



Flash Memory Summit

# Using Native NVMe-oF SSDs to Advance Computational Storage

Shahar Noy

Product Marketing, Marvell



# Introduction to Native NVMe-oF SSDs

	NVMe-oF SSD	NVMe SSD
<b>Interface</b>	PCIe and Ethernet	PCIe
<b>Performance (Single port)</b>	50Gb/s 100Gb/s	4GB/s (Gen3) 8GB/s (Gen4)
<b>Performance (Dual Port)</b>	2x 25Gb/s 2x 50Gb/s	2x 2GB/s (Gen4) 2x 4GB/s (Gen3)
<b>System Power*</b>	X	1.1X ~ 2X
<b>System Cost*</b>	X	1.2X ~ 3X (higher when factoring system reliability)

\* 2U/24 System; Ethernet Switch; High Availability; 200Gb to 600Gb throughput; vs PCIe based fabric with X86 and NICs. The higher the throughput the greater are the savings



Flash Memory Summit





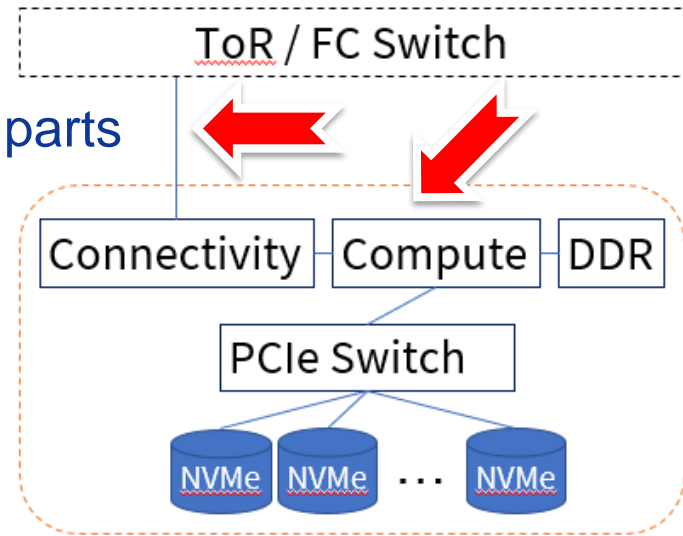
# The Challenges of Storage Controllers

## Centralized System

- One component with non-autonomous parts
- Compute is shared
- Software runs in a single process
- Single point of control
- Single point of failure

## Network

- Oversubscribed





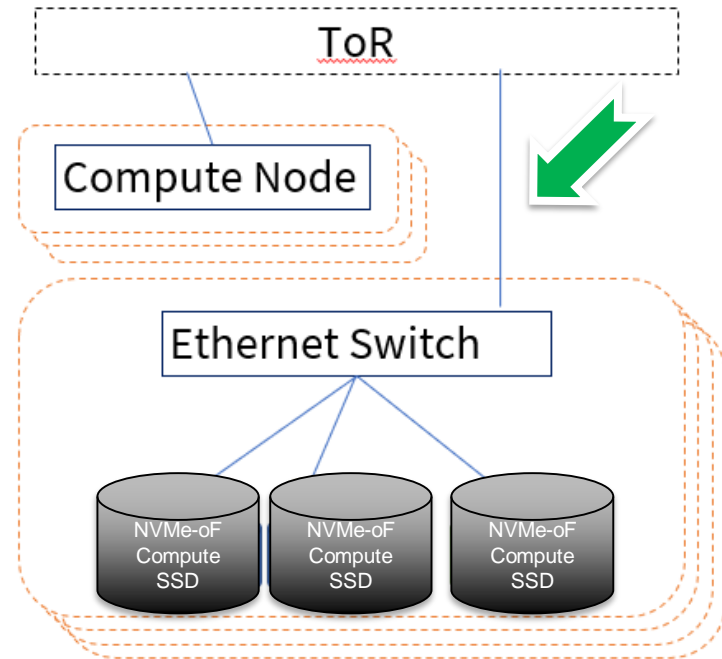
# Computational Storage To The Rescue

## Distributed System

- Multiple autonomous components
- Compute is dedicated
- Software runs in concurrent processes
- Multiple points of control
- Multiple points of failure

## Network

- Subscribed/Oversubscribed





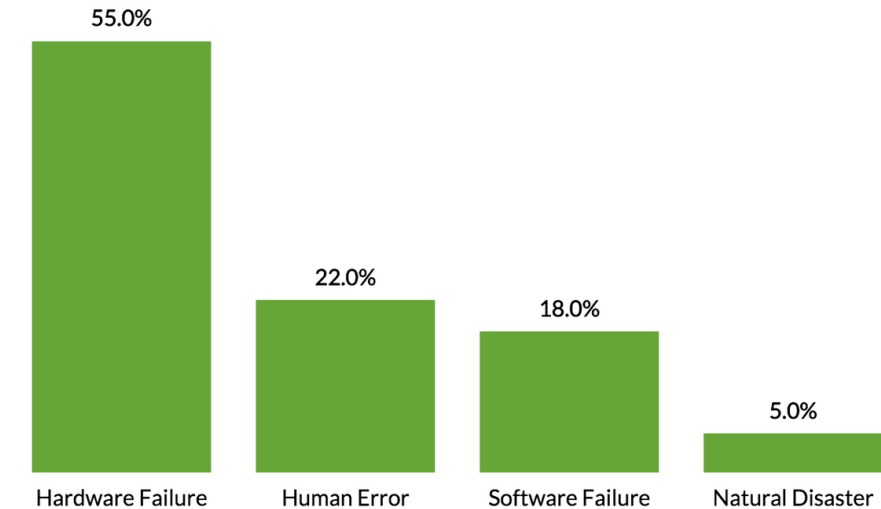
# Distributed Computational NVMe-oF SSDs

## More Efficient Data Center

- Improved Economics
- Full Throughput
- Shared Resources
- Point-to-Point Communication
- Lower Up Front Cost
- High Utilization
- **Better Reliability**

## What Causes Outages?

Percentage estimate of root causes by type



Data source: Quorum via Seagate Report 2017





# A Word on Datacenter Outages

## Statistics

UptimeInstitute®

November 2018

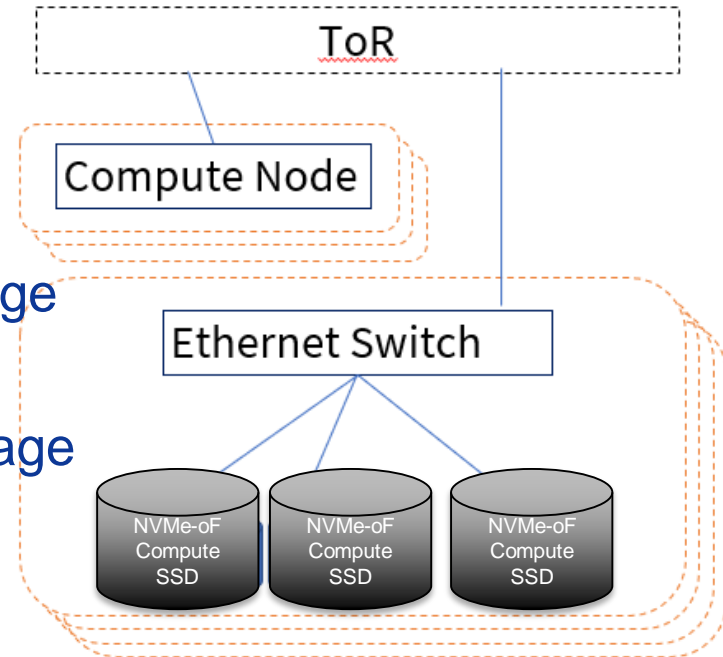
- 33% of Datacenters had at least 1 outage
- 80% of outages were preventable
- HA (2N Architecture) suffered 22% outage

Information Technology Intelligence Consulting

**ITIC**

June 2018

- 1 Hour downtime = \$260,000





# Other Applications

## Additional Benefits of Distributed Computational NVMe-oF SSD

- Object Store: CEPH Daemon
- Database: KV Interface
- Virtualization: Live Migration

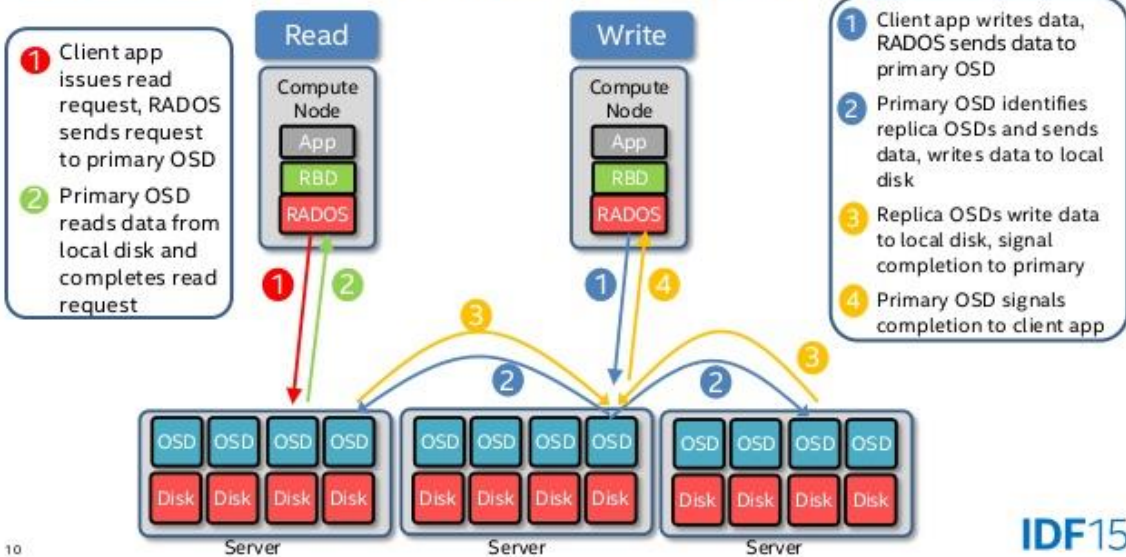




# Object Store: CEPH Daemon



## Object Store Daemon (OSD) Read and Write Flow



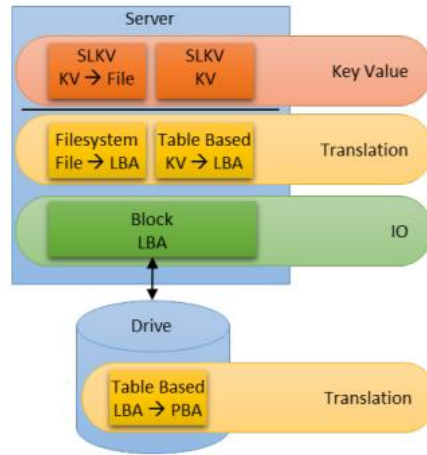
10

IDF15

Source: <https://blog.cypressxt.net/osd-performances-scaling/>

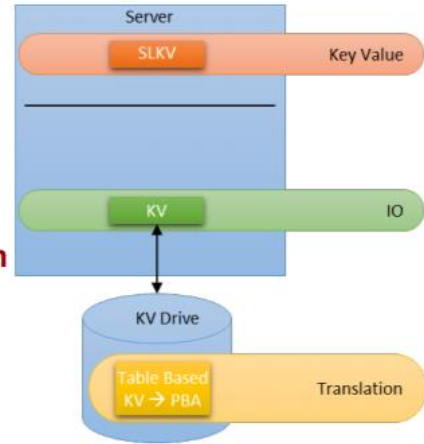


# Database: KV Interface



Previous Drive Protocol Stack

Protocol Optimization



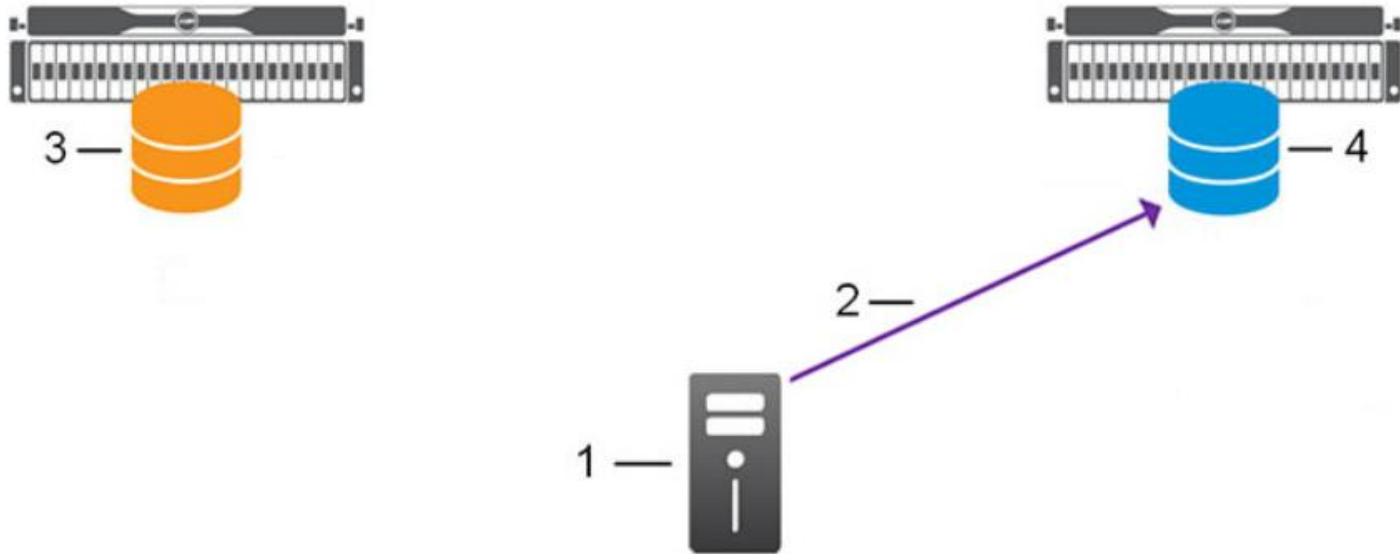
New Drive Protocol Stack

Source: [https://www.samsung.com/semiconductor/global.semi.static/Samsung\\_Key\\_Value\\_SSD\\_enables\\_High\\_Performance\\_Scaling0.pdf](https://www.samsung.com/semiconductor/global.semi.static/Samsung_Key_Value_SSD_enables_High_Performance_Scaling0.pdf)

Source: [https://www.flashmemorysummit.com/English/Collaterals/Proceedings/2018/20180807\\_BMKT-101-1\\_Zha.pdf](https://www.flashmemorysummit.com/English/Collaterals/Proceedings/2018/20180807_BMKT-101-1_Zha.pdf)



# Virtualization: Live Migration





# Summary

## Native Computational NVMe-oF SSD





Flash Memory Summit

**Thank You!**