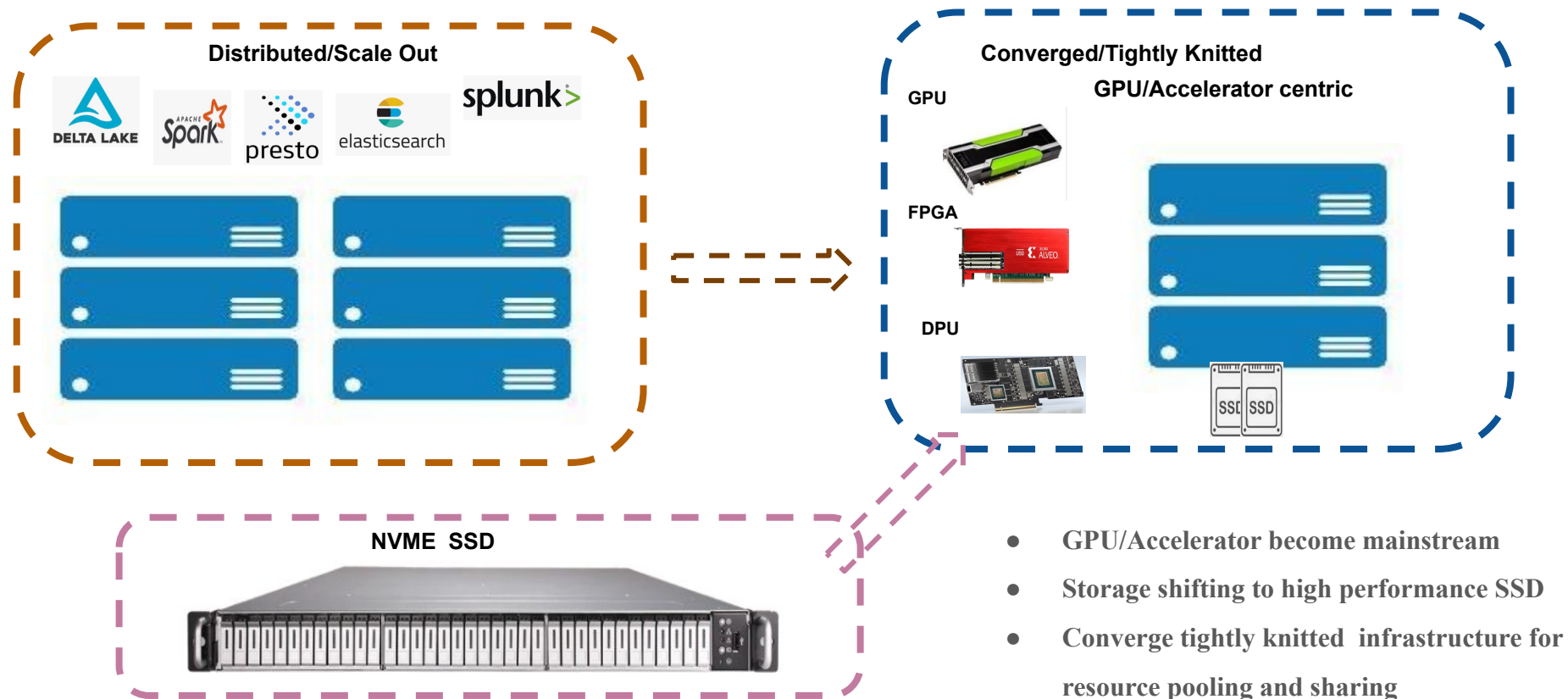


KubeFlash

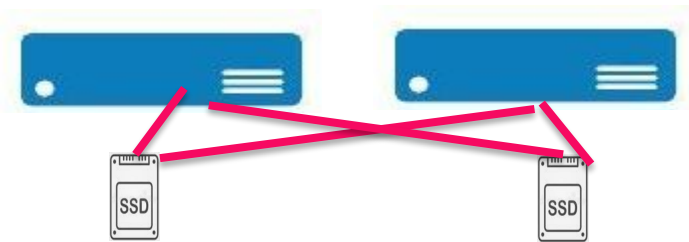
Revolutionary All Flash Technology For HCI and Kubernetes

Computing infrastructure landscape evolving



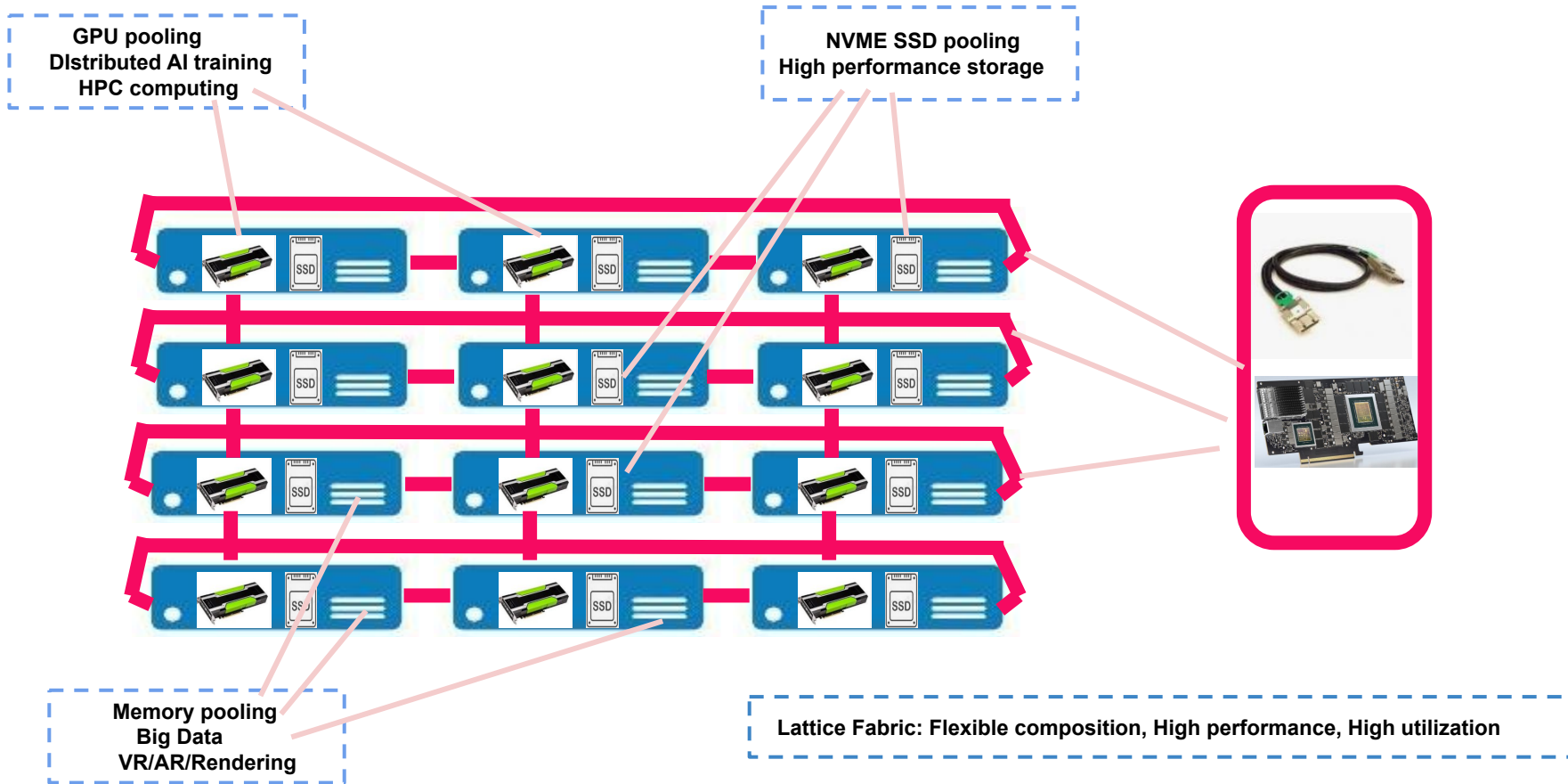
Pain Points in Computing Infrastructure

- All Flash/Enterprise SSD have grown into tens of billions \$ of market, the majority still relies on dual port SSD — Expensive and very few vendors

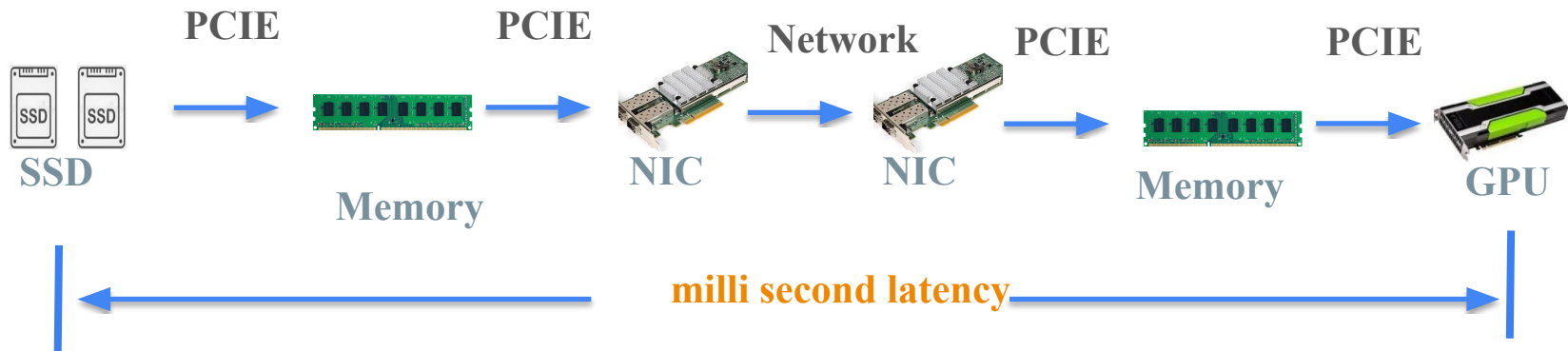


- Kubernetes has become the mainstream computing framework, but even after 20 years development, Kubernetes lack high performance storage solution
- GPU/AI/LLM have high demand for high performance storage, yet current solution fall behind

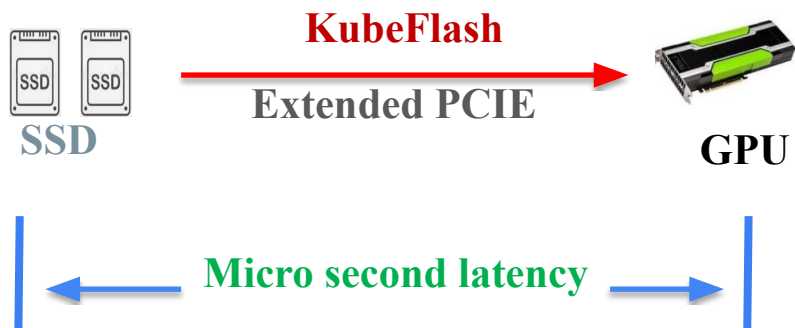
KubeFlash extends PCI Express for server interconnect



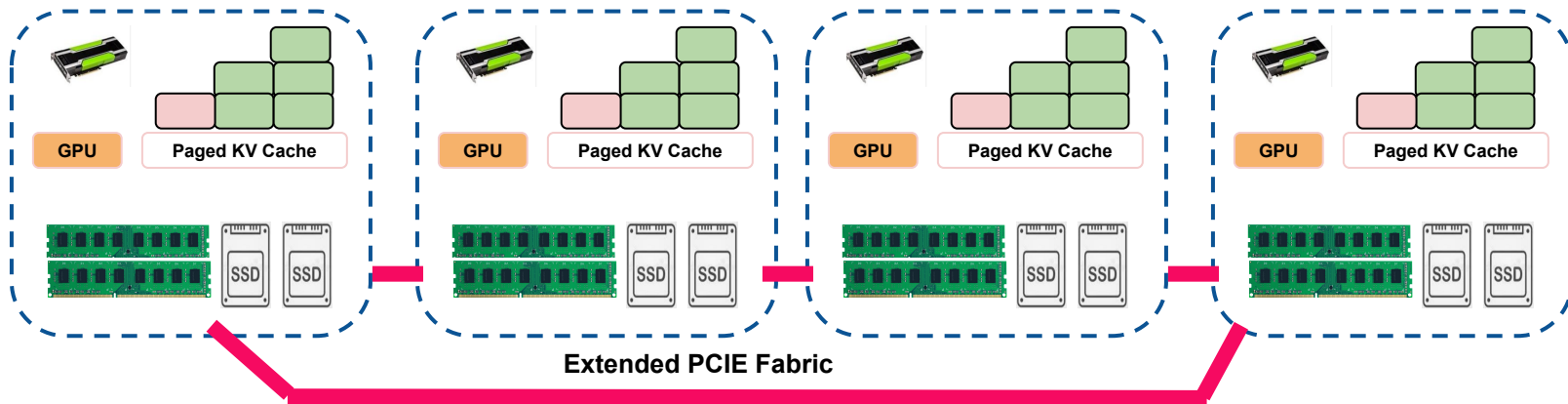
GPU/SSD Data Exchange



KubeFlash Enable

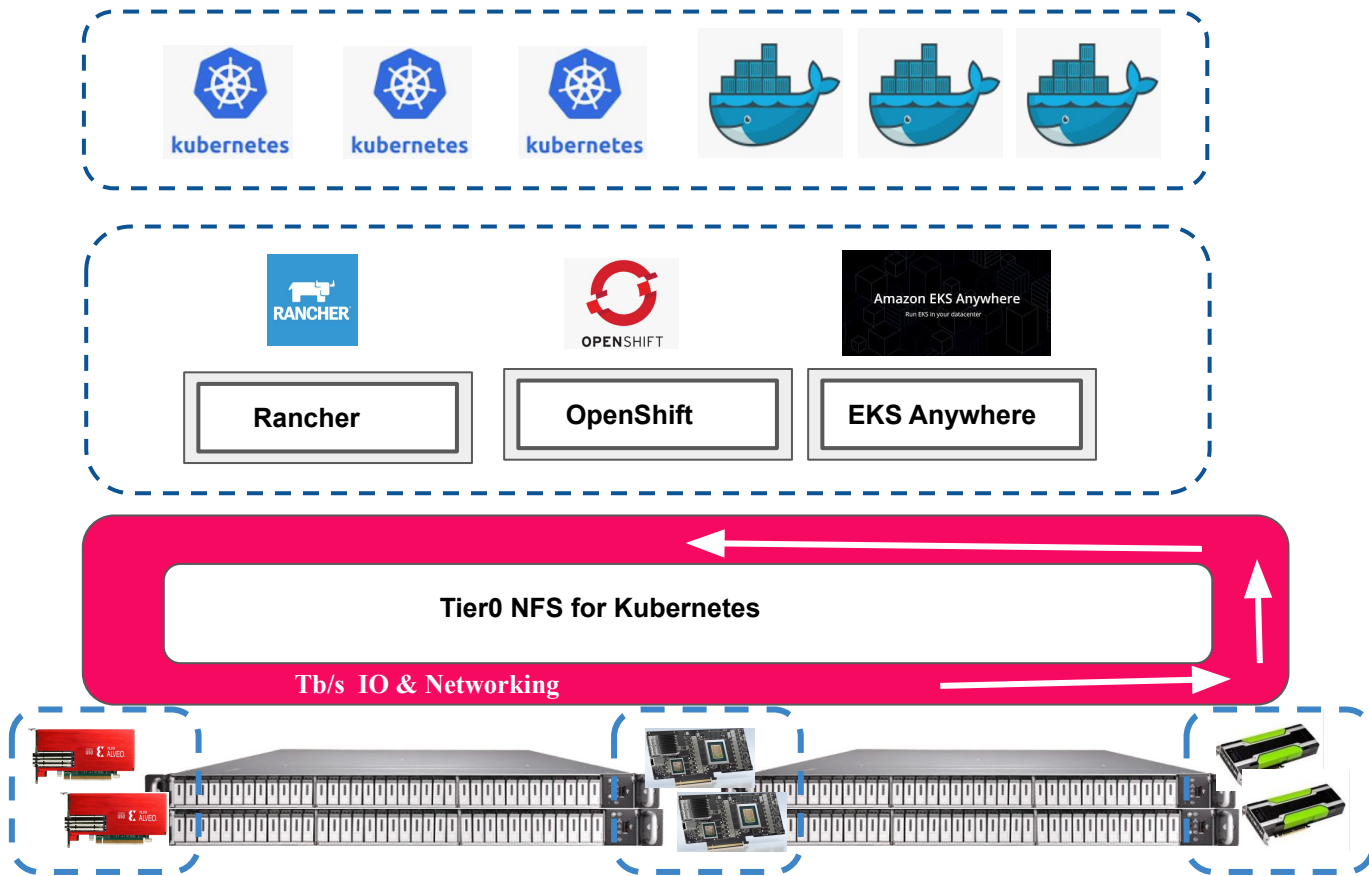


KubeFlash: Tightly coupled GPU/SSD KV Cache for LLM



- Extended PCIE fabric provide pooling for memory and SSD
- **Data movement all in hardware, minimal driver/FW involvement**
- Big improvement in bandwidth and latency compared with traditional RDMA

KubeFlash – Revolutionize Kubernetes Storage



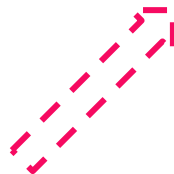
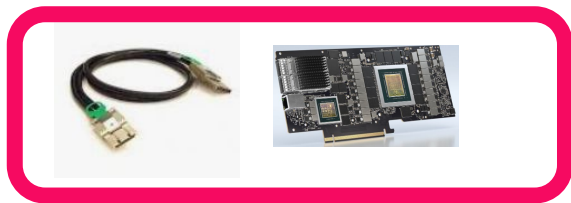
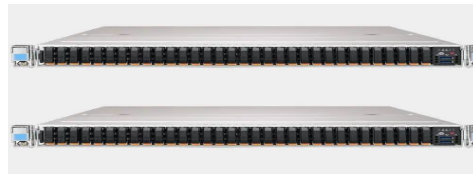
KubeFlash integrating with any platform with ease



Hewlett Packard
Enterprise

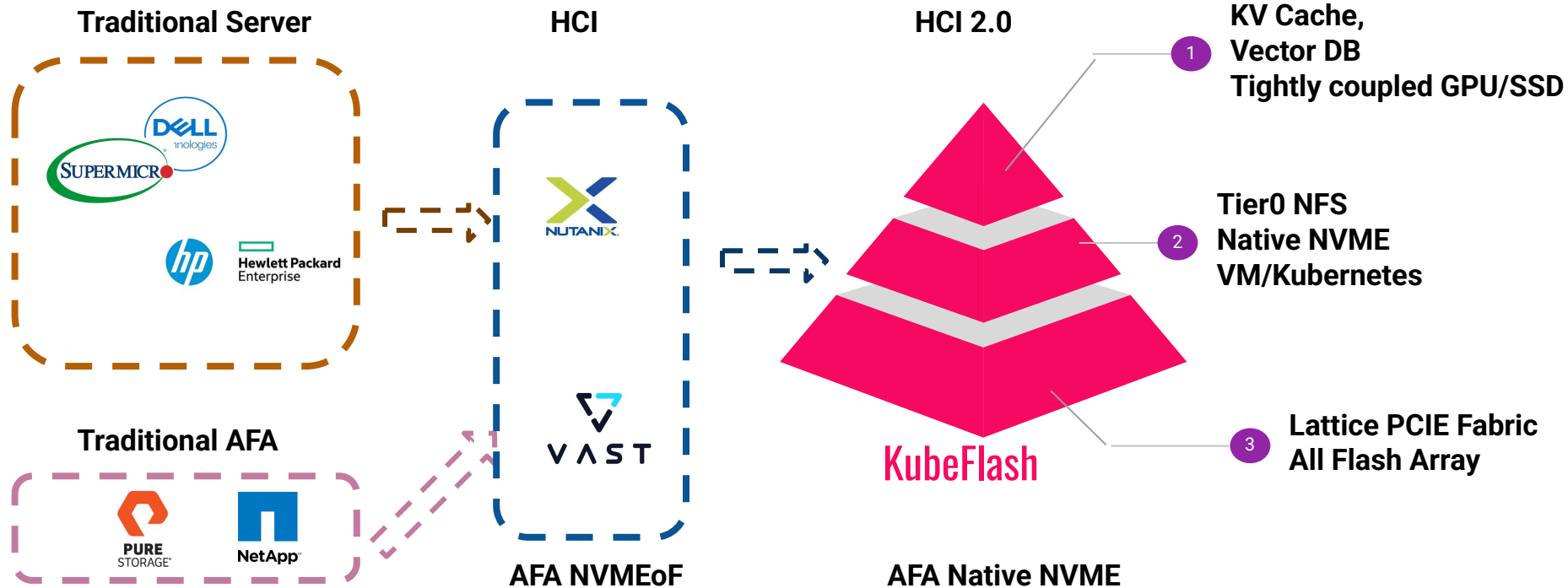


Lenovo



- **Extended PCIE fabric could turn any server platform into high performance cluster**
- All Flash Array built from standard off shelf components
- Converged and native SSD performance across cluster

KubeFlash Computing infrastructure evolving



Market Segments for **KubeFlash**

All Flash Array

HCI

AI & Data Analytics

Market Size

~ 25B

~ 15B

Billions

Major Player



Many Players

KubeFlash
Advantage

- Off-the-Shelf components
- Utilizing server from any vendor
- High performance and low CPU utilization

- Native local NVME performance
- Power efficient
- Revolutionary all flash solution for Kubernetes
-

- Converged SSD/Memory Caching
- Native NVME throughput for KV cache
- Boosting data analytics with native flash storage

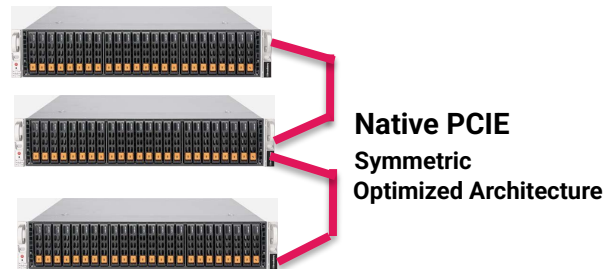
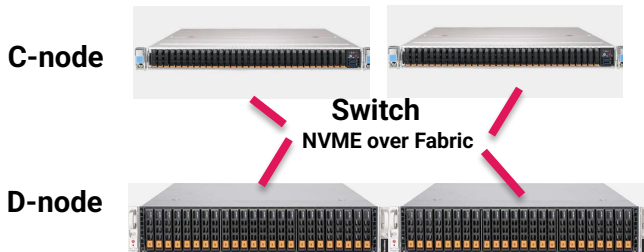
KubeFlash competitor comparison

All Flash Array



KubeFlash

Architecture



Implementation

- NVME over Fabric as internal bus
- Erasure coding across D-node
- Eliminate Dual-port SSD

- Extended PCIE as internal bus
- Erasure coding across symmetric nodes
- Eliminate Dual-port SSD

App Extension

D-node Support big data analytics

HCI mode support any enterprise application

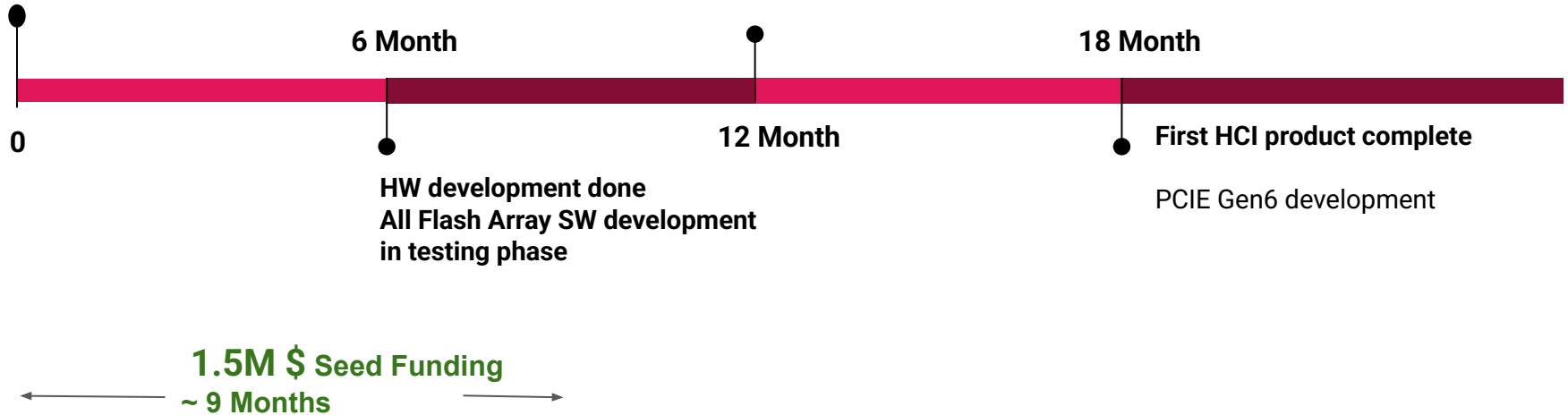
Development Phases

Building Team

Hire ~10 people team
HW/SW development
System integration
Performance tuning

Product sampling

HCI & Kubernetes Integration
complete



Founding Team



- Wei Zhou, worked in Storage industry for over 20 years, have held senior management position at Marvell and SK hynix, leading teams in storage HW/FW/system engineering. Master degree of Applied Physics from Stanford.
- K. Wong, have worked in various fields in storage, have broad experience in SSD FW engineering and linux kernel engineering, have held principle engineer positions at Marvell and Sk Hynix
- X. Chen, 20 years plus experience in storage industry, strong kernel and OS engineering expertise, have held various engineering and management positions at Marvell and Startups