SNIA PERSISTENT MEMORY + SUMMIT 2022 COMPUTATIONAL STORAGE

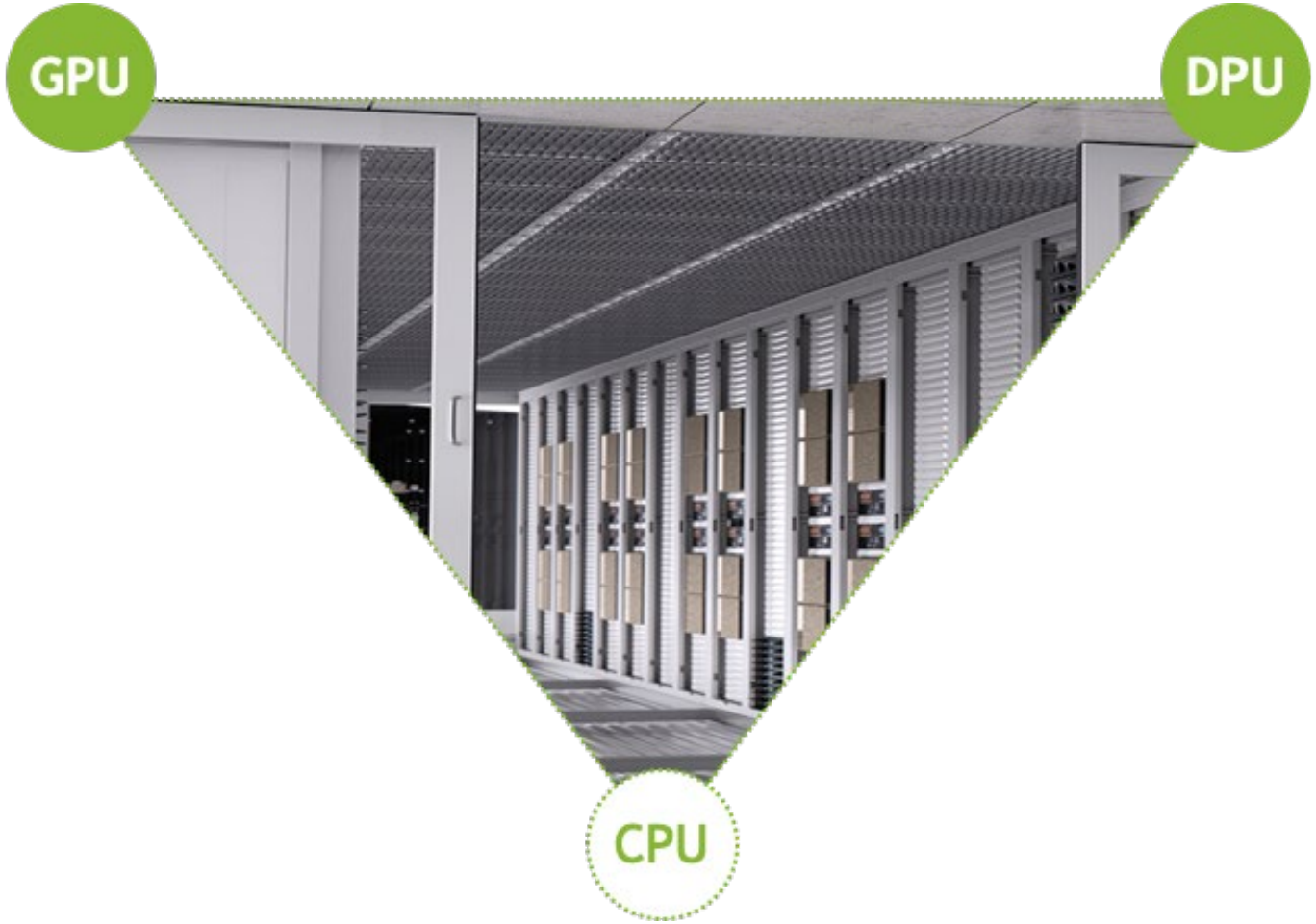
VIRTUAL EVENT • MAY 24-25, 2022

GPU + DPU for Computational Storage

Rob Davis, VP Storage Technology,
NVIDIA Network Platform Group

Storage that Thinks

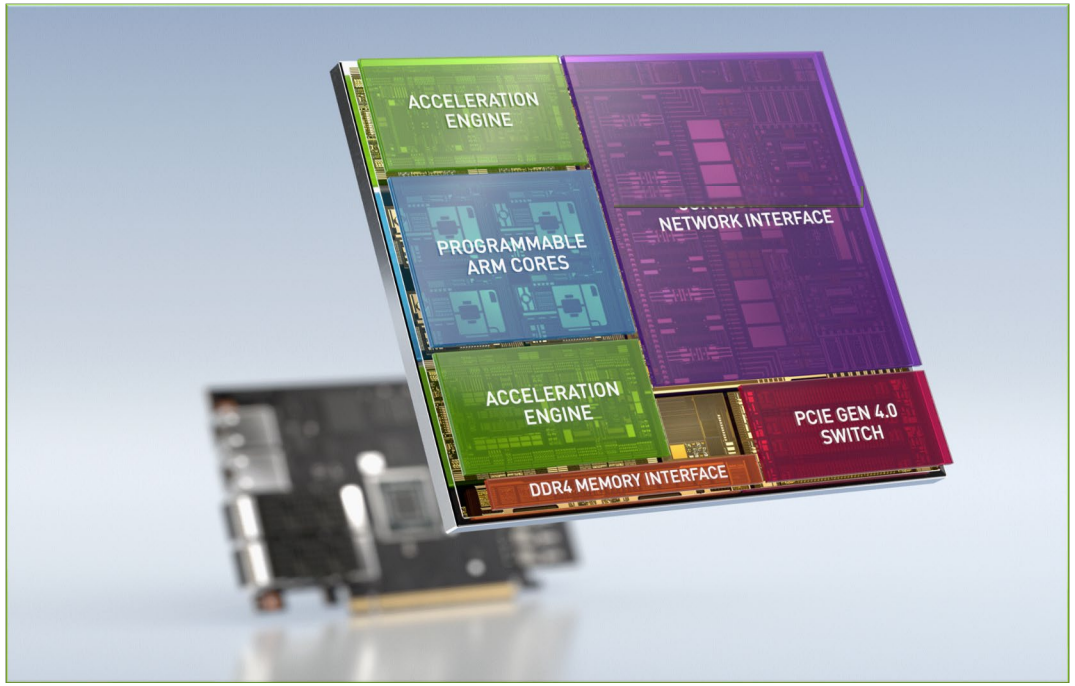
DPU – The New Building Block in a Data Center



https://blogs.nvidia.com/blog/2020/05/20/whats-a-dpu-data-processing-unit/

A DPU Can Accelerate Data Center Workloads

Data Center Infrastructure on a Chip



Offload | Accelerate | Isolate

Enterprise AI

Cloud Native HPC
Accelerated Computing

Cloud Gaming

Cloud Computing
Bare Metal as a Service

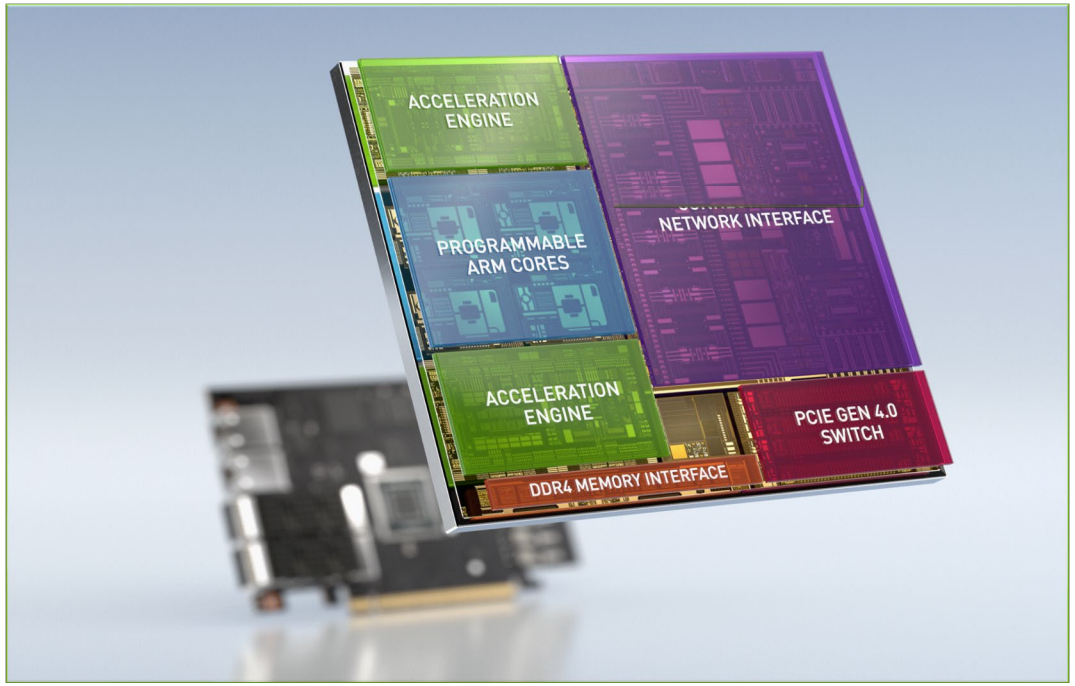
Cybersecurity
AI-Powered

5G
Industrial/Core/Edge

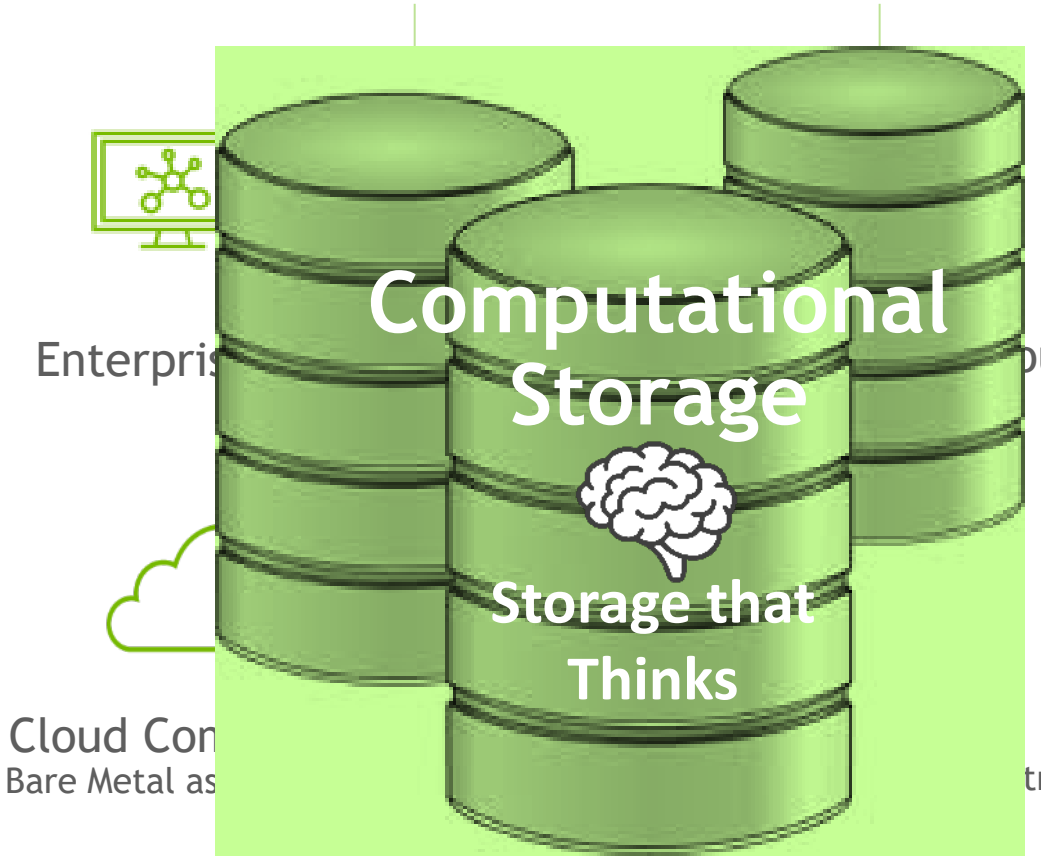
https://www.nvidia.com/en-us/networking/products/data-processing-unit/

A DPU Can Accelerate Data Center Workloads

Data Center Infrastructure on a Chip



Offload | Accelerate | Isolate

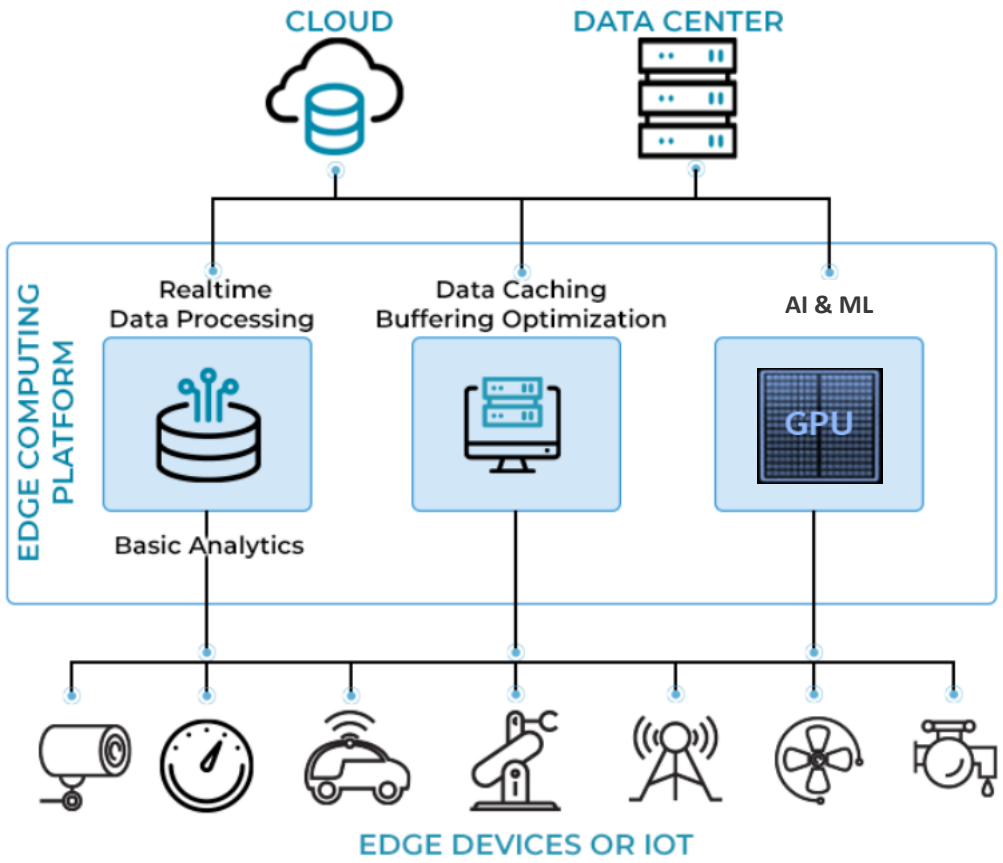


https://www.nvidia.com/en-us/networking/products/data-processing-unit/

Computational Storage and Edge Computing

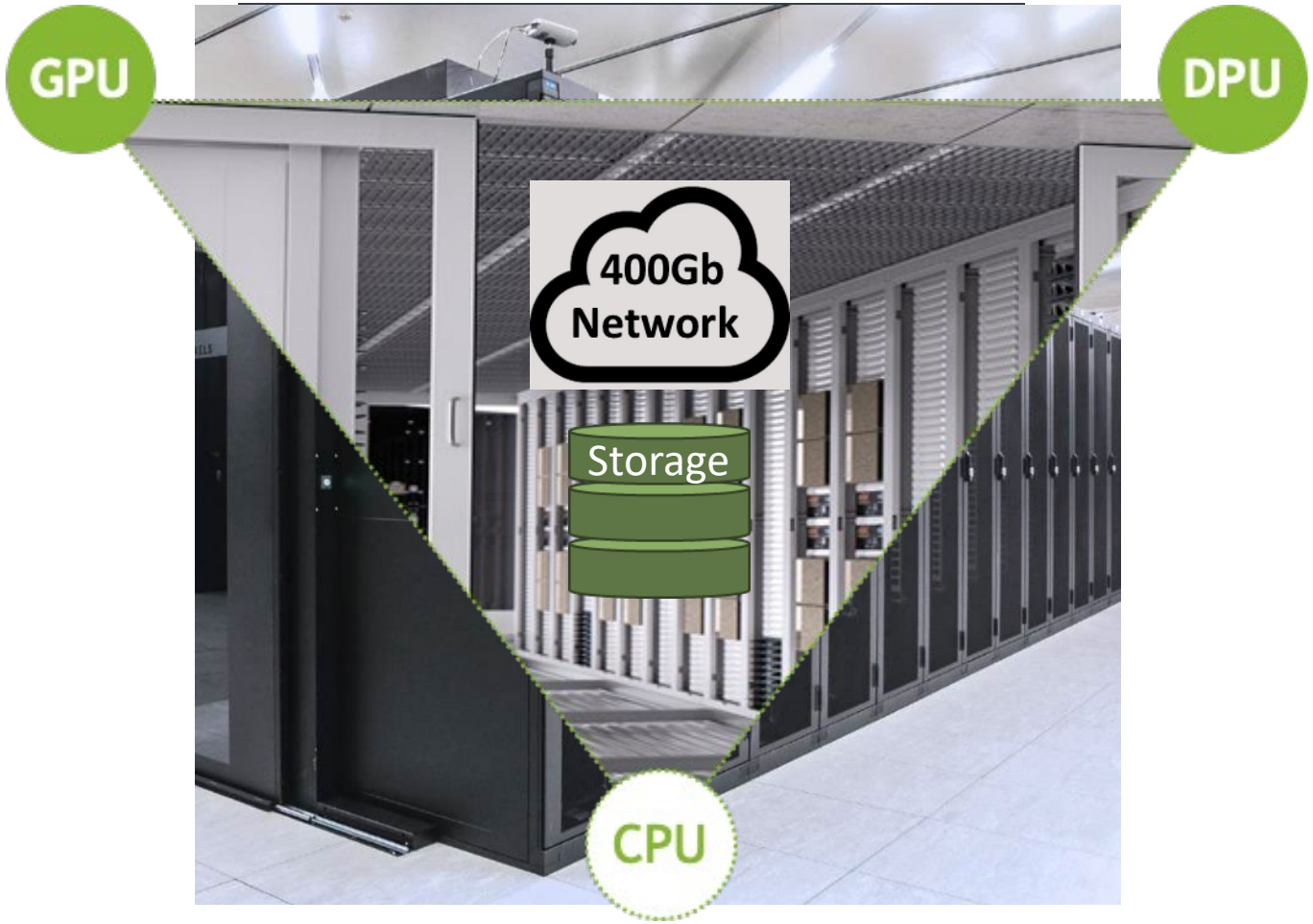
DATA CENTER COMPUTING

EDGE COMPUTING

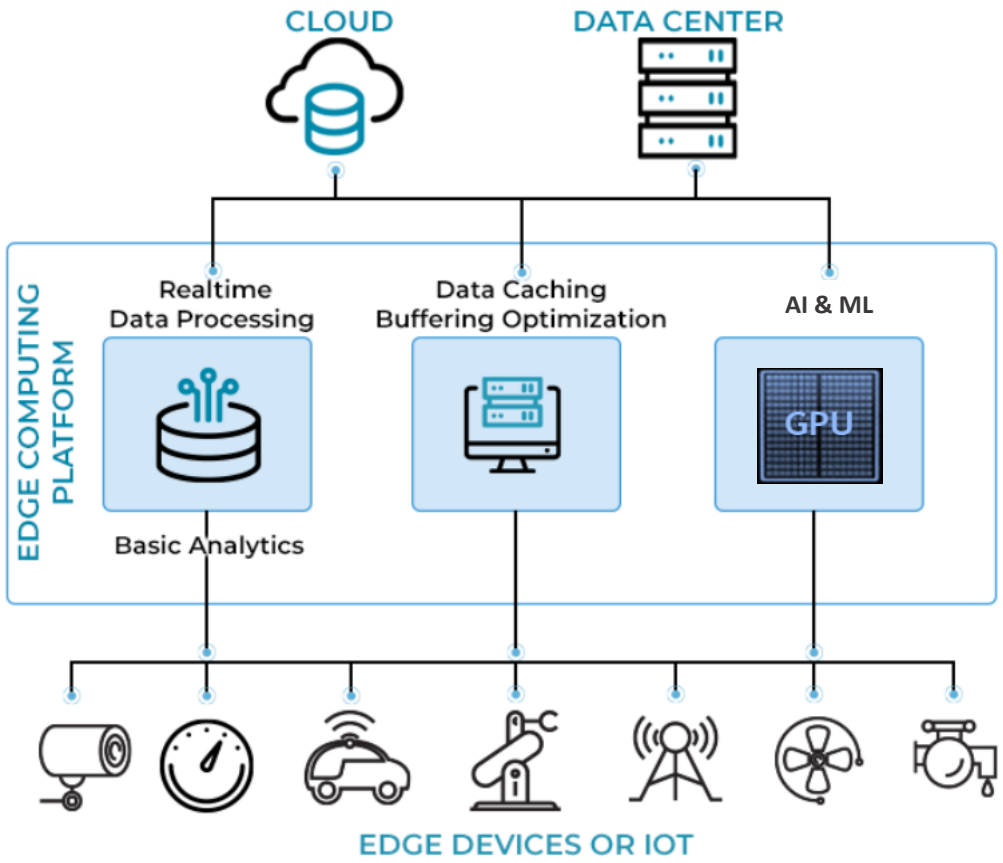


Computational Storage and Edge Computing

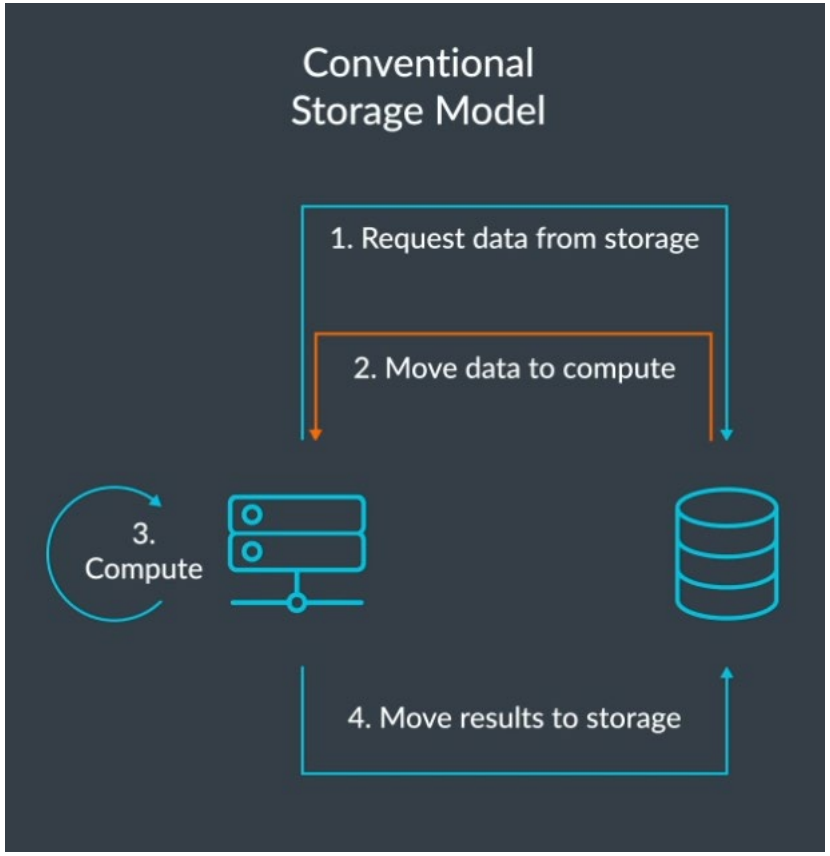
DATA CENTER COMPUTING

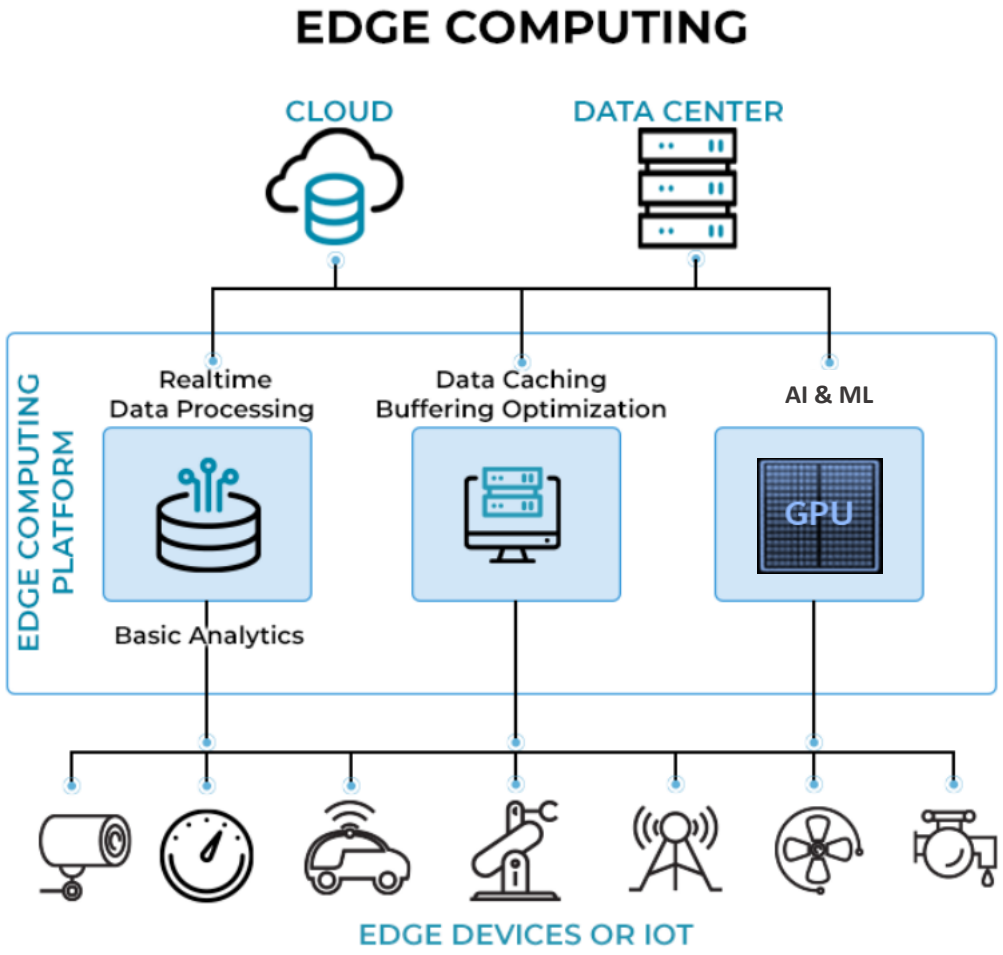


EDGE COMPUTING

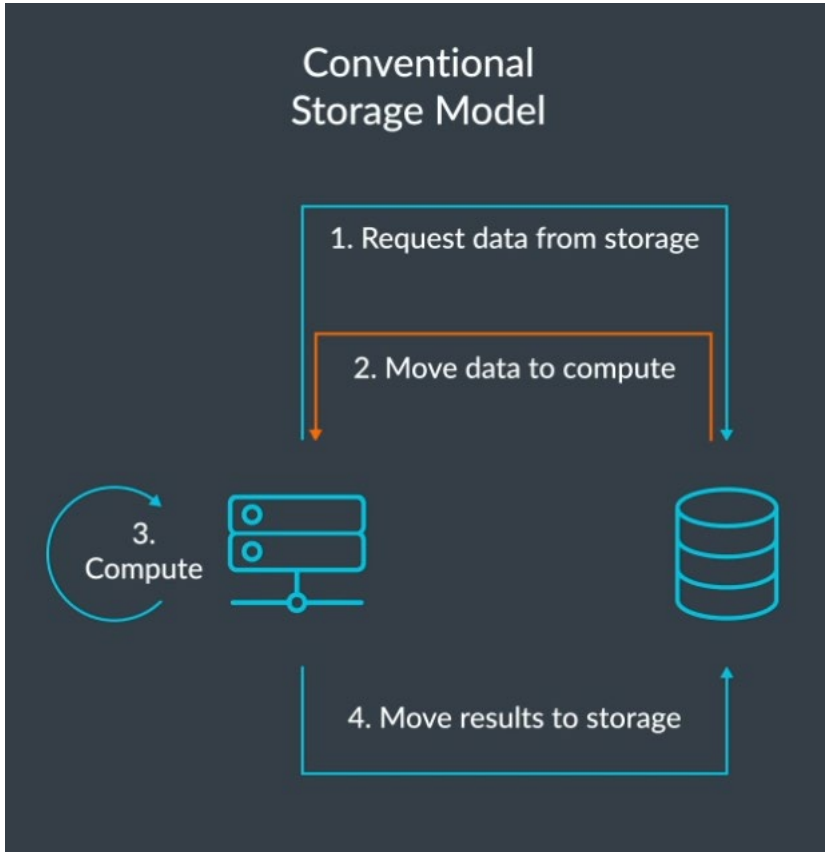


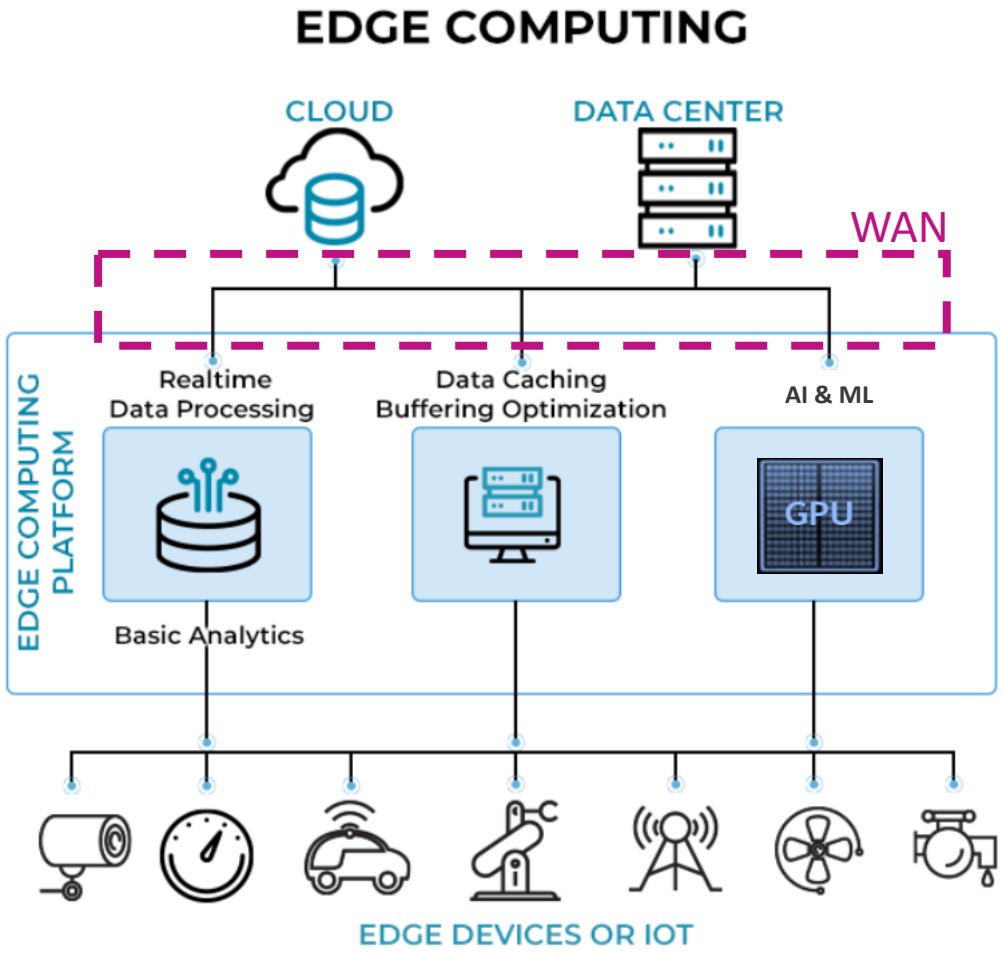
Computational Storage and Edge Computing



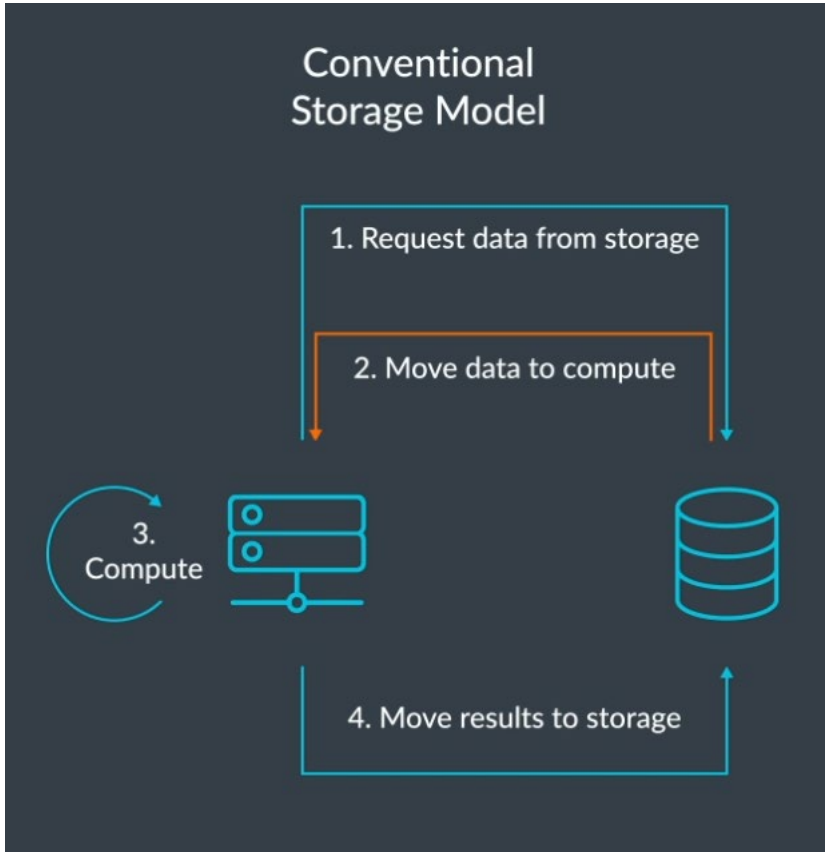


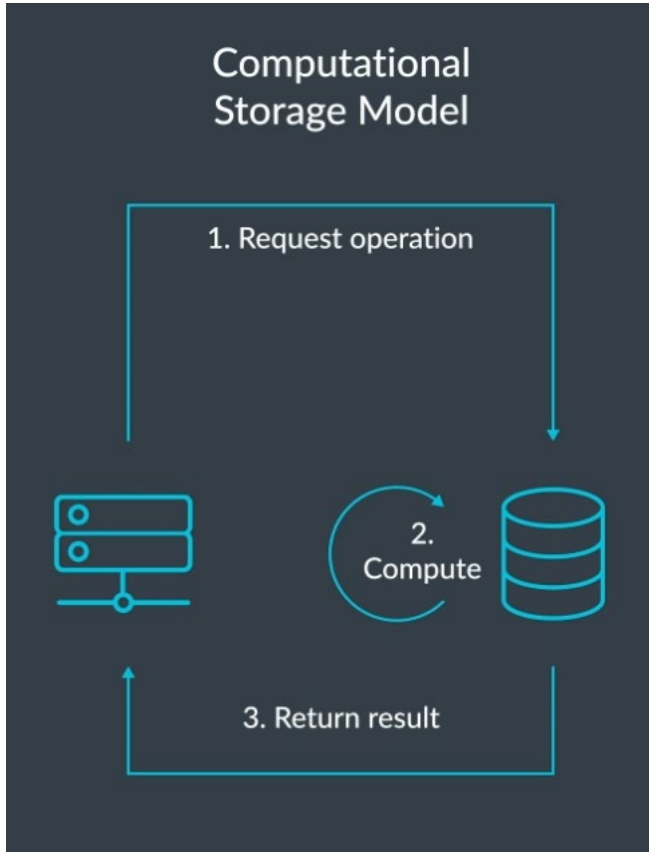
Computational Storage and Edge Computing





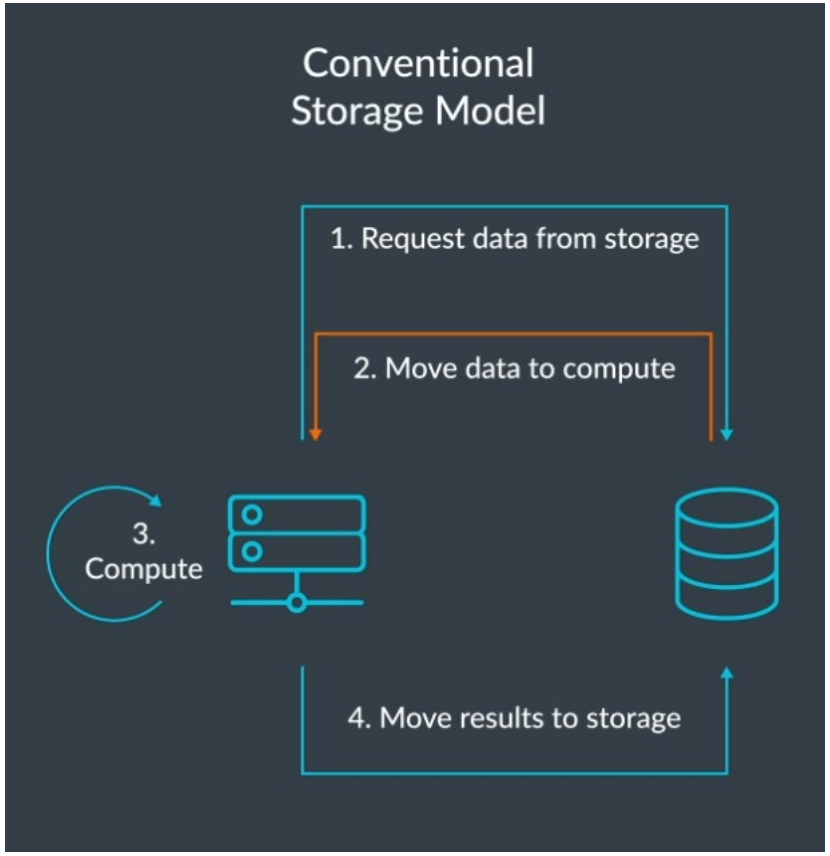
Computational Storage

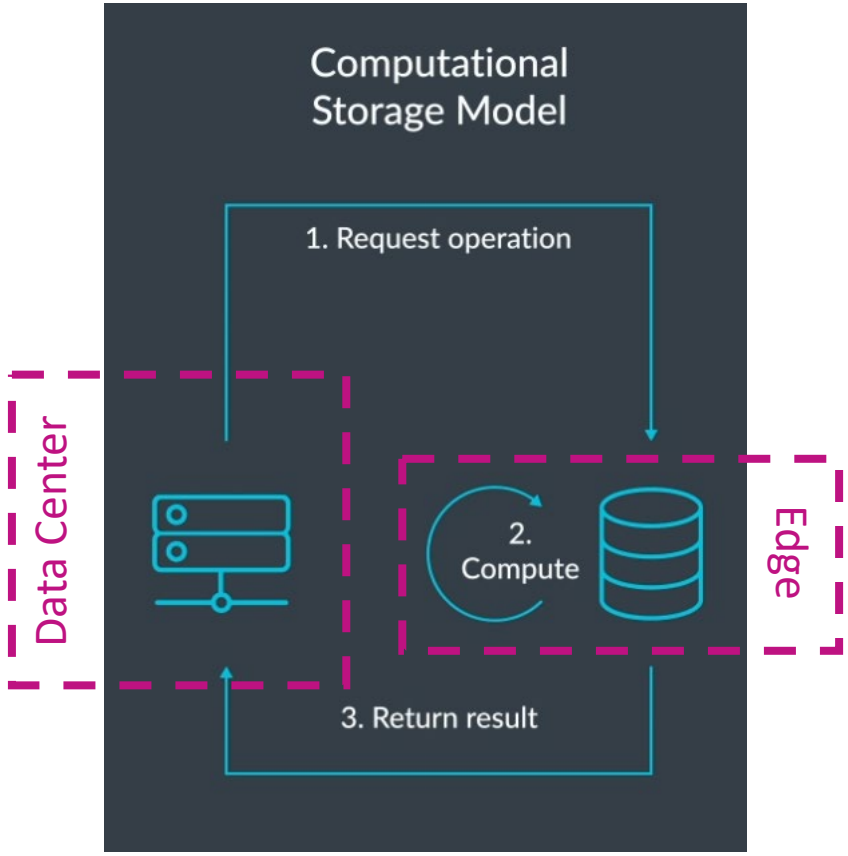




https://www.arm.com/solutions/storage/computational-storage

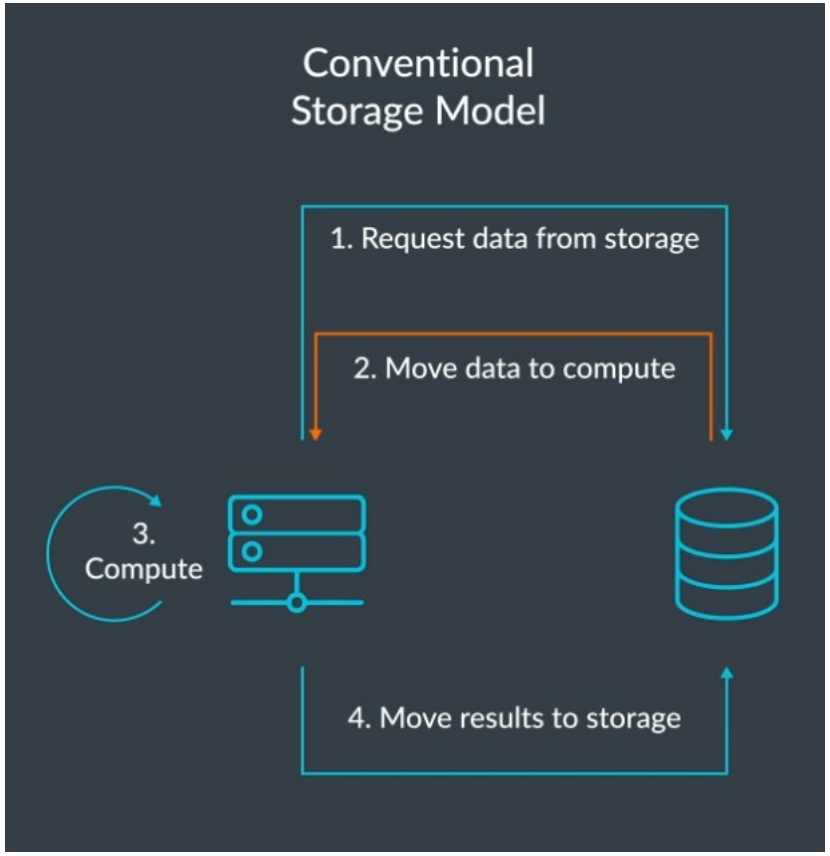
Computational Storage and Edge Computing

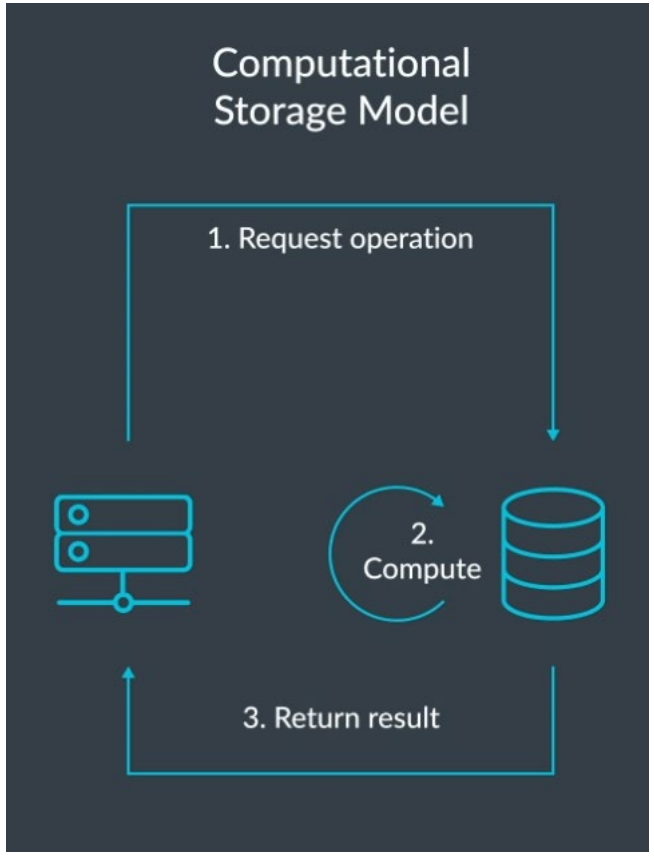




https://www.arm.com/solutions/storage/computational-storage

Computational Storage

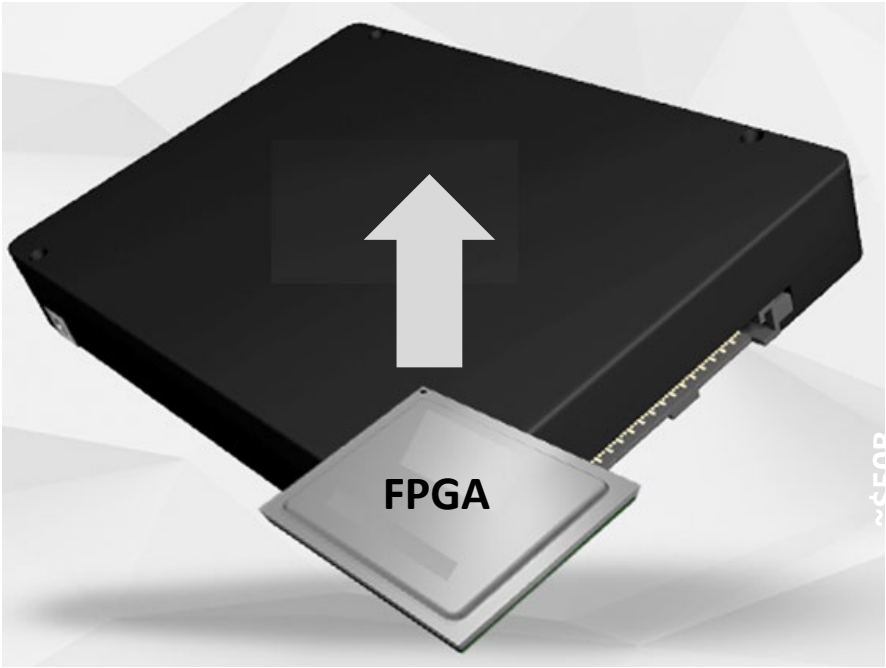




https://www.arm.com/solutions/storage/computational-storage

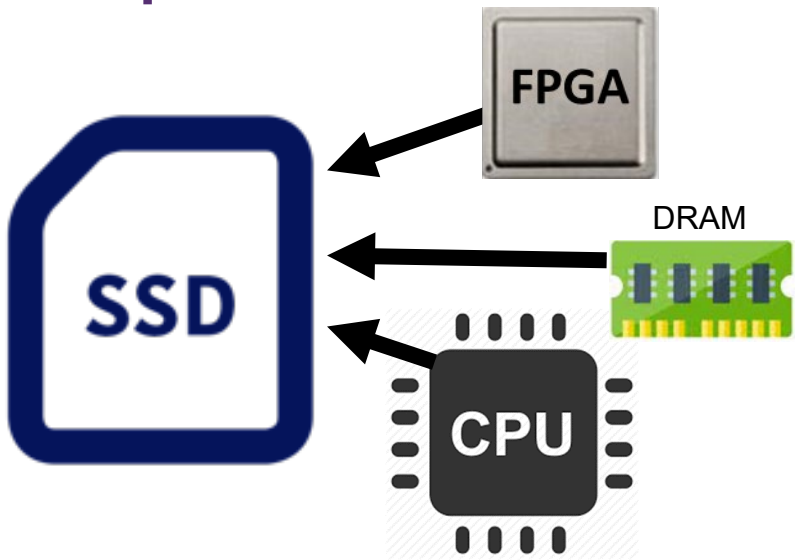
Computational Storage Solution Options

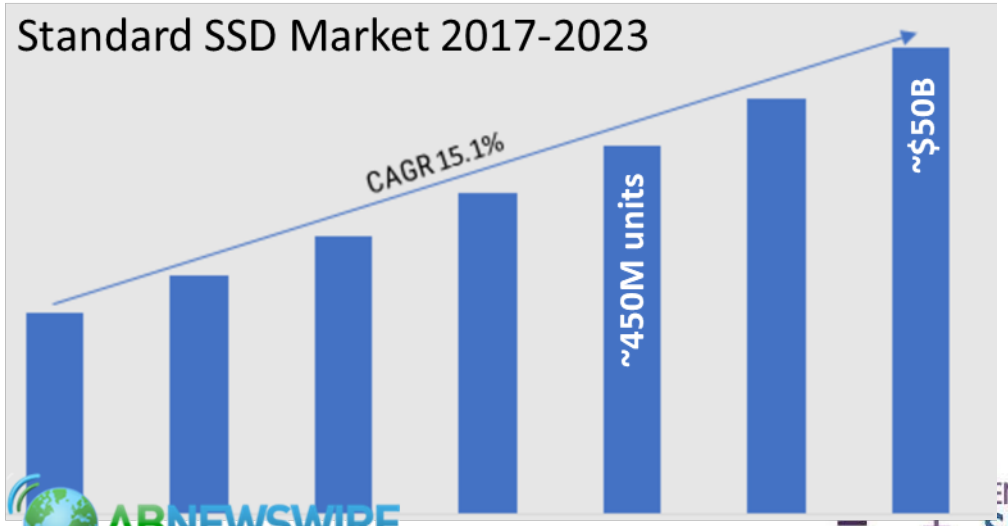
- FPGA based solutions
  - Need control CPU
  - Limited flexibility
    - Special technical expertise needed for reprograming
  - Form Factor?



Computational Storage Solution Options

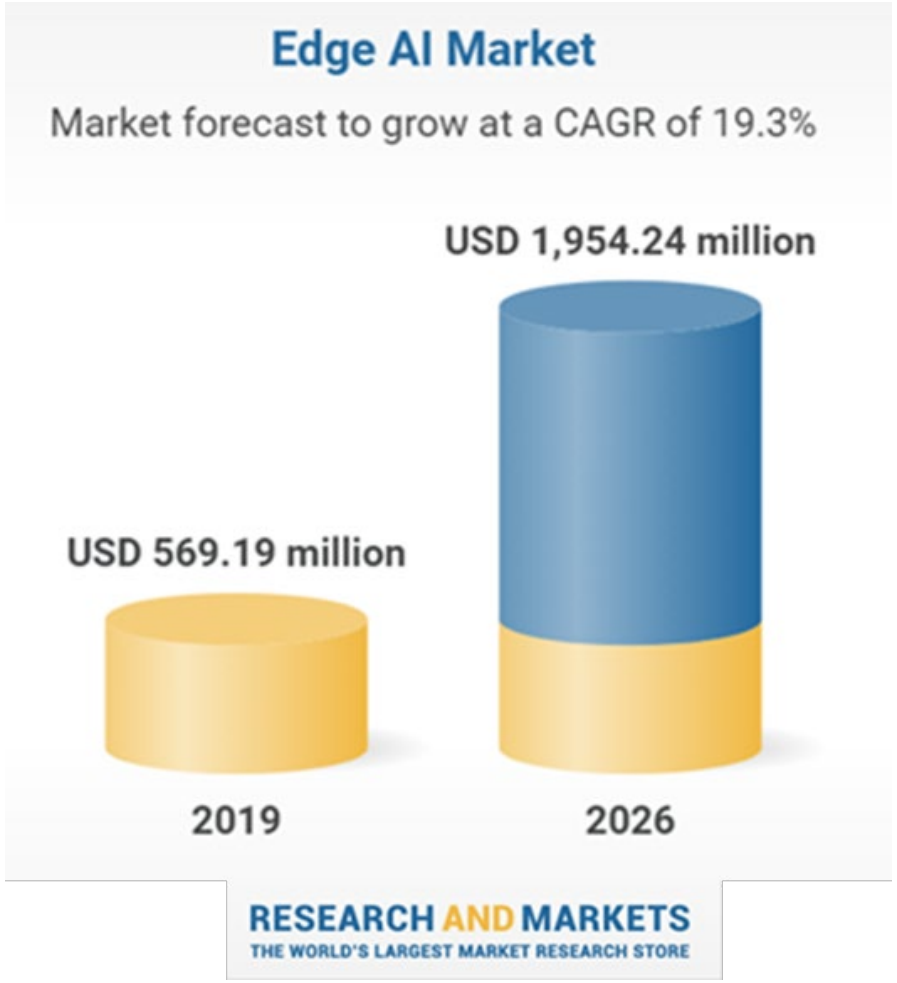
- Smart SSDs
  - Added CPU+DRAM performance or FPGA
  - Power and Cooling?
  - Not vendor volume product line for SSDs





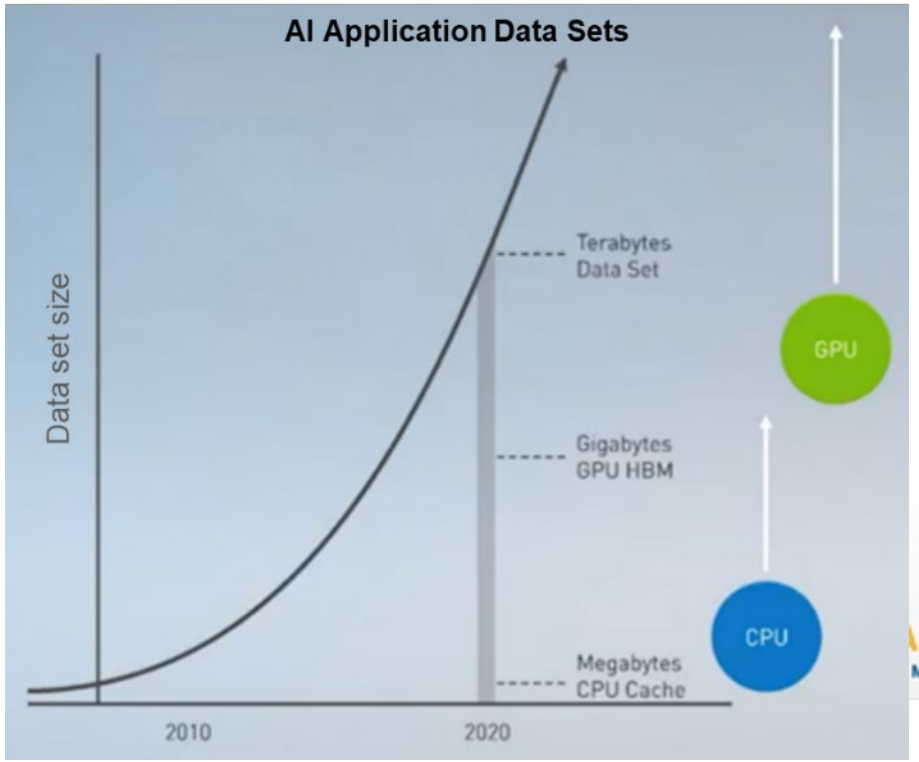
Computational Storage Solution Options

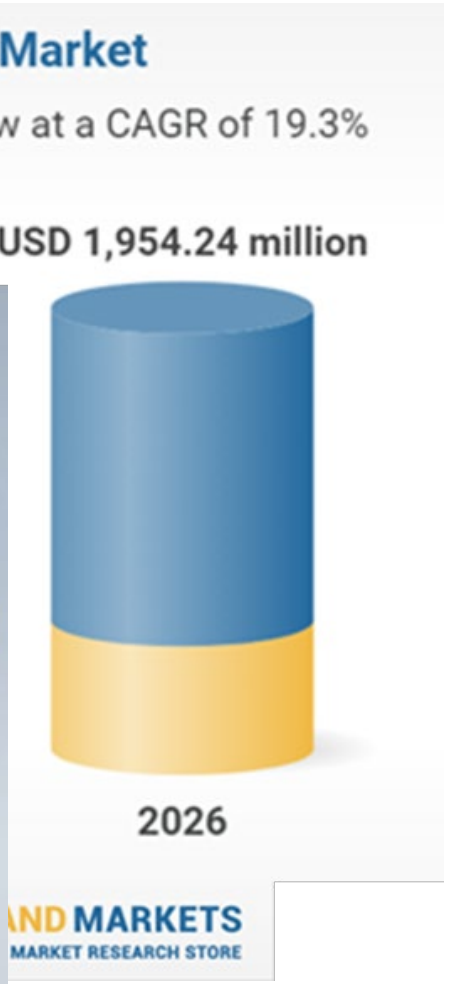
- CPU based solutions
  - Limited AL & ML capabilities
  - Data set size limited



Computational Storage Solution Options

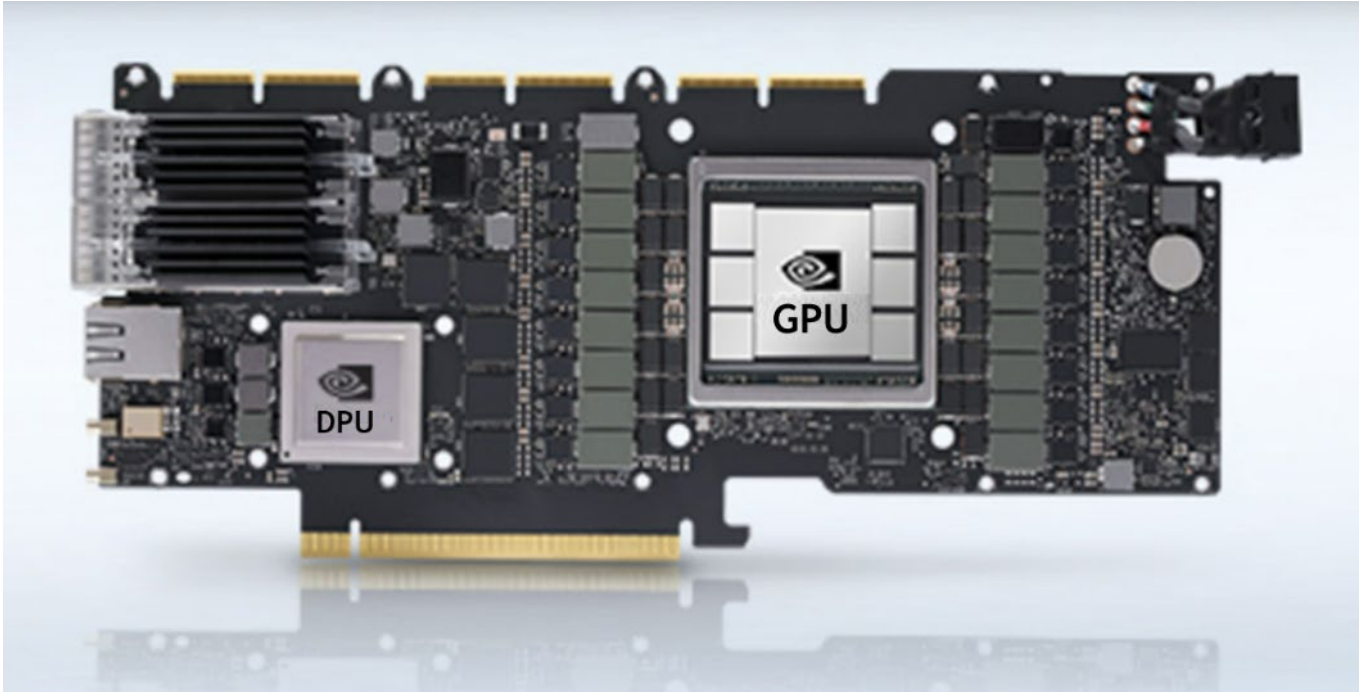
- CPU based solutions
  - Limited AL & ML capabilities
  - Data set size limited





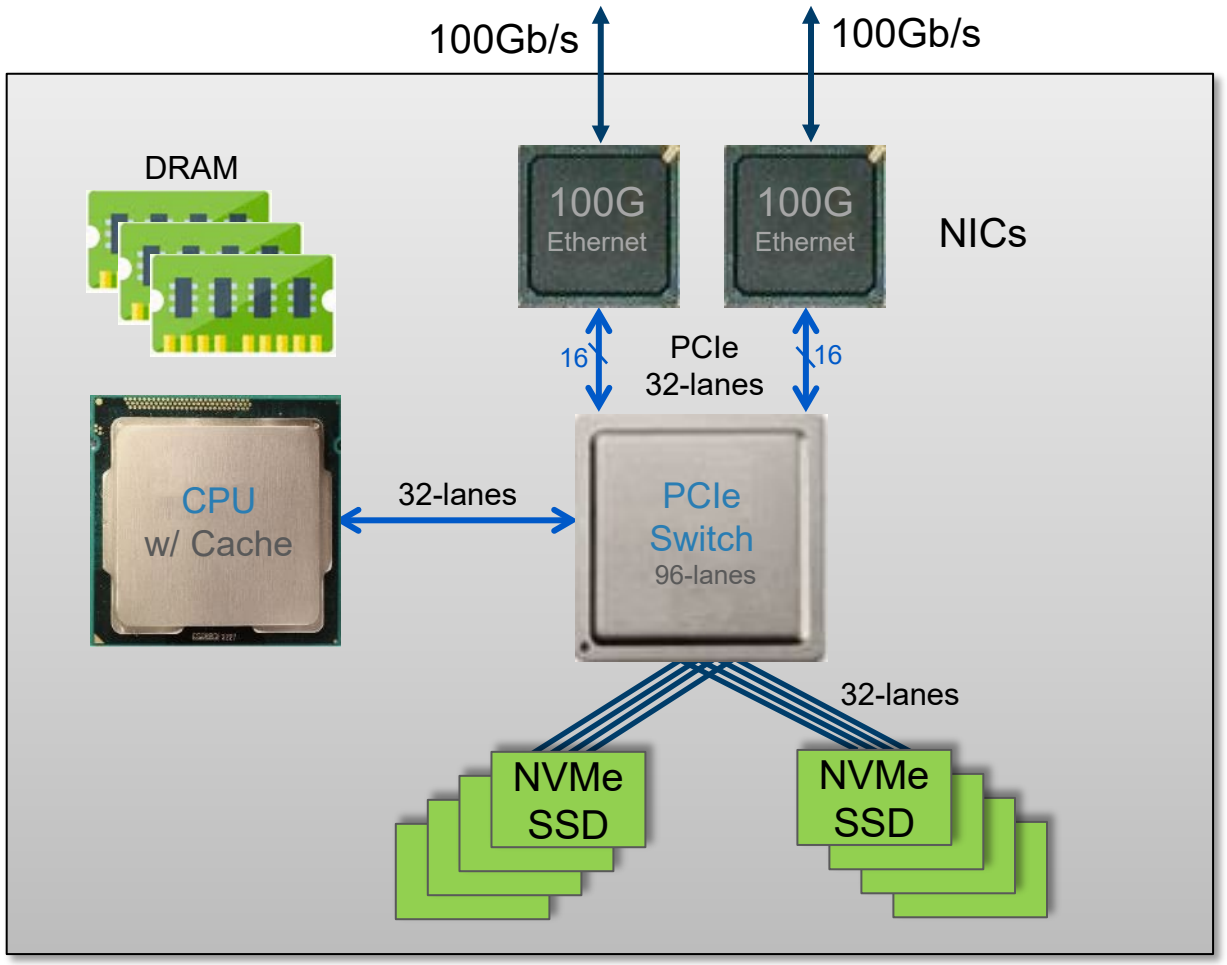
Computational Storage Solution Options

- GPU + DPU based solutions
  - Standard PCIe Form Factor
  - Standard components
  - Very flexible and broad open solution set including VERY strong AI
  - Very high performance and capacity storage IO



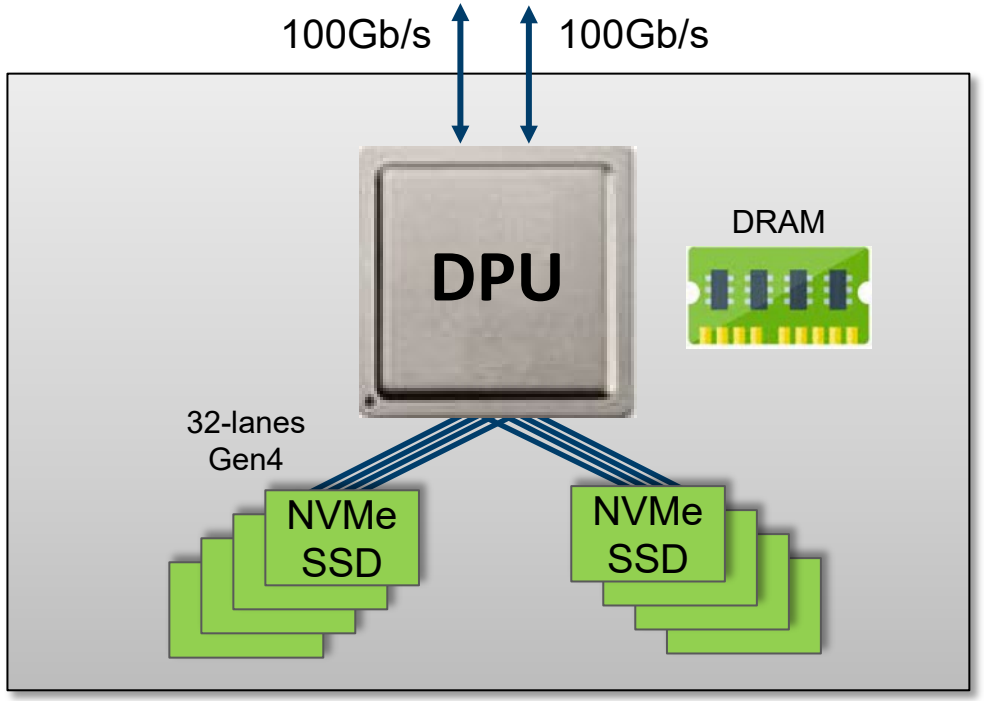
https://www.nvidia.com/en-us/data-center/products/egx-converged-accelerator/

DPU based storge systems



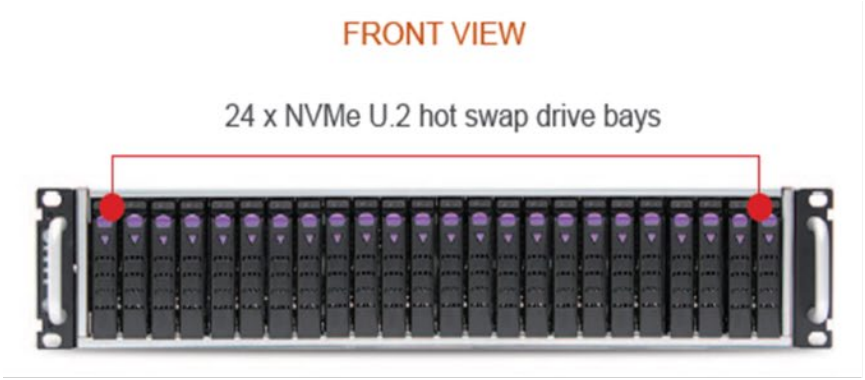
Conventional Storage Array/JBOF

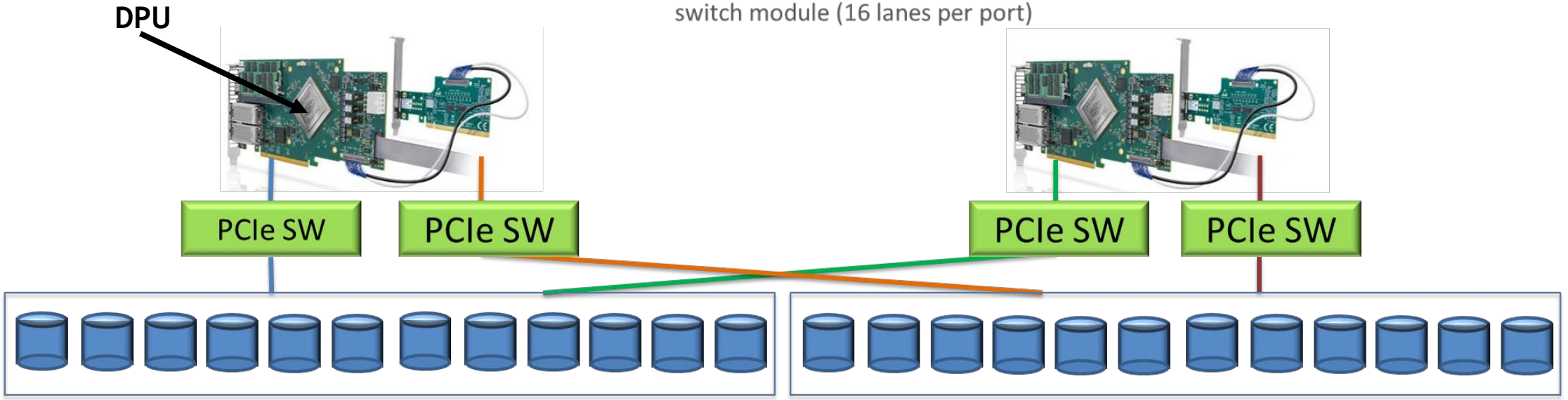
Saves Real Estate, Cost, Power, Cooling and Complexity



DPU Storage Solution

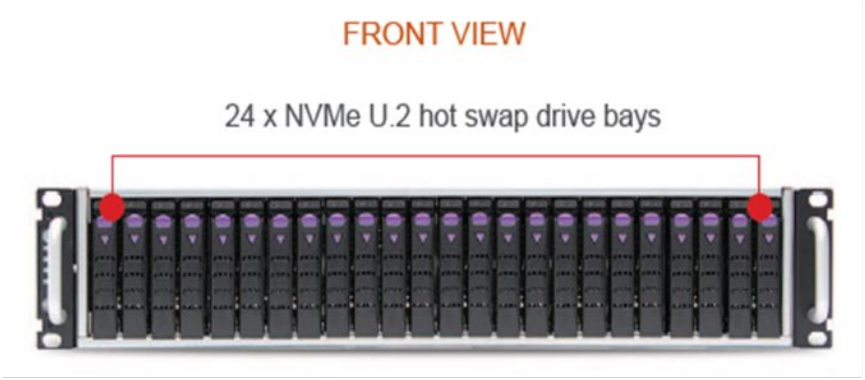
Available DPU based storage system



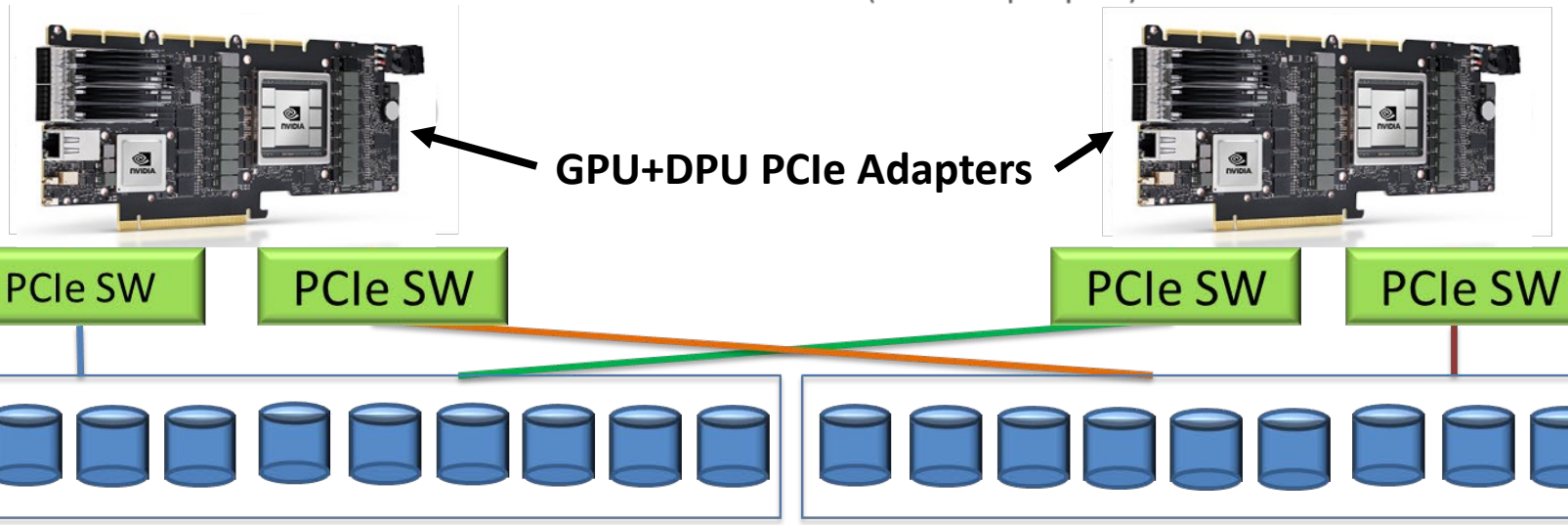


https://www.aicipc.com/en/pagefull/170-1572

GPU + DPU Computational Storage Solution

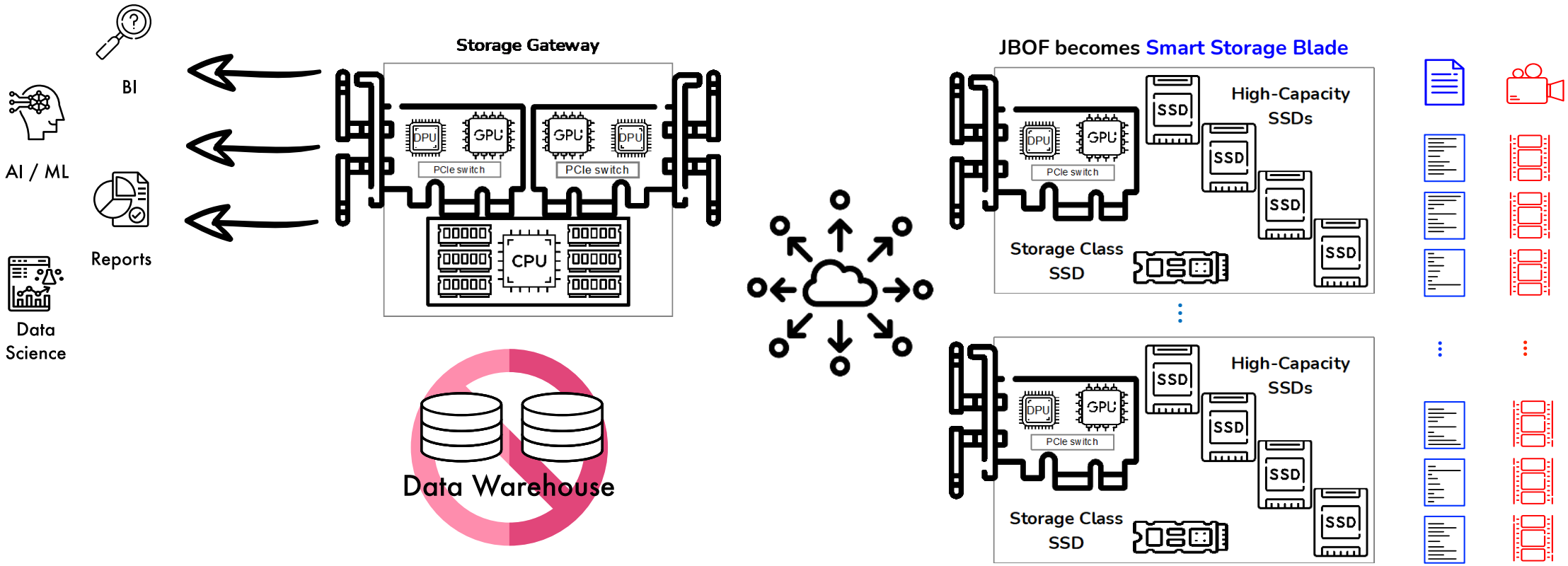


4 x PCIe high speed ports per
switch module (16 lanes per port)



GPU + DPU Computational Storage Solution

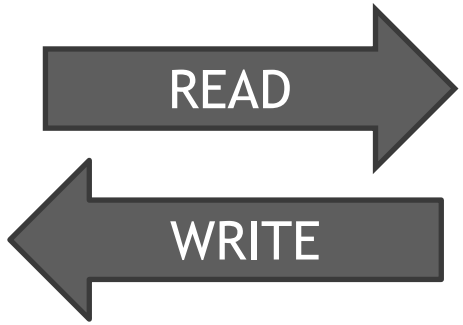
Analyze in-place, Extract / Load / Transform, On demand

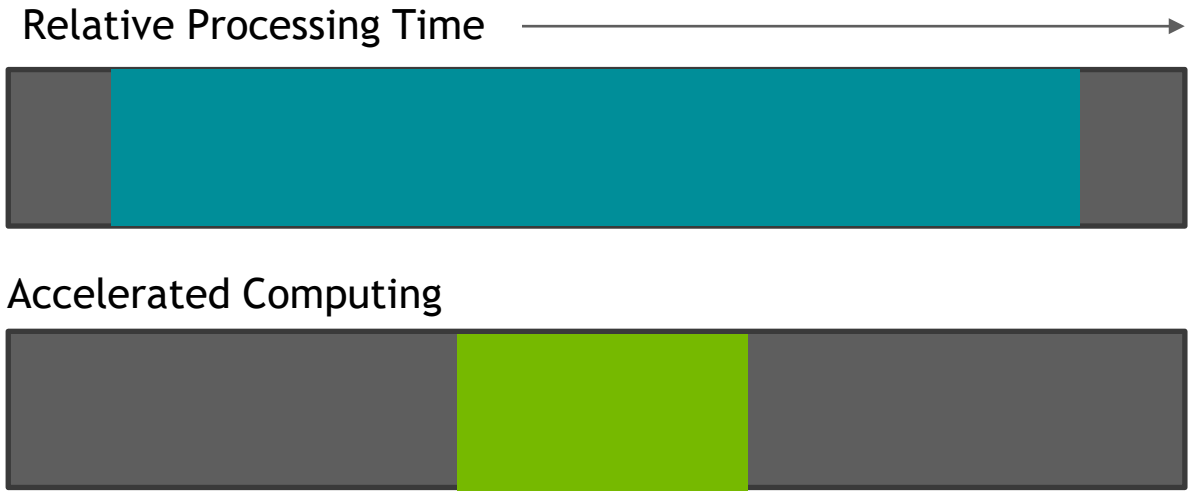


https://www.airmettle.com/

AI Creates a GPU Storage IO Challenge

Big Dataset





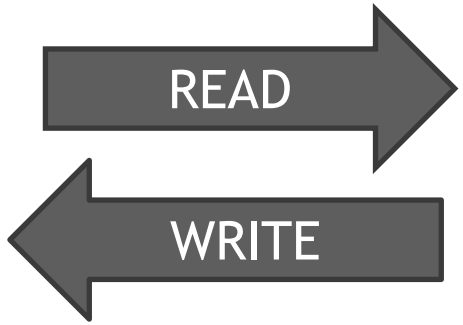
IO

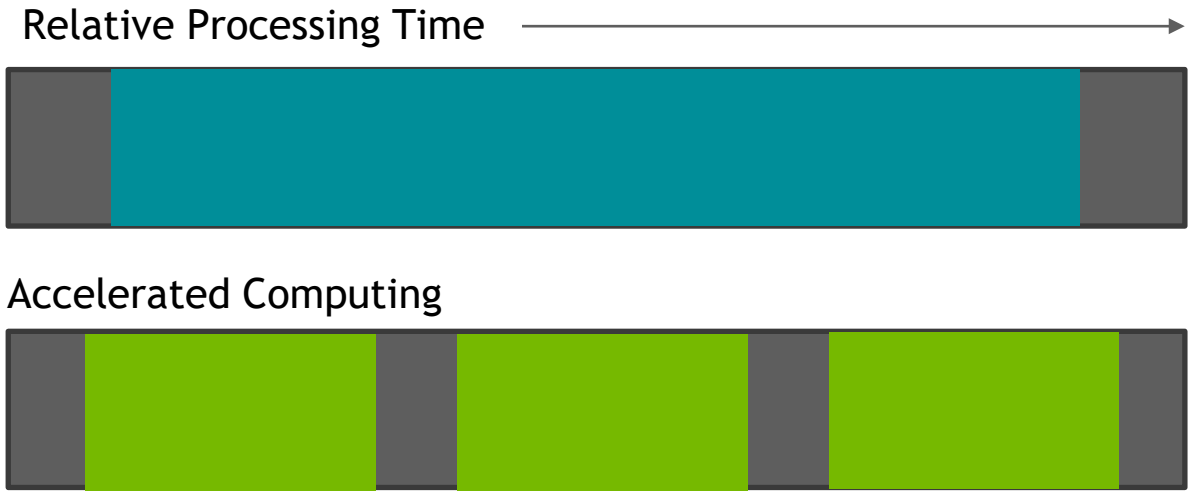
CPU-based compute

GPU-based compute

AI Creates a GPU Storage IO Challenge

Big Dataset



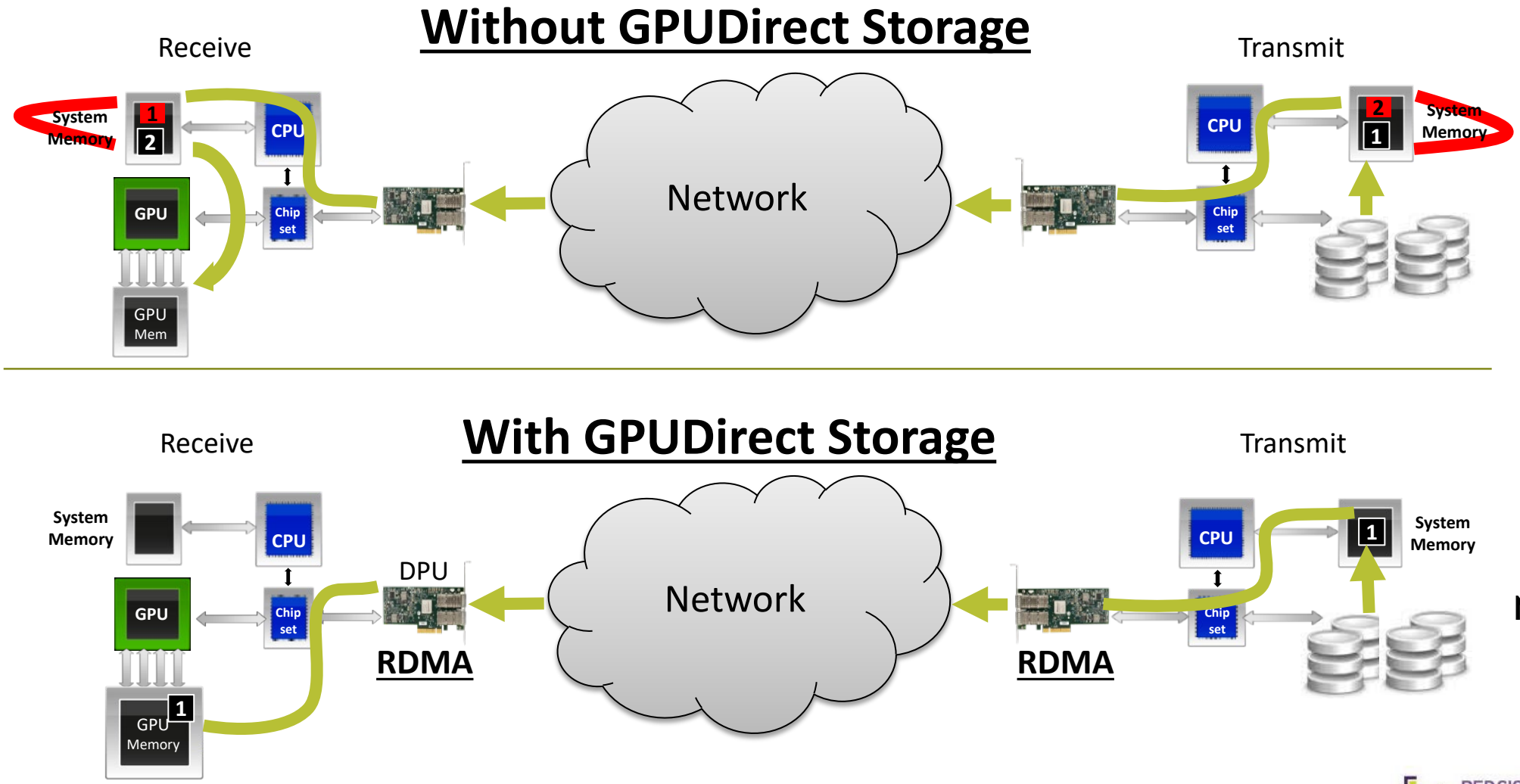


IO

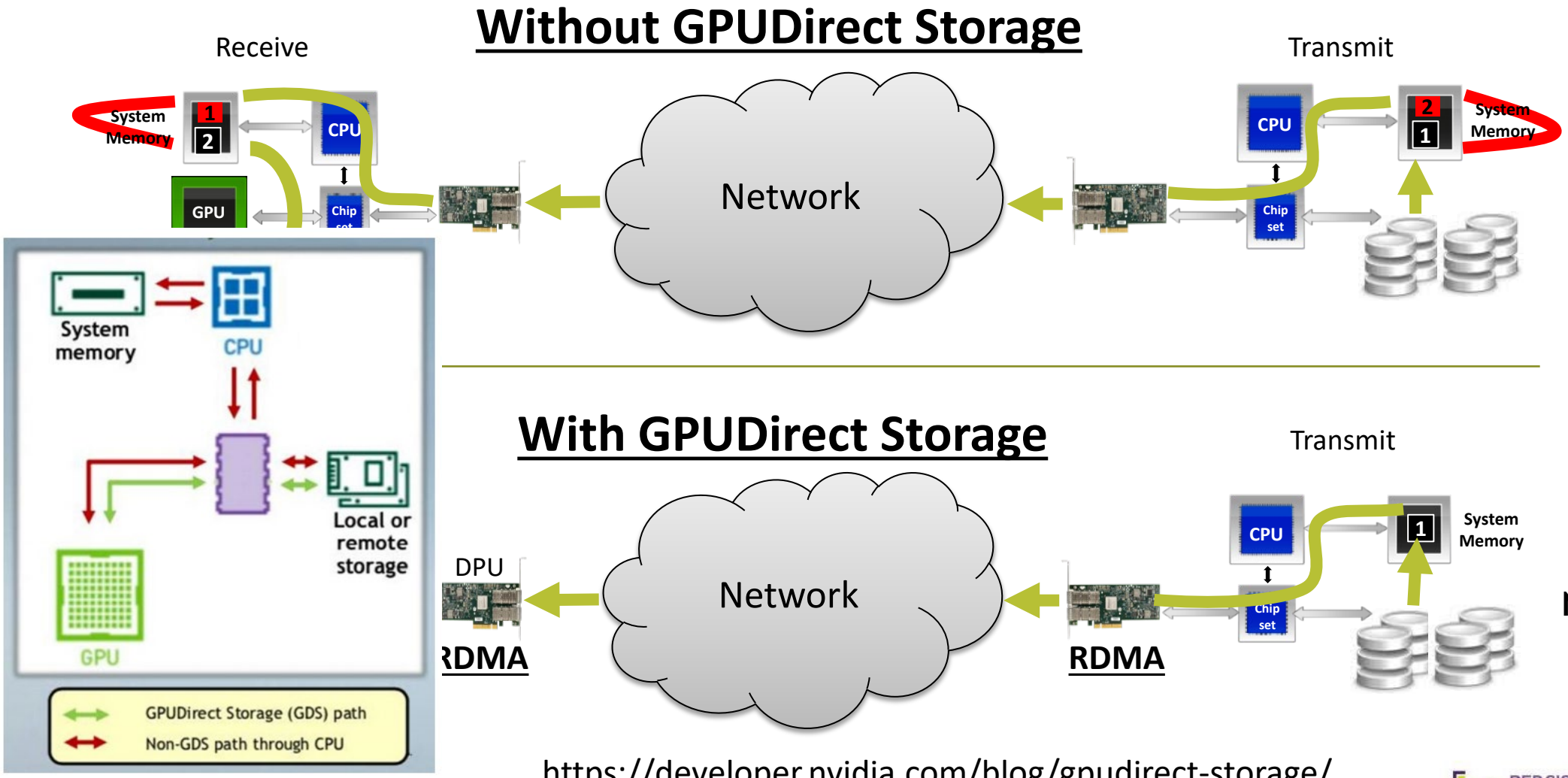
CPU-based compute

GPU-based compute

GPUDIRECT STORAGE(GDS)



GPUDIRECT STORAGE(GDS)

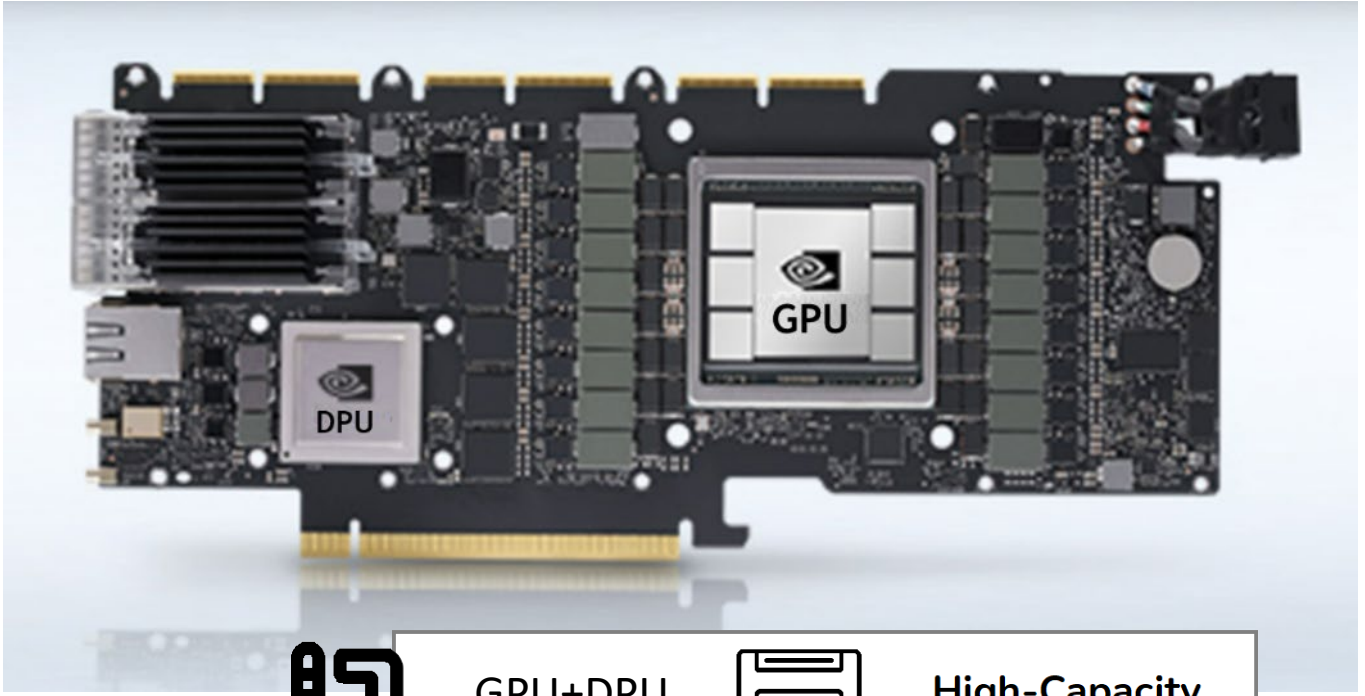


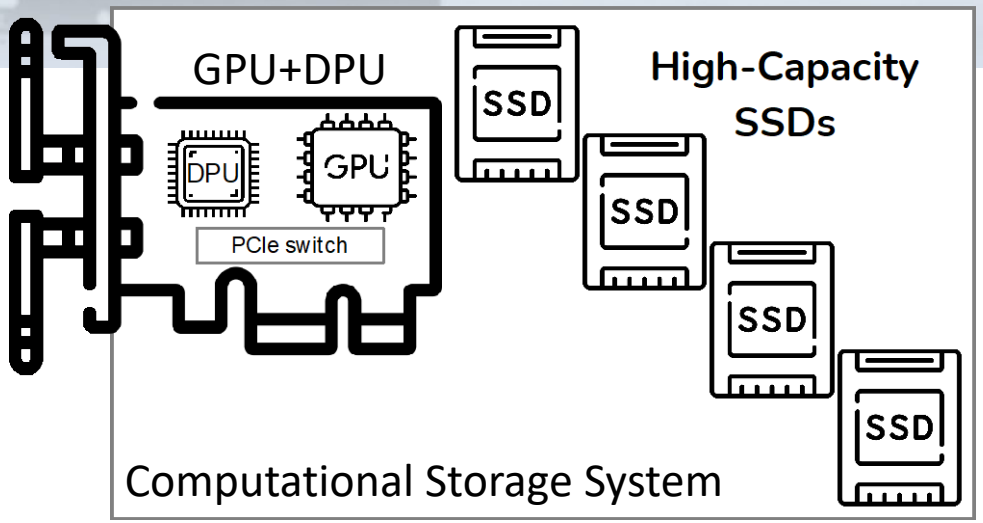
NFSoRDMA
NVMe-oF/RDMA

https://developer.nvidia.com/blog/gpudirect-storage/

Key Takeaways

- Combined GPU and DPU solutions are ideal for Computational Storage AI solutions
- They leverage common components, system architectures, software and expertise
- And enable AI to seamlessly move to the edge





Thank You!

Please take a moment to rate this session.

- Your feedback is important to us.

