at line rate

Agenda

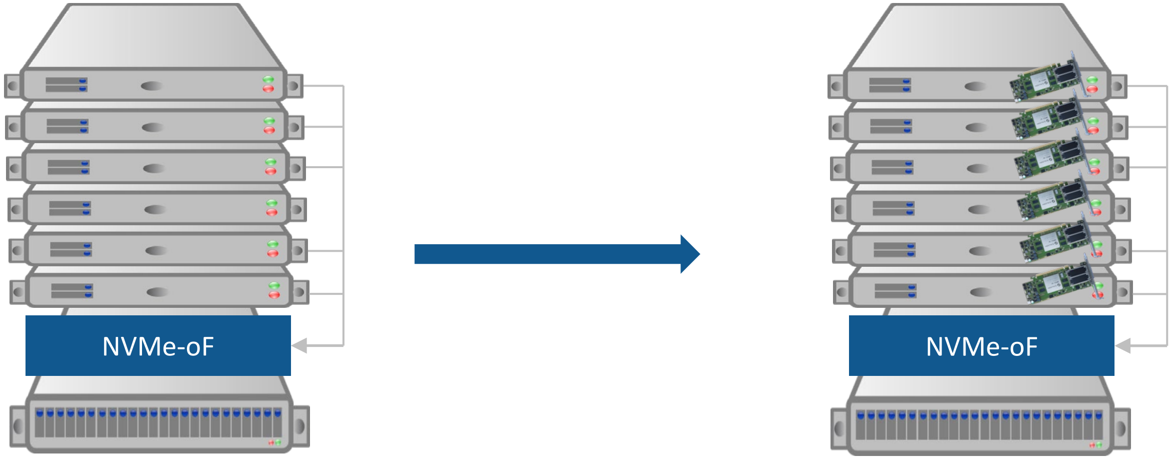
1. Kalray in a Nutshell
2. NVMe-oF Background
3. What is a DPU?
4. **Use Cases**
   - #1: NVMe-oF All-Flash-Array for SSD Disaggregation
   - #2: **Storage Adapter for SSD Disaggregation**
   - **#3**: Storage Adapter for Hyper-Converged Infrastructure
5. Conclusion

USE CASE #2

Storage Adapter for SSD Disaggregation

Disaggregated SSDs

Disaggregated SSDs with Storage Adapter



NVMe-oF

NVMe-oF

- Any OS on host servers supported
- Enables bare-metal storage virtualization
- Enables boot from NVMe-oF disks
- Data services on initiator side offload server CPU

USE CASE #2

Storage Adapter Enables Bare-Metal Storage Virtualization

1 The DPU presents itself as local NVMe disks to the host server

- No requirement on the operating system that will be deployed on the bare-metal server
- Enables boot from NVMe-oF storage
- Transparent for the bare-metal server which sees only NVMe drives

2 The DPU manages the NVMe to NVMe-oF translation

- The configuration of the translation is done by having orchestration software interact directly with the DPU
- No orchestration SW needed on the bare-metal server
- Workloads can move from one bare-metal server to another one

USE CASE #2

Data Services Enabled by Storage Adapter

Basic Services

- Bare-metal storage virtualization
- NVMe emulation
- NVMe initiator
- Networking stack and QoS
- Networking configuration fully SW defined

Data Reduction

- Data reduction on initiator side results in lower network bandwidth

Data Availability

- Distributed Erasure Coding enables better reliability while improving storage efficiency
- Mapping of volumes fully SW defined

Data Security

- Encryption in motion transparent to the host server
- Security Configuration fully SW defined

Agenda

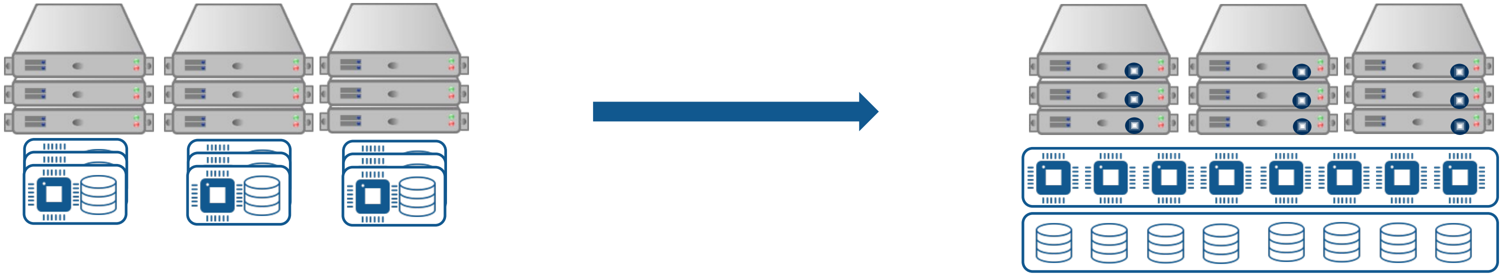
1. Kalray in a Nutshell
2. NVMe-oF Background
3. What is a DPU?
4. **Use Cases**
   - #1: NVMe-oF All-Flash-Array for SSD Disaggregation
   - #2: Storage Adapter for SSD Disaggregation
   - **#3: Storage Adapter for Hyper-Converged Infrastructure**
5. Conclusion

USE CASE #3

Storage Adapter for Hyper-Converged Infrastructure

Standard HCI

HCI with Storage Adapters



- Provide a global data store repository
- Enables bare-metal storage virtualization
- Offload data services from server CPU

USE CASE #3

Storage Adapter for Hyper-Converged Infrastructure

1 The DPU takes ownership of the local drives in HCI server

- Using PCIe peer-to-peer technology

2 The DPU presents itself as local NVMe disks to the host server

- No requirement on the operating system or hypervisor that will be deployed on the bare-metal server

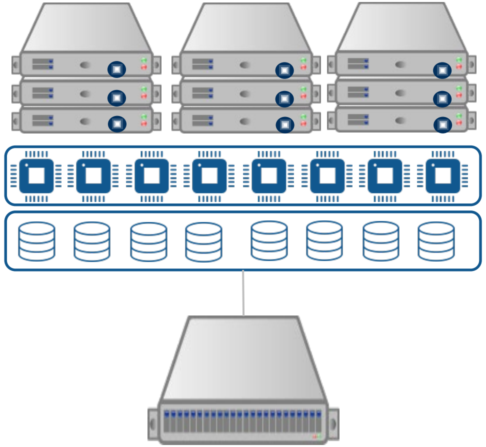
3 The host servers see a global data store

- The global data store is made of all the drives of each HCI server
- Each DPU acts as both a storage target and a storage initiator
- Storage is fully virtualized without needing a hypervisor. The virtualization is fully SW defined

USE CASE #3

Scale Storage and Compute Independently

HCI with Storage Adapters



Increase the global repository size by adding NVMe-oF storage nodes

USE CASE #3
Data Services Enabled by Storage Adapter

Basic Services

- Bare-metal storage virtualization
- NVMe emulation
- NVMe-oF target and initiator
- Networking stack and QoS
- Networking configuration fully SW defined

Data Reduction

- Data reduction on initiator side results in lower network bandwidth

Data Availability

- Distributed Erasure Coding enables better reliability while improving storage efficiency
- Mapping of volumes fully SW defined

Data Security

- Encryption in motion transparent to the host server
- Security Configuration fully SW defined

Agenda

1. Kalray in a Nutshell
2. NVMe-oF Background
3. What is a DPU?
4. Use Cases
   - #1: NVMe-oF All-Flash-Array for SSD Disaggregation
   - #2: Storage Adapter for SSD Disaggregation
   - #3: Storage Adapter for Hyper-Converged Infrastructure
5. **Conclusion**

CONCLUSION

- NVMe SSDs and NVMe-oF will disrupt the storage market: This is already happening!

- Software Defined Storage based on virtualization will become a bottleneck to leverage the full potential of NVMe-oF.

- We believe DPUs are the solution and will allow to unlock the full potential of NVMe-oF while keeping the flexibility of Software Defined Storage.

Building the Next Generation
of Storage Solution With Kalray DPU

KALRAY

STORAGE DEVELOPER CONFERENCE SDC 21 BY Developers FOR Developers

THANK YOU

www.kalrayinc.com

DISCLAIMER

Kalray makes no guarantee about the accuracy of the information contained in this document. It is intended for information purposes only, and shall not be incorporated into any contract. It is not a commitment to deliver any material, code or functionality, and should not be relied upon in making purchasing decisions. The development, release and timing of any features or functionality described for Kalray products remains at the sole discretion of Kalray.

Trademarks and logos used in this document are the properties of their respective owners.