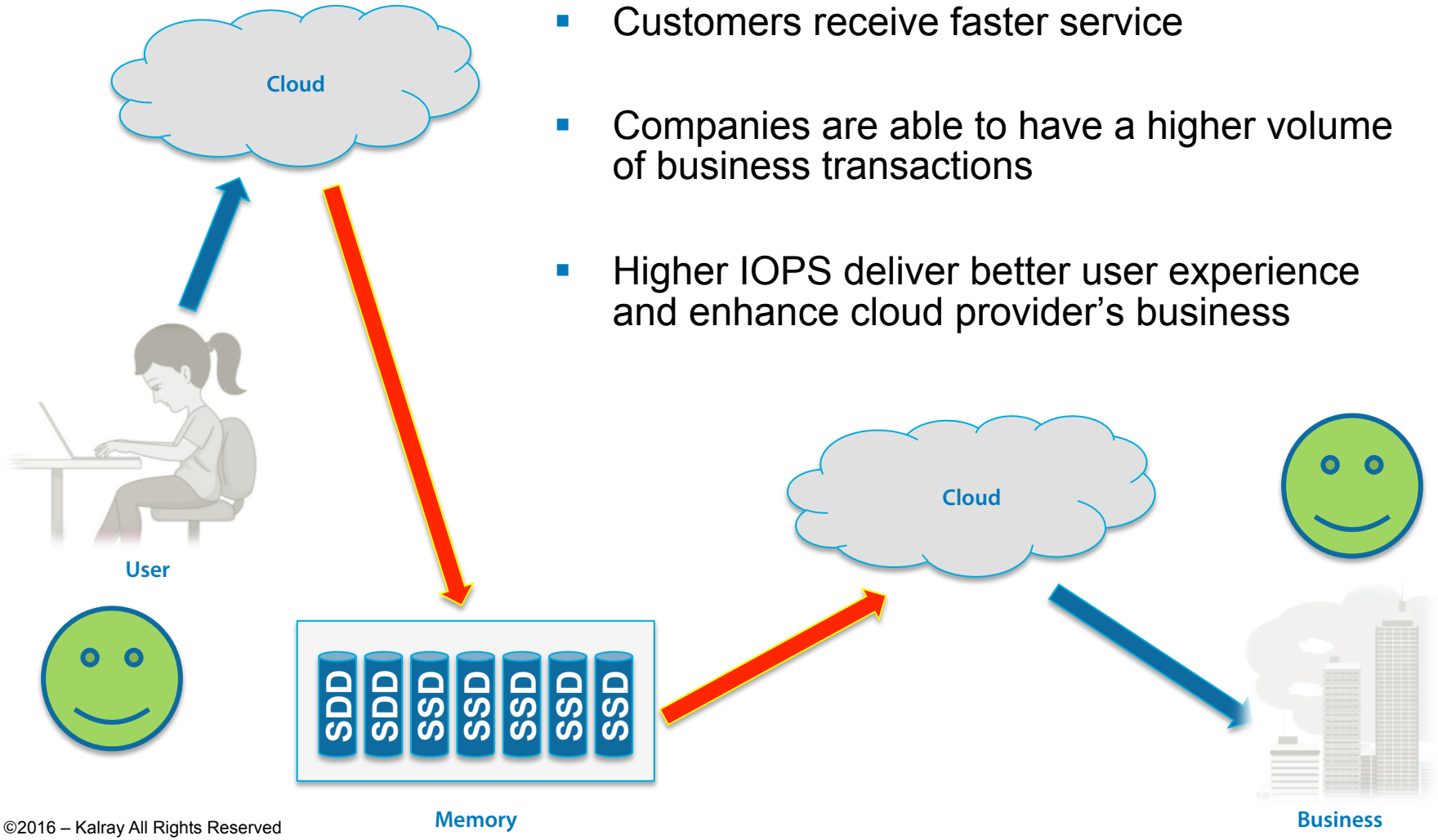
KALRAY

High Speed I/O processor for NVMe over Fabric (NVMeoF)

Patrice COUVERT
Datacenter Product Marketing
pcouvert@kalray.eu

SSD Technology: Faster Storage for Better Business

- Customers receive faster service
- Companies are able to have a higher volume of business transactions
- Higher IOPS deliver better user experience and enhance cloud provider's business



Faster SSD Technology: Positive Impact on Businesses

KAYAK

Caching airline quotes to speed up service

NETFLIX

Personalization for > 50 mill. customers

Domino's Pizza

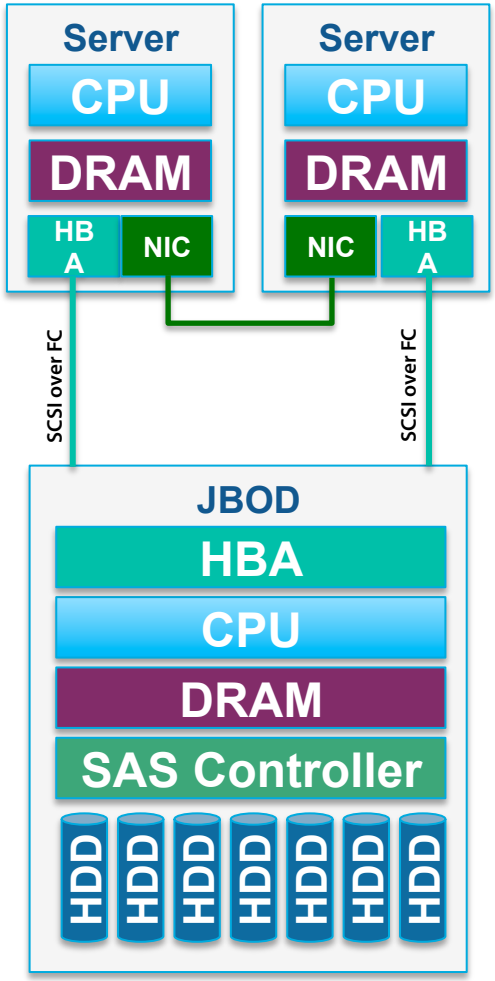
Increase booking during big sports events like the Superbowl

amadeus

3.7 million bookings per day

Demanding Applications that take benefit of High IOPs/ High Throughput SSDs

Phase 1 of SSD Revolution: Evolution of SANs



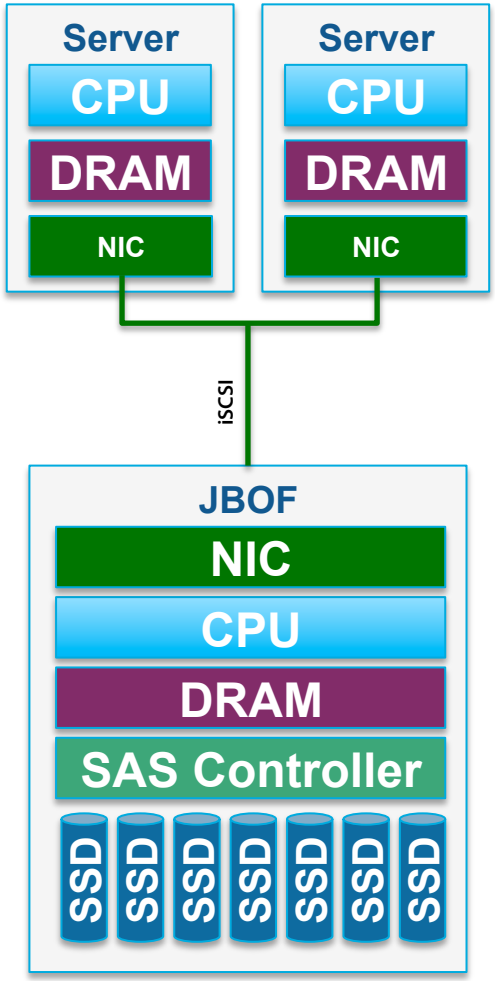
SAN evolution:

- Ethernet as a converged fabric:
  - 40/100 GbE faster than 16G/32G FC
- Flash memory:
  - High IOPS/low latency demanding applications
  - Less expensive

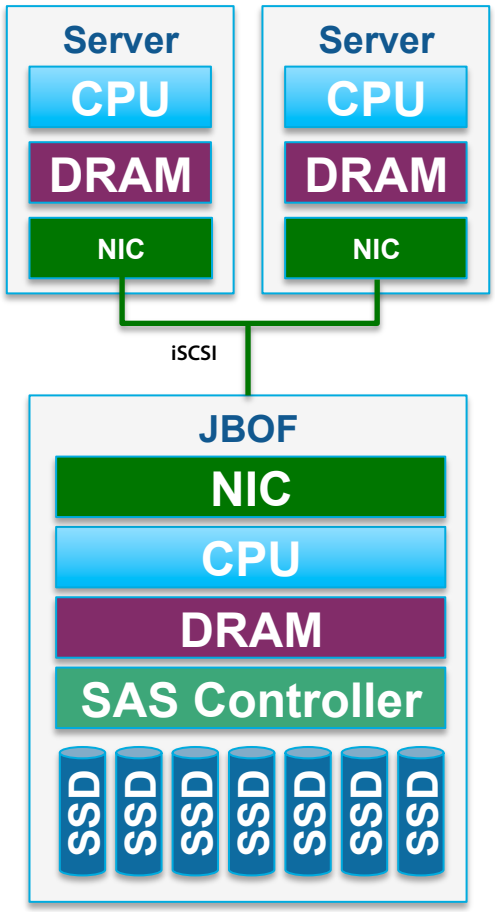
SSDs increase overall performance

Remaining hurdles:

- Still using SCSI protocol
- Still using the SAS protocol
- Legacy protocols for spinning disks limit the performance of SSDs



Phase 2 of the SSD Revolution : Coming Soon…



NVMe replaces SAS protocol:

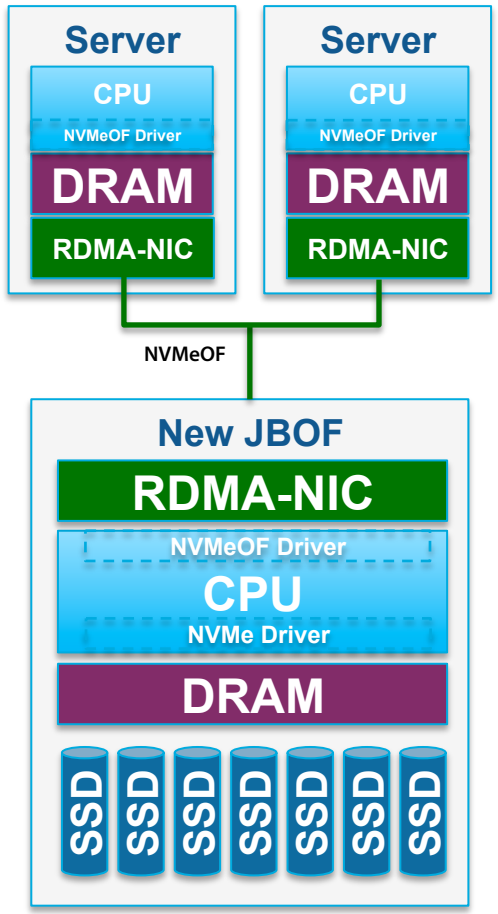
- NVMe vs. SCSI: **3x more IOPS**
- NVMe vs. SAS: **2-3x more bandwidth**

NVMeOF gives SSDs their free reign!

NVMeOF replaces iSCSI protocol:

- End-to-end protocol using NVMe over RDMA
- High IOPS/high bandwidth
- Extremely low latency RDMA protocol

NVMe & NVMeOF: end-to-end communication that allows SSD to reach its full potential



Explosion of MIOPs : Bottleneck at the CPU Level

Introduction of NVMe and NVMeOF puts pressure on the main CPU:

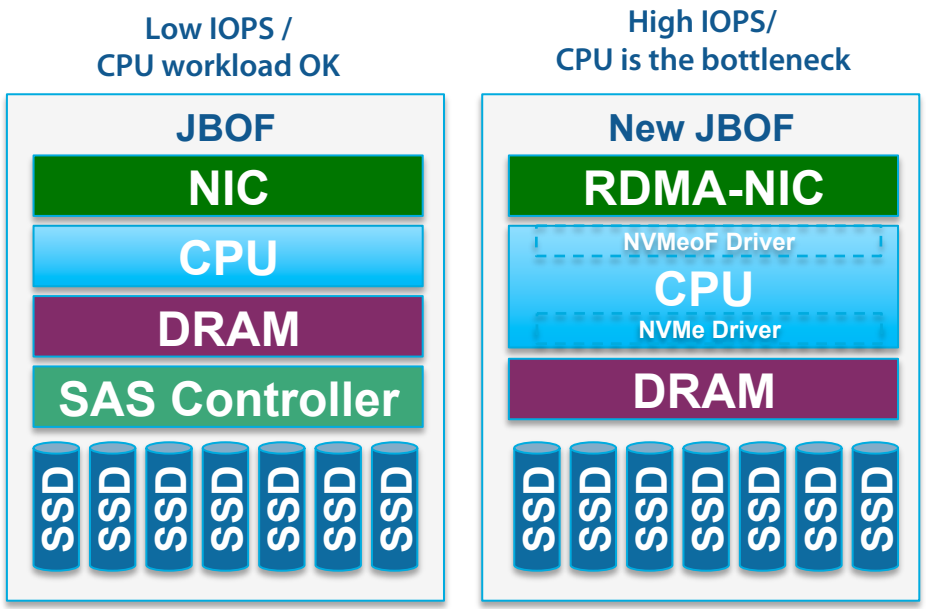
- A 40GbE can require up to 1.6 MIOPS for 100% Read / 30% Write
- A JBOF CPU Controller must handle 3.2 MIOPS:

NVMeOF Driver	1.6 MIOPS
NVMe Driver	1.6 MIOPs
Total	**3.2 MIOPs**

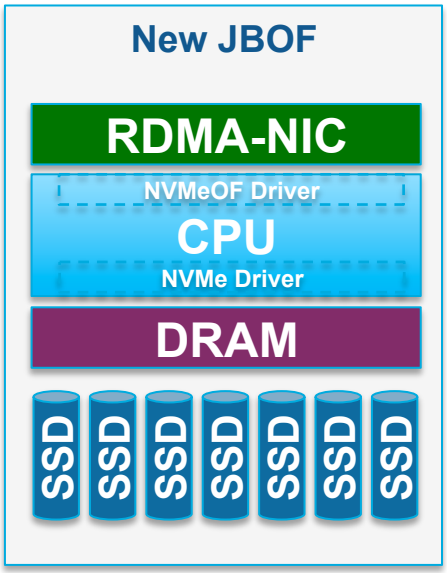
Offloading is required to sustain this explosion of MIOPS while delivering the lowest latency

Disk performances

- 15K HDD 210 IOPs
- 6Gb SATA SSD 90 KIOPs
- 12Gb SAS SSD 155 KIOPs
- NVMe SSD >700 KIOPs



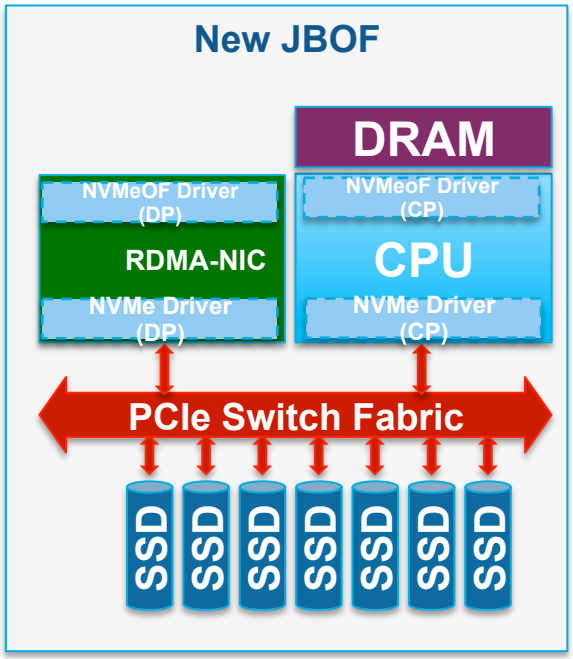
Offloading: the Ideal Solution for JBOF



Offloading

Offloading the NVMeOF/NVMe driver stack:

- NVMeOF/NVMe Data Plane on a SmartNIC
- NVMeOF/NVMe Control Plane on the CPU

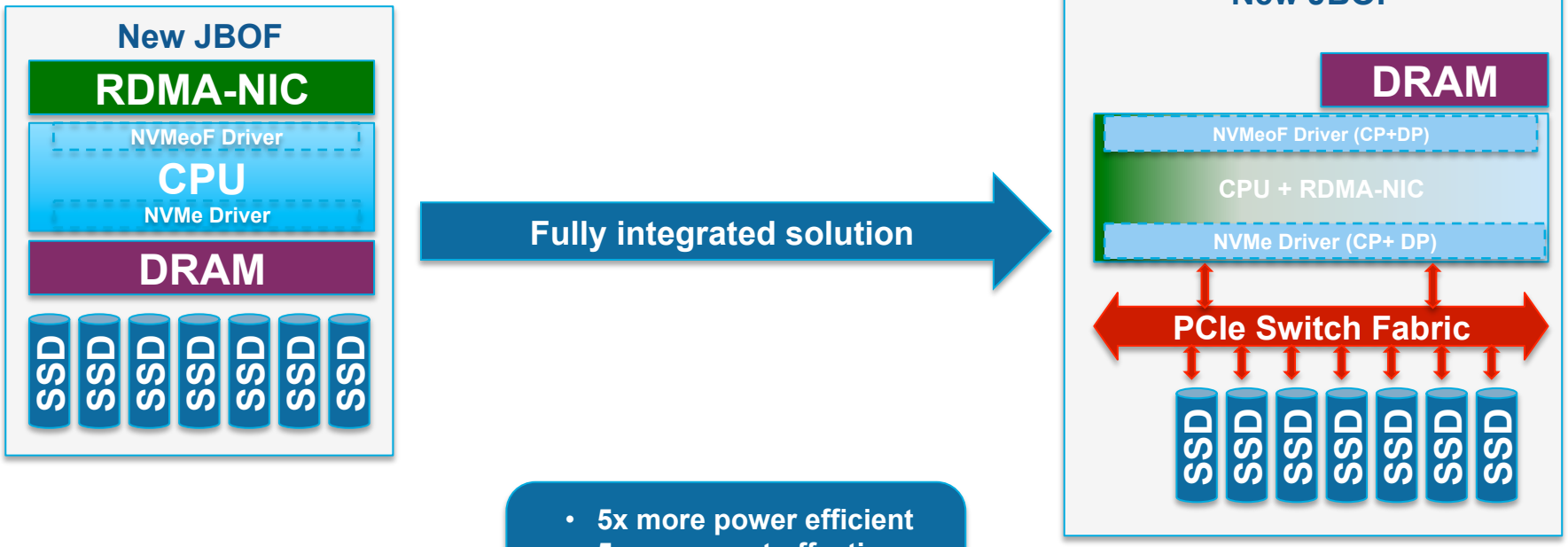


- **Frees up the main CPU for executing the main application**
- **Bypasses the main CPU system memory to save one memory copy, optimizing latency**

CPUs in JBOF: what is their role?

- JBOF are generally simple, meaning storage tasks are outsourced to a I/O proxy servers:
  - RAID / Erasure Coding
  - Logical volume management
  - And more…
- So, you're wondering…
  - …what's remaining on my main CPU in this case ?
  - …and why should I be spending money on an unnecessary device ?

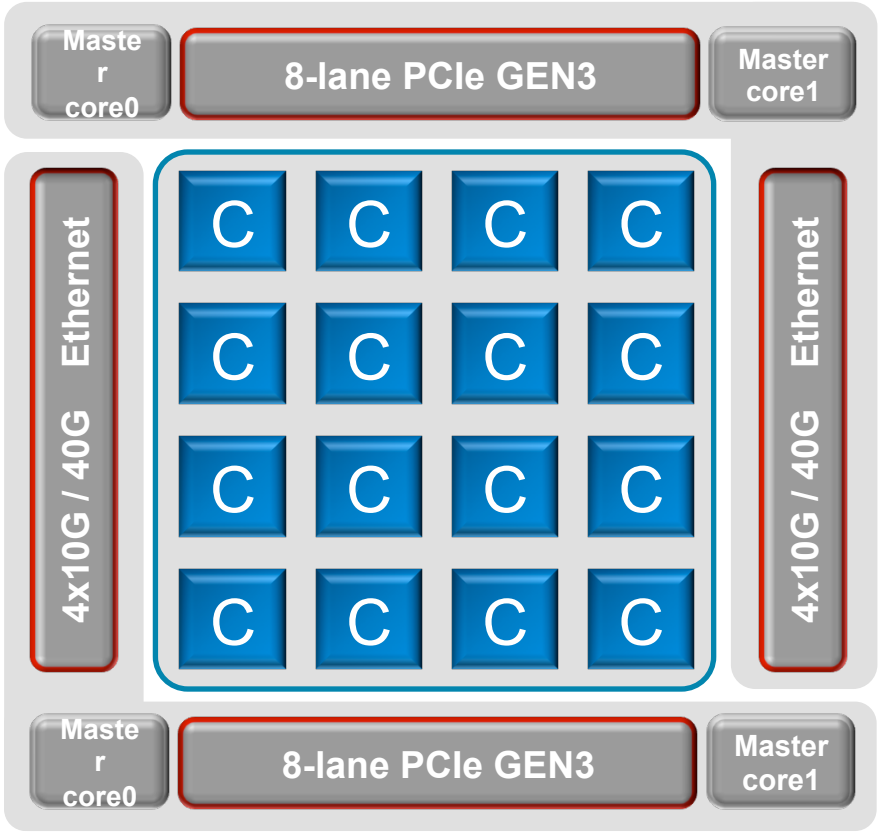
Optimum solution for JBOF:
Fully-Integrated Solution (CPU + NIC)



- 5x more power efficient
- 5x more cost-effective
- Less cooling
- Higher density

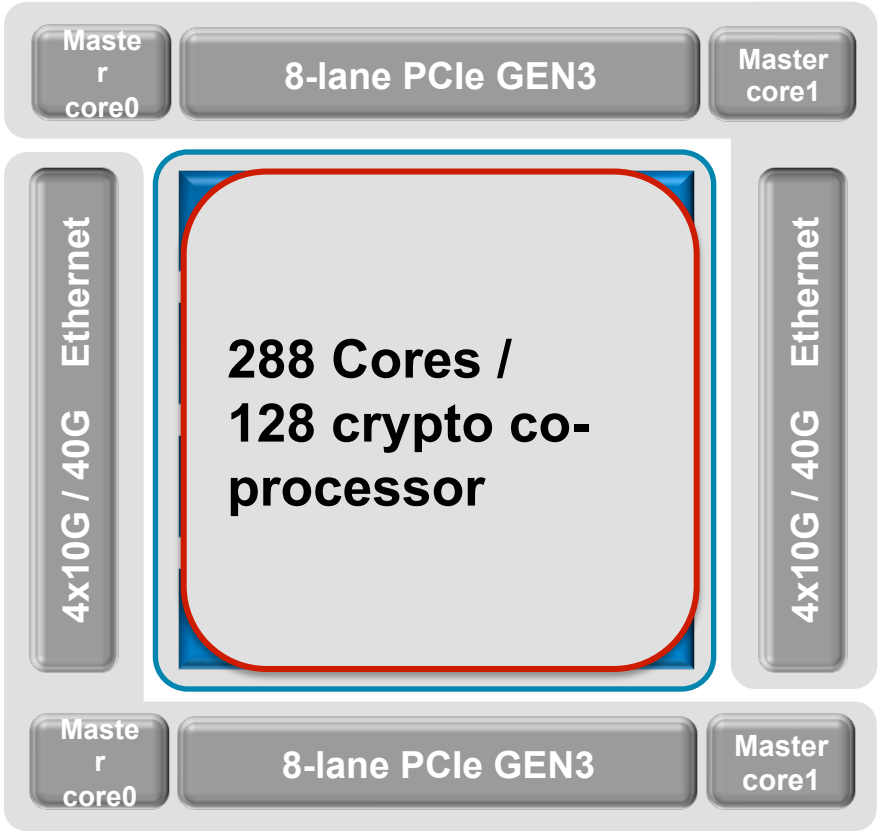
MPPA High Speed I/O Processor: The solution for JBOF

- High Speed Interfaces
  - 2x 40GbE
  - 2x PCIe Gen3 8-lanes (EP/RC)



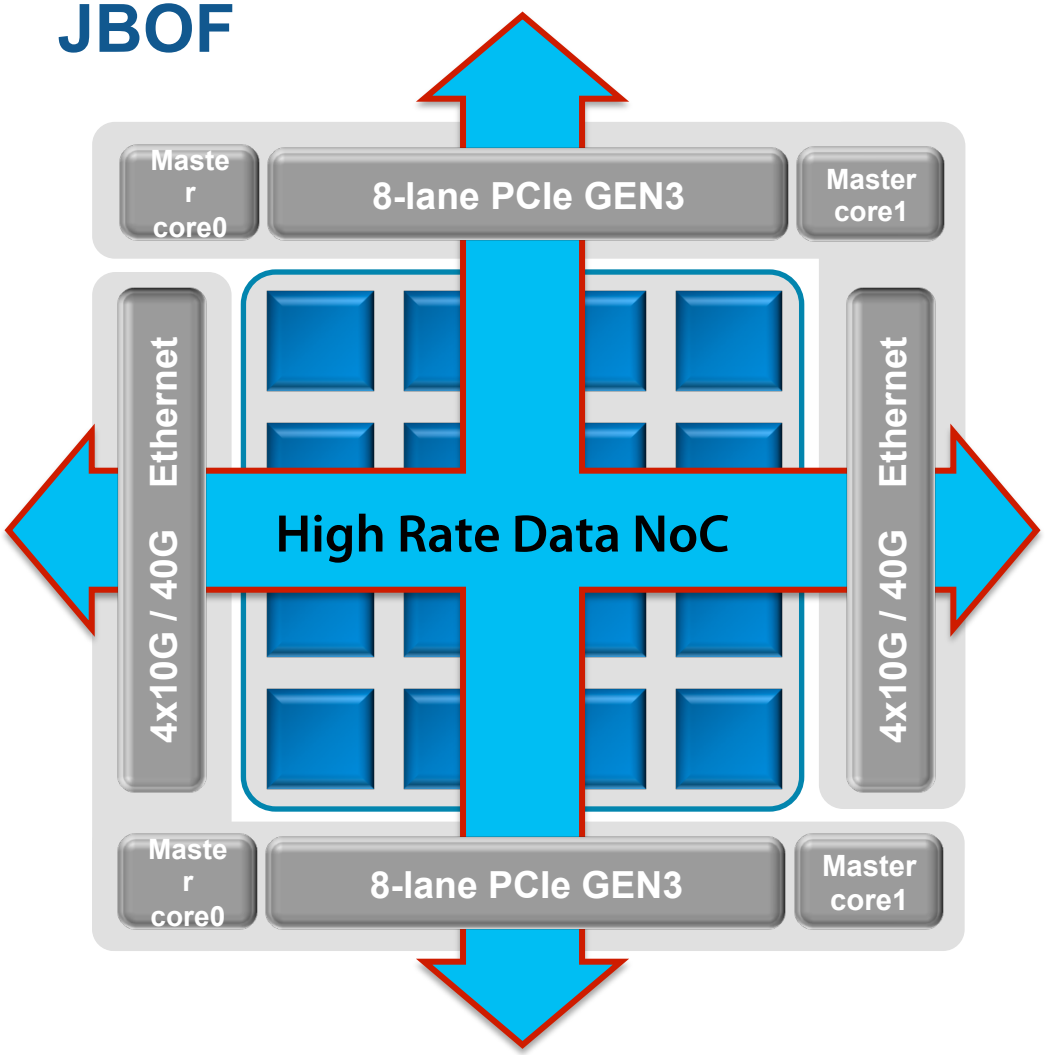
MPPA High Speed I/O Processor: The solution for JBOF

- High Speed Interfaces
  - 2x 40GbE
  - 2x PCIe Gen3 8-lanes (EP/RC)
- Connected to a large array of processing:
  - Full C/C++ Programmable
  - Dataplane execution



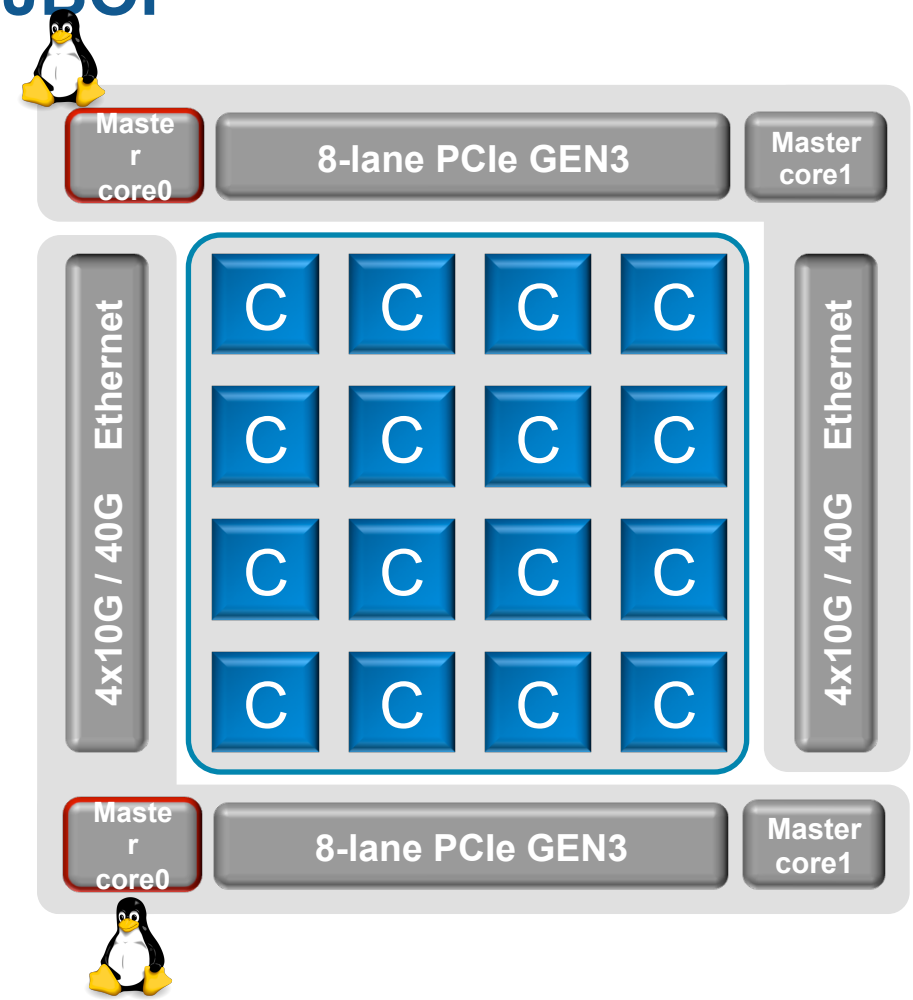
MPPA High Speed I/O Processor: The solution for JBOF

- High Speed Interfaces
  - 2x 40GbE
  - 2x PCIe Gen3 8-lanes (EP/RC)
- Connected to a large array of processing
  - Full C/C++ Programmable
  - Dataplane execution
- High bandwidth NoC
  - Direct packet-to-core delivery
  - High bandwidth / Low Latency



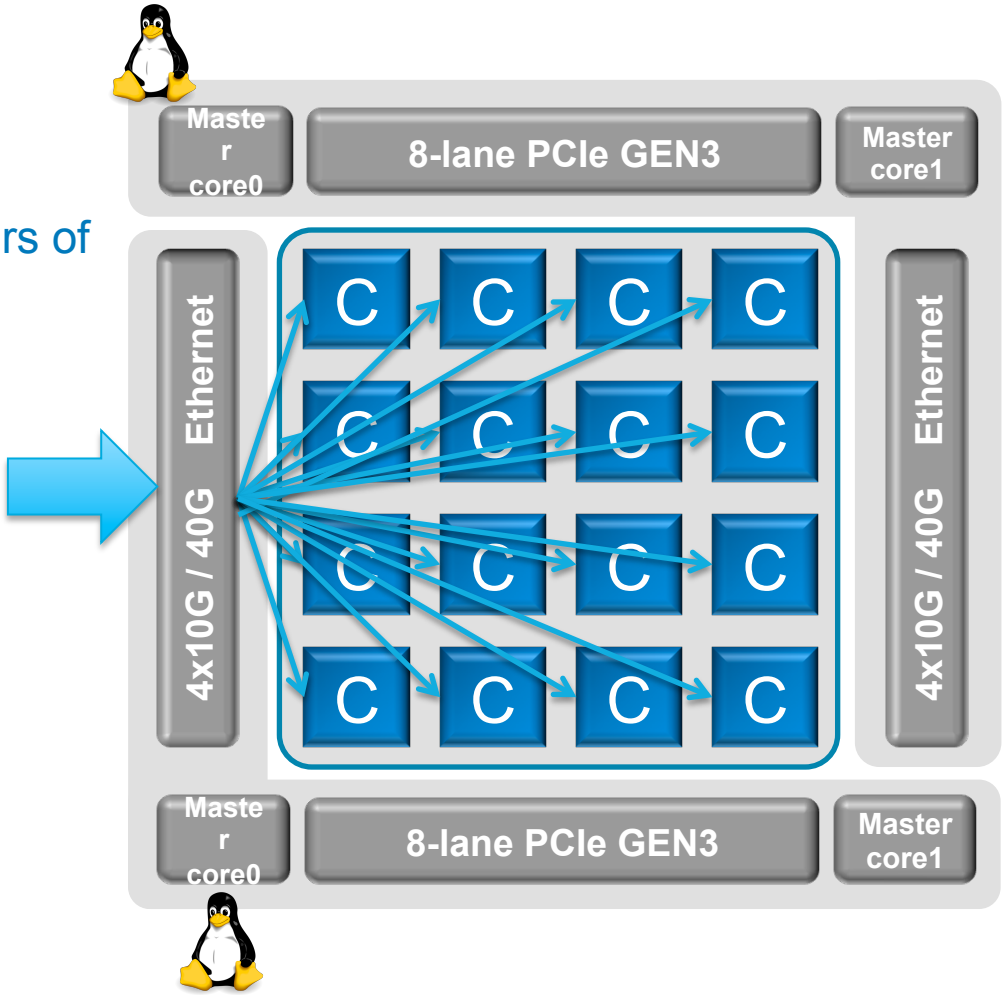
MPPA High Speed I/O Processor: The solution for JBOF

- High Speed Interfaces
  - 2x 40GbE
  - 2x PCIe Gen3 8-lanes (EP/RC)
- Connected to a large array of processing
  - Full C/C++ Programmable
  - Dataplane execution
- High bandwidth NoC
  - Direct packet-to-core delivery
  - High bandwidth / Low Latency
- IO Master Cores
  - Linux + Control Plane



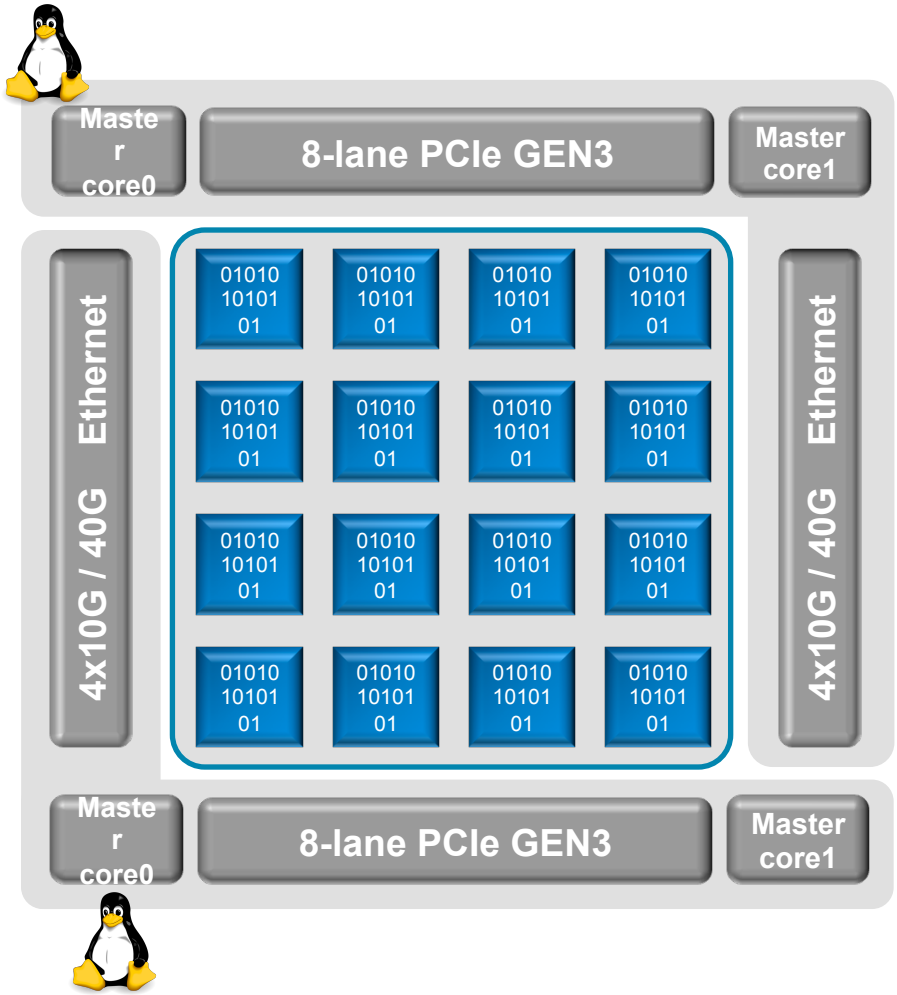
NVMeoF on MPPA IO processor

- MIOPS is an embarrassingly parallel problem
  - Easy to distribute on 16 clusters of 16 cores



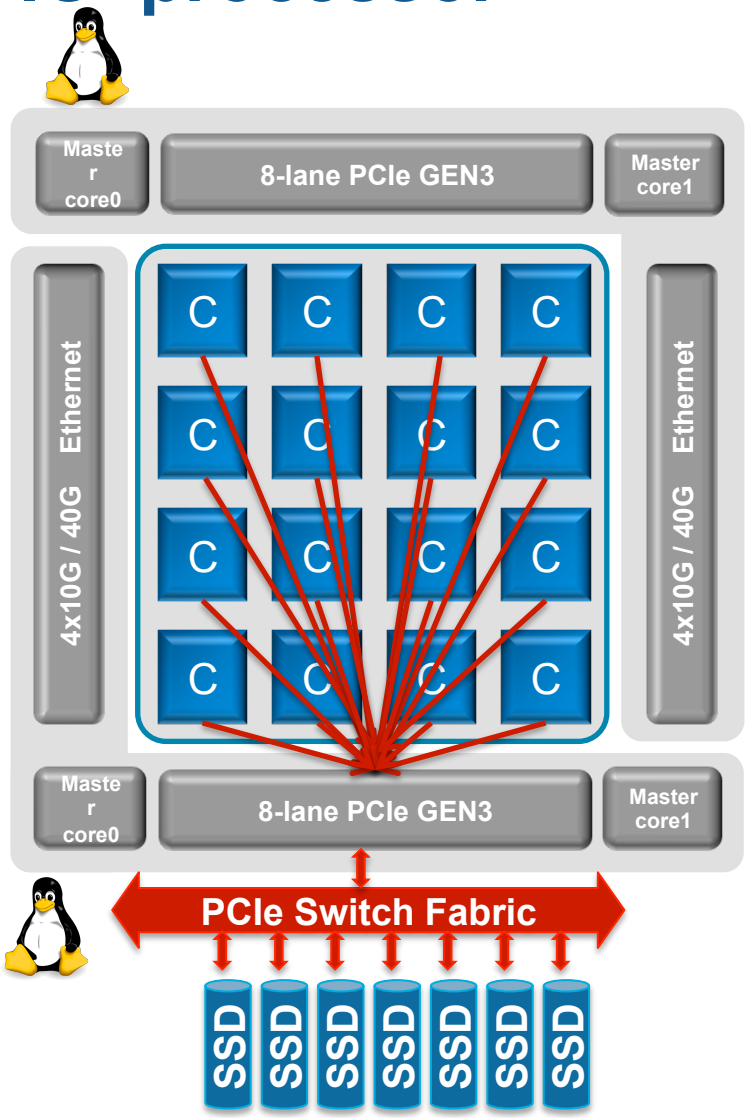
NVMeoF on MPPA IO processor

- MIOPS is an embarrassingly parallel problem
  - Easy to distribute on 16 clusters of 16 cores
- Dataplane executed in the Clusters
  - RoCE decapsulation
  - NVMeoF to NVMe translation
  - Optional encryption/compression/ erasure Coding



NVMeoF on MPPA IO processor

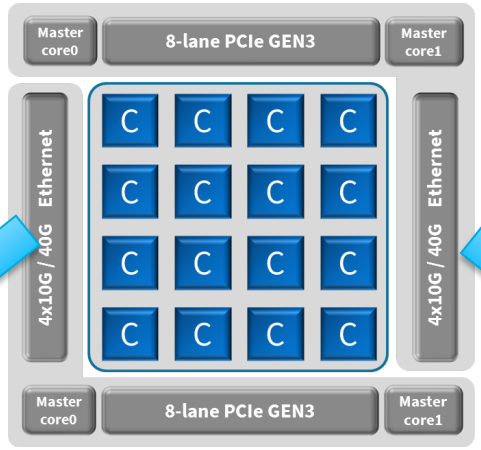
- MIOPS is an embarrassingly parallel problem
  - Easy to distribute on 16 clusters of 16 cores
- Dataplane executed in the Clusters
  - RoCE decapsulation
  - NVMeoF to NVMe translation
  - Optional encryption/compression/ erasure Coding
- NVMe Commands directly executed from the MPPA IO processor
  - PCIe Root Complex
  - PCIe Endpoint using peer-to-peer
  - 1 SQ/CQ per clusters and per SSD

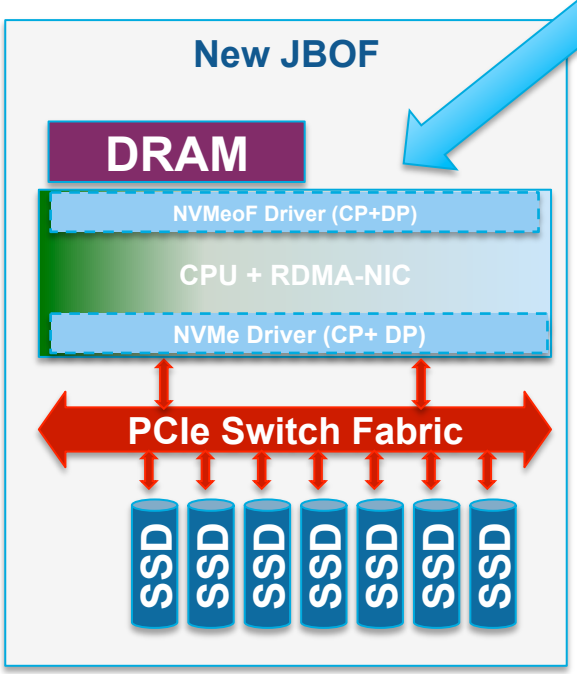


Conclusion

Integrated CPU + NiC

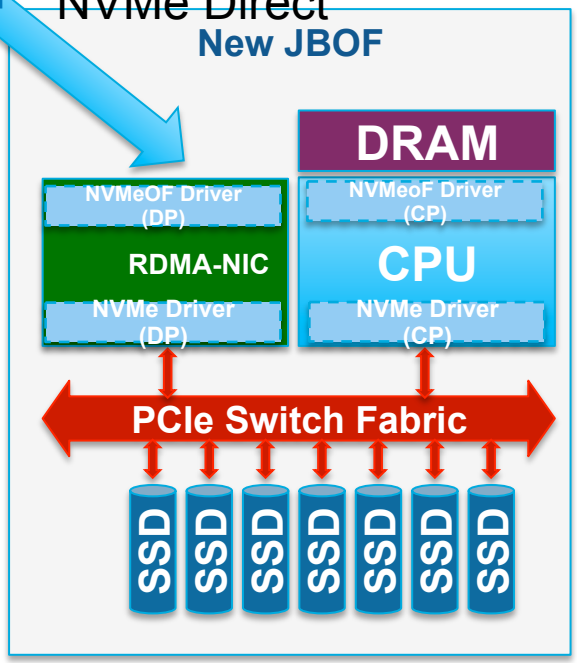
- 5x Power efficiency
- 5x Cost Effective
- Dense solution





Smart NiC for Offloading the main CPUs

- NVMeoF Offload
- NVMe Offload
- NVMe Direct



Kalray's MPPA2®-256 I/O processor

- Offloads networking & NVMe protocol stacks from main CPU
- Combines industry-leading low latency and high Ethernet bandwidth communication
- Optimizes power and data center density

KONIC40

KONIC80

Kalray's KONIC boards

- Deliver seamless integration to support NVMeOF

Contact me: pcouvert@kalray.eu