# Computational Storage
# Real-World Deployments

Stephen Bates
CTO, Eideticom

Flash Memory Summit

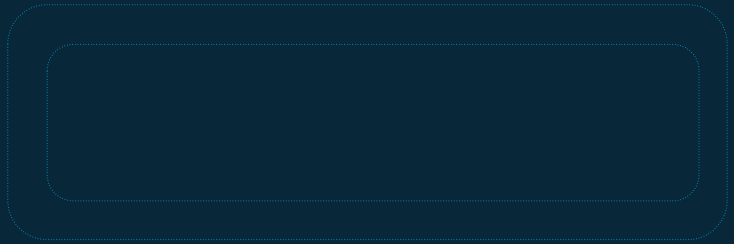
# Computational Storage: State of the Nation

- State of the Nation
- Real-World Deployments
  - NGD Systems
  - Samsung
  - ScaleFlux
  - Eideticom

# Computational Storage: Real World Deployments

- Computational Storage standardization is happening!
  - SNIA Computational Storage: 45+ member companies, 224 members
    - High-level architecture
    - User-space library
  - NVMe Computational Storage: 25+ member companies, 78 members
    - NVMe commands for Computational Storage

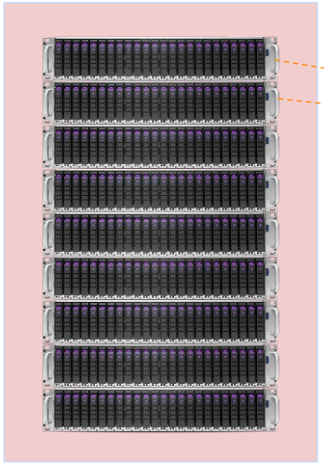
# VMware Demo



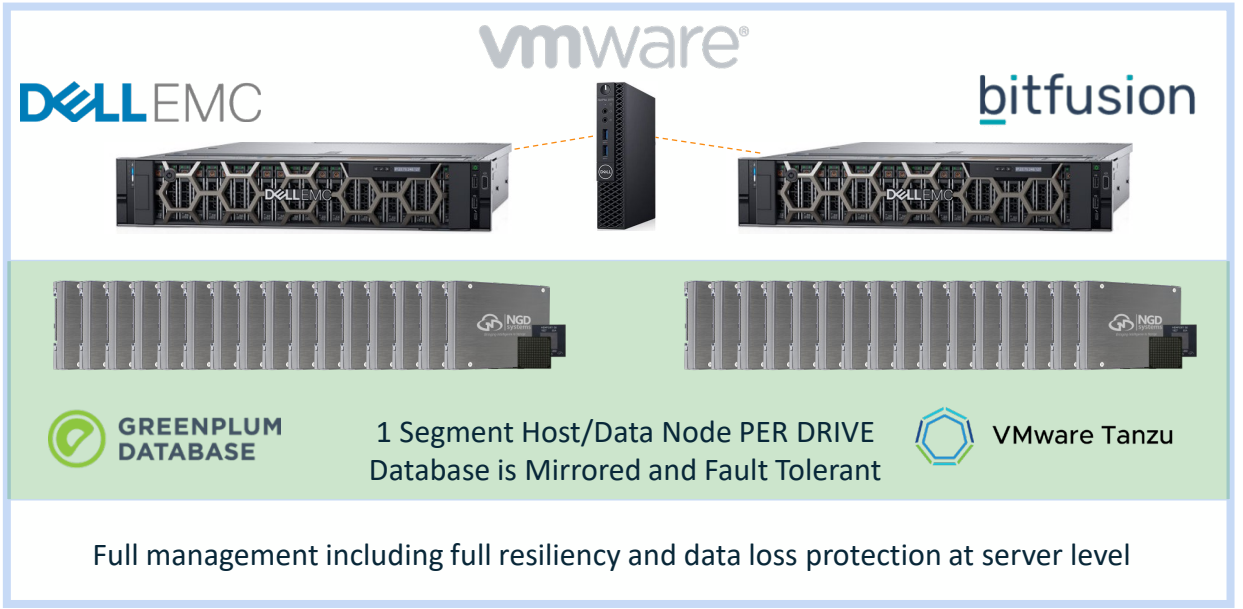
# Edge Analytics –Demo with VMware – xLab Platform

**Computational Storage allows it to be drive level.**

**Reducing footprint, server cost, while still offering full fault tolerance**

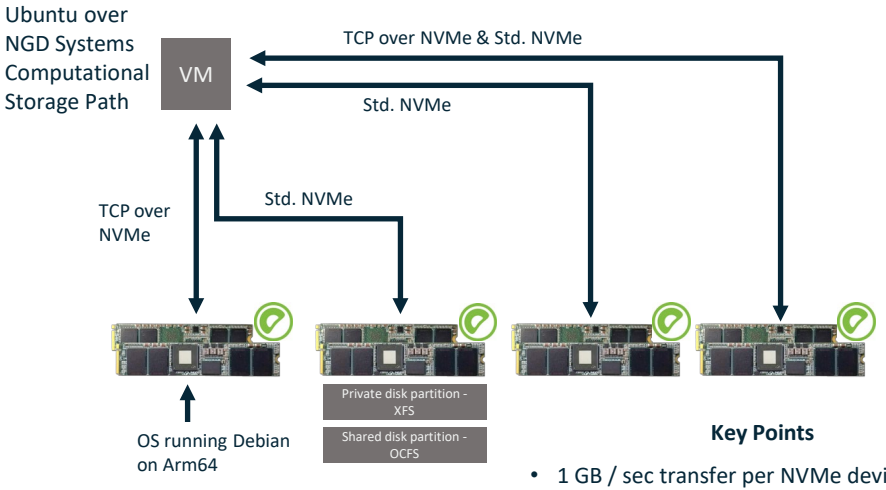



Traditionally Segment Host and Data Nodes were at the **server level**.



# Integration of NGD Systems Devices to vSphere

LINK TO ONLINE DEMO – VMWorld 2020




**Key Points**

- 1 GB / sec transfer per NVMe device
- 16TB capacity per device
- Simultaneous addressing as storage device & as remote compute node
- PCI passthrough allows native use by VMs
- Greenplum running on each node

1. VM Directpath IO for NVMe devices
    1. up to 15 into a VM
2. TCP connected jump box allows addressing of devices from network.
3. Two partitions
    1. one shared w/OCFS
    2. one dedicated to Greenplum

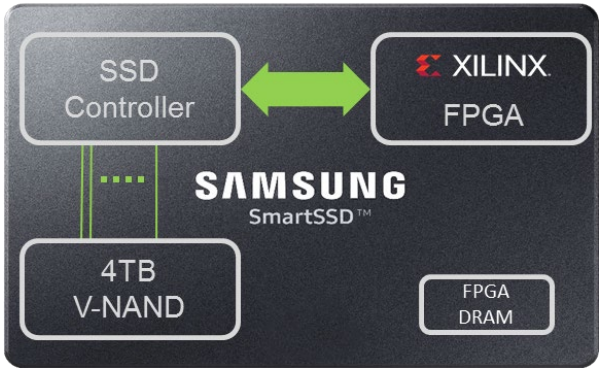
| Computational Storage | Parallel Database with integrated Analytics | vSphere & Bitfusion |
|---|---|---|
| Embed compute with storage, offloading main server, improving performance on smaller systems by reducing data transfer to main system and enabling on-chip intelligence | Query across NVMe devices in parallel, making effective use of computational storage. Embedded analytics allowing analytics free of resources on the main system. Seamless replication of data to backup host. | Ability to offer Edge resiliency with vSAN, HA, FT. GPU acceleration for computational storage w/ Bitfusion. Effective use of limited host resources. |

# Samsung SmartSSD

# SmartSSD® CSD Scales to Accelerate Data-Rich Workloads

Flash Memory Summit

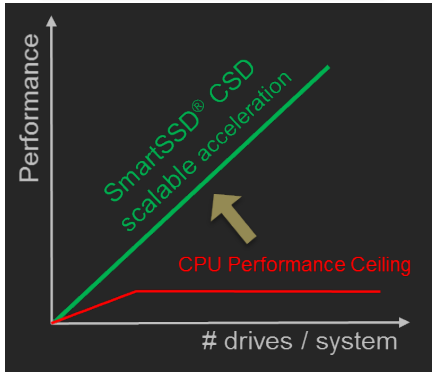
## SmartSSD U.2 Platform




### Computational Storage

- ✓ **3 & 6 GBps internal BW per device:** Minimize external data movement
- ✓ **FPGA:** Each device has 3x~10x core equivalents for offload/acceleration
- ✓ **4TB storage, 4 GB FPGA DRAM:** For Inline and Data@Rest processing

## Acceleration Concept




### Scalable Performance

- ✓ **Near Data Processing:** Data format conversion, Filtering, Metadata management, DB Analytics, Video processing
- ✓ **New Services:** Secure content, Edge acceleration

## Eideticom + SmartSSD CSD

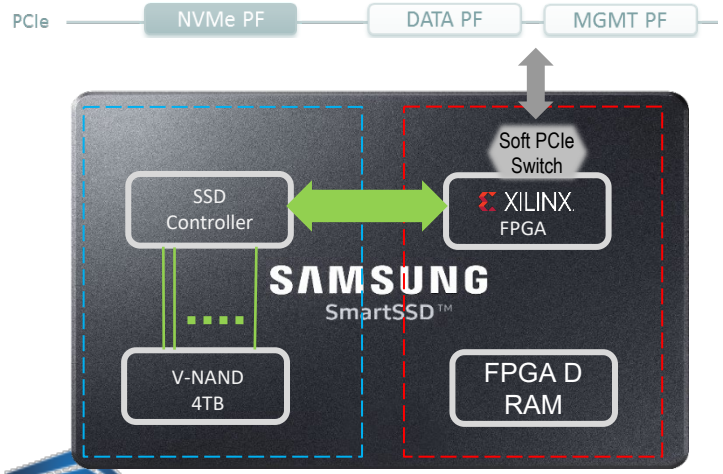

### NVMe Computational Storage

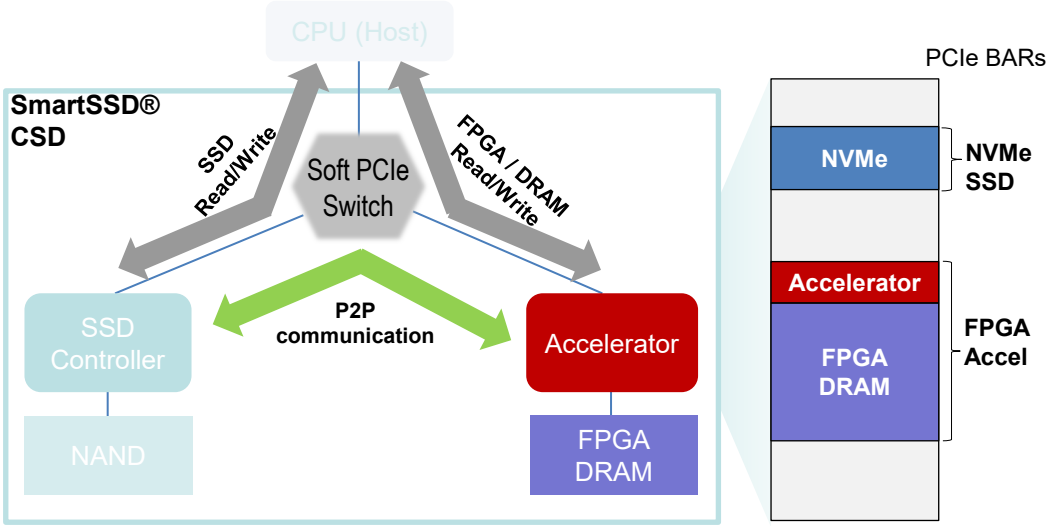
- ✓ **Standards Based:** Uses NVMe to access both the storage and the computation
- ✓ **Consumable:** Leverages the inbox NVMe drivers and software available in all modern OSes

# SmartSSD® CSD HW Architecture

Flash Memory Summit

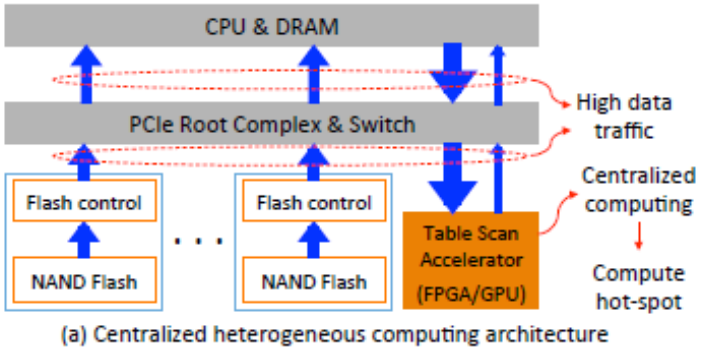
- Peer-to-peer (P2P) communication enables unlimited concurrency
  - SSD:Accelerator data transfers use internal data path
    - Save precious L2:DRAM Bandwidth (Compute Nodes) • Scale without costly x86 frontend (Storage Nodes)
    - Avoid the unnecessary funneling and data movement of standalone accelerators
  - FPGA DRAM is exposed to Host PCIe address space
    - NVMe commands can securely stream data from SSD to FPGA peer-to-peer



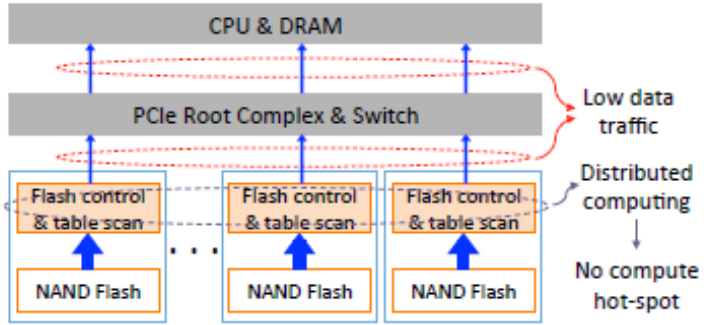



# ScaleFlux

# Data Filtering in the Real World
## Alibaba Cloud & ScaleFlux



(a) Centralized heterogeneous computing architecture



(b) Distributed heterogeneous computing architecture


POLARDB Meets Computational Storage: Efficiently Support Analytical Workloads in Cloud-Native Relational Database

FAST'20 18th USENIX Conference on File and Storage Technologies



POLARDB

**Challenges with baseline architecture:**

1. ***High Data Traffic***
2. ***Data processing hot-spots***

**Solution:**

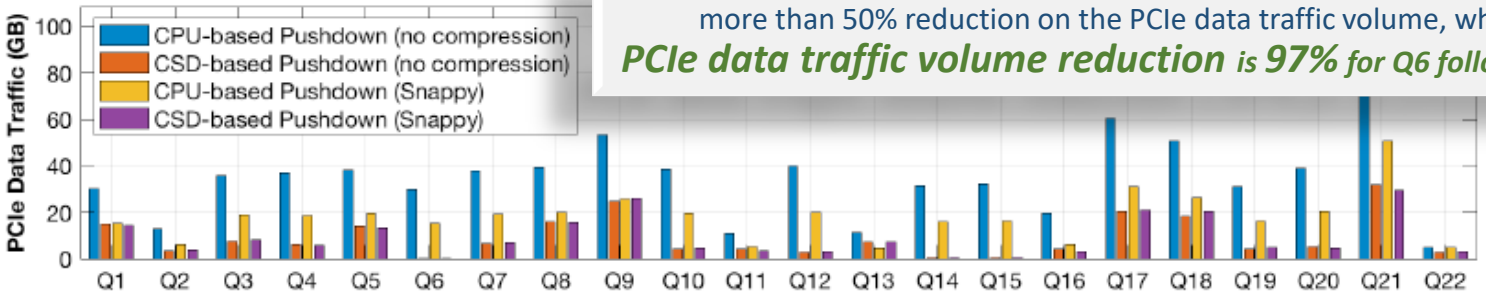
**Distributed heterogenous compute architecture**

**Push table scans to the CSD**

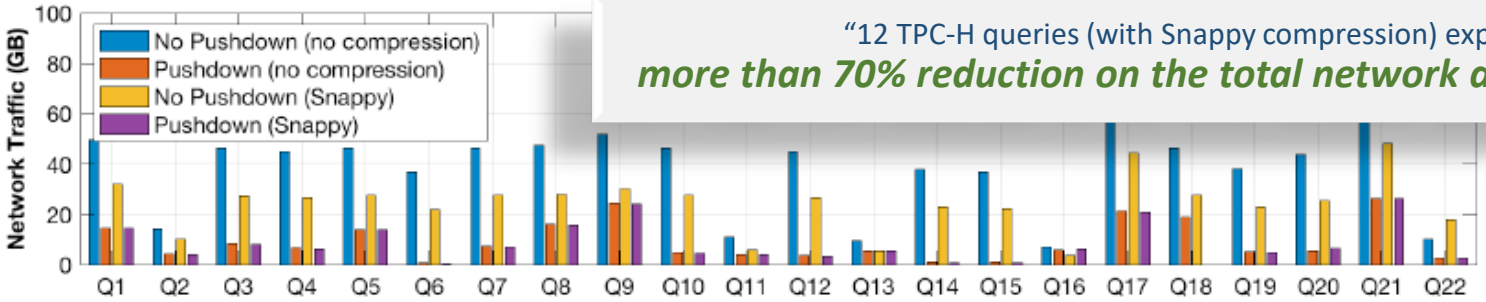
# Data Filtering in the Real World
## Alibaba Cloud & ScaleFlux

FAST'20
18th USENIX Conference on File and Storage Technologies
FEBRUARY 24-27, 2020
SANTA CLARA, CA, USA
Co-located with NSDI'20
Sponsored by USENIX in cooperation with ACM SIGOPS

"Switching from CPU-base push-down to CSD-based pushdown, 7 TPC-H queries experience more than 50% reduction on the PCIe data traffic volume, where the maximum ***PCIe data traffic volume reduction is 97% for Q6 followed by 94% for Q14.***"




"12 TPC-H queries (with Snappy compression) experience ***more than 70% reduction on the total network data traffic*** volume."





POLARDB

Figure 11: (a) PCIe data traffic inside storage nodes and (b) network data traffic in the POLARDB cluster.

# Eideticom

## Eideticom's NoLoad® CSP

Purpose built for acceleration of storage and compute-intensive workloads

### 1) NoLoad Software Stack

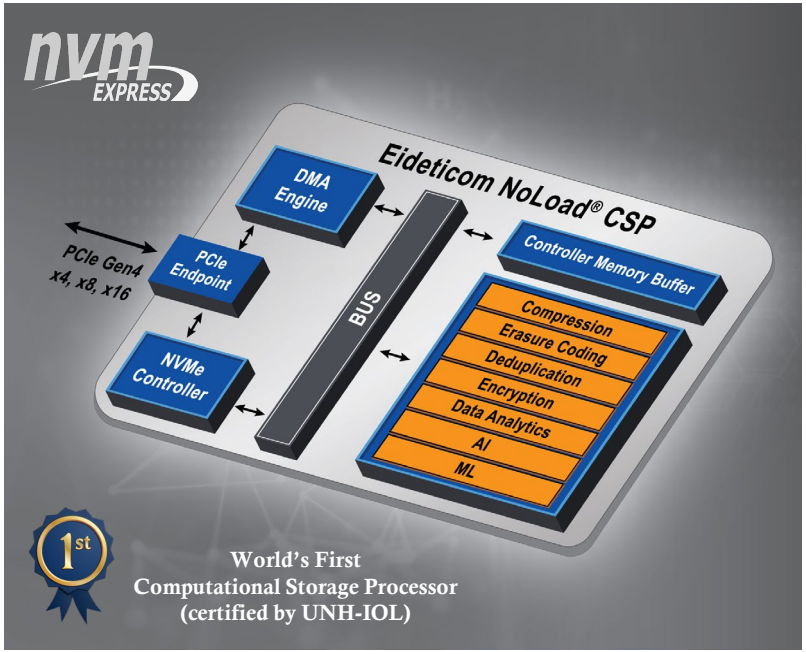
- End-to-end computational storage solution providing transparent computational offload
- Complete Software and IP core stack

### 2) NoLoad NVMe Front End

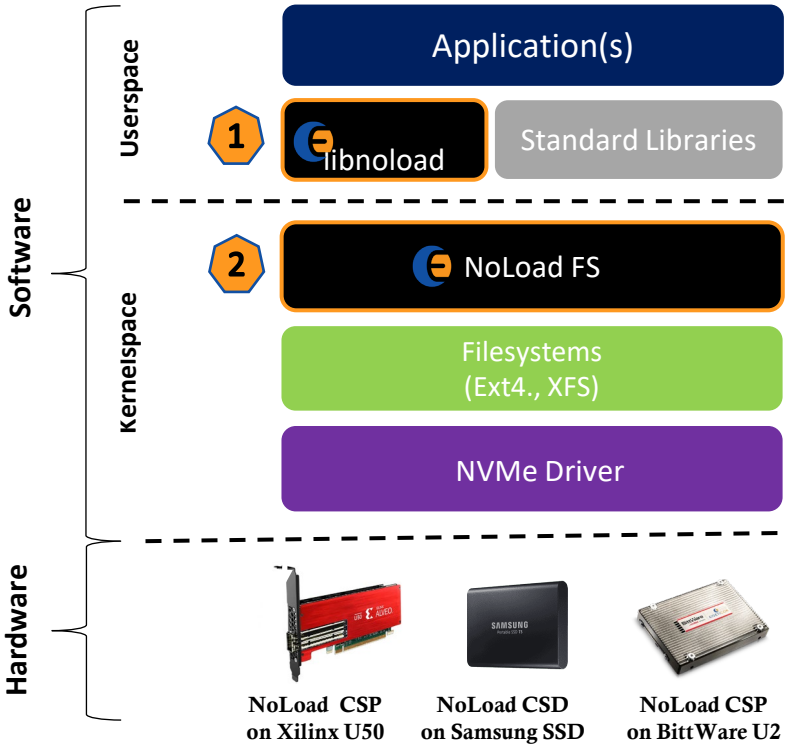
- NVMe compliant, standards-based interface
- High performance interface tuned for computation

### 3) NoLoad Computational Accelerators

- **Storage Accelerators:** Compression, Encryption, Erasure Coding, Deduplication
- **Compute Accelerators:** Data Analytics, Video Codec, AI and ML



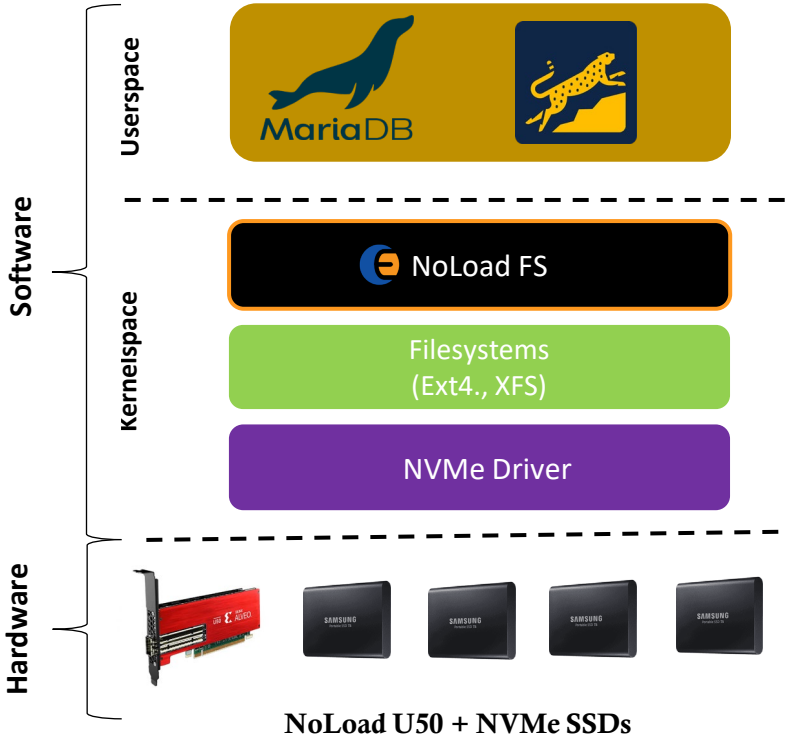
# NoLoad® Software Stack: Transparent Computation



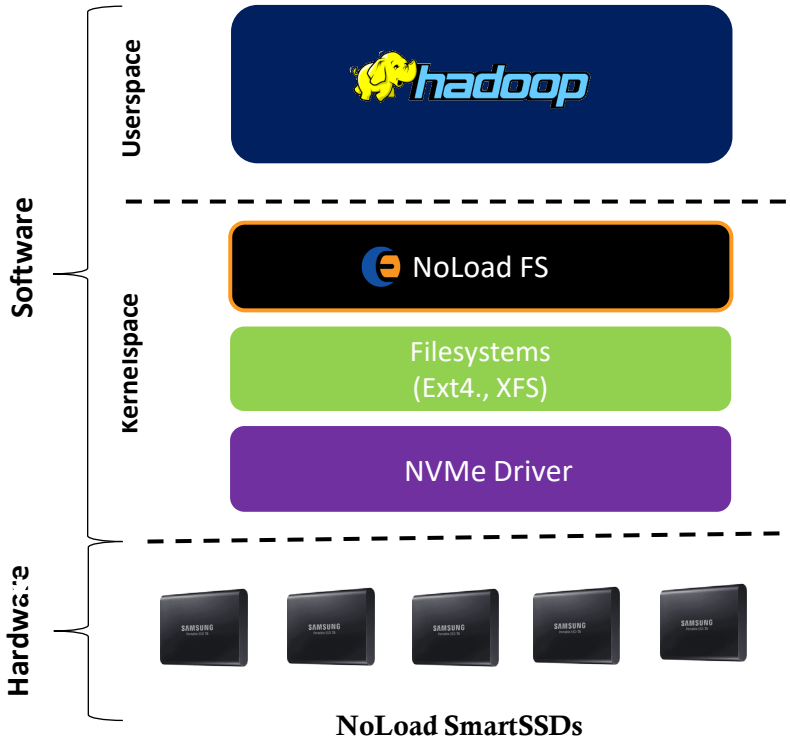

## Eideticom Software Components

1. **libnoload: User-space library**
   - Acceleration w/o Operating System changes
2. **NoLoad FS: Stacked-filesystem**
   - Acceleration w/o Application changes
   - Customer chooses their preferred filesystem

# Hadoop: Transparent Computation with NoLoad FS

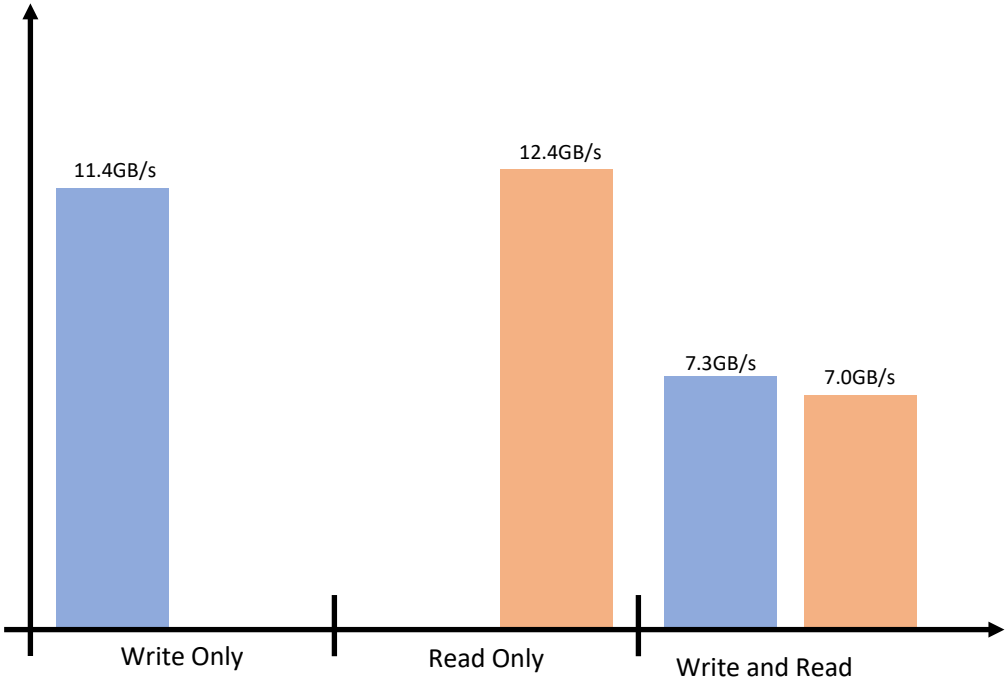


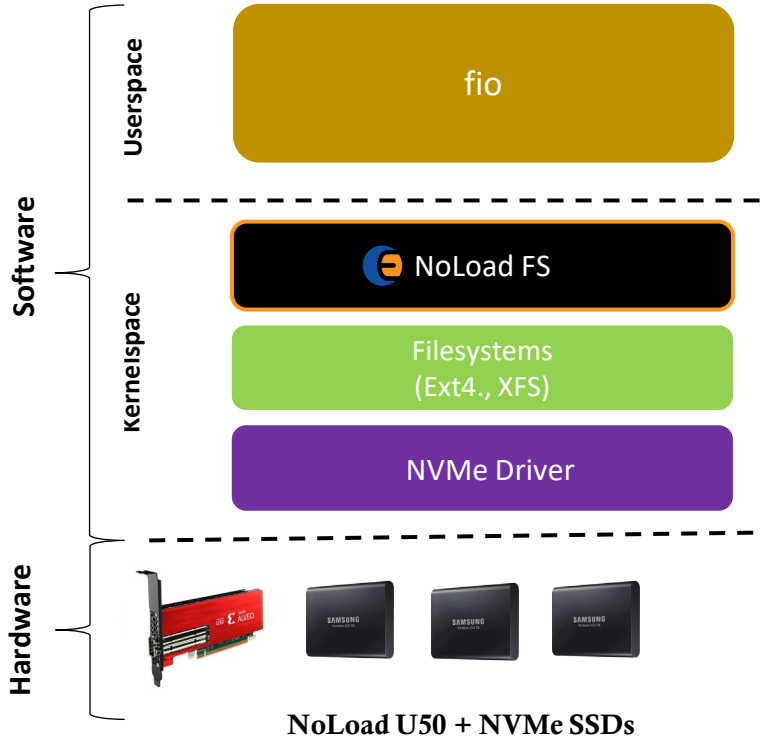

NoLoad U50 + NVMe SSDs




NoLoad SmartSSDs

# Hadoop: Transparent Computation with NoLoad FS




NoLoad U50 + NVMe SSDs

# Computational Storage: Real World Deployments

- Computational Storage product deployment is happening!
- Standardization will increase adoption
  - Vendor-neutral interfaces
  - Open-source software ecosystem
  - NVM Express
- NVM Express market expected to grow at >40% CAGR between 2020 and 2025. Computational storage will be a huge part of that.

# Thank You!

Flash Memory Summit

Everything You Need To Know
For Success